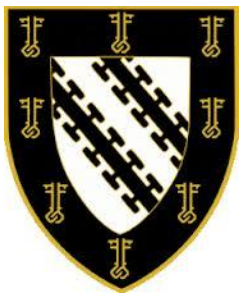
PROTEIN COMPLEXES IN THE GAS PHASE: STRUCTURAL INSIGHTS FROM ION MOBILITY-MASS SPECTROMETRY AND COMPUTATIONAL MODELLING

Zoe Lauren Hall

University of Oxford



Exeter College



This dissertation is submitted for the degree of Doctor of Philosophy, 2013

PROTEIN COMPLEXES IN THE GAS PHASE: STRUCTURAL INSIGHTS FROM ION MOBILITY-MASS SPECTROMETRY AND COMPUTATIONAL MODELLING

Zoe Lauren Hall, Exeter College

This dissertation is submitted for the degree of Doctor of Philosophy, Hilary Term 2013

Structure determination of macromolecular protein assemblies remains a challenge for well-established experimental methods. In this thesis, an emerging structural technique, ion mobility-mass spectrometry (IM-MS) is explored. An assessment of collision cross section (CCS) measurement accuracy using travelling-wave IM (TWIMS) instrumentation was carried out (Chapter 3). Through the collation of a protein complex CCS database and the development of a calibration framework for TWIMS, significant improvements to CCS measurement accuracy have been achieved. Next, the advantages and limitations of using IM-MS to generate restraints for structure characterisation were explored. Computational tools designed to exploit IM-MS data for structural modelling were developed and tested on a training set of systems (Chapter 4). These include two heteromeric protein complexes, and an oligomeric intermediate involved in β2-microglobulin aggregation. Further structural information can be attained by using gas-phase dissociation techniques, such as collision-induced dissociation (CID). The effects of charge state on CCS and the gas-phase dissociation pathway of complexes were investigated (Chapter 5). This highlighted the possibility of using CID in conjunction with supercharging to manipulate dissociation pathways to achieve more useful structural information. Finally, the gas-phase structures of globular and intrinsically disordered protein complexes were probed by IM-MS and molecular dynamics (MD) simulations (Chapter 6). Experimental observations were recapitulated remarkably closely by simulations, including gas-phase structural collapse and the ejection of monomer subunits when the energy of the system was increased sufficiently. Overall, this research has contributed to the IM-MS field by providing the framework for improved CCS measurements of large protein complexes and the use of restraints from IM-MS for structural modelling. Significantly, IM-MS has been used in combination with charge manipulation, CID and MD simulations to reveal further insights into the gas-phase structures, stabilities and dissociation pathways of multimeric protein complexes.

CONTENTS

CONTENTS i

DECLARATION vi

ACKNOWLEDGEMENTS vii

ABBREVIATIONS viii

1 ION MOBILITY-MASS SPECTROMETRY OF PROTEIN COMPLEXES: PRINCIPLES AND TECHNIQUES 1

1.1 INTRODUCTION 1

1.2 MASS SPECTROMETRY 2

1.2.1 Overview 2

1.2.2 Electrospray ionisation 2

1.2.3 Nanoelectrospray ionisation 5

1.2.4 Electrospray mass spectra 7

1.2.5 Ion transmission 10

1.2.6 Quadrupole mass analyser 11

1.2.7 Time-of-flight mass analyser 13

1.2.8 Detectors 15

1.2.9 Tandem mass spectrometry 16

1.3 ION MOBILITY 19

1.3.1 Principles of ion transport 19

1.3.2 Drift tube *versus* travelling-wave ion mobility 22

1.4 THEORETICAL APPROACHES 24

1.4.1 Collision cross section calculations 24

1.4.2 Molecular dynamics 26

1.5 GAS-PHASE PROTEIN STRUCTURE 27

2 MATERIALS AND METHODS 35

2.1 INSTRUMENTATION 35

2.1.1 Absolute collision cross section measurements 35

2.1.2 Travelling-wave ion mobility 38

2.1.3 Charge state distribution parameters 40

2.2 SAMPLE SOURCES AND PREPARATION 41

2.2.1 Peptides, proteins and homomeric complexes 41

2.2.2 Heteromeric complexes 42

2.2.3 ΔN6-β2-microglobulin 43

2.2.4 p53 constructs 44

2.3 REAGENTS 45

2.3.1 Co-factors and chaotropes 45

2.3.2 Charge manipulation 45

2.4 X-RAY CRYSTAL STRUCTURE CALCULATIONS 46

2.4.1 Volume, area and salt bridges 46

2.4.2 Theoretical CCS calculations 47

2.5 COARSE-GRAINED MODELLING 48

2.5.1 CCS and volume calculations 48

2.5.2 Coarse-grained modelling approach 48

2.5.3 Ultrafast shape recognition 49

2.5.4 Hierarchical clustering 50

2.6 MOLECULAR DYNAMICS SIMULATIONS 51

2.6.1 Simulation method 51

2.6.2 Charge assignment algorithm 52

2.6.3 RMSD clustering 53

3 PROTEIN COMPLEXES AS ION MOBILITY STANDARDS: A CALIBRATION FRAMEWORK AND DATABASE FOR GAS-PHASE STRUCTURAL BIOLOGY 55

3.1 INTRODUCTION 55

3.2 RESULTS 57

3.2.1 Characterising the linear drift tube 57

3.2.2 Collision cross section database 60

3.2.3 Effects of drift gas selection 64

3.2.4 Interpreting CCS: Relationship between experimental and calculated values 67

3.2.5 Calibration strategies for travelling-wave ion mobility 70

3.2.6 Minimising ion heating in TWIMS 73

3.2.7 Wave velocity and height affect resolution and CCS accuracy 77

3.3 DISCUSSION 82

3.4 SUMMARY 85

4 DEVELOPING COMPUTATIONAL STRATEGIES FOR STRUCTURAL MODELLING OF PROTEIN ASSEMBLIES 87

4.1 INTRODUCTION 87

4.1.1 Integrative structure characterisation of protein assemblies 87

4.1.2 Structural modelling from ion mobility-mass spectrometry data 90

4.2 DEVELOPING A MODELLING APPROACH 93

4.2.1 Overview 93

4.2.2 Representation 95

4.2.3 Spatial restraints: overlap and volume 96

4.2.4 The CCS restraint 100

4.2.5 Symmetry restraint 101

4.3 RESULTS 103

4.3.1 IM-MS of subcomplexes 103

4.3.2 Evaluating CCS and volume restraints: Benchmarking on a homomeric complex 111

4.3.3 Sensitivity and specificity 114

4.3.4 Application to heteromeric complexes 115

4.3.5 Tryptophan synthase 117

4.3.6 Nitrobenzene dioxygenase 120

4.3.7 Predicting topology of ΔN6-β2m oligomers 124

4.4 DISCUSSION 130

4.5 SUMMARY 133

5 EXPLORING THE EFFECTS OF CHARGE MANIPULATION ON CONFORMATION AND DISSOCIATION OF PROTEIN COMPLEXES 135

5.1 INTRODUCTION 135

5.2 RESULTS 140

5.2.1 Relationship between CCS, charge and surface area 140

5.2.2 Supercharging and charge reduction: Effects on conformation 142

5.2.3 Supercharging and unfolding in solution 148

5.2.4 Supercharged species can access atypical dissociation pathways 151

5.2.5 Gas-phase *versus* solution-phase disassembly 158

5.2.6 Dissociation of heteromeric tryptophan synthase 163

5.2.7 The role of salt bridges 169

5.3 DISCUSSION 174

5.4 SUMMARY 175

6 PROBING GAS-PHASE PROTEIN COMPLEX STRUCTURES BY ION MOBILITY AND MOLECULAR DYNAMICS 177

6.1 INTRODUCTION 177

6.2 RESULTS 180

6.2.1 Ion mobility experiments reveal collision-induced compaction 180

6.2.2 Molecular dynamics simulations corroborate experimental results 188

6.2.3 Effect of charge state on compaction and unfolding *in silico* 191

6.2.4 Insights into collision-induced dissociation 193

6.2.5 Intrinsically disordered p53 collapses in the gas phase 197

6.3 DISCUSSION 206

6.4 SUMMARY 211

7 CONCLUSIONS AND FUTURE DIRECTIONS 213

REFERENCES 219

APPENDIX 243

DECLARATION

This thesis is a summary of research carried out in the Department of Chemistry, University of Oxford, between October 2009 and April 2013. The work described in this thesis is wholly my own and not the result of collaboration, unless specifically stated in the text. It has not as a part, or as a whole, been submitted for a degree, diploma or other qualification at any other university. The length of this thesis does not exceed the word limit.

Zoe Lauren Hall

Oxford, April 2013

ACKNOWLEDGEMENTS

Over the past three and half years, I have received help and support from a number of people to whom I owe thanks. First and foremost, to my supervisor Professor Carol Robinson, who gave me the invaluable opportunity to work in her research group, and provided guidance during my PhD. I have had the pleasure of working alongside a number of people, both within the Robinson and Benesch groups, and external collaborators. In particular I am grateful to Matt Bush, Justin Benesch, Helena Hernández, Kevin Pagel, Carla Schmidt, Todd Mize, Joseph Marsh (University of Cambridge), Eviatar Natan (University of Cambridge) and Kasia Domanska (Vrije Universiteit Brussel), who have provided training, technical help, inspiring discussions, or samples. I would especially like to thank Argyris Politis, with whom I worked alongside on the coarse-grained modelling project, for the many hours spent helping me to write or fix scripts. Similarly, the advice and encouragement received from Mark Thachuk and others, and the excellent support of the Oxford Supercomputer help team was essential whilst learning about molecular dynamics simulations.

Oxford has been a truly great place to live and study, and the warm social environment of the Exeter College MCR has played a big part in this. Fir tree, Turf and KA sessions with Carlos Callejo Delgado, Sonny Dhanda and Henrik Indergaard; college dinners and bops with Giulia Macaro, Andreas Schiffer and Georgios Anastasiades; plus a brief (and painful) exposure to rowing, kept me balanced. As did the support from friends, family, flatmates and colleagues, particularly Shanna Isaacson, Ima-obong Ebong, Gill Hilton, Carla Schmidt, Florian Stengel, Julien Marcoux and Argyris Politis. I am also thankful to those who helped by reading and commenting on the various chapters in my thesis: Toni Borysik, Kevin Pagel, Rune Salbo, Jonathan Hopper, Julien Marcoux, Carla Schmidt, and my sister Alice Hall, whose experience as a journal editor came in handy! Particular thanks must be given to Argyris Politis who took the time to go through the whole thesis.

Finally, at the end of this journey: I look back and am thankful for my education in Zimbabwe, South Africa, London and Oxford. For my introduction to mass spectrometry at LGC, with excellent mentors Gavin O'Connor and Chris Hopley. For friends lost and gained. And now, I look forward to rejoining the world!

ABBREVIATIONS

3D	three-dimensional
AA	ammonium acetate
ACN	acetonitrile
ADH	alcohol dehydrogenase
AMP-PNP	adenosine 5′-(β,γ-imido)triphosphate
Ar	argon
Arg	arginine
Asp	aspartic acid
ATD	arrival time distribution
(ΔN6)–β2m	(ΔN6)–beta-2-microglobulin
CCS	collision cross section
CID	collision-induced dissociation
CSD	charge state distribution
Da	Dalton
DBU	1, 8-diazabicycloundec-7-ene
DC	direct current
DMA	differential mobility analysis
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
DTE	dithioerythritol
DTT	dithiothreitol
DRA	dialysis-related amyloidosis
ECD	electron capture dissociation
E.coli	*Escherichia coli*
EDTA	ethylenediaminetetraacetic acid
EHSS	exact hard spheres scattering
EM	electron microscopy
ESI	electrospray ionisation
FWHM	full width at half maximum
Glu	glutamic acid
HDX	hydrogen-deuterium exchange
He	helium
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
His	histidine
IM	ion mobility
IMP	integrative modelling platform
KCl	potassium chloride

Lys	lysine
m-NBA	*meta*-nitrobenzyl alcohol
MCP	microchannel plate
MD	molecular dynamics
MeOH	methanol
MMOH	methane monooxygenase hydroxylase
Mn	manganese
MOPS	4-morpholinepropanesulfonic acid
MS	mass spectrometry
MS/MS	tandem mass spectrometry
m/z	mass to charge ratio
N_2	nitrogen
Na	sodium chloride
NBDO	nitrobenzene dioxygenase
NECD	native electron capture dissociation
nESI	nanoelectrospray ionisation
NMR	nuclear magnetic resonance spectroscopy
PA	projection approximation
PDB	protein data bank
Q-ToF	quadrupole/time-of-flight
RF	radiofrequency
R_g	radius of gyration
RMSD	root mean square deviation
RNA	ribonucleic acid
ROC	receiver operating characteristics
SAP	serum amyloid P
SASA	solvent accessible surface area
SAXS	small-angle X-ray scattering
SID	surface-induced dissociation
SRIG	stacked ring ion guide
TEA	triethylammonium acetate
ToF	time-of-flight
ToMOH	toluene/*o*-xylene monooxygenase hydroxylase
TRAP	*trp* RNA binding protein
TTR	transthyretin
TWIG	travelling-wave ion guide
TWIMS	travelling-wave ion mobility spectrometry
USR	ultrafast shape recognition

1 ION MOBILITY-MASS SPECTROMETRY OF PROTEIN COMPLEXES: PRINCIPLES AND TECHNIQUES

1.1 INTRODUCTION

Mass spectrometry (MS) and ion mobility (IM) are widely used for small molecule analysis and proteomics, however instrumental and methodological advances have increased the range of analytes amenable to analysis (*1*). This includes large protein assemblies such as ribosomes (*2*) and Mega Dalton virus particles (*3*). Over the last decade, MS and IM-MS have played an increasingly important role in structural biology, complementing traditional biophysical methods such as X-ray crystallography, nuclear magnetic resonance spectroscopy (NMR) and electron microscopy (EM). MS has an advantage over these techniques as it is suitable to probe the structures and dynamics of heterogeneous and low abundance protein assemblies (*4*). As such, by studying the intact protein by MS, the stoichiometry of the complex can be determined; connectivity of individual subunits is mapped by MS of solution-disrupted complex (*5*-*6*). This information can be used in conjunction with restraints derived from IM, enabling the construction of three-dimensional (3D) topological models. In this chapter, the principles behind these techniques and their application to the study of protein complexes are introduced, and evidence for folded protein structure in the gas phase is discussed.

1.2 MASS SPECTROMETRY

1.2.1 Overview

MS is a gas phase separation technique. Ions are produced in the source region, transmitted into the high vacuum region and subsequently separated according to their mass to charge ratios (*m/z*) in the mass analyser. The ions are finally detected, and recorded in a mass spectrum. Protein complexes are typically analysed directly from solution. However, the solution conditions required for storage are often incompatible with MS, due to the presence of non-volatile salts. This can result in ion suppression and/or adduct formation. In order to overcome this samples are transferred to buffered aqueous solutions, most commonly ammonium acetate (AA), at or around neutral pH (*5*). High concentrations of AA displace any non-volatile adduct ions. The high volatility of both ammonia and acetic acid render these easy to remove during the desolvation process.

1.2.2 Electrospray ionisation

The most common ionisation technique for the study of biomolecules is electrospray ionisation (ESI). ESI-MS was first developed by Dole in the late 1960s (*7*) for the determination of polymer molecular masses. Subsequently, Fenn and co-workers applied ESI-MS to the analysis of peptides and proteins (*8*), an achievement for which he was awarded a share of the Nobel prize in Chemistry in 2002. ESI is considered to be a “soft” ionisation method, taking place at atmospheric pressure. During ESI in positive ion mode, analyte in solution is passed through a metal capillary, to which a high voltage (2 - 5 kV) is applied. The highest electric field is at the capillary tip, and results in the enrichment of positive charges at the liquid surface at the tip. This leads to the formation

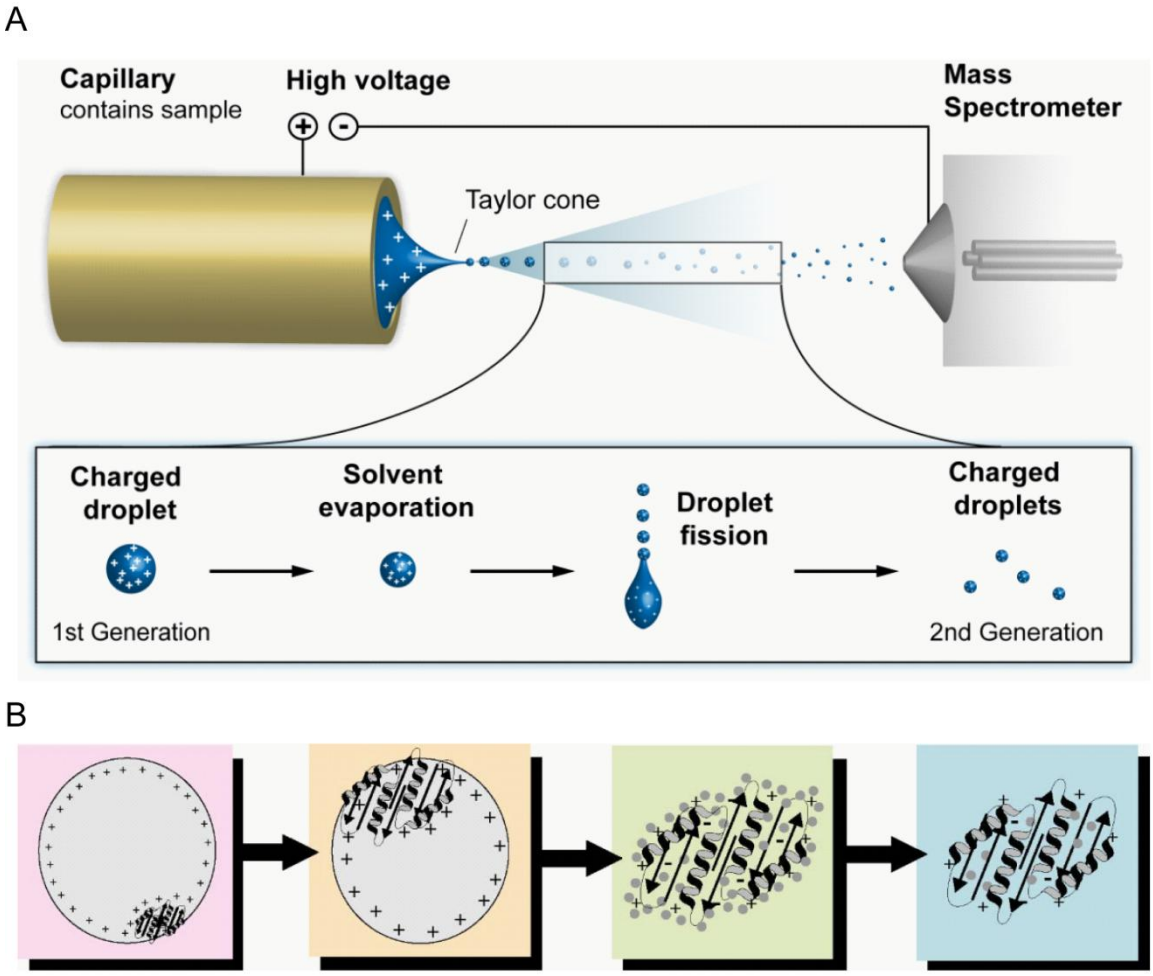


Figure 1.1. Electrospray ionisation schematic. **A** Sample is passed through a metal capillary with high voltage applied. Build-up of charge at the surface of the liquid initiates the formation of a Taylor cone, from which a jet of charged droplets is emitted. Solvent evaporation occurs, increasing the charge density on the surface of the droplet. When this exceeds the Rayleigh limit, asymmetric droplet fission occurs releasing progeny charged droplets. Figure courtesy of Kevin Pagel. **B** In the charged residue model for ionisation, the protein complex remains inside the charged droplet, which gradually decreases in size due to desolvation and repeated fissions. Eventually only the desolvated complex ion remains.

of a "Taylor cone" (*9*) (Figure 1.1A). If the field is high enough, the tip becomes unstable and a jet of solvent emerges from the cone tip, which breaks up into smaller charged droplets with diameters of several μm (*10*). These charged droplets are driven by pressure and potential gradients, assisted by a nebulising gas stream, towards the

sampling orifice of the mass spectrometer. The heated nebulising gas aids desolvation, which subsequently results in droplet shrinkage and an increase in the charge density on the droplet (*9*). This continues until the Rayleigh limit (*11*) is reached, at which point the Coulombic repulsion between charges overcomes the surface tension at the droplet surface. This initiates Coulomb fission, with the formation of progeny droplets which are ~ 2 % of the mass and ~ 15 % of the charge of the parent droplet (*10*). Further desolvation leads to repeated fissions and smaller droplets until gas phase analyte ions are formed. The Rayleigh limit is given by Equation 1.1,

$$ze = 8\pi\sqrt{\varepsilon\sigma r^3} \tag{1.1}$$

where *ze* is the charge on the droplet, ε is the permittivity of a vacuum, σ is the surface tension, and *r* the droplet radius.

Several mechanisms have been proposed to explain the transfer of the analyte to the gas phase during ESI. The ion evaporation model (IEM), proposed by Iribarne and Thomson, purports that ions are directly emitted from the droplet, prior to complete desolvation (*12*). On the other hand, in the charged residue model (CRM) (Figure 1.1B) proposed by Dole *et al.*, the ion remains inside the charged droplet. This droplet gradually decreases in size due to desolvation and repeated fissions until the completely desolvated charged analyte remains (*7*). The consensus is that the IEM describes the ionisation of small molecules, whilst the CRM is the dominant mechanism for macromolecules (*9*). Supporting the CRM, Fernandez de la Mora demonstrated that the maximum charge on a globular protein is governed by the Rayleigh limit of the droplet surrounding the analyte immediately prior to complete desolvation (*13*).

Finally, the conformation-dependent neutralisation theory (CDN) proposed by Grandori and co-workers, suggests that it is the neutralisation of acidic residues, rather than the charging of basic residues, which is important for the observed charge state in MS (*14*). It has been recently suggested however that a combination of CRM and IEM (combined charged residue-field emission model) can best explain experimental observations for macromolecules (*15-16*). In this model, small charge carrying species are emitted from the analyte-containing droplet, a process which also affects charging of the macromolecule.

1.2.3 Nanoelectrospray ionisation

In order to address some of the limitations of ESI, nanoelectrospray ionisation (nESI), an adaptation of conventional ESI, was developed by Wilm and Mann (*17-18*). Ions are formed by nESI in much the same way as described above for conventional ESI. Samples are typically introduced through glass capillaries pulled to a fine tip with an inner diameter of 1- 10 μm (*c.f.* conventional ESI ~ 0.5 mm). The smaller capillary diameter allows lower flow rates of 10 – 50 nL min^{-1}. Compared to ESI, in which flow rates of several μL per min are typically employed, sample consumption is significantly lowered for nESI (1 – 3 μL). Lower flow rates allow the spray tip to be placed closer to the sampling orifice, which further enhances sensitivity. nESI is therefore ideal for biological analyses, where sample volume is often limited, and analyte concentration low. In order to initiate nESI, an electrical potential of 0.8 – 2 kV is applied to the capillary. Electrical conductivity is achieved *via* a conductive ferrule in contact with the gold-coated capillary or a platinum wire insert (*5*). A backing pressure can be applied to initiate/maintain a steady spray.

ESI is also less tolerant to the presence of non-volatile salts. The initial droplet size in nESI however, is approximately ten-fold smaller than in conventional ESI. As a result, desolvation is more efficient, removing the need for heating and requiring fewer fission steps (*19*). This results in nESI having a higher tolerance for non-volatile salt contamination. This is of particular importance for protein complexes, which often have buffer salts present in solution to maintain their stability. A smaller initial droplet size has a further advantage in that the possibility of non-specific aggregation is reduced. Non-specific aggregation can occur when more than one analyte molecule is present in the droplet (*20*). As a consequence, "cleaner" spectra can be achieved for proteins and protein complexes using nESI (Figure 1.2). Finally, the higher charge to volume ratio of droplets formed by nESI enables the use of higher surface tension solvents including water and aqueous buffers, with no need for organic co-solvents (*18*). This has implications for the study of protein complexes, where the preservation of noncovalent interactions and near-native conformations is paramount.

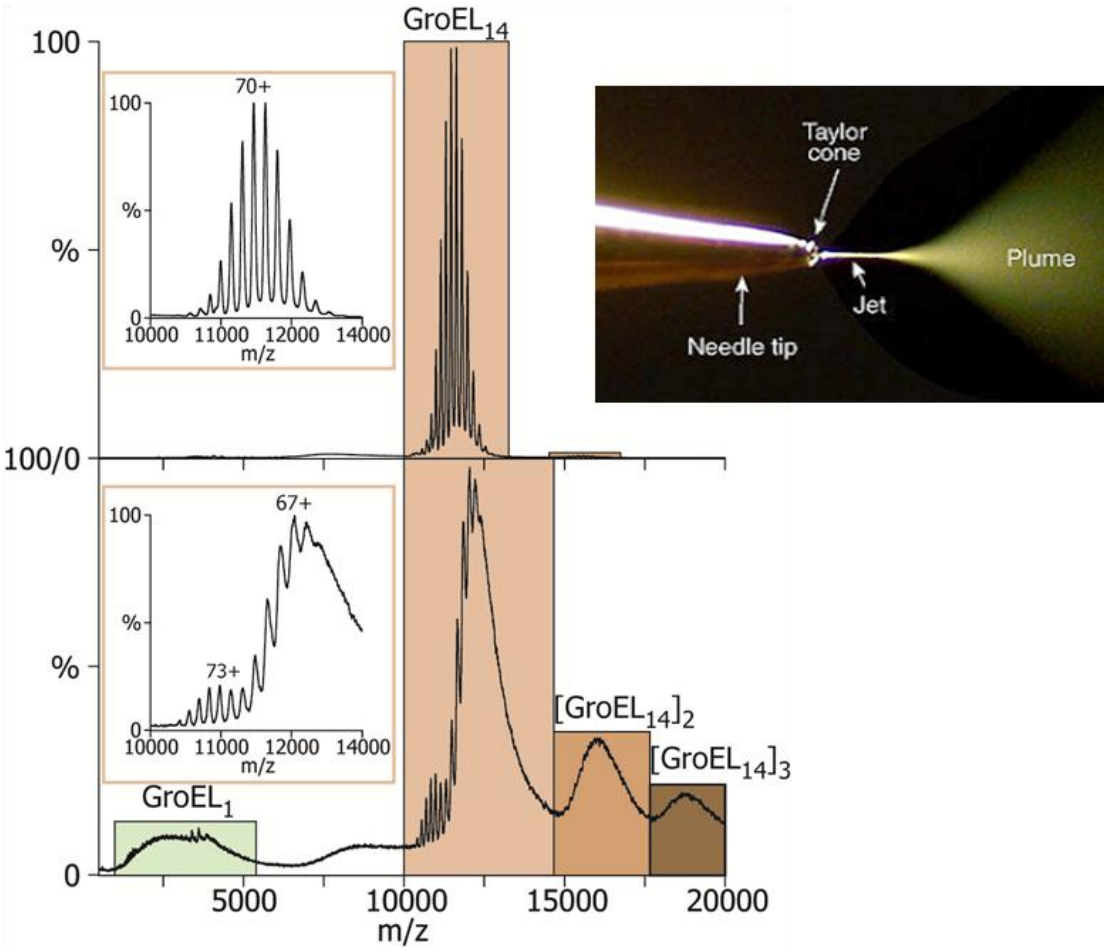


Figure 1.2 Nanoelectrospray ionisation. Comparison of nESI- (top) and ESI- (bottom) MS spectra of GroEL 14-mer. Whilst the ESI-MS spectrum shows evidence for non-specific aggregation of GroEL, the nESI-MS spectrum is characterised by sharp peaks for GroEL 14-mer, and negligible aggregation. Figure adapted from (*10*). The photograph on the right shows the jet of liquid emitted from the nESI capillary (image from www.arcsciences.com).

1.2.4 Electrospray mass spectra

ESI and nESI of macromolecules in the positive or negative ion mode give rise to multiply charged ions, due to the presence of multiple basic and/or acidic sites. Typically a mass spectrum is characterised by a Gaussian-shaped series of peaks at different *m/z* values, corresponding to the analyte at different charge states. This is known as the charge state distribution. Multiple conformations of a protein or protein complex often yield multiple overlapping Gaussian charge state distributions (*21*). Computer software

used to analyse the spectra can derive the molecular mass and charge state by solving the simultaneous equations for adjacent peaks as shown below (*22*):

$$p_1 = \frac{m+ z_1}{z_1} \qquad p_2 = \frac{m + z_2}{z_2} \tag{1.2}$$

Where *m* is the molecular mass, p_1 and p_2 are the *m/z* for adjacent peaks, z_1 and z_2 are the charge of peak 1 and peak 2 respectively. For adjacent peaks, $z_2 = z_1 + 1$. Substituting for *m* in these equations gives:

$$z_1 = \frac{p_2 - 1}{p_1 - p_2} \tag{1.3}$$

Substituting z_1 allows the molecular mass *m* to be determined:

$$m = \frac{(p_2-1)(p_1-1)}{p_1 - p_2} \tag{1.4}$$

The observed charge state distribution is dependent on several factors, including solution conditions (surface tension, pH, ionic concentration, counterions) (*23-24*), instrumental parameters (voltages, pressures, capillary position) (*25-26*), interaction with other molecules in the gas phase and molecular conformation (*21*). Higher charge states for proteins were shown by Chait and colleagues to be correlated with thermal unfolding. This was effected in a stainless steel capillary immediately prior to ESI (*27*). Similarly, the mass spectra for proteins electrosprayed from buffered or denaturing solutions differ significantly (Figure 1.3). The spectra for proteins in their native-like conformations are characterised by few peaks, at high *m/z*. Their maximum charge states match closely (typically 70 - 100 %) the Rayleigh limit for droplets with the same radii as the globular protein (*13*). This can be estimated by means of the molecular mass, *m,* of the protein:

$$Z_R = 0.078\sqrt{m} \tag{1.5}$$

In contrast, spectra for denatured proteins are characterised by many peaks at low *m/z*. Denatured proteins deviate substantially from spherical. This influences the size and shape of precursor droplets and consequently higher charge states than predicted by the Rayleigh limit are observed (*28*). Kaltashov and co-workers noted a close correlation between protein charge states and their solvent accessible surface area, and this is considered to be the primary determinant of how many charges a protein accommodates during ESI (*28*). An interesting example to consider is the protein pepsin, which possesses five basic sites (two arginine, one lysine, one histidine, and the N-terminus). The ESI mass spectrum of pepsin however reveals a charge state distribution centred around 11+, in close agreement with the surface area predicted charge state of 11.2+ (*28*). The wider distribution of charge states typically observed in denatured protein spectra is presumably a consequence of conformational heterogeneity in solution.

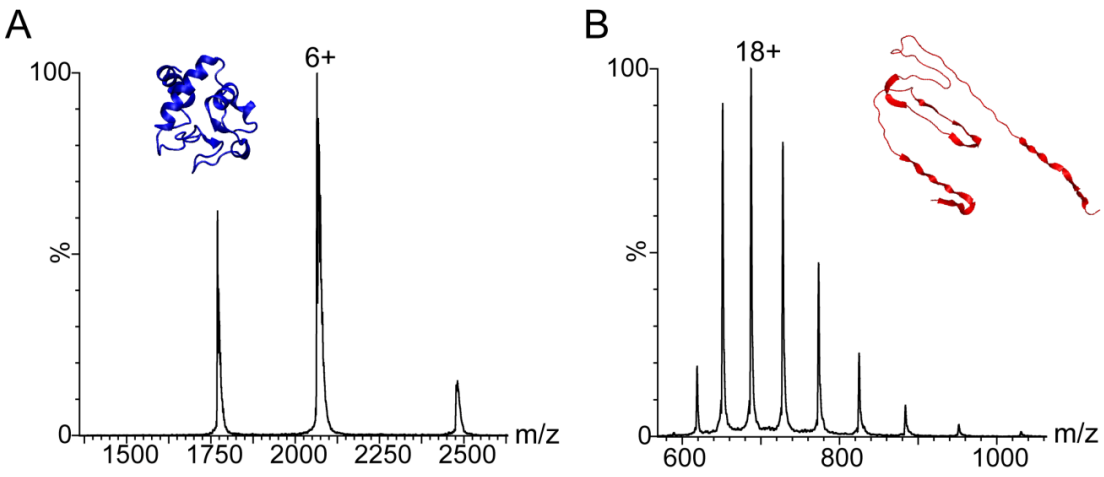


Figure 1.3. Charge state distributions for cytochrome *c* in **A** buffer (200 mM AA), and **B** denaturing solution (49:49:1 water:acetonitrile:acetic acid) reveal higher charge states, and a wider charge state distribution for the denatured protein. Representation for cytochrome *c* in **B** was produced by unfolding the protein manually with the Pymol sculpting feature.

1.2.5 Ion transmission

Ions produced at atmospheric pressure in the source require transmission intact through several pumping stages to the mass analyser(s) under high vacuum. The instruments used here employ a Z-spray nESI ion source (Figure 1.4). The ion beam is placed orthogonal to the sample cone, which in turn is orthogonal to the extractor cone. Potentials in the sample and extractor cone control the path of the ion beam, which must bend in order to enter the high vacuum MS. This has the advantage of preventing the build-up of neutral species.

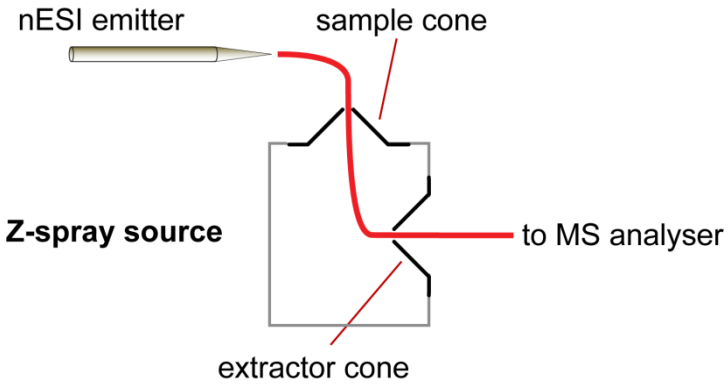


Figure 1.4. Z-spray nESI source schematic. The ion beam is placed orthogonal to the sample cone, which is in turn orthogonal to the extractor cone. The ion beam therefore has to bend twice before entering the high vacuum MS, preventing the build-up of neutral species.

Maintaining weak noncovalent interactions requires careful optimisation of both source accelerating voltages and gas pressures. An increase in pressure in the transfer region between the source and analyser significantly improves the signal intensities of high *m/z* ions (*29*). This can be achieved by partially closing the isolation valve (SpeediValve) of the source vacuum line to the roughing pump. Other methods to achieve increased pressure in the first vacuum stages include increasing the orifice size of the sampling cone, or fitting a sleeve around the front section of the first ion guide (*5*). The higher pressure results in more collisions with neutral gas molecules, which serves to dissipate

excess kinetic energy into the gas, and therefore focusing ion trajectories ("collisional focussing") (*10*).

In this thesis, the instruments used are hybrid quadrupole/ion mobility/time-of-flight mass spectrometers. Ions are directed from the extractor cone into the quadrupole mass analyser through an RF-only hexapole ion guide. Ions then pass through the mobility region before *m/z*-based separation in the time-of-flight (ToF) mass analyser and detection. The principles behind these two mass analysers are discussed below; a more detailed description of the instrumentation used is given in Chapter 2.

1.2.6 Quadrupole mass analyser

The quadrupole mass analyser consists of four parallel rods or electrodes with a time-varying electric potential applied (Figure 1.5). This electric potential has a DC component and an RF component (*30*). Each electrode pair is biased with a DC voltage opposite in sign to that of the other pair, whilst the RF voltage on one pair is 180^{o} out of phase with the other pair. The superposition of these voltages in space results in a quadrupolar electric field (*31*). This field determines whether an ion has a stable trajectory allowing passage through the quadrupole. Ions with instable trajectories, on the other hand, are deflected and collide with the rod surface.

Quadrupoles can be used in two modes: scanning mode and RF-only mode. In scanning mode, mass scanning is accomplished by maintaining a constant frequency, whilst varying the DC voltage and RF amplitude simultaneously. Selection of a narrow window of ions around a particular *m/z* can be achieved. This is particularly important for tandem

MS, in which an ion of interest is selected in the quadrupole and subsequently dissociated in the argon-containing collision cell. Quadrupoles can also be used in RF-only mode, in which the quadrupole acts as an ion guide, aiding the transmission of ions by focussing trajectories. Ions of 0.77 to 5 times a particular *m/z*, as determined by the RF amplitude, are transmitted (*29*). RF-scanning across the area of interest allows the full MS spectrum to be obtained.

Quadrupoles allow transmission of ions up to a maximum *m/z* value determined by:

$$[m/z]_{max} = \frac{7\ x\ 10^6\ V}{f^2 r_o^2} \quad (1.6)$$

Where *V* is the RF amplitude, *f* is the RF frequency, r_o is the inner radii between the rods. In order to select and transmit high mass ions of up to 32 000 *m/z*, quadrupoles are operated at reduced RF frequencies (300 kHz) (*29*).

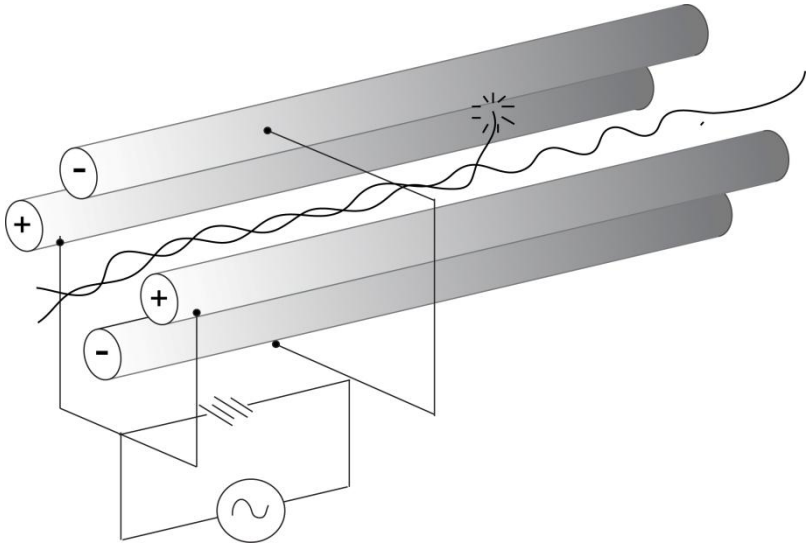


Figure 1.5. Quadrupole mass analyser schematic. Each pair of rods is has an equal but opposite DC charge, and an RF voltage 180^o out of phase with the other pair. The path of an ion with a stable trajectory passing through the quadrupole, and that of an ion with an instable trajectory hitting the wall of the analyser is shown. Figure adapted from (*30*).

1.2.7 Time-of-flight mass analyser

Due to the pulsed nature of the ToF experiment, ions from a continuous beam (as produced in ESI) are first trapped before being pulsed orthogonally into the field-free ToF region at a frequency which matches the extraction timing of the ToF analyser (*10*). All ions are transmitted simultaneously in contrast to the scanning quadrupole mass analyser, providing high sensitivity (Figure 1.6). Ions are accelerated into the ToF flight tube with constant voltage. Ideally all ions are energised to the same kinetic energy, and travel at varying velocities through the flight tube. The velocity is dependent on the ion mass, and therefore ions of differing masses arrive at the detector at different times, with heavier ions taking longer to reach the detector (*32*). This is summarised in the following equations.

Ions with charge *ze* and accelerating potential V_o are accelerated to the same kinetic energy:

$$zeV_o = \frac{mv^2}{2} \tag{1.7}$$

where, *m* is the ion mass, and *v* the ion velocity. Ions enter the flight tube with velocity*:*

$$v = \left(\frac{2zeV_0}{m}\right)^{\frac{1}{2}} \tag{1.8}$$

Flight time, *t,* is based on kinetic energy, mass and the distance travelled, *D*:

$$t = D\left(\frac{m}{2zeV_o}\right)^{\frac{1}{2}} \tag{1.9}$$

This can be rearranged in terms of *m/z*:

$$\frac{m}{z} = 2V_o e\left(\frac{t}{D}\right)^2 \tag{1.10}$$

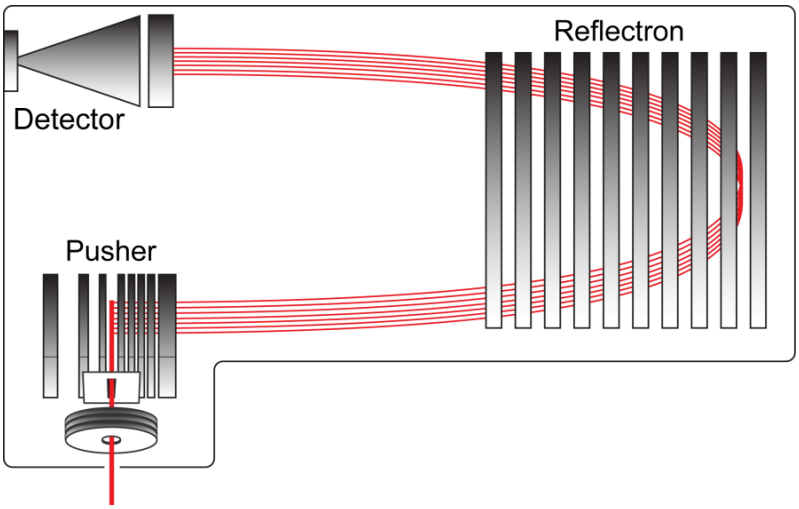


Figure 1.6. Time-of-flight mass analyser schematic. Ions are pulsed into the ToF region by the pusher, and travel through a field-free region to the reflectron. The reflectron corrects for differences in kinetic energy and deflects ions towards the detector. Ions with lower *m/z* reach the detector before those with high *m/z*. Figure adapted from (*10*).

In reality ions have a distribution of kinetic energies, which is corrected for by the use of an electrostatic reflector (reflectron). Ions of the same mass, but differing in their kinetic energies will penetrate the reflectron to different extents. Ions with higher kinetic energy will penetrate the reflectron further, and therefore have marginally longer flight paths, than ions with less kinetic energy (*33*). As a result the ions reach the detector at the same time. The reflectron has the additional advantage of doubling the length of the flight tube, therefore providing enhanced mass resolution. In addition to high mass resolution and sensitivity, ToF mass analysers have the advantage of a theoretically unlimited *m/z* range, making them ideally suited for the analysis of large macromolecules (*32*).

1.2.8 Detectors

The most common detector found in ToF-MS instruments is the micro-channel plate detector (MCP). MCP detectors are electron multipliers, consisting of a plate containing parallel cylindrical channels which are coated with a semi-conductor substance (Figure 1.7). Ions reaching the detector hit the plate, and secondary electrons are emitted, creating an electron cascade though the channel. This results in a signal intensity increase of several orders of magnitude, providing a measurable current. One advantage of the MCP detector is that multiple channels allow the detection of several different *m/z* ions simultaneously. The signal is converted to precise arrival times by means of a time-to-digital converter (TDC), or an analogue-to-digital converter (ADC). The main difference between the two is that ADC can differentiate between single and multiple ions of a given *m/z* arriving at a particular time. The use of the ADC signal converter thus improves the instrument's dynamic range (*34*).

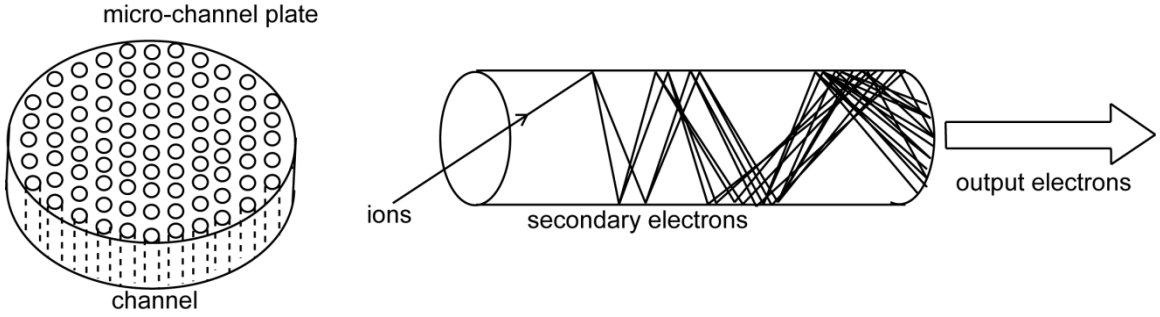


Figure 1.7. Schematic of the micro-channel plate detector. Ions hit the plate and pass through channels, lined with semi-conductor material. As ions hit this material, secondary electrons are emitted, causing an electron cascade through the channel, multiplying the signal by several orders of magnitude, resulting in a measurable current.

1.2.9 Tandem mass spectrometry

Tandem MS or MS/MS is achieved by placing two mass analysers in series. The instrument of choice for large biomolecules (*34*) is the quadrupole/ToF hybrid MS (Q-ToF). The Q-ToF consists of a quadrupole and ToF mass analyser placed in series, and separated by a collision cell (*10*). During tandem MS, ions at particular *m/z* are selected and transmitted through the quadrupole and into a collision cell in which the ion is dissociated. The ion selected in the quadrupole is known as the precursor ion, and the dissociated charged fragments as the product ions. These product ions are transmitted to the ToF mass analyser, where they are separated according to *m/z* and subsequently detected. Dissociation of protein complexes may be induced using a variety of techniques. Analyte ions can be collided with a surface (surface-induced dissociation, SID) (*35*) or neutral gas molecules (collision-induced dissociation, CID) (*36*), bombarded with electrons (electron capture dissociation, ECD) (*37*) or infrared radiation (black-body infrared dissociation, BIRD) (*38*).

In this thesis, CID was employed for activation and dissociation of protein complexes. During the collisions of ions with neutral gas molecules (typically argon), a portion of the ions kinetic energy is converted into internal energy. If the internal energy increases sufficiently dissociation can occur (*10*). The strength of the collisions can be controlled by the pressure in the collision cell, the diameter of the neutral gas molecule and the accelerating voltage into the collision cell (collision energy). As the acceleration of the collisions gradually increases, the protein complex ions undergo “collisional cleaning”, in which further desolvation and removal of adduct ions occurs (*36*). This is advantageous in that narrower *m/z* peaks can be achieved, and consequently more accurate masses measured, however the protein complex may also undergo

conformational changes (*36, 39*) as a result of activation. Further increases in activation typically result in the unfolding of one or more monomers (*40*) and charge migration to the newly formed solvent-accessible regions occurs in a cooperative manner (*41*). This culminates in the ejection of a monomer subunit as a result of Coulomb repulsion (*42*). CID of multimeric complexes typically results in an asymmetric partitioning with respect to mass and charge (*42*). This characterised by the loss of a monomer which carries a disproportionately high number of charges with respect to its mass, in relation to the remaining "stripped complex" (Figure 1.8). The partitioning of surface area however is believed to be proportional to charge, with the ejected monomer exhibiting a highly extended conformation (*43*).

Tandem MS has been highly successful for the analysis of polydisperse analytes such as heat shock proteins. MS spectra of these heterogeneous systems are difficult to resolve due to overlapping *m/z* peaks (*44*). Tandem MS spectra of the charge-stripped complexes following monomer ejection are characterised by peaks at higher *m/z* and consequently with greater separation between adjacent charge states. Finally, from a structural perspective CID can provide information on the identity and masses of peripheral subunits (*45*).

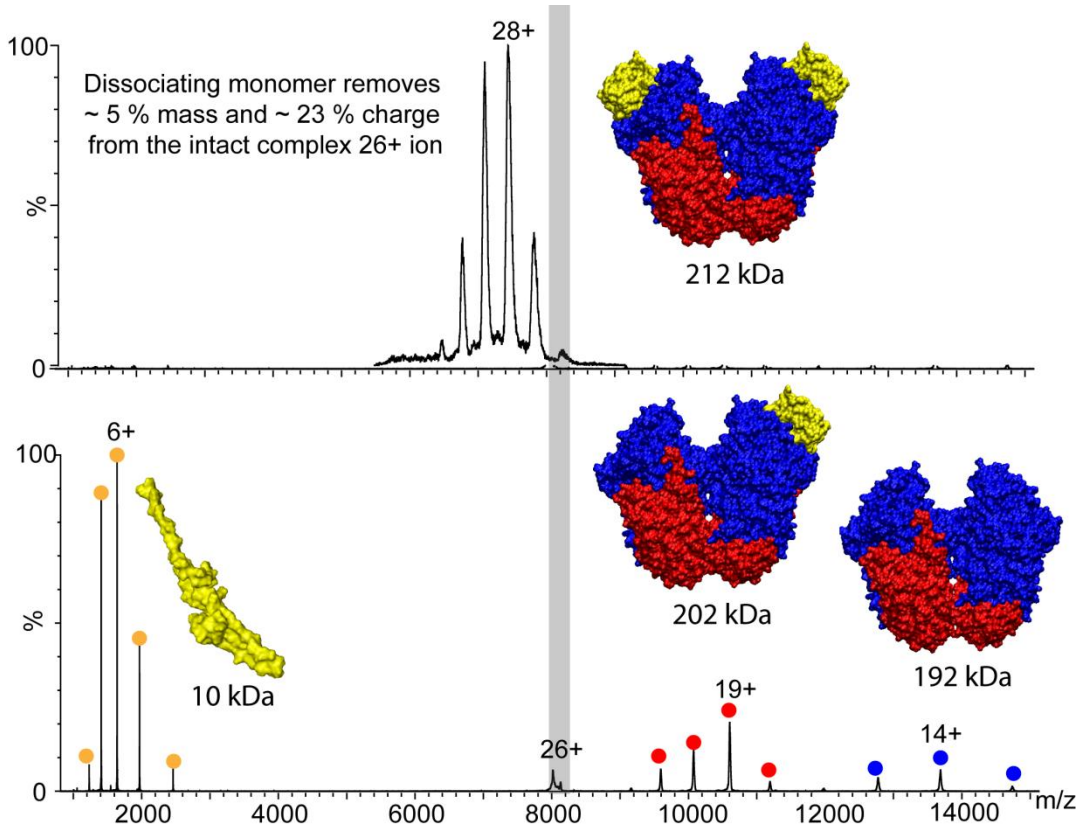


Figure 1.8. Collision-induced dissociation of a heteromeric complex. **A** Mass spectrum of $\alpha_2\beta_2\gamma_2$ ToMOH under mild instrument conditions. **B** MS/MS spectrum of 26+ ToMOH in which γ-monomers (yellow) are ejected from the intact 6-mer to form $\alpha_2\beta_2\gamma$ 5-mer and subsequently $\alpha_2\beta_2$ 4-mer. Representation for the dissociated γ-monomer was unfolded manually using Pymol sculpting tool.

1.3 ION MOBILITY

1.3.1 Principles of ion transport

Ion mobility is a separation technique effected in a gas-filled chamber or tube. Ions with the same *m/z*, but different geometries can be distinguished by their gas-phase mobilities (Figure 1.9). Drift times in these experiments depend on interactions between ions and drift-gas molecules in the presence of a weak uniform electric field (*46*). Ion velocities (v_D) are proportional to the electric field (E) and the mobility of the ion (K) under the conditions in the drift tube.

$$v_D = KE \tag{1.11}$$

The transit time for the ion is governed by collisions with the neutral drift gas. Therefore, the mobility of the ion is inversely proportional to the gas pressure, p, and the buffer gas density, N *(47)*. The drift velocity, v_D, is thus proportional to the ratio E/N or E/p. The mobility also depends on the ion-neutral forces; this takes the form of a diffusion collision integral or average collision cross section, Ω_{avg} (*47-48*). Provided the field is weak, below that of the low-field limit, mobility is field independent and can be calculated using Eq. 1.12.

$$K = \frac{1}{N}\frac{(18\pi)^{\frac{1}{2}}}{16}\frac{ze}{(k_B T)^{\frac{1}{2}}}\left[\frac{1}{m_I} + \frac{1}{m_N}\right]^{\frac{1}{2}}\frac{1}{\Omega_{avg}} \tag{1.12}$$

where m_I and m_N are the masses of the ion and drift gas respectively. k_B is the Boltzman constant, T temperature of the drift gas, z the number of elementary charges on the ion, e the charge on one electron, and N the drift gas number density.

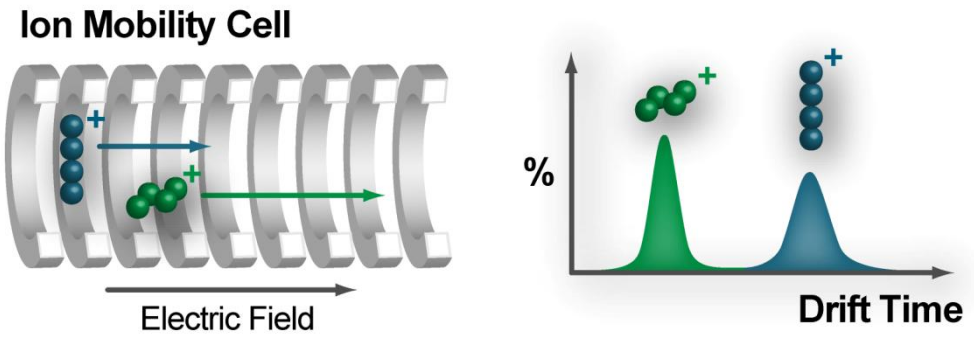


Figure 1.9. Ion mobility separates ions with the same *m/z* but differing topologies. Ions with more compact conformations undergo fewer collisions with the drift gas, and therefore exhibit a higher gas-phase mobility and a shorter transit time through the drift tube. Figure courtesy of Kevin Pagel.

There are several implicit assumptions in the kinetic theory for ion transport described above. The ions and neutrals are treated as hard spheres, with perfectly elastic collisions. Only binary collisions are considered between neutrals and ions, with no ion-ion collisions. The most important assumption is that the thermal energy of the ion is much greater than the energy acquired by the ion as a result of the field. The drift gas density is therefore high enough to collisionally dampen ions such that the internal ion temperature is that of the bulk buffer gas. As a consequence ions are not expected to align with the field and the collision cross section (CCS) measurement corresponds to the average of all random orientations of the ion as it passes through the drift region (*46*) (Figure 1.10). The assumption is only valid for field strengths below that of the low field limit. Field strength is given by the ratio *E/N* which is typically reported in Townsends (Td). The field is considered weak if the following inequality is satisfied (*47, 49*).

$$\frac{E}{N} < \left[\frac{m_I}{m_I + m_N}\right]^{\frac{1}{2}} \frac{d^2}{z} \qquad (1.13)$$

where *d* is the sum of the ion and neutral radii in measured in angstroms.

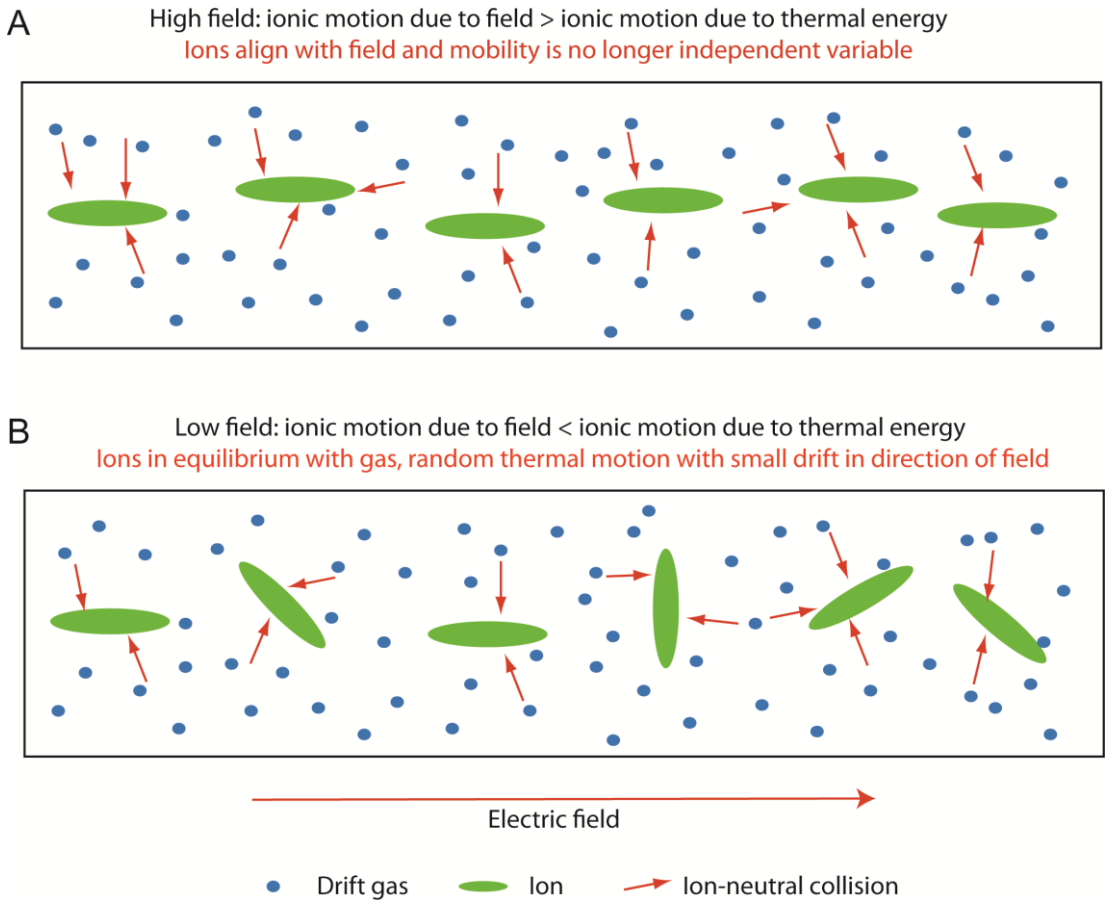


Figure 1.10 Ion mobility above and below the low field limit. **A** High field strength in the drift tube result in ions aligning with the field. Mobility becomes a field-dependent property. **B** Low field strength allows ions to equilibrate with the bulk gas. Mobility is field independent and CCS measurement corresponds to the average of all random orientations of the ion as it passes through the drift tube.

1.3.2 Drift tube *versus* travelling-wave ion mobility

The two types of IM used in this thesis are drift tube (or static field) IM and travelling-wave IM. In traditional drift tube IM the electric field is uniform along the drift tube and CCS can be directly calculated from drift times. Typically CCS values are determined from the slopes of drift time *versus* reciprocal drift voltage plots (Eq. 1.14).

$$t_D = \frac{L}{KE} + t_0 \qquad (1.14)$$

The intercept is given by the transport time of ions from the exit of the drift region to the ToF mass analyser (t_0), and the slope is inversely proportional to K. L is the length of the drift tube.

To enable direct comparison of mobility values over a range of experimental conditions, reduced mobilities (K_0), corrected to standard pressure (p_0 = 760 Torr) and temperature (T_0 = 273.15 K) are reported:

$$K_0 = \frac{p}{p_0}\frac{T_0}{T}K \qquad (1.15)$$

The CCS (Ω) of an ion is related to the reduced mobility (K_0) through the Mason-Schamp equation (*48*).

$$\Omega = \frac{3ze}{16N}\sqrt{\frac{2\pi}{\mu k_B T}}\frac{1}{K_0} \qquad (1.16)$$

where μ is the reduced mass of the ion and drift gas.

The relationship between drift times and CCS is more complex for travelling-wave IM experiments, for which electric fields vary with time and CCS are instead determined by calibration using ions with known CCS (Figure 1.11) (*50*). In travelling-wave IM, ions are propelled by superimposing a DC voltage pulse on the confining RF of an electrode,

and then moving the pulse to an adjacent electrode and so on, to provide a travelling voltage wave (*51*). Ions with lower mobility fall behind the travelling wave more often than ions with higher mobility and this forms the basis for separation (*52*). Whilst travelling-wave IM has a lower resolving power than traditional drift tube IM, there is significantly improved ion transmission, rendering it suitable for routine use.

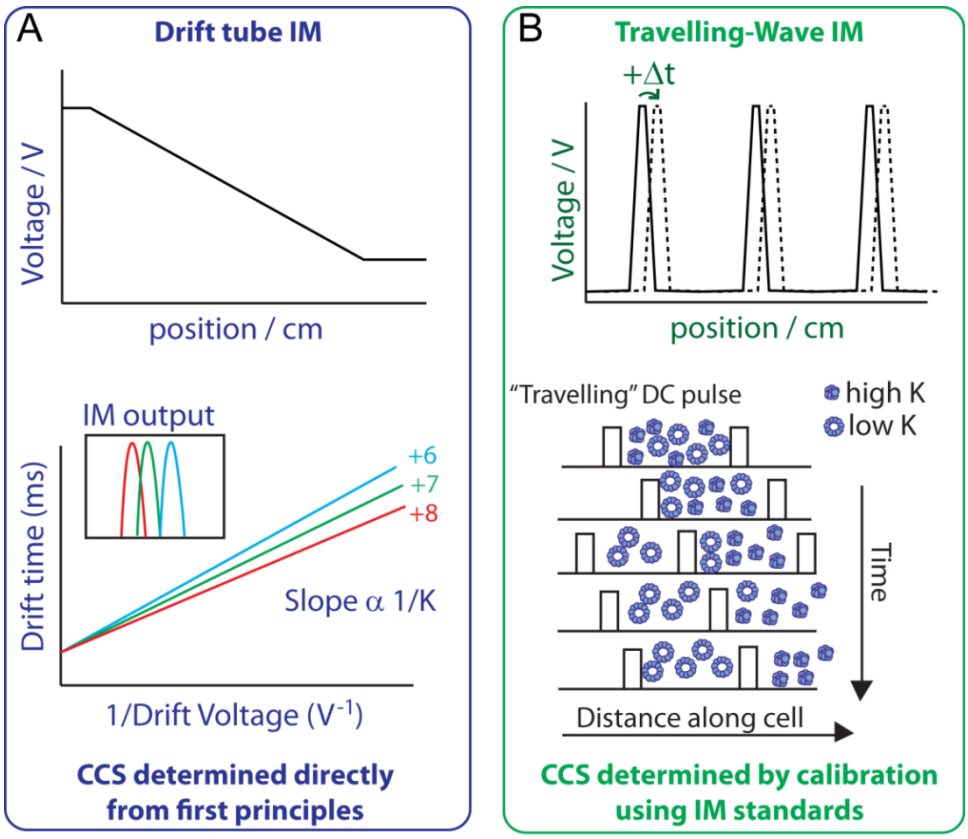


Figure 1.11. Drift tube *versus* travelling wave IM. **A** In traditional drift tube IM, the field is uniform across the drift tube. CCS measurements are determined from plots of drift time as a function of inverse drift voltage. **B** In travelling-wave IM, the field varies with time, consisting of a "travelling" DC pulse. Ions with lower mobility fall behind the travelling wave more often than ions with higher mobility. CCS are determined by calibration using IM standards.

1.4 THEORETICAL APPROACHES

1.4.1 Collision cross section calculations

CCSs derived from IM are often interpreted by comparing with CCS calculated from atomic or low-resolution models. Calculated CCSs can be obtained for atomic coordinates from NMR, X-ray crystallography and homology modelling, or for spherical-based models using coarse-grained approaches (*53*). In this way, structural information can be obtained from IM and CCS-derived restraints can be used for modelling analysis.

The simplest method for estimating CCS values is the projection approximation (PA) algorithm. This treats the polyatomic ion as a set of hard spheres, with defined collision radii for the respective atoms with the buffer gas. The average CCS can be obtained by averaging the projected area of these spheres onto a plane, over all possible orientations (*54*). This approach assumes hard sphere interactions between the ion and buffer gas and was first implemented by Mack (*55*). Subsequently Bowers and co-workers adapted this approach in order to allow calculation of mobilities at different temperatures (*56*). However the PA approach ignores long-range interactions between the ions and buffer gas. Additionally, the PA also ignores details of scattering processes between the buffer gas and ions. This becomes increasingly important for large complex ions, particularly those with concave surface topologies (*57*). As a consequence PA typically underestimates CCS.

In order to account for contributions from multiple collisions between ions and buffer gas molecules, Shvartsburg and Jarrold developed the exact hard spheres scattering (EHSS) method (*57*). Here the Ω_{avg} in Eq. 1.12 is treated as a momentum transfer integral, which

is related to the scattering angle. This is the angle between the incoming and outgoing trajectory in a collision between the ion and buffer gas. Rigorous derivation of Ω_{avg} takes a series of integrations that average all possible ion-neutral collisions over varying scattering angles and collision energies. In this method, even though scattering is treated correctly, within a hard-sphere limit, the long-range interactions between the buffer gas and ions are not considered (*54*).

Another way to calculate CCS is the trajectory method. This is the most reliable and accurate method for calculating the collision integral Ω_{avg}. This takes into account long-range interactions, using a realistic intermolecular potential. Mesleh *et al.* employ a potential consisting of a sum of two-body Lennard-Jones and ion-induced dipole interactions (*58*). Although accurate, this approach is computationally intensive, and therefore unsuitable for the mobility calculation of large biological molecules (*59*).

These three algorithms to calculate CCS are included in MOBCAL, an open-source software program, developed by Shvartsburg and Jarrold (*57-58*). These algorithms have been parametised on small organic and inorganic molecules, and may suffer from inaccuracies when applied to large and complex biomolecules. Recently Bleiholder *et al.* have developed the projected superposition approximation algorithm (*60*). This accurately accounts for size and shape effects on the intermolecular interaction potential, and is computationally efficient. However this algorithm is not yet in general circulation. None of these methods however consider the effects of inelastic collisions, or the possibility of gas-phase rearrangements. A preliminary energy relaxation or molecular dynamics (MD) simulation *in vacuo,* prior to CCS calculation, may reduce errors.

Comparing experimental and theoretical CCS is still by no means a solved problem and remains the source of greatest controversy within the gas-phase structural biology community.

1.4.2 Molecular dynamics

Originally developed for theoretical physics, MD is now widely used in biological chemistry, to study the structures and dynamics of large biomolecules. MD simulations follow the trajectories of atoms and molecules by numerically solving Newton's equations of motion for a system of interacting particles. The trajectory is defined as the change in position coordinates with time. Each particle experiences a total force from interactions with other particles as defined by the force field. The force field is a set of parameters describing the potential energy of a system of particles and defines a set of parameters for each atom type. This includes atomic mass, van der Waals radii, and partial charge. Derived empirically and from quantum mechanics calculations, the force field considers both bonded forces (chemical bonds, bond angles, bond dihedrals) and non-bonded forces (van der Waals and long range electrostatic interactions). From the force acting on each particle, the acceleration can be calculated using Newton's law. Combined with knowledge of the particle's prior position and velocity, this allows the new position after a defined time step to be calculated.

The combination of MD simulations with gas-phase experiments, such as IM-MS, can provide valuable insights into the structure and interactions of biomolecular ions (*61*). Early studies of small proteins (*62-63*) and peptides (*64-65*) revealed candidate structures for these gas-phase species that have been informative both in terms of folding and the

effects of dehydration. *In vacuo* MD simulations of globular proteins revealed residual water molecules, aggregating on the protein surface, which assist the preservation of folded structure (*66*). Upon dehydration, limited conformational changes were observed. These included the inward-folding of hydrophilic side chains, which tended to form new hydrogen bonds (*67*). These new hydrogen bonds are likely to play a key role in the stabilisation of secondary structure as the protein is transferred into vacuum conditions.

1.5 GAS-PHASE PROTEIN STRUCTURE

Since the introduction of ESI and the first experiments to demonstrate the preservation of noncovalent binding in the mass spectrometer (*68*), one of the most controversial issues has been the extent to which gas-phase protein structures mirror those found in solution. Breuker and McLafferty used native electron capture dissociation (NECD) of cytochrome *c* protein ions in order to probe folded structure (*37, 69*). NECD monitors site-specific thermal unfolding in the gas phase. Using this approach, the unfolding of regions interacting hydrophobically in the native structure could be shown, despite the fact that these regions were the last to unfold in solution (*70*). The results of these experiments were compared with MD simulations of cytochrome *c*, surrounded by a monolayer of water and placed in a vacuum (Figure 1.12) (*62*). These simulations are consistent with evaporation of water from the native protein, with the last water molecules to leave being those which had been attached to the ionic side chains (*71*). Complete loss of water leads to the collapse of exterior ionic groups, which form new interactions with each other, protecting native-like interactions within the protein ion. Storage of ions for millisecond time periods can result in loss of hydrophobic and electrostatic interactions.

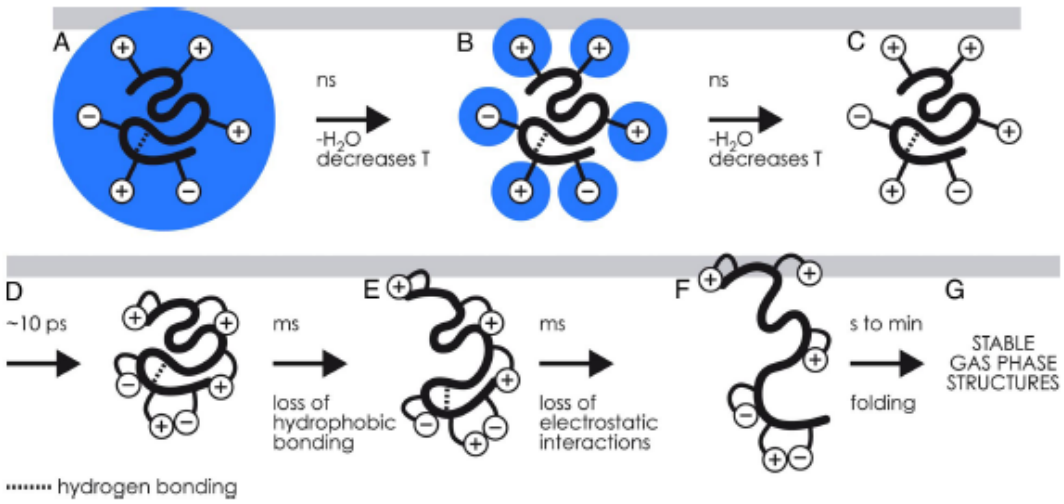


Figure 1.12. Evolution of protein structure after ESI. **A** Protein covered in shell of water. **B** Protein undergoes loss of water and cooling. **C-D** Dry protein undergoes collapse of exterior ionic groups on a picosecond time scale. **D-G** Near-native protein undergoes millisecond loss of hydrophobic bonding and electrostatic interactions, forming new noncovalent interactions to give a stable gas phase structure. Figure adapted from (*71*).

Recently further experimental evidence from ECD was presented in an investigation of a three-helix bundle protein KIX. These ECD experiments rely on unselective backbone cleavage, with the retention of noncovalent intramolecular interactions, which prevent the separation of backbone cleavage products. In this way, the abundance of ECD fragments provides a measure of the extent of gas phase unfolding. The experimental data showed the most stable regions of the protein to be those forming salt bridges and ionic hydrogen bonds (*72*). The wider implications of this study is that if proteins are stabilised by a favourable network of electrostatic charges their native structure is more likely to survive in the gas phase than a protein with less electrostatic interactions.

An alternative means of assessing the folded states of proteins in the gas phase is by hydrogen-deuterium exchange (HDX) in the gas phase. These experiments report on

protection from exchange and provide a read out of the number of exchangeable sites involved in hydrogen bonding interactions. HDX experiments showed that labile protons in cytochrome *c* (5+ to 17+) were exchanged for deuterium, in different patterns consistent with at least six gas-phase conformations (*73*). Infrared laser heating and fast collisions were found to induce ions to unfold and exchange to a greater extent, while charge-stripping ions to lower charge values yielded apparent folding. Combining HDX with IM-MS to probe the gas-phase structure suggests that the 5+ charge state of cytochrome *c* ions have near native-like structure while the 9+ charge state contains an extended helix (*74*).

Further support for compact structure of low charge states comes from recent IM-MS studies of ubiquitin (Figure 1.13). Solutions that stabilise the native state of ubiquitin yielded essentially one family of tightly folded desolvated ubiquitin structures with CCS that match the size of the native structure (*75*). By contrast, solution conditions favouring the partially folded A state yielded CCS matching in size to conformations between the A state and the fully unfolded state. The A state, which is charged more extensively than the native-like state, decays rapidly (≤50 ms) unfolding to more extended structures. By contrast the native-like conformation of ubiquitin survives for >100 ms in a 294 K solvent-free environment (*75*). These experiments establish that compact structures for low charge states can survive for extended periods of time in the gas phase. This prompted further investigation with recent ECD experiments of ubiquitin (*76*). The results of these ECD experiments establish that the order of unfolding of the various regions is dictated by the extent to which hydrophobic bonding is compensated for by electrostatic interactions.

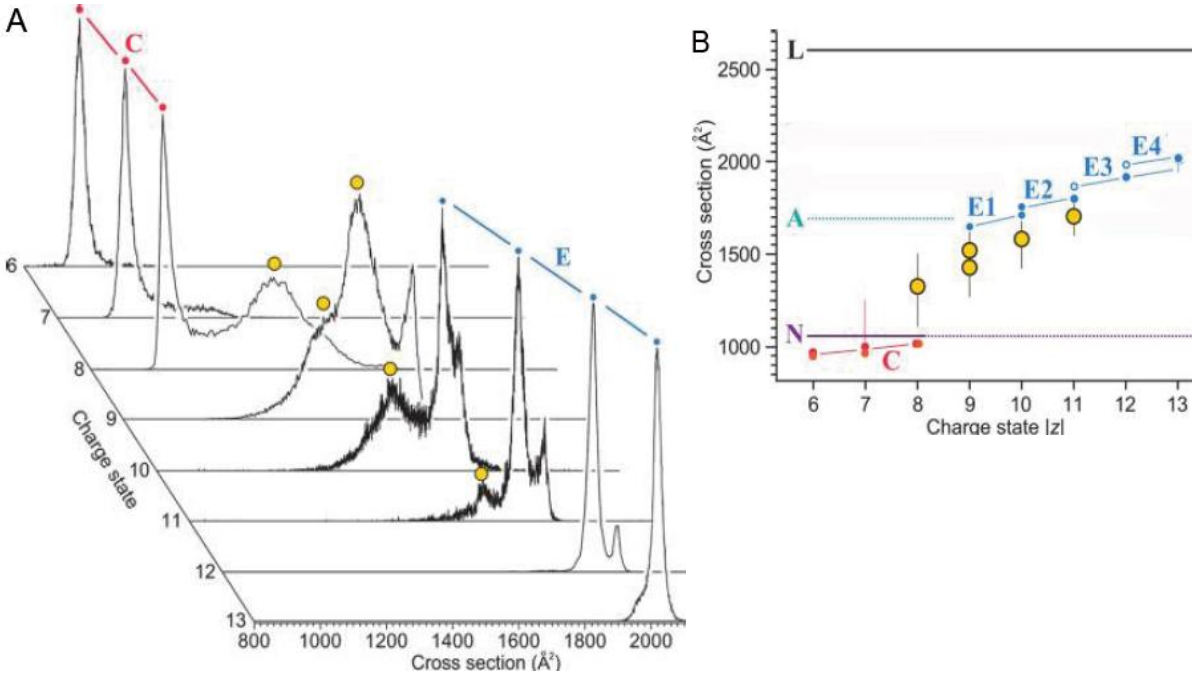


Figure 1.13. Ion mobility data for ubiquitin. **A** CCS distributions for ubiquitin ions as a function of charge state. Peaks marked in red (6-8+) and blue (10-13+) correspond to a family C and E of ions respectively. Broad features observed for ions of intermediate charge are marked with a yellow dot. **B** Compact structures, C have CCS in close agreement with that calculated for the native structure (N). Also shown are CCS for the A state (A) and for a theoretical linear (L) conformation. Figure adapted from (*75*).

The balance between hydrophobic and electrostatic interactions was also critical in determining the outcome of early investigations of protein-ligand interactions. For small proteins investigated with a number of ligands, it was often the case that more favourable binding was observed for ionic ligands, rather than for hydrophobic ones (*77-78*). The early observations are however particularly interesting in the light of recent observations of hydrophobic ligand binding within an accessible protein cleft (*79*) and intact membrane protein complexes (Figure 1.14), wherein the overwhelming number of subunit interactions are hydrophobic (*80-82*). Evidence from MD simulations suggests that the micelle shields the native structure during transfer into the gas phase (*83*). However such an explanation is not possible for the hydrophobic ligand binding

interaction and its preservation suggests that a van der Waals component must contribute to the stability of these complexes once in the gas phase. This implies that electrostatic interactions important in determining preservation of the folded state of individual proteins must also be complimented by van der Waals forces when ionic interactions are absent.

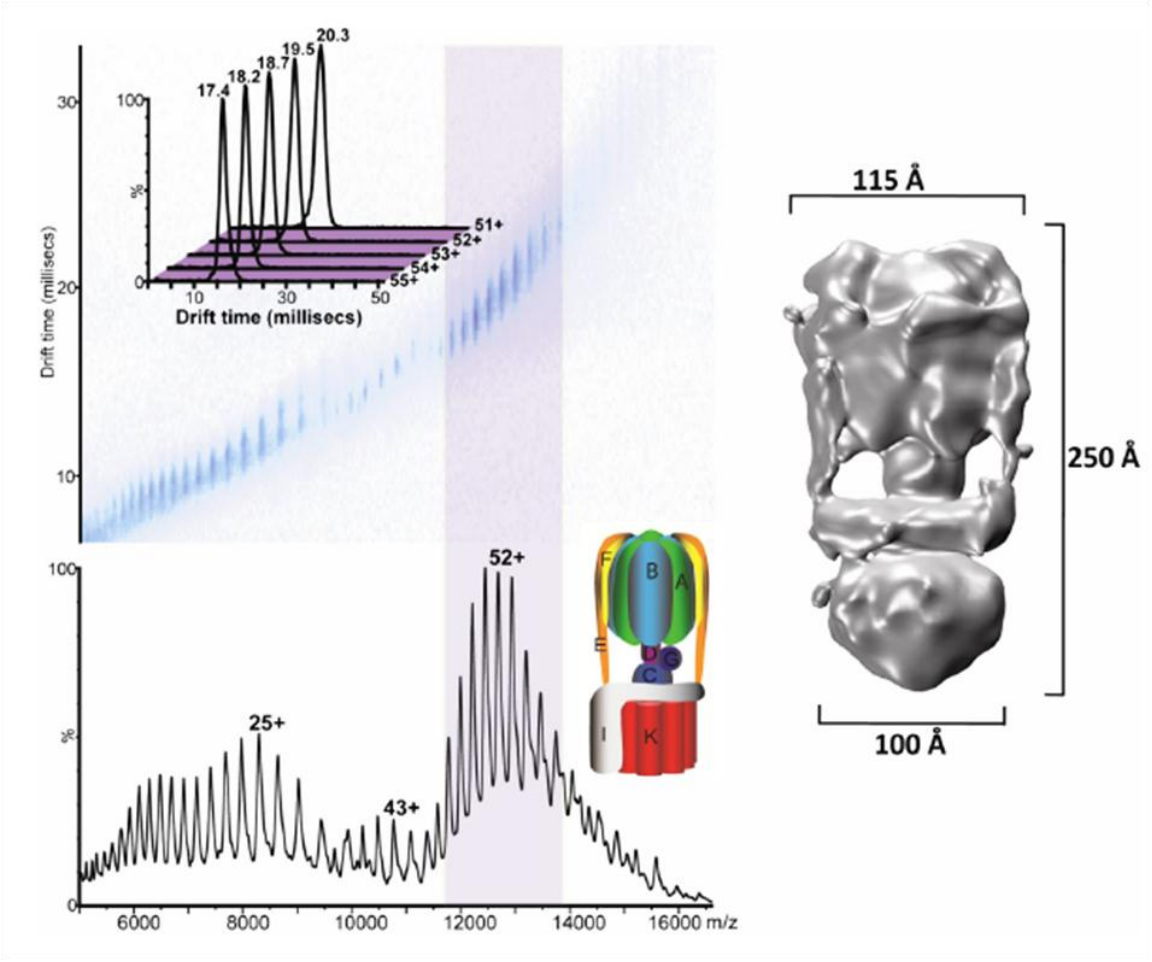


Figure 1.14. Ion mobility-mass spectrometry of the intact *Thermus thermophilus* ATPase. IM-MS data for intact ATPase (highlighted). Calculated CCS value based on the EM density is consistent with the experimentally measured values. Figure adapted from (*82*).

The first direct evidence for the conservation of native-like architecture in a gas-phase multimeric complex came from IM-MS experiments (*39*). These experiments demonstrate that the ring topology of 11-mer *trp* RNA binding protein (TRAP) is maintained in the solvent-free environment of the mass spectrometer. Furthermore, it was shown that binding of tryptophan and RNA increased the stability of the ring structure, and prevented structural collapse. The preservation of an internal binding cavity within a protein assembly was also suggested by tandem MS experiments. These studies revealed that substrates of GroEL remain buried within the internal cavity upon transfer to the gas phase (*84-85*). Monomer subunits are preferentially lost over substrate molecules during CID. CCS measurements, in which no difference was observed between the empty GroEL and GroEL-GroES complexes and those bound to substrates, point to the same conclusion. Evidence supporting the collapse of GroEL during ESI-MS however has also been reported (*86*).

Recently, IM-MS has been applied to membrane proteins (*82*). CCS measurements for a 130 kDa potassium ion channel were consistent with a compact structure, providing further support that native-like quaternary structure can be maintained in the gas phase (*87*). This is despite the harsh conditions required to remove the surrounding detergent micelle. Similar measurements for the tetrameric transporter complex $BtuC_2D_2$ indicated the presence of a broad range of conformations, encompassing the theoretical CCS (*87*).

Intriguingly, retention of enzyme activity and virus infectivity, following soft-landing after transmission through a mass spectrometer has been reported (*88-89*). With function closely related to structure, these data strongly support the hypothesis that extensive

irreversible structural changes do not occur during the time scale of the MS experiment. Overall, there is clear evidence that many of the interactions which maintain the quaternary structure of protein complexes persist in the gas phase. Whilst these may not precisely reflect the protein complex in its native state, the preservation of distinctive architectures and specific binding of ligands suggest that gas-phase measurements can provide information on the structure in solution.

2 MATERIALS AND METHODS

2.1 INSTRUMENTATION

2.1.1 Absolute collision cross section measurements

Absolute CCS measurements were performed on a modified Synapt HDMS (Waters Corp., Manchester, UK) instrument (*52*) in which the travelling-wave ion mobility cell was replaced with an 18 cm drift cell that has a radial RF ion confinement (RF frequency of 2.7 MHz and peak-to-peak amplitude of 200 V) and a linear voltage gradient to direct ions along the axis of transmission to the ToF mass analyser (Figure 2.1). This cell uses no travelling waves, but was fabricated in much the same way as the travelling-wave ion mobility cell used in the commercial instrument (*52*). The drift cell consists of 122 gold-coated circular electrodes that have an internal diameter of 7 mm, are 0.5 mm thick, and have a centre-to-centre spacing of 1.5 mm. These electrodes are connected to printed circuit boards that have voltage divider networks composed of 10 MΩ surface-mount resistors. Potentials of the electrodes in this cell reduce in constant decrements from the entrance to exit electrodes; the difference in potentials between the entrance and exit electrodes is referred to as the drift voltage. The cell is enclosed in a 185 mm housing composed of two printed circuit board side plates, aluminium top and bottom plates, and stainless-steel end plates with 1.0 mm orifices for ion transmission.

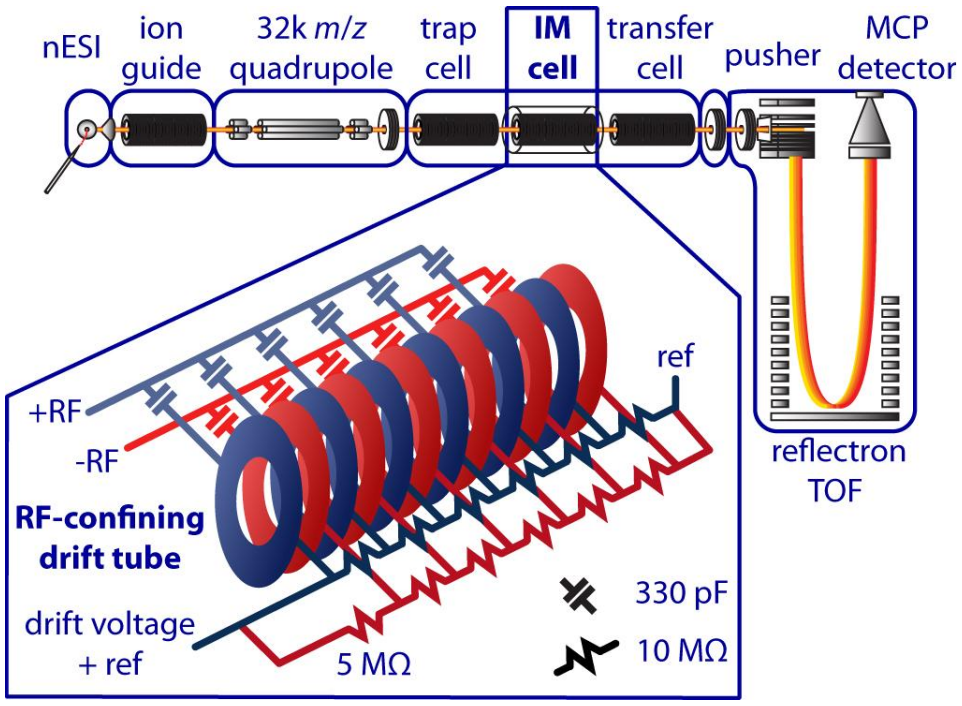


Figure 2.1. Modified quadrupole/ion mobility/time-of-flight hybrid mass spectrometer used for absolute collision cross section measurements. The ion mobility cell in this instrument has a linear voltage gradient along the axis of ion transmission and RF confinement in the radial direction (RF-confining drift tube). In contrast to the standard Synapt HDMS, travelling waves are not used in this ion mobility cell.

CCS values were determined directly from the slopes of drift time *versus* reciprocal drift voltage plots (*90*), using drift voltages ranging from 50–200 V in 2–2.5 Torr of either helium or nitrogen gas. Typically, drift times were measured at 8-10 drift voltages, plotted as a function of reciprocal drift voltage, and yield R^2 correlation coefficients close to 1. Drift cell field energies for these experiments ranged from < 3.1–17 Td, values which are significantly less than the low-field limit reported (60–140 Td) for a large set of peptide ions (*91*); the low-field limit for native-like protein and protein complex ions is even greater (*48-49*). Arrival time distributions for single ion species reported typically yield resolutions of 15–20 (calculated using $t_D/\Delta t_D$, where Δt_D is the full width at half maximum, FWHM). These values are slightly greater than those reported in some

travelling-wave ion mobility experiments on similar instruments, consistent with the use of higher mobility gas pressures (enabled by smaller IM cell apertures).

Pressures were measured using a calibrated absolute pressure transducer (MKS Baratron Model 626A, Wilmington, MA) that is connected directly to the IM cell. Temperatures were measured using a calibrated thermocouple (Omega Engineering Ltd., Manchester, UK) which was installed directly on the outside of the drift cell. The amplitude of the RF in the drift tube was 200 V for all measurements. For protein complexes, instrument conditions were optimised for high *m*/*z* ions (*5, 92*), with typical capillary voltages of 1.4-1.6 kV and pressures of 4-6 mbar in the source transfer region. Ion activation conditions, which depend strongly on the sample cone voltage, the injection voltage into the trap cell, and the injection voltage into the IM cell, were minimised to conditions near the threshold for ion transmission. These values were typically 15-20 V, 5-10 V, and 20 V, respectively. Higher cone voltages were required to observe the highest-mass protein complexes. For native-like complexes, low collision energies were used (5-10 V); for collision-induced unfolding and dissociation experiments, the collision energy was increased (40-180 V). All spectra were calibrated externally using caesium iodide (100 mg mL^{-1}). Unless stated otherwise, experimental CCS values are measured in helium, and are quoted as an average of all charge states and replicate measurements. CCS error is given as the greater of the standard deviation of measurements, or 3 %.

2.1.2 Travelling-wave ion mobility

Travelling-wave ion mobility experiments were performed on standard Synapt G2 HDMS (Waters Corp., Manchester, UK) equipped with a travelling-wave ion mobility cell containing nitrogen gas (Figure 2.2) (*51-52, 93*). In the second generation TWIMS, the mobility separator consists of a gas-filled stacked ring ion guide (SRIG), with opposite phases of RF applied to adjacent ring electrodes, providing radial confinement. In the presence of a gas, ions are propelled through the device by the use of superimposing DC voltages on the RF. This is applied to two pairs of adjacent electrodes, and stepped though the device in single plate pair steps. The helium cell consists of four RF-only ring electrodes, and is ~7 mm in length. The TWIMS device consists of 168 electrodes and is 254 mm in length. Each plate is 0.5 mm thick and spaced by 1 mm. Immediately before the helium cell, and after the TWIMS cell, lie the respective trap and transfer travelling-wave ion guides (TWIGs). No travelling-wave is used in the trap TWIG: the ions are trapped by RF voltages and periodically gated to allow release into the mobility separator region. The transfer TWIG has a continuous high amplitude wave present to maintain mobility separation prior to analysis on the ToF. Both regions are connected to the main vacuum system, and are pressurised with Ar gas (~10^{-2} mbar).

Typical instrument settings used include elevated source transfer region pressure (4 - 6 mbar), capillary voltage 1.4 - 1.6 kV, sample cone voltage 15 V, trap/transfer collision energy 0-5 V. Unless stated otherwise, nitrogen and helium flow rates were 60 mL min^{-1} and 180 mL min^{-1} respectively, resulting in measured pressures of ~3 Torr. Injection voltages into the He cell, TWIMS cell and transfer region were lowered to the threshold for ion transmission in order to minimise ion heating effects (*94*). Wave velocities and heights were varied as described in the text.

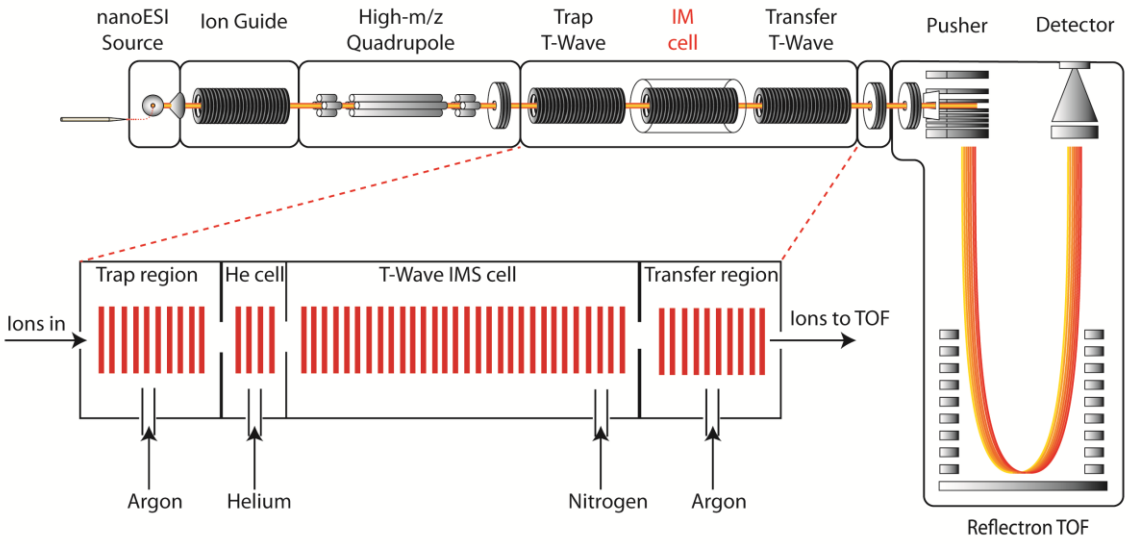


Figure 2.2. Travelling-wave ion mobility mass spectrometer. Schematic showing the Synapt G2 HDMS, a quadrupole/ion mobility/ToF hybrid instrument. The tri-wave region, consisting of the trap region, helium cell, travelling-wave IMS cell and transfer region, is shown in more detail.

Travelling-wave drift times were calibrated using a modified form of a protocol reported previously (*92*). Observed drift times (t_D) are corrected for *m*/*z* dependent flight times:

$$t'_D = t_D - c\sqrt{m/z} \tag{2.1}$$

The constant *c* is determined empirically as described previously (*92*).

Literature CCS in helium are corrected for charge, *z* and reduced mass of the ion and drift gas, μ:

$$\Omega' = \frac{\Omega_{literature}\sqrt{\mu}}{z} \tag{2.2}$$

An exponential factor (X) is determined from the slope of the plot of ln(Ω') as a function of ln(t'_D). Calibrated CCS_{He} values are determined by plotting literature CCS_{He} as a function of the final corrected drift times (t''_D):

$$t''_D = t'^{X}_D \left(\frac{z}{\sqrt{\mu}}\right) \tag{2.3}$$

2.1.3 Charge state distribution parameters

Charge state distribution (CSD) parameters Z_{max} and Z_{min} refer to the minimum and maximum charge states observed in MS. Z_{av} and Z_W, the average charge state and the CSD width at half-maximum height, were calculated by Gaussian fitting of the charge state envelope (*28*) (Figure 2.3).

In order to calculate the Z_{av} for multiple overlapping CSDs, the Z_{av} for each distribution can be found by fitting multiple Gaussian distributions. An overall average charge state can be calculated using Eq. 2.4, where I_i is the intensity of the *i*th charge state, Z_i is the net charge on the *i*th charge state, and N is the total number of charge states observed (*24*).

$$Z_{av} = \frac{\sum_{i=1}^{N} Z_i I_i}{\sum_{i=1}^{N} I_i} \tag{2.4}$$

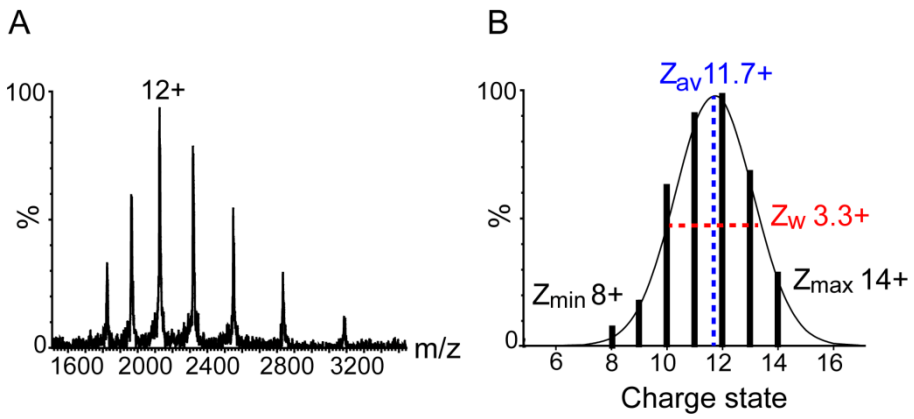


Figure 2.3. MS charge state distribution for SAP monomer (**A**) and corresponding Gaussian fitting (**B**). Parameters Z_{av}, Z_{min}, and Z_{max} are defined as the average, minimum and maximum charge states in the CSD respectively. Z_W is the width of the CSD at half-maximum height.

2.2 SAMPLE SOURCES AND PREPARATION

2.2.1 Peptides, proteins and homomeric complexes

Ions are formed by nESI of analyte-containing solution (1-10 μM) using gold-coated borosilicate micropipettes prepared in-house. Sample details are given in Table 2.1.

Table 2.1. Sample preparation for peptides, proteins, and homomeric complexes

Sample	Supplier	Item	Protocol
Gly-Arg-Gly-Asp-Ser	Sigma	G4391	DN
Ser-Asp-Gly-Arg-Gly	Sigma	S3771	DN
additional peptides	Waters	186002337	DN
ubiquitin	Sigma	U6253	DN
cytochrome *c*	Fluka	30396	DN
			Buffer
myoglobin	Sigma	M1882	DN
			Buffer
β-lactoglobulin A	Sigma	L7880	Buffer
transthyretin (TTR)	Sigma	P1742	Buffer, BS
avidin	Sigma	A9275	Buffer
bovine serum albumin	Sigma	P7656	Buffer
concanavalin A	Sigma	C2010	Buffer, BS
serum amyloid P (SAP)	CalBioChem	565190	Buffer, BS
alcohol dehydrogenase (ADH)	Sigma	A7011	Buffer, BS
pyruvate kinase	Sigma	P9136	Buffer, BS
glutamate dehydrogenase	Sigma	G7882	Buffer, BS
β-galactosidase	Sigma	G5635	Buffer, BS twice
GroEL	Sigma	C7688	Refolded[a], Buffer, BS twice

DN = Samples prepared in water/methanol/acetic acid (49/49/2) solutions.

Buffer = Samples prepared in 200 mM AA solution.

BS = On day of analysis, sample buffer exchanged using a Micro Bio-Spin 6 column (Bio-Rad, Hercules, CA) that has been equilibrated with 200 mM AA.

[a] Ref (*5*)

2.2.2 Heteromeric complexes

Samples for heteromeric complexes were received from collaborators (Table 2.2).

Table 2.2 Sample preparation details for heteromeric complexes

Sample	PDB	Expression/ Purification	Provided by	Storage buffer
tryptophan synthase *Salmonella typhimurium*	1WBJ	Ref (*95*)	I.Schlichting Max Planck Institute, Heidelberg	50 mM bicine pH 7.8 10 mM EDTA 1 mM DTE 0.02 mM pyridoxal phosphate
acetyl coA carboxylase carboxyltransferase *Escherichia coli*	2F9Y	Ref (*96*)	G. Waldrop Louisiana State University	10 mM HEPES pH 7.0 500 mM KCl
nitrobenzene dioxygenase (NBDO) *Comamonas sp. JS765*	2BMO	Ref (*97-98*)	R. Friemann University of Uppsala	50 mM bis-Tris pH 6.8 5 % glycerol 1 mM DTT trace ammonium sulfate
methane monooxygenase hydroxylase (MMOH) *Methylococcus capsulatus*	1MTY	Ref (*99*)	S.J. Lippard Massachusetts Institute of Technology	25 mM MOPS pH 7.0
toluene/*o*-xylene monoxygenase hydroxylase (ToMOH) *Pseudomonas stutzeri*	2INC	Ref (*99*)	S.J. Lippard Massachusetts Institute of Technology	25 mM MOPS pH 7.0
carbamoyl phosphate synthetase *Escherichia coli*	1BXR	Ref (*100*)	F. Raushel Texas A&M University	100 mM phosphate buffer pH 7.6 10 mM ornithine

NBDO, MMOH, ToMOH, acetyl coA carboxylase carboxyltransferase, tryptophan synthase and carbamoyl phosphate synthetase were buffer exchanged into 150-250 mM AA, pH 7, using pre-equilibrated ultracentrifugation filter devices (Amicon-Ultra 05, MWCO 10 kDa or Micro Bio-Spin-6). Ornithine (1 μM), manganese acetate (34 μM) and AMP-PNP (23 μM) were added as co-factors to carbamoyl phosphate synthetase. The buffer-exchanged heteromeric complexes were observed by MS under native-like conditions and where possible, subcomplexes were observed in buffer-only conditions. The complexes were then disrupted in solution using changes in ionic strength, pH, or by adding organic solvents to the sample. Subcomplexes were generated using 25 % butanol (NBDO), 4 M AA (ToMOH), 30 % MeOH/10 % DMSO (tryptophan synthase), 37 % acetonitrile (MMOH), 30 % MeOH/0.01 % acetic acid or 35% MeOH/1% ammonium hydroxide (acetyl coA carboxylase carboxyltransferase), and 25% DMSO (carbamoyl phosphate synthetase).

2.2.3 ΔN6-β2-microglobulin[1]

ΔN6-β2-microglobulin (ΔN6-β2m) was expressed in *E. coli* and purified using ion exchange chromatography and size exclusion chromatography (*101*). Fibrils of ΔN6-β2m were prepared by incubating 100 μM of the purified protein at 37 °C in 50 mM acetate buffer, pH 5.0 and 100 mM NaCl. The quality of the fibrils was analysed by negative stain electron microscopy. Protein samples (5 μL) were applied on carbon-coated grids (Formvar/carbon on 400 Mesh Copper), and negatively stained with 1% (wt⁄vol) uranyl acetate. Micrographs were recorded on a Jeol Jem-2100 electron microscope at 200 kV. Prior to IM-MS the lyophilised ΔN6-β2m was dissolved in 100 mM AA, pH 5.0.

[1] ΔN6-β2m expression and purification, and the electron microscopy of fibrils was performed by K. Domanska, Vrije University Brussels.

2.2.4 p53 constructs[2]

p53 constructs containing the core domain (C, CT and FL) were modified at four residues (M133L/V203A/N239Y/N268D) (*102*-*103*) and cloned into pET-24a with an N-terminal 6×His/lipoamyl domain tag and tobacco-etch-virus (TEV) protease cleavage-site. Each cloned plasmid was transformed into *E. coli* BL21(DE3) and expressed and purified following published protocols (*104*). The tetramerisation domain (residues 325-355) was produced as described recently (*105*). The cDNA clones were purchased from Geneservice Ltd. (Cambridge, UK) and transformed in *E. coli* BL21 cells using a pRSET vector without any tag. Cells were grown and induced at 37 °C. The construct was diluted (1-10, in 25 mM Tris buffer pH 7.5) loaded on anion exchange chromatography (Q Sepharose) and eluted protein loaded on a HiLoad 26/60 Superdex 75 column (GE Healthcare, UK) using phosphate buffer (25 mM phosphate buffer pH 7.2, 300 mM NaCl). The DNA used is the p53 response element (MW 1594 Da, 5' CGCGGACATGTCCGGACATGTCC-CGC'3). Complexes were buffer exchanged into 200-500 mM AA, using Micro Bio-Spin 6 columns (Bio-Rad, UK) or Slide-A-Lyzer dialysis cassettes (Fisher Scientific, UK). Protein sequences for the constructs are given below (red = tetramerisation domain T, dark blue = core domain C, cyan = disordered regions; FL = full length).

T (325-355)

EYFTLQIRGRERFEMFRELNEALELKDAQAG

C (94-312)

SSVPSQKTYQGSYGFRLGFLHSGTAKSVTCTYSPALNKLFCQLAKTCPVQLWVDSTPPPG
TRVRAMAIYKQSQHMTEVVRRCPHHERCSDSDGLAPPQHLIRVEGNLRAEYLDDRNTFR
HSVVVPYEPPEVGSDCTTIHYNYMCYSSCMGGMNRRPILTIITLEDSSGNLLGRDSFEVR
VCACPGRDRRTEEENLRKKGEPHHELPPGSTKRALPNNT

[2] Protein expression and purification for p53 constructs was carried out by Dr Eviatar Natan, University of Cambridge.

CT (94-360)

SSVPSQKTYQGSYGFRLGFLHSGTAKSVTCTYSPALNKLFCQLAKTCPVQLWVDSTPPPG
TRVRAMAIYKQSQHMTEVVRRCPHHERCSDSDGLAPPQHLIRVEGNLRAEYLDDRNTFR
HSVVVPYEPPEVGSDCTTIHYNYMCYSSCMGGMNRRPILTIITLEDSSGNLLGRDSFEVR
VCACPGRDRRTEEENLRKKGEPHHELPPGSTKRALPNNTSSSPQPKKKPLDGEYFTLQIR
GRERFEMFRELNEALELKDAQAGKEPG

FL(1-393)

MEEPQSDPSVEPPLSQETFSDLWKLLPENNVLSPLPSQAMDDLMLSPDDIEQWFTEDPGP
DEAPRMPEAAPPVAPAPAAPTPAAPAPAPSWPLSSSVPSQKTYQGSYGFRLGFLHSGTAK
SVTCTYSPALNKLFCQLAKTCPVQLWVDSTPPPGTRVRAMAIYKQSQHMTEVVRRCPH
HERCSDSDGLAPPQHLIRVEGNLRAEYLDDRNTFRHSVVVPYEPPEVGSDCTTIHYNYM
CYSSCMGGMNRRPILTIITLEDSSGNLLGRDSFEVRVCACPGRDRRTEEENLRKKGEPHH
ELPPGSTKRALPNNTSSSPQPKKKPLDGEYFTLQIRGRERFEMFRELNEALELKDAQAGK
EPGGSRAHSSHLKSKKGQSTSRHKKLMFKTEGPDSD

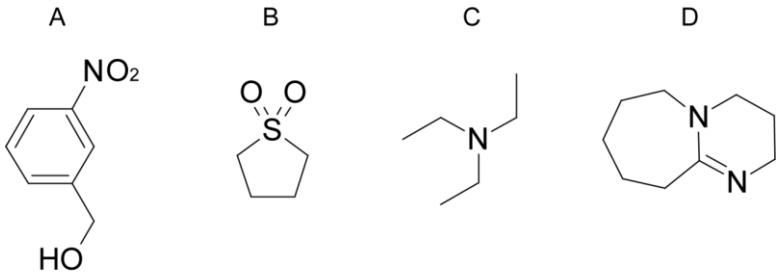
2.3 REAGENTS

2.3.1 Co-factors and chaotropes

L-ornithine, manganese acetate, AMP-PNP and high purity solvents DMSO, methanol, acetonitrile and butanol were obtained from Sigma Aldrich (St Louis, MO, USA). Acetic acid and ammonium hydroxide were obtained from Sigma Aldrich (St Louis, MO, USA) and were of high purity (Fluka, TraceSelectUltra).

2.3.2 Charge manipulation

AA solution, triethylammonium acetate buffer (TEA), 1, 8-diazabicycloundec-7-ene (DBU), imidazole, *m*-nitrobenzyl alcohol (*m*-NBA) and sulfolane were obtained from Sigma Aldrich (St Louis, MO, USA). Charge reduction was carried out by addition of TEA, imidazole or DBU (10 – 20 mM). Supercharging was achieved by the addition of 1 % (*v/v*) *m*-NBA or 7 % (*v/v*) sulfolane. Structures and important physicochemical properties for these reagents are shown below (Figure 2.4).



	Boiling point (°C)	Gas phase basicity (kJ/mol)
DBU	80-83	1015
Triethylamine	89	951
Ammonia	-33	819
Water	100	660
m-NBA	175-180	800
Sulfolane	285	801

Figure 2.4. Charge manipulation reagents. Supercharging is carried out using low volatility reagents such as *m*-NBA (**A**) and sulfolane (**B**). Charge reduction is achieved through additives with high gas phase basicity, including TEA buffer (**C**) and DBU base (**D**) (*106-107*).

2.4 X-RAY CRYSTAL STRUCTURE CALCULATIONS

2.4.1 Volume, area and salt bridges

The volumes and surface areas of all-atom structures were determined using the program MSMS (*108*), which utilises a rolling probe, radius 1.4 Å, to calculate the solvent accessible surface area and volume. PISA (*109*) was used to calculate interface areas and interfacial salt bridges. Salt bridges were permitted between basic (Lys, Arg) and acidic (Glu, Asp) residues, with a bond distance < 4 Å. Many computational methods are available to detect internal pockets and cavities in proteins (*110-112*). The volumes of the cavities in the complexes were estimated using the Roll algorithm (*113*), implemented in POCASA (rolling sphere probe radius 1 Å, grid size 1 Å) available at http://altair.sci.hokudai.ac.jp/g6/service/pocasa/. The cavity volume was reported as a percentage of the solvent excluded volume of the complex. Gap volume indices were calculated using SURFNET (*114*) on http://www.biochem.ucl.ac.uk. Structures were visualised using VMD (*115*).

2.4.2 Theoretical CCS calculations

Atomic coordinates were obtained from the protein data bank (PDB). All calculations include hydrogen atoms, which were added explicitly using psfgen in the molecular graphics program VMD (*115*). In order to perform all-atom CCS calculations for subcomplexes, subunits were manually removed from the coordinate file of the X-ray structure for the intact complex. Helium CCSs were calculated using the PA and EHSS algorithms implemented in MOBCAL (*57-58*).

Globular complexes: Although the PA method is known to underestimate the experimental CCS by neglecting multiple collisions between the ion and buffer gas (*57, 60*), it is highly correlated with experimental CCS for protein complexes (R^2> 0.99) (*116*). Therefore in Chapters 4-6, a scaled CCS_{PA} (Eq. 2.5) is reported for comparison with CCS from IM, whereby the experimental CCS can be predicted (± 3%) by multiplying the CCS_{PA} by a factor of 1.14 (see Chapter 3.2.4 for derivation).

$$CCS_{calc} = 1.14\, CCS_{PA} \left(\frac{m_{exp}}{m_{pdb}}\right)^{2/3} \quad (2.5)$$

In Eq. 2.5, m_{exp} is the experimentally determined mass and m_{pdb} is the mass calculated from the crystal structure. This scales the CCS for any missing atoms or residues, assuming spherical growth.

p53 constructs: The scaled PA CCS estimate described above was derived from globular proteins and their complexes, and does necessarily apply to intrinsically disordered proteins, whose structures may behave differently in the gas phase. Therefore experimental CCS for p53 constructs are compared with theoretical CCSs calculated

using the both PA (*55, 117*) and EHSS (*57*) algorithms, paying greater heed to relative CCS changes, rather than the absolute values. The following coordinates from the PDB were used: $(T)_4$, 1PES; C, 2FEJ; $(C)_4$DNA, 2AC0. The p53 segments for which no high-resolution structural information is available, $(CT)_4$, $(CT)_4$DNA, $(FL)_4$, were represented using coordinates from proposed models derived from a combination of X-ray, NMR, EM and SAXS data (*118*).

2.5 COARSE-GRAINED MODELLING

2.5.1 CCS and volume calculations

Helium CCS were calculated using the PA algorithm implemented in a version of MOBCAL (*57-58*) adapted in-house for coarse-grained models (*92*) and scaled using Equation 2.5 for comparison with experimental CCS. Several methods are available to calculate CCS for atomic structures (*55, 57, 60*); the PA method (*55*) is the only one appropriate for calculating CCS of coarse-grained models based on spheres. Volumes of coarse-grained models were determined using a Monte Carlo approach. The contribution of the buffer gas to volume was accounted for in all volume calculations, by subtracting the radius of the buffer gas (r_{He} ~ 0.1 nm) from spherical radii in the models. Coarse-grained models were visualised using VMD.

2.5.2 Coarse-grained modelling approach

Proteins were represented as spheres, and protein complexes and their subcomplexes as a set of overlapping spheres, with radius r defined using the subunit CCS, $r = (\Omega/\pi)^{1/2} - r_{He}$ or from the subunit mass (M) and density (ρ), $r = (3M/4\pi\rho)^{1/3}$. Adequate sampling is critical for this modelling approach. When building models without using atomic

information, for ≤ 3 subunits, the coordinates of the spheres were generated manually using in-house python scripts. These sampled many different relative orientations and conformations, in a planar search. When more degrees of freedom were present (>3 subunits), the sampling of model structures was obtained within the Integrative Modelling Platform by incorporating a Monte Carlo conjugate gradient approach (*119*). This makes use of Monte Carlo steps followed by a conjugate gradient optimisation to search for candidate structures. All degrees of freedom are considered rendering this approach unbiased towards any particular solution.

2.5.3 Ultrafast shape recognition

Candidate structures are scored (Chapter 4) and clustered based on their structural similarities. To identify structural similarities between coarse-grained models a shape descriptor, the Ultrafast Shape Recognition (USR) algorithm, is used (*120*). This approach is based on distribution-based descriptors (*121*), and was developed previously for fast screening of compound and ligand databases (*122*). In this work, it was adapted for use in coarse-grained structures[3]. It uses the four moments of distribution (mean, variance, skewness and kurtosis) to quantify shape similarities between models. The four distributions of distances were generated by selecting four locations in model structures: the centroid, the closest subunit to the centroid, the farthest subunit to the centroid (*y*) and the farthest subunit to *y*. The four moments for each distribution were calculated and compared with those from the native/reference structure. The level of similarity was quantified using a scoring function (S_{USR}):

$$\boldsymbol{S_{USR}} = \frac{\mathbf{1}}{\mathbf{1}+\frac{\mathbf{1}}{\boldsymbol{N_{sph}}}\frac{\mathbf{1}}{\boldsymbol{N}}\sum_{\boldsymbol{n}=\mathbf{1}}^{\boldsymbol{N}}\left|\boldsymbol{M_n^{Ref}}-\boldsymbol{M_n^{i}}\right|} \tag{2.6}$$

[3] USR algorithm and IMP sampling codes were adapted for use with coarse-grained structures by Argyris Politis, University of Oxford

where S_{USR} is the USR score, N_{sph} are the number of spheres in the structure, N is the number of moments of distribution used, M^{Ref} and M^{i} denote the moments of distribution for the reference and model structures respectively. This function takes scores from 0 to 1, where 1 corresponds to maximum similarity and 0 to minimum. For the purpose of statistical analysis of restraints, a "native-like" model was defined as that with S_{USR} > 0.95 relative to the native structure (Figure 2.5).

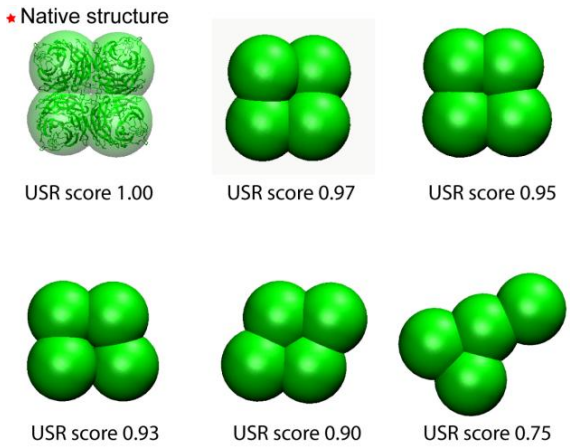


Figure 2.5. Models of a homotetramer with associated USR scores relative to the native structure.

2.5.4 Hierarchical clustering

Candidate model solutions are analysed in terms of their assembly features, including the overall shape, subunit positions and contacts (*123*). The variability of these structures is assessed by hierarchical clustering (*124*). The hierarchical clustering assigns set of structures into groups where structures in the same cluster are more similar to each other than structures within other groups. The 1% lowest scoring models are hierarchically clustered with MATLAB (The Mathworks Inc.), based on USR score for all pairs of structures, thereby identifying discrete clusters of structures with similar geometric features. Clusters were defined using a cut-off of > 0.05 (± 0.01) USR score. Each identified cluster is represented by the model with the lowest overall score.

2.6 MOLECULAR DYNAMICS SIMULATIONS

2.6.1 Simulation method

The simulations were run in double floating-point precision, with GROMACS (*125*) 4.5.3. Pentameric SAP (PDB 1SAC) was simulated in a solvent-free system using the OPLS-AA/L force field (*126*). The simulation method was based on a previously published procedure (*66*) for simulating proteins *in vacuo*. A steepest descent energy minimisation was performed, followed by 1 ns vacuum equilibration simulations, with randomly generated initial velocities. Initial conditions for the 10 ns vacuum simulations were taken from the final structures and velocities of the 1 ns equilibration simulations. Neither periodicity nor cut-offs were employed in the calculations, as is standard for vacuum simulations. Energy conservation was achieved using a 1 fs integration step, and constraining bonds to hydrogen with the LINCS algorithm (*127*). A dielectric constant of $\varepsilon = 2\varepsilon_o$ (where ε_o is the dielectric permittivity of free space) was found to be appropriate to describe the gas-phase basicity of cytochrome *c (63)* and was used here. 10 ns production simulations were run over a linear temperature gradient of 300–800 K. As a control, 10 ns simulations were also performed at constant temperature (300 K).

Several methods have been used to determine the protonation states for the titratable groups (*128-129*) of protein ions, most of which do not take into account the solvent accessibility of the residues. It is well known that the charge states of protein ions observed during MS are highly correlated to the solvent accessible surface area (*130*). Solvent accessibility is closely related to residue depth (*131*). Therefore a search algorithm was employed (see section 2.6.2), which evenly distributed a user-defined number of charges over solvent accessible basic residues (< 5 Å from the surface). In this way, charges were assigned for 18+, 25+ and 30+ SAP. In addition to this, six alternative

configurations were prepared (two for each charge state), with charges located on different residues, including an asymmetric distribution of charges, determined from the ejected monomer in tandem MS experiments. Nine trajectories, for each charge state, were analysed with respect to their radius of gyration (R_g). In addition, CCSs were calculated for structures every 20 ps.

p53 constructs: $(CT)_4$DNA and $(CT)_4$ complexes were simulated in a solvent-free system as above, with the following modifications: the AMBER99 force field (*132*) (parameterised for DNA) was used and following energy minimisation, 1 ns vacuum simulations were carried out at constant temperature (300 K). Three trajectories, for each complex were analysed with respect to their root mean square deviation (RMSD) from the initial structure.

2.6.2 Charge assignment algorithm

In order to assign charges to the protein complex, the solvent accessible basic residues were identified. These were defined as Lys, Arg, and His residues that were <5 Å from the surface. The depth of residues was calculated using http://mspc.bii.a-star.edu.sg/tankp/ (*131*). In order to minimise Coulombic repulsion, and attain an approximately even distribution of charges, an A* search[4] algorithm was used (*133*). This algorithm finds the least-cost path between points, called nodes, using a heuristic function to determine the order in which the search visits nodes in the tree. In this case, the nodes represented the solvent-accessible basic residues, and the least-cost path was

[4] A* search algorithm was implemented for use with atomic structures by Cesar Tron-Lozai, University of Oxford

that which minimised Coulombic repulsion, by maximising the sum of distances between charged residues.

2.6.3 RMSD clustering

In order to identify the most populated conformations in trajectories at different stages, an RMSD-clustering algorithm based on the method of Daura *et al.* (*134*) was used. The general clustering approach is as follows: the backbone RMSD between all pairs of structures was determined at defined intervals in the trajectory. For each structure the number of other structures with backbone RMSD within 0.1 nm was counted. The structure with the largest number of neighbours was taken as the centre of a cluster, and forms a cluster of similar structures with its neighbours. These structures were then removed from the pool of structures, and the process was repeated until the pool of structures was empty.

3 PROTEIN COMPLEXES AS ION MOBILITY STANDARDS: A CALIBRATION FRAMEWORK AND DATABASE FOR GAS-PHASE STRUCTURAL BIOLOGY

3.1 INTRODUCTION

IM is frequently coupled with MS, allowing the interrogation of ion structures (*46, 135-136*). Applications of IM-MS are diverse and range from the analysis of explosives (*137*), phospholipids (*138*), pharmaceuticals (*139*), peptides (*140*), to intact protein assemblies (*82, 84*). Pioneering IM-MS experiments demonstrated that highly distinctive protein architecture could be maintained in the gas phase (*39*). Subsequently, IM-MS has been applied to increasingly complex and challenging biological targets (*141-142*). IM-MS data for Aβ (*143*) and β2-microglobulin oligomers (*144*) have revealed detailed insights into the initial self-assembly of these amyloid forming species. IM-MS is also being used to characterise ever larger biological molecules and has provided insights into the assemblies and topologies of Mega-Dalton virus capsids (*3*), with masses up to 10^6-10^7 Da. An exciting application of IM-MS in the field of structural biology is the use of CCS as restraints to assist structural modelling of protein assemblies (*145-146*).

Applications of IM-MS have increased dramatically since the introduction of commercial TWIMS mass spectrometers (*52*), which have enabled the use of IM-MS by a rapidly-growing user base. Although the transport of ions through drift tubes with weak, uniform electric fields is well understood and characterised (*46, 48, 147*), the transport of ions

through ion mobility cells with travelling waves is less well understood. CCS for ions transported by travelling waves are typically determined by calibrating drift times for ions of interest against those measured for ions of known CCS (*92, 148-149*). One problem with this calibration approach when applied to noncovalent protein complexes, is that the reference CCS values in the literature (*91, 150-151*) that are used for calibration, are based on monomeric denatured proteins (*40, 84, 149*). Thus, the calibrants have significantly higher mobilities and lower masses than the macromolecular assemblies under scrutiny. As a result, large extrapolations are needed to reach the required mass range, which can result in the derivation of erroneous CCS (*50*). There is great interest, therefore in obtaining new calibration frameworks that are more relevant to protein complexes and large assemblies.

In this chapter, CCS measurements both in helium and nitrogen gases are reported for a large set of biomolecular ions, including denatured peptides and proteins, and native-like proteins and protein complex ions. CCS in these experiments were determined directly from drift times measured in a low-pressure (~2 Torr) drift tube with RF-confinement, without the need for mobility calibration. These ions have masses from 0.5 to 801 kDa and CCS from 1 to >200 nm^2, thus are ideal for calibrating travelling-wave ion mobility results for a wide range of ions. Access to this information enables precise quantification of the errors associated with travelling-wave CCS measurements. From this new database and error analysis, a set of best practices are proposed for CCS calibration. Furthermore, the instrumental parameters critical for CCS accuracy in travelling-wave IM are discussed. Overall, these results enable higher accuracy CCS measurements, allowing the generation of more accurate 3D models of biomolecular assemblies, such as protein complexes, using gas-phase structural biology techniques.

3.2 RESULTS

3.2.1 Characterising the linear drift tube

Drift times in these experiments were measured using a modified Synapt HDMS instrument in which the travelling-wave ion mobility cell was replaced with a drift tube, with radial RF-confinement and an axial linear voltage gradient (described fully in Chapter 2). In an effort to characterise the accuracy of the linear drift tube, drift times for avidin 16+ (4-mer, 64 kDa) were measured at 10 drift voltages, ranging from 20–200 V, in 2.0 Torr of helium gas (Figure 3.1A). Linear regression of the drift time *versus* reciprocal drift voltage data result in R^2 value of 0.998, however analysis of the residuals (observed - predicted drift time) revealed non-random scatter, indicating higher-order effects (*152*). At low drift voltages, there is considerable deviation towards longer drift times than expected. These effects increase with decreasing drift voltage. In order to understand these effects in greater detail, the Cook's distance test (*152-153*) was applied. This test gives an estimate of the influence of a data point when performing least squares regression analysis. Data points with large residuals and/or high leverage (Cook's distance $D_i > 1$) may distort the outcome and accuracy of a regression analysis.

$$D_i = \frac{\sum_{j=1}^{n}\left(\hat{Y}_j - \hat{Y}_{j(i)}\right)^2}{p\ MSE} \tag{3.1}$$

where D_i is the Cook's distance, $\hat{Y}_j$ is the prediction from the full regression model for observation j, $\hat{Y}_{j(i)}$ is the prediction for the observation j from a refitted regression in which observation i has been excluded, *MSE* is the mean square error of the regression model, p is the number of fitted parameters in the model.

Data measured at the lowest drift voltage (20 V) was shown to exhibit significant leverage, compared with those drift times measured at higher voltages (Figure 3.1B). Re-testing following the removal of this data point resulted in the drift time measured at the second lowest drift voltage exhibiting strong leverage, and so on. No statistical evidence for these effects however was observed for data obtained at drift voltages > 50 V.

IM-MS of avidin 16+ was therefore repeated excluding all data obtained < 50 V. Linear regression of drift time *versus* reciprocal drift voltage resulted in R^2 value of 0.9998, with very small and randomly scattered residuals (Figure 3.1). The exclusion of data collected at low drift voltages, resulted in 4 % decrease in the slope, and corresponding mobility. Similar results were obtained for other protein complexes, with improvements in mobility (and therefore CCS) accuracy of up to 10 %, by taking measurements only at drift voltages > 50 V.

Somewhat analogous effects have been observed in ion funnels following drift tubes (*90*). In this case, the non-linear effects arise because the extent of transverse diffusion in the tube increases with lower drift voltages, resulting in the average displacement of ions off-centre. This results in slightly longer trajectories in the ion funnel and longer apparent drift times.

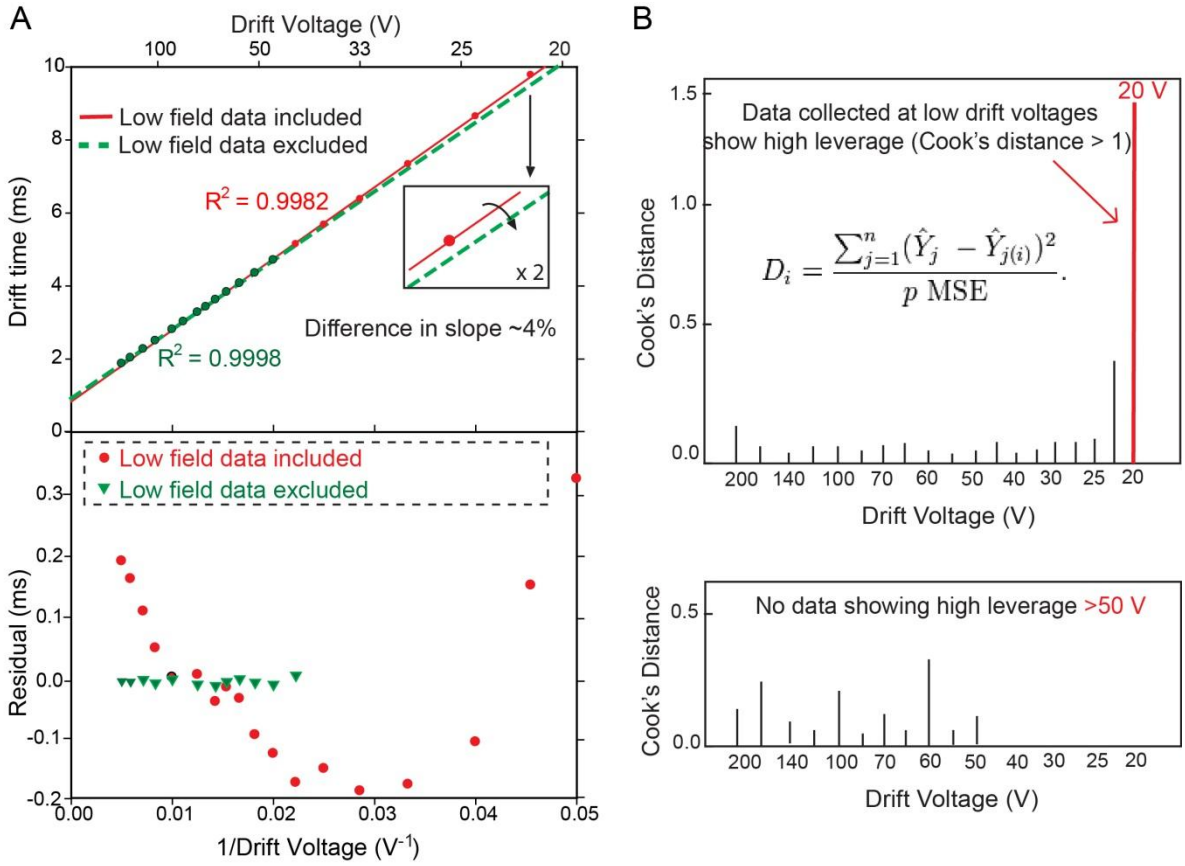


Figure 3.1. Non-linear behaviour observed for ions at low fields. **A** Drift times for avidin 16+ were measured over drift voltages 20-200 V (red) and 50-200 V (green). Linear regression of the data results in high correlation coefficients (R^2 > 0.99). Analysis of the residuals (observed – expected drift time) showed significant higher-order effects when low field data (< 50 V) was included. **B** The Cook's distance test showed the data acquired at the lowest drift voltages exhibited significant leverage (Cook's distance > 1), distorting the accuracy of the regression analysis. No statistical evidence for leverage was found for data collected at drift voltages > 50 V. Removal of the low field data improved the accuracy of the slope, and associated mobility value, by 4 %.

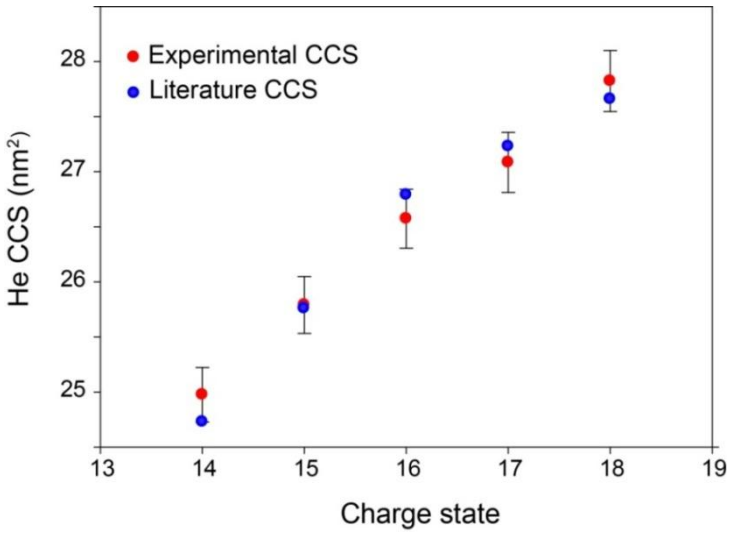


Figure 3.2. Comparison with literature values. CCS for denatured cytochrome *c* (49/49/2 water/methanol/acetic acid), measured on the RF-confining drift tube, and those reported previously (*150*) show good agreement. Error bars reflect measurement precision (±1 %).

In order to validate this CCS measurement approach, CCS were measured for cytochrome *c* using the RF-confining drift tube and compared with literature values (*150*) (Figure 3.2), with excellent agreement.

3.2.2 Collision cross section database

Having established the optimum conditions for measuring drift times, mobilities in helium and nitrogen gases were measured for a wide range of peptide, protein and protein complex ions. The charge states reported in Tables 3.1 and 3.2 are readily achieved using nESI from the respective solutions, and have single arrival time distributions (or one predominant distribution) under the gentle (minimal activation) conditions used in these experiments. Gentle experimental conditions, near the threshold for ion transmission, are desirable for these experiments to mitigate the effects of structural collapse, structural unfolding, and other transitions that have been observed for native-like protein complex ions that have been activated in the gas phase (*40, 154-155*).

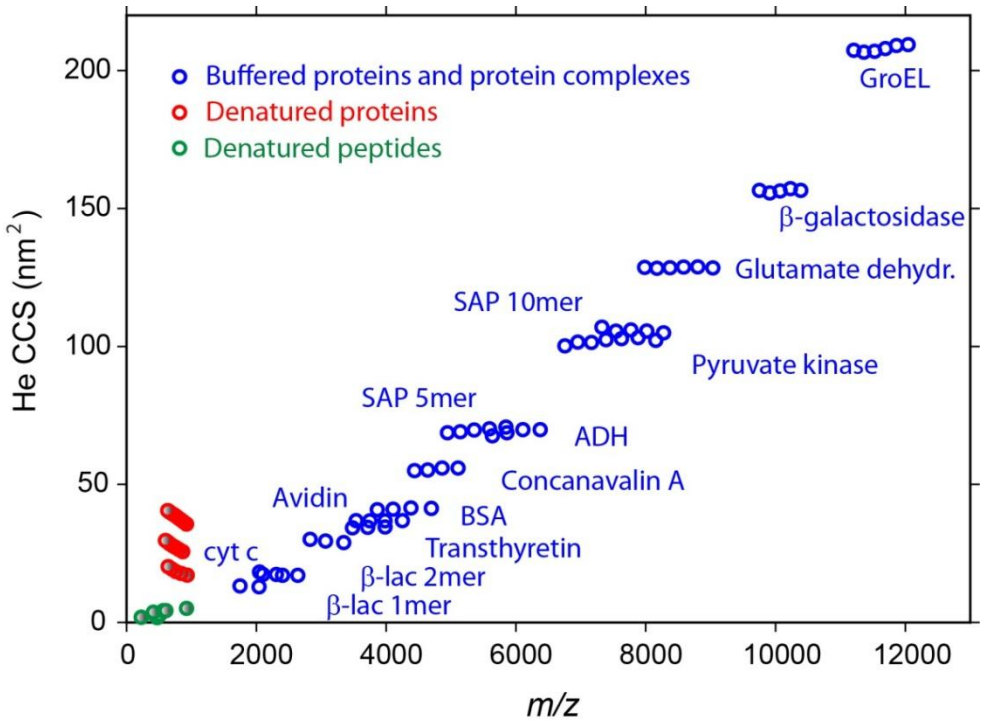


Figure 3.3. Collision cross sections in helium gas for the observed charge states of denatured peptide, denatured protein, native-like protein, and native-like protein complex ions as a function of *m*/*z*.

CCS values are reported in Tables 3.1 and 3.2, and results in helium gas are also shown in Figure 3.3. As expected, CCS_{N2} are consistently greater than CCS_{He}. This is due to the greater polarisability of nitrogen gas, which results in the long-range attractive potential having greater effects on CCS (*156*). CCS values are the average of three replicate measurements of mobility, each made on separate days using ten drift time measurements. Errors in the slopes of individual drift time versus reciprocal drift voltage plots are typically 0.5 – 1.0 %, but the standard deviation of measurements performed over long periods of time can be slightly larger (typically 0.5 – 2 %). Additional sources of error include pressures, which were measured using a calibrated absolute pressure transducer (manufacturer reported accuracy 0.25 %), and the temperature in the ion mobility cell, measured using a calibrated thermocouple with probes attached to either end of the drift tube exterior (error ≈ 0.5 %). Therefore, the errors for the CCS in this database are estimated to be < 3 %.

Table 3.1. Collision cross sections (Ω) for denatured ions[5]

	m / Da	Z	Ω_{He} / nm^2	Ω_{N2} / nm^2
GRGDS	490	1	1.32	2.06
		2	1.39	2.56
SDGRG	490	1	1.30	2.04
		2	1.42	2.59
angiotensin fragment 1-7	898	2	2.26	3.34
RASG-1	1 000	2	2.25	3.31
angiotensin II	1 046	2	2.45	3.35
bradykinin	1 060	2	2.37	3.44
angiotensin I	1 296	3	3.28	4.74
renin substrate	1 758	3	3.80	5.22
enolase T35	1 872	3	3.80	5.19
enolase T37	2 827	3	4.65	-
ubiquitin bovine erythrocytes	8 k	7	-	19.1
		8	-	19.9
		9	16.7	20.9
		10	17.3	22.0
		11	18.0	23.4
		12	18.9	24.8
		13	19.8	26.0
cytochrome *c* equine heart	12 k	13	-	30.8
		14	25.2	32.0
		15	26.0	33.3
		16	26.7	34.5
		17	27.4	36.0
		18	28.0	36.7
		19	28.7	37.9
		20	29.2	-
***apo* myoglobin** equine heart	17 k	15	-	40.6
		16	-	41.8
		17	-	43.1
		18	35.2	44.4
		19	36.0	45.7
		20	36.8	47.0
		21	37.5	48.2
		22	38.2	49.2
		23	38.7	50.1
		24	39.2	50.9
		25	39.6	-
		26	40.0	-

[5] CCS measurements in Tables 3.1 and 3.2 were performed by Zoe Hall and Matthew Bush (University of Oxford)

Table 3.2. Collision cross sections (Ω) for native-like ions[5]

	n	*m* / Da	*z*	Ω_{He} / nm^2	Ω_{N2} / nm^2
cytochrome *c* equine heart	1	12 k	6	12.4	14.9
			7	12.8	15.9
***apo* myogloblin** equine heart	1	17 k	7	16.6	19.7
			8	16.7	-
***β*-lactoglobulin** bovine milk	1	18 k	7	16.6	19.5
			8	16.9	20.3
			9	17.8	-
	2	37 k	11	28.5	32.3
			12	29.0	33.1
			13	29.6	34.3
TTR human plasma	4	56 k	14	34.1	38.4
			15	34.0	38.5
			16	33.8	38.8
avidin egg white	4	64 k	15	36.4	41.5
			16	36.4	41.5
			17	36.4	41.6
			18	36.4	-
serum albumin bovine	1	69 k	14	40.9	44.9
			15	41.0	44.9
			16	40.6	44.7
			17	40.4	44.9
concanavalin A *Canavalia ensiformis*	4	103 k	19	-	60.6
			20	55.5	60.8
			21	55.5	60.9
			22	54.8	60.5
			23	54.5	-
SAP human serum	5	125 k	22	70.3	76.3
			23	69.7	76.0
			24	69.3	74.6
			25	68.6	73.1
			26	68.3	72.8
ADH *Saccharomyces cerevisiae*	4	143 k	23	69.4	74.2
			24	69.4	74.5
			25	68.3	74.4
			26	67.2	75.0
pyruvate kinase rabbit heart	4	237 k	30	103	-
			31	103	111
			32	103	111
			33	102	110
			34	102	110
			35	100	110
SAP human serum	10	250 k	31	104	111
			32	105	112
			33	106	112
			34	105	111
			35	107	-

glutamate dehydrogenase bovine liver	6	336 k	37 38 39 40 41 42 43	- 128 128 128 128 128 128	134 134 134 134 135 - -
β-galactoctosidase *Escherichia coli*	4	465 k	44 45 46 47 48	155 156 155 155 155	162 163 163 162 -
GroEL *Escherichia coli*	14	801 k	65 66 67 68 69 70 71 72	- - 209 209 207 207 206 207	218 220 220 219 219 218 219 -

3.2.3 Effects of drift gas selection

Most travelling-wave ion mobility separations using Synapt and Synapt G2 HDMS instruments are performed using nitrogen gas, but CCS_{He} are desirable to enable comparisons with the literature and calculated CCS for structural models (*57, 157*). This is typically achieved by calibrating travelling-wave drift times in nitrogen gas using CCS_{He} values. This has been justified because calibration should correct for any systematic differences between K_0^{He} and K_0^{N2}. Moreover, non-correlated differences between K_0^{He} and K_0^{N2} should be less significant for large biomolecular complexes with low charge densities than for small ions with high charge densities (*158*). However, there are relatively few ions for which both CCS_{He} and CCS_{N2} have been measured directly under similar experimental conditions. This has made it challenging to assess the errors associated with using drift times measured in nitrogen to estimate CCS_{He}.

A plot of reciprocal mobilities in helium ($1/K_0^{He}$) and nitrogen ($1/K_0^{N2}$) gases shows that these values are well correlated for the biomolecular ions investigated (Figure 3.4A). The relationship between these two quantities provides a means to convert K_0^{N2} values to estimated K_0^{He} values, which in turn can be used to generate estimated CCS_{He} values (CCS_{He}^{est}). CCS_{He}^{est} values obtained using this approach are very close to those measured directly (CCS_{He}^{exp}) over a wide range of CCS (Figure 3.4B). The relative errors of this approach can be significant for the denatured peptides studied (up to 15 %), but decrease substantially with increasing CCS (Figure 3.4C). For ions generated under native-like conditions, the relative errors are small (standard deviation = 2.2 %). Therefore, errors associated with using drift times measured in nitrogen to estimate CCS_{He} are small for the ions generated under native-like conditions and similar to the associated errors in K_0^{N2} and K_0^{He}.

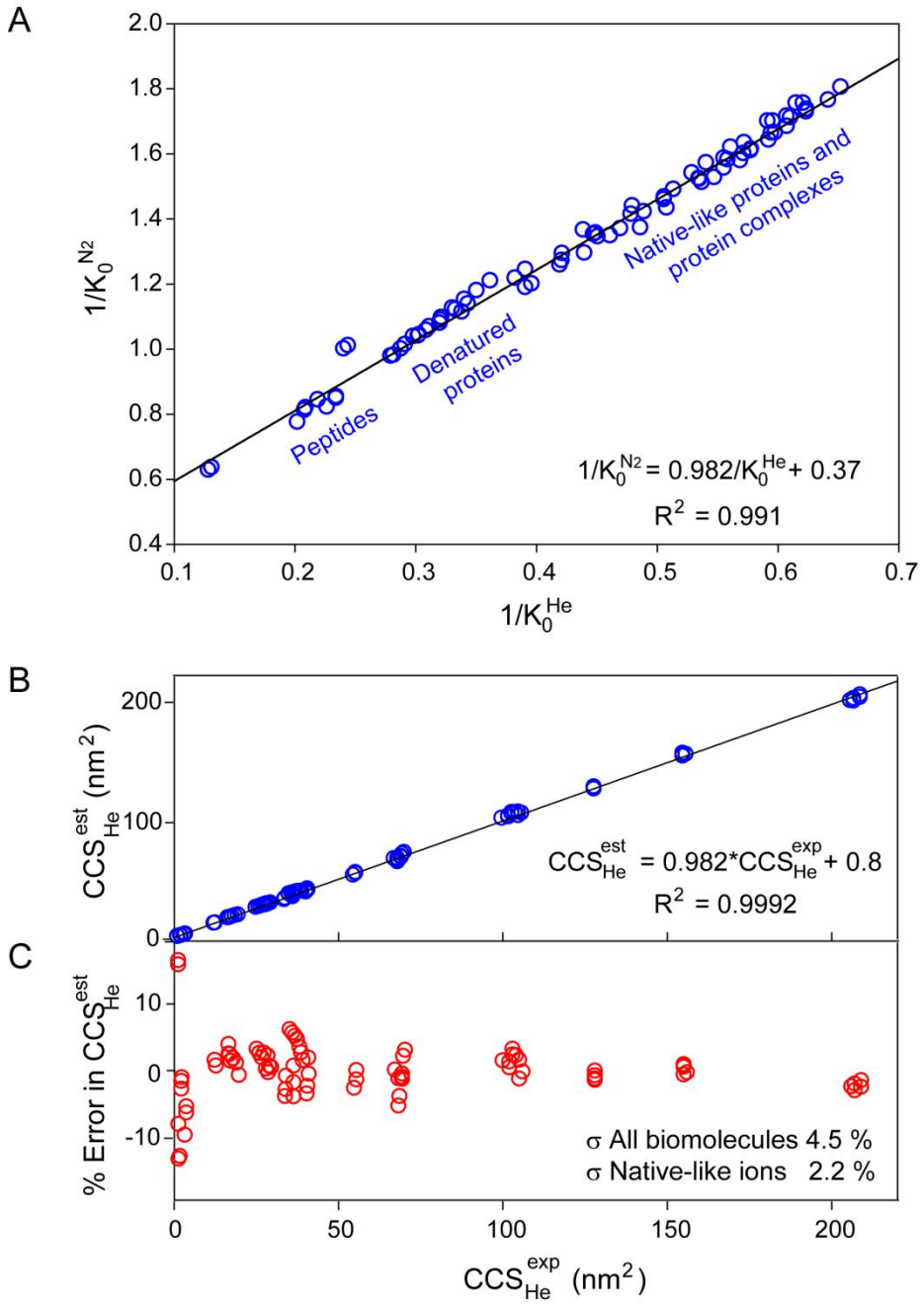


Figure 3.4. Effects of drift gas selection. **A** Correlation between reciprocal reduced mobilities in helium and nitrogen drift gas. Values for biomolecular ions exhibit a linear relationship, with a high correlation coefficient. **B** CCS in helium gas estimated from mobilities measured in nitrogen gas (CCS_{He}^{est}), using the relationship for biomolecules determined in **A**, are well correlated with those determined directly using mobilities measured in helium (CCS_{He}^{exp}). **C** The standard deviation between CCS_{He}^{est} and CCS_{He}^{exp} values is 4.5 % for all ions, and 2.2 % for the native-like ions.

3.2.4 Interpreting CCS: Relationship between experimental and calculated values

As discussed in Chapter 1, there are several methods available to calculate CCS. In brief, the simplest method for estimating CCS values is the PA method (*157*), where the orientationally-averaged CCS is determined by averaging the geometric cross-section over all possible collision geometries (*55*). The advantage of the PA approach is that it is compatible with coarse-grained representations of protein complexes, which are used in modelling strategies where no high resolution structures are available. The EHSS method accounts for contributions from multiple collisions between ions and buffer gas molecules (*57*), which are increasingly important for ions with concave surface topologies. Amongst all, the most accurate method for calculating CCS is the trajectory method (*58*), which models the ion-buffer interaction as a Lennard Jones potential, however this becomes computationally intractable for systems with many atoms.

Here, a critical comparison of helium CCS_{exp}, CCS_{PA}, and CCS_{EHSS} values are reported for a large set of proteins and protein complexes, from 12 to 801 kDa (Table 3.3, Figure 3.5). CCS_{PA} and CCS_{EHSS} values were calculated for models obtained from the PDB and which have had hydrogen atoms added. The CCS values calculated for the atomic structures were scaled to account for residues not present:

$$CCS' = (m_{expected} / m_{pdb})^{2/3} \times CCS \quad (3.2)$$

where $m_{expected}$ is the mass determined for the UniProt sequence and m_{pdb} is the mass determined from the crystal structure. This correction should be appropriate for residues absent from the surfaces of globular models, but may underestimate contributions for residues that would extend significantly from the model surface, *e.g.*, flexible loops.

Table 3.3. Proteins and Complexes used for Figure 3.5

Name	***n***	**PDB ID**	m_{pdb} **/kDa**	**Uniprot ID**	$m_{expected}$ **/kDa**
cytochrome *c*	1	1HRC	11.7	P00004	11.7
β-lactoglobulin	1	2Q2M	17.2	P02754	18.3
myoglobin	1	1WLA	17.6	P68082	17.6
β-lactoglobulin	2	2Q2M	34.3	P02754	36.6
TTR	4	1F41	50.5	P02766	55.0
avidin	4	1VYO	54.1	P02701	66.5
serum albumin	1	1E78	65.7	P02769	66.5
concanavalin A	4	3D4K	102	P02866	102
SAP	5	1SAC	116	P02743	126
ADH	4	2HCY	147	P00330	147
pyruvate kinase	4	1F3W	227	P11974	232
SAP	10	2A3X	232	P02743	253
glutamate dehydrogenase	6	1NR7	330	P00366	334
β-galactosidase	4	3IAP	460	P00722	465
GroEL	14	1SS8	771	P0A6F5	801

For almost all ions, Figure 3.5A shows the helium CCS_{exp} values are larger than the corresponding CCS_{PA} values and smaller than the corresponding CCS_{EHSS} values. The helium CCS_{exp} values for GroEL are slightly less than the corresponding CCS_{PA} value, which is very different than the behaviour for all of the other proteins and complexes investigated. These differences are consistent with a gas-phase structure that is significantly smaller than that suggested by the crystal structure. One possibility is that GroEL undergoes substantial collapse (*39, 155*) during IM analysis or ESI, relative to the other native-like ions investigated. This is supported by a differential mobility analysis (DMA) study of GroEL (*86*), which measured similar mobilities. DMA measures the mobility of the ion at atmospheric pressure, and therefore this suggests that any structural rearrangement occurs early in the electrospray process.

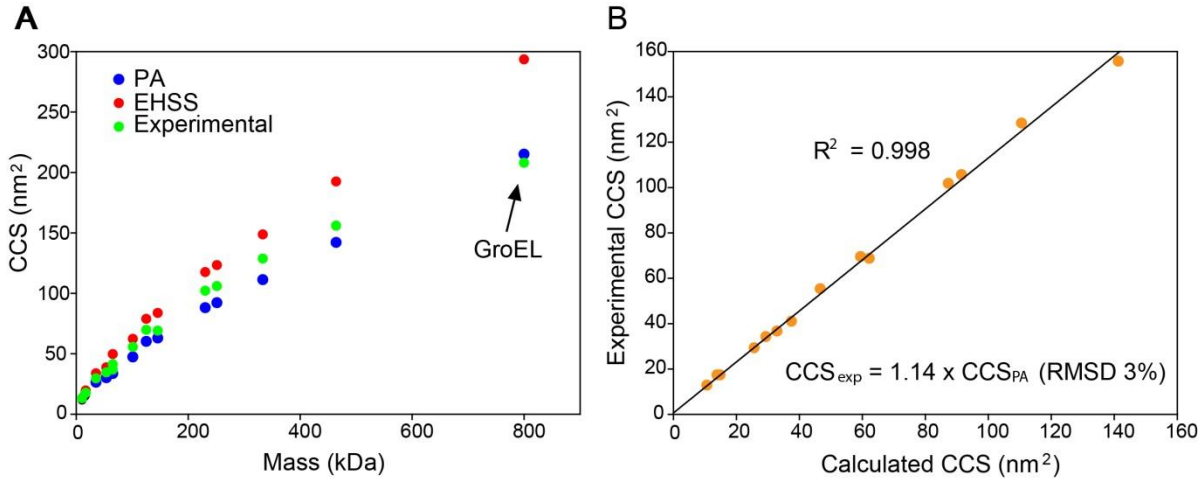


Figure 3.5. Comparing experimental and calculated CCS_{He}. **A** The experimental helium CCS (CCS_{exp}, green) of buffered proteins and protein complexes are consistently lower than those calculated from their corresponding crystal structures using the EHSS method (CCS_{EHSS}, red), and greater than those calculated using the PA method (CCS_{PA}, blue). The exception is GroEL, which has CCS_{exp} slightly less than CCS_{PA}. This suggests structural collapse of this large ion in the gas phase. **B** The experimental CCS of buffered proteins and protein complexes (GroEL has been excluded from this analysis) are well correlated with calculated values, and can be estimated by multiplying the CCS_{PA} by a factor of 1.14 (RMSD 3%).

The ability to relate experimental CCS to theoretical values is critical for the use of CCS as restraints in modelling approaches to determine protein complex topologies. Figure 3.5B shows that helium CCS_{exp} are well correlated with CCS_{PA} (GroEL has been excluded from the dataset), and can be estimated by the use of a scaling factor, such that $CCS_{exp} = 1.14 \times CCS_{PA}$, with an associated RMSD of 3%.

3.2.5 Calibration strategies for travelling-wave ion mobility

CCS in travelling-wave IM experiments are not determined directly from measured drift times, as is done for the RF-confining drift tube used in the experiments reported above. Instead they are determined by calibrating drift times for ions of interest against those measured for ions of known CCS. Calibration is necessary because the relationship between travelling-wave drift time and mobility is non-linear, and the specific non-linear factors that relate the two quantities depend strongly on experimental conditions and are difficult to determine from first principles (*50, 92*).

To evaluate the effects of calibrant ion selection on the accuracy of the calibration, drift times for denatured peptides, denatured proteins, and native-like protein complexes were measured using a standard Synapt G2 HDMS instrument with a travelling-wave ion mobility cell. Experiments were performed in ~3 Torr nitrogen at constant wave velocity (200 m/s) and wave height (20 V). Numerous protocols have been reported for calibrating travelling-wave ion mobility drift times (*92, 148, 159*). For the analysis below a modified form of a protocol reported previously (*92*) was used (section 2.1.2).

Most reported CCS_{He} for protein complexes have been based on calibrations using denatured monomeric proteins, typically ubiquitin, cytochrome *c*, and myoglobin. To evaluate the accuracy of these typical calibration approaches for measuring CCS of protein complexes, the travelling-wave data described above was calibrated using only results for denatured ubiquitin, cytochrome *c*, and myoglobin (Figure 3.6A). Data used to obtain the calibration functions are well correlated with the associated best-fit line.

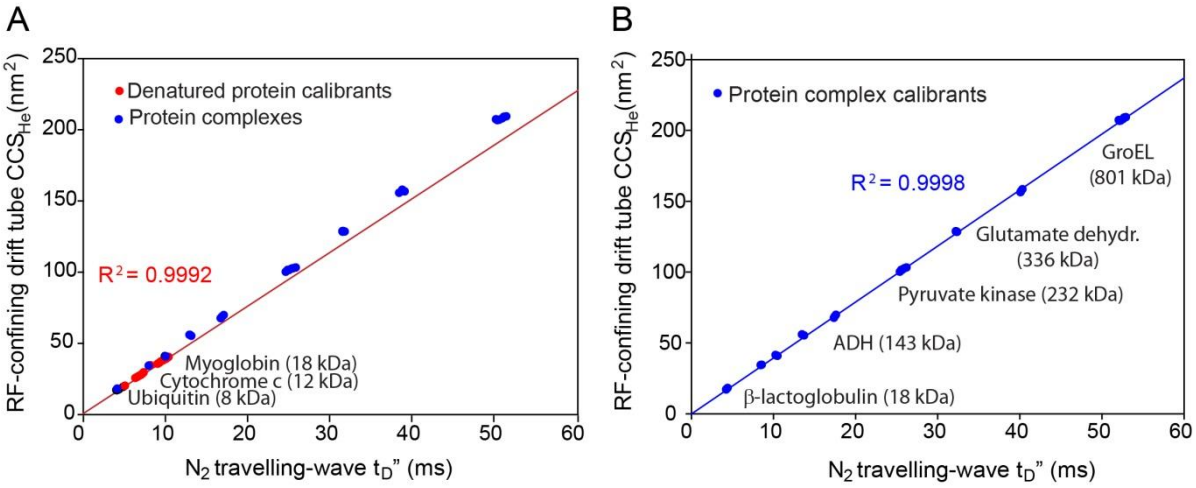


Figure 3.6. Accuracy of calibration scheme for TWIMS. Travelling-wave data was calibrated using denatured proteins only (**A**) or protein complexes only (**B**). Whilst the best-fit-lines in each case are well correlated with the data used to define the calibration functions, increasing errors for the determination of protein complex CCS arise when denatured ions are used as calibrant ions.

Calibrated CCS_{He} for native-like protein complexes however exhibit increasingly large errors with increasing mass, *i.e.*, denatured protein ions are less effective for calibration of larger native-like ions than the smaller native-like ions investigated. Using this calibration approach would result in protein complex CCS_{He} that are 2.5 - 15 % smaller than those measured directly using the RF-confining drift tube, depending on the size of the ion. The same travelling-wave data was also calibrated using only protein complex ions (Figure 3.6B). Data used to optimise the calibration function are well correlated with the best-fit line (R^2 = 0.9998), and yield protein complex CCS that are very similar, (0.9 ± 1.0 %), to those measured directly using the RF-confining drift tube. Significantly, there is no systematic deviation between the direct and indirect CCS and the random deviation is comparable to the uncertainties in the direct measurements.

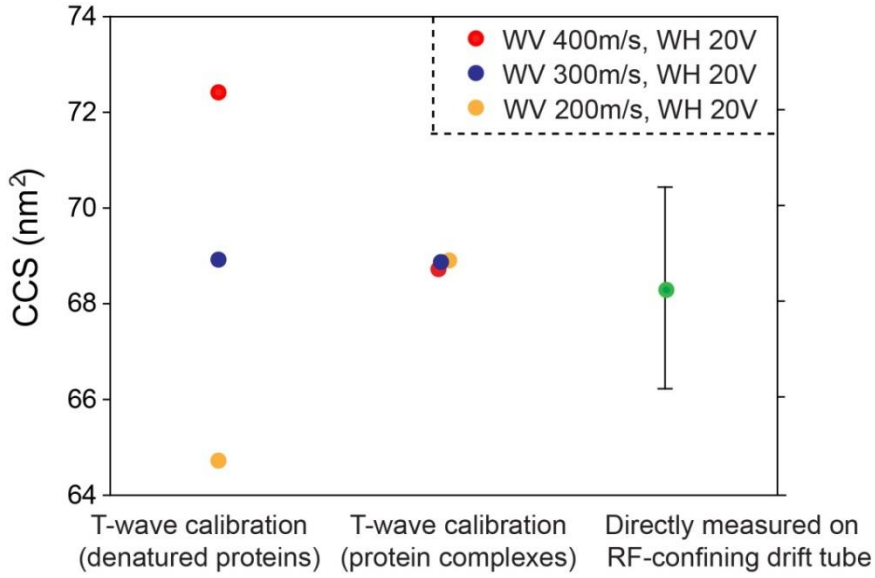


Figure 3.7 Precision of calibration schemes for TWIMS. CCS_{He} for ADH was determined using TWIMS at three wave velocities, using either denatured protein ions or protein complex ions as calibrants. The results shown are compared to the CCS measured for ADH on the RF-confining drift tube (green). Error bar reflects ± 3 %. WV = wave velocity; WH = wave height.

In order to characterise CCS measurement accuracy and precision of each calibration scheme under conditions of varying wave velocity, the CCS was determined for native-like ADH (143 kDa) ions at three wave velocities and constant wave height. These results are summarised in Figure 3.7, where the measured CCS_{He} for ADH (charge-averaged) is shown for two calibration strategies, and compared to that measured directly on the RF-confining drift tube. CCS_{He} for ADH ions obtained *via* calibration using only protein complex ions appear to be essentially independent of the wave velocity selected, with high CCS accuracy (deviation from RF-confining drift tube CCS < 1 %). In contrast, CCS obtained *via* calibration using only denatured protein ions depend strongly on the wave velocity selected, with variable CCS accuracy (CCS deviation 1 – 6 %).

Overall, it is clear that using native-like protein complex ions to obtain the calibration plots results in more accurate CCS for native-like protein complex ions and less sensitivity to the specific travelling-wave conditions used. These results, and others (*160-161*), indicate that calibration will yield the most accurate CCS when drift times for ions of interest are corrected using an exponential factor determined for a set of ions with similar properties.

3.2.6 Minimising ion heating in TWIMS

Having optimised the calibration strategy, the next step was to turn to the detailed instrumental settings of the Synapt G2 HDMS, which had not yet been fully characterised for interrogating large protein complexes[6]. The increased mobility resolution which the new generation of TWIMS device affords, is primarily due to the ability to run at increased operating pressures (> 2 mbar), which allows the use of higher travelling-wave velocities and amplitudes (*93*). In order to prevent bulk flow of nitrogen out of the IMS cell, a helium chamber is positioned prior to the IMS section of the device. Ions have higher mobility in helium, and therefore lower fields are required to inject ions into the gas. This has the advantage of "cooling" the ions before entry to the high-pressure nitrogen IMS cell, reducing unwanted scattering, and ion activation or heating (*93-94*).

High injection voltages and gas pressures in the tri-wave region result in more energetic collisions between analyte and buffer gas molecules. Sufficiently energetic collisions may result in ion activation and unfolding, similar to changes observed in protein

[6] Subsequent studies by Salbo et al. (2012) and Zhong et al. (2011) have made significant contributions towards this.

complexes by increasing the sample cone voltage (*162*), or collision energy (*154*, *163*). Few studies, however, have assessed the effect of individual parameters in the second generation TWIMS on protein complex structures (*94*, *164*). Therefore, injection voltages in the tri-wave region and gas pressures in the He and TWIMS cells were systematically adjusted, whilst monitoring the effect on the arrival time distribution (ATD) for ADH. Travelling-wave velocities and heights were kept constant (200 m/s and 20 V respectively), and one parameter changed at a time. An increase in the width of the distribution, or the appearance of multiple peaks, is taken as evidence of ion activation, resulting in unfolding.

First the effects of increasing the trap DC bias were examined. This sets the potential difference between the trap TWIG and He cell, and therefore controls the injection of ions into the He cell. As the trap DC bias was increased from 40 V (~ 30V required for ion transmission) to 80 V, the appearance of a second conformation for ADH 25+ is observed, at higher drift times (Figure 3.8A). This is consistent with a partially unfolded conformation.

Similar observations were made whilst increasing the He cell DC from 50 V (~30 V required for transmission) to 120 V (Figure 3.8B). He cell DC is responsible for the transfer of ions from the He cell to the high pressure N_2-containing TWIMS cell. Optimisation of the trap DC bias and He cell DC is therefore crucial in order to attain a balance between maintaining ion transmission and minimising ion activation. The choice of setting is also highly dependent on the size and charge of the system being studied and the gas pressures in the He and TWIMS cells.

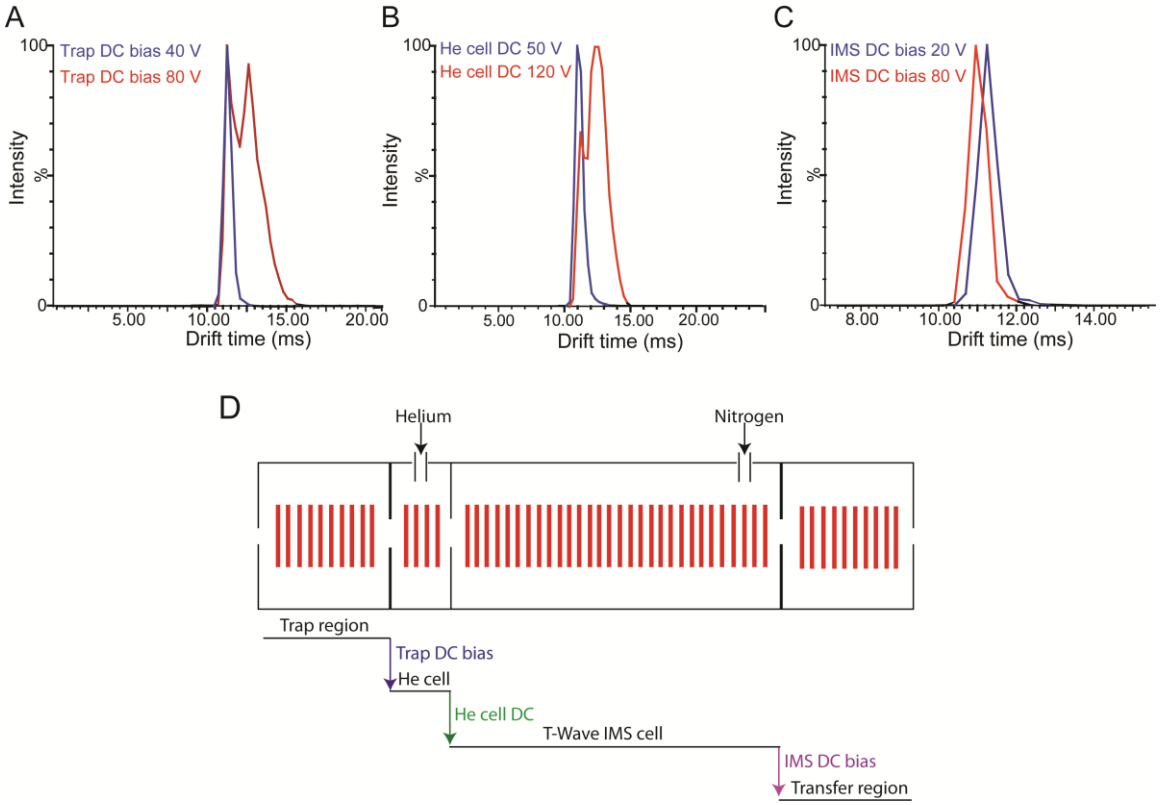


Figure 3.8 Effect of ion injection energies and gas pressures on ion heating. Arrival time distributions for ADH 25+ under conditions of varying **A** Trap DC bias, **B** He cell DC and **C** IMS DC bias. **D** Tri-wave region of Synapt G2, and associated bias voltages.

In contrast to the previous two parameters, the IMS DC bias has little effect on drift times, or ion transmission. The passage of ions from the TWIMS cell to the transfer region is governed by the IMS DC bias. However ions move easily from the high-pressured environment of the IMS cell into the low-pressure TWIG. Interestingly, increasing this bias sufficiently resulted in slightly shorter drift times (Figure 3.8C). Higher fields here may have allowed ions to pass through several travelling waves in the transfer region, before losing sufficient energy to "surf" the wave. This is effectively reducing the length of the transfer region. However, as no structural changes occur, provided the same IMS DC bias is used for both calibrants and analyte, this parameter does not affect CCS measurement accuracy.

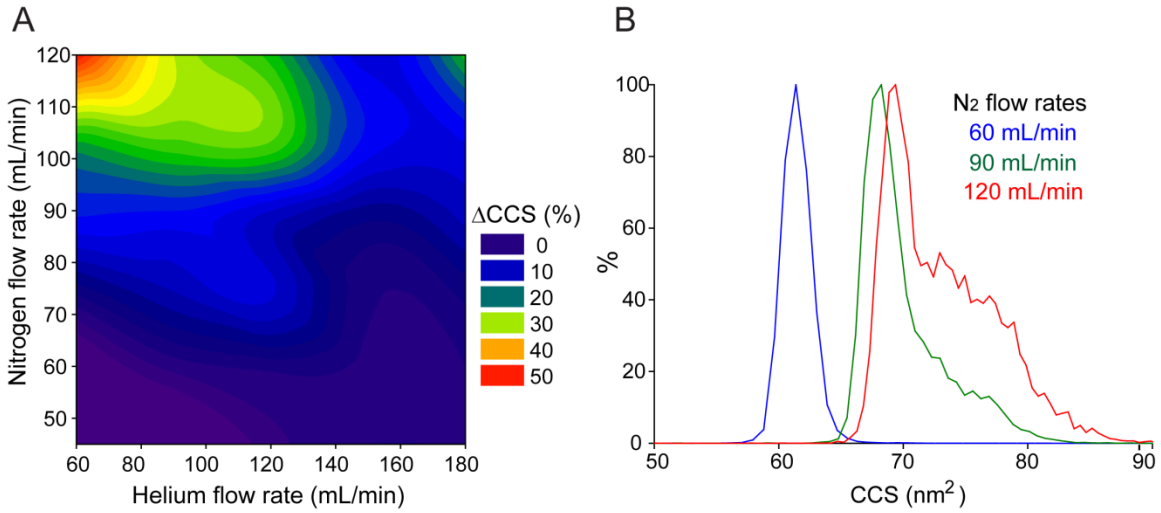


Figure 3.9. Effect of gas pressures on CCS accuracy. **A** CCS of ADH 25+ was measured whilst varying helium and nitrogen flow rates in the IM cell. Denatured cytochrome *c* (14-20+), myoglobin (18-25+) and ubiquitin (9-13+) ions were used as calibrants. These ions were chosen over protein complexes in this instance, as they are less susceptible to heating effects, being in an unfolded conformation initially. ΔCCS is the % deviation from the measured CCS value at 180 mL min^{-1} He and 60 mL min^{-1} N_2. **B** Arrival time distribution on CCS scale for ADH 25+ at different N_2 flow rates, with constant He flow rate (180 mL min^{-1}). A constant wave height (20 V) and wave velocity (200 m/s) was used for all measurements.

Finally, the effect of gas pressures used in the He chamber and N_2-contaning IMS cell were examined. The CCS of ADH 25+ was measured whilst varying helium (60 – 180 mL min^{-1}) and nitrogen (45 – 120 mL min^{-1}) flow rates. The measured CCS accuracy was found to be strongly dependent on N_2 flow rate (Figure 3.9). By increasing the N_2 pressure in the IMS cell, higher wave velocities and amplitudes become accessible and enhanced resolution is achievable (*93*). However, this comes at a cost: the higher nitrogen pressures can result in ion heating. The effects of ion heating as a result of high nitrogen pressures can be mitigated to a certain extent, by keeping high (> 140 mL min^{-1}) He flow rates to "collisionally cool" the ions, prior to injection into the N_2-containing IM cell (Figure 3.9). This is in agreement with a study on the dissociation of "thermometer

ions", in which greater dissociation was observed at higher nitrogen, and lower helium flow rates (*94*).

From these experiments, it can be concluded that optimum conditions for the analysis of protein complexes are those in which the sample cone voltage, the injection voltage into the He cell, and the injection voltage into the IMS cell are minimised to conditions near the threshold for ion transmission. In addition, high He flow rates of > 140 mL min^{-1} should be employed, and nitrogen flow rates kept below ~ 90 mL min^{-1}. This results in measured pressures of ≤ 3 Torr. Under these conditions ion heating in TWIMS is negligible for large protein complexes**.**

3.2.7 Wave velocity and height affect resolution and CCS accuracy

Next, the effects of the travelling-wave velocity and wave height on resolution were considered. Whilst resolution in conventional IM has been well defined (*54*), the factors governing resolution in TWIMS are less well understood (*50*) and have not been explored systematically. Therefore the effects of wave velocity and wave height for drift time resolution ($t_D/\Delta t_D$) of ADH ions were assessed. Travelling-wave drift times were measured for ADH ions at constant wave height, whilst varying the wave velocity from 125 to 1100 m/s. This was repeated for many different wave heights. Similarly, the wave velocity was held constant, whilst varying wave height from 10 to 40 V. This was also repeated at several wave velocities. Over 120 datasets were collected in this way, and are shown below (Figure 3.10). Data extracted at wave heights 20 and 40 V (Figure 3.10A and B), shows a clear increase in drift time resolution with increasing wave velocity, up to ~ 500 and 800 m/s respectively. This is followed by a decrease in resolution with

increasing wave velocity. Overall however, a complex relationship between the resolution and travelling wave parameters exists (Figure 3.10C) with much depending on the ratio of wave height to wave velocity. Similar findings were subsequently published by Ruotolo and co-workers (*164*). The highest resolution found here for ADH was at wave height 25 V and wave velocity 400 - 500 m/s.

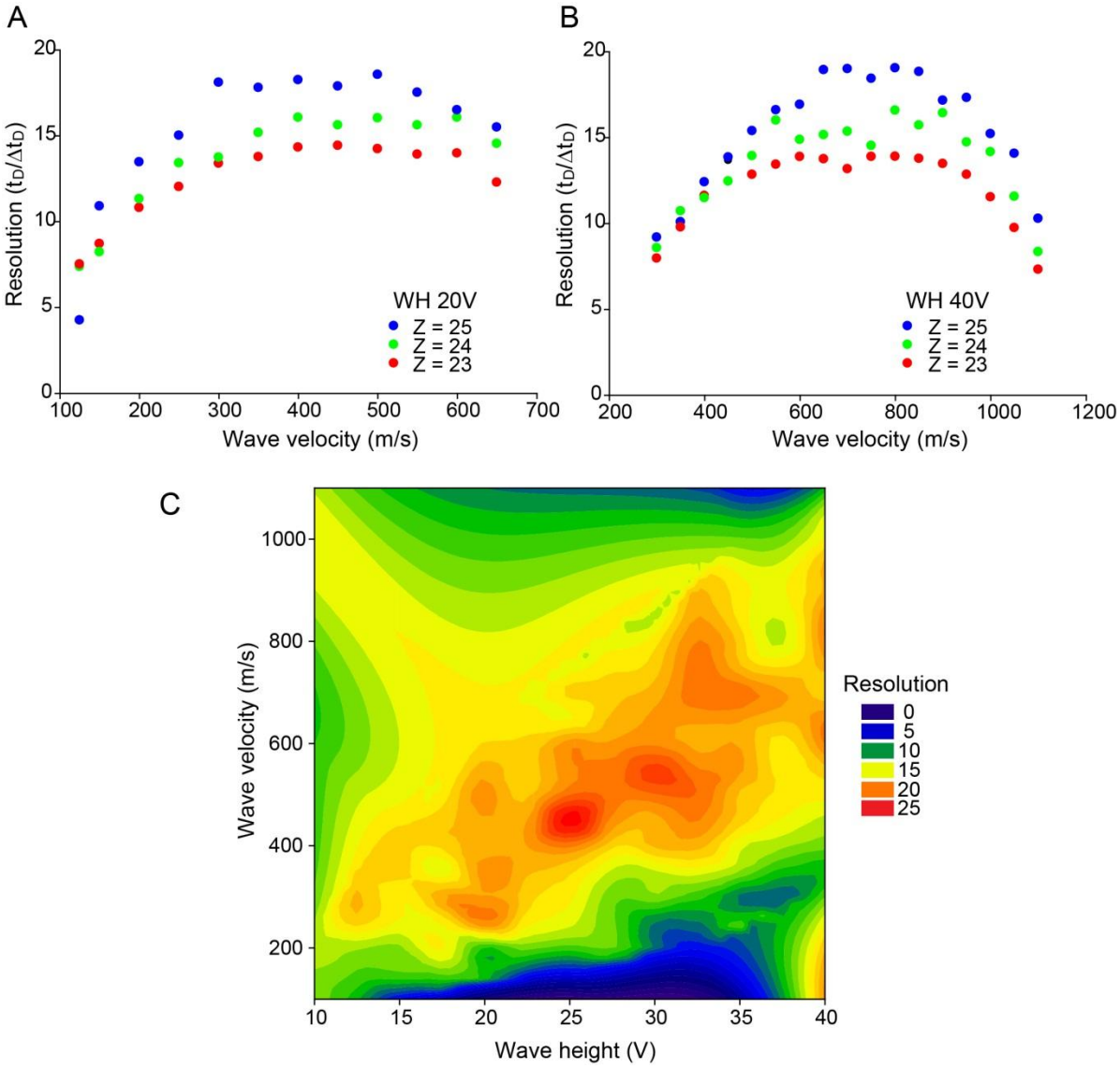


Figure 3.10. Effect of travelling-wave settings on resolution. The relationship between wave velocity and drift time resolution for ADH 23-25+ ions ($t_D/\Delta t_D$) is shown at wave height 20 V (**A**) and 40 V (**B**). Contour plot for ADH 25+ resolution with increasing wave height and velocity (**C**). Experiments were performed at ~ 3 Torr in the IMS cell.

Whilst these conditions offer the optimum resolution, it is interesting to note the effects of wave velocity and height on CCS accuracy. In order to fully characterise this, CCS_{He} were determined for pyruvate kinase, a 232 kDa tetramer, over a wide range of wave velocities and heights (Figure 3.11). Once again, two calibration strategies were compared: using either denatured protein ions (myoglobin and cytochrome *c*) or protein complex ions as calibrants. This time, only two protein complexes were chosen as calibrants: ADH (143 kDa) and glutamate dehydrogenase (336 kDa). These closely bracketed the mass and mobility of pyruvate kinase.

The wave velocity was varied at constant wave height (20 V). When using denatured ions as calibrants (Figure 3.11A, red), the measured CCS for pyruvate kinase increased from ~90 to 117 nm^2 as the wave velocity was increased from 150 to 400 m/s. At intermediate wave velocities (~200 m/s), the CCS were within that expected from the RF-confining drift tube experiments. However at higher and lower wave velocities, CCS errors were significant. Similarly, the wave velocity was held constant (200 m/s) and the wave height varied. Use of denatured ion calibrants resulted in the measured CCS for pyruvate kinase decreasing from ~ 105 to 98 nm^2 as the wave height was increased from 15 - 35 V (Figure 3.11B, red). When using bracketing protein complex ions as calibrants however, pyruvate kinase CCS was within 3% of drift tube CCS, over the range of wave velocities and heights investigated (Figure 3.11, blue).

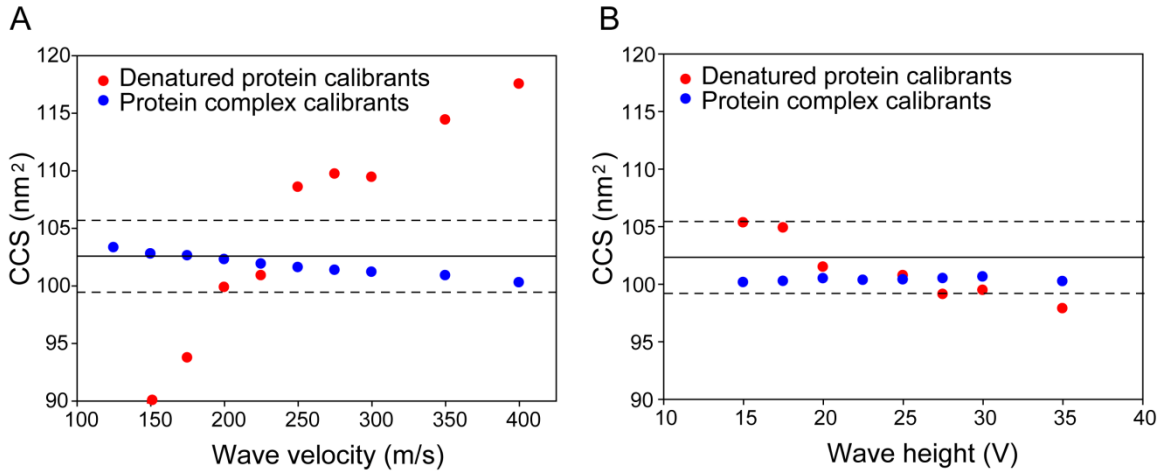


Figure 3.11. Effect of travelling-wave parameters on CCS accuracy. Pyruvate kinase CCS_{He} is shown with increasing wave velocity (**A**). Wave height and IMS pressure was 20 V and 3 Torr respectively. Pyruvate kinase CCS_{He} is shown with increasing wave height (**B**). Wave velocity and IMS pressure was 200 m/s and 3 Torr respectively. CCS was determined using denatured protein ions (red), or bracketing protein complex ions (blue) as calibrants. The solid and dashed lines show the absolute CCS of pyruvate kinase measured on a RF-confining drift tube and the estimated error (±3%).

Despite careful calibration, using bracketing protein complex ions, a trend can be observed: the measured CCS decreases with increasing wave velocity (Figure 3.11A). This was further investigated by evaluating the travelling-wave drift times of four protein complexes at various wave velocities, over many experiments (wave height 20 V, IMS pressure 3 Torr). With increasing wave velocity, the four protein complexes exhibited differences in their relative travelling-wave drift times. For example, at wave velocity 200 m/s, complexes could be well defined by a single power relationship, between CCS′ and t'_D (Figure 3.12C). However at higher wave velocities, subtle differences in the protein complexes become apparent, and applying one power function to define all data is clearly a simplification of the true relationship between travelling-wave drift times and CCS.

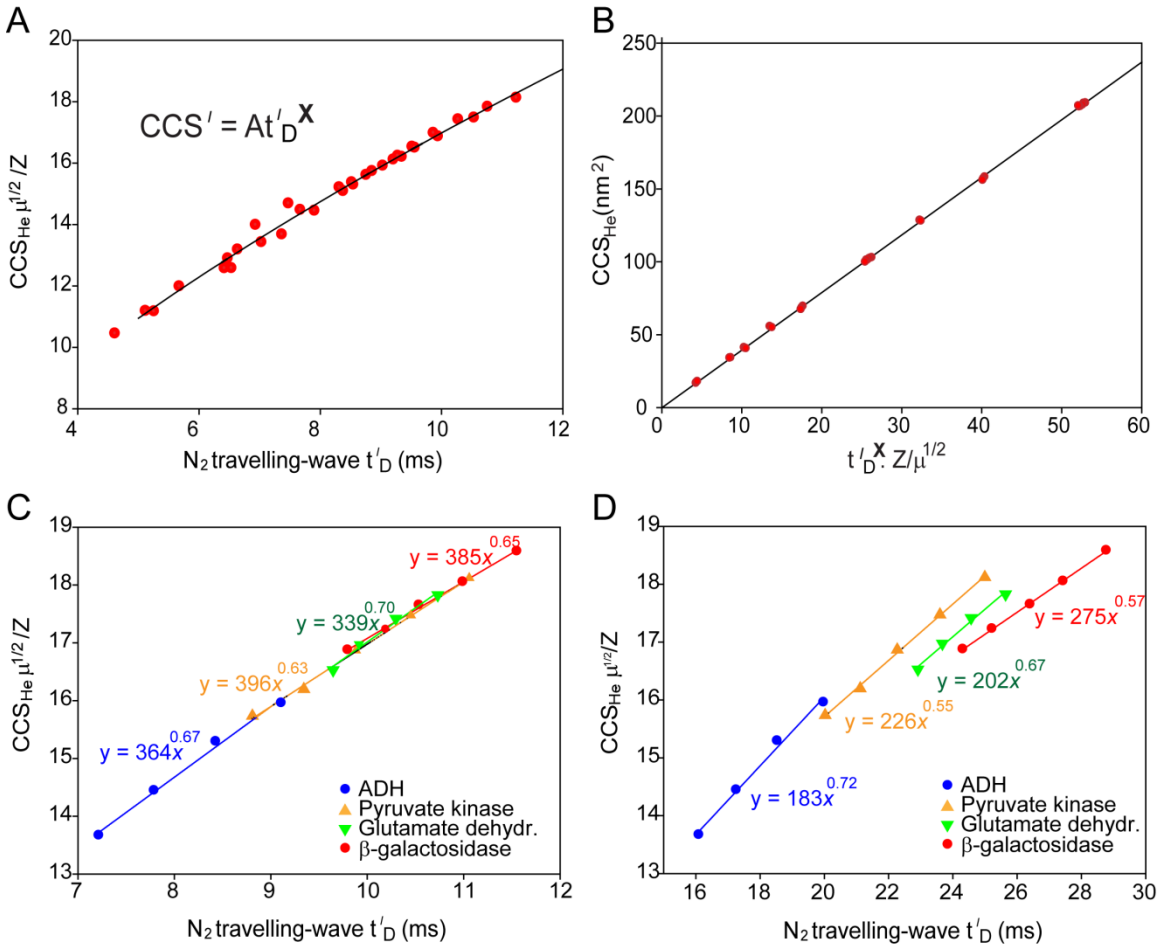


Figure 3.12. Travelling-wave drift times and CCS are related by a power function (**A**). The exponential factor, X, from this relationship is used to correct travelling-wave drift times. These corrected drift times are linearly correlated with CCS_{He} and this forms the basis of calibration (**B**). A single power relationship between travelling-wave drift times and CCS becomes increasingly insufficient to describe data from multiple species, with different *m/z* and mobilities, at higher wave velocities (**C** = WV 200 m/s; **D** = WV 400 m/s). A constant wave height 20 V and IMS pressure 3 Torr was used.

3.3 DISCUSSION

The absolute CCS_{He} measurements made on the RF-confining drift tube can be compared with values in the literature. For the 16 denatured peptide and protein ions reported both in Table 3.1 and in a database[7] obtained from traditional drift tube measurements (reported errors of 1 %) (*91, 150*), values here are 0.3 ± 1.4 % larger. There are relatively few reported CCS_{He} for native-like protein complexes. For example, CCS_{He} determined from travelling-wave drift times in nitrogen gas and calibration with denatured protein ions for tetrameric TTR (z = 15) (*40*) and tetradecameric GroEL (charge-state averaged) (*84*) are 29 ± 2 and 244 ± 3 nm^2, respectively, whereas the RF-confining drift tube measured values were 34.0 and 207 nm^2 for these ions, respectively. "Surfing-the-wave" travelling-wave drift time measurements (*165*) indicate that the charge-state averaged CCS_{He} for tetrameric TTR, pentameric SAP, and decameric SAP are 31.5, 58.8, and 88.9 nm^2 (estimated error of 5 %) (*145*), compared to the RF-confining drift tube values of 34.0, 69.2 and 105 nm^2 respectively. Differences between the CCS_{He} values obtained here and those in the literature, which all used instrumentation with similar geometries, but less direct means to determine CCS_{He}, illustrate the need for a unified framework for obtaining CCS in different laboratories.

Calibration approaches for converting between IM drift times and CCS predate travelling-wave technologies (*158*), and carry a number of advantages over direct CCS measurements. Calibration enables measurements of high accuracy without detailed knowledge of instrument conditions, *i.e.*, temperature, drift-gas pressure, drift-gas composition, and ion behaviour. Consequently, calibration has the potential to enable

[7] http://www.indiana.edu/~clemmer/Research/Cross%20Section%20Database/cs_database.php

robust and accurate CCS measurements that can be readily reproduced by different laboratories, providing a common frame of reference for all measurements. Through this work, the errors associated with travelling-wave CCS measurements have been more precisely quantified. Errors for the calibrated CCS include contributions from the uncertainties of CCS for the calibrant ions (3 % for the values reported in Tables 3.1 and 3.2), the reproducibility of the drift times measured for the calibrant and analyte ions, and other factors. Errors associated with using drift times measured in nitrogen to estimate CCS_{He} are small for the ions generated under native-like conditions and similar to the associated mobility measurement error. These results indicate that careful experimental design and execution can result in calibrated CCS with propagated errors of less than 5%.

With this database and error analysis, a set of best practices for CCS calibration can now be proposed. The most important factor for CCS accuracy is the choice of appropriate calibrant ions, which should closely match the analyte mobility and mass. In addition, instrumental settings at the threshold for ion transmission should be selected in order to minimise ion heating. Optimisation of travelling-wave parameters should be carried out on a case-by-case basis. Where resolution is the more important aspect of the experiment, higher wave velocities and pressures would be applicable. The mobility resolution of the second generation travelling-wave ion mobility separator has been shown to have improved from around 10 in the original Synapt HDMS to 45 for two reverse peptides (*93*), however the mobility resolution achieved for protein complexes does not reach full potential of the instrument. This is likely due to inherent conformational heterogeneity in solution and/or a degree of flexibility in the gas phase. Whilst optimum resolution is attained at higher wave velocities, the most accurate CCS measurements require lower wave velocities.

Reports in the literature (*166*) have suggested that wave velocities and heights affect protein complex structures, however from this work, it is clear that it is CCS measurement accuracy, rather than the structures themselves, which is adversely affected. These effects become more prominent at higher wave velocities and when inappropriate calibrants are used. At higher wave velocities, it appears that differences in the gas-phase behaviour of ions of various sizes and mobilities become more pronounced. This could be explained by that fact that ions do not constantly drift and undergo repeat accelerations when the travelling-wave pulse reaches them. There is a certain time taken to reach a steady drift velocity and this can be significant for large ions. Therefore, at higher wave velocities, this acceleration time may become a notable fraction of the drift time, changing the nature of the separation to have a more *m/z* dependent character. Closely bracketing the native-like ion is therefore preferable to using a larger set of IM standards with a broad range of mobilities and masses. The closely bracketed set mitigates the need to characterise further ion transport properties in the instrument and results in smaller extrapolation errors. Whilst this and other studies (*140, 160, 164*) show that accurate CCS can be measured using travelling-wave IM, the nature of ion transport through the travelling-wave mobility separator is yet to be fully understood. Further experiments combined with SIMION modelling (*167*) could help address these issues.

Finally, the CCS database provides a foundation for understanding the general relationship between gas-phase and condensed-phase structures for native-like proteins and protein complexes. This is important for determining how IM-MS results should be used to restrain structural models. The excellent correlation between CCS from ion mobility and those calculated from atomic structures suggests that proteins retain "memory" of their native structures in the gas phase (*116*). This encouraging result

indicates that measurements made in the gas phase can be related to protein complex structures in solution.

3.4 SUMMARY

This work provides a detailed framework for assessing and reducing the errors associated with using travelling-wave drift times for determining CCS for ions spanning masses and CCS of more than two orders of magnitude. Through the CCS database reported here, the accurate CCS determination of large protein assemblies has been made possible, offering significant improvements over previous calibration strategies employing denatured protein ion calibrants. Optimisation of instrumental parameters was found to be critical for CCS accuracy. Gas pressures and ion injection energies were adjusted to minimise ion heating. Optimum CCS accuracy was achieved with the use of lower wave velocities (*c.f.* maximum resolution at higher wave velocities). To avoid *m/z* dependent effects at higher wave velocities, closely bracketing the analyte with calibrants is preferable to a larger set of IM calibrants of broad mass and mobilities.

Overall, small calibration errors are readily achieved when using at least two native-like IM standards that closely bracket the masses and mobilities of the analyte ions of interest. In the context of gas-phase structural biology, these results provide the framework for using restraints generated by IM-MS to generate more accurate 3D models of protein complexes and other biomolecular assemblies.

Part of this chapter has been adapted with permission from Bush et al. (2010), Analytical Chemistry, 82, 9557-9565, © 2010 American Chemical Society doi:10.1021/ac1022953.

4 DEVELOPING COMPUTATIONAL STRATEGIES FOR STRUCTURAL MODELLING OF PROTEIN ASSEMBLIES

4.1 INTRODUCTION

4.1.1 Integrative structure characterisation of protein assemblies

The cell consists of hundreds of macromolecular assemblies, responsible for diverse processes such as RNA synthesis, translation, protein folding and degradation (*4, 168*). Understanding the mechanisms of these assemblies requires structural characterisation. However, detailed structure determination of assemblies is challenging using a single experimental or computational approach (*123*). Over the last few years, significant advances in integrative structural modelling have been made, with structures for the nuclear pore complex (*169*), and 26S proteasome suggested (*170-171*). Recently the release of the Integrative Modelling Platform (IMP) (*119*), a software suite for building and characterising models of macromolecular assemblies, was a significant contribution towards integrative structure modelling. Integrative modelling entails encoding data from a variety of sources, and can characterise assemblies at several levels of resolution, depending on the data available. Data from individual techniques may offer varied accuracy and precision (*123*). In the integrative approach, all available data is simultaneously considered, allowing improved accuracy, precision and efficiency than using any one data type on its own (*123*).

Generation of data

The first step in the process of integrative modelling is the collection of high quality data. This may be in the form of experimental datasets, statistical information, molecular force fields, physical laws and available databases (*119*). Whilst X-ray crystallography (*172*) and NMR (*173*) can provide high resolution data, that of the atomic/residue positions, they have certain limitations. For instance, X-ray crystallography requires the growth of suitable crystals which diffract, and lacks information on dynamics. NMR is limited by the size of the assembly (*4*). Data from computational approaches, such as docking and comparative modelling, often have low accuracy (*174*). Other biophysical and biochemical methods provide lower resolution data. For instance, the stoichiometry and assembly composition can be determined by immunoblotting, and MS (*5*). The yeast two hybrid assay, and affinity purification, can inform on the interactions between assembly components (*175-176*). Several techniques can offer information on the relative orientation of components: cryo-EM, HDX, hydroxyl-radical footprinting and chemical cross-linking, to name a few (*177-181*). Finally, the overall assembly shape and symmetry can be inferred from small-angle X-ray scattering (SAXS) and EM (*182-183*).

Representation

An appropriate representation scheme must be selected based on the resolution of the available data. Each assembly component can be represented by a particle, defined by its position and size. The particle may correspond to atoms, atomic groups, residues, secondary structure elements, domains, proteins, subcomplexes, or the whole assembly (*123*). If data are provided in different resolutions, a hybrid approach where models are represented by multiple schemes can be adopted.

Spatial restraints and optimisation

The data, once collected and collated, can then be used to restrain features of the assembly. These may be the positions of proteins, protein contacts, symmetry, shape and volume (*184*). A restraint can be thought of as force acting on the components of the assembly in order to produce a configuration which satisfies the experimental data. Restraints frequently take the form of a harmonic function of the restrained feature (*184*). Ideal solutions take a value of 0, whereas values greater than 0 correspond to violated restraints. Models consistent with input data are calculated by optimising a scoring function, which consists of individual spatial restraints. The scoring function is a probability density function of the Cartesian coordinates of the assembly proteins (*C*) given information (*I*) on a restrained feature, P_f (*169*).

$$p(C / I) = \prod_f p_f(C / I_f) \tag{4.1}$$

Practically, the scoring function, S_F is calculated by summing individual restraints, *r*.

$$S_F = \sum_f r_f \tag{4.2}$$

The scoring function is therefore optimised to obtain all possible models that satisfy input restraints. Adequate sampling is critical, in order to exhaustively search the conformational space of structures fitting the data. For example, IMP makes use of Monte Carlo sampling algorithms to generate tens of thousands of random configurations. These are then optimised by simultaneously minimising violations of all restraints. This is achieved with the use of conjugate gradients, and simulated annealing molecular dynamics, which refine the position of particles (*123, 168, 184*). Ideally, the global optimum corresponds to the native assembly structure.

Ensemble analysis

Multiple configurations may satisfy the input data. This could indicate that there is insufficient data to define the native state, or that there are multiple native states. In this case, ensemble analysis is performed on the candidate structures (*169*). This analysis can be in terms of similarity of the assembly features, which may be the protein contacts, overall shape, or for high resolution models, the RMSD of atomic/residue positions. Clustering based on structural similarity can allow assessment of the variability in the ensemble of solutions (*123*, *171*).

4.1.2 Structural modelling from ion mobility-mass spectrometry data

The use of data from proteomics, EM, SAXS, NMR and X-ray crystallography is now well established (Figure 4.1) within integrated modelling approaches (*119, 185-188*). The possibility of including data from IM-MS in the context of such hybrid approaches is attractive, however poorly addressed. Noncovalent-MS is beginning to play a more important role in structural biology (*189-190*), having a particular advantage for studying the structures of heterogeneous, low-abundance protein assemblies. Typical MS data includes the stoichiometry of intact complexes and subcomplexes, their subunit connectivity, and their assembly dynamics (*5, 191*). One of the major limitations is the fact that data are typically low-resolution, compositional information. Coupling IM to MS provides additional topological information in the form of an orientationally-averaged CCS (*87, 149*). IM therefore provides complementary information to mass measurements from MS*;* the resulting hybrid technology is becoming a powerful method for structural interrogation.

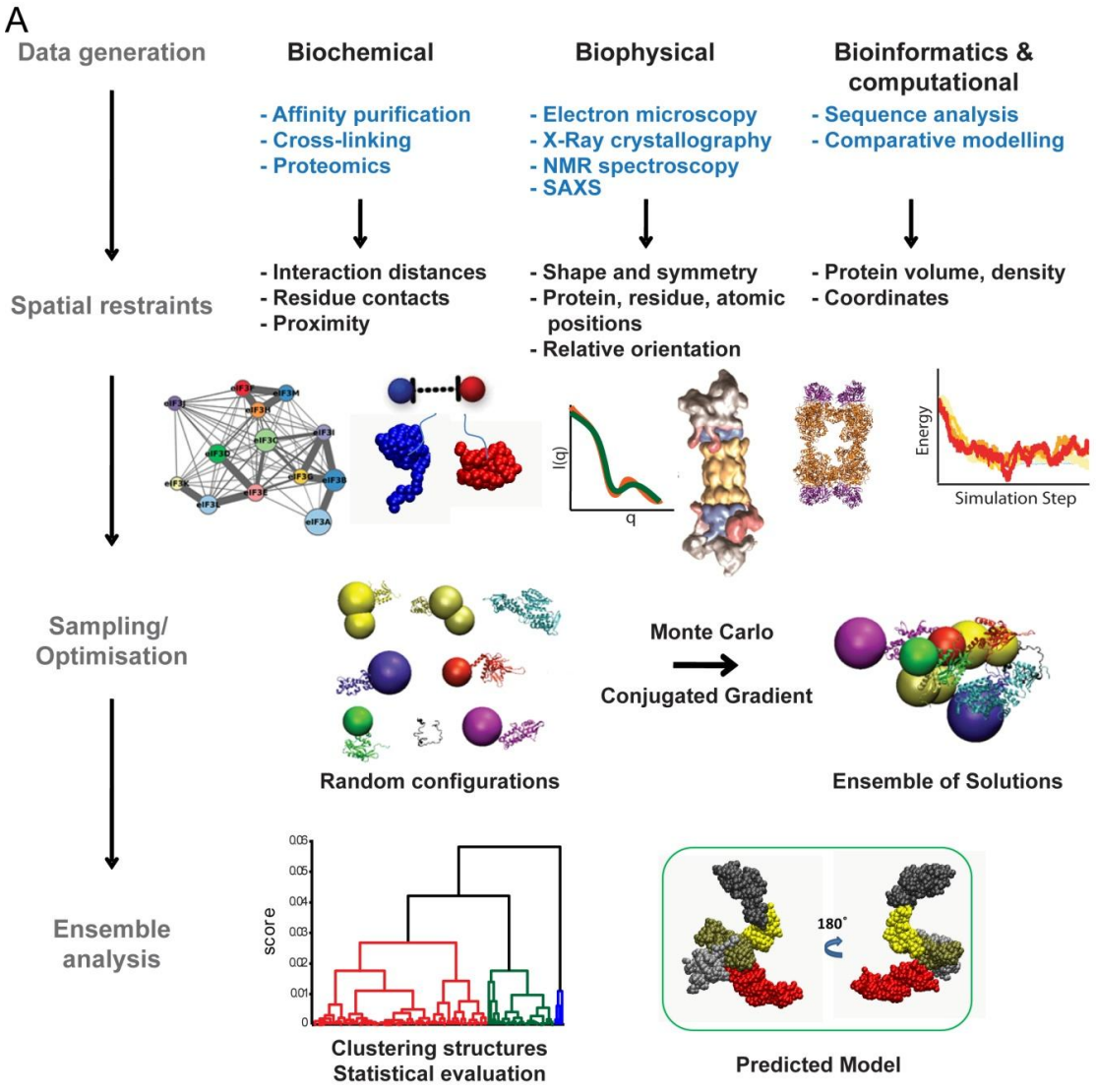


Figure 4.1 Integrated structural modelling. **A** Data from a variety of techniques can be converted to spatial restraints, which are used to define a scoring function. The scoring function is optimised by simultaneously minimising violations of all restraints, to produce an ensemble of candidate solutions which satisfy the input data. These can be evaluated in terms of their assembly features by clustering and other statistical tools, in order to predict the native structure. **B** Incorporation of IM-MS into such hybrid approach would be advantageous. The types of restraints which IM-MS provide are shown.

IM-MS, in conjunction with coarse-grained modelling approaches, has enabled distinction between topologies for two subcomplexes in eukaryotic initiation factor 3 (*145*), fitting of substructures of RNA polymerase III into EM density maps (*192*) and modelling of the single stranded binding protein of the *E. coli* replisome (*53*). Recently the architecture of polydisperse αB-crystallin was interrogated using IM-MS to filter candidate structures, which were built using restraints from NMR, MS and symmetry (*142*).

In this chapter, the advantages and limitations are explored for using IM-MS to build 3D models of protein complexes from information derived from the CCS of intact complexes and subcomplexes. These subcomplexes are obtained from solution-phase experiments designed to perturb protein interfaces using chaotropic agents or manipulation of ionic strength. Information on the overall topology of a protein complex can be obtained from CCS of the intact complex. Stable subcomplexes however can be used to reveal the building blocks of the complex. These building blocks can then be assembled into a model of the intact complex, guided by CCS restraints.

The CCS restraint was incorporated into a scoring function, with volume restraints which were derived from crystal structures of protein complexes. The performance of the scoring function was critically assessed, using a homotetramer for benchmarking. This validated the use of CCS for building low-resolution models of protein assemblies. To illustrate this modelling approach, models were built and scored for heteromers tryptophan synthase and NBDO, based on volume and experimentally determined CCS.

Finally a topological model was predicted for the tetrameric intermediate in the aggregation of ΔN6-truncated β2-microgobulin, whose structure is still unknown.

4.2 DEVELOPING A MODELLING APPROACH

4.2.1 Overview

To apply the information generated from IM-MS of protein assemblies and their subcomplexes to structure characterisation, a coarse-grained modelling approach is used. Proteins are represented as single spheres and subcomplexes as a set of overlapping spheres with radii derived from subunit CCS, or where unavailable, from subunit mass, given the spherical density. Models for protein complexes are built in a stepwise manner, by combining subunits and subcomplexes, which make up the assembly. Candidate models are generated by an exhaustive search of conformational space, using a Monte Carlo approach (>3 subunits) or a local planar search (≤3 subunits). Experimental data are translated into spatial restraints, which are used to define a scoring function. This enables evaluation of the ensemble of structures that satisfy a given set of restraints. By minimising the scoring function, and through cluster analysis of the best scoring models, predictions can be made about the native structure (Figure 4.2).

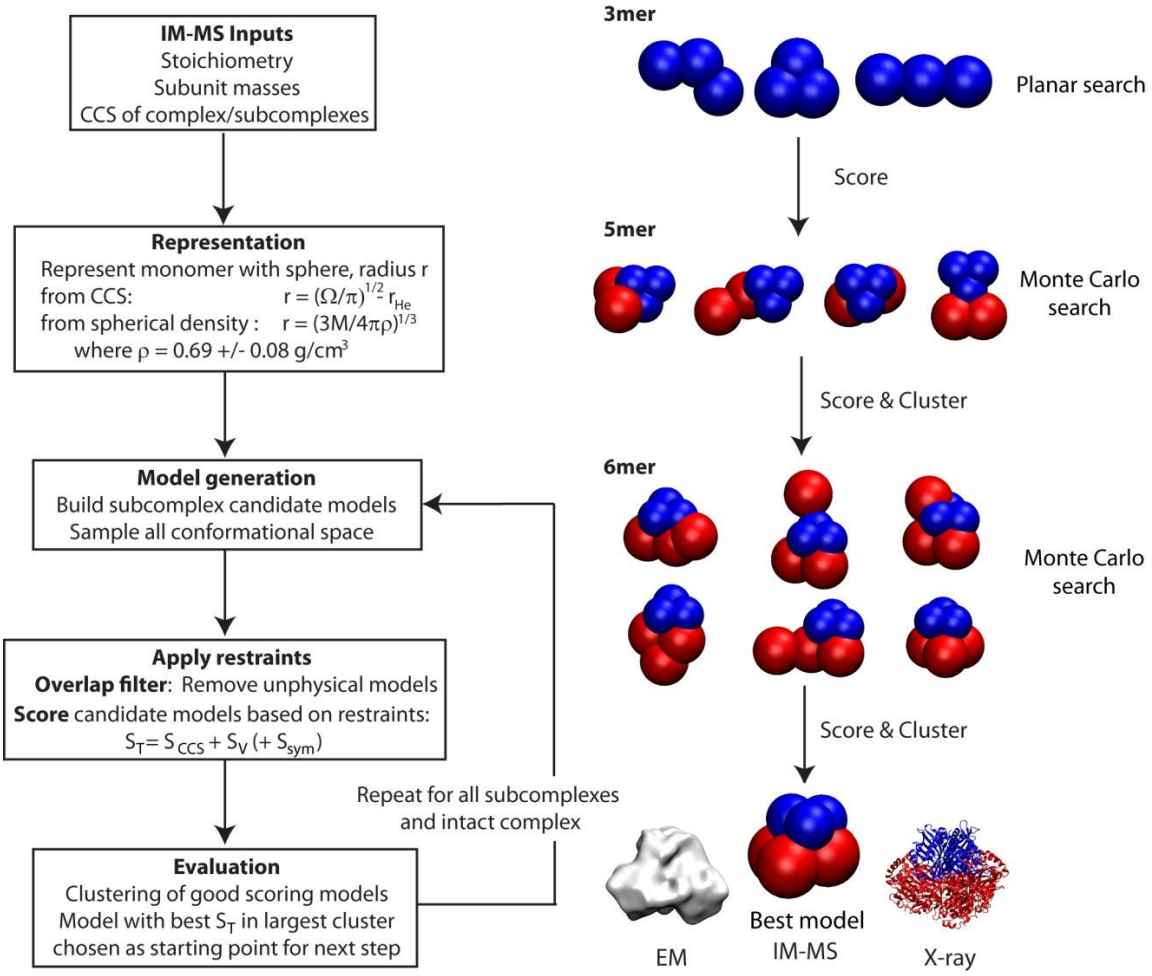


Figure 4.2. Overview of the modelling strategy. Information generated from IM-MS are converted into spatial restraints which are applied to a computational approach for generating and evaluating model structures. Coarse-grained model structures were first built for subunits and subcomplexes, before being assembled into a model for the intact complex. Different relative orientations and positions were sampled and models scored, based on violations of CCS (S_{CCS}), volume (S_V), and where applicable, symmetry (S_{sym}) restraints. The total score, S_T, is given by the sum of individual restraints. Good scoring models are clustered in order to identify the best scoring model within the largest cluster to take to the next step.

4.2.2 Representation

A coarse-grained representation was chosen for modelling protein assemblies as IM-MS experiments provide low-resolution structural information; these are the 3D-orientation of subunits within an assembly and/or the overall topology. This modelling approach has the additional advantage that calculations are computationally inexpensive, which is particularly important for the study of very large and complex systems. In order to validate this approach for representing protein complexes, low-resolution models were built from coordinates in the PDB and used as a training set of complexes. Fifteen protein complexes, both homomeric and heteromeric, were selected that span a wide range of masses, topologies and subunit numbers. These complexes were represented as a set of overlapping spheres (*145*). The centres of the spheres were determined by the subunit centre of mass, and sphere radius defined by the subunit calculated CCS.

To evaluate the accuracy of coarse-grained representations (i.e. closeness of CCS to atomic structure), the CCS of the low-resolution models were computed and compared with the CCS calculated from the corresponding all-atom structures (*92*). Deviation between these values was 2.0 ± 1.1 % (Figure 4.3) indicating that a set of overlapping spheres offers a valid representation of the protein complex structure.

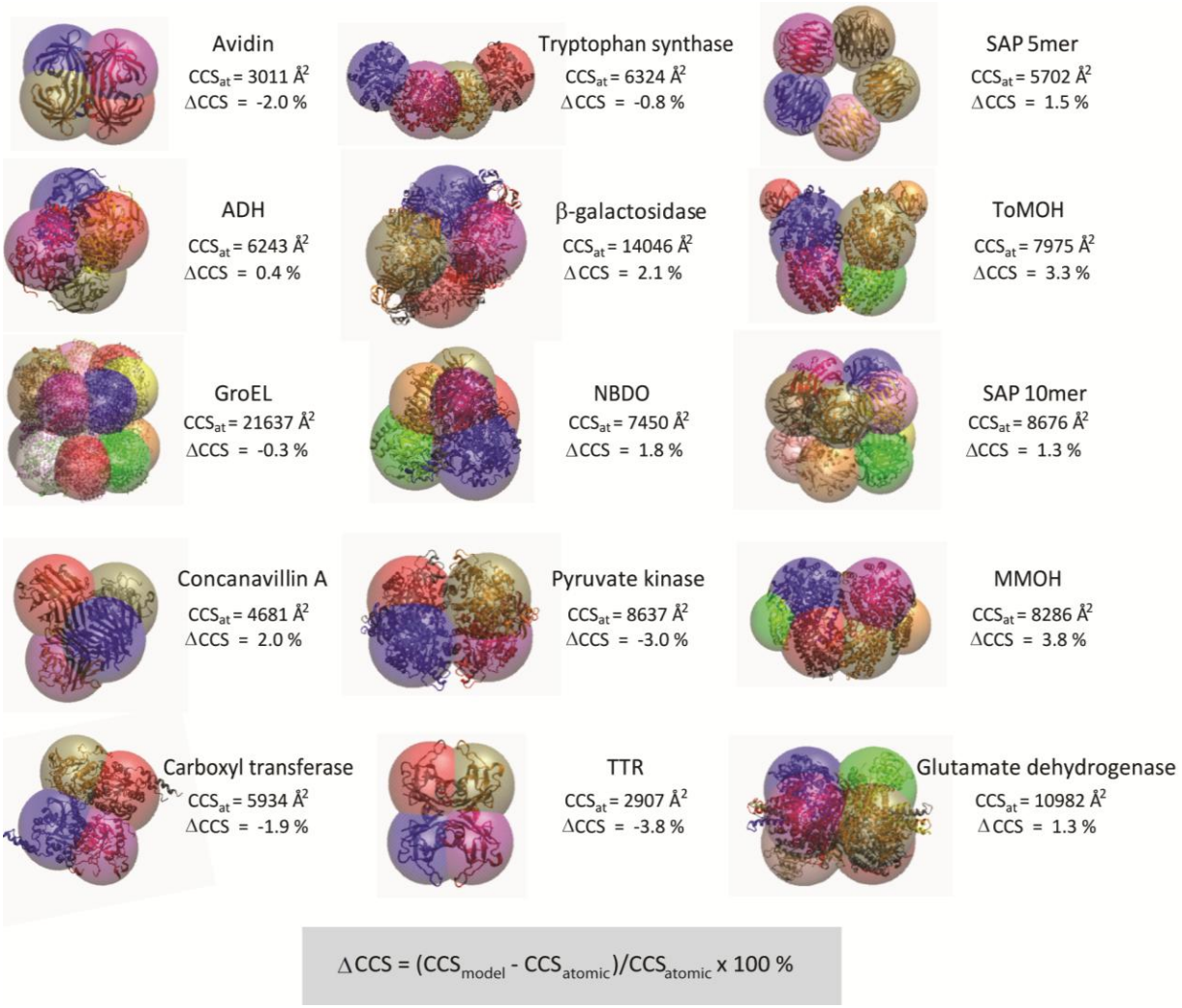


Figure 4.3. Validating coarse-grained representation for protein complexes. Coarse-grained models were prepared for fifteen protein complexes using their X-ray crystallographic information from the PDB. The sphere centres were determined by the subunit centre of mass, and the sphere radius defined by the subunit calculated CCS. The calculated CCS of the atomic structure is given, and compared with that of the low resolution model. Model accuracy is denoted as $\Delta CCS = (CCS_{model} - CCS_{atomic})/CCS_{atomic} \times 100$ %.

4.2.3 Spatial restraints: overlap and volume

The translation of the available experimental data into restraints is one of the most crucial aspects of the modelling process (*184*). When building models from IM-MS with no atomic-level information, the positions for the spheres representing subunits are randomly generated, and sample many different relative orientations and conformations.

The extent to which the spheres overlap can be used as a filter to remove models with spheres loosely attached or unattached. Similarly those models with spheres overlapping to an unrealistic extent, which would result in steric clashes, can also be discounted. From the models above generated using the crystallographic positions of atoms, the average linear overlap (O) of interacting spheres A and B was calculated using the following equation, where r_A and r_B are the radii of two spheres, and d is the distance between the sphere centres.

$$O = \frac{r_A + r_B - d}{r_A + r_B} \text{ x } 100\ \% \tag{4.3}$$

The average distance overlap of interacting spheres in the set of fifteen models examined here was 29 ± 10 %. This information, with an additional tolerance, was applied as a filter, whereby models with interacting spheres which overlap by < 15 % and > 45% were considered to be "unphysical" and excluded with no further analysis.

Alternative arrangements of spheres can take up different volumes. In order to incorporate volume restraints into the model building process, an assessment of the volume of the models for the fifteen complexes used above was carried out. This provided a measure of the packing densities of the spheres representing proteins in the coarse-grained models. Models were prepared for the fifteen complexes at varying levels of resolution (representations for TTR are shown in Figure 4.4). At the lowest level, the whole complex was represented by a single sphere. At higher levels of resolution, the protein complex was represented as a set of overlapping spheres, where one sphere represented a subunit, a domain, or a residue. At the highest level of resolution was the atomic representation from the X-ray crystal structure. The centres of the spheres were determined using the centre of mass of the group of atoms to be represented by a single

sphere (whole complex, subunit, domain), the α-carbon (residue level) or atomic positions, depending on the level of resolution required. The radii of the spheres were scaled such that the CCS of the assembly model matched that of the CCS for the atomic structure.

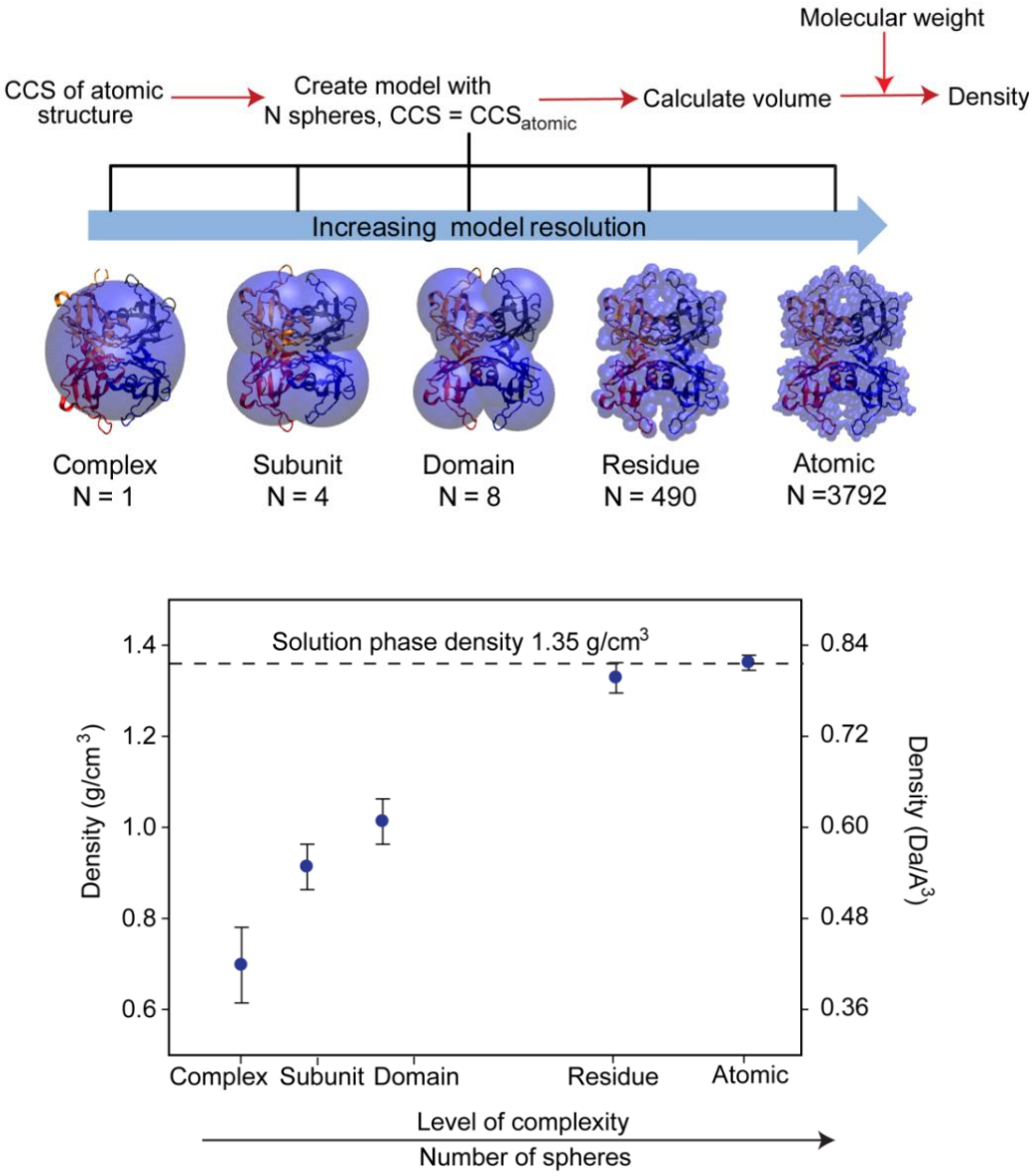


Figure 4.4. Relationship between average spherical density and model resolution (degree of complexity). The density can be used to set volume restraints when building models with different levels of information available. Error bars represent the standard deviation for the 15 protein complexes. Representations for TTR are shown with increasing model resolution.

Volumes, *V*, were calculated for the models and spherical packing densities, ρ_{sph}, derived using the following equation, where N_A is Avogadro's number, and MW the molecular weight in Da.

$$\rho_{sph} = \frac{MW}{V.N_A} \tag{4.4}$$

The average spherical packing density increased with model resolution, reaching a maximum at the atomic-level resolution (Figure 4.4). When representing a protein or protein assembly as a single sphere, the average spherical packing density was calculated as 0.69 ± 0.08 g cm^{-3} (0.42 ± 0.05 Da Å^{-3}). Surprisingly the standard deviation of packing densities was low (5-12 %), despite complexes showing a wide range of complexity, heterogeneity and size. Experimental studies using gas-phase electrophoretic mobility (*193*) and linear drift tube IM-MS studies (*194*) suggest effective ion densities of 0.58 ± 0.05 g cm^{-3} and 0.61 ± 0.07 g cm^{-3} respectively, which agree well with spherical packing densities calculated from these single sphere models.

It is important to note that whilst these models represent the atomic structure of the protein complex accurately with respect to CCS, the same is not true of total volume. Low-resolution spherical representations, built using CCS data to define the radius, overestimate the volume, giving a lower density than in solution. This behaviour is attributed to the nature of the sphere surface, which is smooth, unlike the protein complex surface. In this way, the surface area to volume ratio increases with increasing model resolution. Thus, at the subunit level, spherical packing density increases to a mean of 0.91 ± 0.05 g cm^{-3} (0.55 ± 0.03 Da Å^{-3}). The packing density at the residue level of resolution was 1.33 ± 0.03 g cm^{-3}, in agreement with protein solution phase density of 1.35 g cm^{-3} (*195*).

These values can be used when building and scoring models. For example, consider a protein assembly where the CCS of a monomer is not available, in cases where it is not formed readily in solution. The radius of the sphere representing this monomer can be defined using its mass, and the spherical packing density value appropriate to the model, defined above. Similarly, the expected volume of the assembly can be used as a restraint in order to score candidate model structures through the implementation of a harmonic penalty (Eq. 4.5).

$$S_V = \left(\frac{V_{exp} - V_{calc}}{\sigma_{exp}}\right)^2 \quad (4.5)$$

where the volume score, S_v, is the closeness-of-fit between the expected and calculated volume, V_{exp} and V_{calc} respectively. The parameter σ_{exp} regulates the strength of the restraint and is determined by the uncertainty in the data. In this case the error is defined as two standard deviations from the mean (± 10%). In this way overlap between spheres, spherical packing density and volume of the assembly are used to filter and restrain the candidate models.

4.2.4 The CCS restraint

The CCS restraint is defined using the experimentally determined CCS, implemented as a harmonic penalty, where perfect agreement between model and experimental CCS would take a value of 0; violations of the restraint would result in higher values. The CCS restraint is implemented as shown in Equation 4.6.

$$S_{CCS} = \left(\frac{CCS_{exp} - CCS_{calc}}{\sigma_{exp}}\right)^2 \quad (4.6)$$

where the CCS score, S_{CCS} is given by the closeness-of-fit between the experimental CCS and calculated values (scaled CCS_{PA}), CCS_{exp} and CCS_{calc} respectively. σ_{exp} is the

experimental error in the data. In these IM experiments, in which a linear drift tube was used, the measured CCS accuracy is estimated to be <3 %. In order to ensure realistic errors, σ_{exp} of ± 6 % is used.

The total score, S_T is calculated by summing the scores resulting from the individual restraints.

$$S_T = w_V S_V + w_{CCS} S_{CCS} \tag{4.7}$$

where *w* is the weighting factor for each restraint. For volume and CCS restraints however, the strength of each restraint is regulated by σ_{exp}, the uncertainty in the data. Greater uncertainty in the data results in a lower score, and therefore these functions have in-built weighting, and consequently *w* = 1 is appropriate for these restraints.

4.2.5 Symmetry restraint

Symmetry within protein complexes is very common (*196*). From an investigation of 930 heteromeric protein complexes, ranging in size and number of subunits (N = 3-26), 86.1 % with multiple subunit copies showed symmetrical connectivity (*197*). Thus a symmetry restraint was incorporated into the scoring function. The symmetry scoring function implemented for use on spherical structures relies on assessing the symmetry of the models from the distribution of distances between the subunits. S_{sym} is the sum of the root mean squares of the differences between equivalent distances in symmetry units, where *N* is the number of subunits, d_{ir} and d_{jr} are distances between equivalent subunits, *i* and *j*, and a reference point, *r*.

$$S_{sym} = \sqrt{\frac{1}{N}\sum_{i<j}^{N}(d_{ir} - d_{jr})^2} \tag{4.8}$$

Thus models are scored in terms of the differences in the distances of equivalent subunits to the centre of mass (COM). For example, consider a complex consisting of two types of subunit, α and β (Figure 4.5):

$$S_{sym} = S_{sym(a-COM)} + S_{sym(b-COM)} \tag{4.9}$$

Structures with symmetrical connectivity are scored close to 0, whilst those with higher scores are asymmetrical. The symmetry restraint is incorporated into the scoring function as follows:

$$S'_T = w_V S_V + w_{CCS} S_{CCS} + w_{sym} S_{sym} \tag{4.10}$$

where *w* is the weighting of each restraint. In this study, an assumption has been made that all restraints are equally weighted ($w = 1$).

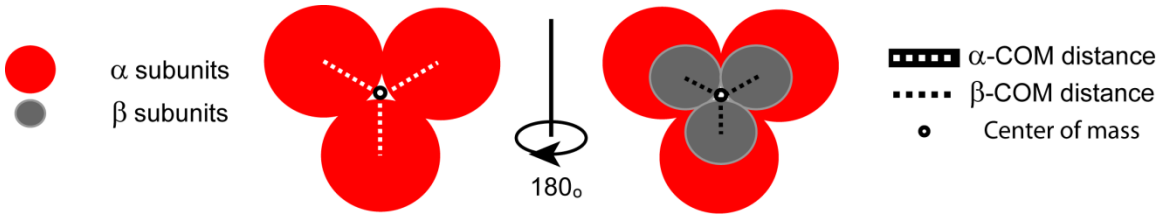


Figure 4.5. Symmetry restraint for $\alpha_3\beta_3$ complex.

4.3 RESULTS

4.3.1 IM-MS of subcomplexes

The use of the CCS of subcomplexes as a restraint in this modelling approach requires that the solution conditions used to perturb the inter-subunit interactions do not affect their tertiary structure. To investigate the effect of solution conditions on CCS, six heteromeric complexes were selected, with varying topologies and interface areas. X-ray crystal structures were available for all six; IM-MS data was obtained for the same protein constructs that were used to obtain the X-ray crystallographic data. Subcomplexes were generated for these six heteromeric protein complexes by disruption in solution. In some cases, subcomplexes could be detected from the same solution conditions as the intact complex. For others, subcomplexes were formed by the addition of organic solvent, by manipulating the pH, or by increasing the ionic strength.

The experimental CCS of all subcomplexes and parent complexes were compared to CCS values calculated from their atomic structures (Table 4.1). Despite an overall good agreement for 18 of the 21 complexes and subcomplexes, some anomalies were found. For example, the measured CCS for intact MMOH (6-mer 251 kDa) complex was greater than predicted (Figure 4.6A). The experimental CCS was however within 10 % deviation from the calculated CCS. Despite differences in CCS for the intact hexameric complex, the $\alpha\beta$ subcomplex and the α monomer showed 5.2 and 3.0 % deviation respectively from the calculated CCS.

Table 4.1. Experimental and calculated CCS for heteromeric complexes and their subcomplexes.

Complex/ Subcomplex	Mass (kDa)	[a]CCS_{exp} ($Å^2$)	[b]Error ($Å^2$)	CCS_{PA} ($Å^2$)	CCS_{EHSS} ($Å^2$)	[c]CCS_{calc} ($Å^2$)	[d]ΔCCS (%)
MMOH							
$\alpha_2\beta_2\gamma_2$	251	10586	318	8304	11135	9618	10.1
$\alpha\beta$	105	5531	166	4550	5990	5256	5.2
α	60	3785	114	3367	4367	3904	-3.0
NBDO							
$\alpha_3\beta_3$	218	8765	263	7450	10024	8578	2.2
$\alpha_2\beta_3$	168	7462	224	6621	8882	7616	-2.0
β_3	69	4043	121	3732	4926	4254	-5.0
α	49	3302	99	2905	3779	3361	-1.8
ToMOH							
$\alpha_2\beta_2\gamma_2$	212	9663	290	7975	10669	9211	4.9
$\alpha\beta\gamma$	106	5578	167	4639	6124	5357	4.1
γ	10	1107	33	936	1168	1091	1.5
tryptophan synthase							
$\alpha_2\beta_2$	143	7629	229	6324	8397	7260	5.1
$\alpha\beta_2$	114.5	6181	185	5157	6829	5932	4.2
β_2	85	4717	142	3978	5233	4584	2.9
α	28	2203	66	1879	2402	2147	2.6
acetyl coA carboxylase carboxyltransferase	135	6708	201	5934	7595	7076	-5.2
$\alpha_2\beta_2$	67	4127	124	3604	4630	4297	-4.0
$\alpha\beta$	37.5	2762	83	2402	3099	2897	-4.7
α	31	2343	70	2147	2745	2562	-8.6
β							
carbamoyl phosphate synthetase							
$\alpha_4\beta_4$	640	18766	563	19968	26659	22764	-17.6
$\alpha_2\beta_2$	320	12835	385	11927	15910	13597	-5.6
$\alpha\beta$	160	7380	221	6322	7207	7207	2.4

[a] CCS_{exp} is the average of all charge states (standard deviation <1-3 %), and replicates taken on different days and from different solution conditions.

[b] CCS measurement error on the RF-confining linear drift tube is estimated to be ± 3%

[c] $CCS_{calc} = 1.14\ CCS_{PA} .(m_{exp}/m_{pdb})^{2/3}$, where m_{exp} is experimental mass and m_{pdb} is mass calculated from the crystal structure

[d] $\Delta CCS = (CCS_{exp} - CCS_{calc})/CCS_{calc} \times 100\ \%$

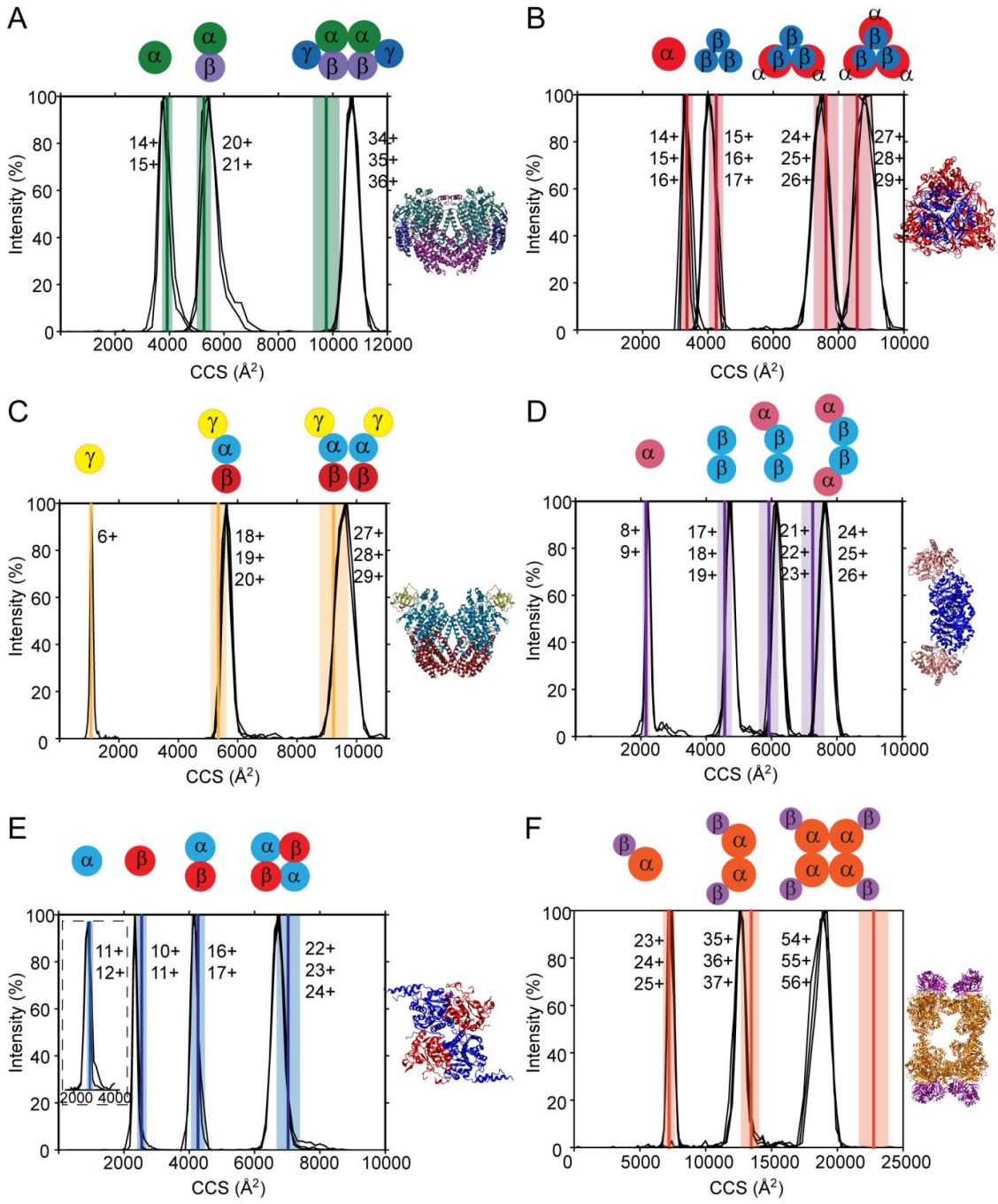


Figure 4.6. Ion mobility arrival time distributions, shown on a CCS axis, for six heteromeric complexes and their subcomplexes. Different charge states observed have been overlaid. The solid line and associated shaded region represents the calculated CCS ± 6 %. **A** MMOH (200 mM AA) and subcomplexes (200 mM AA, 37 % acetonitrile); **B** NBDO (250 mM AA) and subcomplexes (250 mM AA, 25 % butanol); **C** ToMOH (200 mM AA) and subcomplexes (4 M AA); **D** Tryptophan synthase (200 mM AA) and subcomplexes (200 mM AA, 30% MeOH, 10% DMSO); **E** Acetyl coA carboxylase carboxyltransferase (250 mM AA) and subcomplexes (250 mM AA, 30% MeOH, 0.01 %

acetic acid); **F** Carbamoyl phosphate synthetase and subcomplexes (150 mM AA, 1 μM ornithine, 23 μM AMP-PNP, 34 μM Mn^{2+}). X-ray crystal structures for the intact complexes are shown alongside. The complexes from **A** to **F** are in order of decreasing relative interface strength.

NBDO (6-mer, 218 kDa), composed of a dimer of two trimers ($\alpha_3\beta_3$), forms three subcomplexes in 25 % butanol, two of which contain the close-packed β_3 trimer (Figure 4.6B). For the intact complex in buffer, the experimental value deviated 2.2 % from the calculated CCS. Experimental CCSs for subcomplexes $\alpha_2\beta_3$ and β_3 deviated 2.0 and 5.0 % from their corresponding calculated values. The measured CCS of the α monomer, released in solution, was 1.8 % lower than that calculated for the α chain when bound in the complex. These results imply that even in 25 % butanol, the overall topology and structure of the subcomplexes is not perturbed.

ToMOH (6-mer, 212 kDa) is composed of a dimer of trimers $(\alpha\beta\gamma)_2$ in a linear arrangement (Figure 4.6C). ToMOH undergoes disassembly into two αβγ trimers, followed by the loss of γ upon increase of ionic strength (> 1 M AA). CCS measurements for ToMOH and associated subcomplexes showed 1.5 - 4.9 % deviation from the calculated CCSs.

Similar results were obtained for two tetramers, tryptophan synthase (143 kDa), a symmetrical αββα complex with a near-linear conformation and acetyl coA carboxylase carboxyltransferase, a close-packed (135 kDa) complex formed from two αβ dimers (Figures 4.6D and E). Disruption in solution resulted in the formation of various subcomplexes including an extended trimer, as well as dimers and monomeric subunits.

All these showed excellent agreement between experimental and calculated values, with < 5.2 % deviation for seven of the eight species generated. The only exception was the β monomer of the carboxyltransferase, which deviated by 8.6 % from the calculated CCS indicating some collapse, perhaps of the unstructured loop region observed in the X-ray structure of this subunit.

Interestingly the largest complex examined, the carbamoyl phosphate synthetase (8-mer, 640 kDa), with a rectangular planar topology of four α subunits each with an associated β subunit, indicated a smaller conformation in the gas phase than the X-ray crystal structure (Figure 4.6F). A combination of factors, including weak interfaces (Figure 4.7) and an extensive internal cavity, may have led to some degree of structural rearrangement or collapse in the gas phase (*86, 155*). The CCSs of $\alpha_2\beta_2$ and αβ subcomplexes in which the central cavity is no longer present, however, showed excellent agreement with the calculated CCSs (deviation within 2.4 - 5.6 %).

Close agreement was noted for CCSs measured for different charge states of each species (standard deviation <1-3 %) and for replicate measurements (N = 3). Noteworthy is the observation that CCS of complexes or subcomplexes measured in buffer and buffer/organic solution conditions, revealed no differences in experimental values (Figure 4.8). For example, CCS measurements for intact NBDO and subcomplexes were carried out in 250 mM AA and in buffer containing 25 % butanol. CCS measurements in the different solutions for the intact complex, the β_3 subcomplex and the unbound α monomer differed by < 1 % (well within experimental error). This was also the case for other complexes.

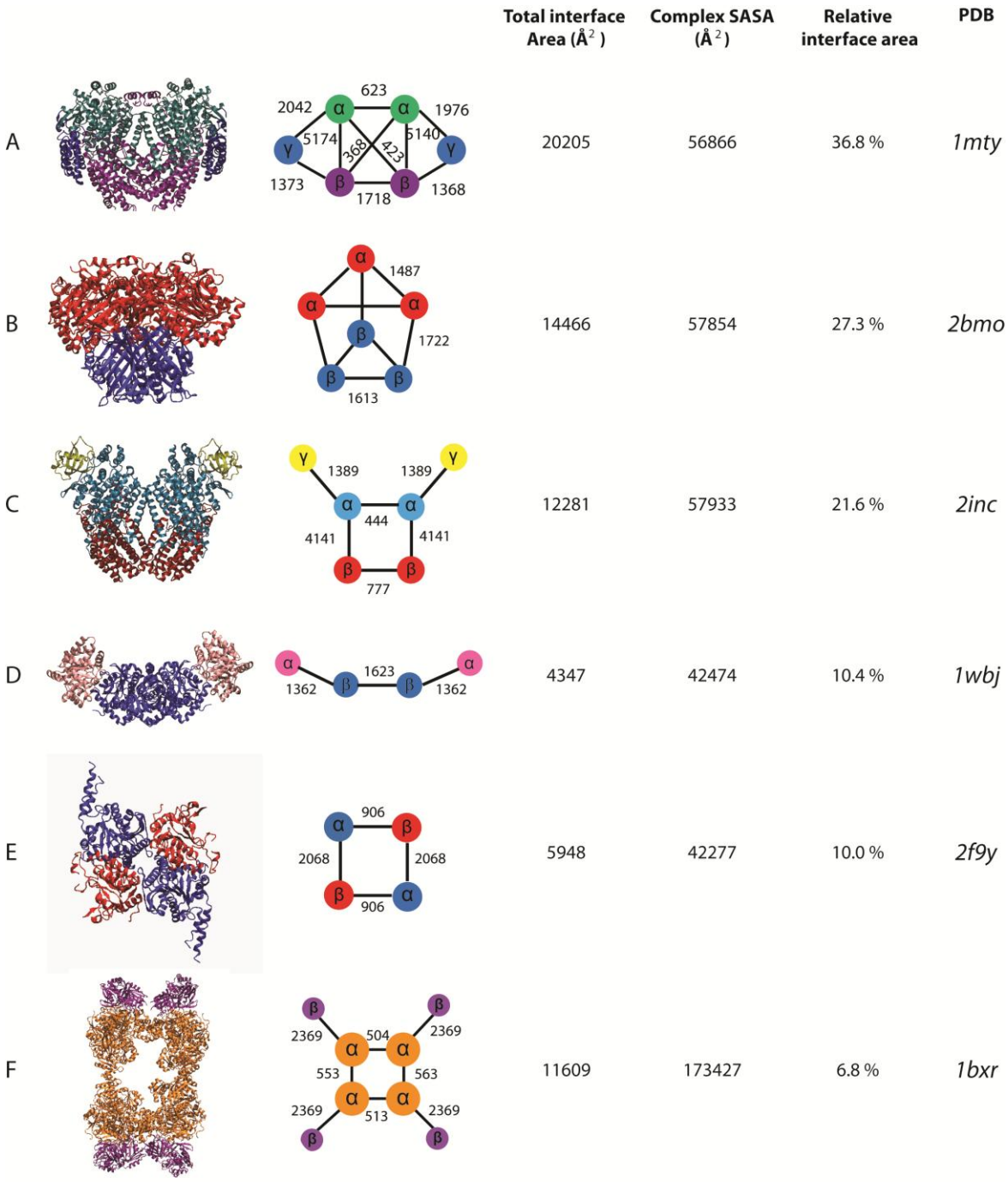


Figure 4.7. Analysis of absolute and relative interfacial areas of heteromeric protein complexes. Complexes from A to F are in order of decreasing interface strength, as defined by the ratio of the total interface area to the solvent accessible surface area (SASA) of the complex. **A** = MMOH, **B** = NBDO, **C** = ToMOH, **D** = tryptophan synthase, **E** = acetyl coA carboxylase carboxyltransferase, **F** = carbamoyl phosphate synthetase.

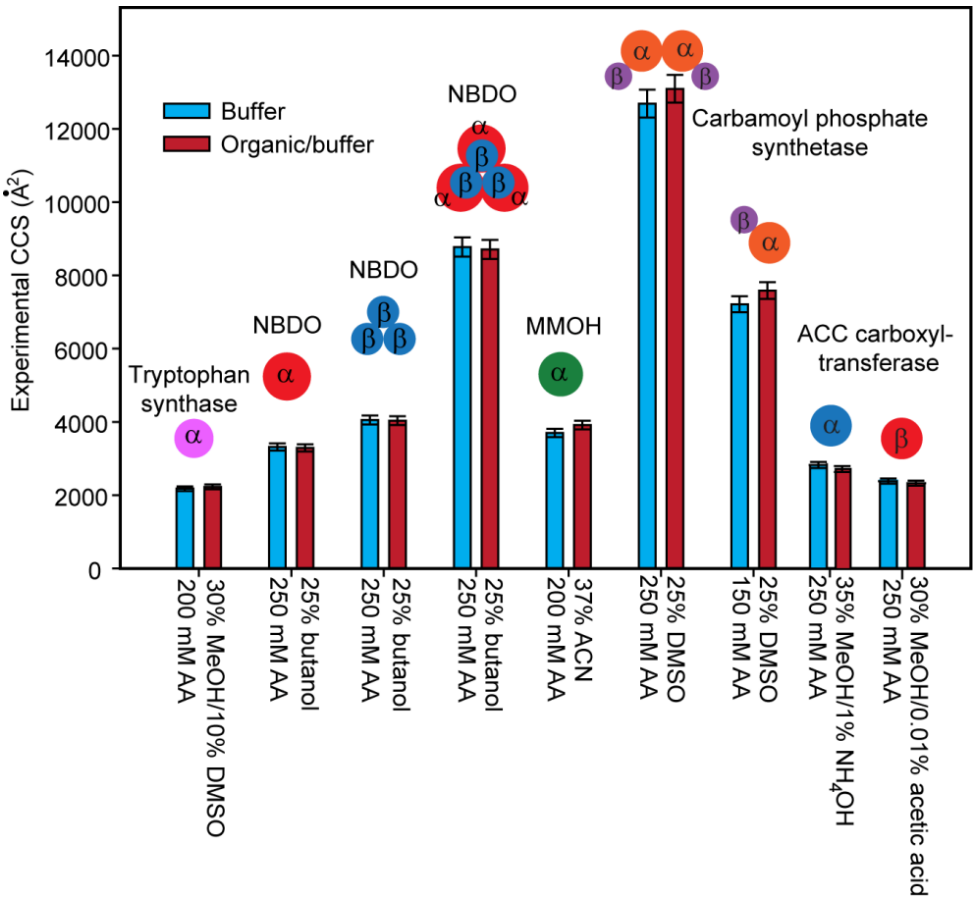


Figure 4.8. Comparison of CCS measurements of subcomplexes in buffer and organic-containing solutions. CCS measurements were either taken in buffer-only (blue) solution conditions or under conditions suitable to perturb subunit-subunit interactions (red). Disruption in solution required the addition of organic solvent to the sample in AA buffer. Experimental CCS is the average of all charge states. Error bars represent the experimental error, which is estimated to be ± 3 %. The species measured and the various subunits/ subcomplexes are shown in the graph.

To further investigate the effects of solution disruptants on protein structure, tryptophan synthase was titrated with destabilising agent (methanol). The relative peak intensities of each species observed in the MS spectra were plotted against increasing organic content in solution (Figure 4.9A). As the % organic is increased, the peaks corresponding to the intact complex decreased in intensity, corresponding to dissociation to the $\alpha\beta_2$ trimer and β_2 dimer, which showed corresponding increases in intensity. As the concentration of organic solvent increased further, only individual subunits were observed. The CCS for

each species was examined with increasing concentration of methanol and was found to be invariant for the intact complex and subcomplexes. In contrast to the analogous gas-phase process where subunit unfolding is typically observed prior to dissociation, solution-phase disruption is characterised by the perturbation of inter-subunit interactions, without unfolding. As the organic content reaches a critical point (in this case 70 % methanol), unfolding of individual subunits was observed. This is characterised by an increase in CCS for α and β, and a broad range of charge states. It is interesting to note, that whilst the CCS of folded species is relatively insensitive to charge state (standard deviation < 1 – 3 %), there is a strong dependence on charge state for the CCS of unfolded α and β species, with CCS increasing with charge state. These experiments demonstrate that in all the cases studied here the CCS of subcomplexes are retained under the different solution conditions used for their generation.

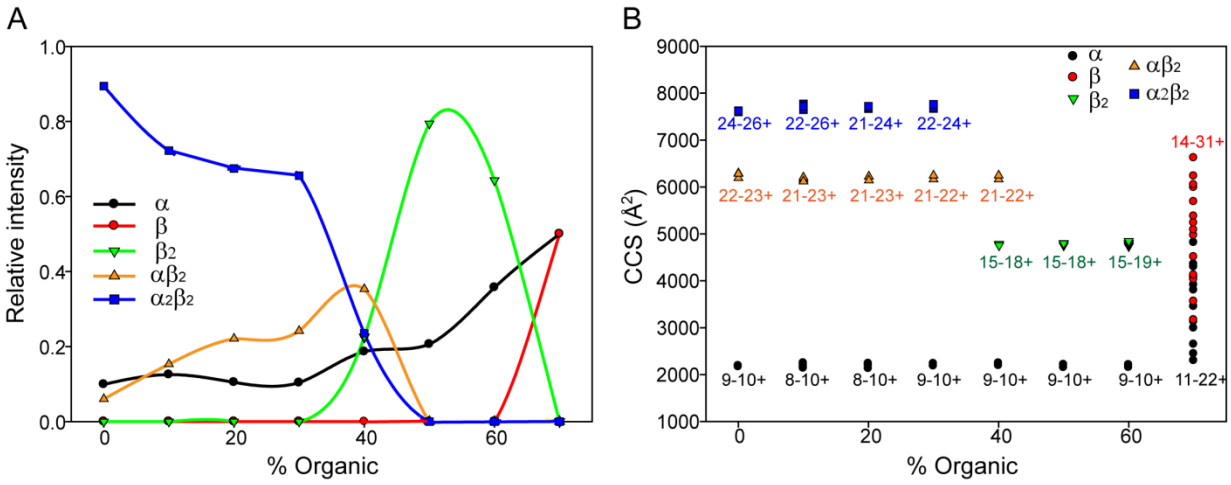


Figure 4.9. Titration of tryptophan synthase with destabilising agent (methanol) reveals a preference for complex and subcomplex dissociation over the unfolding of subunits. **A** Relative peak intensities of each species in the MS spectra are plotted against increasing organic content in solution. **B** CCS is shown for each species as the % organic is increased. Each charge state is plotted individually, with the charge states observed listed below.

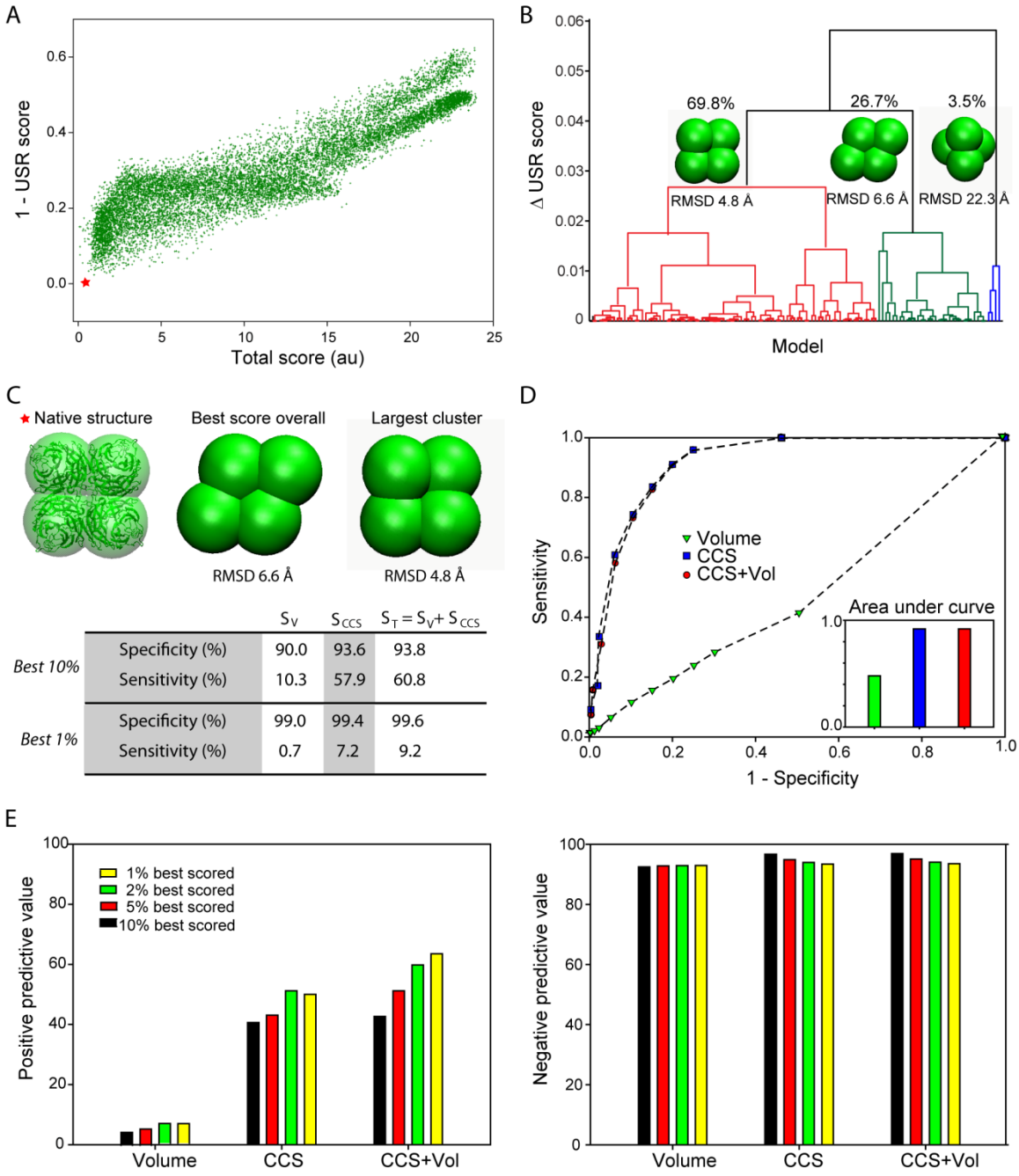
Overall 18 out of 21 (86 %) subcomplexes and complexes had CCSs that differed by < 6 % from their predicted CCS. However, in a minority of cases, gas-phase structural rearrangements, such as the possible collapse of cavity-containing complexes (carbamoyl phosphate synthetase) highlight the limitations of this technique, particularly for large assemblies as noted previously (*86*). Here, combining IM-MS with data from other experimental techniques such as EM or SAXS will be of critical importance. In all cases however, subunit connectivities and topological arrangements of the subcomplexes were conserved under different solution conditions, highlighting their potential as restraints for building and evaluating coarse-grained models of protein assemblies.

4.3.2 Evaluating CCS and volume restraints: Benchmarking on a homomeric complex

For accurate structural modelling of protein assemblies, it is important to capture and evaluate all the available information for the complexes (*123*). This can be achieved by introducing a scoring function, capable of evaluating quantitatively the consistency of the models derived from experimental data and other restraints. To establish the information content of the restraints, both individually and combined, benchmarking was carried out on Influenza B virus neuraminidase (PDB 1VCJ). This is a homotetramer (174 kDa), with a simple square-like topology. Using a Monte Carlo sampling algorithm, 100 000 model structures were generated, with each sphere radius derived from the calculated monomeric CCS. Following a filtering step, to remove any models with overlap < 15% or > 45%, 39 796 models remained, which were subsequently scored, based on CCS and volume restraints.

First the performance of the scoring function was evaluated, by identifying the relationship between the score distribution of the models and their similarity to the native structure, as determined by the USR algorithm (*120*). The USR score compares two models, and ranges from 0 (non-related) to 1 (identical). Based on tests, models with USR > 0.95 were defined as "native-like". A good correlation was found between the score, S_T, and (1-USR score) (Figure 4.10A), whereby native-like models had low S_T.

The 1% lowest scoring (S_T) models were clustered according to their pair-wise USR scores (Figure 4.10B), using a hierarchical tree approach. It is important to note that no knowledge of the native structure is required for clustering, as models are compared with each other. Three clusters were identified, corresponding to distinct configurations. The largest cluster (69.8 %) reflected square-like arrangement of subunits, the second largest (26.7 %) a rhomboidal arrangement, and the smallest cluster (3.5 %), a tetrahedral arrangement. Each cluster is represented by the model with the lowest S_T within the cluster. The RMSD of subunit centres of mass from the native structure was calculated and found to be 4.8 Å, 6.6 Å and 22.3 Å for the lowest scored (S_T) model in the first, second and third largest clusters respectively. This suggests that the native-like structure can be identified as the most frequently occurring predicted model. An additional advantage of clustering analysis is that it provides an indication of the variability in the ensemble of best scoring models. Where a large number of clusters are present, or when no pre-dominant cluster arises, this is an indication that the data is not sufficient to predict the native structure, or there are multiple native states. This is not the case here where close to 70% of the structures were grouped within one cluster.



		S_V	S_{CCS}	$S_T = S_V + S_{CCS}$
Best 10%	Specificity (%)	90.0	93.6	93.8
	Sensitivity (%)	10.3	57.9	60.8
Best 1%	Specificity (%)	99.0	99.4	99.6
	Sensitivity (%)	0.7	7.2	9.2

Figure 4.10. Evaluating the scoring function: benchmarking on a homotetramer. Models for the Influenza B virus neuraminidase were evaluated and scored based on violations of the CCS and volume restraints. The structural similarity of the models to the native structure was evaluated using the USR algorithm and (1-USR score) was plotted against total score, S_T (**A**). The native structure (1-USR score = 0) is indicated by a red star. The 1% lowest scoring models were subjected to a hierarchical cluster analysis (**B**), which revealed three distinct clusters. Clusters are represented by the model with the lowest S_T; centroid RMSD to the native structure is shown. **C** The coarse-grained model built from atomic coordinates is overlaid on the crystal structure, and compared with the model

with the lowest S_T overall, and the model representing the largest cluster. Specificity, (TN/(TN+FP)), and sensitivity (TP/(TP+FN)) were calculated for the 10% and 1% lowest scoring models, where TN is true negative, TP true positive, FP false positive and FN false negative. A "positive" or "native-like" model is defined as a model with USR score > 0.95 relative to the native structure. ROC analysis (**D**) and analysis of positive and negative predictive values (**E**) for the individual and combined restraints. Positive and negative predictive values are given by TP/(TP+FP), and TN/(TN+FN).

4.3.3 Sensitivity and specificity

Evaluation of the individual contribution of each restraint to the scoring function was carried out by performing tests on the sensitivity and specificity (Figure 4.10C). In this context sensitivity may be thought of as the ability of restraints to identify native-like models, while specificity refers to their ability to identify poor models. The 10% and 1% lowest scoring models for Influenza B virus neuraminidase were evaluated. Generally, low sensitivities (<60 %) were observed for the volume, CCS, and combined CCS/volume restraints. On the other hand, all restraints showed very high specificity (>90 %), both individually and combined. Although the scoring function does not capture all good models (low sensitivity), far more critical to the success of the method is a low false positive rate (high specificity). These restraints have proven efficient at excluding poor models, with specificity > 99.0 % in the 1% lowest scoring models.

The ability of the restraints to predict the native structure can be further assessed using receiver operating characteristics (ROC) (*184, 198*). Here the sensitivity is plotted against the false positive rate (1 – specificity) at different scoring cut-off values. The area under the curve gives a measure of the information content of each restraint. An area of 0.5 indicates the restraint cannot discriminate between correct and incorrect topologies. The closer the integrated area is to 1, the more likely the method is able to predict the native

structure. From analysis of the ROC curves volume performs very poorly as a restraint (area under curve 0.48) in contrast to the value of 0.92 for CCS (Figure 4.10D). Combining the volume and CCS restraints does not appear to offer any advantage (area under curve 0.92) further confirming that volume is a poor restraint compared to CCS.

When evaluating model scores (Figure 4.10A), the confidence or probability that a low scoring model is native-like, is described by the positive predictive value. Conversely, the probability that a high-scoring model is not similar to the native structure is given by the negative predictive value. Positive and negative predictive values were plotted for the volume, CCS and combined volume/CCS restraints (Figure 4.10E). All restraints show very high negative predictive value (>90 %), consistent with the high specificity described above. In contrast, the positive predictive value for the restraints is lower, although this increases significantly as the search for good scoring models is narrowed from the best-scored 10 % to 1 %. Interestingly, although the volume restraint offers no positive predictive value on its own, when combined with the CCS restraint higher positive predictive values are obtained than for CCS alone. This is particularly noticeable for the 1% lowest scoring models, where the positive predictive value increases from 50 % (CCS restraint) to 64 % (CCS and volume restraints). Thus the addition of volume restraints to the CCS restraint does offer an advantage.

4.3.4 Application to heteromeric complexes

Having demonstrated that it is possible to determine the topology of a simple homomeric assembly using CCS restraints, and explored the limitations of the technique, the next step was to model heteromeric assemblies, using experimental data. Two complexes were

selected to evaluate the modelling approach: tryptophan synthase, an αββα complex assembled in an extended linear conformation, and NBDO consisting of stacked α_3 and β_3 rings assembled into a compact $\alpha_3\beta_3$ heterohexamer. The distinct topological arrangements and the availability of high resolution X-ray data for the intact assemblies make these two complexes excellent candidates to evaluate the power, as well as the pitfalls, of this approach. IM-MS experiments were carried out on these two complexes as described above, in aqueous buffer and in destabilising solution conditions. CCS was measured for the intact complexes and their associated subcomplexes and subunits (Figure 4.11).

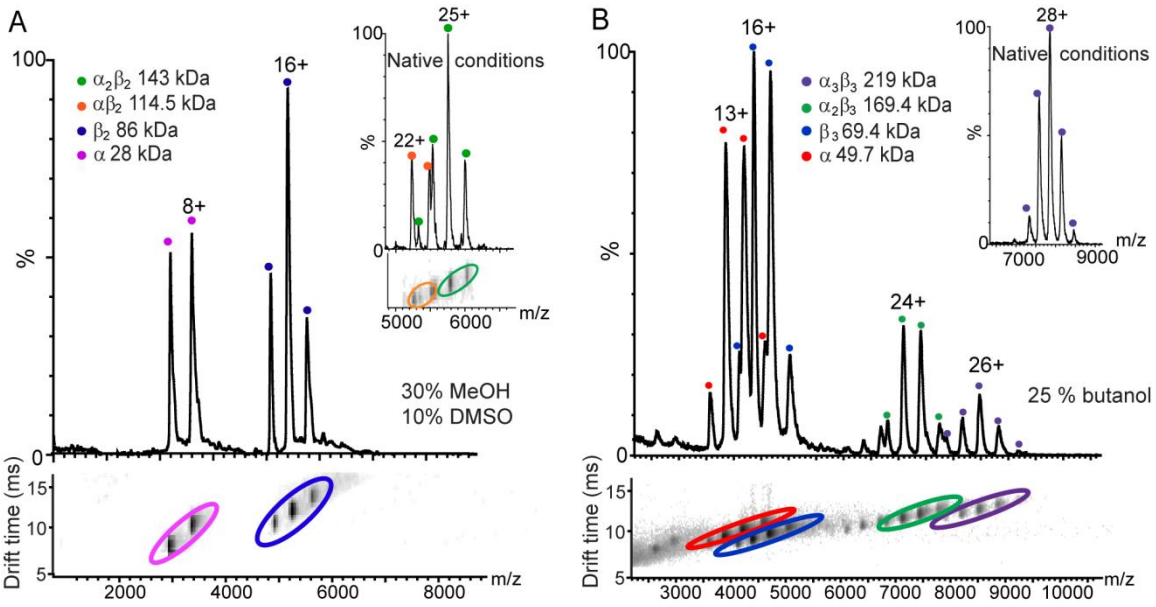


Figure 4.11. Solution disassembly of tryptophan synthase and NBDO. MS spectrum and drift time *versus m/z* contour plot for tryptophan synthase in 30% methanol, 10% DMSO (**A**). The intact complex (green), the $\alpha\beta_2$ subcomplex (orange), the β_2 subcomplex (blue) and the α subunit (pink) are shown. MS spectrum and drift time *versus m/z* contour plot for NBDO in 25% butanol (**B**). Charge state series were identified as corresponding to the intact complex (purple), the $\alpha_2\beta_3$ subcomplex (green), the β_3 subcomplex (blue) and the α subunit (red). MS spectra for the complex in native-like conditions (200-250 mM AA buffer) are shown in the insets.

4.3.5 Tryptophan synthase

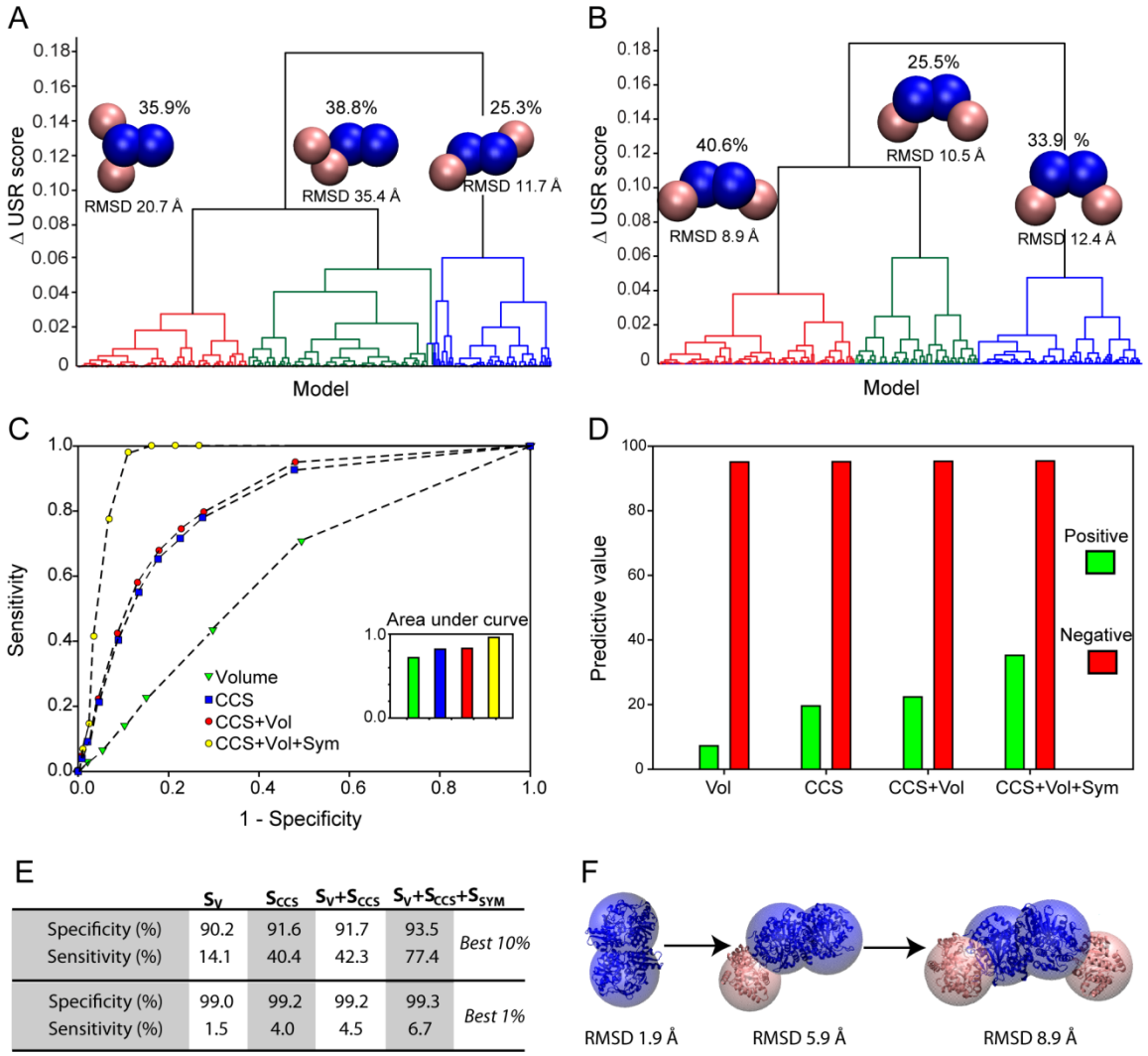
The radius of the α–subunit sphere was derived from the experimental CCS. No CCS information was available for β monomer however. Therefore the radius of the sphere representing the β subunit was defined using the average spherical packing density (0.69 ± 0.08 g cm^{-3} for a protein represented by a single sphere as detailed above). Models were generated for the β_2 subcomplex, by varying the overlap between β-subunits within the upper and lower bounds set by the overlap filter. The 35 resulting models were scored based on violations of CCS and volume restraints. The lowest scored model was chosen as the starting point to dock the α subunit and consequently to build models for the $\alpha\beta_2$ trimer subcomplex. By varying the overlap between spheres representing α and β subunits, whilst simultaneously rotating the outer two subunits about the central subunit, from a linear arrangement (180 ^{o}) to a compact trigonal arrangement (60 ^{o}), 252 models were generated. The model ensemble was then scored, based on CCS and volume restraints. Due to the low number of models generated in planar searches, no clustering of best scoring structures was necessary and the model with the lowest overall score was taken to the next step, to form the $\alpha_2\beta_2$ subcomplex, by the addition of a further α subunit.

In order to sample conformational space for the tetramer, a Monte Carlo approach was employed. The relative position of the $\alpha\beta_2$ trimer subunits was fixed, whilst the second α–subunit coordinates were randomly generated to cover all possible connectivities and orientations. After filtering 100 000 models for overlap, a total of 32 095 candidate models for the $\alpha_2\beta_2$ tetramer remained. Models were scored, based on violations of the CCS and volume restraints, using a harmonic penalty function. Clustering of the 1%

lowest scoring models revealed three distinct clusters, each of a similar size and with distinct subunit connectivity (Figure 4.12A). Poor RMSD from the native structure (35.4 Å) was obtained for the model representing the largest cluster. This demonstrates that for this complex, the experimental CCS data is insufficient to determine the structure. Further information is required, particularly with respect to the connectivity of subunits.

Knowledge of symmetry may offer an advantage here. One method to obtain information on symmetry is through EM, which provides information on size, shape and symmetry. Whilst cryo-EM is the preferred method for structure determination, 3D construction from negative stain images can be particularly useful for small structures below 100 – 200 kDa (*178*). For example, negative stain EM elucidated the structural organisation of 100 kDa tetramer, geminin, with a resolution of 17.5 Å (*199*). Resulting EM maps can enable the fitting of substructures into assembly density (*200*).

Hypothetical EM density maps were therefore simulated (20 Å resolution) for tryptophan synthase, from which two-fold rotational symmetry in an extended conformation was clearly defined (Figure 4.13). Lower resolution EM maps however cannot inform on the order of subunits within the electron density. For this, CID experiments were used. Here identification of peripheral subunits, which are preferentially lost during dissociation, can be achieved (*45*). During CID, the intact tryptophan synthase complex dissociated exclusively *via* the loss of α-monomer. No loss of a β-subunit was observed suggesting the peripheral nature of α-subunits (Figure 4.13). Combining information on the symmetry with the connectivity from MS and CID experiments, suggests a symmetrical connectivity, with two peripheral α-subunits and a β_2-dimer core.



	S_V	S_{CCS}	S_V+S_{CCS}	$S_V+S_{CCS}+S_{SYM}$	
Specificity (%)	90.2	91.6	91.7	93.5	*Best 10%*
Sensitivity (%)	14.1	40.4	42.3	77.4	
Specificity (%)	99.0	99.2	99.2	99.3	*Best 1%*
Sensitivity (%)	1.5	4.0	4.5	6.7	

Figure 4.12. Structure characterisation of tryptophan synthase 4-mer. The 1% lowest scoring models were clustered on their pair-wise USR scores. 3 clusters were identified for the models scored based on CCS and volume restraints (**A**). The addition of a symmetry restraint reduced variability in the ensemble of best scoring models (**B**). Clusters are represented by the model with the lowest S_T within the cluster; centroid RMSD to the native structure is shown. **C** ROC analysis shows by incorporating additional restraints, the predictive power of the modelling approach is increased. This is also illustrated in **D**, where both positive and negative predictive values increase when more restraints are combined (data for 1% lowest scoring models is shown). **E** Restraint sensitivity and specificity for the 10% and 1% lowest scoring models. **F** The best model for tryptophan synthase (all restraints) and subcomplexes (CCS and volume restraints), is shown, and superimposed on the X-ray crystal structure with corresponding centroid RMSD.

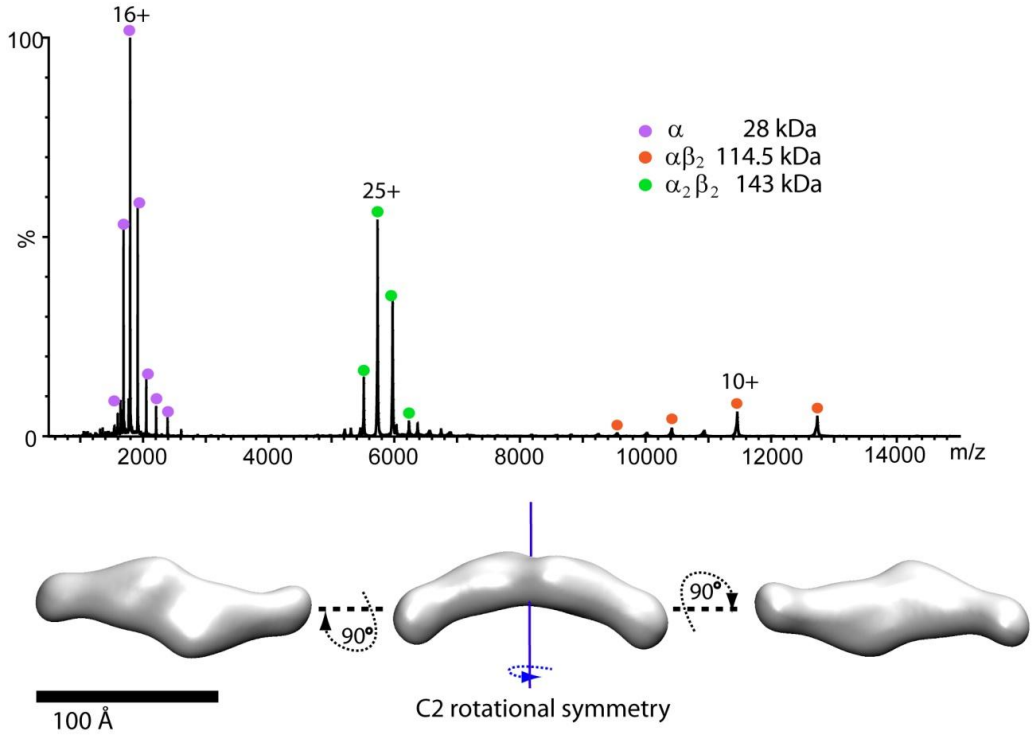


Figure 4.13. CID spectra and simulated EM density maps (20 Å resolution) for tryptophan synthase. Different orientations are shown for the EM maps; the scale bar represents 100 Å.

Re-evaluation of the tryptophan synthase models employing the additional symmetry restraint, based on symmetrical connectivity, revealed three clusters in the 1% lowest scoring models (Figure 4.12B). The RMSDs relative to the native structure were 8.9 – 12.4 Å for these three clusters. In comparison to the results when only CCS and volume restraints were used, there was less structural variability in the ensemble of best scoring models. Moreover the lowest scored model in the largest cluster showed high topological similarity to the native structure.

4.3.6 Nitrobezene dioxygenase

The radii of spheres representing α and β subunits were derived from the experimental CCS, and the average spherical packing density respectively. First, models were built for the β_3 subcomplex. To achieve this, the spherical overlap and the angle between subunits

were simultaneously varied, leading to a total of 595 models satisfying the initial restraints. These models were then evaluated using a scoring function, based on CCS and volume restraints. The β_3-model with the lowest score, with a trigonal topology, was taken to the next step, to form the $\alpha_2\beta_3$-subcomplex, by the addition of two α subunits.

In order to sample conformational space for the pentamer, a Monte Carlo approach was used keeping the relative position of the three β subunits fixed. After filtering 100 000 models for overlap violation, a total of 52 728 models for the $\alpha_2\beta_3$ pentamer remained. Models were scored, based on violations of the CCS and volume restraints. Clustering of the 1% lowest scoring models (Figure 4.14) revealed two distinct clusters. The lowest scored model in the largest cluster was used in the next step, to build the hexamer.

Following the same sampling method as for the pentamer, the relative position of the three β subunits and two α subunits was fixed, while coordinates for the third α subunit were randomly generated. After filtering for overlap violation, 38 465 out of 100 000 models remained for the heterohexamer. Using the CCS and volume restraints, all models were scored based on their deviations from the experimental data. Clustering of the 1% lowest scoring models revealed three clusters (Figure 4.15A). The lowest scored model in the largest cluster (62.1 %) showed excellent similarity to the native structure (centroid RMSD 8.5 Å). Interestingly, and in contrast to the tryptophan synthase example above, in this case of NBDO the CCS and volume restraints are sufficient to define the topological structure.

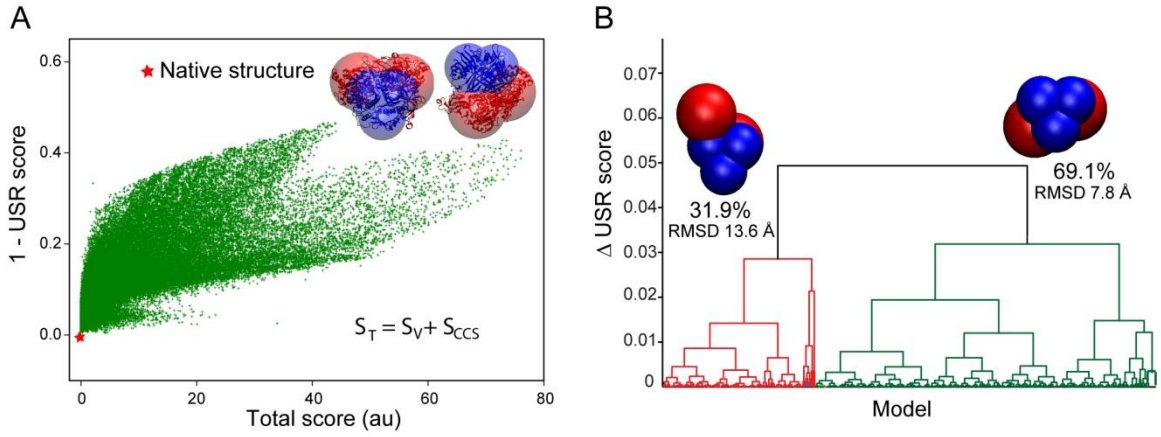
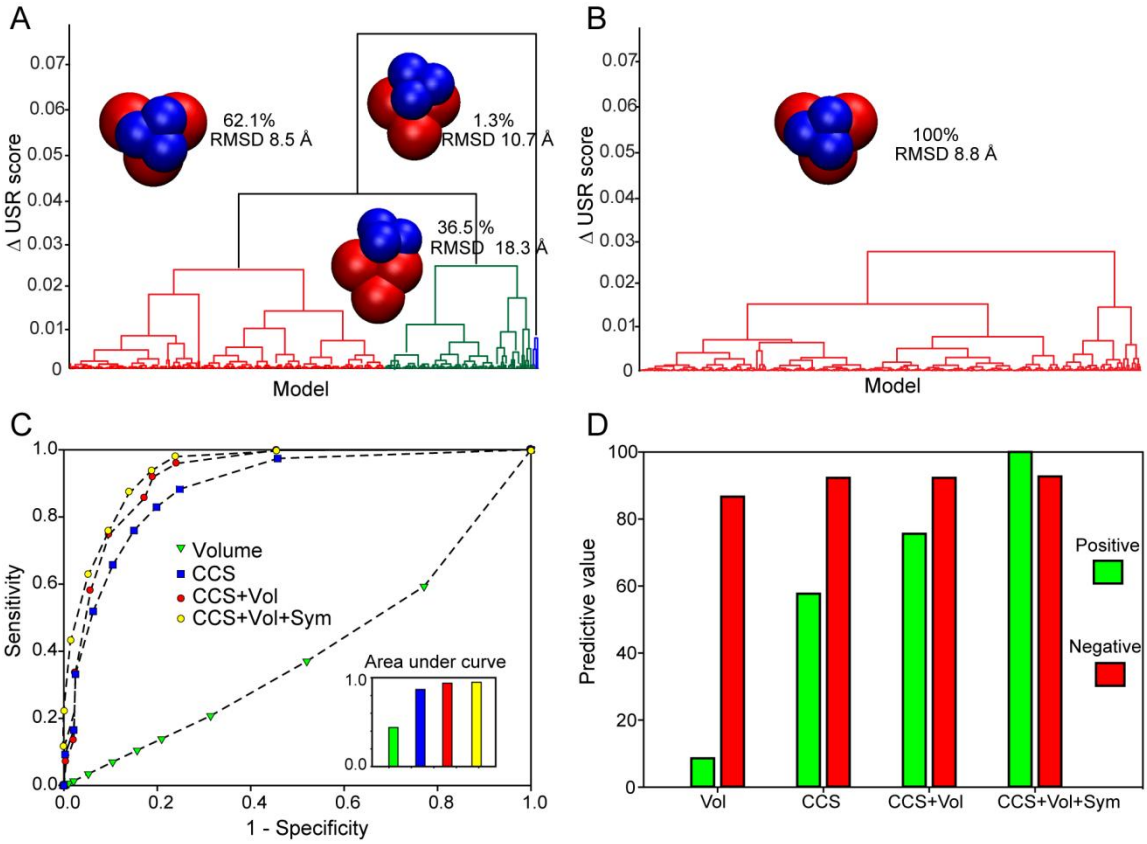


Figure 4.14. A The total score (S_T) distribution for the NBDO 5-mer is plotted against 1-USR score, relative to the native structure. **B** The 1 % lowest scoring models for the NBDO 5-mer were subjected to clustering analysis. Two clusters were identified. Each cluster is represented by the model within the cluster with the lowest S_T and centroid RMSD to the native structure for these are shown.

Knowledge of symmetry was found to be highly informative for refining the search for models for tryptophan synthase. Therefore the improvement that incorporation of a symmetry restraint would hold for modelling NBDO was assessed. Thus EM density maps were simulated (resolution 20 Å) for this complex, from which 3-fold rotational symmetry (C3) was clearly observed. CID experiments, in which the dissociation of α and β monomers were observed, suggest that both α and β subunits are peripherally located (Figure 4.16). Therefore a symmetry restraint, based on symmetrical connectivity, was incorporated into the scoring function, and NBDO 6-mer models were re-evaluated. This time, only one cluster was observed in the 1% lowest scoring models (Figure 4.15B). The lowest scored model in this cluster is closely similar to that found using only CCS and volume restraints (RMSD 8.8 Å from native structure). However, the inclusion of a symmetry restraint to the scoring function greatly reduces the structural variability among the ensemble of best scoring models, thereby increasing confidence when selecting models for an assembly.



E

	S_V	S_{CCS}	S_V+S_{CCS}	$S_V+S_{CCS}+S_{SYM}$	
Specificity (%)	89.5	93.7	94.3	94.7	*Best 10%*
Sensitivity (%)	7.0	51.9	58.0	62.9	
Specificity (%)	98.9	99.5	99.7	100.0	*Best 1%*
Sensitivity (%)	0.7	7.1	9.2	12.2	

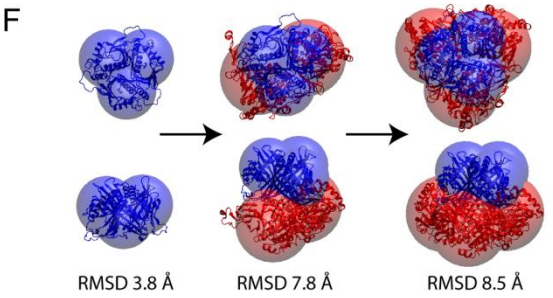


Figure 4.15. Structure characterisation of NBDO 6-mer. The 1% lowest scoring models were hierarchically clustered on their pair-wise USR scores. 3 clusters were identified for the models scored based on CCS and volume restraints (**A**). The addition of a symmetry restraint reduced variability in the ensemble of best scoring models (**B**). **C** ROC analysis shows that by incorporating additional restraints, the predictive power of the modelling approach is increased. In **D**, both positive and negative predictive values increase when more restraints are combined (data for 1% lowest scoring models shown). **E** Restraint sensitivity and specificity for the 10% and 1% lowest scoring models. **F** The best model (top and side view), for intact NBDO and subcomplexes (CCS and volume restraints only), is shown, and superimposed on the X-ray crystal structure with corresponding centroid RMSD.

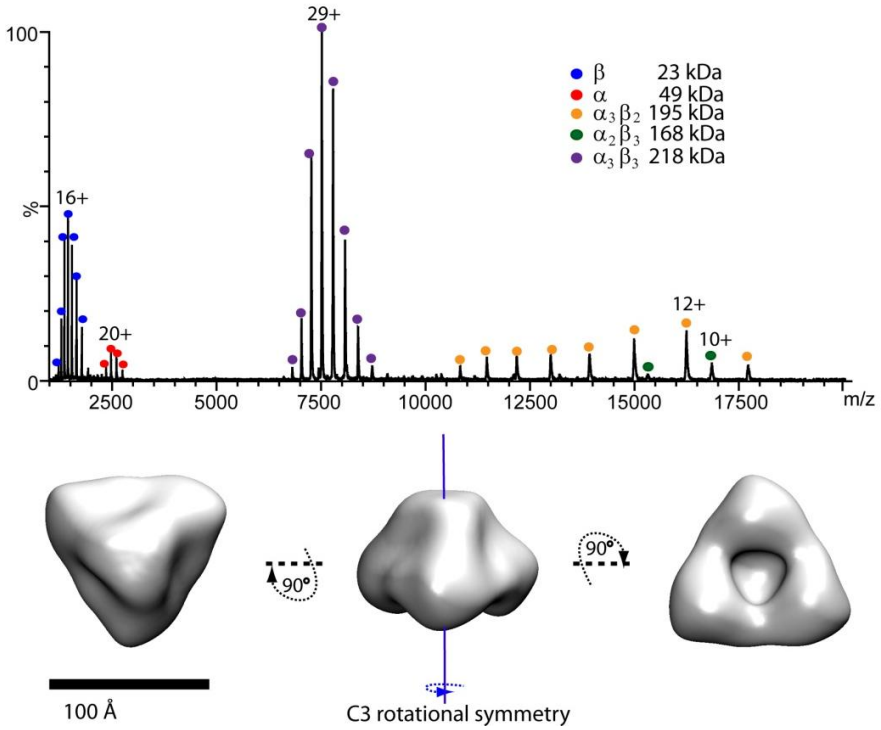


Figure 4.16. CID spectra and simulated EM density maps (20 Å resolution) for NBDO. Different orientations are shown for the EM maps; the scale bar represents 100 Å.

4.3.7 Predicting topology of ΔN6-β2m oligomers

Finally, IM-MS data was used to predict the structure of an assembly whose detailed structure has yet to be resolved. Early aggregation intermediates of ΔN6-truncated β2m isoform were examined. β2m (99-residues, 11.8 kDa) is a key component of the major histocompatibility class I complex. The build-up of β2m, up to 25- to 60-fold, can result as a consequence of long-term haemodialysis (*201*). This leads to amyloid fibril deposition, also known as dialysis-related amyloidosis (DRA). In amyloid deposits extracted from DRA patients, up to 30% of the constituting β2m is ΔN6-truncated (*101*, *202*). ΔN6-β2m has been shown to be highly amyloidogenic at neutral pH, also acting as a fibrillar seed for full-length β2m aggregation (*203*). Understanding the detailed mechanism of the aggregation process requires the identification of all the conformational states and oligomeric structures (*101*). However, the use of traditional

high-resolution structural techniques to study amyloidogenic proteins is challenging due to the transient and heterogeneous nature of the intermediate species, which readily interconvert (*143*, *204-205*). Here IM-MS was applied to study ΔN6-β2m oligomers.

IM-MS of ΔN6-β2m (10 μM, pH 5) revealed equilibria between early oligomeric species, from monomeric up to tetrameric (Figure 4.17). At higher concentrations higher-order oligomers were detected, however differentiating between specific and non-specific aggregation becomes challenging. In general, it has been shown that at low protein concentrations (10 μM and below), non-specific interactions, which arise as a consequence of the ESI process, become negligible (*20*).

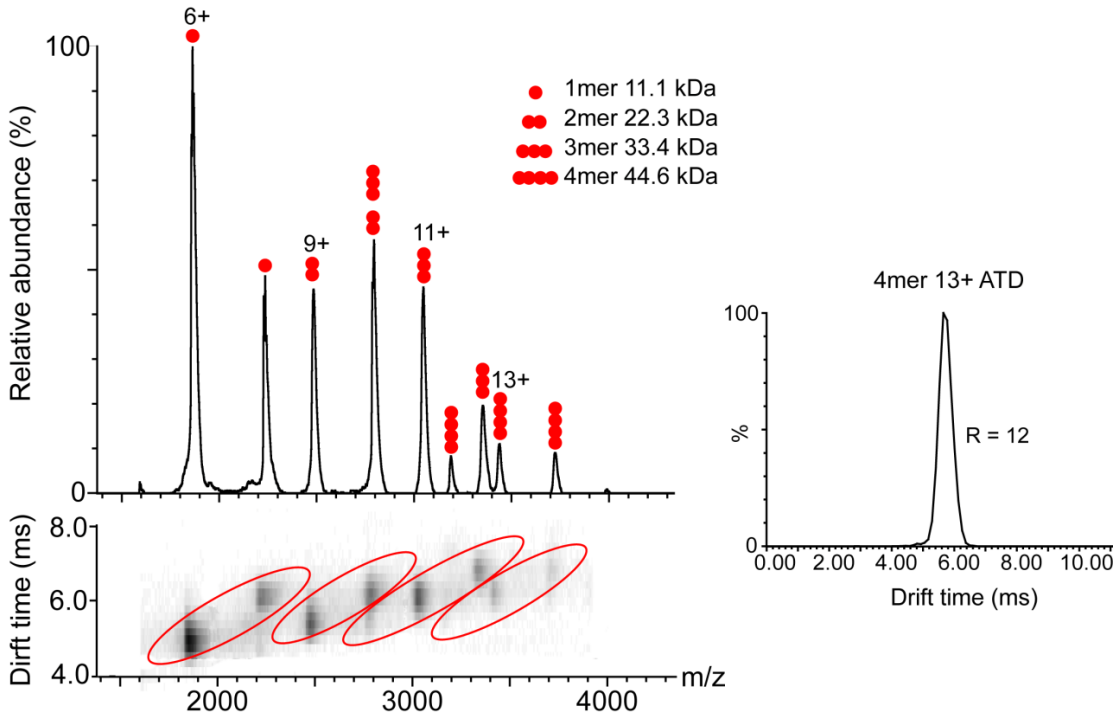


Figure 4.17. IM-MS data for early intermediates in ΔN6-β2m aggregation. Four distributions are observed in the spectrum, corresponding to the 1-mer, 2-mer, 3-mer and 4-mer of ΔN6-β2m. Representative ATD for the 4-mer (13+). $t_D/\Delta t_D$ ~ 12, and peak shape, indicate that the ATD corresponds to a single conformer, or group of closely-related conformations.

Due to overlapping *m/z* peaks for oligomers in MS, separation in mobility space was critical for the correct assignment of species in the MS spectrum. One of the advantages of working with ΔN6-β2m is that aggregation and fibrillation occur readily at less acidic or even neutral pH. This in contrast to full-length β2m which requires highly acidic pH (pH ~2) and/or fibrillar seeds in order to form fibrils. Therefore mobility arrival time distributions show compact folded species (Figure 4.17), rather than the potentially acid-unfolded forms reported in other studies (*144, 204, 206*).

The CCS for the monomer and dimer were measured and compared with the CCS calculated for their corresponding atomic coordinates. There is no reported crystal structure for ΔN6-β2m monomer, however PDB entry 1LDS represents the full-length β2m monomer. The first six N-terminal residues were removed prior to CCS calculation in order to more closely reflect ΔN6-β2m. The calculated and experimental CCS for the monomer were found to be in good agreement (CCS_{exp} 1200 ± 36 $Å^2$; CCS_{calc} 1290 $Å^2$). Crystal structure 2X89 shows ΔN6-β2m dimer in complex with a nanobody. This shows that ΔN6-β2m forms a domain-swapped dimer, with the short NHVTLSQ peptide (residues 83– 89) serving as the hinge region (*101*). The CCS calculated for this dimer was compared with the dimer CCS from IM-MS. In this case, the experimental value was slightly lower (~ 10 %) than anticipated from the crystal structure (CCS_{exp} 1900 ± 57 $Å^2$; CCS_{calc} 2138 $Å^2$). This could be the result of subtle rearrangements in the gas phase. Another explanation is that the dimer in solution in these experiments is not the domain-swapped form observed during X-ray crystallography.

The coarse-grained modelling approach described in this chapter was applied to ΔN6-β2m. The radius of the sphere representing the monomer was derived from the

experimental CCS. Models were then built for the dimer and trimer (CCS_{exp} 2530 ± 76 $Å^2$) by sequentially adding subunits, whilst simultaneously varying the spherical overlap (dimer and trimer), and the angle between subunits (trimer only). Models were scored, based on their violations of the CCS and volume restraints. The model with the lowest total score at each stage was taken to the next step, to form the (n+1) oligomer. In order to sample conformational space for the tetramer (CCS_{exp} 3057 ± 92 $Å^2$), a Monte Carlo approach was used keeping the relative position of the three other subunits fixed. After filtering 50 000 models for overlap violation, a total of 18 651 models for the tetramer remained. Models were scored, based on violations of the CCS and volume restraints. Clustering of the 1% lowest scoring models revealed three clusters (Figure 4.18). The largest cluster (84.4 %) is represented by a compact topological arrangement of subunits. Therefore based on the experimental data, a compact structure is proposed for the ΔN6-β2m tetramer.

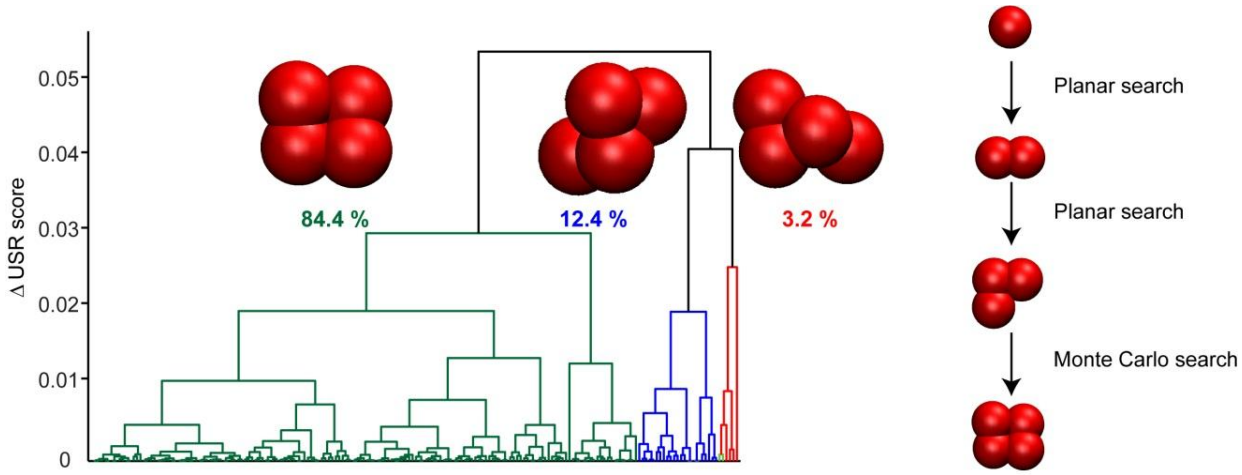


Figure 4.18. Modelling the ΔN6-β2m tetramer from IM-MS data. The tetramer was assembled from the monomer, by the stepwise addition of subunits. The model with lowest S_T at each stage was chosen as the starting point for the subsequent addition of further monomer units. The 1% lowest scoring models for the tetramer were clustered on their pair-wise USR scores. Three clusters were identified for the models which were scored based on CCS and volume restraints. Clusters are represented by the model with the lowest S_T within the cluster.

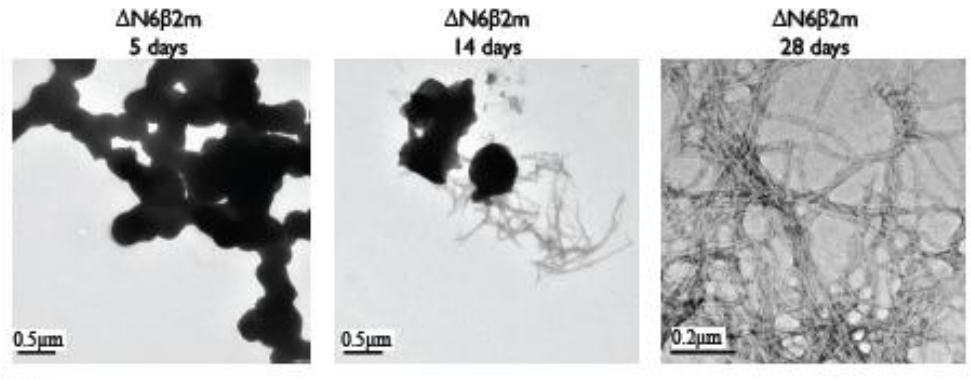


Figure 4.19. Fibrillogenesis of ΔN6-β2m followed by electron microscopy. ΔN6-β2m was incubated for 28 days at 37 °C and pH 5.0 (50 mM acetate buffer). Samples were taken after 5, 14, and 28 days and subjected to negative stain EM imaging. After 14 days of incubation, the first amyloid fibrils started to grow out of these aggregates. At 28 days, all aggregates were converted into amyloid fibrils (*101*).

Compact conformations for the trimer and tetramer intermediates of β2m have been proposed previously for a non-nucleated, monomer-addition mechanism resulting in fibrils with worm-like morphology (*207*). Other studies, operating at lower pH, identified elongated oligomers using IM-MS (*144*). It has been established that different β2m aggregation pathways exist which depend on the solution conditions. Some of these pathways lead to fibril formation of varying morphologies (*208*) and some may be off-pathway for fibril formation (*209*). Fibril formation[8] however was observed for ΔN6-β2m, under similar conditions to those used in the IM-MS experiments (Figure 4.19).

Compact tetramers have also been proposed by an X-ray crystallography study of β2m mutants (*210*) and by covalent-labelling of wild-type β2m oligomers, in the presence of copper (*211*). Moreover, the model from IM-MS is consistent with a recently published EM study (*212*), in which cryo-EM showed that β2m fibrils were assembled from

[8] Electron microscopy of ΔN6-β2m fibrils was carried out by K. Domanska (Vrije University), with whom I collaborated for the β2m study.

compact globular tetrameric units. This EM fibril map (EMD 1613) was used to fit the atomic structures for two domain-swapped ΔN6-β2m dimers (PDB 2X89). According to fittings (Figure 4.20) into the fibril density map, ΔN6-β2m dimers could form dimer-of-dimers by interacting through their hinge regions. The ΔN6-β2m tetramer coarse-grained model matched closely the EM density for the globular repeating unit (Figure 4.20D).

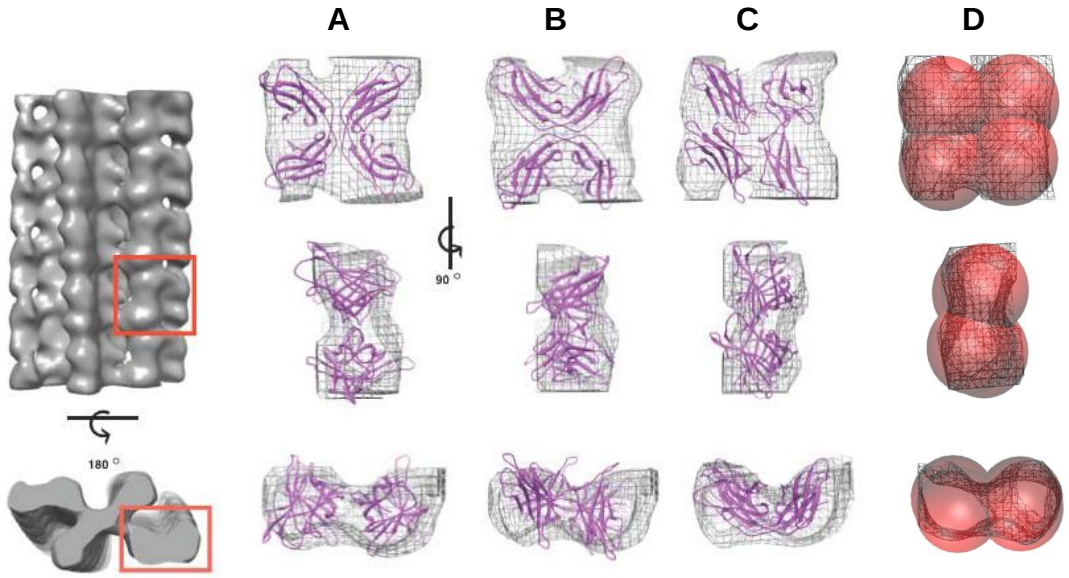


Figure 4.20. Fitting of a domain-swapped ΔN6-β2m dimer (PDB 2X89) into the cryo-EM density map of β2m fibrils (EMD 1613) (*212*) using Chimera (UCSF) (*185*). Segments (red box) of the EM map are shown as surface-and-mesh representations in three different orientations. ΔN6-β2m dimers (purple) could form dimer-of-dimers by interacting through their hinge regions. **A** Hinges run parallel to the fibril axis. **B** Hinges run perpendicular to the fibril axis. **C** Hinges are located at the surface of the globular regions of density and run perpendicular to the fibril axis. **D** Coarse-grained model (best-scored model in largest cluster) for ΔN6-β2m tetramer overlaid on a segment of the fibril EM map.

Overall the proposed compact model for ΔN6-β2m tetramer from IM-MS is consistent with data from a variety of techniques, including EM, X-ray crystallography and covalent labelling MS. The integration of such data with computational studies (for example docking, molecular dynamics) could reveal higher-level structural details for oligomers and ultimately offer insights into the aggregation process.

4.4 DISCUSSION

Integrative computational methods have been used previously to develop structural models for the 26S proteasome (*170*) and the nuclear pore complex (*169*). In these examples, restraints derived from diverse experimental data were applied to define a scoring function. IM-MS, an emerging technique, can play a role in such integrative methods. Whilst IM-MS provides low-resolution information, its predictive power can be enhanced when combined with other structural biology techniques, for example EM, which reveals the symmetry and overall shape of assemblies (*213*). Incorporating a symmetry restraint into the scoring function was essential to determine the structure for tryptophan synthase, and greatly reduced structural variability in the best scoring models for NBDO. Clear improvements were observed, in the correlation between total score and similarity to the native structure, when symmetry was included as additional restraint (Figure 4.21).

The impact of combining the symmetry restraint with CCS can also be shown quantitatively through analysis of ROC curves, whereby the area under the curve increased when all restraints are imposed (Figure 4.12C, 4.15C). For example, for tryptophan synthase, the area under the curve increased from 0.82 for the volume/CCS restraint to 0.97 with the addition of a symmetry restraint. Similarly, for NBDO 6-mer, within the 1% lowest scoring models, the positive predictive value of the method increases from 58 % for the CCS restraint, to 76 % when the volume restraint is added, to 100 % when all restraints are imposed (Figure 4.15D). Such results demonstrate the significant increase in predictive power by integrating different experimental information into a computational approach.

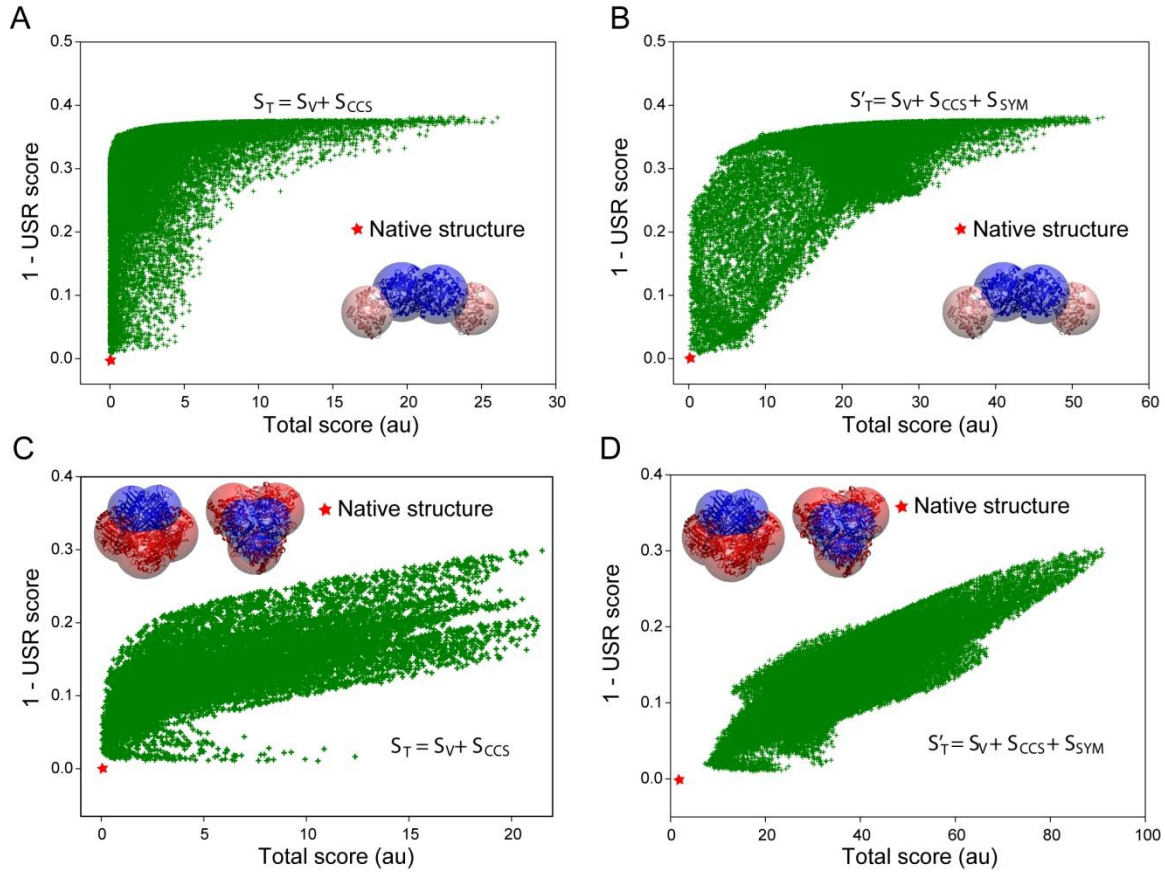


Figure 4.21. The total score (S_T) distributions for the tryptophan synthase 4-mer (**A, B**) and NBDO 6-mer (**C, D**) and are plotted against 1-USR score, relative to the native structure. The native structure (1-USR = 0) is indicated by a red star. For NBDO and its subcomplexes, lower scored models are associated with high structural similarity to the native structure. In **A**, the CCS and volume restraints are not sufficient for structure prediction. In **B** and **D**, the predictive power of the approach is further improved by the addition of a symmetry restraint.

It is necessary for scoring to take into account uncertainties in the different techniques (*123*). The quality of experimental data, for example, is crucial. While improved experimental accuracy and resolution can improve the predictive power of IM, this may be limited by the accuracy of theoretical CCS values, and how these relate to the experimental values (*59*). Even with good quality data, it should also be noted that structurally different models can have similar scores, for example in NBDO S_{CCS} is not affected by rotation of the β_3-trimer relative to α_3. Although this method does not

necessarily predict a unique best scoring model, it has proved effective in filtering poor models, as shown by the high specificity and negative predictive value of the restraints.

In order to include IM-MS datasets in hybrid approaches, other sources of uncertainty in the data must be considered. Most important are the two underlying assumptions: the first is that the solution conditions required to perturb subunit interactions do not influence the overall subunit shape. This has been demonstrated previously for homomeric protein complexes (*145*) and for solvent-disrupted subunits (*214*). In this chapter, in the great majority of cases CCSs of heteromeric subcomplexes were shown to be unaffected by the solution conditions required to perturb inter-subunit interactions. This is provided the solution conditions are optimised, by careful titration of the chaotrope until subcomplexes can be detected, before unfolding and complete denaturation occur.

The second assumption is that there have been no extensive conformational changes occurring to the protein assembly when passing from solution into the gas phase. Whilst the precise packing of the side chains and secondary structural elements of the gas-phase structure are unlikely to correspond to the folded structure in solution (*71*), there is a growing body of evidence to suggest that the overall native-like topology can be preserved in the gas phase (*75, 215-216*). Deviations from the crystal structure, however, have been noted in certain cases (*86*) and may depend on the experimental conditions or on particular topological arrangements. In the data set presented in this chapter, the CCS for one cavity-containing complex was significantly lower than expected. One possible explanation is that this complex undergoes gas-phase collapse. This suggests that future research to attempt to stabilise such complexes in the gas-phase would be of high

importance. It is interesting to consider the alternative possibility however, that the crystal structure may not be an accurate representation of the protein in solution (*217-218*). Additional experiments, such as EM, SAXS, and HDX could be used to probe this further. The extent to which the native-like structures of protein complexes are maintained in the gas phase remains a controversial topic. A deeper understanding of the gas-phase structures of proteins and their complexes will be key for this modelling approach and its integration in future hybrid structure determination strategies.

4.5 SUMMARY

In this chapter, a modelling approach is described for generating models consistent with experimental data from IM-MS of intact assemblies and their subcomplexes. Disruption of protein complexes in solution did not affect the overall conformation of subcomplexes for several heteromers, making CCS of subcomplexes suitable for use as a restraint. Combined with spatial restraints, derived from crystal structures of protein complexes, the feasibility of using CCS for predicting native structures has been critically assessed. The strength of the individual and combined restraints was tested, using a variety of statistical tools, for a simple model system. The methodology was then applied to generate an ensemble of low-resolution models for heteromeric protein complexes tryptophan synthase and NBDO using experimental data. Candidate models were evaluated by applying a scoring function, which uses restraints from IM experiments, volume restraints, and symmetry, where applicable. Cluster analysis revealed structural variability within the ensemble of best scoring models, which was reduced when including the symmetry restraint. The best models for the intact complexes, and their subcomplexes, showed good structural similarity to the native structure, with centroid RMSD 1.9 – 8.9 Å. Finally a topological model was predicted for an intermediate in the

aggregation of ΔN6-β2m, a tetrameric complex whose structure has yet to be resolved in detail. A compact topology was proposed, consistent with data from other studies which use diverse structural approaches. Overall in this chapter, the power of CCS restraints for building models of protein complexes has been assessed, showing that IM-MS can play a role in structure characterisation, particularly when integrated with data/information from other sources.

The work described in this chapter has been adapted with permission from Hall et al. (2012), Structure, 20, 1596-1609, © 2012 Elsevier Ltd. doi:10.1016/j.str.2012.07.001.

5 EXPLORING THE EFFECTS OF CHARGE MANIPULATION ON CONFORMATION AND DISSOCIATION OF PROTEIN COMPLEXES

5.1 INTRODUCTION

ESI results in multiply charged macromolecular ions (*8, 219*). The observed charge state distribution is dependent upon many factors, including solution conditions (*23-24*), instrumental factors (*25-26*), interaction with other molecules in the gas phase and molecular conformation (*27*). The charge carried by a protein or protein complex can be lowered by "charge reduction", or increased through "supercharging". Charge manipulation of protein complexes has several useful applications. Charge reduction can be used to resolve overlapping mass spectra, as the peak spacing between neighbouring charge states increases at lower charge states. Furthermore, charge reduction has been shown to confer additional stability to gas phase ions by reducing Coulombic effects (*220-222*). Charge reduction can be achieved in a number of ways. These include the use of ^{210}Po α-particle source which ionises air (*220*), ion-molecule reactions with proton scavengers including solvents such as acetonitrile (*223-225*), and ion-ion reactions (*226*). The most readily accessible method employed to lower the charge state is the use of solution additives, such as TEA and imidazole, which have higher gas-phase basicities than typical MS buffers (*106, 222, 227*). In the final stages of desolvation, these species effectively compete with ionised sites on the protein for charge (*16*). It is not clear however how these reagents affect the folded structure of a protein or complex.

A study in which lysozyme was progressively charge reduced from 10+ to 3+ by solution additives showed, by circular dichroism spectroscopy, there to be no loss in secondary structure for any of the charge states (*106*). This is in direct contrast to experiments performed by Bornschein *et al.* (*228*), who propose that charge reduction of protein complexes using solution additives can result in unfolding, leading to an increase in CCS. No CCS increase was observed for species charge-reduced in the gas phase. Higher cone voltages however were used for the solution additive experiments.

Supercharging experiments are used to produce higher charge states than normally accessible by globular proteins and their complexes. Supercharging extends the mass range of MS without the need for hardware modifications, by shifting peaks of interest to lower *m/z* regions; this allows the analysis of very large protein assemblies (*229-230*). Supercharging can be achieved by the addition of low volatility reagents with low solution-phase basicity, typically *m*-NBA or sulfolane (*231-232*). These accumulate in the electrospray droplet, and can prolong the droplet life time. The mechanism for supercharging is still not fully understood. Originally thought to be due to changes in surface tension (*233*), the commonly accepted theory is that the corresponding increase in charge is due to denaturation in the final stages of desolvation in the electrospray droplet (*234-236*). This could be chemical denaturation (due to the increase in reagent concentration during electrospray) or thermal denaturation (the droplet can access higher temperatures). An alternative mechanism however has been proposed in which the supercharging reagent directly interacts with ionisable residues (*237*). As a result of its high dipole moment, the reagent can delocalise charge, therefore lowering the electrostatic barrier for proximal charging along the amino acid chain.

The evidence for folded structure under supercharging conditions is highly controversial. In a recent study by Loo and co-workers, an increase in average charge state is reported, from 10+ to 15+, for the 29 kDa carbonic anhydrase-zinc complex. The fact that supercharging does not result in loss of protein-ligand interactions (*229*) is evidence that the ligand binding cavity is retained, implying that the structure is native-like. A series of experiments at odds with these results have shown that for the protein *holo* myoglobin, addition of sulfolane leads to unfolding of the protein as assessed by partial loss of heme binding and loss of secondary structure using circular dichroism spectroscopy (*236*). Moreover supercharging was affected by chemically modifying cytochrome *c* with up to seven cross-links or ubiquitin with up to two cross-links. While these cross-links did not affect the average charge states of these proteins in the absence of *m*-NBA, the extent of supercharging induced by *m*-NBA increased with decreasing numbers of cross-links (*235*). Similarly for protein complexes, the addition of *m*-NBA to concanavalin A perturbed the solution phase tetramer/dimer equilibrium and increased the tetramer CCS (*234*).

Several studies have suggested that the charge state of the complex ion can influence its dissociation pathway (*221, 238-239*). Gas-phase dissociation of protein assemblies is most often achieved by subjecting the ions to high energy collisions with a gas. The Coulombically-driven ejection of a highly charged, unfolded monomer results (*42*), leaving behind the “stripped complex”. Whilst CID of multimeric protein complexes generally results in an asymmetric partitioning with respect to mass and charge, the partitioning of surface area is believed to be more symmetric (*43*).

In some cases however, atypical dissociation has been observed, where subcomplexes and/or compact monomers are released instead of highly charged, unfolded monomers (*38, 221, 239-240*). For example, charge-reduced TTR dissociated *via* the loss of compact monomer, or peptide fragments (*221*). Higher charge states of stable protein 1 were found to eject subcomplexes rather than individual monomers. This was attributed to an unusually high charge density (*239*) and conformational restrictions for unfolding. On the other hand, anthrax toxin prechannel complex dissociated extensively into subcomplexes and monomers in the presence of *m*-NBA (*234*). This was rationalised by assuming the supercharging reagent effected structural changes, not evident through CCS measurements, but sufficient to influence the dissociation. Despite these intriguing results, a systematic study has not been carried out to date on how charge states, particularly those accessed by supercharging and charge reduction, influence the dissociation route of protein complexes.

In this chapter, the relationship between charge states and folding is investigated, focussing on the manipulation of charge in solution and its effect on protein complex native structure. In order to achieve this, charge reduction and supercharging experiments were carried out on a set of protein complexes, and their CCS monitored as a function of charge state. Whilst charge reduction had very little impact on the overall CCS of the low charge states, a significant increase in CCS was detected for the highest charge states of the supercharged complexes, with the exception of pentameric SAP. Furthermore, the effect of charge state on the dissociation of two homomeric complexes, SAP and TTR, and heteromeric tryptophan synthase was systematically investigated. To achieve this, the CCS of the CID products were monitored as a function of precursor ion charge state, including both charge-reduced and supercharged complex ions. Interestingly,

supercharged SAP and TTR were both able to access atypical dissociation pathways, with the loss of compact monomers and dimers. Tryptophan synthase simultaneously accessed typical and atypical dissociation routes, with both compact and extended forms of the α-monomer ejected in parallel. Extending the study to a further eight complexes, and comparing typical and atypical dissociation routes identified in this study and reported in the literature, a link was identified in which atypical dissociation routes are favoured by a low number of interfacial salt bridges and high overall charge density.

5.2 RESULTS

5.2.1 Relationship between CCS, charge and surface area

With a large IM-MS dataset of proteins and protein complexes at our disposal from Chapter 3, relationships were established between CCS, charge state and surface area. The average charge state observed in MS is thought to be closely related to the SASA, which can be calculated from the crystal structure (*28*). The exact relationship may be affected by the solution conditions, nanospray settings, and instrument. In order to establish a relationship between the SASA and the average charge state, Z_{av}, under these specific experimental conditions, ln(SASA) was plotted as a function of ln(Z_{av}). Excellent linear correlation (R^2 = 0.99) was found, over nearly two orders of magnitude of mass (Figure 5.1A). The average charge state therefore can be used to predict the surface area of the macromolecular ion.

The relationship between protein charge states and their CCS has been well characterised for individual proteins, such as ubiquitin, myoglobin and cytochrome *c (149-150)*. In these examples, the charge state of the protein ion was found to correlate with an increase in CCS (Figure 5.1B). An increase in CCS with charge state is consistent with the unfolding of the more highly charged ions in the gas phase, although it is equally likely to result from unfolding in the solution conditions used to generate these high charge states, which include significant quantities of acid and organic solvent. The CCS of the native-like ions measured in Chapter 3 were plotted as a function of charge state. Interestingly, in contrast to the denatured proteins, CCS of native-like ions are only subtly affected by charge state (Figure 5.1C). This suggests that these large ions are native-like in solution and Coulombically-stable in the gas phase, at least for the time required to pass through the drift tube.

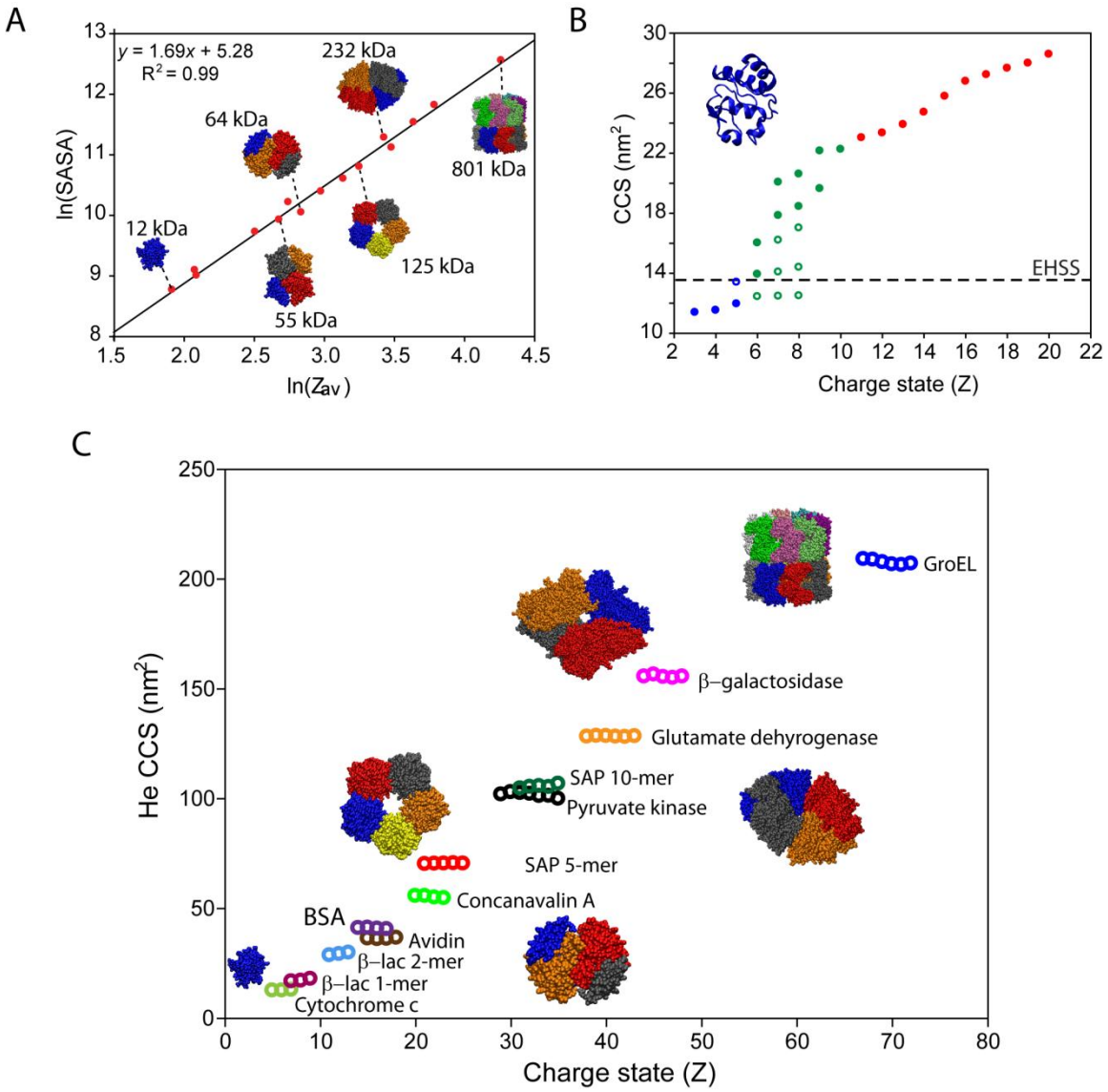


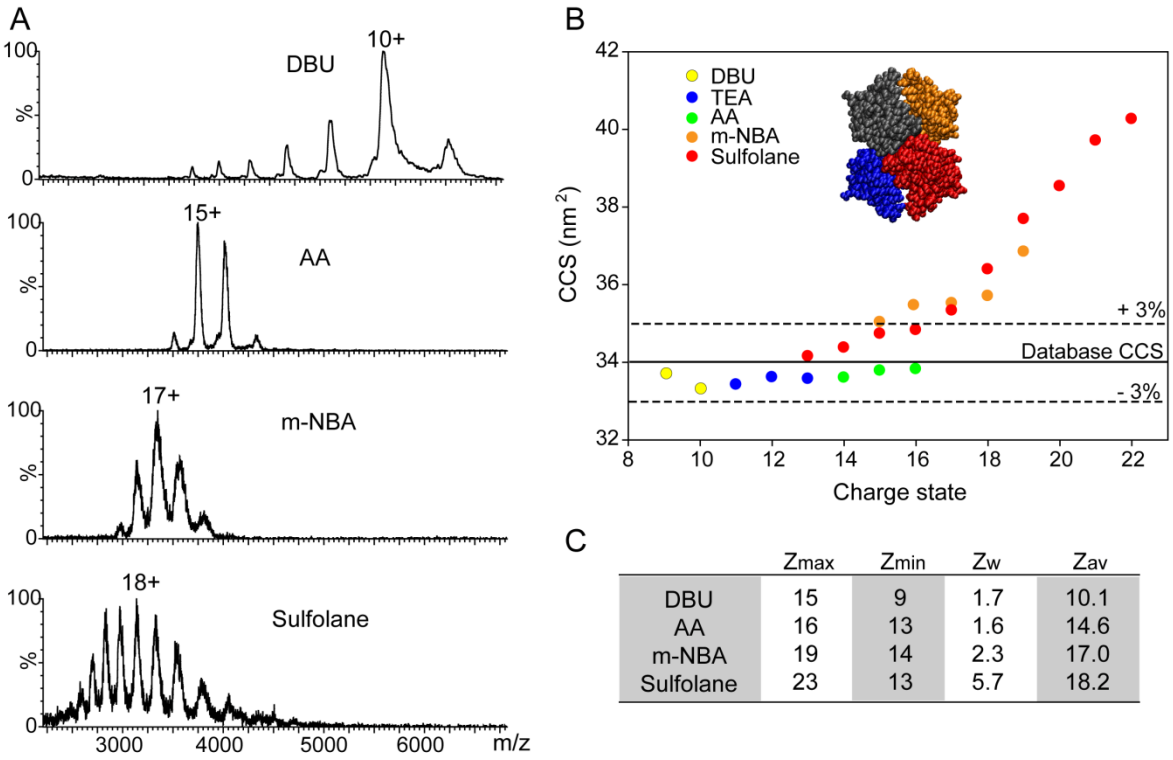
Figure 5.1. Relationship between CCS, charge state and surface area. **A** The average charge state from noncovalent MS (Z_{av}) can be used to predict protein complex solvent accessible surface area (SASA). Data shows buffered proteins and protein complexes from 12 to 801 kDa in 200 mM AA. **B** CCS of cytochrome *c* (75/25 water/acetonitrile) is influenced by charge state, with higher charge states occupying more extended conformations. A plot of CCS *versus* charge state, adapted from (*150*), is shown. Charge states 11-20+ (red) were generated by acidifying the solution with 2.5 % acetic acid, whereas charge states <6+ (blue) were generated by charge stripping, through the introduction of a base into the desolvation region. The filled and open points show the CCS for the features dominating at high and low injection energies respectively. **C** The CCS of individual proteins and protein complexes in AA buffered solutions shows little variation with charge state.

5.2.2 Supercharging and charge reduction: Effects on conformation

Having observed the effect of charge state on CCS of native-like protein complexes in AA, this study was extended to consider charge-reduced and supercharged species. Charge reduction was carried out by the addition of TEA buffer or DBU base (10 - 20 mM) to the complexes in AA (200 mM). Supercharging was achieved by the addition of fixed amounts of *m*-NBA (1 %) or sulfolane (7 %). Seven protein complexes were selected, which vary in size, subunit number, and topology. These were TTR (55 kDa, 4-mer), avidin (64 kDa, 4-mer), ADH (143 kDa, 4-mer), SAP (125 kDa, 5-mer), pyruvate kinase (232 kDa, 4-mer), glutamate dehydrogenase (336 kDa, 6-mer) and β-galactosidase (465 kDa, 4-mer). CCS were measured for these complexes in AA, and compared to the CCS of equivalent charge-reduced or supercharged ions. All CCS measurements in this chapter were made on the RF-confining linear drift tube.

MS of the smallest complex, TTR, from an AA buffered solution revealed one charge state distribution, centred on 14.6+ and ranging from 13-16+ (Figure 5.2A). By the addition of DBU, charge states as low at 9+ were attained. CCS measurements of ions from 9+ to 16+ were all well within the experimental error of the database CCS (charge-averaged) value (Chapter 3, Table 3.2). The addition of *m*-NBA (1 %), increased the average charge state to 17.0+, with a range 14-19+. CCS measurements for these ions, however were greater than expected from the database CCS, and increase with increasing charge state (Figure 5.2B). Similar observations were made following the addition of sulfolane (7 %), whose presence increased the average charge state to 18.2+. Additionally a much wider distribution of charges was attained, ranging from 13+ to 23+. Interestingly, the lower charge states of TTR in the presence of sulfolane (13-16+) were

in agreement with the database CCS, albeit slightly higher than CCS measured for those same ions in AA.



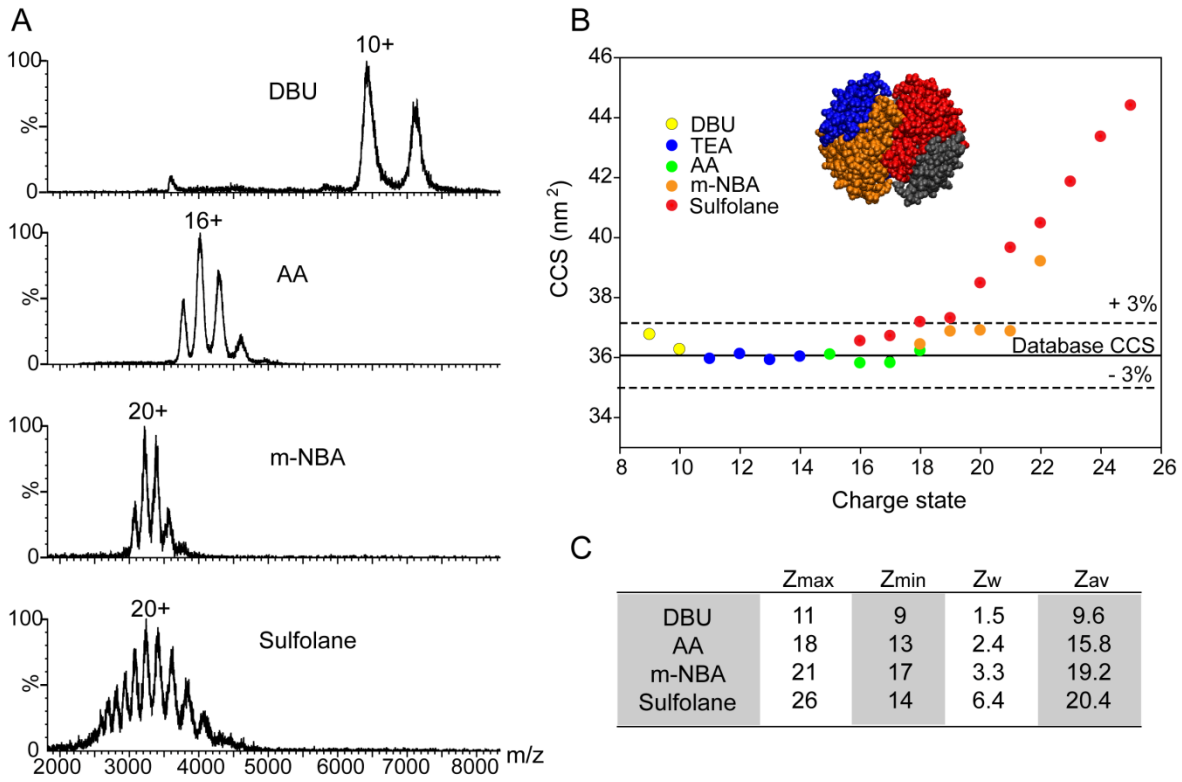
C

	Zmax	Zmin	Zw	Zav
DBU	15	9	1.7	10.1
AA	16	13	1.6	14.6
m-NBA	19	14	2.3	17.0
Sulfolane	23	13	5.7	18.2

Figure 5.2. Charge manipulation of TTR. **A** MS spectra of TTR (55 kDa, 4-mer) in 200 mM AA, or 200mM AA with either 20 mM DBU, 1 % *m*-NBA or 7 % sulfolane. **B** The CCS of TTR ions are plotted against charge state for various solution conditions. Average CCS and associated error, as determined in Chapter 3 using the RF-confining drift tube, are shown as solid and dashed lines. **C** The charge state distribution parameters Z_{av} (average charge state), Z_W (charge state distribution width at half height), Z_{max} and Z_{min} (maximum and minimum charge states observed) are shown in table form.

Similarly for avidin (64 kDa, 4-mer), MS of a solution electrosprayed from AA resulted in one charge state distribution (Figure 5.3), with an average charge state 15.8+ (range 13-18+). The addition of TEA lowered the charge state to a minimum of 11, and further charge reduction to 9+ was achieved by the addition of DBU base. The CCS for ions electrosprayed from AA or from charge-reducing solutions were indistinguishable. Supercharging with *m*-NBA or sulfolane resulted in an increase in the average charge state to ~20+ in both cases, however the width of the charge state distribution was wider

for sulfolane (Z_W 3.3 *versus* 6.4). Whilst CCS of the lower charge states of avidin from the supercharged solutions were close to the database CCS value, a dramatic increase in CCS with increasing charge state was seen for charge states >19-21+, corresponding to unfolding.

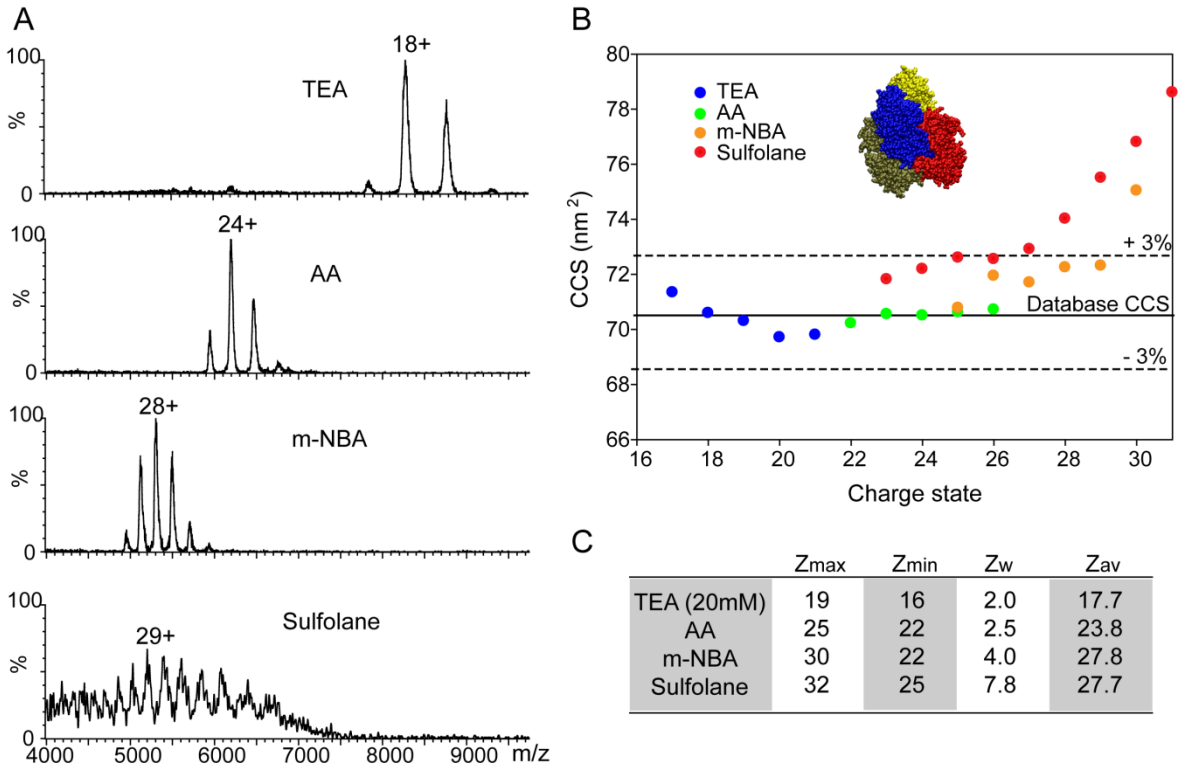


	Zmax	Zmin	Zw	Zav
DBU	11	9	1.5	9.6
AA	18	13	2.4	15.8
m-NBA	21	17	3.3	19.2
Sulfolane	26	14	6.4	20.4

Figure 5.3. Charge manipulation of avidin. **A** MS spectra of avidin (64 kDa, 4-mer) in 200 mM AA, or 200mM AA with either 20 mM DBU, 1 % *m*-NBA or 7 % sulfolane. **B** The CCS of avidin ions are plotted against charge state for various solution conditions. Average CCS and associated error, as determined in Chapter 3, are shown as solid and dashed lines. **C** The charge state distribution parameters are shown in table form.

Using the same approach, a wide range of charge states were generated for ADH (143 kDa, 4-mer), using different electrospray solution compositions (Figure 5.4). Charge states as low as 16+, and as high as 32+, were observed from the solutions containing TEA buffer and sulfolane respectively. Supercharged states accessed *via* the addition of *m*-NBA had CCS similar to those charge states generated from the AA or charge-

reducing buffered solutions. In addition, the CCS of these supercharged species were generally within the error of the database CCS for ADH. A stable electrospray was difficult to achieve in the presence of sulfolane, however a wide number of charge states were generated (Z_W = 7.8), with the average charge state increasing from 23.8+ in AA to 27.7+ in sulfolane-containing solution. As noted for TTR and avidin, there was a strong relationship between the charge state and CCS for the highest charge sates, indicative of unfolding.

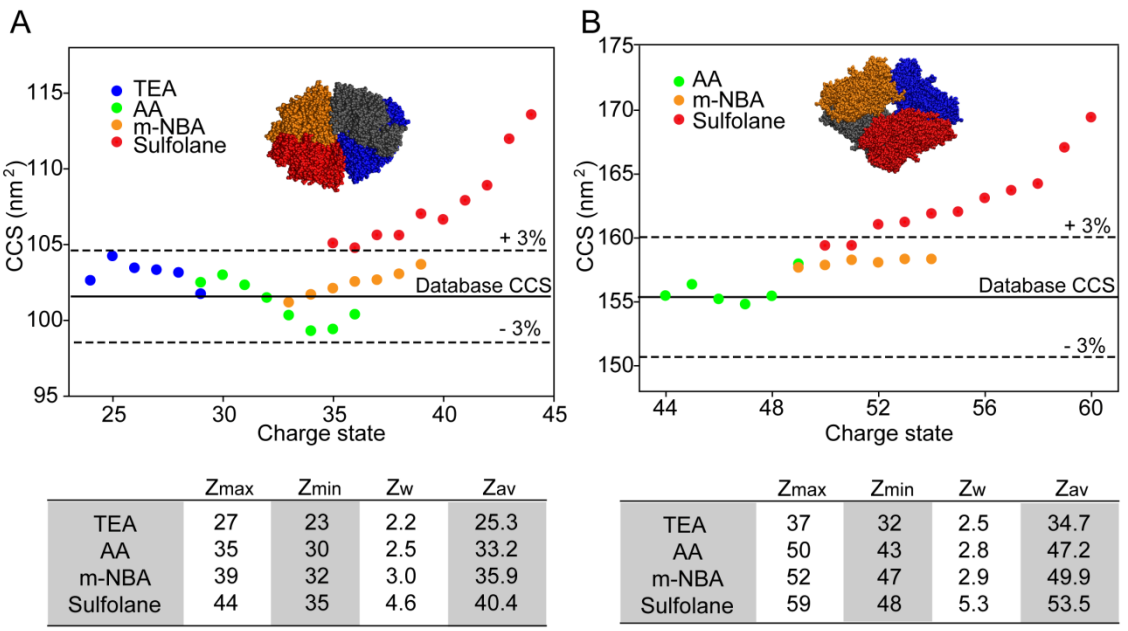


	Zmax	Zmin	Zw	Zav
TEA (20mM)	19	16	2.0	17.7
AA	25	22	2.5	23.8
m-NBA	30	22	4.0	27.8
Sulfolane	32	25	7.8	27.7

Figure 5.4. Charge manipulation of ADH. **A** MS spectra of ADH (143 kDa, 4-mer) in 200 mM AA, or 200mM AA with either 20 mM TEA, 1 % *m*-NBA or 7 % sulfolane. **B** The CCS of ADH ions are plotted against charge state for various solution conditions. Average CCS and associated error, as determined in Chapter 3, are shown as solid and dashed lines. **C** The charge state distribution parameters are shown in table form.

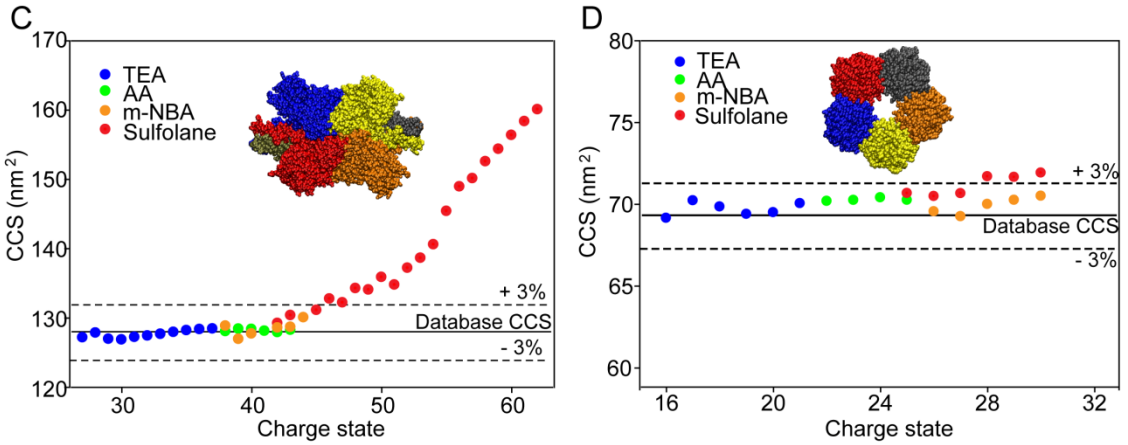
Similar experiments were carried out for pyruvate kinase (232 kDa, 4-mer), *β*-galactosidase (465 kDa, 4-mer), glutamate dehydrogenase (336 kDa, 6-mer) and SAP (125 kDa, 5-mer). Results are summarised in Figure 5.5 (spectra available in Appendix, Figures A1-A4). In these complexes, as with the three discussed in detail, charge reduction did not affect the CCS, indicating conformational stability. Interestingly, for pyruvate kinase (Figure 5.5A), a general decrease in CCS with increasing charge state was noted for ions generated from AA or AA/TEA buffered solutions. These ions (23+ to 35+) had CCS within the error of the database CCS, however the observed trend indicates subtle compaction, or rearrangement in the gas phase, becoming more apparent at higher charge states.

Overall, in the majority of cases supercharging with 1 % *m*-NBA resulted in charge state distributions of protein complexes with comparable widths to those observed in AA. Additionally, the CCS observed for these species were generally within the error of the corresponding database value for CCS (with the exception of TTR, and the highest observed charge state for avidin and ADH). In many cases however, the CCS of ions from *m*-NBA solutions were on average greater than that observed for similar ions in AA (e.g. Figure 5.5B). This is likely due to a greater number of adducts (*241*), although could also be the result of minor structural changes. Supercharging from sulfolane-containing solutions, on the other hand, generally resulted in wide charge state distributions. In extreme cases, as many as 26 charge states were simultaneously observed (Figure 5.5C; Appendix Figure A3). For six out of the seven complexes, the CCS of species supercharged using sulfolane showed a strong dependence on charge state, with CCS increasing with charge state. CCS of all but the lowest charge states are significantly greater than the database CCS for the corresponding complex.



	Zmax	Zmin	Zw	Zav
TEA	27	23	2.2	25.3
AA	35	30	2.5	33.2
m-NBA	39	32	3.0	35.9
Sulfolane	44	35	4.6	40.4

	Zmax	Zmin	Zw	Zav
TEA	37	32	2.5	34.7
AA	50	43	2.8	47.2
m-NBA	52	47	2.9	49.9
Sulfolane	59	48	5.3	53.5



	Zmax	Zmin	Zw	Zav
TEA	20	25	1.9	27.3
AA	40	36	3.0	38.4
m-NBA	43	39	2.7	40.6
Sulfolane	69	44	15.5	55.0

	Zmax	Zmin	Zw	Zav
TEA	20	16	2.2	18.1
AA	27	22	2.6	24.6
m-NBA	30	25	2.5	28.2
Sulfolane	30	24	3.1	27.1

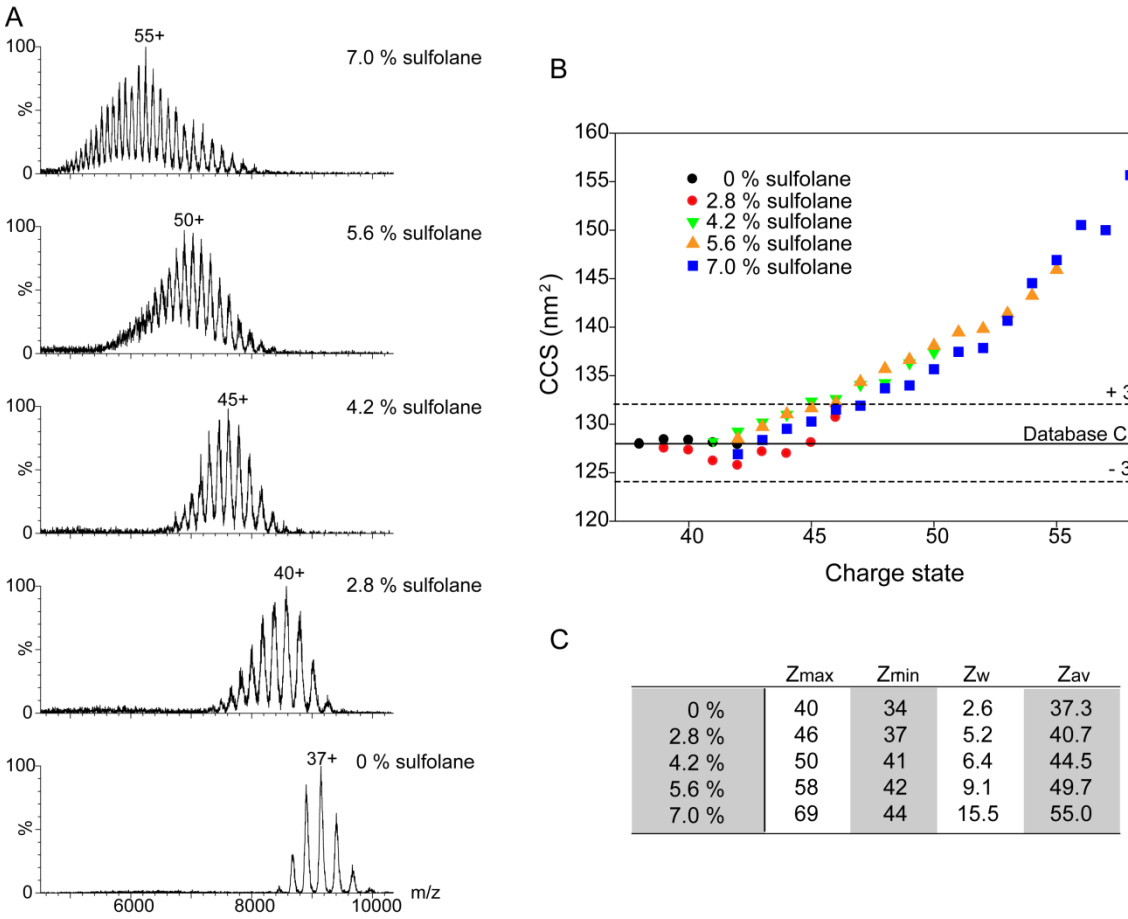
Figure 5.5. Charge manipulation of protein complexes. The CCS of pyruvate kinase (**A**), *β*-galactosidase (**B**), glutamate dehydrogenase (**C**) and SAP 5-mer (**D**) ions are plotted against charge state for various solution conditions (200 mM AA, or 200mM AA with either 20 mM TEA, 1 % *m*-NBA or 7 % sulfolane). Average CCS and associated error, as determined in Chapter 3, are shown as solid and dashed lines. The charge state distribution parameters are shown below each plot in table form. MS spectra are shown in Appendix (Figures A1-A4).

Interestingly, the SAP pentamer did not follow the same pattern (Figure 5.5D). Similar charge state distributions were observed for SAP electrosprayed from solutions containing *m*-NBA or sulfolane. In line with this, all supercharged species (26+ to 30+) had similar CCS to the complex following electrospray from AA buffer (22+ to 25+) or even in its charge-reduced state following addition of TEA (16+ to 21+). The fact that for all 15 charge states, the CCS measurements vary by not more than a few percent provides compelling evidence that the charge on the surface of the complex has little or no effect on its conformation. The highest charge states from the sulfolane solution had CCS slightly greater than expected from the database CCS. The lack of any trend between CCS and charge state, suggests that this could be the result of increased adduction, arising from low volatility supercharging reagents and the gentle MS conditions employed (*241*).

5.2.3 Supercharging and unfolding in solution

The increased CCS for higher charge states accessed *via* supercharging with sulfolane suggests extensive unfolding. This may occur in solution, during the final stages of electrospray, or after transfer to the vacuum. During electrospray, accumulation of the low volatility supercharging reagent occurs. The resulting high concentration of reagent could lead to chemical and/or thermal denaturation (*234-236*). On the other hand, higher charge states can lead to increasing Coulombic instability; gas-phase unfolding therefore occurs to mitigate the repulsive forces (*69, 76*). It is also possible that a combination of solution and gas-phase unfolding occurs.

In the experiments described thus far, fixed concentrations of reagents were used. However it is interesting to examine the effects of increasing the supercharging reagent concentration. Glutamate dehydrogenase was titrated with increasing sulfolane concentration (Figure 5.6). The extent of supercharging was related to the reagent concentration, as anticipated. The CCS was monitored for each charge state, under conditions of increasing sulfolane.



	Zmax	Zmin	Zw	Zav
0 %	40	34	2.6	37.3
2.8 %	46	37	5.2	40.7
4.2 %	50	41	6.4	44.5
5.6 %	58	42	9.1	49.7
7.0 %	69	44	15.5	55.0

Figure 5.6. Titration with increasing sulfolane. Glutamate dehydrogenase was exposed to increasing concentrations (0 – 7 %) of sulfolane (**A**). CCS was plotted as a function of charge state for each solution condition. The database CCS and its associated error are shown as solid and dashed lines respectively (**B**). As the initial sulfolane concentration increased, the average and maximum charge states increased, as did the charge state distribution widths (**C**).

Interestingly, the CCS increased with charge state, independent of the initial reagent concentration. There are two possible explanations:

i) The conformation (CCS) is determined by the charge state. Higher charge states result in unfolding in the gas phase due to greater Coulombic instability.

ii) The charge state is determined by the conformation. Proteins which unfold in solution, or during electrospray, are able to access a greater number of charges, directly related to the surface area of the protein at the final stage of desolvation.

Further insights were obtained by monitoring charge state distribution widths for the seven protein complexes during supercharging with *m*-NBA or sulfolane. Perturbation of structure was measured by the deviation of the most intense charge state (Z_{Base}) CCS from the database CCS (charge-averaged) for the corresponding complex. It is notable that complexes whose structures were perturbed to a greater extent had significantly wider charge state distributions (Figure 5.7). This is analogous to the wide charge state distributions commonly observed for denatured individual proteins.

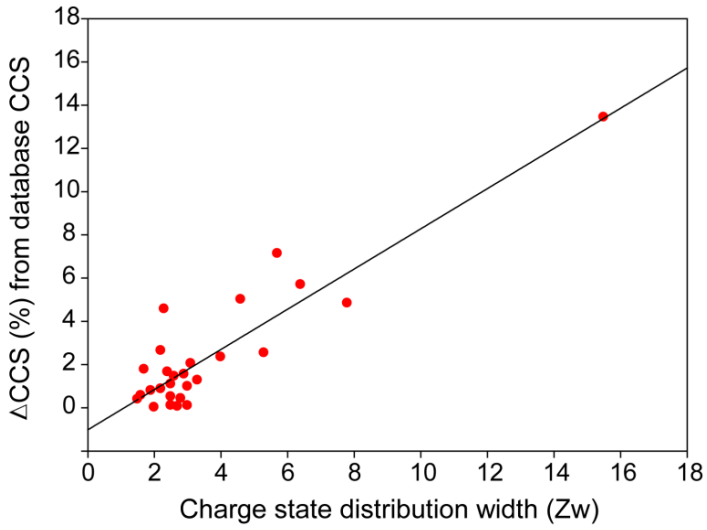


Figure 5.7. Charge state distribution width as an indicator of solution unfolding. Deviation of CCS (Z_{Base}) from the database CCS is plotted as function of the charge state distribution width, Z_W.

Unfolding in the gas phase however, is not typically accompanied by any change in the charge state distribution. Therefore, a shift to a wide distribution of charge states is indicative of unfolding in solution or during the final stages of desolvation during ESI. A wide charge state distribution therefore, implies greater conformational heterogeneity than a narrow charge state distribution.

Supercharging of glutamate dehydrogenase with ~3 % sulfolane increases the maximum charge state in AA from 40+ to 46+ (Figure 5.6). Interestingly, the CCS for these supercharged species are similar to those for the lower charge states, and to the database CCS. This observation implies that the supercharging effect at lower reagent concentrations is not due to the presence of extended conformations in solution, resulting from denaturation. This is supported by observations that supercharged protein complexes from 1% *m*-NBA solutions generally result in ions with CCS similar to lower charge states, from AA or charge-reducing solutions. Therefore unfolding in solution is unlikely to be the sole cause of the charge increase observed during supercharging experiments.

5.2.4 Supercharged species can access atypical dissociation pathways

In order to assess the effects of charge state on dissociation, two protein complexes were examined in detail. Firstly, SAP 5-mer, whose overall conformation was shown to be unaffected by either charge reduction or supercharging. This was illustrated by the small deviation in CCS between charge states 16+ and 30+. Secondly TTR, a protein complex whose dissociation has been extensively studied by both CID and SID (*40, 221, 242-243*). In one such study, the charge-reduction of TTR was shown to promote dissociation

via the loss of compact monomers and peptide fragments (*221*). This was proposed to be due to the lower intramolecular Coulombic repulsion experienced by lower charge states, which effectively raises the energy barrier for unfolding. The effects on TTR dissociation of increasing the charge state by supercharging however have not been assessed.

First the dissociation of homomeric SAP was studied, making use of charge reduction with TEA, and supercharging using *m*-NBA. A wide range of charge states were generated for SAP and different precursor ions were selected in the quadrupole for tandem MS (Figure 5.8A). The lowest charge states were highly stable and not susceptible to dissociation. Higher charge states however readily dissociated *via* the loss of monomers. For the lower charge states the charge was asymmetrically distributed, with respect to mass, between the dissociated monomer and the remaining tetramer. The dissociated monomer carried approximately half the total charge of the intact complex. Interestingly, as the charge state increases further, through the addition of supercharging reagent, an atypical dissociation pathway was accessed in which the loss of low charged monomers, and dimers was observed.

To investigate these observations further, the dissociated Z_{av} for the monomer was plotted against precursor ion charge state (Figure 5.8B). From 18-24+, as the precursor ion charge state increases, so does the Z_{av} for the monomer. For Z > 24+ two distinct populations of monomer were observed, the emerging population having significantly lower charge states. As the precursor ion charge state increased from 25+ to 30+, the low-charge monomer population increased in intensity relative to that of the high-charged. Increasing proportions of dimer were also observed with an increase in

precursor ion charge state (Z > 24+). These observations suggest that for Z=25+ and above, a second dissociation pathway becomes accessible and with increasing precursor ion charge state, this alternate pathway becomes more favourable.

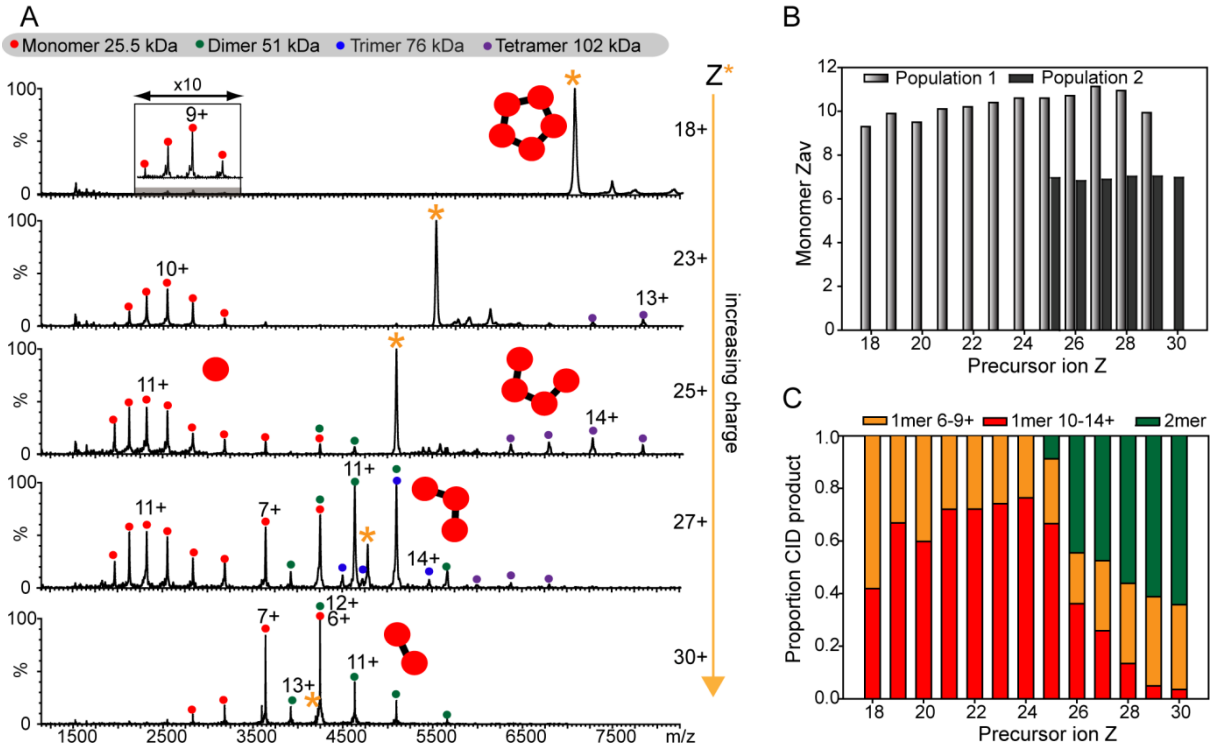


Figure 5.8. CID of SAP 5-mer. **A** MS/MS of SAP 5-mer with increasing precursor ion charge state. Initially higher charge states readily dissociate *via* the loss of highly charged monomer. As the charge state increases further, through the addition of *m*-NBA, an alternate dissociation pathway is accessed. MS/MS spectra were acquired at similar laboratory frame energies (collision energy x Z). **B** The dissociated monomer Z_{av} is plotted against precursor ion charge state. Above Z=25+, two distinct populations of monomer are observed. **C** The relative intensities of monomer and dimer CID products in the MS/MS spectra are plotted against precursor ion charge state. Precursor charge states were generated using the following additives to 200 mM AA: 18-21+ (20 mM TEA), 22-25+ (no additive), 26-30+ (1 % *m*-NBA).

Using the same approach as for SAP, the dissociation of TTR was studied for a wide range of charge states (Figure 5.9). The charge state of the dissociated monomer was found to increase with the charge state of the precursor ion (Z = 10+ to 18+). Generally, the ejected monomer removed approximately half the number of charges which were present for the intact ion. For example for Z=10+, the lowest charge state observed following charge reduction with DBU, the dissociated monomer had Z_{av} ~ 5+.

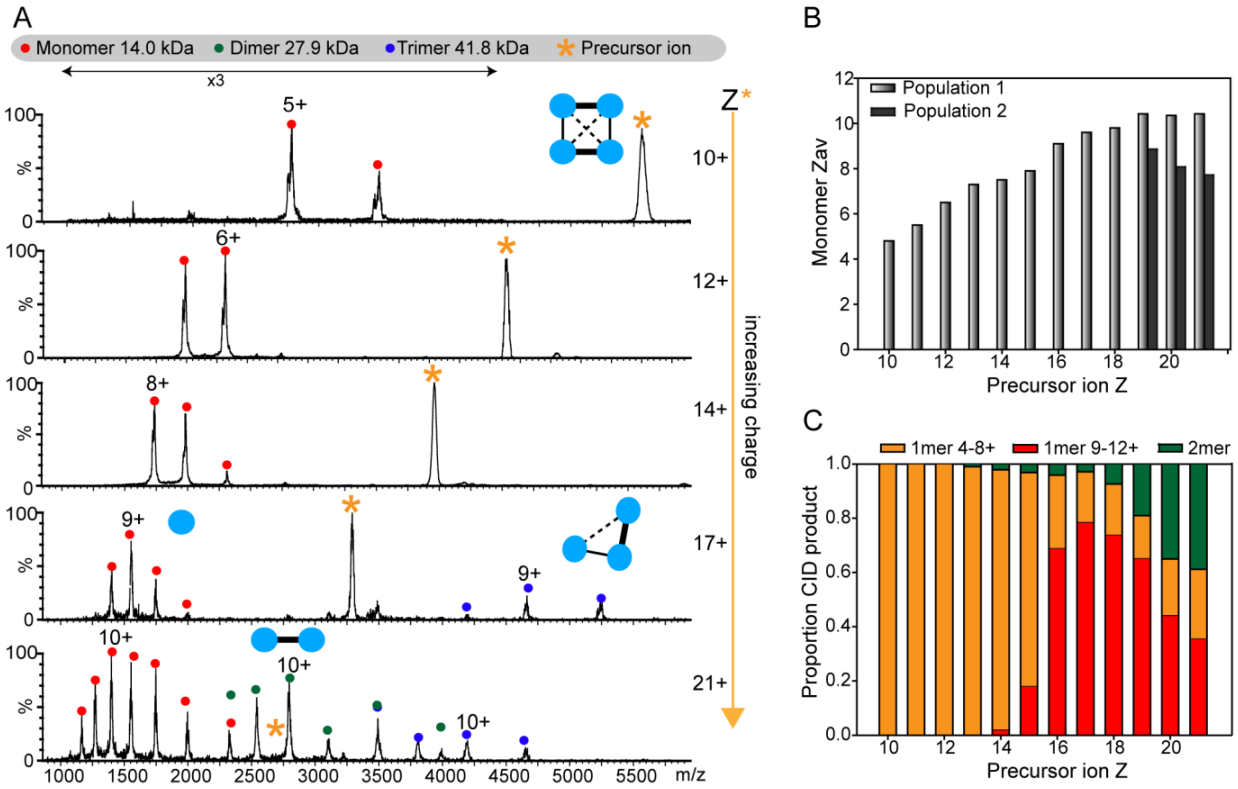


Figure 5.9. CID of TTR. **A** MS/MS of TTR with increasing precursor ion charge state. Initially, the dissociated monomer charge state increases with the precursor ion charge state. As TTR charge state is increased further by the addition of *m*-NBA, TTR can dissociate *via* alternate pathways, losing lower charged monomers and dimers. MS/MS spectra were acquired at similar laboratory frame energies (collision energy x Z). **B** The dissociated monomer average charge state is plotted against precursor ion charge state. Above Z=19+, two distinct populations of monomer are observed. **C** The relative intensities of monomer and dimer CID products in the MS/MS spectra are plotted against precursor ion charge state. Precursor charge states were generated using the following additives to 200 mM AA: 10-13+ (20 mM DBU), 14-16+ (no additive), 17- 21+ (1 % *m*-NBA).

As TTR charge state was increased further (> 18+) by the addition of supercharging *m*-NBA, TTR was found to dissociate atypically, losing lower charged monomers as well as dimers. The dissociated monomer Z_{av} was plotted against precursor ion charge state. Above Z=19+, two populations of monomer were observed, with distinct charge state distributions. Furthermore, for Z > 17+ increasing populations of dimer were observed with an increase in precursor ion charge state.

Observation of the dimer could also result from the sequential loss of two monomers. The charge state signature of the dimers suggests this is not the case for the high intensity dimer peaks. For example, TTR 21+ dissociates to form a dimer with Z_{av} ~ 10+. If this dimer (10+) arises as a consequence of the sequential loss of two monomers, the sum of charges on the two ejected monomers must be 11+ for the total charge to equal that of the intact complex with 21+. The dissociated monomers however have charge states 6+ to 12+. The low intensity peaks for charge states 7+ and 8+ of the dimer however could arise from the sequential loss of two monomers, and therefore we tentatively assign these charge states to the "stripped complex". Based on charge state however we conclude therefore that a major population of dimer is expelled intact from the tetrameric complex.

The assignment of SAP and TTR MS/MS spectra was aided by IM measurements. Peaks corresponding to nM^{nz+} and $(n+1)M^{(n+1)z+}$ have identical *m/z* values, however typically experience different drift times through the mobility device, as illustrated for SAP 26+ and TTR 21+ CID (Figure 5.10).

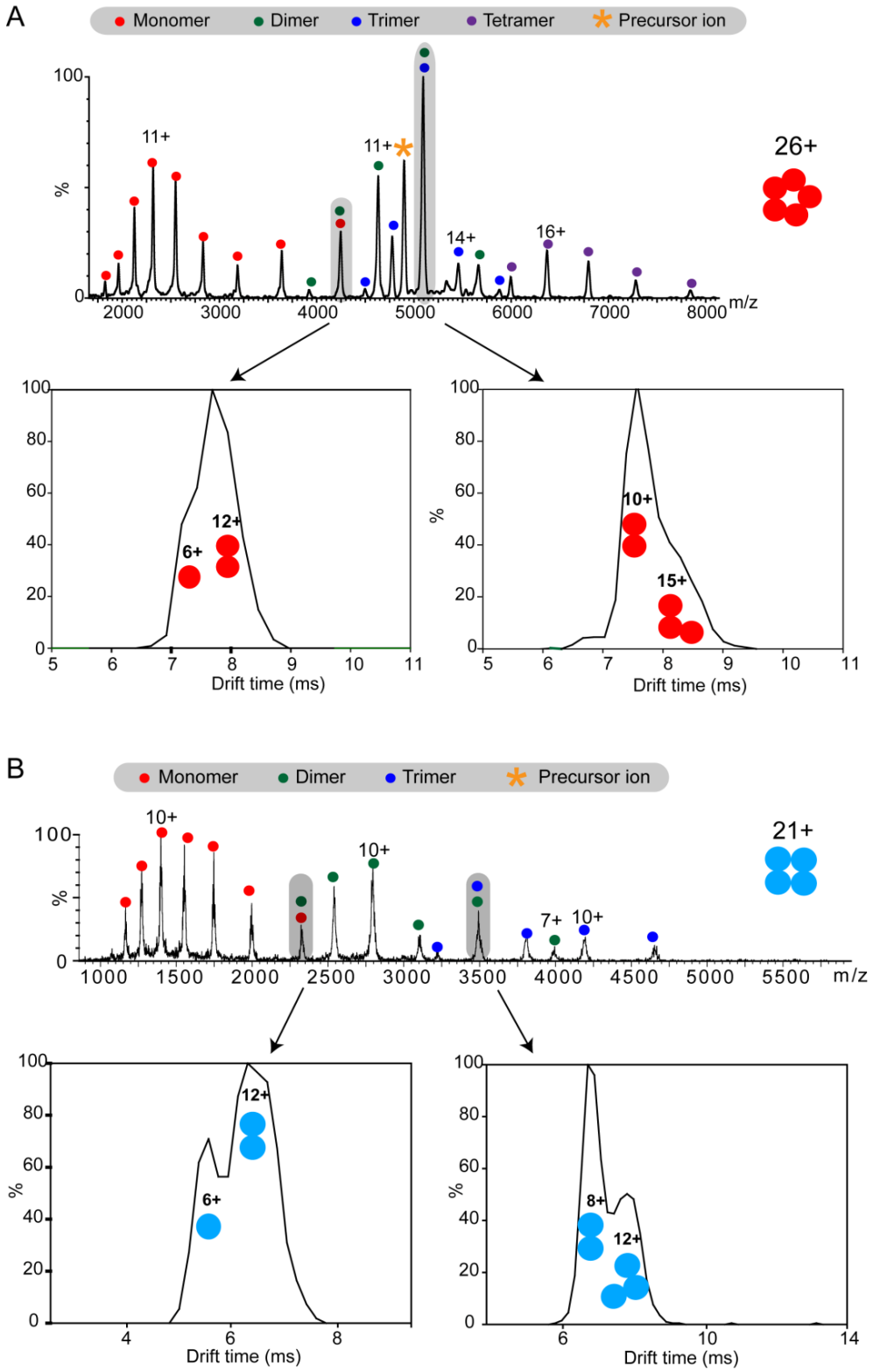


Figure 5.10. Ion mobility aids charge state assignment in MS/MS spectra. Peaks corresponding to nM^{nz+} and $(n+1)M^{(n+1)z+}$ have identical *m/z* values, however typically experience different IM drift times, as shown for SAP 26+ (**A**) and TTR 21+ (**B**).

The atypical dissociation resulting in the ejection of dimers and low-charged monomers was observed for high charge states of both SAP and TTR, accessed by supercharging. Therefore there are two possible explanations for this behaviour: either the high charge state of the precursor ion dictates the atypical behaviour, or the supercharging reagent itself affects the complex by weakening subunit interactions. In order to test these two possible scenarios, the dissociation of SAP 26+ was examined following ESI from either AA buffered or *m*-NBA containing solutions (Figure 5.11). The MS/MS spectra for this ion were closely similar, regardless of whether the ion resulted from supercharging with *m*-NBA or was electrosprayed directly from AA buffer. In each case, two dissociation pathways could be observed: the typical dissociation of highly charged monomers, and the atypical dissociation resulting in loss of dimers, and low charged monomers. These results imply that in this case it is the charge state of the ion which governs its dissociation route rather than the presence or absence of the supercharging reagent.

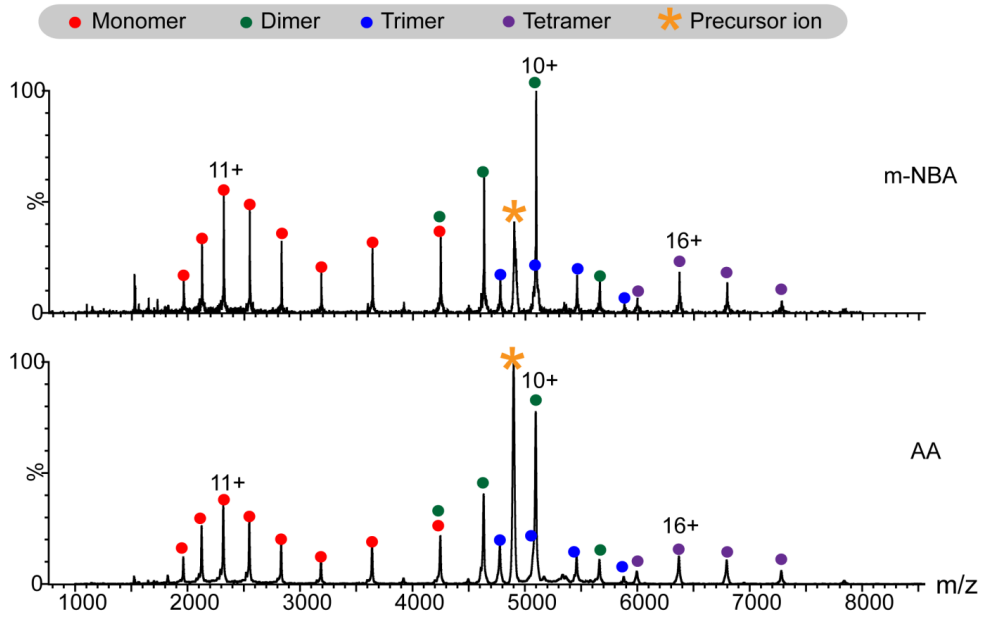


Figure 5.11. MS/MS of SAP 5-mer 26+ in the presence or absence of *m*-NBA. Similar dissociation patterns are observed for the 26+ ion, regardless of whether it was generated from AA buffered solutions, or from those solutions containing *m*-NBA.

5.2.5 Gas-phase *versus* solution-phase disassembly

To obtain insights into the structures of the dissociation products observed for SAP and TTR, IM-MS was employed. CCSs were measured for the dissociated monomers and dimers, and compared with CCS calculated from the corresponding X-ray crystal structures. In addition, SAP and TTR were disrupted in solution, by the addition of organic solvent. As shown in Chapter 4 for several heteromeric complexes, disruption in solution does not generally affect the CCS of the subcomplexes and subunits which make up the assembly. The CCS for the solution disassembly products for these two homomers were therefore measured and compared with the analogous gas-phase product.

SAP was disrupted in solution to form monomers, by the addition of 20 % acetonitrile (Figure 5.12). Different solution conditions were attempted, however no dimer was observed. The CCS was therefore measured for the monomer formed in 20 % acetonitrile solution, and compared with the CCS of monomer ions resulting from gas-phase dissociation. The CCS for the gas-phase monomer increases with monomer charge state. Whilst two conformations for the 8+ monomer were observed, the remaining charge states were characterised by one major feature each in their respective IM arrival time distributions. The CCS of the solution-formed monomer (charge states 8-10+) were in good agreement with the CCS of the lowest charge states (6-7+, 8+(*b*)) of the gas-phase dissociated monomer. These were also in excellent agreement with CCS calculated from the X-ray crystal structure. As no dimer was formed in solution, the gas-phase dimer was compared with that of a dimer from the X-ray structure. Two conformations were evident for the 11+ charge state, compact and slightly unfolded. The compact structure exhibited good agreement with that calculated from the X-ray structure. Higher charge states (12 and 13+) showed some unfolding while CCS of the low charged dimers (9+ and 10+)

were within 4-5 % of the calculated CCS. We conclude that the low charge dimers and monomers, formed in the gas phase, have compact structures in line with CCS predicted from the X-ray structure.

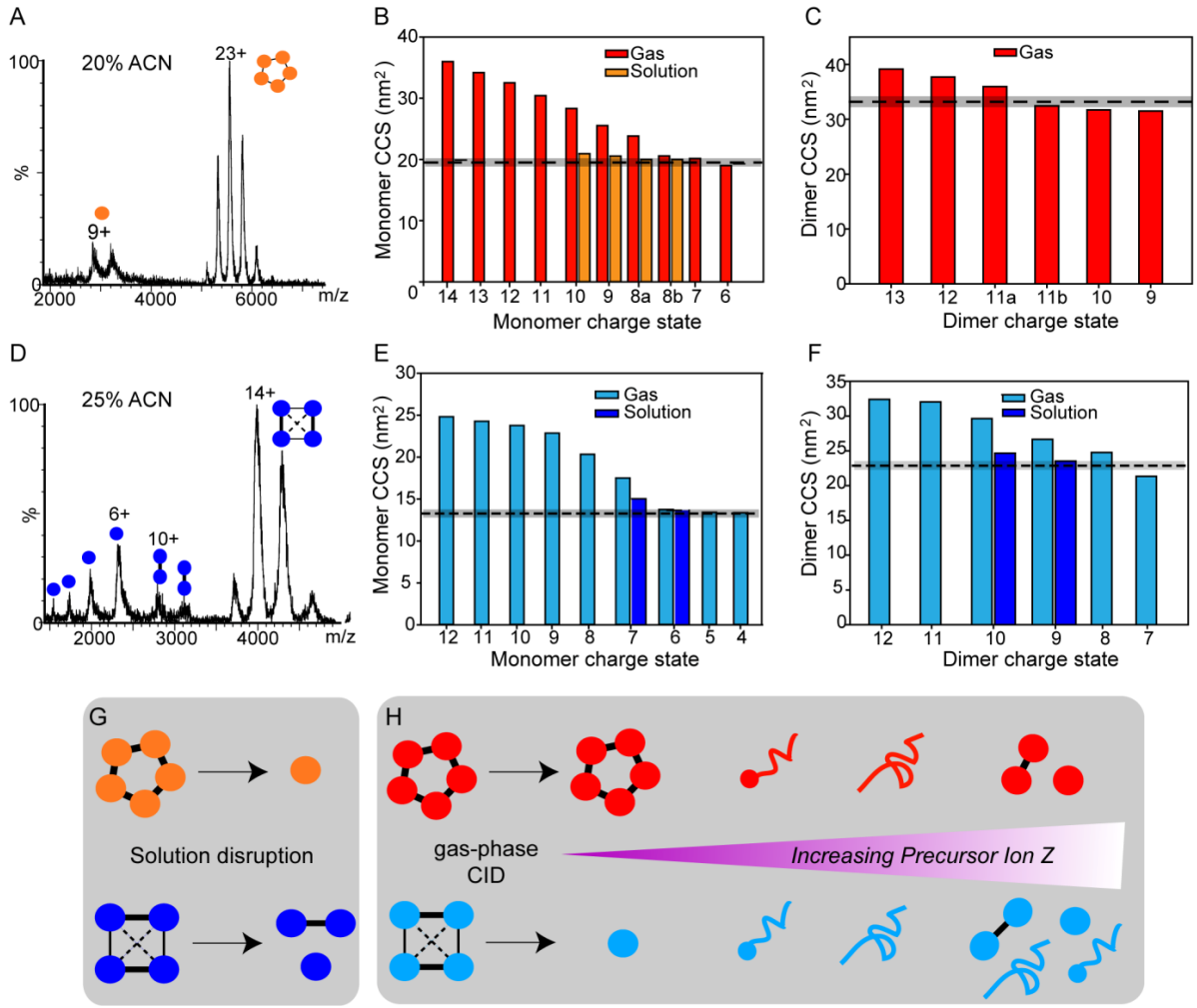


Figure 5.12. Comparison of gas-phase dissociation and solution disassembly. SAP and TTR were disrupted in solution by the addition of acetonitrile (**A** and **D**). CCS were measured for the monomer formed in solution, and compared with CCS of the gas-phase dissociated monomer (**B** and **E**), and values calculated from the X-ray crystal structure. CCS were also measured for dimer formed in solution and compared with CCS of the corresponding gas-phase product, and values calculated from the X-ray crystal structure (**C** and **F**). Calculated CCS from the X-ray crystal structure (scaled CCS_{PA} ±3 % error) are shown as the dashed line and grey shaded region. Schematic showing solution disruption (**G**) and CID, with increasing precursor ion Z (**H**).

Similarly, TTR was disrupted in 25 % acetonitrile solution to form monomers and dimers (Figure 5.12D). CCS were measured for the monomer (6-7+) and dimer (9-10+) formed in solution, and compared with CCS of the corresponding gas-phase products, and with values calculated from the X-ray crystal structure. The lowest charge states for the gas-phase monomer (4-6+) were in good agreement with both the CCS of the solution-disrupted monomer and the X-ray crystal structure. Interestingly, these low charge states of the monomer could be accessed by dissociation of either charge-reduced or supercharged TTR (Figure 5.12H). The CCS for the lowest charge state of the gas-phase TTR dimer (7+) and that expected from the X-ray crystal structure differed by ~ 6 %. On the other hand, higher dimer charge states had CCS greater than that expected, consistent with unfolding of one or more monomers.

It is noteworthy that the CCS of the CID products described here are independent of the charge state for the precursor ion from which they are ejected (Figure 5.13). For example, the CCS for the 7+ monomer, dissociated from the lowest and highest charge states of TTR, is indistinguishable within the precision of the measurement.

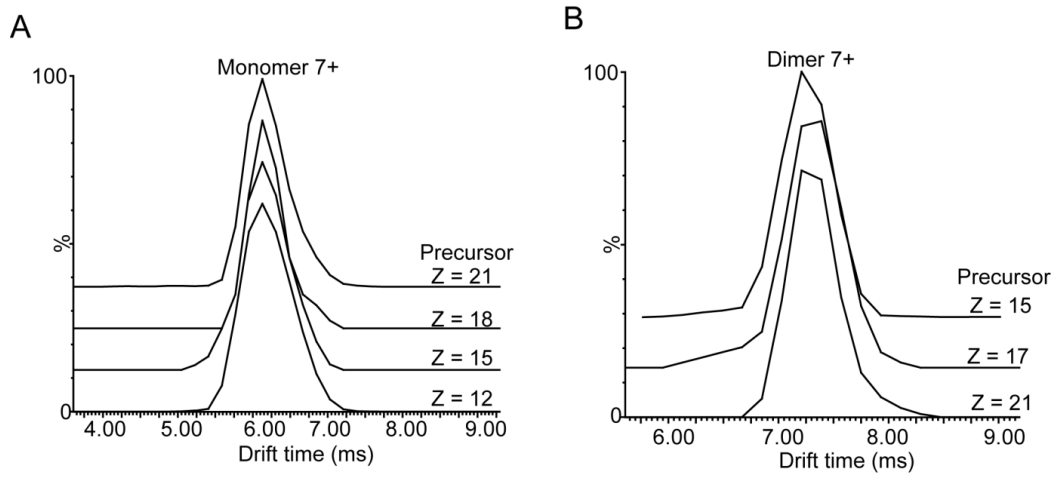


Figure 5.13. CCS of CID products are independent of the charge state for the precursor ion from which they arise. **A** and **B** show arrival time distributions for TTR monomer (7+) and dimer (7+), ejected from a range of precursor ion charge states.

CCS for higher order oligomers, i.e. the SAP 3-mer, SAP 4-mer, and the TTR 3-mer were measured and compared with CCS calculated for the corresponding X-ray crystal structures (Figure 5.14). These structures were obtained by removing monomer subunits from the complex crystal structures, and therefore are unlikely to reflect the most stable conformations in solution. In spite of this, the CCS for the lowest charge states of both SAP and TTR trimers were in good agreement with calculated CCS. Rearrangement to more stable structures, with similar CCS to the "clipped" X-ray crystal structures however cannot be ruled out. In contrast, the lowest charge states for the SAP 4-mer have significantly lower CCS than expected from the crystal structure. This suggests that substantial collapse occurs in the gas phase, burying the newly exposed surface area resulting from monomer dissociation. These results are in line with previous studies in which "stripped" complexes, particularly those whose subunits are arranged in ring-like architectures, have been shown to undergo significant collapse following dissociation in the gas phase (*244-245*). As with the monomers and dimers above, CCS of the higher order oligomers increase with charge state, consistent with the unfolding of one or more monomers.

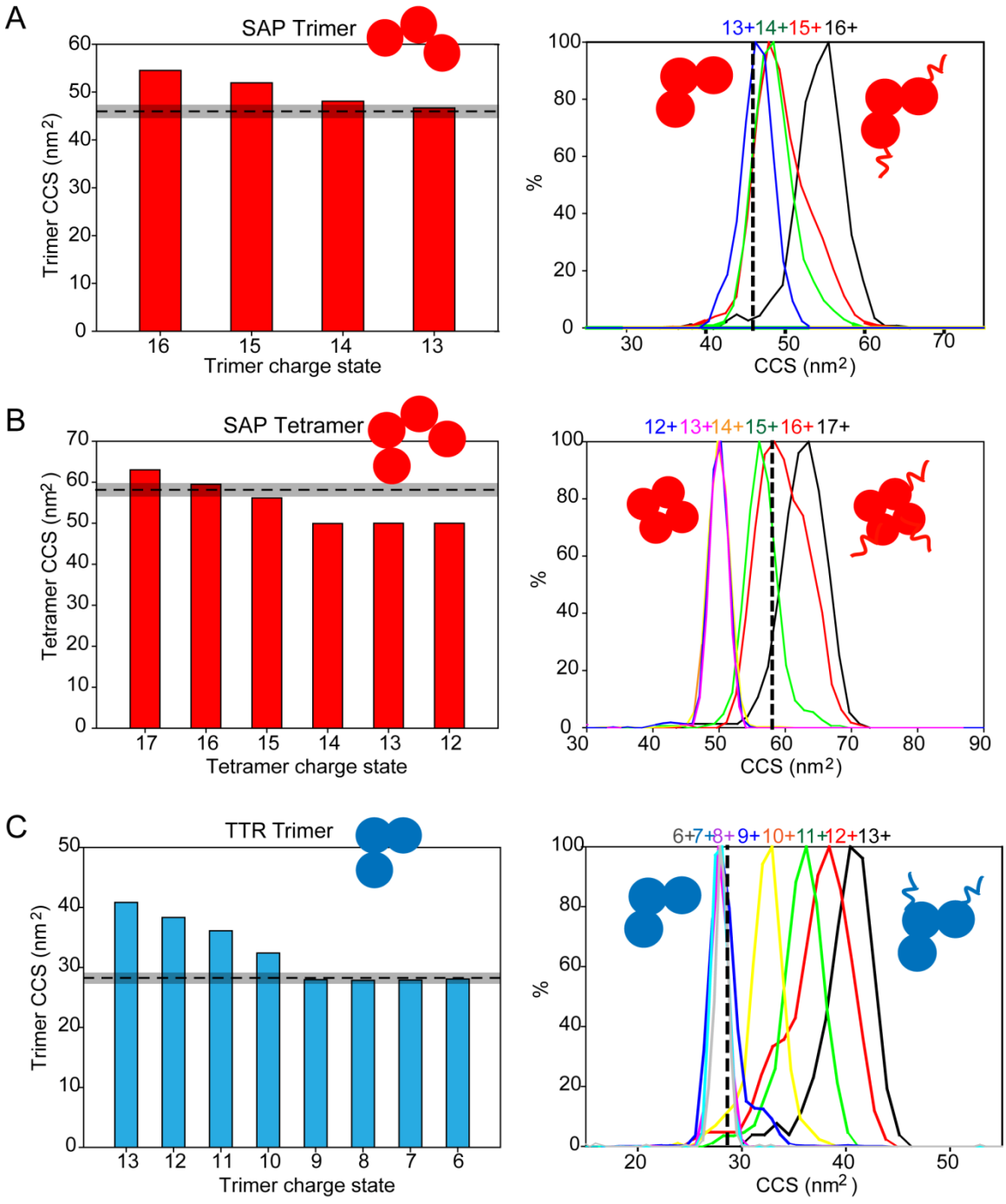


Figure 5.14. Higher order oligomers show evidence for structural rearrangement. CCS were measured for SAP 3-mer (**A**), SAP 4-mer (**B**) and TTR 3-mer (**C**) and compared with the CCS calculated for the corresponding X-ray crystal structures. Arrival time distributions on a CCS axis for the individual charge states are shown on the right hand side. The dashed line and shaded region represents the scaled CCS_{PA} ± 3 %.

Importantly, this study of TTR and SAP reveals that the lower charge states of monomer and dimer ejected during CID are compact. These compact CID products are favoured by very high precursor ion charge states (both SAP and TTR), and by very low precursor ion charge states (TTR). IM-MS however can only provide information on overall shape and not on the detailed tertiary structure. Although a compact and non-native state cannot be ruled out for these dissociation products, the excellent agreement between the CCS of the gas and solution-disrupted products, and their corresponding agreement to the X-ray crystal structure, suggest that these gas-phase products retain compact structures which may reflect native tertiary structure.

5.2.6 Dissociation of heteromeric tryptophan synthase

Having demonstrated that supercharged species of SAP and TTR can access atypical dissociation pathways resulting in compact gas-phase products, the next step was to test this on a heteromeric complex. Tryptophan synthase (143 kDa), with its established solution disassembly (Chapter 4) was chosen. The CCS for a wide range of charge states of the intact complex were measured (Figure 5.15). This included those formed from charge reduction with TEA (19-22+), from AA buffered solutions (22-27+) and those generated *via* supercharging with *m*-NBA (27-31+). These CCS were compared with previous measurements made for the same complex in buffered solution (Chapter 4). All charge states from 19-29+ had CCS within the error of previous CCS measurements. Charge states 30+ and 31+ however had larger CCS than anticipated, consistent with unfolding.

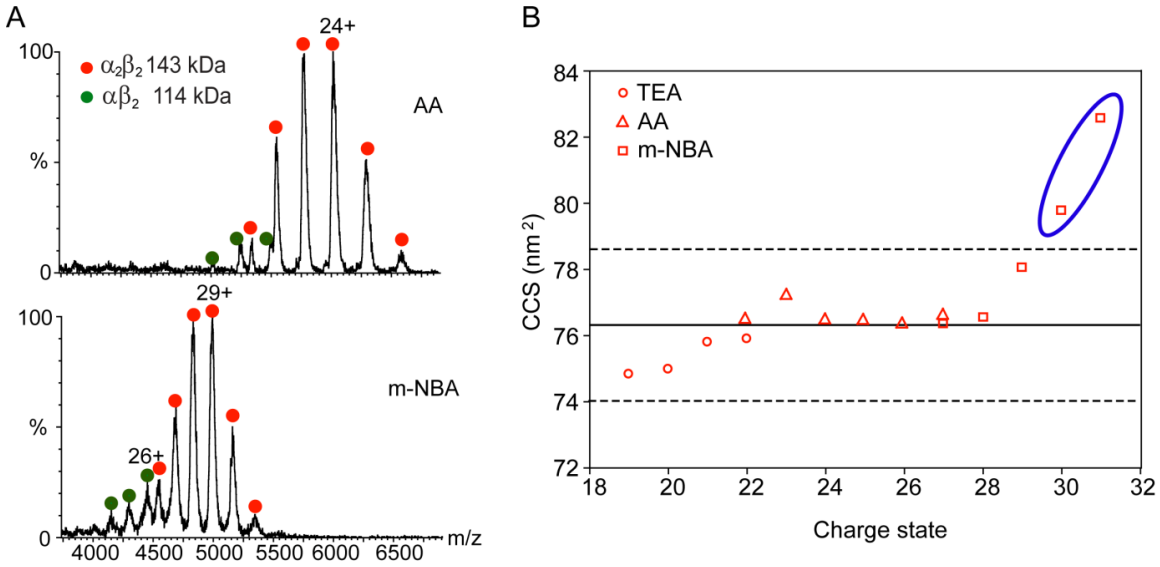


Figure 5.15. Charge manipulation of heteromer tryptophan synthase. **A** MS spectra for tryptophan synthase in AA, and with 1 % supercharging reagent *m*-NBA. Minor proportion of trimer is present in solution, in addition to the intact complex. **B** CCS of tryptophan synthase plotted against charge state. Data from TEA, AA and *m*-NBA-containing solutions are shown. Unfolding is observed for charge states > 30+. Previous CCS measurements (charge-averaged, Chapter 4) and associated error (± 3 %) are shown as solid and dashed lines.

Tandem MS experiments were performed on the intact tryptophan synthase in AA buffer. CID of the 24+ ion resulted in the loss of α-monomers (Figure 5.16). Unexpectedly, the monomer had a range of charge states, from 6+ to 16+. Closer inspection of MS/MS and IM data revealed three distinct populations of α-monomer, with overlapping Gaussian charge state distributions centred on 14+, 11+ and 7+. This could be the result of non-interconverting conformational isomers of the intact tryptophan synthase complex (*238*). Close examination of the arrival time distributions for the precursor ion however showed no evidence for multiple conformations. Therefore this result suggests that there are three parallel dissociation pathways, resulting in loss of highly unfolded, partially unfolded and compact α-monomer.

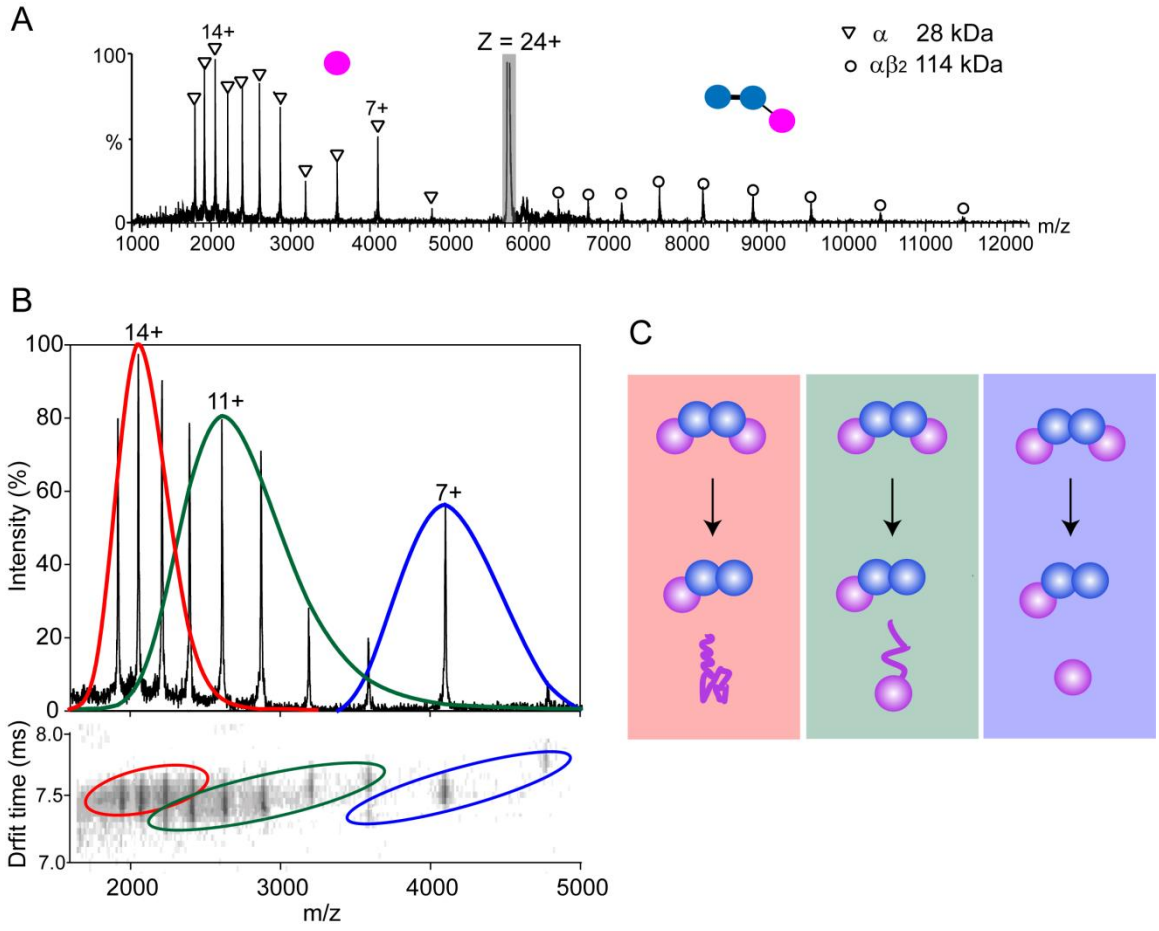


Figure 5.16. CID of tryptophan synthase. **A** MS/MS of tryptophan synthase (24+) reveals loss of α-monomers. **B** Closer examination reveals three distinct populations of monomer, with respect to their charge state distributions. Drift time *vs m/z* contour plot shows three populations of monomer with respect to their conformations (CCS). **C** Schematic for the proposed parallel routes for gas-phase dissociation: loss of a highly unfolded (high charge states, red), partially unfolded (intermediate charge states, green) and compact (low charge states, blue) α-monomer.

To explore how the precursor ion charge state affected the charge on the monomeric CID products, MS/MS was performed on the range of tryptophan synthase charge states resulting from charge reduction and supercharging experiments (Figure 5.17A). The change in relative intensities of the charge state distributions of the α-monomer defined as low (6-9+), intermediate (10-13+) and high (14-17+) were plotted as a function of the precursor ion charge state (Figure 5.17B). Interestingly the maximum population of the

highest charged ejected monomer occurs from an intermediate precursor charge state of 25+. Plotting Z_{av} of the dissociated α-monomer against the precursor ion charge states reveals three distinct trends. i) Initially as the precursor ion charge state increases, the monomer Z_{av} increases. This corresponds to a reduction in intensities of the lowest charge states, and an increase in the higher charge states. ii) For Z > 26+, a corresponding decrease in monomer Z_{av} is observed. This is consistent with an increase in tryptophan synthase charge state favouring the dissociation of lower charged monomers. This shift in trend is analogous to observations made for supercharged TTR and SAP wherein high charge states resulted in ejection of compact low-charged monomers and dimers. iii) For Z > 29+, a slight increase in monomer Z_{av} is observed with increasing precursor ion charge state. This observation can be explained due to the fact that these ions (30+ and 31+) are partly unfolded prior to MS/MS experiments (Figure 5.15).

The CCS of the gas-phase dissociation products were then measured and compared with the corresponding solution disassembly species, and with the CCS calculated from the X-ray crystal structure. Disruption in solution was carried out by the addition of 30 % methanol or 30 % methanol/ 10 % DMSO as described in Chapter 4. This resulted in sequential loss of α-subunits from the β_2-dimer core (Figure 5.18). At higher laboratory frame energies, the loss of two monomers was observed during CID, and therefore CCS measurements could also be undertaken for the β_2-dimer, in addition to the α-monomer and $\alpha\beta_2$-trimer. CCS for the lowest charge states (6-7+) of the gas-phase α-monomer were in good agreement (<1 % deviation) with the α-monomer in solution (8-10+), and with the CCS calculated from the X-ray crystal structure (2 - 4 % deviation). Similarly all observed charge states of the gas-phase β_2-dimer (10-13+) had CCS in excellent

agreement with that of the dimer formed in solution (14-17+) and calculated from the crystal structure (1 - 4 % deviation). The lowest charge states for the trimer had CCS lower than expected, consistent with post-dissociation rearrangement to more stable structures, as discussed above for the higher order oligomers of SAP. Increasing CCS for higher charge states suggests unfolding. This was observed for higher charge states of the monomer and trimer, although was not noted for the dimer.

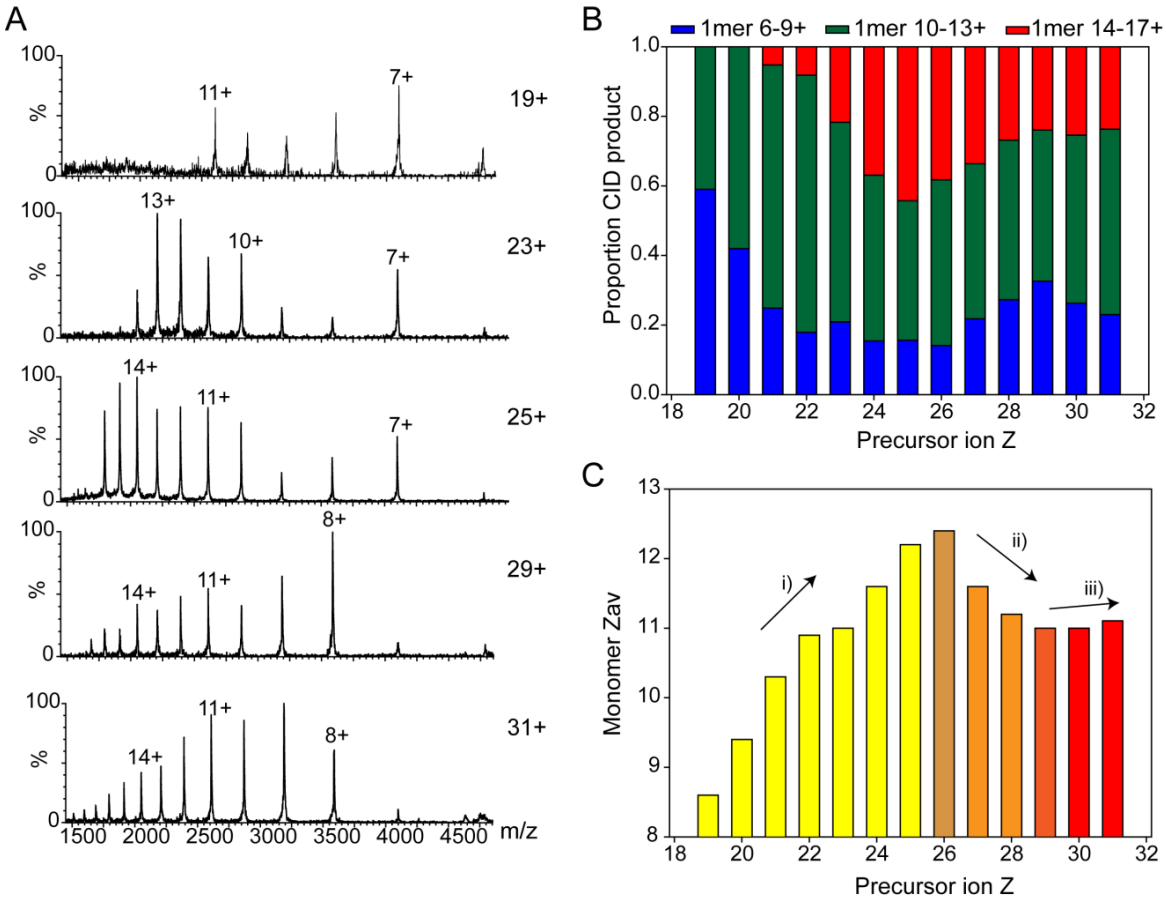


Figure 5.17. Charge-state dependent dissociation of tryptophan synthase. **A** MS/MS of tryptophan synthase with increasing precursor ion charge state. **B** The relative intensities of low (6-9+, blue), intermediate (10-13+, green) and high (14-17+, red) charged α-monomer are plotted against precursor ion charge state. **C** The Z_{av} for the dissociated α-monomer is plotted against precursor ion charge state. Three trends are evident, and are described in the text. Charge states were generated using the following additives to 200 mM AA: 19-21+ (20 mM TEA), 22-26+ (no additive), 27-31+ (1 % *m*-NBA).

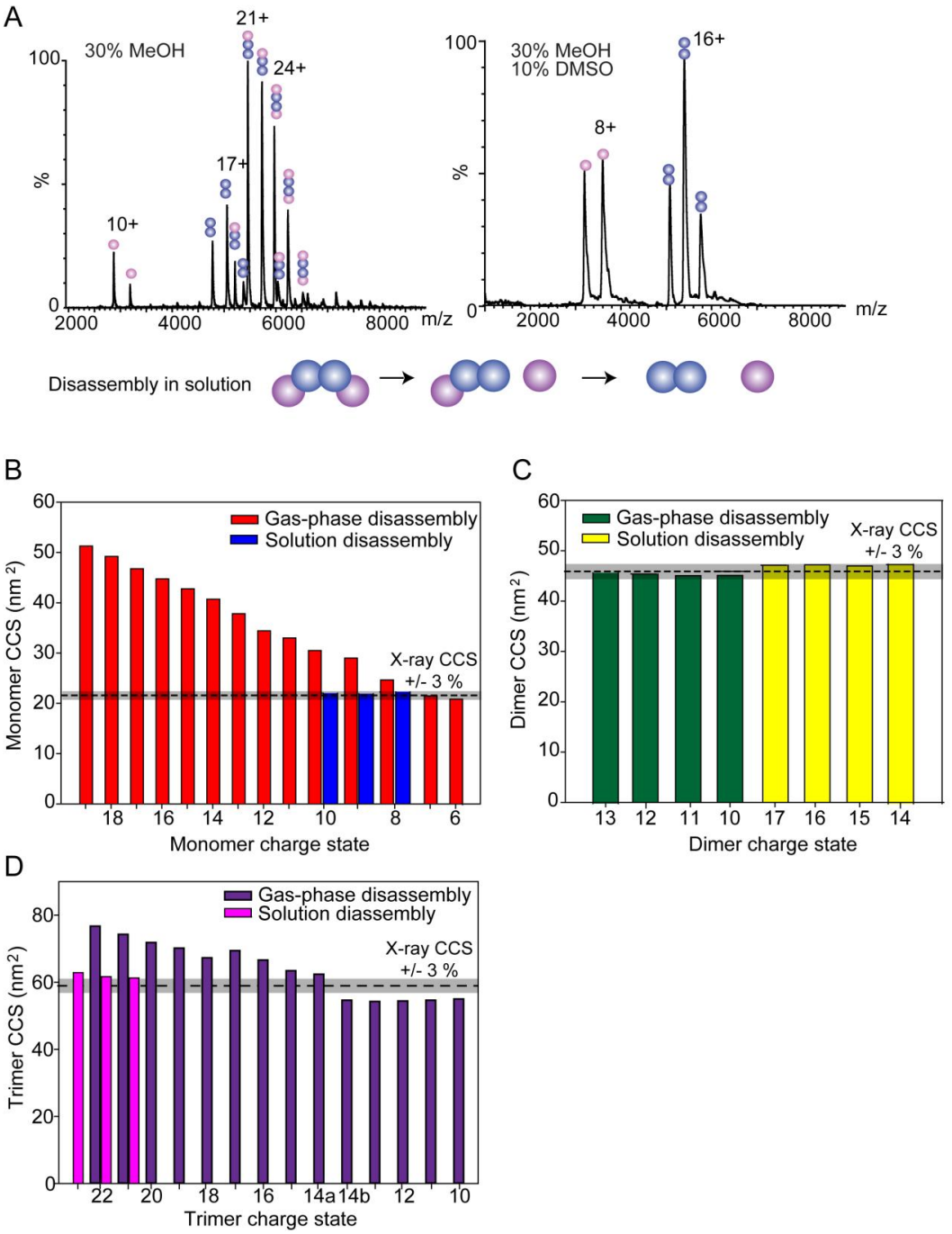


Figure 5.18. Comparing gas- and solution-phase disassembly pathways. **A** Tryptophan synthase disassembles in solution to give $\alpha\beta_2$-trimer, β_2-dimer and α–monomer. CCS were measured for the α–monomer (**B**), β_2-dimer (**C**) and $\alpha\beta_2$-trimer (**D**) formed in solution (blue, yellow, pink respectively), and compared with CCS of the corresponding gas-phase product (red, green, purple). CCS were compared with calculated values. Calculated CCS from the X-ray crystal structure (scaled $CCS_{PA} \pm 3$ %) are shown as the dashed line and grey shaded region.

5.2.7 The role of salt bridges

In order to investigate the general applicability of using supercharging to promote alternate dissociation pathways, the study was extended to include additional protein complexes. These included avidin, ADH, pyruvate kinase, glutamate dehydrogenase, MMOH, ToMOH, NBDO and β-galactosidase. Interestingly, the atypical pathways observed for SAP and TTR were also observed for avidin. The remaining seven complexes however dissociated exclusively through the typical dissociation pathway with loss of a highly-charged unfolded monomer even under supercharging conditions. These observations suggest that not all complexes will undergo atypical dissociation despite accessing high charge states. This prompts the question as to why high charge states of certain protein complexes become prone to alternate dissociation pathways while others retain the typical dissociation pathway. To establish a link between the structure and the dissociation pathway several options were considered including the number of disulfide bridges, the average charge density and the number of interfacial salt bridges. To increase the dataset complexes previously reported as undergoing atypical dissociation were included in the analysis (Table 5.1).

Considering first the number of disulfide bridges: the atypical dissociation of textilotoxin (*247*) and phospholipase A_2 (7 disulfides per subunit each) (*42*) has been attributed to the high proportion of disulfides, which inhibit unfolding and promote dissociation of compact species. Interestingly, the remaining complexes in Table 5.1 undergoing atypical CID had only 0-1 disulfide bridges per subunit, indicating that additional factors are important for this phenomenon.

Table 5.1 Investigating factors directing atypical dissociation

Complex	PDB	SASA (Å^2)	Z_{av} [a]	Z_{av}/SASA (x 10^3 Å^{-2})	Interfacial SB[b]	Disulfide bridges	Atypical CID	Ref
TTR 55 kDa, 4-mer	1F41	19290	17	0.88	0	0	Y	This study
avidin 64 kDa, 4-mer	1VYO	18190	19	1.04	0	4	Y	This study
SAP 125 kDa, 5-mer	1SAC	38190	28	0.73	6	5	Y	This study
tryptophan synthase 143 kDa, 4-mer	1WBJ	43290	29	0.67	1	0	Y	This study
shiga-like toxin I 38.5 kDa, 5-mer	1BOS	13960	13	0.93	6	5	Y	(*38*)
2-keto-3-deoxy-arabinoate dehydr. 133 kDa, 4-mer	3BQB	42708	27	0.63	8	4	Y	(*240*)
stable protein 1 147 kDa, 12-mer	1TR0	45270	31	0.68	4	0	Y	(*239*)
ecotin 32 kDa, 2-mer	1ECY	16678	15	0.90	2	2	Y	(*246*)
calcineurin 62 kDa, 2-mer	1AUI	23620	15	0.64	2	0	Y	(*238*)
phospholipase A_2 26.4 kDa, 2-mer	1MH2	11410	11	0.96	4	14	Y	(*42*)
concanavalin A 102 kDa, 4-mer	3D4K	32680	28	0.86	2	0	Y	(*244*)
textilotoxin I/II 84/7 kDa, 6-mer	-	-	17	-	-	42	Y	(*247*)
Average				**0.81 ± 0.15**	**3.2 ± 2.6**	**6.9 ± 12.3**		
glutamate dehydrogenase 336 kDa, 6-mer	1NR7	107290	41	0.38	3	0	N	This study
β-galactosidase 465 kDa, 4-mer	3IAP	136712	50	0.37	4	0	N	This study
ToMOH 212 kDa, 6-mer	2INC	59750	31	0.52	5	0	N	This study
MMOH 251 kDa, 6-mer	1MTY	60640	35	0.58	12	0	N	This study
NBDO 218 kDa, 6-mer	2BMO	58630	30	0.51	8	0	N	This study
ADH 143 kDa, 4-mer	2HCY	48930	28	0.57	6	4	N	This study
pyruvate kinase 232 kDa, 4-mer	1F3W	69290	36	0.52	8	0	N	This study
Average				**0.49 ± 0.09**	**6.6 ± 3.0**	**0.6 ± 1.5**		

[a] Z_{av} is the average charge state observed upon addition of 1 % *m*-NBA (this study) or reported in the literature in varying solution conditions.

[b] Number of interfacial salt bridges broken during first step of *in silico* disassembly, based on breaking the minimum interfacial surface area. Salt bridges were calculated using PISA.

Previously, high charge states of cytochrome *c* dimer were shown to dissociate similarly to those dimers whose individual subunits had been conformationally restricted by cross-linking (*42*). This suggests that the charge on the precursor complex ion is important. This was further explored in a study of stable protein 1, in which it was suggested that in addition to structural features restricting unfolding, a high charge density contributed to the unusual dissociation of this complex *(239)*. Therefore, the charge per unit area for the average observed charge state in MS (Table 5.1) was investigated. The complexes for which atypical dissociation pathways were observed had significantly higher charge densities on average ($(0.81 \pm 0.15) \times 10^{-3}$ Å^{-2} *versus* $(0.49 \pm 0.09) \times 10^{-3}$ Å^{-2}), than the complexes dissociating typically (Wilcoxin rank-sum test, $p = 0.00029$, 1-sided).

Next, the interfacial salt bridges were examined. This was considered important, as electrostatic interactions increase in strength in the gas phase (*77*), and have been implicated in stabilising protein structure *in vacuo (72)*. The absolute number of salt bridges which would be broken during dissociation could therefore influence the likelihood of dissociation with no prior unfolding, i.e. "direct dissociation" into subcomplexes and/or subunits. In order to test this, the most likely route for direct dissociation without unfolding was considered. Previous studies have predicted evolution and assembly pathways of homomeric protein complexes from their crystal structures (*248*). These studies propose that the interface size dictates the assembly pattern of subunits into a complex *in vivo*. Similarly during disassembly of the complex *in vitro*, the weakest interfaces are broken first. Using the so-called "interface model", the strength of each interface is assumed to be proportional to the surface area buried between the two subunits. Therefore, using this simple model, the complexes were disassembled *in silico*, with each step requiring the disruption of the smallest total interface area (Figure 5.19).

The number of salt bridges that would be broken during the first disassembly step was calculated. It is interesting to note that the complexes for which an alternate gas-phase dissociation pathway was observed, tend to be those in which the minimum number of salt bridges were broken (Figure 5.19 and Table 5.1) during *in silico* disassembly (3.2 ± 2.6 *versus* 6.6 ± 3.0). Using the Wilcoxin rank-sum test (*249*), this was found to be statistically significant (p = 0.0178, 1-sided). This suggests that a higher number of salt bridges increase the barrier to dissociate, without unfolding.

A trend is also apparent for the size (molecular weight, or SASA) of the complex, and its preferred dissociation route. Complexes which dissociate atypically are generally smaller than those dissociating typically. However, no significant difference in the interface size between the typical and atypical complexes was observed. Furthermore the number of interfacial salt bridges and interface area are poorly correlated. These observations suggest that the interfacial salt bridge trend is unrelated to the complex size.

Interestingly, smaller complexes tend to have a greater % increase in Z_{av} when exposed to a given concentration of *m*-NBA (Figures 5.2-5.5) than larger complexes. This may be related to the fact that smaller complexes have a greater surface area to volume ratio, and this has implications for the mechanism of supercharging. Whilst charge density can be considered to be a driving force behind atypical *versus* typical dissociation, it is the combination of accessible surface area and supercharging reagent concentration (Figure 5.6) that determines the charge density. Based upon this, a higher concentration of supercharging reagent might be expected to induce atypical dissociation in some of the typically dissociating complexes, however unfolding in solution may also increase.

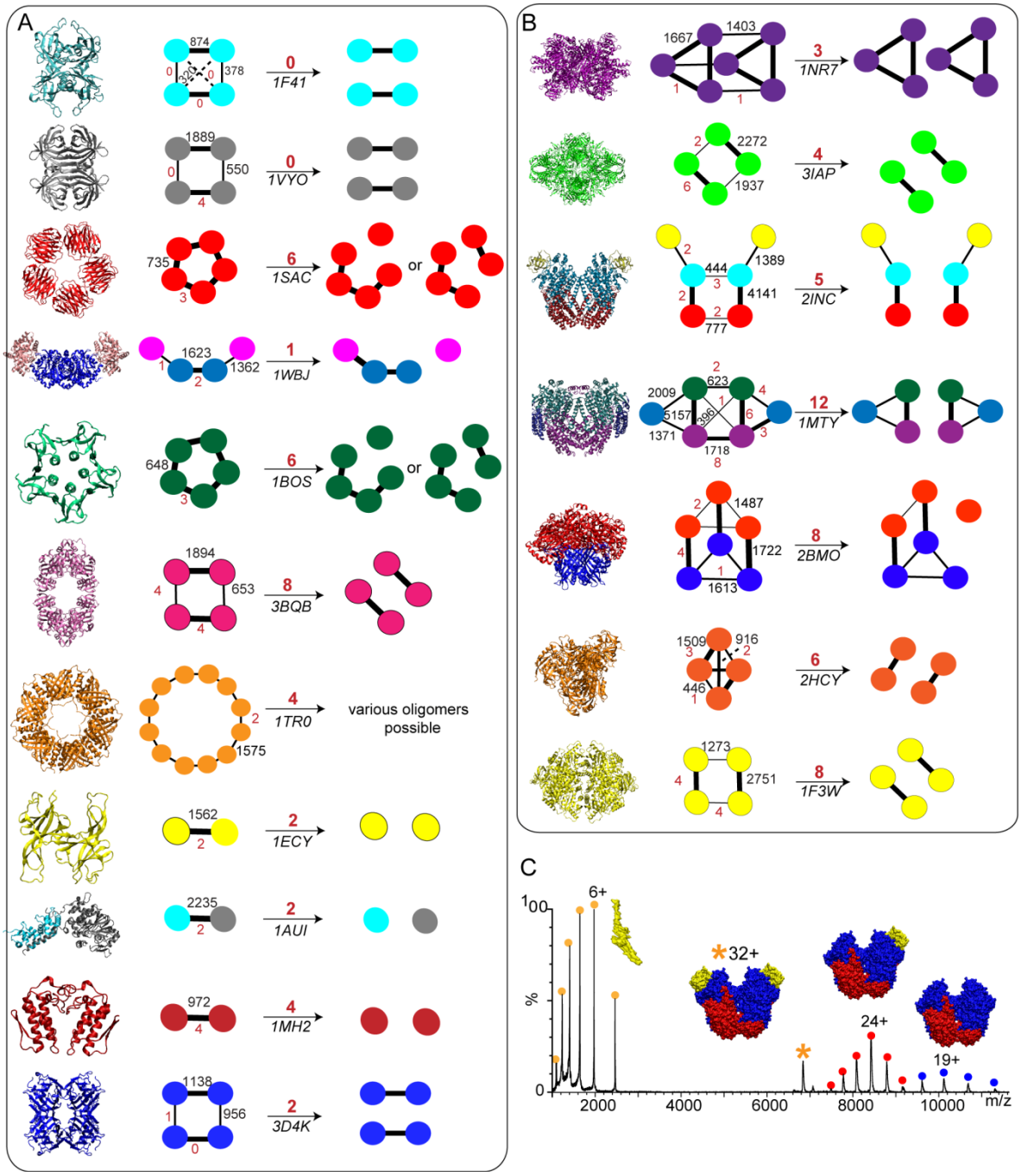


Figure 5.19. *In silico* disassembly of complexes using the interface model. Disassembly proceeds by the route in which the least amount of interface area is broken. The crystal structures of the complexes (atypical CID = **A**, typical CID = **B**) are shown with their corresponding PDB numbers, and ball-and-stick representations (ball = monomer; stick = interface). Interface areas in $Å^2$ (black) and number of salt bridges (red) calculated using PISA are given. The number of salt bridges broken in each simulated disassembly step is given above the reaction arrows. **C** Example of typical CID is shown in the MS/MS spectrum of ToMOH 32+.

5.3 DISCUSSION

Manipulation of charge states of protein complexes offers several advantages. One advantage which has not been explored in detail is the effect on dissociation of multimeric complexes. In this chapter, the dissociation of protein complexes was shown to be highly charge-state dependent. Of particular interest is the behaviour of very high charge states, accessed by supercharging, where the dissociation of compact monomers and dimers was observed for several complexes. The lowest charge states for these dissociation products had similar CCS to the corresponding solution disassembly products and their X-ray crystal structures. IM-MS experiments however cannot report on the precise packing of secondary structural elements or of the side chain packing of the gas-phase structure; all which are necessary to define the fully folded native state in solution. However available experimental evidence implies that a compact state of a protein can survive following dissociation by gas-phase CID.

The structural information typically available from CID is the identity of peripheral subunits, which dissociate preferentially (*45*). Whilst the time scale of CID is in the order of microseconds, SID, in which dissociation results from the protein complex colliding with a solid surface, occurs much faster, on the picosecond time scale. The rapid deposition of energy in SID results in dissociation occurring on a faster time scale than unfolding, and as a result dissociation tends to occur in a charge-symmetric fashion (*35, 243, 250*). Recently IM experiments have shown SID products to be compact, with CCS in good agreement with calculated values (*162*). SID typically results in subcomplex dissociation, rather than the loss of single monomers as typically observed in CID (*35*). For example SID was used to disassemble heterohexamer nitrile hydratase into its constituent trimers (*250*). By the dissociation of complexes into their component

subcomplexes, detailed information can be attained on subunit connectivity. Furthermore, if these species remain in their folded states, CCS measurements can be highly informative for the structures of these subassemblies.

The results presented in this chapter show that similar to SID, subcomplexes and compact monomers can be observed following CID of supercharged species. It is possible that the increased intermolecular repulsive forces in the higher charge states reduce the energy barrier for dissociation (*251*). However, it is more likely the subtle energetic balance between the competing processes of unfolding and dissociation are the key to this phenomenon (*252*). Individual features of complexes undoubtedly play a role: the ability to attain higher charge densities and the number of salt bridges between subunits are just two factors identified and explored here. Whatever the origin of this dissociation, its observation prompts further investigation, particularly as it may yield heterodimers and trimers from heteromeric complexes.

5.4 SUMMARY

The CCS for several homomeric protein complexes was monitored as a function of charge state. Charge reduction using solution additives, was achieved without perturbing the structure, in that unfolding was not detected. As the charge state was increased, through the use of supercharging reagents, the CCS increased significantly with charge. These effects were more pronounced for sulfolane than for *m*-NBA, and with higher concentrations of reagent. However for some protein complexes supercharging does not appear to disrupt the structure. Consequently, a higher charge state does not necessarily imply unfolding.

Additionally, a systematic investigation was carried out on the effects of manipulating charge states in solution on the gas-phase dissociation of multimeric protein complexes. Supercharged species for four of the investigated complexes dissociated atypically, ejecting compact monomers and/or dimers. These gas-phase dissociation products had CCS in good agreement with the corresponding solution disassembly products and with those calculated from X-ray crystal structures. A dataset comprising of complexes dissociating atypically was compared to a set of protein complexes that dissociate through the more typical route. Analysis of their interfaces and charge densities suggest that complexes with fewer interfacial salt bridges and higher charge densities are more likely to undergo atypical CID. This study highlights the potential for generating subcomplexes in the gas phase with charge manipulation, thus affording structurally informative dissociation routes for CID of multimeric assemblies.

6 PROBING GAS-PHASE PROTEIN COMPLEX STRUCTURES BY ION MOBILITY AND MOLECULAR DYNAMICS[9]

6.1 INTRODUCTION

Noncovalent MS is becoming an increasingly important adjunct to the structural biology of protein complexes, enabling elucidation of the topology and architecture of both soluble and membrane-embedded macromolecular protein assemblies (*3, 82*). MD studies carried out *in vacuo* demonstrate that the major features of protein structures are preserved when transferred from solution into the gas phase (*63, 66, 83, 253-254*). These protein structures are thought to be kinetically trapped in near-physiological states, close to those of their X-ray crystal structures (*67*). Recently IM-MS has provided evidence that, under appropriate conditions, native-like topologies can be maintained for protein complexes in the absence of bulk solvent (*39, 84*). Additional conformational stability can be conferred by ligand binding (*163*) and the addition of counterions (*155, 255*). Collapse to compact states in the gas phase, with smaller CCS than calculated for their native-like structures, has been observed for some protein complexes (Chapters 3 and 4) (*86*), however this phenomenon is not fully understood.

In this chapter, a combined experimental and theoretical approach, using IM-MS and MD simulations, was used to investigate the structures of globular and intrinsically disordered multi-subunit protein complex ions. All the globular complexes chosen for this study had

[9] Part of this work (p53 study) was in collaboration with Drs Kevin Pagel (IM-MS experiments) and Eviatar Natan (protein expression and purification), for which I carried out data analysis and computational simulations. I performed all other experiments and simulations detailed in this chapter.

similar secondary structural features, with low α-helical content and high β-sheet content (Figure 6.1). Due to the different relative gas-phase stabilities of α-helices and β-sheets (*256-258*), this was considered important. Whilst SAP and TRAP are ring-like structures with large central cavities, avidin and TTR resemble more globular compact structures, with few pockets and internal cavities.

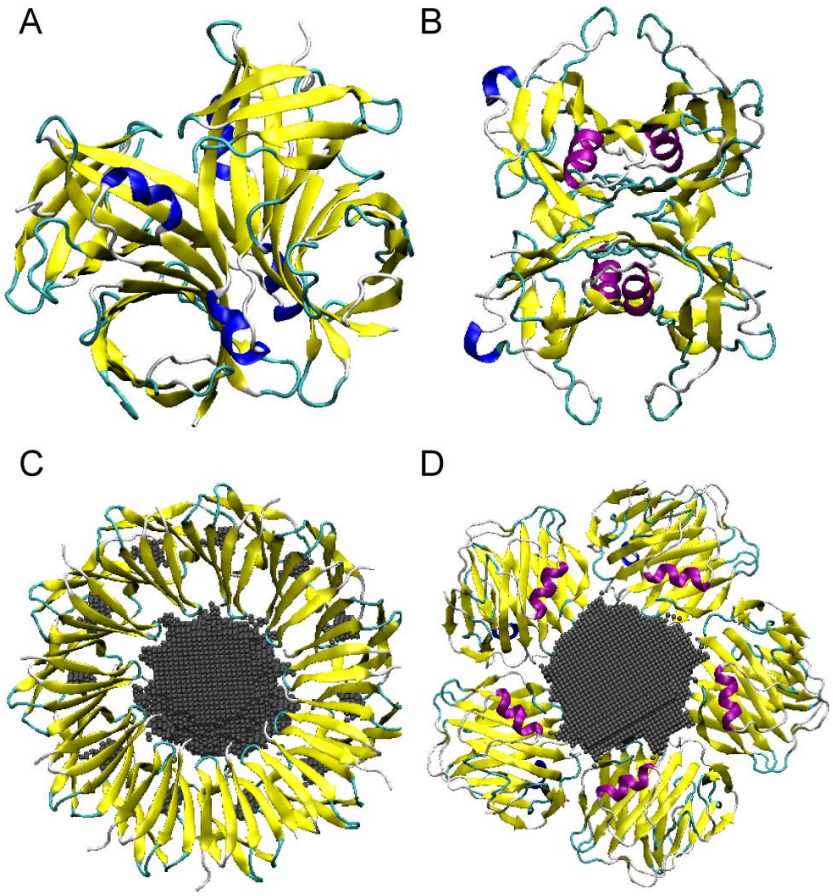


Figure 6.1. Structural features of globular protein complexes. The protein complexes avidin, TTR, TRAP, SAP (**A** to **D**) are coloured according to their secondary structure: yellow = β-sheet; purple = α-helix; dark blue = 3_{10}-helix; light blue = loop. Internal cavities for SAP and TRAP calculated using POCASA are coloured grey.

In contrast to globular proteins, intrinsically disordered proteins lack a well-defined folded structure. Native-like topologies have been shown for globular protein complex ions at low activation energies. It is not clear however how transfer to the gas phase would affect the structures of complexes containing intrinsically disordered domains, with their high proportion of charged residues (*259*). Analogous to the extended conformations of highly charged individual proteins, these regions could remain elongated in the gas phase. Alternatively in the absence of solvent these unstructured regions could collapse onto folded domains. To examine these possible scenarios IM-MS was applied to one of the most studied natively disordered complexes, that of the tumour suppressor protein p53.

To address the effect of charge state on the structure of protein complex ions, charge manipulation was carried out in solution. This was achieved using additives with high gas-phase basicity for charge reduction (*221*) and low volatility supercharging reagents (*233*) to increase the charge. For protein complexes containing a central cavity, such as SAP and TRAP, the lowest charge states were able to access more compact conformations, upon activation. By contrast, all charge states of protein complexes containing intrinsically disordered regions exhibited a spontaneous compaction. In all cases, the magnitude of compaction observed experimentally using IM-MS was closely matched by that observed in MD simulations.

In Chapter 5, dissociation of compact monomers and dimers was observed for the highest charge states of SAP, accessed *via* supercharging. This is in contrast to the more typical CID pathway, which culminates in the ejection of a highly unfolded monomer. To investigate this further, MD was used to simulate the effects of collisional activation on

pentameric SAP using a linear temperature gradient to increase gradually the internal energy of the system. The major experimental observations, including the release of an unfolded monomer in CID of intermediate charge states, and atypical dissociation of the highest charge states were all mirrored by MD simulations. The surprising parallel of the two approaches highlights how IM-MS when combined with high temperature solvent-free simulations, can provide detailed insights into the gas-phase structures of macromolecular ions, as well as the conformational changes occurring during collisional activation and dissociation.

6.2 RESULTS

All CCS measurements were made on the RF-confining linear drift tube described previously. Unless stated otherwise, calculated CCS are scaled CCS_{PA} (Eq. 2.5).

6.2.1 Ion mobility experiments reveal collision-induced compaction

In order to access a wide range of charge states for the studied complexes, a charge reduction strategy was developed and optimised on pentameric SAP (125 kDa). Five charge states (22-26+) were observed for SAP in AA buffer (Figure 6.2A). The same protein complex in TEA buffer also had five charge states (16-20+). To maximise the width of the charge state distribution, a titration with TEA in the presence of AA, was carried out. Addition of 10 mM TEA resulted in a wide charge state distribution 17-25+ (9 charge states). Higher charge states of 26-30+ were achieved by the addition of *m*-NBA or sulfolane (*232*). In Chapter 5, the gas-phase structure for SAP at low activation energy was found to be highly similar across a wide range of charge states (18+ to 30+), including those which result from the addition of supercharging reagents. In addition, the

width and shape of the arrival time distributions recorded for each charge state of this pentameric assembly were examined. These indicated the presence of one major conformation, or one ensemble of closely related conformations, thus validating the charge manipulation approaches.

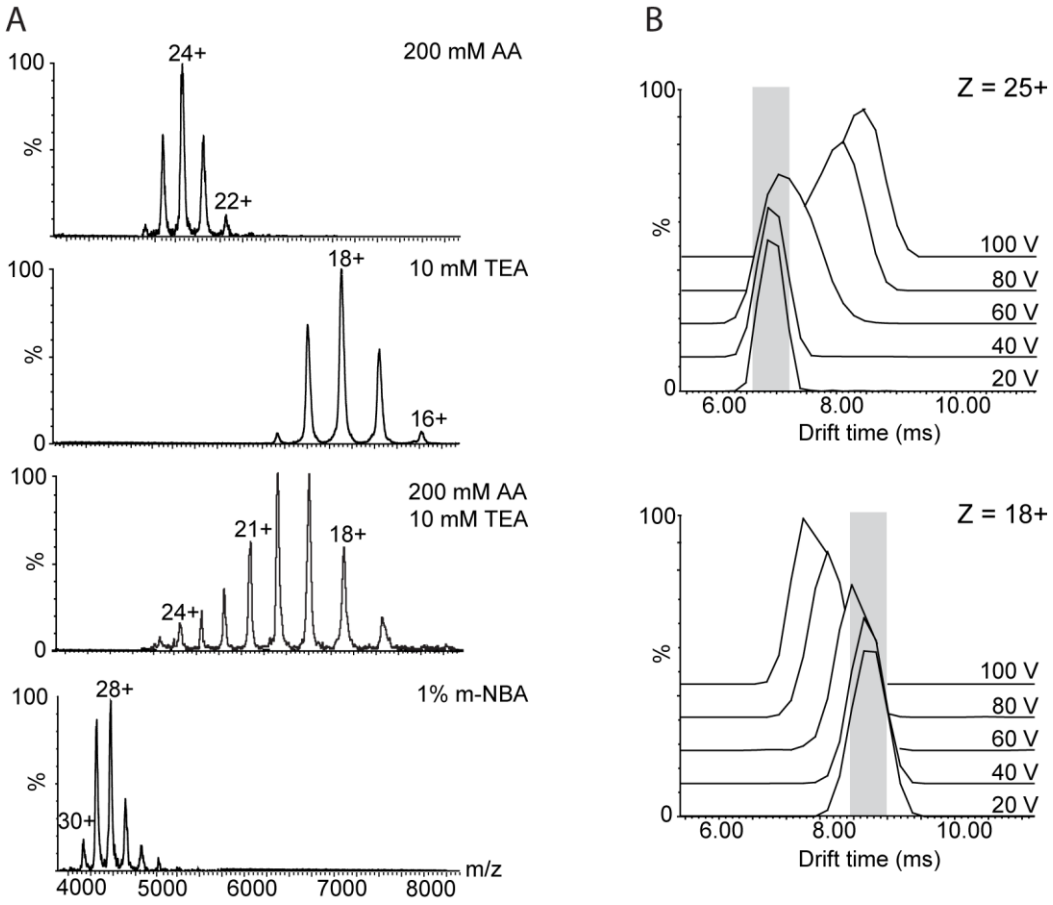


Figure 6.2. Charge manipulation and charge-dependent unfolding of SAP. **A** SAP 5-mer was electrosprayed from different buffer solutions in order to maximise the width of the charge state distribution (top three panels). MS spectra are shown for SAP in 200 mM AA, pH 7.0; 10 mM TEA buffer pH 7.0 and a mixed buffer solution of 200 mM AA with 10 mM TEA, pH 7.0. A maximum spread of 9 charge states (17-25+) was achieved for the mixed buffer solution. Charge states 26-30+ were achieved using 1% *m*-NBA (bottom panel). **B** SAP 25+ (top) experiences an increase in drift time with increasing accelerating voltage into the collision cell. In contrast, SAP 18+ (bottom) experiences a decrease in drift time with increasing accelerating voltage.

The next step was to examine the various charge states of SAP and their response to activation. By increasing the accelerating voltage into the collision cell, the amount of collisional activation can be controlled. Arrival time distributions are shown for two charge states of SAP (Z=25+ and Z=18+) (Figure 6.2B). From 20 to 100 V, the drift time for SAP 25+ increases dramatically, indicating the formation of extended conformations, consistent with collisional unfolding (*36*). Surprisingly, a decrease in drift time was observed for the SAP 18+ charge state as the accelerating voltage was increased. This significant shift in drift time indicates conformational changes consistent with compaction.

The activation energy experienced by the ion can be defined using the laboratory frame energy E_{lab}, which takes into account the charge state, Z of the ion (Eq. 6.1).

$$E_{lab} = Accelerating\ voltage \times Z \quad (6.1)$$

Plotting the CCS for SAP 18-30+, as a function of laboratory frame energy, reveals a general decrease in CCS prior to a large increase (Figure 6.3A). This can be interpreted as the formation of a more compact state than the native-like structure, prior to unfolding into more extended conformers. Interestingly the degree of compaction before unfolding was greatest for the lowest charge states. At the lowest achievable activation energy, the CCS for the 18+ ion of SAP was 6900 $Å^2$. This decreased to a minimum CCS of 6200 $Å^2$ at E_{lab} 1620 eV, corresponding to ~10 % compaction. The CCS for SAP 22+ decreased from 7020 $Å^2$ at low activation energy to a minimum of 6760 $Å^2$ (4 % compaction) at E_{lab} 1480 eV. The highest charge states (25+ to 30+) did not show any decrease but rather at E_{lab} > 1000 eV began to increase due to collisional unfolding. Overall these results demonstrate a clear compaction of the lowest charge states of the SAP pentamer.

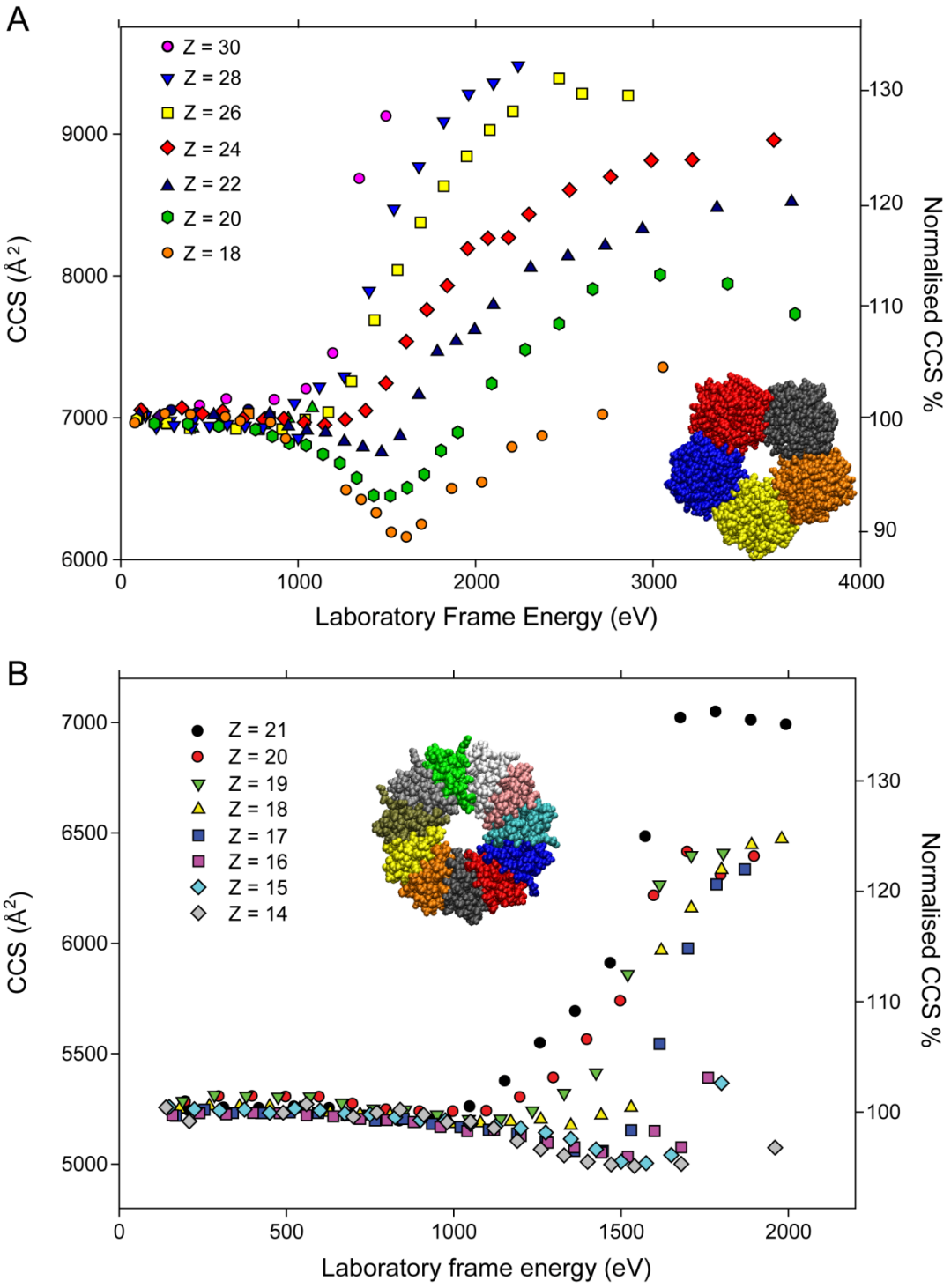


Figure 6.3. Activation profiles for cavity-containing protein complexes. The CCS of a wide range of charge states of the SAP 5-mer (**A**) and TRAP (**B**) was monitored using IM, whilst increasing the laboratory frame energy. For clarity, only even charge states of SAP are shown.

In order to determine whether this charge-state dependent compaction is specific to SAP or can be observed for other protein complexes, CCSs were measured for three additional protein complexes (TTR, TRAP and avidin). The same mixed buffer approach (AA/TEA) was used to generate a broad charge state distribution and the activation energy increased incrementally. In this way, a similar trend to SAP was obtained for TRAP (Figure 6.3B). A 4% decrease in CCS was observed prior to unfolding for the lowest charge state generated (14+). This compaction was less significant as the charge state increased from 14+ to 17+, until 19+ when unfolding proceeded with no prior compaction. Thus, similar to SAP, TRAP undergoes compaction of the lowest charge states. Interestingly the extent of compaction measured is less for TRAP than for SAP.

Turning to TTR, the average CCS for the charge states 9+ to 15+ at the lowest activating conditions was 3370 $Å^2$ with < 1% standard deviation between all charge states (Figure 6.4A). As the E_{lab} was raised, higher CCSs were observed consistent with the ions populating increasingly extended conformations. No compaction was observed for any of the charge states investigated. Whilst at low energy there was no difference between CCS of different charge states, the CCS increased with escalating charge state at higher activation energies. For example, at E_{lab} 1000 eV, the CCS for TTR 11+, 13+ and 15+ were 3820 $Å^2$, 4610 $Å^2$ and 5170 $Å^2$ respectively. The addition of 20 mM DBU, which has higher gas-phase basicity than TEA, was carried out in order to access further lower charge states for TTR. Similar to previous observations (*221*), TTR 10+ and 9+ did not experience unfolding, even under the maximum accelerating voltages accessible.

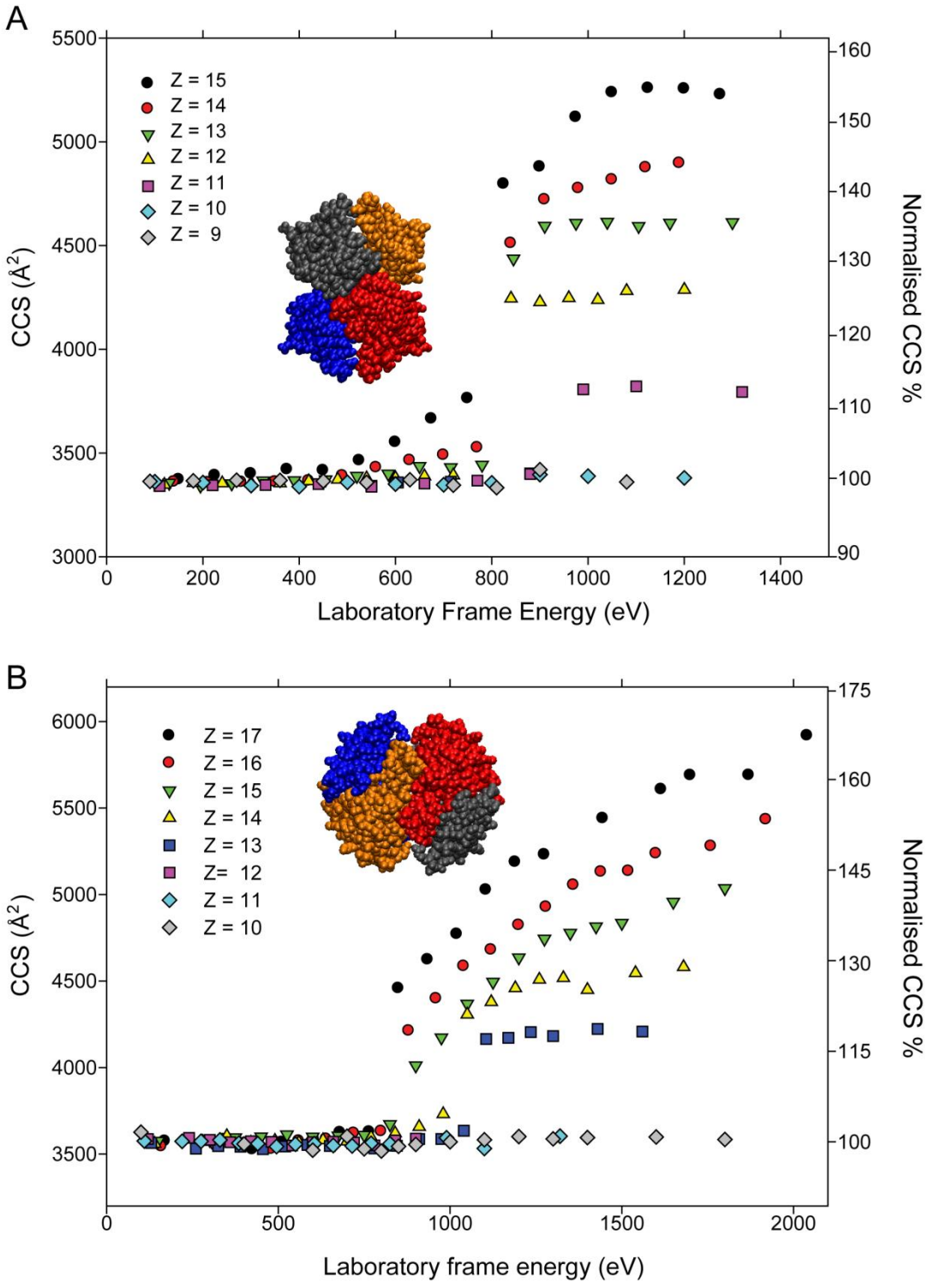


Figure 6.4. Activation profiles for protein complexes without central cavity. The CCS of a wide range of charge states for TTR (**A**) and avidin (**B**) was monitored using ion mobility, whilst increasing the laboratory frame energy.

Similar results were obtained for avidin (Figure 6.4B), which did not show evidence for a compact conformation when collisionally activated. As with TTR, the lowest charge state for avidin (10+) did not undergo unfolding or compaction even at maximum activation energy.

In order to assess whether the additive has any additional effect on the structure, the CCS of SAP 22+ was monitored, following electrospray from a variety of solution conditions. No differences in the compaction/unfolding pathways of the 22+ ions were observed in AA, AA with 10 mM TEA or AA with 10 mM imidazole. In all cases, compaction followed by unfolding was observed (Figure 6.5A). Similarly the addition of 10 mM TEA, 20 mM DBU or 10 mM imidazole had no effect on the unfolding pathway of avidin 13+ (Figure 6.5B). Thus it can be concluded that the solution additives themselves do not affect compaction/unfolding in response to activation. It is rather the charge state of the ion itself that is the critical factor here.

Furthermore, the drift times were monitored for SAP in negative ion mode during collisional activation (Figure 6.6). An initial decrease in drift time was observed, consistent with compaction, as was observed for the experiments performed in positive ion mode. This was followed by an increase in drift time at higher collision energies, consistent with unfolding. The agreement between the two ionisation modes indicate that both the conformation, and the gas-phase behaviour during collisional activation, are the same for SAP pentamer, regardless of whether it is has an excess of positive or negative charges.

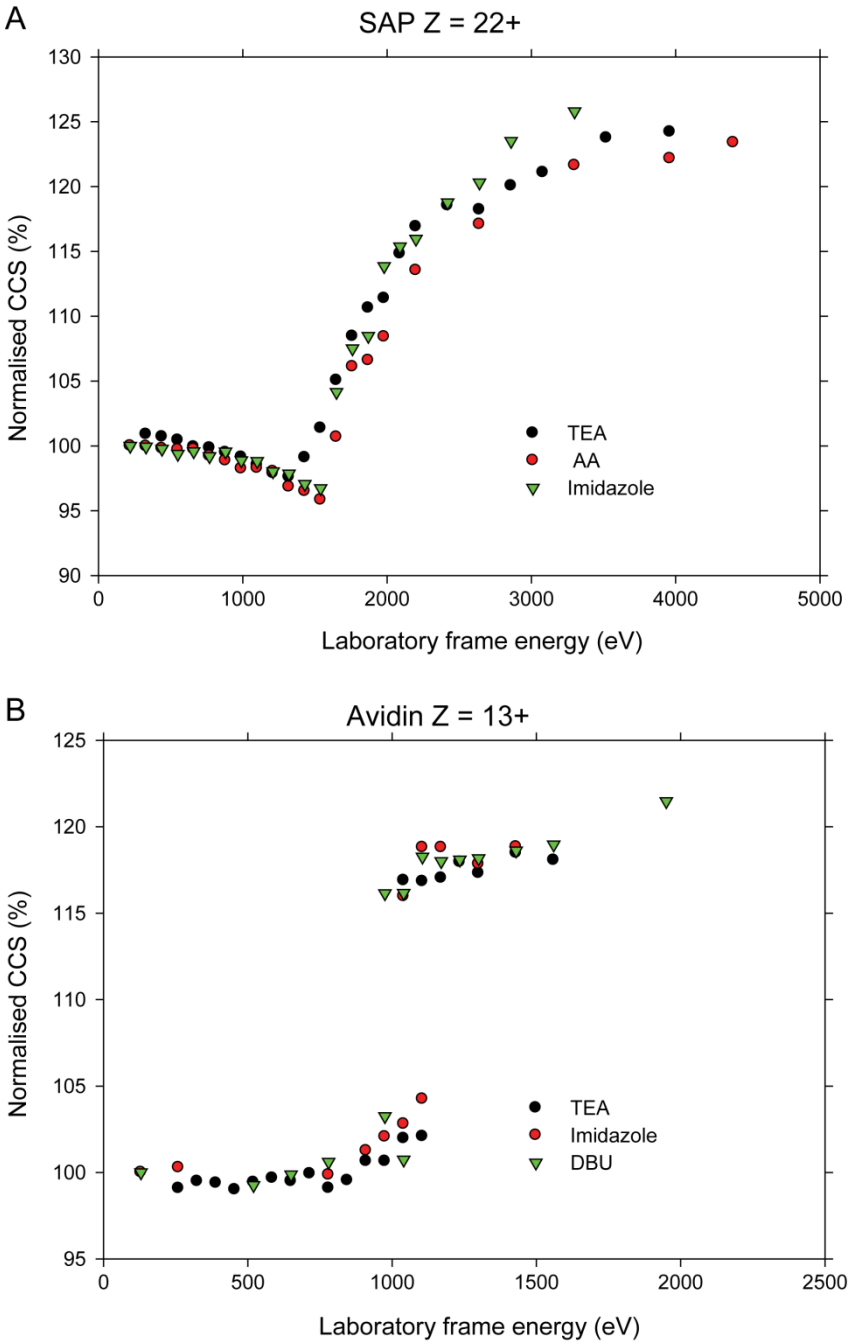


Figure 6.5. Evaluating the effect of the additive on gas-phase ion structure in response to activation. The CCS of ions from different solution conditions was monitored at increasing laboratory frame energy. **A** SAP 22+ in 200mM AA with 10mM TEA (black), 10 mM imidazole (green) and 200 mM AA only (red). **B** Avidin 13+ in 200 mM AA with 10 mM imidazole (red), 10 mM TEA (black), 20 mM DBU (green).

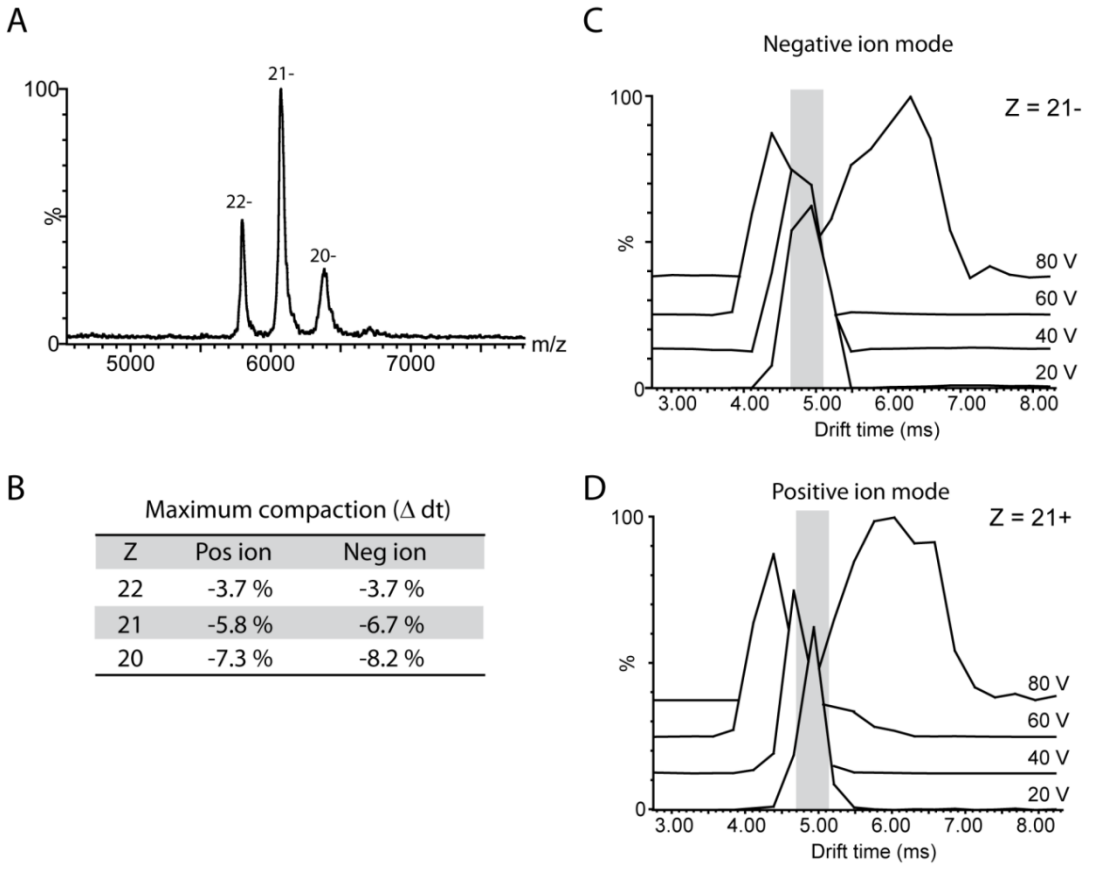


Figure 6.6. Investigating conformational changes during collisional activation of SAP in negative ion mode. The MS spectrum for SAP in negative ion mode is shown in **A**. Similar trends and extent of compaction was observed in both positive and negative ion modes (**B**). Drift time traces over increasing collision energies for SAP 21- (**C**) and SAP 21+ (**D**).

6.2.2 Molecular dynamics simulations corroborate experimental results

Although MD has been used previously to examine the structure of proteins, peptides and even small complexes, it is not clear how these methods would recapitulate macromolecular complexes in the gas phase, particularly during activation. In order to investigate this, *in vacuo* MD simulations for pentameric SAP 18+ were performed. 10 ns simulations were run over a linear temperature gradient of 300–800 K to attempt to mimic IM-MS activation. As a control, 10 ns simulations were also performed at constant temperature (300 K). The R_g for SAP 18+ was computed along the trajectories and found to be constant over the 10 ns simulation at 300 K (Figure 6.7).

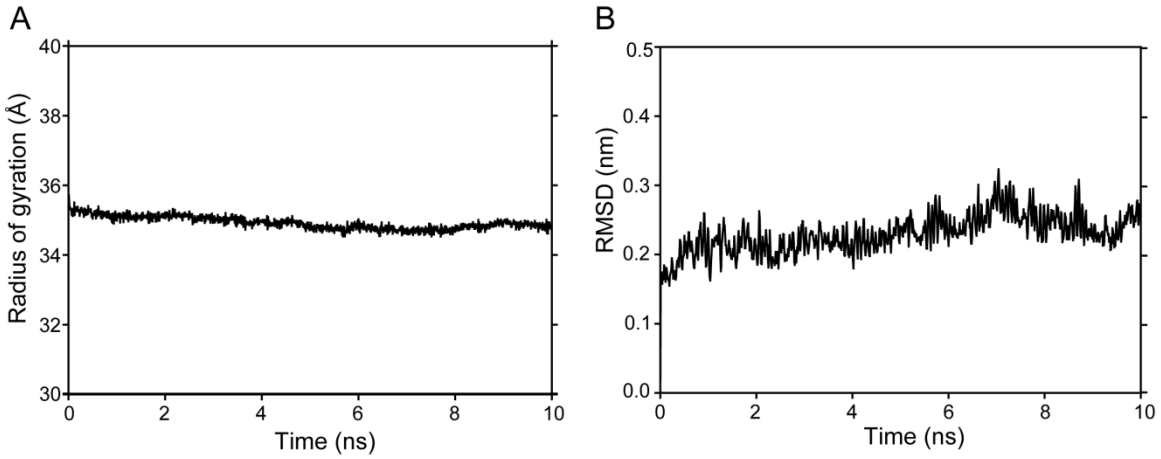


Figure 6.7. MD simulations of SAP *in vacuo* at constant temperature (300 K). **A** Radius of gyration for SAP 18+ over 10 ns simulations at constant temperature (300 K). **B** Backbone RMSD of the SAP 18+ pentamer, from the starting structure at 0 ns, over 10 ns at constant temperature (300 K).

In contrast, during the 10 ns simulation of SAP 18+ with increasing temperature (300 – 800 K) (Figure 6.8), large conformational changes were evident. From analysis of R_g the simulation was divided into three phases. In the first phase (0 – 3 ns), at lower temperatures, there is little change in R_g indicating the initial ring-like structure is retained. This is followed by a decrease in R_g (3 – 7.5 ns) of up to 15 %. The final phase of the simulation (7.5 - 10 ns) is marked by an increase in R_g, to ~10 % above its initial value. In order to observe the conformational changes to the assembly, coordinates were extracted from a representative trajectory and their CCS calculated.

RMSD-based clustering was used to identify the most populated conformations at different stages in the trajectory (Figure 6.8). Representative structures at different time points were thus compared to the structure at 0 ns, which corresponded to the initial native-like ring topology (CCS_{calc} 6800 $Å^2$). The R_g decrease was associated with the symmetrical ring topology distorting into a buckled ring (~5 ns, CCS_{calc} 6657 $Å^2$), gradually becoming more collapsed and reaching a minimum value at ~7.5 ns (CCS_{calc}

6326 Å^2). The CCS of the smallest conformation was 7 % lower than the CCS at 0 ns. The subsequent increase in R_g after 8 ns is due to considerable conformational change, consistent with the extension of one or more monomers (~10 ns, CCS_{calc} 8048 Å^2).

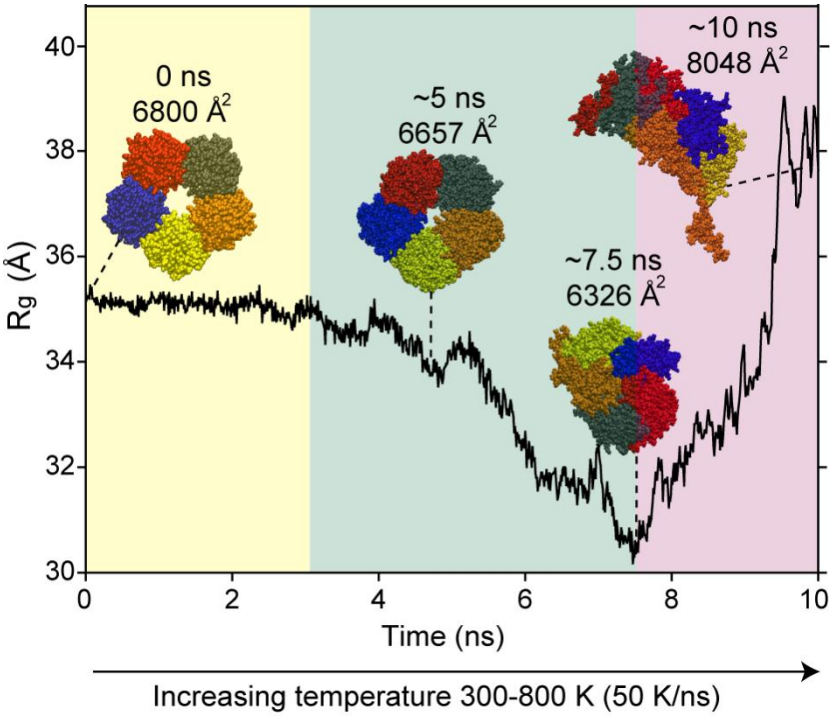


Figure 6.8. *In vacuo* molecular dynamics simulations of SAP 18+. Radius of gyration for SAP over 10 ns simulation with a linear temperature gradient (300 to 800 K). A representative trajectory is shown for SAP 18+, with charges evenly distributed over the surface-accessible basic residues. Additional trajectories are shown in Figure 6.9. RMSD-clustering was performed to identify the most populated conformational states during different stages of the trajectory. These illustrate the native-like ring topology (0 – 3 ns, yellow) becoming distorted into a buckled ring, followed by collapse of the ring (3 – 7.5 ns, green). Finally, significant unfolding, consistent with the extension of a monomer was observed (7.5 – 10 ns, purple). The given CCSs were calculated from the corresponding structures.

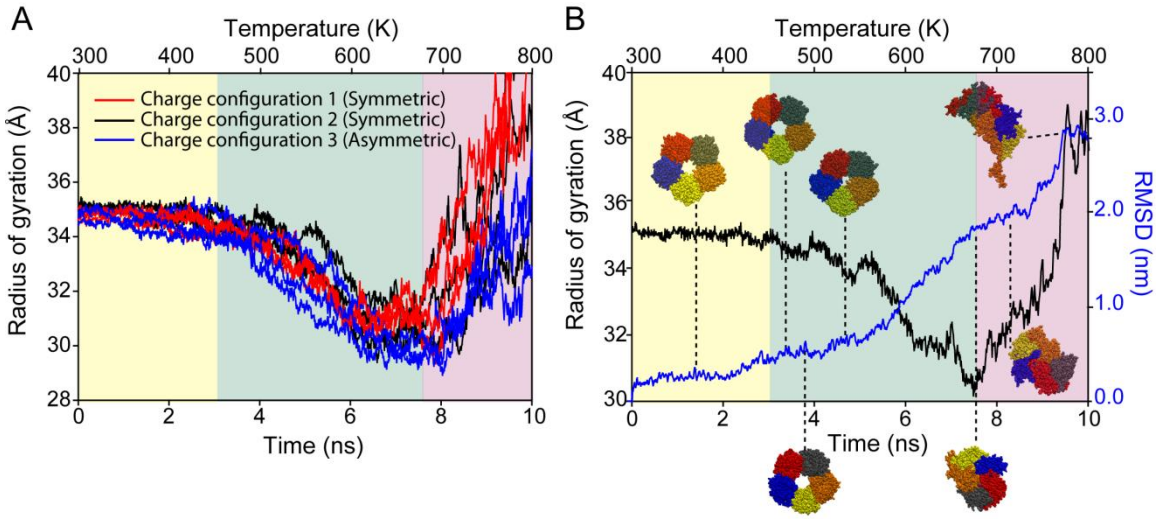


Figure 6.9. Analysis of *in vacuo* molecular dynamics simulations of SAP 18+. **A** Radius of gyration over simulations is shown for nine independent trajectories of SAP 18+. The three different charge configurations used, including one asymmetric arrangement, are shown in red, black and blue. Different regions were identified as constant R_g (~0–3 ns, yellow), decreasing R_g (~3–7.5 ns, green) and increasing R_g (~7.5–10 ns, purple). **B** The R_g and the backbone RMSD of the pentamer from the starting structure at 0 ns are shown, for a representative trajectory of SAP 18+, in black and blue respectively. RMSD clustering was performed as described; the most populated clusters in each region were identified and are shown.

6.2.3 Effect of charge state on compaction and unfolding *in silico*

Having established simulation conditions for SAP 18+, the effect of charge state on the conformational changes observed was investigated by carrying out simulations for two additional charge states of SAP (25+ and 30+). From IM-MS experiments, SAP 18+ and 25+ were selected to represent low (charge-reduced) and intermediate charge states respectively. SAP 30+ was chosen for MD simulations to reflect a supercharged state. Furthermore, in order to evaluate the importance of the specific arrangement of charges on the *in silico* unfolding pathway, two additional charge configurations, including an asymmetric arrangement, were prepared for each of the three charge states investigated.

Table 6.1. Degree of compaction observed in simulations and experiments for low, intermediate and high charge states of SAP

Charge state	MD simulations CCS_{calc} (Å^2)		MD simulations [a]ΔCCS_{calc} (%)	Experimental [b]ΔCCS_{exp} (%)
18+	Initial	6828 ± 62	-7.3± 1.3 %	-10.2 ± 0.6 %
	Minimum	6328 ± 93		
25+	Initial	6882 ± 44	-3.0 ± 1.6 %	-0.2 ± 0.2 %
	Minimum	6677 ± 118		
30+	Initial	6946 ± 45	-1.0 ± 0.5 %	N/A
	Minimum	6875 ± 60		

[a]ΔCCS_{calc} is the % difference between initial and minimum CCS for nine MD simulations with a linear temperature gradient.

[b]ΔCCS_{exp} is the % difference between initial and minimum CCS observed in three repeat IM-MS experiments with increasing activation energy.

Analysis of nine trajectories (three charge configurations with three independent replicates each) for each charge state was conducted. All nine trajectories for SAP 18+ are shown in Figure 6.9A. There was relatively low variability in the extent of compaction observed for these different charge configurations (Table 6.1). In contrast, distinct pathways could be observed when the overall charge states were different (18+, 25+, 30+) (Figure 6.10). The lowest charge state (Z = 18+) experienced the greatest decrease in CCS (ΔCCS_{calc} -7 %) before unfolding with a corresponding increase in CCS. The average CCS decrease for the intermediate charge state (Z = 25+) was 3 %. The highest charge state (Z = 30+) did not undergo any significant decrease in CCS, only a rapid increase in CCS after 4 ns. These results clearly show that despite different arrangements of the charges on the protein complexes, it is the overall charge state that is the key determinant of collapse for a protein complex with an internal cavity.

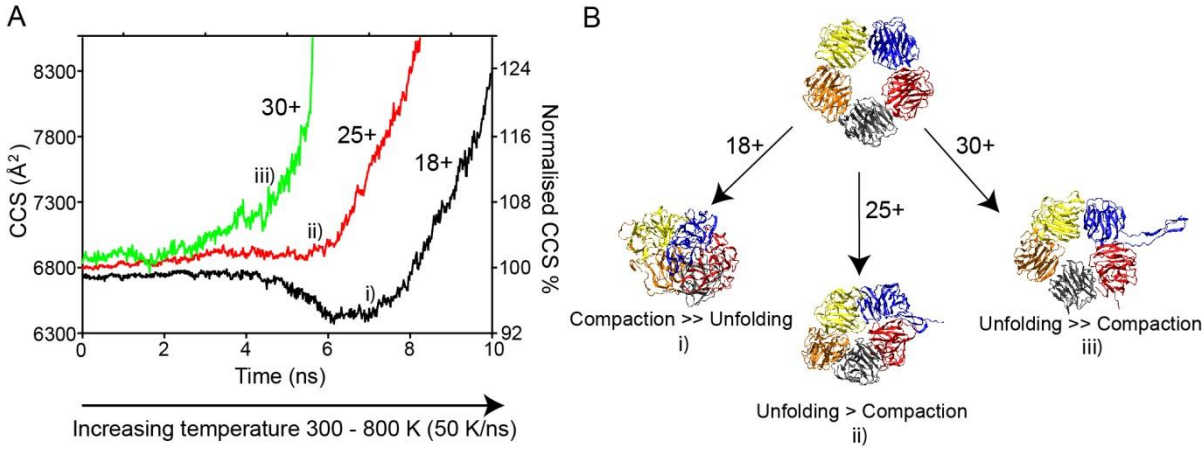


Figure 6.10. Effect of charge state on unfolding pathways *in silico*. **A** CCSs were calculated for SAP over 10 ns MD simulations with a linear temperature gradient (300 to 800 K). Average values for all trajectories are shown in green, red and blue for three different charge states of SAP: 30+, 25+, and 18+ respectively. **B** During the high temperature simulations, SAP 18+ collapsed to compact conformational states, prior to unfolding. SAP 25+ compacted to a lesser extent than 18+ with a greater degree of unfolding. SAP 30+ proceeded to unfold with no prior compaction. The structures shown are the most populated conformations (based on RMSD-clustering) in the regions (i)-(iii), from a representative trajectory.

6.2.4 Insights into collision-induced dissociation

In the MD simulations the charges are fixed, and cannot migrate to the newly-exposed surface area that is formed as a result of unfolding. Therefore in order to simulate CID *in silico*, the initial charge distribution is important. CID experiments on the SAP 5-mer 25+ ion reveal the dissociation of monomer subunits (25.5 kDa) with average charge state 11.7+. Therefore, when preparing SAP 25+ for MD simulations, 12 charges were distributed evenly over the surface-accessible basic residues of one monomer, with the remaining 13 charges distributed over the accessible basic sites of the remaining four monomers. Simulations were carried out as described previously. After ~9 ns, dissociation of a highly charged monomer was observed. The CCS of the monomer gradually increased over the course of the simulation (Figure 6.11), corresponding to

progressive loss of structure. The average CCS of the newly dissociated monomer was ~3500 $Å^2$, significantly greater than the bound monomer (CCS_{calc} 1990 $Å^2$) in the native-like structure at 0 ns. This shows close agreement to the experimental CCS of the dissociated monomer, which was measured to be 3200 $Å^2$ (12+). In contrast, MD simulations revealed that the tetramer, which remains intact after dissociation of the monomer, did not undergo significant changes in CCS, increasing just 5 % from its initial value of 5870 $Å^2$.

During CID of supercharged SAP 30+, the dissociation of monomers (25.5 kDa) and dimers (51.0 kDa) was observed. The average charge state of the monomer (~ 6+) indicated a significantly different structure to the monomer (12+) released during CID of SAP 25+ above. Experimental CCS were measured and found to be consistent with CID pathways resulting in the dissociation of compact monomers (6+, 1880 $Å^2$) and compact dimers (12+, 3760 $Å^2$), as noted in Chapter 5. This was reproduced using either *m*-NBA or sulfolane as the supercharging reagent. Interestingly, during MD simulations of SAP 30+, where each subunit carried an equal number of charges (6+), the dissociation of compact dimers (CCS_{calc} 3795 $Å^2$) and monomers (CCS_{calc} 2220 $Å^2$) was observed. This is in line with experimental observations. By following the backbone RMSD of individual subunits, relative to the starting structure at 0 ns, distinct pathways for unfolding and dissociation can be clearly seen for the 25+ and 30+ charge states of SAP (Figure 6.12) during simulations.

In contrast, MS/MS experiments on SAP 18+ (Figure 6.12B) resulted in the majority of the precursor ions remaining intact, even at the highest collision energies accessible. A very small population of dissociated monomer was observed, with charge state centred

on 9+. Thus in order to simulate CID of SAP 18+, 9 charges were distributed evenly over the surface-accessible basic residues of one monomer, with an additional 9 charges distributed over the accessible basic sites of the remaining four monomers. During the simulations, SAP 18+ underwent conformational changes consistent with compaction and unfolding, as described above. However, no dissociation was observed. The fact that all simulations of the three different charge states of the complex recapitulate the experimental observations is remarkable given the size and complexity of the macromolecular ions investigated and the limited understanding of the effects of gas-phase activation on 3D structure.

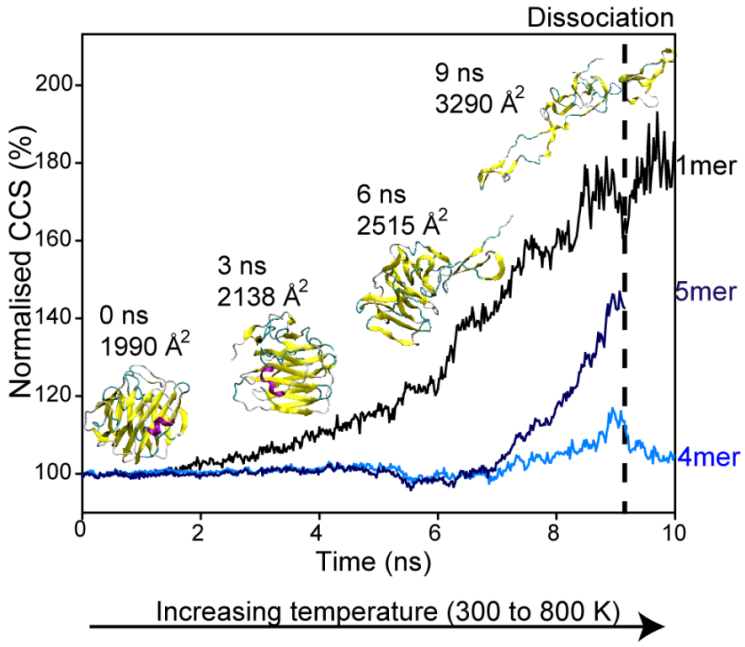


Figure 6.11. Collision-induced dissociation of SAP 25+. MD simulations of SAP 5-mer (25+), with a linear temperature gradient (300 to 800 K) were run. The precursor ion had charges asymmetrically distributed; one subunit had 12+ (as determined by MS/MS), one 4+, and the remaining three subunits each had 3+. The average CCS was monitored for the SAP monomer (12+), 4-mer (13+) and the intact pentameric complex (black, light blue, dark blue respectively). Prior to dissociation, monomer and 4-mer coordinates were extracted from the intact complex to calculate the CCS. RMSD clustering was carried out and representative structures for the monomer at different stages in one trajectory are shown at 0, 3, 6 and 9 ns, with their associated CCSs.

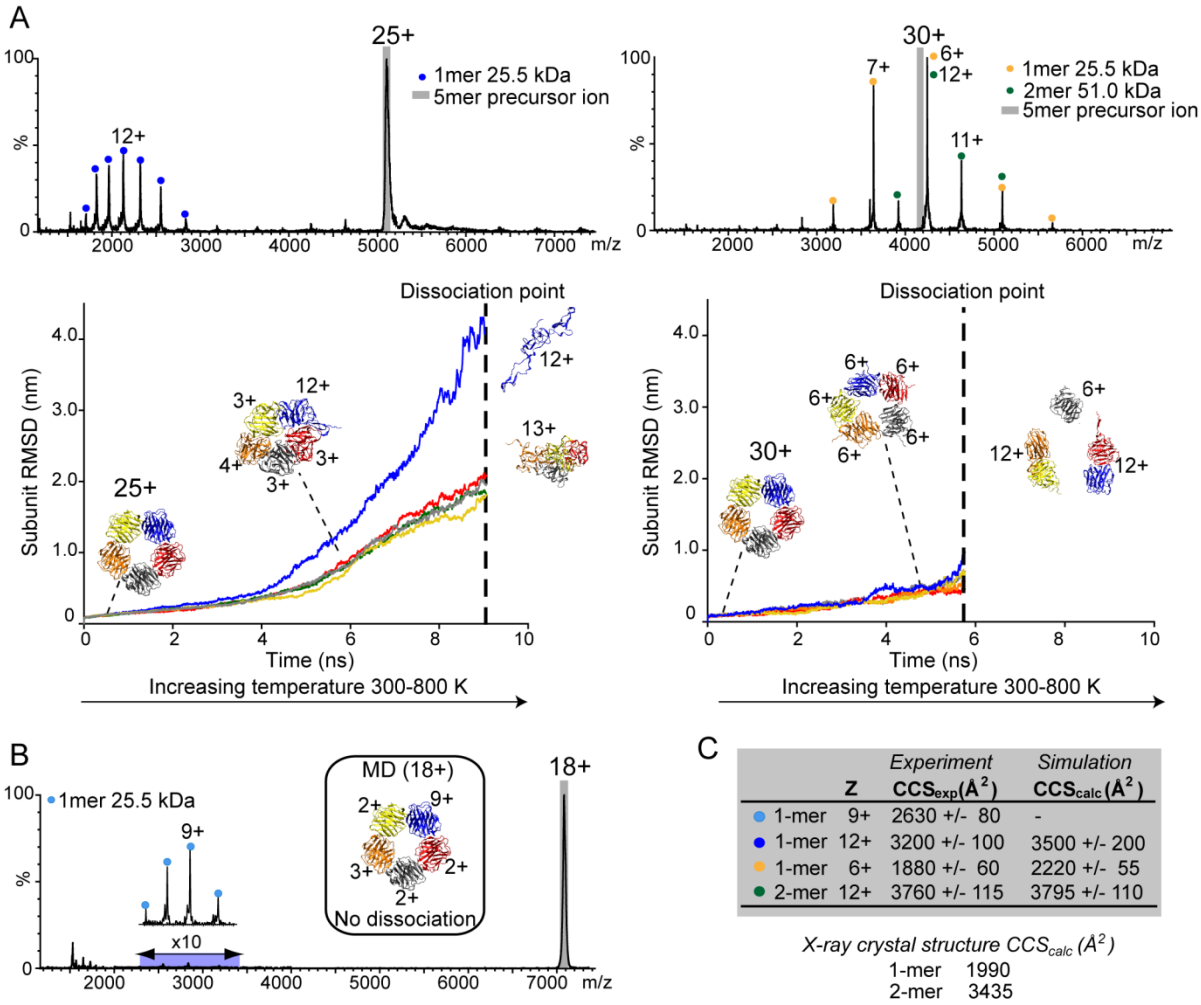


Figure 6.12. Charge-dependent dissociation pathways for SAP. **A** MS/MS spectra for SAP 25+ and 30+ are shown. CID of SAP 25+ results in loss of a highly charged monomer, whilst SAP 30+ releases dimers, and lower charged monomers. The average charge states observed for CID products were used to assign the charge distributions in MD simulations. Thus SAP 25+ had charges asymmetrically distributed, with one subunit carrying 12 charges, while the remaining 13 charges were evenly distributed over the other four subunits. For SAP 30+, each subunit had a charge state of 6+. The backbone RMSD of individual subunits locally superimposed, from the starting structure at 0 ns, was monitored during simulations, and an average for all the equivalent trajectories is shown. Representative structures are shown at different time points, for one trajectory each, during simulations of SAP 25+ and 30+. **B** MS/MS spectrum for SAP 18+, indicating very little dissociation of the precursor ion. A small population of monomer, with charge state centred around 9+ can be observed in the magnified region around *m/z* 3000. No dissociation was observed in MD simulations

of SAP 18+, with one monomer carrying 9 charges. **C** The CCS of dissociation products from MD and measured using IM-MS following CID experiments.

6.2.5 Intrinsically disordered p53 collapses in the gas phase

Homotetrameric p53 consists of folded core (C) and tetramerisation (T) domains, which are linked by intrinsically disordered regions and flanked by disordered N and C termini (Figure 6.13). Whilst the individual domains of p53 have been well characterised, the full-length protein has eluded structural biologists for decades. One of the major challenges is that large natively unfolded regions, accounting for ~ 40 % of the full-length protein, impede structural analysis by traditional methods (*118*). For this study, the effects of introducing disordered regions (flexible linker, N and C termini) to the folded core and tetramerisation domains were probed using IM-MS and MD simulations.

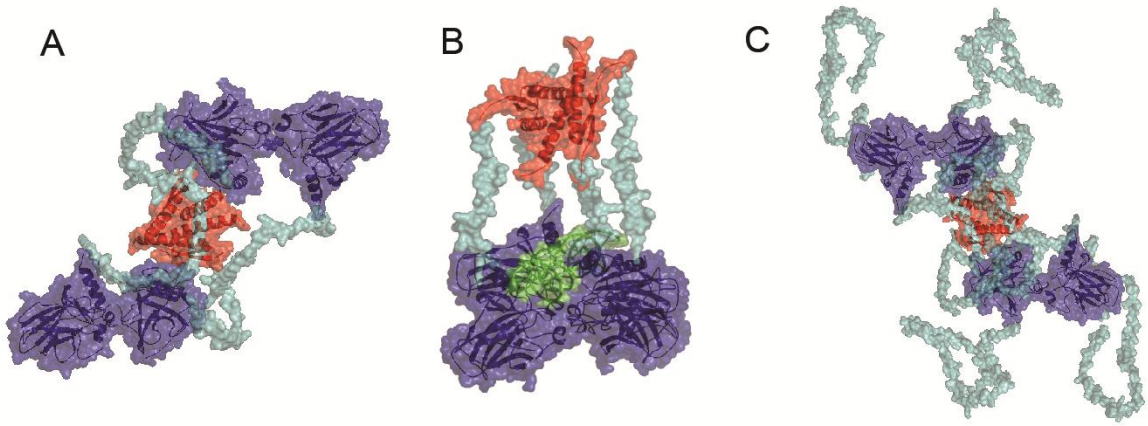


Figure 6.13. The p53 tetramer is composed of a weakly associated dimer-of-dimers. In the absence of DNA, the oligomeric state of p53 is controlled by the tetramerisation domain (T), whilst the core domain (C) is responsible for binding DNA. Upon DNA binding, p53 wraps around the double helix and the structure becomes more rigid and compact (*118*, *260*). Model structures for $(CT)_4$ (**A**), $(CT)_4$DNA (**B**), and the full-length p53 (**C**) were obtained from a previous multi-instrumental study (*118*). Colouring: red = T; blue = C; cyan = intrinsically disordered regions; green = DNA.

First, IM-MS[10] was applied to the smallest 4.1 kDa tetramerisation domain. This was found to exist exclusively as tetramers, $(T)_4$ (Figure 6.14A). The average measured CCS (1425 $Å^2$) was comparable to theoretical values (CCS_{PA} 1457 $Å^2$, CCS_{EHSS} 1862 $Å^2$) calculated from the atomic structure (Table 6.2). Next the 26.2 kDa core domain was examined. The core domain is known not to form higher oligomers in the absence of DNA (*261*). As expected, only one charge state series was identified in MS, corresponding to C monomer (Figure 6.14B). An average CCS_{exp} of 1835 $Å^2$ for this monomer is slightly below the CCS calculated using the PA algorithm (1944 $Å^2$), and in good agreement with previous IM measurements for the isolated core domain (*262*). The next step was to add DNA (Figure 6.14C) to the core domain. Three charge state series were identified, corresponding to C, $(C)_2$DNA and $(C)_4$DNA. An average CCS of 5384 $Å^2$ was measured for the 120 kDa $(C)_4$DNA complex, in close agreement with the theoretical CCS_{PA} (5377 $Å^2$).

Having established reasonable agreement between the measured and theoretical CCS of C, $(T)_4$ and $(C)_4$DNA complexes, the effect of intrinsically disordered regions in the (CT) constructs was examined. Here, core and tetramerisation domains are connected *via* a flexible linker, consisting of 14 amino acids. IM-MS of the 120 kDa tetramer $(CT)_4$ revealed three populations present in solution: CT, $(CT)_2$ and $(CT)_4$, consistent with the assembly of dimer-of-dimers (Figure 6.14D). Interestingly, the average CCS measured for $(CT)_4$ was 5515 $Å^2$, ~60 % of the theoretical value (CCS_{PA} 8022 $Å^2$, CCS_{EHSS} 10319 $Å^2$), indicating substantial collapse. Similarly, for the 137.5 kDa DNA-bound form, $(CT)_4$DNA (Figure 6.14E), an experimental CCS (6047 $Å^2$) much lower than calculated values was obtained (CCS_{PA} 8022 $Å^2$, CCS_{EHSS} 10352 $Å^2$).

[10] IM-MS experiments for p53 constructs were performed by Kevin Pagel, University of Oxford

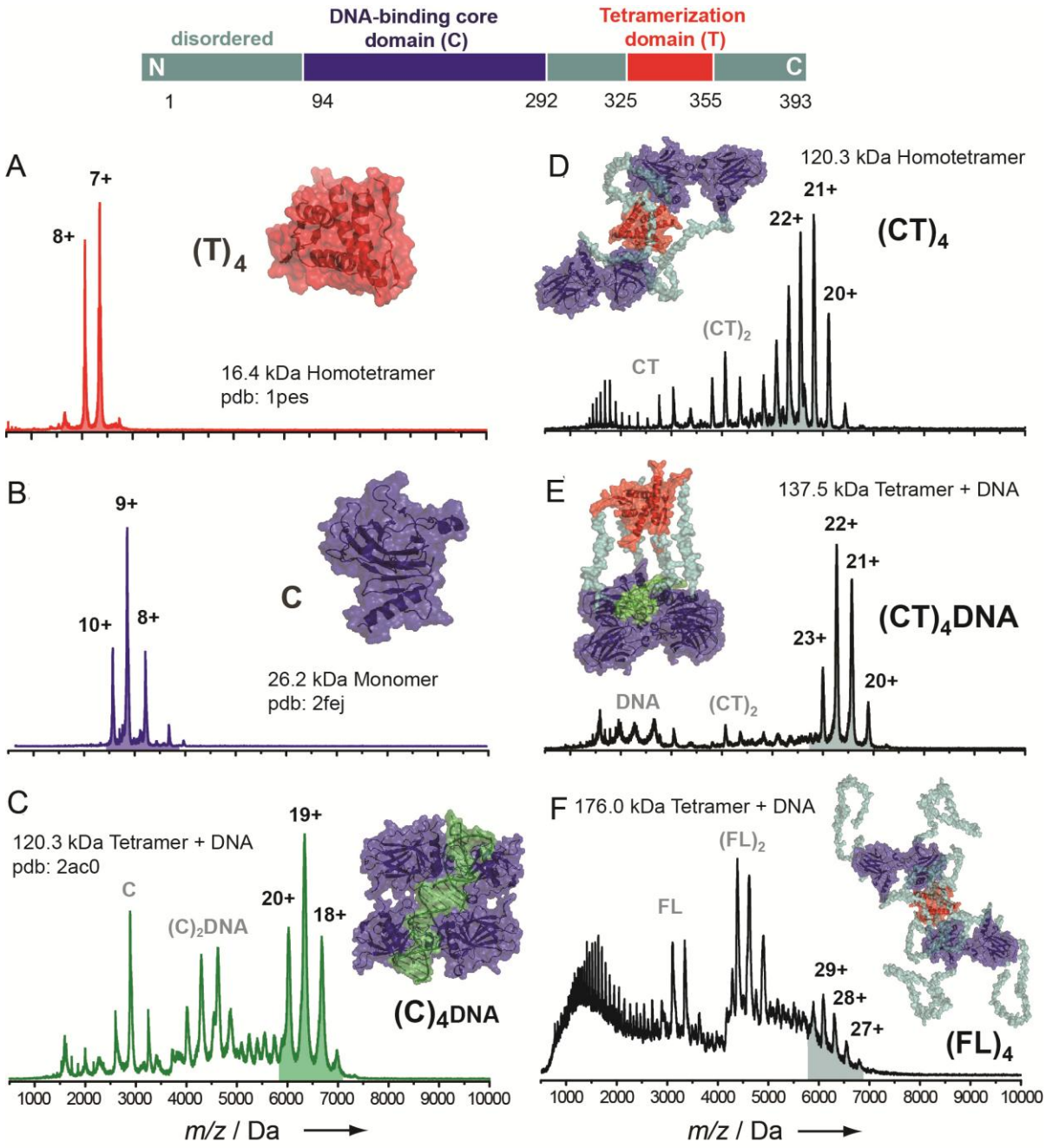


Figure 6.14. Schematic representation of p53 and mass spectra of individual domains and constructs. All spectra were recorded from AA buffered solution and at similar instrument conditions. Structural models for $(T)_4$, C and $(C)_4$DNA correspond to X-ray or NMR structures with the given PDB codes (**A**-**C**). Model structures for $(CT)_4$, $(CT)_4$DNA and FL_4 (**D**-**F**) were obtained from a previous study (*118*).

Table 6.2. Experimental and theoretical CCSs for p53 constructs and their DNA complexes. Values obtained from charge reduced (TEA) solutions are highlighted in grey.

Complex	Residues	MW (Da)	PDB	*Z*	CCS_{exp} ($Å^2$)	#CCS_{calc} ($Å^2$) PA	EHSS
$(T)_4$	325-355	16400	1PES	8	1453	1457	1862
				7	1397		
				6	1376		
				5	1353		
C	94-312	26150	2FEJ	10	1746	1944	2463
				9	1851		
				8	1907		
				7	1803		
				6	1840		
$(C)_4$**DNA**	94-312 + DNA	120100	2AC0	20	5327	5397	7098
				19	5392		
				18	5434		
$(CT)_4$	94-360	120300	Ref (*118*)	22	5449	8184	10319
				21	5536		
				20	5554		
				19	5521		
				16	5668		
				15	5709		
$(CT)_4$**DNA**	94-360 + DNA	137500	Ref (*118*)	23	6012	8022	10352
				22	5978		
				21	6092		
				20	6105		
$(FL)_4$	1-393	176000	Ref (*118*)	29	7160	15942	19288
				28	7145		
				27	7139		
				26	7107		

The scaled CCS_{PA} estimate derived from globular proteins was not used here, as intrinsically disordered proteins have structures which may behave differently to globular proteins in the gas phase. CCS_{exp} are compared instead with CCS_{PA} and CCS_{EHSS} paying greater heed to relative CCS changes, rather than the absolute values.

Finally, the full-length p53 was considered, which in addition to the flexible linker, is flanked by extended disordered N and C termini. Due to its low thermal stability (*261*), peaks corresponding to the tetrameric full-length protein $(FL)_4$ were low intensity, with a high proportion of $(FL)_2$ and FL monomer present in the solution (Figure 6.14F). IM-MS for $(FL)_4$ revealed a dramatic structural collapse, with measured CCS (7138 $Å^2$) that correspond to just ~40 % of the theoretical values (CCS_{PA} 15942 $Å^2$, CCS_{EHSS} 19288 $Å^2$).

In order to test whether a reduced charge state has an effect on the structural collapse observed here for the p53 constructs, charge reduction was carried out using TEA buffer. This allowed CCS measurement for lower charge states of $(T)_4$, C and $(CT)_4$. However, the obtained nanoelectrospray was not stable enough to enable direct CCS measurements for $(C)_4$DNA, $(CT)_4$DNA or $(FL)_4$. In all cases, the charge state appeared to have very little effect on CCS (Table 6.2), which is in contrast to the charge state-dependent collapse in response to activation described above for the SAP pentamer.

It is interesting to note that the experimental CCS for the isolated domains (C, $(T)_4$ and $(C)_4$DNA) are lower than expected from the scaled CCS_{PA} (Eq. 2.5), being in very close agreement with the unscaled PA CCS. Although these particular constructs do not contain large intrinsically disordered domains, the presence of flexible arms and loops may increase their susceptibility to gas-phase rearrangements. This highlights a potential limitation for the general application of a scaled CCS_{PA} estimate derived from a small set of globular complexes (section 3.2.4), and illustrates the need for improved approaches to comparing experimental and calculated CCS. This might include subjecting atomic coordinates to *in vacuo* energy minimisations or MD simulations prior to CCS calculation. Following this approach (Appendix, Figure A5) for C, $(T)_4$ and $(C)_4$DNA

resulted in minimised structures with average CCS up to 15 % lower (CCS_{PA} 1683, 1253 and 4631 $Å^2$ respectively)[11]. Scaling these values by the empirically determined factor 1.14 (Eq. 2.5) gives calculated CCS with improved agreement with the experimentally measured values (deviation 4.3, 0.2, and 2.0 % respectively).

On the other hand, the striking difference between experimental and theoretical CCS values for the p53 constructs containing disordered regions was unprecedented and warranted further investigation. To identify the origin of this dramatic difference, MD simulations were performed *in vacuo* using models of $(CT)_4$ and the $(CT)_4$DNA as starting points. Three replicate 1 ns simulations were carried out at constant temperature (300 K) in a solvent-free environment. Interestingly, simulations revealed the rapid collapse of $(CT)_4$ and $(CT)_4$DNA complexes to compact conformations on a sub-nanosecond time scale (~ 0.1 ns) (Figure 6.15). The most populated conformational state during each trajectory was determined using RMSD-clustering, and subsequently CCS were calculated. Good agreement was found between experimental CCS and the average values calculated for $(CT)_4$ and $(CT)_4$DNA during MD.

To test which regions experience the most extensive structural changes *in vacuo*, the backbone RMSD from corresponding starting structures was calculated over the simulation period. The RMSD of the individual core and tetramerisation domains within the $(CT)_4$ and $(CT)_4$DNA complexes was found to increase by ~3-4 Å. By contrast the full complexes exhibited a significant RMSD from their initial conformations (16-20 Å) (Figure 6.15). This value is considerably larger than the 2-3 Å RMSD observed for other proteins in solvent-free environments (*83*). These results imply that rearrangement of p53

[11] In contrast, *in vacuo* minimisation (300 K) of SAP pentamer results in CCS change of < 3 %.

constructs arises from convergence of the essentially unaltered core and tetramerisation domains *via* collapse of the flexible linkers.

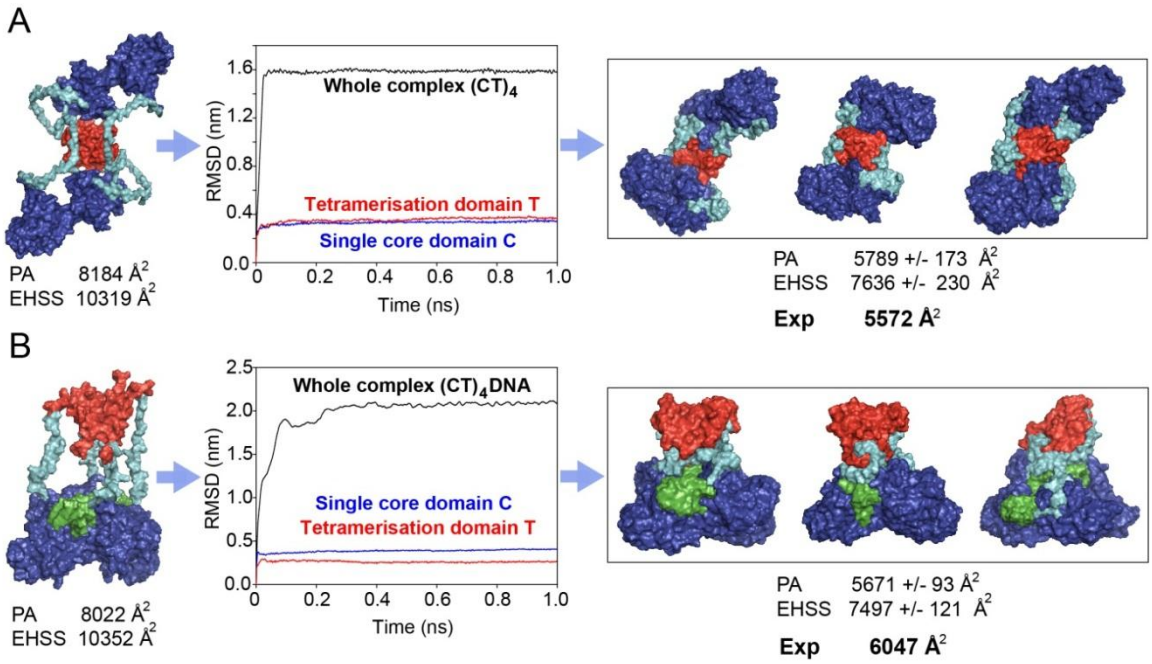


Figure 6.15. Molecular dynamics simulations of selected p53 constructs. MD simulations (1 ns) were carried out for (**A**) the $(CT)_4$ and (**B**) the $(CT)_4$DNA complexes *in vacuo*, at 300 K. Based on the $(CT)_4$ and $(CT)_4$DNA starting structures derived from previous experimental studies (*118*) (left) an RMSD-based clustering analysis was carried out in order to identify the most populated conformations within each trajectory. Representative structures for three replica trajectories are shown in each box. Average calculated CCSs agree well with experimental values.

The average gas phase density of native-like, globular proteins is ~ 0.61 g cm^{-3} enabling the CCS of an idealised spherical protein to be calculated (*92*). Experimental CCSs for p53 constructs were compared with the constant-density estimate for CCS (Figure 6.16). This revealed a clear correlation between measured values and CCSs estimated for spherical particles. Additionally, CCS for the compact MD-derived structures, which were in good agreement with the experimental values, also fitted closely to the constant-density estimate. Thus despite extensive collapse, p53 constructs exhibit typical packing densities of native-like globular proteins.

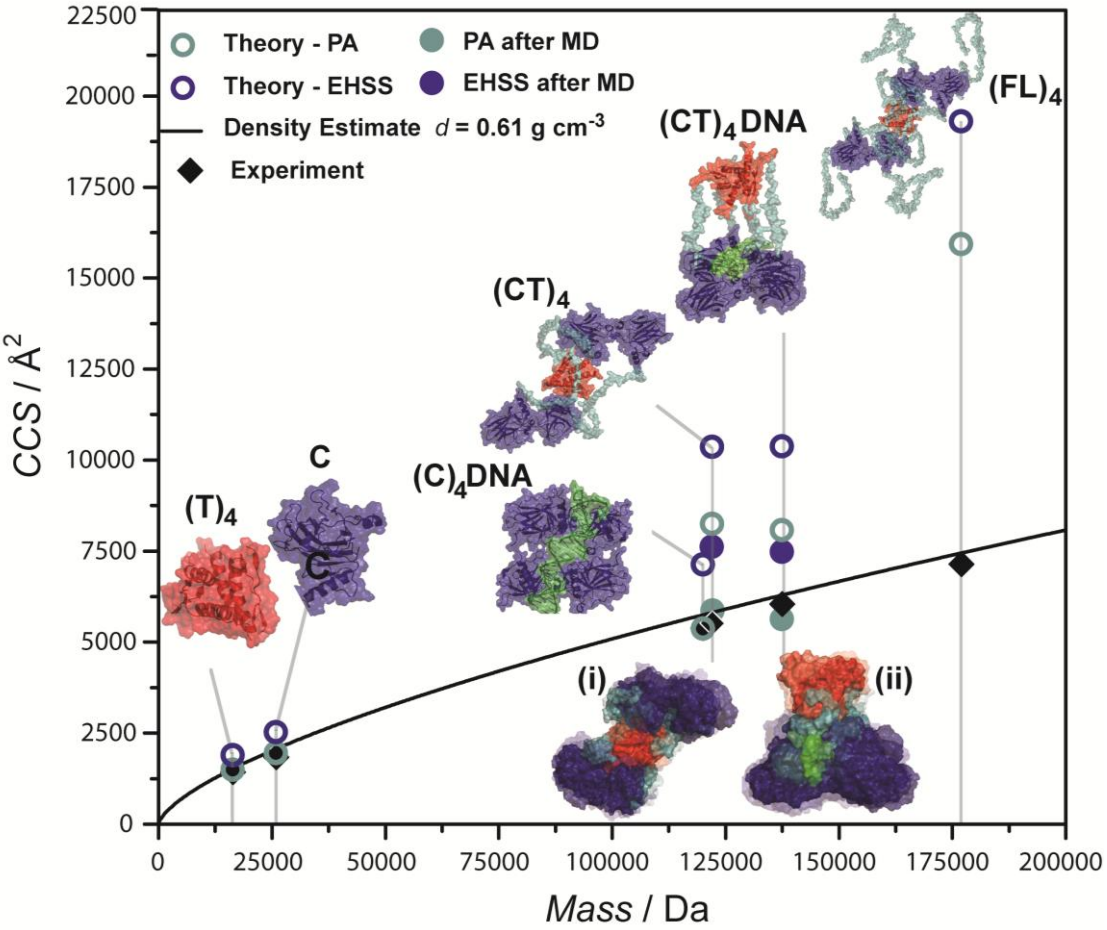


Figure 6.16. Plot of mass against CCS for p53 domains and constructs together with representative structures from MD simulations. Experimental values (black diamonds) are compared with theoretical CCS calculated using the PA (light blue open circles) or EHSS algorithm (dark blue open circles). CCSs are calculated using atomic information from PDB files (Table 6.2) or proposed models (*118*). Analogous values are calculated after MD simulation (filled circles). The estimated mass to CCS-correlation of idealised spherical particles with the average density of a native-like gas-phase protein is shown (solid line). Three of the most populated conformers formed during MD simulations of $(CT)_4$ and $(CT)_4$DNA are overlaid (i) and (ii) respectively.

Table 6.3. Charge states predicted from surface area estimates (structures from MD in grey).

Complex	SASA[a] (Å^2)	Predicted[b] Z_{av}	Experimental[c] Z_{av}
$(T)_4$	8872	9.5	7.1
C	15065	13.0	9.0
$(C)_4$DNA	39909	23.2	19.1
$(CT)_4$	62106	30.2	21.4
$(CT)_4$ (MD)	36203	21.9	21.4
$(CT)_4$DNA	65276	31.1	21.7
$(CT)_4$DNA (MD)	35568	21.6	21.7
$(FL)_4$	121133	44.8	28.8

[a] SASA is the solvent accessible surface area calculated from the atomic structure

[b] The average charge state (Z_{av}) was predicted from the empirical relationship defined in section 5.2.1

[c] Experimental charge state averages were calculated by Gaussian fitting of the charge state envelope

Finally the average charge states (from MS) of all complexes were examined. These were compared with predicted values, based on the surface area correlation (*28*) described in section 5.2.1. The folded subdomains $(T)_4$, C and $(C)_4$DNA all show good agreement to predicted values, however when constructs included disordered regions, the observed charge states were significantly lower than anticipated. Thus for $(CT)_4$, $(CT)_4$DNA and $(FL)_4$, average charge states observed in MS were 8-16 charge states lower than predicted (Table 6.3). This is in contrast to isolated, monomeric intrinsically disordered proteins, for which the charge states were shown to correlate well with surface area estimates (*263*).

It is noteworthy that the surface areas calculated for the collapsed structures $(CT)_4$ and $(CT)_4$DNA formed during MD simulations are in excellent agreement with the experimental average charge state. The loss of surface area of the native-like p53

constructs implies that collapse occurs concomitantly with charging. This is in accord with recent proposals for the electrospray mechanism in which complexes reside in the interior of droplet and solvent and electrolyte ions expel the excess surface charge of droplets. Charging of the protein takes place at the very last stages of desolvation, *via* transfer of charges onto the gas-phase complex (*15*). Consequently, fewer charges than expected from the surface area estimate are obtained for the p53 constructs with large intrinsically disordered regions, which must therefore collapse very late in the electrospray process.

6.3 DISCUSSION

One of the major limitations of MD is its inability to match the time scale over which the protein undergoes structural rearrangements (often >1 ms). The computational cost of running all-atom calculations for these time scales makes such simulations impractical. One way to overcome this is to run simulations at elevated temperatures. This approach is commonly used to study protein unfolding in bulk-solvent (*83, 258, 264*). Elevated temperature simulations, including simulated annealing and replica-exchange molecular dynamics, have also been applied to *in vacuo* studies to allow enhanced sampling (*65, 265*) and to correlate CCS data from IM-MS with candidate structures (*266*). In this chapter, a linear temperature gradient was chosen for the *in vacuo* MD simulations of globular protein complexes. This was in order to impart increasing amounts of energy to the system to mimic IM-MS activation experiments. This gradient was also chosen to enhance sampling, such that simulations could be carried out over a reasonable time scale.

A further important consideration for the simulations is the choice of charge distribution. In this chapter "symmetric" and "asymmetric" distributions are differentiated. In reality, the arrangement of charges is unlikely to be perfectly symmetrical. Local areas of slightly higher charge density are thought to play a role in triggering unfolding, which in turn prompts charge migration to maintain a uniform surface charge density (*36*). Charge migration is not yet possible in such large-scale simulations and therefore the use of fixed charges and the choice of charge distribution reflects one of the limitations of the theoretical model. Whilst it is difficult to model precisely what happens to protein complexes in the gas phase, key steps observed in IM-MS experiments were captured in MD simulations, where experimentally determined charge states were used.

i) Observation of a compact state prior to unfolding of cavity-containing complexes

The collapse of ring-like structures in the gas phase has been documented previously (*39*). In this previous study, higher charge states of TRAP were found to be more compact. However, this was carried out on a prototype instrument, using a higher energy regime than used in this study. Here, IM-MS experiments reveal that the lower charge states of SAP and TRAP passed through an intermediate compact state, prior to unfolding. Both SAP and TRAP complexes had cavities calculated to be ~7 % of their total volumes. The lowest charge states of these complexes exhibited a decrease in experimental CCS of 10 % (SAP) and 4 % (TRAP). Thus SAP experienced CCS decrease 3 % greater, and TRAP 3 % less, than expected from the size of the cavity alone.

Since the CCS decrease does not correlate solely with cavity size, an assessment of the interface packing density was carried out for the two protein complexes using the gap

volume packing index (*114*). The gap volume index corresponds to the volume of interface cavities, normalised with respect to the buried surface area (*267*). Here, an interface cavity is defined as a packing defect in protein tertiary structure, which can accommodate a probe of a defined size (*268*). Typical values for protein complexes are 2.5 ± 1.0 Å (*269*). The gap volume indices for TRAP and SAP were calculated; TRAP had a gap volume index of 0.99 Å, indicating that subunits are twice as well packed as those in SAP (gap volume index 2.32 Å). Consequently for a structure with high interface packing density, where subunit interactions are more rigid, less compaction is observed than would be anticipated for a totally collapsed structure. This could therefore account for the differences observed for these two complexes between the cavity size and degree of compaction during collisional activation.

ii) Lowest charge states experience the greatest degree of compaction

The charge-state dependence of the compaction observed for cavity-containing proteins was investigated. The ability to modulate charge state without affecting the conformation was critical for this. In order to access a wide range of charge states for the complexes, a charge reduction strategy, based on the addition of TEA to AA buffer, was employed (*106, 222*). In IM-MS experiments with activation, a relationship was observed between decreasing charge state and the extent of compaction for SAP and TRAP. This trend was mirrored by the MD simulations, where SAP 18+ showed a greater degree of compaction than equivalent simulations of SAP 25+ and 30+. It seems likely that by reducing the charge, the energy barrier to unfolding is increased and thus more energy is required in order to initiate unfolding (*252*). For low charge states, compact conformations become increasingly accessible, provided there has been sufficient energy input to overcome the barrier for collapse.

iii) Less energy is required to unfold higher charge states

Higher charge states require less energy to initiate unfolding. This is supported by both experimental observations and simulations. A clear relationship between charge state and the extent of unfolding was demonstrated by IM-MS for all four globular complexes studied, with higher charge states being able to access more highly extended conformations than lower charge states. This is in agreement with early theoretical and experimental studies, which noted that higher charge states of monomeric lysozyme and cytochrome *c* were more prone to unfolding *in vacuo (63, 150)*. Interestingly, the lowest charge states of avidin and TTR did not unfold, even at maximum activation energies accessible. One explanation for this observation is that unfolding competes with dissociation (*41, 252*). For the lowest charge states of avidin and TTR, the energy barrier for unfolding is increased, such that the energy required for dissociation is lower than that for unfolding. This is consistent with a previous study (*221*) which showed that lower charge states of TTR which had been charge-reduced using crown ethers ejected folded, rather than extended, monomers during CID.

iv) Loss of compact monomers and dimers occur from symmetrically charged complex ions

Atypical CID pathways, which do not result in monomer dissociation, have been described for several protein complexes (*38, 238-240*). The detailed mechanism by which this unusual form of dissociation occurs is not fully understood, however in Chapter 5 the number of interfacial salt bridges and overall charge density were proposed to play an important role. Dissociation of intact charge-symmetric subcomplexes has also been reported during SID (*162, 250, 270*), where dissociation occurs on a faster time scale than unfolding due to the rapid high-energy deposition. Here, atypical dissociation

resulting in ejection of compact monomers and dimers was observed during CID experiments of the highest charge states of SAP that were accessed by supercharging. Lower and intermediate charge states dissociated *via* the loss of an unfolded monomer.

Similar observations were noted in MD simulations of SAP 30+ and 25+. Interestingly, compact monomers and dimers were dissociated during simulations in which the precursor ion had a symmetric charge distribution. In contrast, an initial asymmetric charge distribution resulted in dissociation of highly charged unfolded monomers. This is consistent with theoretical calculations by Thachuk and co-workers (*41*), who propose that perfect symmetric charge partitioning in multimeric complexes results in the largest intermolecular repulsion, lowering the barrier for dissociation. Asymmetric charge partitioning, on the other hand, lowers the barrier for unfolding through increased intramolecular repulsion. It is the balance between inter- and intra- molecular Coulomb energy, which controls the dissociation path of complexes. This implies that it is the partitioning and migration of charge which is pivotal. Fixed charged simulations cannot inform on this. It is tempting however to speculate that charge migration, which occurs concomitantly with unfolding, may be inhibited in precursor ions with higher charge states, and fewer free basic residues. At some critical point, dissociation becomes energetically and/or kinetically favourable over the cooperative processes of charge migration and unfolding.

v) Intrinsically disordered proteins populate compact conformations in the gas phase

Contrary to the cavity-containing globular proteins described above, complexes containing intrinsically disordered domains collapsed in a charge-independent manner. Importantly, this collapse was spontaneous, occurring in the final stages of electrospray

prior to transfer into the vacuum. Elevated temperatures in MD simulations were not required to induce the dramatic conformational collapse, which occurred rapidly, on a sub-nanosecond time scale. In principle, all proteins or protein complexes might undergo a compaction when transferred into a solvent-free environment (*59*). This collapse is characterised by minor rearrangements of the charged residues at the protein exterior, which fold back to the protein surface to form strong electrostatic interactions upon removal of water (*71*). For large protein complexes such a compaction would produce only a subtle effect. More significant changes may be anticipated when the protein structure comprises flexible arms or loops. The magnitude of the collapse observed here however is unprecedented among protein complexes considered to date. It can be concluded that the vast majority of the gas-phase rearrangement of p53 occurs at the flexible linker and the disordered termini. Thus the $(CT)_4$ and $(FL)_4$ p53 complexes do not retain the same conformational state in the gas phase as observed in solution. The inner core, by contrast, which governs the intermolecular organisation of the complex, appears to retain its structural integrity.

6.4 SUMMARY

IM-MS experiments were carried out on globular protein complexes and those which contain intrinsically disordered domains. Whilst the four globular complexes studied maintained their native-like structures in the gas phase at lower energies, those containing intrinsically disordered regions populated collapsed conformations, which were 40-60 % smaller than predicted from their atomic structures. This collapse is a spontaneous process, independent of the charge state of the ion, or the energy regime employed. On the other hand those globular complexes which contained a large central cavity could access significantly more compact conformations during collisional

activation. The degree of compaction here was found to be strongly charge-state dependent with the lowest charge states able to access conformations with up to 10 % lower CCS than the native-like structure. Relative changes in CCS during IM-MS of p53 constructs, and during IM-MS activation experiments for SAP, were also closely reflected in MD simulations. These simulations were carried out *in vacuo*, and, where activation was required, under conditions of increasing temperature.

MD was also used to simulate CID of SAP. MS was used to measure the charge state of the ejected monomer and this charge was assigned to one subunit, with residual charges distributed evenly among the remaining subunits. Under these conditions, MD simulations captured the unfolding and ejection of a single subunit for intermediate charge states of SAP. The highest charge states recapitulated the ejection of compact monomers and dimers, which were also observed in CID experiments of high charge states of SAP accessed by supercharging. Overall this study highlights the potential of vacuum MD simulations of protein complexes, in combination with experimental data from IM-MS and CID, to provide additional insights into the processes of gas-phase collapse, unfolding and dissociation.

The work described in this chapter has been adapted with permission from Hall et al. (2012), Journal of the American Chemical Society, 134, 3429-3438, © 2012 American Chemical Society doi:10.1021/ja2096859 and from Pagel et al. (2013), Angewandte Chemie International Edition, 52, 361-365, © 2013 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim doi:10.1002/anie.201203047.

7 CONCLUSIONS AND FUTURE DIRECTIONS

One of the challenges in structural biology is to obtain three-dimensional information of increasingly complex heterogeneous assemblies. In this thesis, I have developed IM-MS based strategies for the structural characterisation of multimeric protein complexes. This relies on accurate CCS measurements, a validated modelling approach, and sensible interpretation of CCS data. Further structural information can be attained from disassembling the complex in solution or the gas-phase. Therefore the effect of solution disruption on structure, and the role of charge state on the gas-phase dissociation pathways of multimeric complexes, has been systematically investigated. Finally, the gas-phase structures of globular and intrinsically disordered protein complexes were probed by IM-MS and MD simulations, revealing new insights into the stabilities of these important macromolecules in the absence of solvent.

A database of buffered protein and protein complex CCS has been completed that has enabled significant improvements to current approaches for the CCS determination of large protein assemblies using TWIMS (Chapter 3). In addition to the database, a calibration framework was suggested, in which small calibration errors are readily achieved by closely bracketing the analyte of interest with IM standards of similar mass and mobility. Interestingly, this study revealed deterioration of the calibration at higher wave velocities, resulting in inaccurate CCS measurements. This *m/z* dependent effect is not fully understood and the nature of ion transport through the travelling-wave mobility separator would be a good target for future investigations. A comparison of helium CCS from IM with those calculated from atomic structures revealed a close correlation. This suggests that proteins retain “memory” of their native structures in the gas phase. The

improvement to CCS measurement accuracy, and the ability to relate measurements made in the gas phase to protein complex structures in solution, provides the basis for using IM-MS to elucidate biomolecular assembly structures and is a subject of continuing investigation.

An assessment of IM-MS for building 3D models of protein complexes was presented (Chapter 4). Experimental CCS of heteromeric complexes and their subcomplexes, which were formed from solution disruption, were compared with calculated values for CCS. Generally, disruption in solution did not affect the overall conformation, validating the use of subcomplex CCS restraints for assembling models of the intact complex. The presumed gas-phase collapse of a large cavity-containing complex was observed however, highlighting one of the limitations of IM-MS for structural characterisation. A coarse-grained modelling approach, based on spheres, was developed and validated. CCS-derived restraints from IM-MS were incorporated into a scoring scheme, and successfully used to predict model structures in a benchmark of homomeric and heteromeric protein complexes. This approach was then applied to predict the structures of early oligomeric intermediates in ΔN6-β2m aggregation. Overall, IM-MS is therefore shown to have good potential for structure determination, particularly when integrated with data/information from other sources, such as symmetry from EM.

Exciting future directions in this field include the incorporation of additional sources of structural information into the scoring function, to improve the predictive value of this method. Such complementary techniques include HDX, cross-linking MS, and EM. By increasing the information content of the scoring function, this modelling approach

would be tested on more challenging biological systems, including larger and more heterogeneous assemblies. Further investigation into the gas-phase structures of complexes, including potential conformational changes and the stabilisation of cavity-containing proteins to prevent gas-phase collapse will also be critical to the successful adoption of IM-MS as an established method for structural biology. The ΔN6–β2m oligomers would also make excellent targets for future studies. These studies might utilise a cross-linking and CCS-restrained docking strategy, with molecular dynamics refinement, to propose higher resolution structural models and thus provide detailed insights into the aggregation process.

A systematic investigation was carried out on the effects of manipulating charge states in solution on the conformation and gas-phase dissociation of multimeric protein complexes (Chapter 5). Charge reduction did not affect structure, whilst supercharging often resulted in unfolding, particularly at higher reagent concentrations. However increases in charge state were achieved in some cases without any CCS increase, implying that other factors than unfolding are at least partly responsible for the supercharging phenomenon. Interestingly several supercharged species dissociated atypically, ejecting compact monomers and/or dimers. A correlation was explored between the high charge density on the precursor ion, and a low number of interfacial salt bridges, with the ability to access atypical CID routes.

The ability to manipulate the dissociation pathway, by altering the charge state is an intriguing prospect. The potential for unfolding of the intact complex however during ESI, particularly in the presence of supercharging reagents, is one serious limitation that

has already been proposed (*234*). Increasing the charge state, without unfolding is therefore a major objective. One attractive option could be the use of cross-linking strategies to hinder unfolding, as used in early investigations into the origin of the asymmetric charge partitioning in CID (*42*). Further development of alternative supercharging strategies, such as ion-ion reactions or cation adduction (*271*) which do not unfold complexes would also be of major significance. Similarly the discovery of supercharging reagents that operate at lower concentrations, and are less prone to promote unfolding during ESI could be advantageous. Overall an improved understanding of the dissociation of multimeric complexes is required, particularly with respect to their charge-state dependence.

Using a combination of IM-MS experiments and MD simulations, the gaseous structures of protein complexes were probed, including some of the most challenging targets for structural biology (Chapter 6). The striking gas-phase collapse of flexible regions in intrinsically disordered p53 was discovered, with a surprising correlation found between IM-MS and MD simulations. Similarly, through simulations and experiments the collapse of the ring-topology for pentameric SAP was unequivocally observed as the energy of the system is increased. The implication here is that at lower energies the native-like topology is maintained: a vital requirement for the incorporation of IM-MS in current and future structural biology approaches. MD simulations of complexes were combined with experimental data from CID to provide additional insights into the processes of collision-induced unfolding and gas-phase dissociation. Whilst excellent agreement between experiments and simulations were obtained, the model has several limitations. This includes the short time scale of the simulation, the lack of proton migration, and the use of temperature to simulate increased energy from collisions. The

incorporation of charge migration, and collisions with gas molecules is theoretically possible, but is currently too computationally intensive for such large systems. Other improvements could include the use of a coarse-grained force field, which would allow longer simulation times. Importantly, the interpretation of IM-MS data is substantially enhanced by the use of *in vacuo* MD simulations prior to CCS calculations. With increasing computational power, this could become routine in the future, in order to relate gas-phase experimental measurements to solution structures.

In summary, this thesis makes a contribution to improving the measurement and interpretation of protein complex CCS derived from IM. The use of CCSs as restraints for generating structural models of biomolecular assemblies is examined, and a computational approach to exploit such data developed. The potential for generating subcomplexes in the gas phase with charge manipulation has been highlighted and will hopefully stimulate further research that could benefit gas-phase structural biology. Finally, this thesis has contributed significantly to the discussion on gas-phase protein structure and the application of IM-MS to protein complexes, and has opened new avenues for future research.

REFERENCES

1. Hilton, G. R., and Benesch, J. L. (2012) Two decades of studying non-covalent biomolecular assemblies by means of electrospray ionization mass spectrometry, *J. R. Soc. Interface* **9**, 801-816.
2. McKay, A. R., Ruotolo, B. T., Ilag, L. L., and Robinson, C. V. (2006) Mass measurements of increased accuracy resolve heterogeneous populations of intact ribosomes, *J. Am. Chem. Soc.* **128**, 11433-11442.
3. Uetrecht, C., Barbu, I. M., Shoemaker, G. K., van Duijn, E., and Heck, A. J. R. (2011) Interrogating viral capsid assembly with ion mobility-mass spectrometry, *Nat. Chem.* **3**, 126-132.
4. Robinson, C. V., Sali, A., and Baumeister, W. (2007) The molecular sociology of the cell, *Nature* **450**, 973-981.
5. Hernández, H., and Robinson, C. V. (2007) Determining the stoichiometry and interactions of macromolecular assemblies from mass spectrometry, *Nat. Protoc.* **2**, 715-726.
6. Sharon, M., Mao, H. B., Erba, E. B., Stephens, E., Zheng, N., and Robinson, C. V. (2009) Symmetrical modularity of the COP9 signalosome complex suggests its multifunctionality, *Structure* **17**, 31-40.
7. Dole, M., Mack, L. L., Hines, R. L., Mobley, L. D., Ferguson, L. D., and Alice, M. B. (1968) Molecular beams of macroions, *J. Chem. Phys.* **49**, 2240-2249.
8. Fenn, J. B., Mann, M., Kai Meng, C., Fu Wong, S., and Whitehouse, C. M. (1989) Electrospray ionisation for mass spectrometry of large biomolecules, *Science* **246**, 64-70.
9. Kebarle, P., and Verkerk, U. H. (2009) Electrospray: From ions in solution to ions in the gas phase, what we know now, *Mass Spectrom. Rev.* **28**, 898-917.
10. Benesch, J. L., Ruotolo, B. T., Simmons, D. A., and Robinson, C. V. (2007) Protein complexes in the gas phase: Technology for structural genomics and proteomics, *Chem. Rev.* **107**, 3544-3567.
11. Rayleigh, L. (1882) On the equilibrium of liquid conducting masses charged with electricity, *Philos. Mag. S. 5* **14**, 184-186.
12. Iribarne, J. V., and Thomson, B. A. (1976) On the evaporation of small ions from charged droplets, *J. Chem. Phys.* **64**, 2287-2294.

13. Fernandez de la Mora, J. (2000) Electrospray ionization of large multiply charged species proceeds via Dole's charged residue mechanism, *Anal. Chim. Acta* **406**, 93-104.
14. Grandori, R. (2003) Origin of the conformation dependence of protein charge-state distributions in electrospray ionization mass spectrometry, *J. Mass Spectrom.* **38**, 11-15.
15. Hogan Jr., C. J., Carroll, J. A., Rohrs, H. W., Biswas, P., and Gross, M. L. (2009) Combined charged residue-field emission model of macromolecular electrospray ionization, *Anal. Chem.* **81**, 369-377.
16. Hogan Jr., C. J., Carroll, J. A., Rohrs, H. W., Biswas, P., and Gross, M. L. (2008) Charge carrier field emission determines the number of charges on native state proteins in electrospray ionization, *J. Am. Chem. Soc.* **130**, 6926-6927.
17. Wilm, M. S., and Mann, M. (1994) Electrospray and Taylor-cone theory, Dole's beam of macromolecules at last?, *Int. J. Mass Spectrom.* **136**, 167-180.
18. Wilm, M., and Mann, M. (1996) Analytical properties of the nanoelectrospray ion source, *Anal. Chem.* **68**, 1-8.
19. Karas, M., Bahr, U., and Dülcks, T. (2000) Nano-electrospray ionization mass spectrometry: Addressing analytical problems beyond routine, *Fres. J. Anal. Chem.* **366**, 669-676.
20. Lane, L. A., Ruotolo, B. T., Robinson, C. V., Favrin, G., and Benesch, J. L. (2009) A Monte Carlo approach for assessing the specificity of protein oligomers observed in nano-electrospray mass spectra, *Int. J. Mass Spectrom.* **283**, 169-177.
21. Dobo, A., and Kaltashov, I. A. (2001) Detection of multiple protein conformational ensembles in solution via deconvolution of charge-state distributions in ESI MS, *Anal. Chem.* **73**, 4763-4773.
22. Mann, M., Meng, C. K., and Fenn, J. B. (1989) Interpreting mass spectra of multiply charged ions, *Anal. Chem.* **61**, 1702-1708.
23. Mirza, U. A., and Chait, B. T. (1994) Effects of anions on the positive ion electrospray ionization mass spectra of peptides and proteins, *Anal. Chem.* **66**, 2898-2904.
24. Iavarone, A. T., Jurchen, J. C., and Williams, E. R. (2000) Effects of solvent on the maximum charge state distribution of protein ions produced by electrospray ionization, *J. Am. Soc. Mass Spectrom.* **11**, 976-985.

25. Sterling, H. J., Cassou, C. A., Susa, A. C., and Williams, E. R. (2012) Electrothermal supercharging of proteins in native electrospray ionization, *Anal. Chem.* **84**, 3795-3801.
26. Li, Y., and Cole, R. B. (2003) Shifts in peptide and protein charge state distributions with varying spray tip orifice diameter in nanoelectrospray fourier transform ion cyclotron resonance mass spectrometry, *Anal. Chem.* **75**, 5739-5746.
27. Mirza, U. A., Cohen, S. L., and Chait, B. T. (1993) Heat-induced conformational changes in proteins studied by electrospray ionization mass spectrometry, *Anal. Chem.* **65**, 1-6.
28. Kaltashov, I. A., and Mohimen, A. (2005) Estimates of protein surface areas in solution by electrospray ionisation mass spectrometry, *Anal. Chem.* **77**, 5370-5379.
29. Sobott, F., Hernández, H., McCammon, M. G., Tito, M. A., and Robinson, C. V. (2002) A tandem mass spectrometer for improved transmission and analysis of large macromolecular assemblies, *Anal. Chem.* **74**, 1402-1407.
30. Hofstadler, S. A., Bakhtiar, R., and Smith, R. D. (1996) Electrospray ionization mass spectroscopy: Part I. Instrumentation and spectral interpretation, *J. Chem. Edu.* **73**, A82.
31. Dawson, P. H. (1986) Quadrupole mass analyzers: Performance, design and some recent applications, *Mass Spectrom. Rev.* **5**, 1-37.
32. Cotter, R. J. (1992) Time-of-flight mass spectrometry for the structural analysis of biological molecules, *Anal. Chem.* **64**, 1027A-1039A.
33. Mamyrin, B. A. (2001) Time-of-flight mass spectrometry (concepts, achievements, and prospects), *Int. J. Mass Spectrom.* **206**, 251-266.
34. Chernushevich, I. V., Loboda, A. V., and Thomson, B. A. (2001) An introduction to quadrupole–time-of-flight mass spectrometry, *J. Mass Spectrom.* **36**, 849-865.
35. Wysocki, V. H., Jones, C. M., Galhena, A. S., and Blackwell, A. E. (2008) Surface-induced dissociation shows potential to be more informative than collision-induced dissociation for structural studies of large systems, *J. Am. Soc. Mass Spectrom.* **19**, 903-913.
36. Benesch, J. L. P. (2009) Collisional activation of protein complexes: Picking up the pieces, *J. Am. Soc. Mass Spectrom.* **20**, 341-348.
37. Breuker, K., and McLafferty, F. W. (2003) Native electron capture dissociation for the structural characterization of noncovalent interactions in native cytochrome c, *Angew. Chem. Int. Ed.* **42**, 4900-4904.

38. Felitsyn, N., Kitova, E. N., and Klassen, J. S. (2001) Thermal decomposition of a gaseous multiprotein complex studied by blackbody infrared radiative dissociation. Investigating the origin of the asymmetric dissociation behavior, *Anal. Chem.* **73**, 4647-4661.
39. Ruotolo, B. T., Giles, K., Campuzano, I., Sandercock, A. M., Bateman, R. H., and Robinson, C. V. (2005) Evidence for macromolecular protein rings in the absence of bulk water, *Science* **310**, 1658-1661.
40. Ruotolo, B. T., Hyung, S. J., Robinson, P. M., Giles, K., Bateman, R. H., and Robinson, C. V. (2007) Ion mobility-mass spectrometry reveals long-lived, unfolded intermediates in the dissociation of protein complexes, *Angew. Chem. Int. Ed.* **46**, 8001-8004.
41. Wanasundara, S. N., and Thachuk, M. (2007) Theoretical investigations of the dissociation of charged protein complexes in the gas phase, *J. Am. Soc. Mass Spectrom.* **18**, 2242-2253.
42. Jurchen, J. C., and Williams, E. R. (2003) Origin of asymmetric charge partitioning in the dissociation of gas-phase protein homodimers, *J. Am. Chem. Soc.* **125**, 2817-2826.
43. Benesch, J. L., Aquilina, A., Ruotolo, B. T., Sobott, F., and Robinson, C. V. (2006) Tandem mass spectrometry reveals the quaternary organisation of marcomolecular assemblies, *Chem. Biol.* **13**, 597-605.
44. Benesch, J. L., and Robinson, C. V. (2006) Mass spectrometry of macromolecular assemblies: Preservation and dissociation, *Curr. Op. Struct. Biol.* **16**, 245-251.
45. Taverner, T., Hernández, H., Sharon, M., Routolo, B. T., Matak-Vinkovic, D., Devos, D., Russell, R. B., and Robinson, C. V. (2008) Subunit architecture of intact protein complexes from mass spectrometry and homology modeling, *Acc. Chem. Res.* **41**, 617-627.
46. Böhrer, B. C., Mererbloom, S. I., Koeniger, S. L., Hilderbrand, A. E., and Clemmer, D. E. (2008) Biomolecule analysis by ion mobility spectrometry, *Annu. Rev. Anal. Chem.* **1**, 293-327.
47. Revercomb, H. E., and Mason, E. A. (1975) Theory of plasma chromatography/gaseous electrophoresis - A review, *Anal. Chem.* **47**, 970-983.
48. Mason, E. A., and McDaniel, E. W. (1998) *Transport properties of ions in gases*, John Wiley & Sons, Inc., New York, NY.

49. Verbeck, G. F., Ruotolo, B. T., Gillig, K. J., and Russel, D. (2004) Resolution equations for high-field ion mobility, *J. Am. Soc. Mass Spectrom.* **15**, 1320-1324.
50. Shvartsburg, A. A., and Smith, R. D. (2008) Fundamentals of traveling wave ion mobility spectrometry, *Anal. Chem.* **80**, 9689-9699.
51. Giles, K., Pringle, S. D., Worthington, K. R., Little, D., Wildgoose, J. L., and Bateman, R. H. (2004) Applications of a travelling wave-based radio-frequency-only stacked ring ion guide, *Rap. Commun. Mass Spectrom.* **18**, 2401-2414.
52. 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/ToF instrument, *Int. J. Mass Spectrom.* **261**, 1-12.
53. Politis, A., Park, A. Y., Hyung, S.-J., Barsky, D., Ruotolo, B. T., and Robinson, C. V. (2010) Integrating ion mobility mass spectrometry with molecular modelling to determine the architecture of multiprotein complexes, *PLoS ONE* **5**, e12080.
54. Clemmer, D. E., and Jarrold, M. F. (1997) Ion mobility measurements and their applications to cluster and biomolecules, *J. Mass Spectrom.* **32**, 577-592.
55. Mack, E. (1925) Average cross-sectional areas of molecules by gaseous diffusion methods, *J. Am. Chem. Soc.* **47**, 2468-2482.
56. von Helden, G., Wyttenbach, T., and Bowers, M. T. (1995) Inclusion of a MALDI ion source in the ion chromatography technique: Conformational information on polymer and biomolecular ions, *Int. J. Mass Spectrom.* **146–147**, 349-364.
57. Shvartsburg, A. A., and Jarrold, M. F. (1996) An exact hard-spheres scattering model for the mobilities of polyatomic ions, *Chem. Phys. Lett.* **261**, 86-91.
58. Mesleh, M. F., Hunter, J. M., Shvartsburg, A. A., Schatz, G. C., and Jarrold, M. F. (1996) Structural information from ion mobility measurements: Effects of the long-range potential, *J. Phys. Chem.* **100**, 16082-16086.
59. Jurneczko, E., and Barran, P. E. (2010) How useful is ion mobility mass spectrometry for structural biology? The relationship between protein crystal structures and their collision cross sections in the gas phase, *Analyst* **136**, 20-28.
60. Bleiholder, C., Wyttenbach, T., and Bowers, M. T. (2011) A novel projection approximation algorithm for the fast and accurate computation of molecular collision cross sections (I). Method, *Int. J. Mass Spectrom.* **308**, 1-10.

61. Sun, N., Soya, N., Kitova, E. N., and Klassen, J. S. (2010) Nonspecific interactions between proteins and charged biomolecules in electrospray ionization mass spectrometry, *J. Am. Soc. Mass Spectrom.* **21**, 472-481.
62. Steinberg, M. Z., Breuker, K., Elber, R., and Gerber, R. B. (2007) The dynamics of water evaporation from partially solvated cytochrome c in the gas phase, *Phys. Chem. Chem. Phys.* **9**, 4690-4697.
63. Reimann, C. T., Valizquez, I., and Tapia, O. (1998) Proteins in vacuo. Denaturation of highly-charged lysozyme studied by molecular dynamics simulations, *J. Phys. Chem. B* **102**, 9344-9352.
64. Baumjetner, A., Bernstein, S. L., Wyttenbach, T., Bitan, G., Teplow, D. B., Bowers, M. T., and Shea, J.-E. (2006) Amyloid β-protein monomer structure: A computational and experimental study, *Protein Sci.* **15**, 420-428.
65. Chen, L., Shao, Q., Gao, Y.-Q., and Russell, D. H. (2011) Molecular dynamics and ion mobility spectrometry study of model beta-hairpin peptide trpzip 1, *J. Phys. Chem. A* **115**, 4427-4435.
66. Marklund, E., Larsson, D. S. D., van der Spoel, D., Patriksson, A., and Caleman, C. (2009) Structural stability of electrosprayed proteins: Temperature and hydration effects, *Phys. Chem. Chem. Phys.* **11**, 8069-8078.
67. van der Spoel, D., Marklund, E., Larsson, D. S. D., and Caleman, C. (2011) Proteins, lipids and water in the gas phase, *Macromol. Biosci.* **11**, 50-59.
68. Katta, V., and Chait, B. T. (1991) Observation of the heme-globin complex in native myoglobin by electrospray-ionization mass spectrometry, *J. Am. Chem. Soc.* **113**, 8534-8535.
69. Breuker, K., and McLafferty, F. W. (2005) The thermal unfolding of native cytochrome c in the transition from solution to gas phase probed by native electron capture dissociation, *Angew. Chem. Int. Ed.* **44**, 4911-4914.
70. Steinberg, M. Z., Elber, R., McLafferty, F. W., Gerber, R. B., and Breuker, K. (2008) Early structural evolution of native cytochrome c after solvent removal, *Chem. Bio. Chem.* **9**, 2417-2423.
71. Breuker, K., and McLafferty, F. W. (2008) Stepwise evolution of protein native structure with electrospray into the gas phase, 10^{-12} to 10^{2} s, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 18145-18152.

72. Breuker, K., Bruschweiler, S., and Tollinger, M. (2011) Electrostatic stabilization of a native protein structure in the gas phase, *Angew. Chem. Int. Ed.* **50**, 873-877.
73. Wood, T. D., Chorush, R. A., Wampler, F. M., 3rd, Little, D. P., O'Connor, P. B., and McLafferty, F. W. (1995) Gas-phase folding and unfolding of cytochrome c cations, *Proc. Natl. Acad. Sci. U.S.A.* **92**, 2451-2454.
74. Valentine, S. J., and Clemmer, D. E. (2002) Temperature-dependent H/D exchange of compact and elongated cytochrome c ions in the gas phase, *J. Am. Soc. Mass Spectrom.* **13**, 506-517.
75. Wyttenbach, T., and Bowers, M. T. (2011) Structural stability from solution to the gas phase: Native solution structure of ubiquitin survives analysis in a solvent-free ion mobility-mass spectrometry environment, *J. Phys. Chem. B* **115**, 12266-12275.
76. Skinner, O. S., McLafferty, F. W., and Breuker, K. (2012) How ubiquitin unfolds after transfer into the gas phase, *J. Am. Soc. Mass Spectrom.* **23**, 1011-1014.
77. Robinson, C. V., Chung, E. W., Kragelund, B. B., Knudsen, J., Aplin, R. T., Poulsen, F. M., and Dobson, C. M. (1996) Probing the nature of noncovalent interactions by mass spectrometry. A study of protein-coA ligand binding and assembly, *J. Am. Chem. Soc.* **118**, 8646-8653.
78. Wu, Q.-Y., Gao, J.-M., Joseph-McCarthy, D., Sigal, G. B., Bruce, E., Whitesides, G. M., and Smith, R. D. (1997) Carbonic anhydrase-inhibitor binding: From solution to the gas phase, *J. Am. Chem. Soc.* **119**, 1157-1158.
79. Liu, L., Michelsen, K., Kitova, E. N., Schnier, P. D., and Klassen, J. S. (2010) Evidence that water can reduce the kinetic stability of protein-hydrophobic ligand interactions, *J. Am. Chem. Soc.* **132**, 17658-17660.
80. Barrera, N. P., Di Bartolo, N., Booth, P. J., and Robinson, C. V. (2008) Micelles protect membrane complexes from solution to vacuum, *Science* **321**, 243-246.
81. Barrera, N. P., Isaacson, S. C., Zhou, M., Bavro, V. N., Welch, A., Schaedler, T. A., Seeger, M. A., Miguel, R. N., Korkhov, V. M., van Veen, H. W., Venter, H., Walmsley, A. R., Tate, C. G., and Robinson, C. V. (2009) Mass spectrometry of membrane transporters reveals subunit stoichiometry and interactions, *Nat. Methods* **6**, 585-587.
82. Zhou, M., Morgner, N., Barrera, N. P., Politis, A., Isaacson, S. C., Matak-Vinkovic, D., Murata, T., Bernal, R. A., Stock, D., and Robinson, C. V. (2011) Mass spectrometry of intact V-type ATPases reveals lipid binding and the effects of nucleotide binding *Science* **344**, 380-385.

83. Friemann, R., Larsson, D. S. D., Wang, Y., and van der Spoel, D. (2009) Molecular dynamics simulations of a membrane protein-micelle complex in vacuo, *J. Am. Chem. Soc.* **131**, 16606-16607.

84. van Duijn, E., Barendregt, A., Synowsky, S., Versluis, C., and Heck, A. J. R. (2009) Chaperonin complexes monitored by ion mobility mass spectrometry, *J. Am. Chem. Soc.* **131**, 1452-1459.

85. van Duijn, E., Simmons, D. A., van den Heuvel, R. H. H., Bakkes, P. J., van Heerikhuizen, H., Heeren, R. M. A., Robinson, C. V., van der Vies, S. M., and Heck, A. J. R. (2006) Tandem mass spectrometry of intact GroEL-substrate complexes reveals substrate-specific conformational changes in the trans ring, *J. Am. Chem. Soc.* **128**, 4694-4702.

86. Hogan Jr., C. J., Ruotolo, B. T., Robinson, C. V., and Fernandez de la Mora, J. (2011) Tandem differential mobility analysis-mass spectrometry reveals partial gas-phase collapse of the GroEL complex, *J. Phys. Chem. B* **115**, 3614-3621.

87. Wang, S., Politis, A., Di Bartolo, N., Bavro, V. N., Tucker, S. T., Booth, P. J., Barrera, N. P., and Robinson, C. V. (2010) Ion mobility mass spectrometry of two tetrameric membrane protein complexes reveals compact structures and differences in stability and packing, *J. Am. Chem. Soc.* **132**, 15468-15470.

88. Siuzdak, G., Bothner, B., Yeager, M., Brugidou, C., Fauquet, C. M., Hoey, K., and Chang, C. M. (1996) Mass spectrometry and viral analysis, *Chem. Biol.* **3**, 45-48.

89. Ouyang, Z., Takts, Z., Blake, T. A., Gologan, B., Guymon, A. J., Wiseman, J. M., Oliver, J. C., Davisson, V. J., and Cooks, R. G. (2003) Preparing protein microarrays by soft-landing of mass-selected ions, *Science* **301**, 1351-1354.

90. Kemper, P. R., Dupuis, N. F., and Bowers, M. T. (2009) A new, higher resolution, ion mobility mass spectrometer, *Int. J. Mass Spectrom.* **287**, 46-57.

91. Valentine, S. J., Counterman, A. E., and Clemmer, D. E. (1999) A database of 660 peptide ion cross sections: Use of intrinsic size parameters for bona fide predictions of cross sections, *J. Am. Soc. Mass Spectrom.* **10**, 1188-1211.

92. Ruotolo, B., Benesch, J. L., Sandercock, A. M., Hyung, S.-J., and Robinson, C. V. (2008) Ion mobility-mass spectrometry analysis of large protein complexes, *Nat. Protoc.* **3**, 1139-1152.

93. Giles, K., Williams, J. P., and Campuzano, I. (2011) Enhancements in travelling wave ion mobility resolution, *Rap. Commun. Mass Spectrom.* **25**, 1559-1566.

94. Merenbloom, S. I., Flick, T. G., and Williams, E. R. (2012) How hot are your ions in TWAVE ion mobility spectrometry?, *J. Am. Soc. Mass Spectrom.* **23**, 553-562.
95. Schneider, T. R., Gerhardt, E., Lee, M., Liang, P.-H., Anderson, K. S., and Schlichting, I. (1998) Loop closure and intersubunit communication in tryptophan synthase, *Biochemistry* **37**, 5394-5406.
96. Bilder, P., Lightle, S., Bainbridge, G., Ohren, J., Finzel, B., Sun, F., Holley, S., Al-Kassim, L., Spessard, C., Melnick, M., Newcomer, M., and Waldrop, G. L. (2006) The structure of the carboxyltransferase component of acetyl-coA carboxylase reveals a zinc-binding motif unique to the bacterial enzyme, *Biochemistry* **45**, 1712-1722.
97. Lessner, D. J., Johnson, G. R., Parales, R. E., Spain, J. C., and Gibson, D. T. (2002) Molecular characterization and substrate specificity of nitrobenzene dioxygenase from comamonas sp. Strain js765, *Appl. Environ. Microbiol.* **68**, 634-641.
98. Parales, R. E., Huang, R., Yu, C. L., Parales, J. V., Lee, F. K., Lessner, D. J., Ivkovic-Jensen, M. M., Liu, W., Friemann, R., Ramaswamy, S., and Gibson, D. T. (2005) Purification, characterization, and crystallization of the components of the nitrobenzene and 2-nitrotoluene dioxygenase enzyme systems, *Appl. Environ. Microbiol.* **71**, 3806-3814.
99. McCormick, M. S., Sazinsky, M. H., Condon, K. L., and Lippard, S. J. (2006) X-ray crystal structures of manganese(II)-reconstituted and native toluene/o-xylene monooxygenase hydroxylase reveal rotamer shifts in conserved residues and an enhanced view of the protein interior, *J. Am. Chem. Soc.* **128**, 15108-15110.
100. Mareya, S. M., and Raushel, F. M. (1994) A molecular wedge for triggering the amidotransferase activity of carbamoyl phosphate synthetase, *Biochemistry* **33**, 2945-2950.
101. Domanska, K., Vanderhaegen, S., Srinivasan, V., Pardon, E., Dupeux, F., Marquez, J. A., Giorgetti, S., Stoppini, M., Wyns, L., Bellotti, V., and Steyaert, J. (2011) Atomic structure of a nanobody-trapped domain-swapped dimer of an amyloidogenic beta2-microglobulin variant, *Proc. Natl. Acad. Sci. U.S.A.* **108**, 1314-1319.
102. Nikolova, P. V., Henckel, J., Lane, D. P., and Fersht, A. R. (1998) Semirational design of active tumor suppressor p53 DNA binding domain with enhanced stability, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 14675-14680.

103. Joerger, A. C., Allen, M. D., and Fersht, A. R. (2004) Crystal structure of a superstable mutant of human p53 core domain. Insights into the mechanism of rescuing oncogenic mutations, *J. Biol. Chem.* **279**, 1291-1296.
104. Arbely, E., Natan, E., Brandt, T., Allen, M. D., Veprintsev, D. B., Robinson, C. V., Chin, J. W., Joerger, A. C., and Fersht, A. R. (2011) Acetylation of lysine 120 of p53 endows DNA-binding specificity at effective physiological salt concentration, *Proc. Natl. Acad. Sci. U.S.A.* **108**, 8251-8256.
105. Natan, E., and Joerger, A. C. (2012) Structure and kinetic stability of the p63 tetramerization domain, *J. Mol. Biol.* **415**, 503-513.
106. Catalina, M. I., van den Heuvel, R. H. H., van Duijn, E., and Heck, A. J. R. (2005) Decharging of globular proteins and protein complexes in electrospray, *Chem. Eur. J.* **11**, 960-968.
107. Sterling, H. J., Prell, J. S., Cassou, C. A., and Williams, E. R. (2011) Protein conformation and supercharging with DMSO from aqueous solution, *J. Am. Soc. Mass Spectrom.* **22**, 1178-1186.
108. Sanner, M. F., Olson, A. J., and Spehner, J. C. (1996) Reduced surface: An efficient way to compute molecular surfaces, *Biopolymers* **38**, 305-320.
109. Krissinel, E., and Henrick, K. (2007) Inference of macromolecular assemblies from crystalline state, *J. Mol. Biol.* **372**, 774-797.
110. Sonavane, S., and Chakrabarti, P. (2008) Cavities and atomic packing in protein structures and interfaces, *PloS Comput. Biol.* **4**, 1-15.
111. Voss, N. R., and Gerstein, M. (2010) 3V:Cavity, channel and cleft volume calculator and extractor, *Nucleic Acids Res.* **38**, W555-W562.
112. Till, M. S., and Ullmann, G. M. (2010) McVol - a program for calculating protein volumes and identifiying cavities by a Monte Carlo algorithm *J. Mol. Model.* **16**, 419-429.
113. Yu, J., Zhou, Y., Tanaka, I., and Yao, M. (2010) Roll: A new algorithm for the detection of protein pockets and cavities with a rolling probe sphere, *Bioinformatics* **26**, 46-52.
114. Laskowski, R. A. (1995) Surfnet: A program for visualizing molecular surfaces, cavities, and intermolecular interactions, *J. Mol. Graph.* **13**, 307-308.
115. Humphrey, W., Dalke, A., and Schulten, K. (1996) VMD: Visual molecular dynamics, *J. Mol. Graph.* **14**, 33-38.

116. Benesch, J. L. P., and Ruotolo, B. T. (2011) Mass spectrometry: Come of age for structural and dynamical biology, *Curr. Op. Struct. Biol.* **21**, 641-649.
117. Gotts, N. G., von Helden, G., and Bowers, M. T. (1995) Carbon cluster anions: Structure and growth from C_5 to C_{62}, *Int. J. Mass Spectrom.* **149-150**, 217-229.
118. Tidow, H., Melero, R., Mylonas, E., Freund, S. M., Grossmann, J. G., Carazo, J. M., Svergun, D. I., Valle, M., and Fersht, A. R. (2007) Quaternary structures of tumor suppressor p53 and a specific p53 DNA complex, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 12324-12329.
119. Russel, D., Lasker, K., Webb, B., Velazquez-Muriel, J., Tjioe, E., Schneidman-Duhovny, D., Peterson, B., and Sali, A. (2012) Putting the pieces together: Integrative modeling platform software for structure determination of macromolecular assemblies, *PLoS Biol.* **10**, e1001244.
120. Ballester, P. J., and Richards, W. G. (2007) Ultrafast shape recognition to search compound databases for similar molecular shapes, *J. Comput. Chem.* **28**, 1711-1723.
121. Zyrianov, Y. (2005) Distribution-based descriptors of the molecular shape, *J. Chem. Inf. Model.* **45**, 657-672.
122. Ballester, P. J., Westwood, I., Laurieri, N., Sim, E., and Richards, W. G. (2010) Prospective virtual screening with ultrafast shape recognition: The identification of novel inhibitors of arylamine N-acetyltransferases, *J. R. Soc. Interface* **7**, 335-342.
123. Alber, F., Förster, F., Korkin, D., Topf, M., and Sali, A. (2008) Integrating diverse data for structure determination of macromolecular assemblies, *Annu. Rev. Biochem.* **77**, 443-477.
124. Johnson, S. (1967) Hierarchical clustering schemes, *Psychometrika* **32**, 241-254.
125. van der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A. E., and Berendsen, H. J. C. (2005) Gromacs: Fast, flexible, and free, *J. Comput. Chem.* **26**, 1701-1718.
126. Kaminski, G. A., Friesner, R. A., Tirado-Rives, J., and Jorgensen, W. L. (2001) Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides, *J. Phys. Chem. B* **105**, 6474-6487.
127. Hess, B., Bekker, H., Berendsen, H. J. C., and Fraaije, J. G. E. M. (1997) LINCS: A linear constraint solver for molecular simulations, *J. Comput. Chem.* **18**, 1463-1472.

128. Miteva, M., Demirev, P. A., and Karshikoff, A. D. (1997) Multiply-protonated protein ions in the gas phase: Calculation of the electrostatic interactions between charged sites, *J. Phys. Chem. B* **101**, 9645.
129. Schnier, P. D., Gross, D. S., and Williams, E. R. (1995) On the maximum charge state and proton transfer reactivity of peptide and protein ions formed by electrospray ionization, *J. Am. Soc. Mass Spectrom.* **6**, 1086-1097.
130. Kaltashov, I., and Abzalimov, R. (2008) Do ionic charges in ESI MS provide useful information on macromolecular structure?, *J. Am. Soc. Mass Spectrom.* **19**, 1239-1246.
131. Chakravarty, S., and Varadarajan, R. (1999) Residue depth: A novel parameter for the analysis of protein structure and stability, *Structure* **7**, 723-732.
132. Wang, J., Cieplak, P., and Kollman, P. A. (2000) How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules?, *J. Comput. Chem.* **21**, 1049-1074.
133. Hart, P. E., Nilsson, N. J., and Raphael, B. (1968) A formal basis for the heuristic determination of minimum cost paths, *IEEE Trans. Sys. Sci. Cyb.* **4**, 100-107.
134. Daura, X., Gademann, K., Jaun, B., Seebach, D., van Gunsteren, W. F., and Mark, A. E. (1999) Peptide folding: When simulation meets experiment, *Angew. Chem. Int. Ed.* **38**, 236-240.
135. Wyttenbach, T., and Bowers, M. T. (2007) Intermolecular interactions in biomolecular systems examined by mass spectrometry, *Annu. Rev. Phys. Chem.* **58**, 511-533.
136. Uetrecht, C., Rose, R. J., van Duijn, E., Lorenzen, K., and Heck, A. J. R. (2010) Ion mobility mass spectrometry of proteins and protein assemblies, *Chem. Soc. Rev.* **39**, 1633-1655.
137. Ewing, R. G., Atkinson, D. A., Eiceman, G. A., and Ewing, G. J. (2001) A critical review of ion mobility spectrometry for the detection of explosives and explosive related compounds, *Talanta* **54**, 515-529.
138. Trimpin, S., Tan, B., Bohrer, B. C., O'Dell, D. K., Merenbloom, S. I., Pazos, M. X., Clemmer, D. E., and Walker, J. M. (2009) Profiling of phospholipids and related lipid structures using multidimensional ion mobility spectrometry-mass spectrometry, *Int. J. Mass Spectrom.* **287**, 58-69.
139. O'Donnell, R. M., Sun, X., and Harrington, P. d. B. (2008) Pharmaceutical applications of ion mobility spectrometry, *TrAC* **27**, 44-53.

140. Bush, M. F., Campuzano, I. D. G., and Robinson, C. V. (2012) Ion mobility mass spectrometry of peptide ions: Effects of drift gas and calibration strategies, *Anal. Chem.* **84**, 7124-7130.
141. Grabenauer, M., Wyttenbach, T., Sanghera, N., Slade, S. E., Pinheiro, T. J. T., Scrivens, J. H., and Bowers, M. T. (2010) Conformational stability of syrian hamster prion protein prp(90−231), *J. Am. Chem. Soc.* **132**, 8816-8818.
142. Baldwin, Andrew J., Lioe, H., Hilton, Gillian R., Baker, Lindsay A., Rubinstein, John L., Kay, Lewis E., and Benesch, J. L.P. (2011) The polydispersity of αB-crystallin is rationalized by an interconverting polyhedral architecture, *Structure* **19**, 1855-1863.
143. Bernstein, S. L., Dupuis, N. F., Lazo, N. D., Wyttenbach, T., Condron, M. M., Bitan, G., Teplow, D. B., Shea, J.-E., Ruotolo, B. T., Robinson, C. V., and Bowers, M. T. (2009) Amyloid-B protein oligomerization and the importance of tetramers and dodecamers in the aetiology of Alzheimer's disease, *Nat. Chem.* **1**, 326-331.
144. Smith, D. P., Radford, S. E., and Ashcroft, A. E. (2010) Elongated oligomers in β-2-microglobulin amyloid assembly revealed by ion mobility-mass spectrometry, *Proc. Natl. Acad. Sci. U.S.A.* **107**, 6794-6798.
145. Pukala, T. L., Ruotolo, B. T., Zhou, M., Politis, A., Stefanescu, R., Leary, J. A., and Robinson, C. V. (2009) Subunit architecture of multiprotein assemblies determined using restraints from gas-phase measurements, *Structure* **17**, 1235-1243.
146. van Duijn, E., Barbu, I. M., Barendregt, A., Jore, M. M., Wiedenheft, B., Lundgren, M., Westra, E. R., Brouns, S. J. J., Doudna, J. A., van der Oost, J., and Heck, A. J. R. (2012) Native tandem and ion mobility mass spectrometry highlight structural and modular similarities in clustered-regularly-interspaced short-palindromic-repeats (CRISPR)-associated protein complexes from Escherichia coli and Pseudomonas aeruginosa, *Mol. Cell. Proteomics.* **11**, 1430-1441.
147. Wyttenbach, T., and Bowers, M. (2003) Gas-phase conformations: The ion mobility/ion chromatography method, in *Modern mass spectrometry* (Schalley, C., Ed.), Springer Berlin, Heidelberg.
148. Smith, D. P., Knapman, T. W., Campuzano, I., Malham, R. W., Berryman, J. T., Radford, S. E., and Ashcroft, A. E. (2009) Deciphering drift time measurements

from travelling wave ion mobility spectrometry-mass spectrometry studies, *Eur. J. Mass Spectrom.* **15**, 113-130.

149. Scarff, C. A., Thalassinos, K., Hilton, G. R., and Scrivens, J. H. (2008) Travelling wave ion mobility mass spectrometry studies of protein structure: Biological significance and comparison with X-ray crystallography and nuclear magnetic resonance spectroscopy measurements, *Rap. Commun. Mass Spectrom.* **22**, 3297-3304.
150. Shelimov, K. B., Clemmer, D. E., Hudgins, R. R., and Jarrold, M. F. (1997) Protein structure in vacuo: Gas-phase conformations of BPTI and cytochrome c, *J. Am. Chem. Soc.* **119**, 2240-2248.
151. Tao, L., McLean, J. R., McLean, J. A., and Russell, D. H. (2007) A collision cross-section database of singly-charged peptide ions, *J. Am. Soc. Mass Spectrom.* **18**, 1232-1238.
152. Cook, R. D., and Weisberg, S. (1982) *Residuals and influence in regression*, Chapman & Hall, New York, NY.
153. Cook, R. D. (1977) Detection of influential observations in linear regression, *Technometrics* **19**, 15-18.
154. Hyung, S. J., Robinson, C. V., and Ruotolo, B. T. (2009) Gas-phase unfolding and disassembly reveals stability differences in ligand-bound multiprotein complexes, *Chem. Biol.* **16**, 382-390.
155. Freeke, J., Robinson, C. V., and Ruotolo, B. T. (2010) Residual counter ions can stabilise a large protein complex in the gas phase, *Int. J. Mass Spectrom.* **298**, 91-98.
156. Jurneczko, E., Kalapothakis, J., Campuzano, I. D. G., Morris, M., and Barran, P. E. (2012) Effects of drift gas on collision cross sections of a protein standard in linear drift tube and traveling wave ion mobility mass spectrometry, *Anal. Chem.* **84**, 8524-8531.
157. von Helden, G., Hsu, M. T., Gotts, N., and Bowers, M. T. (1993) Carbon cluster cations with up to 84 atoms: Structures, formation mechanism, and reactivity, *J. Phys. Chem.* **97**, 8182-8192.
158. Ruotolo, B. T., McLean, J. A., Gillig, K. J., and Russell, D. H. (2004) Peak capacity of ion mobility mass spectrometry: The utility of varying drift gas polarizability for the separation of tryptic peptides, *J. Mass Spectrom.* **39**, 361-367.

159. Williams, J. P., and Scrivens, J. H. (2008) Coupling desorption electrospray ionisation and neutral desorption/extractive electrospray ionisation with a travelling-wave based ion mobility mass spectrometer for the analysis of drugs, *Rap.Commun. Mass Spectrom.* **22**, 187-196.
160. Salbo, R., Bush, M. F., Naver, H., Campuzano, I., Robinson, C. V., Pettersson, I., Jörgensen, T. J., and Haselmann, K. F. (2012) Traveling-wave ion mobility mass spectrometry of protein complexes: Accurate calibrated collision cross-sections of human insulin oligomers, *Rap. Commun. Mass Spectrom.* **26**, 1181-1193.
161. Ridenour, W. B., Kliman, M., McLean, J. A., and Caprioli, R. M. (2010) Structural characterization of phospholipids and peptides directly from tissue sections by MALDI traveling-wave ion mobility-mass spectrometry, *Anal. Chem.* **82**, 1881-1889.
162. Zhou, M., Huang, C., and Wysocki, V. H. (2012) Surface-induced dissociation of ion mobility-separated noncovalent complexes in a quadrupole/time-of-flight mass spectrometer, *Anal. Chem*. **84**, 6016-6023.
163. Hopper, J. T. S., and Oldham, N. J. (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, *J. Am. Soc. Mass Spectrom.* **20**, 1851-1858.
164. Zhong, Y., Hyung, S. J., and Ruotolo, B. T. (2011) Characterizing the resolution and accuracy of a second-generation traveling-wave ion mobility separator for biomolecular ions, *Analyst* **136**, 3534-3541.
165. Giles, K., Wildgoose, J. L., Langridge, D. J., and Campuzano, I. (2010) A method for direct measurement of ion mobilities using a travelling wave ion guide, *Int. J. Mass Spectrom.* **298**, 10-16.
166. Michaelevski, I., Eisenstein, M., and Sharon, M. (2010) Gas-phase compaction and unfolding of protein structures, *Anal. Chem.* **82**, 9484-9491.
167. Lai, H., McJunkin, T. R., Miller, C. J., Scott, J. R., and Almirall, J. R. (2008) The predictive power of SIMION simulation software for modeling ion mobility spectrometry instruments, *Int. J. Mass Spectrom.* **276**, 1-8.
168. Lasker, K., Phillips, J. L., Russel, D., Velazquez-Muriel, J., Schneidman-Duhovny, D., Tjioe, E., Webb, B., Schlessinger, A., and Sali, A. (2010) Integrative structure modeling of macromolecular assemblies from proteomics data, *Mol. Cell. Proteomics.* **9**, 1689-1702.

169. Alber, F., Dokudovskaya, S., Veenhoff, L. M., Zhang, W., Kipper, J., Devos, D., Suprapto, A., Karni-Schmidt, O., Williams, R., Chait, B. T., Sali, A., and Rout, M. P. (2007) The molecular architecture of the nuclear pore complex, *Nature* **450**, 695-701.
170. Förster, F., Lasker, K., Nickell, S., and Sali, A. (2010) Toward an integrated structural model of the 26S proteasome, *Mol. Cell. Proteomics.* **9.8**, 1666-1677.
171. Lasker, K., Förster, F., Bohn, S., Walzthoeni, T., Villa, E., Unverdorben, P., Beck, F., Aebersold, R., Sali, A., and Baumeister, W. (2012) Molecular architecture of the 26S proteasome holocomplex determined by an integrative approach, *Proc. Natl. Acad. Sci. U.S.A.* **109**, 1380-1387.
172. Miao, J., Ishikawa, T., Shen, Q., and Earnest, T. (2008) Extending X-ray crystallography to allow the imaging of noncrystalline materials, cells, and single protein complexes, *Annu. Rev. Phys. Chem.* **59**, 387-410.
173. McDermott, A. (2009) Structure and dynamics of membrane proteins by magic angle spinning solid-state NMR, *Annu. Rev. Biophys.* **38**, 385-403.
174. Schneidman-Duhovny, D., Rossi, A., Avila-Sakar, A., Kim, S. J., Velazquez-Muriel, J., Strop, P., Liang, H., Krukenberg, K. A., Liao, M., Kim, H. M., Sobhanifar, S., Dotsch, V., Rajpal, A., Pons, J., Agard, D. A., Cheng, Y., and Sali, A. (2012) A method for integrative structure determination of protein-protein complexes, *Bioinformatics*. **28**, 3282-3289.
175. Walhout, A. J. M., and Vidal, M. (2001) High-throughput yeast two-hybrid assays for large-scale protein interaction mapping, *Methods* **24**, 297-306.
176. Bauer, A., and Kuster, B. (2003) Affinity purification-mass spectrometry, *Eur. J. Biochem.* **270**, 570-578.
177. Stengel, F., Aebersold, R., and Robinson, C. V. (2012) Joining forces: Integrating proteomics and cross-linking with the mass spectrometry of intact complexes, *Mol. Cell Proteomics.* **11**, doi: 10.1074/mcp.R111.014027.
178. Orlova, E. V., and Saibil, H. R. (2011) Structural analysis of macromolecular assemblies by electron microscopy, *Chem. Rev.* **111**, 7710-7748.
179. Tsutsui, Y., and Wintrode, P. L. (2007) Hydrogen/deuterium exchange-mass spectrometry: A powerful tool for probing protein structure, dynamics and interactions, *Curr. Med. Chem.* **14**, 2344-2358.
180. Wang, L., and Chance, M. R. (2011) Structural mass spectrometry of proteins using hydroxyl radical based protein footprinting, *Anal. Chem.* **83**, 7234-7241.

181. Rappsilber, J. (2011) The beginning of a beautiful friendship: Cross-linking/mass spectrometry and modelling of proteins and multi-protein complexes, *J. Struct. Biol.* **173**, 530-540.
182. Mertens, H., and Svergun, D. I. (2010) Structural characterization of proteins and complexes using small-angle X-ray solution scattering, *J. Struct. Biol.* **172**, 128-141.
183. Zhou, Z. H. (2008) Towards atomic resolution structural determination by single-particle cryo-electron microscopy, *Curr. Op. Struct. Biol.* **18**, 218-228.
184. Alber, F., Kim, M. F., and Sali, A. (2005) Structural characterisation of assemblies from overall shape and subcomplex compositions, *Structure* **13**, 435-445.
185. Yang, Z., Lasker, K., Schneidman-Duhovny, D., Webb, B., Huang, C. C., Pettersen, E. F., Goddard, T. D., Meng, E. C., Sali, A., and Ferrin, T. E. UCSF Chimera, Modeller, and IMP: An integrated modeling system, *J. Struct. Biol.*, **179**, 269-278.
186. Webb, B., Lasker, K., Schneidman-Duhovny, D., Tjioe, E., Phillips, J., Kim, S. J., Velazquez-Muriel, J., Russel, D., and Sali, A. (2011) Modeling of proteins and their assemblies with the integrative modeling platform, *Methods Mol. Biol.* **781**, 377-397.
187. Förster, F., Webb, B., Krukenberg, K. A., Tsuruta, H., Agard, D. A., and Sali, A. (2008) Integration of small-angle X-ray scattering data into structural modeling of proteins and their assemblies, *J. Mol. Biol.* **382**, 1089-1106.
188. Schneidman-Duhovny, D., Kim, S. J., and Sali, A. (2012) Integrative structural modeling with small angle X-ray scattering profiles, *BMC Struct. Biol.* **12**, 1-12.
189. Heck, A. J. R. (2008) Native mass spectrometry: A bridge between interactomics and structural biology, *Nat. Methods* **5**, 927-933.
190. van Duijn, E. (2010) Current limitations in native mass spectrometry based structural biology, *J. Am. Soc. Mass Spectrom.* **21**, 971-978.
191. Stengel, F., Baldwin, A. J., Painter, A. J., Jaya, N., Basha, E., Kay, L. E., Vierling, E., Robinson, C. V., and Benesch, J. L. P. (2010) Quaternary dynamics and plasticity underlie small heat shock protein chaperone function, *Proc. Natl. Acad. Sci. U.S.A.* **107**, 2007-2012.
192. Lane, L. A., Fernández-Tornero, C., Zhou, M., Morgner, N., Ptchelkine, D., Steuerwald, U., Politis, A., Lindner, D., Gvozdenovic, J., Gavin, A.-C., Müller, C.

W., and Robinson, C. V. (2011) Mass spectrometry reveals stable modules in holo and apo RNA polymerases I and III, *Structure* **19**, 90-100.

193. Kaddis, C., Lomeli, S. H., Yin, S., Berhane, B., Apostol, M., Kickhoefer, V. A., Rome, L., and Loo, J. A. (2007) Sizing large proteins and protein complexes by electrospray ionization mass spectrometry and ion mobility, *J. Am. Soc. Mass Spectrom.* **18**, 1206-1216.
194. Bush, M. F., Hall, Z., Giles, K., Hoyes, J., Robinson, C. V., and Ruotolo, B. T. (2010) Collision cross sections of proteins and their complexes: A calibration framework and databse for gas-phase structural biology, *Anal. Chem.* **82**, 9557-9565.
195. Fischer, H., Polikarpov, I., and Craievich, A. F. (2004) Average protein density is a molecular weight dependent function, *Protein Sci.* **13**, 2825-2828.
196. Goodsell, D. S., and Olson, A. J. (2000) Structural symmetry and protein function, *Annu. Rev. Biophys. Biomol. Struct.* **29**, 105-153.
197. Levy, E. D., Pereira-Leal, J. B., Chothia, C., and Teichmann, S. A. (2006) 3D complex: A structural classification of protein complexes, *PLoS Comput. Biol.* **2**, e155.
198. Theodoridis, S., and Koutroumbas, K. (1999) *Pattern recognition*, Academic Press, London, U.K.
199. Okorokov, A. L., Orlova, E. V., Kingsbury, S. R., Bagneris, C., Gohlke, U., Williams, G. H., and Stoeber, K. (2004) Molecular structure of human geminin, *Nat. Struct. Mol. Biol.* **11**, 1021-1022.
200. Topf, M., and Sali, A. (2005) Combining electron microscopy and comparative protein structure modeling, *Curr. Op. Struct. Biol.* **15**, 578-585.
201. Eichner, T., and Radford, S. E. (2011) Understanding the complex mechanisms of β2-microglobulin amyloid assembly, *FEBS J.* **278**, 3868-3883.
202. Eichner, T., Kalverda, A. P., Thompson, G. S., Homans, S. W., and Radford, S. E. (2011) Conformational conversion during amyloid formation at atomic resolution, *Mol. Cell* **41**, 161-172.
203. Borysik, A. J. H., Radford, S. E., and Ashcroft, A. E. (2010) Co-populated conformational ensembles of β2-microglobulin uncovered quantitatively by electrospray ionisation mass spectrometry, *J. Biol. Chem.* **279**, *27069-27077.*
204. Smith, A. M., Jahn, T. R., Ashcroft, A. E., and Radford, S. E. (2006) Direct

observation of oligomeric species formed in the early stages of amyloid fibril formation using electrospray ionisation mass spectrometry, *J. Mol. Biol.* **364**, 9-19.

205. Ashcroft, A. E. (2010) Mass spectrometry and the amyloid problem - how far can we go in the gas phase?, *J. Am. Chem. Soc.* **21**, 1087-1096.
206. Smith, D. P., Giles, K., Bateman, R. H., Radford, S. E., and Ashcroft, A. E. (2007) Monitoring copopulated conformational states during protein folding events using electrospray ionization-ion mobility spectrometry-mass spectrometry, *J. Am. Soc. Mass Spectrom.* **18**, 2180-2190.
207. Smith, D. P., Woods, L. A., Radford, S. E., and Ashcroft, A. E. (2011) Structure and dynamics of oligomeric intermediates in β2-microglobulin self-assembly, *Biophys. J.* **101**, 1238-1247.
208. Kad, N. M., Thomson, N. H., Smith, D. P., Smith, D. A., and Radford, S. E. (2001) β2-microglobulin and its deamidated variant, N17D form amyloid fibrils with a range of morphologies in vitro, *J. Mol. Biol.* **313**, 559-571.
209. Woods, L. A., Radford, S. E., and Ashcroft, A. E. (2012) Advances in ion mobility spectrometry-mass spectrometry reveal key insights into amyloid assembly, *Biochim. Biophys. Acta*. doi: 10.1016/j.bbapap.2012.10.002.
210. Colombo, M., de Rosa, M., Bellotti, V., Ricagno, S., and Bolognesi, M. (2012) A recurrent D-strand association interface is observed in β2-microglobulin oligomers, *FEBS J.* **279**, 1131-1143.
211. Mendoza, V. L., Baron-Rodriguez, M. A., Blanco, C., and Vachet, R. W. (2011) Structural insights into the pre-amyloid tetramer of β2-microglobulin from covalent labeling and mass spectrometry, *Biochemistry* **50**, 6711-6722.
212. White, H. E., Hodgkinson, J. L., Jahn, T. R., Cohen-Krausz, S., Gosal, W. S., Müller, S., Orlova, E. V., Radford, S. E., and Saibil, H. R. (2009) Globular tetramers of β2-microglobulin assemble into elaborate amyloid fibrils, *J. Mol. Biol.* **389**, 48-57.
213. Poliakov, A., van Duijn, E., Lander, G., Fu, C.-y., HJonhson, J. E., Prevelige, P. E., and Heck, A. J. R. (2007) Macromolecular mass spectrometry and electron microscopy as complementary tools for investigation of the heterogeneity of bacteriophage portal assemblies, *J. Struct. Biol.* **157**, 371-383.
214. Leary, J. A., Schenauer, M. R., Stefanescu, R., Andaya, A., Ruotolo, B. T., Robinson, C. V., Thalassinos, K., Scrivens, J. H., Sokabe, M., and Hershey, J. W.

(2009) Methodology for measuring conformation of solvent disrupted protein subunits using T-wave ion mobility MS: An investigation into eukaryotic initiation factors, *J. Am. Soc. Mass Spectrom.* **20**, 1699-1706.

215. Benesch, J. L. P., and Robinson, C. V. (2009) Dehydrated but unharmed, *Nature* **462**, 576-577.
216. Ruotolo, B. T., and Robinson, C. V. (2006) Aspects of native proteins are retained in vacuum, *Curr. Op. Chem. Biol.* **10**, 402-408.
217. Rozbesky, D., Man, P., Kavan, D., Chmelik, J., Cerny, J., Bezouska, K., and Novak, P. (2011) Chemical cross-linking and H/D exchange for fast refinement of protein crystal structure, *Anal. Chem.* **84**, 867-870.
218. Rozbesky, D., Sovova, Z., Marcoux, J., Man, P., Ettrich, R., Robinson, C. V., and Novak, P. (2012) Structural model of lymphocyte receptor NKR-PLC revealed by mass spectrometry and molecular modeling, *Anal. Chem.* doi: 0.1021/ac302860m
219. Loo, J. A., Loo, R. R., Light, K. J., Edmonds, C. G., and Smith, R. D. (1992) Multiply charged negative ions by electrospray ionization of polypeptides and proteins, *Anal. Chem.* **64**, 81-88.
220. Maisser, A., Premnath, V., Ghosh, A., Nguyen, T. A., Attoui, M., and Hogan Jr., C. J., (2011) Determination of gas phase protein ion densities via ion mobility analysis with charge reduction, *Phys. Chem. Chem. Phys.* **13**, 21630-21641.
221. Pagel, K., Hyung, S. J., Ruotolo, B. T., and Robinson, C. V. (2010) Alternate dissociation pathways identified in charge-reduced protein complex ions, *Anal. Chem.* **82**, 5363-5372.
222. Lemaire, D., Marie, G., Serani, L., and Laprevote, O. (2001) Stabilization of gas-phase noncovalent macromolecular complexes in electrospray mass spectrometry using aqueous triethylammonium bicarbonate buffer, *Anal. Chem.* **73**, 1699-1706.
223. Touboul, D., Jecklin, M., and Zenobi, R. (2008) Investigation of deprotonation reactions on globular and denatured proteins at atmospheric pressure by ESSI-MS, *J. Am. Soc. Mass Spectrom.* **19**, 455-466.
224. Hopper, J. T., and Oldham, N. J. (2011) Alkali metal cation-induced destabilization of gas-phase protein-ligand complexes: Consequences and prevention, *Anal. Chem.* **83**, 7472-7479.
225. Hopper, J. T. S., Sokratous, K., and Oldham, N. J. (2012) Charge state and adduct reduction in electrospray ionization–mass spectrometry using solvent vapor exposure, *Anal. Biochem.* **421**, 788-790.

226. Xia, Y., Chrisman, P. A., Erickson, D. E., Liu, J., Liang, X., Londry, F. A., Yang, M. J., and McLuckey, S. A. (2006) Implementation of ion/ion reactions in a quadrupole/time of flight tandem mass spectrometer, *Anal. Chem.* **78**, 4146-4154.
227. Sun, J., Kitova, E. N., and Klassen, J. S. (2006) Method for stabilizing protein−ligand complexes in nanoelectrospray ionization mass spectrometry, *Anal. Chem.* **79**, 416-425.
228. Bornschein, R. E., Hyung, S. J., and Ruotolo, B. T. (2011) Ion mobility-mass spectrometry reveals conformational changes in charge reduced multiprotein complexes, *J. Am. Soc. Mass Spectrom.* **22**, 1690-1698.
229. Lomeli, S. H., Yin, S., Ogorzalek Loo, R. R., and Loo, J. A. (2009) Increasing charge while preserving noncovalent protein complexes for ESI-MS, *J. Am. Soc. Mass Spectrom.* **20**, 593-596.
230. Yin, S., and Loo, J. A. Top-down mass spectrometry of supercharged native protein-ligand complexes, *Int. J. Mass Spectrom.* **300**, 118-122.
231. Iavarone, A. T., Jurchen, J. C., and Williams, E. R. (2001) Supercharged protein and peptide ions formed by electrospray ionization, *Anal. Chem.* **73**, 1455-1460.
232. Lomeli, S. H., Peng, I. X., Yin, S., Ogorzalek Loo, R. R., and Loo, J. A. (2010) New reagents for increasing ESI multiple charging of proteins and protein complexes, *J. Am. Soc. Mass Spectrom.* **21**, 127-131.
233. Iavarone, A. T., and Williams, E. R. (2003) Mechanism of charging and supercharging molecules in electrospray ionization, *J. Am. Chem. Soc.* **125**, 2319-2327.
234. Sterling, H., Kintzer, A., Feld, G., Cassou, C., Krantz, B., and Williams, E. (2012) Supercharging protein complexes from aqueous solution disrupts their native conformations, *J. Am. Soc. Mass Spectrom.* **23**, 191-200.
235. Sterling, H. J., Cassou, C. A., Trnka, M. J., Burlingame, A. L., Krantz, B. A., and Williams, E. R. (2011) The role of conformational flexibility on protein supercharging in native electrospray ionization, *Phys. Chem. Chem. Phys.* **13**,18288-18296.
236. Sterling, H. J., Daly, M. P., Feld, G. K., Thoren, K. L., Kintzer, A. F., Krantz, B. A., and Williams, E. R. (2010) Effects of supercharging reagents on noncovalent complex structure in electrospray ionization from aqueous solutions, *J. Am. Soc. Mass Spectrom.* **21**, 1762-1774.

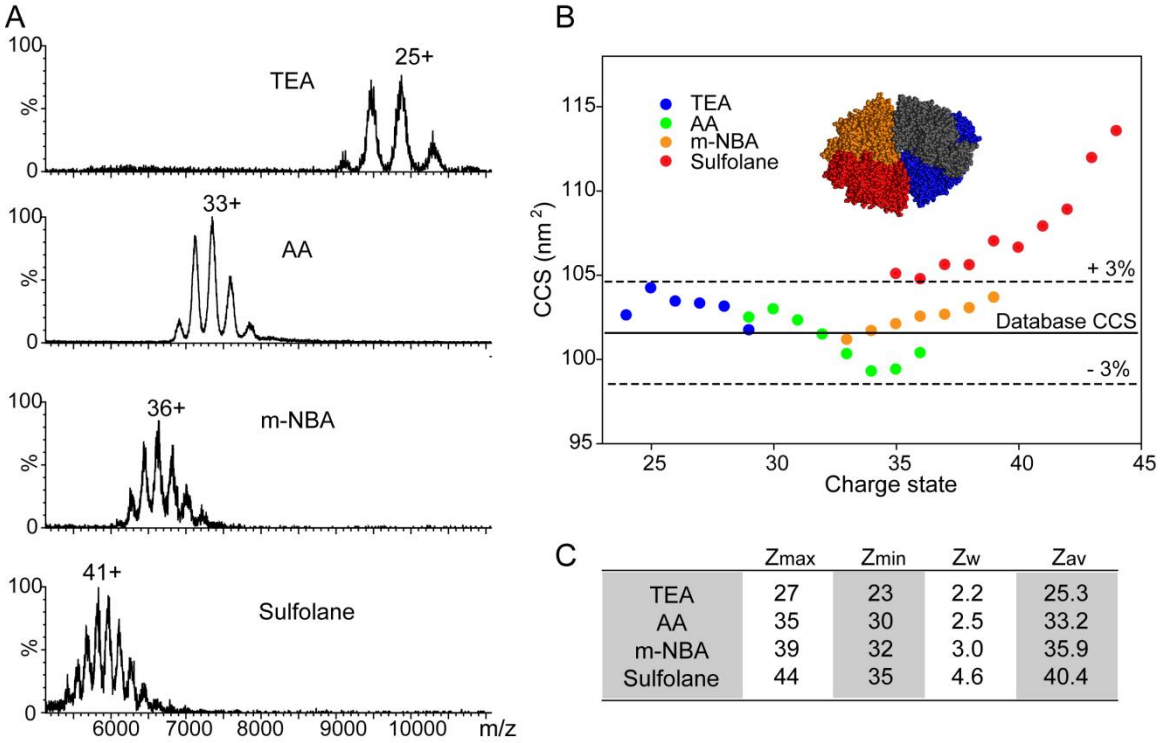
237. Douglass, K. A., and Venter, A. R. (2012) Investigating the role of adducts in protein supercharging with sulfolane, *J. Am. Soc. Mass Spectrom.* **23**, 489-497.
238. Kukrer, B., Barbu, I. M., Copps, J., Hogan, P., Taylor, S. S., van Duijn, E., and Heck, A. J. (2012) Conformational isomers of calcineurin follow distinct dissociation pathways, *J. Am. Soc. Mass Spectrom.* **23**, 1534-1543.
239. Boeri Erba, E., Ruotolo, B. T., Barsky, D., and Robinson, C. V. (2010) Ion mobility mass spectrometry reveals the influenece of subunit packing and charge on the dissociation of multiprotein complexes, *Anal. Chem.* **82**, 9702-9710.
240. van den Heuvel, R. H. H., van Duijn, E., Mazon, H., Synowsky, S. A., Lorenzen, K., Versluis, C., Brouns, S. J. J., Langridge, D., van der Oost, J., Hoyes, J., and Heck, A. J. R. (2006) Improving the performance of a quadrupole time-of-flight instrument for macromolecular mass spectrometry, *Anal. Chem.* **78**, 7473-7483.
241. Hogan Jr., C. J., Ogorzalek Loo, R. R., Loo, J. A., and Fernandez de la Mora, J. (2010) Ion mobility-mass spectrometry of phosphorylase B ions generated with supercharging reagents but in charge-reducing buffer, *Phys. Chem. Chem. Phys.* **12**,13476-13483.
242. Sobott, F., McCammon, M. G., and Robinson, C. V. (2003) Gas-phase dissociation pathways of a tetrameric protein complex, *Int. J. Mass Spectrom.* **230**, 193-200.
243. Beardsley, R. L., Jones, C. M., Galhena, A. S., and Wysocki, V. H. (2009) Noncovalent protein tetramers and pentamers with "n" charges yield monomers with n/4 and n/5 charges, *Anal. Chem.* **81**, 1347-1356.
244. Zhou, M., Dagan, S., and Wysocki, V. H. (2013) Impact of charge state on gas-phase behaviors of noncovalent protein complexes in collision induced dissociation and surface induced dissociation, *Analyst*. doi: 10.1039/C2AN36525A
245. Knapman, T. W., Morton, V. L., Stonehouse, N. J., Stockley, P. G., and Ashcroft, A. E. (2010) Determining the topology of virus assembly intermediates using ion mobility spectrometry–mass spectrometry, *Rap. Commun. Mass Spectrom.* **24**, 3033-3042.
246. Felitsyn, N., Kitova, E. N., and Klassen, J. S. (2002) Thermal dissociation of the protein homodimer ecotin in the gas phase, *J. Am. Soc. Mass Spectrom.* **13**, 1432-1442.
247. Aquilina, J. A. (2009) The major toxin from the Australian common brown snake is a hexamer with unusual gas-phase dissociation properties, *Proteins: Struct. Funct. Bioinf.* **75**, 478-485.

248. Levy, E. D., Erba, E. B., Robinson, C. V., and Teichmann, S. A. (2008) Assembly reflects evolution of protein complexes, *Nature* **453**, 1262-1265.
249. Hollander, M., and Wolfe, D. A. (1999) *Nonparametric Statistical Methods.* John Wiley & Sons, Inc., New York, NY.
250. Blackwell, A. E., Dodds, E. D., Bandarian, V., and Wysocki, V. H. (2011) Revealing the quaternary structure of a heterogeneous noncovalent protein complex through surface-induced dissociation, *Anal. Chem.* **83**, 2862-2865.
251. Rockwood, A. L., Busman, M., and Smith, R. D. (1991) Coulombic effects in the dissociation of large highly charged ions, *Int. J. Mass Spectrom.* **111**, 103-129.
252. Wanasundara, S. N., and Thachuk, M. (2009) Free energy barrier estimation for the dissociation of charged protein complexes in the gas phase, *J. Phys. Chem. A* **113**, 3814-3821.
253. Patriksson, A., Marklund, E., and van der Spoel, D. (2007) Protein structures under electrospray conditions, *Biochemistry* **46**, 933-945.
254. Thirumuruganandham, S. P., and Urbassek, H. M. (2010) Water evaporation and conformational changes from partially solvated ubiquitin, *Biochem. Res. Int.* **2010**.
255. Han, L., Hyung, S.-J., Mayers, J. J. S., and Ruotolo, B. T. (2011) Bound anions differentially stabilize multiprotein complexes in the absence of bulk solvent, *J. Am. Chem. Soc.* **133**, 11358-11367.
256. Vijayakumar, S., Vishveshwara, S., Ravishanker, G., and Beveridge, D. L. (1993) Differential stability of β-sheets and α-helices in beta-lactamase: A high temperature molecular dynamics study of unfolding intermediates, *Biophys. J.* **65**, 2304-2312.
257. Benson, N. C., and Daggett, V. (2008) Dynameomics: Large-scale assessment of native protein flexibility, *Protein Sci.* **17**, 2038-2050.
258. Daggett, V. (2002) Molecular dynamics simulations of the protein unfolding/folding reaction, *Acc. Chem. Res.* **35**, 422-429.
259. Uversky, V. N. (2011) Intrinsically disordered proteins from A to Z, *Int. J. Biochem. Cell Biol.* **43**, 1090-1103.
260. Natan, E., Hirschberg, D., Morgner, N., Robinson, C. V., and Fersht, A. R. (2009) Ultraslow oligomerization equilibria of p53 and its implications, *Proc. Natl. Acad. Sci. U.S.A.* **106**, 14327-14332.

261. Joerger, A. C., and Fersht, A. R. (2008) Structural biology of the tumor suppressor p53, *Annu. Rev. Biochem.* **77**, 557-582.
262. Faull, P. A., Florance, H. V., Schmidt, C. Q., Tomczyk, N., Barlow, P. N., Hupp, T. R., Nikolova, P. V., and Barran, P. E. (2010) Utilising ion mobility-mass spectrometry to interrogate macromolecules: Factor H complement control protein modules 10-15 and 19-20 and the DNA-binding core domain of tumour suppressor p53, *Int. J. Mass Spectrom.* **298**, 99-110.
263. Testa, L., Brocca, S., and Grandori, R. (2011) Charge-surface correlation in electrospray ionization of folded and unfolded proteins, *Anal. Chem.* **83**, 6459-6463.
264. Beck, D. A. C., and Daggett, V. (2004) Methods for molecular dynamics simulations of protein folding/unfolding in solution, *Methods* **34**, 112-120.
265. Segev, E., Wyttenbach, T., Bowers, M. T., and Gerber, R. B. (2008) Conformational evolution of ubiquitin ions in electrospray mass spectrometry: Molecular dynamics simulations at gradually increasing temperatures, *Phys. Chem. Chem. Phys.* **10**, 3077-3082.
266. Tao, L., Dahl, D. B., Perez, L. M., and Russell, D. H. (2009) The contributions of molecular framework to IMS collision cross-sections of gas-phase peptide ions, *J. Am. Soc. Mass Spectrom.* **20**, 1593-1602.
267. Bahadur, R., and Zacharias, M. (2008) The interface of protein-protein complexes: Analysis of contacts and prediction of interactions, *Cell. Mol. Life Sci.* **65**, 1059-1072.
268. Hubbard, S. J., and Argos, P. (1994) Cavities and packing at protein interfaces, *Protein Sci.* **3**, 2194-2206.
269. Bahadur, R. P., Chakrabarti, P., Rodier, F., and Janin, J. (2003) Dissecting subunit interfaces in homodimeric proteins, *Protein Struct. Func. Bioinf.* **53**, 708-719.
270. Dodds, E. D., Blackwell, A. E., Jones, C., M., Holso, K. L., O'Brien, D. J., Cordes, M. H. J., and Wysocki, V. H. (2011) Determinants of gas-phase disassembly behaviour in homodimeric protein complexes with related yet divergent structures, *Anal. Chem.* **83**, 3881-3889.
271. Flick, T., and Williams, E. (2012) Supercharging with trivalent metal ions in native mass spectrometry, *J. Am. Soc. Mass Spectrom.* **23**, 1885-1895.

APPENDIX

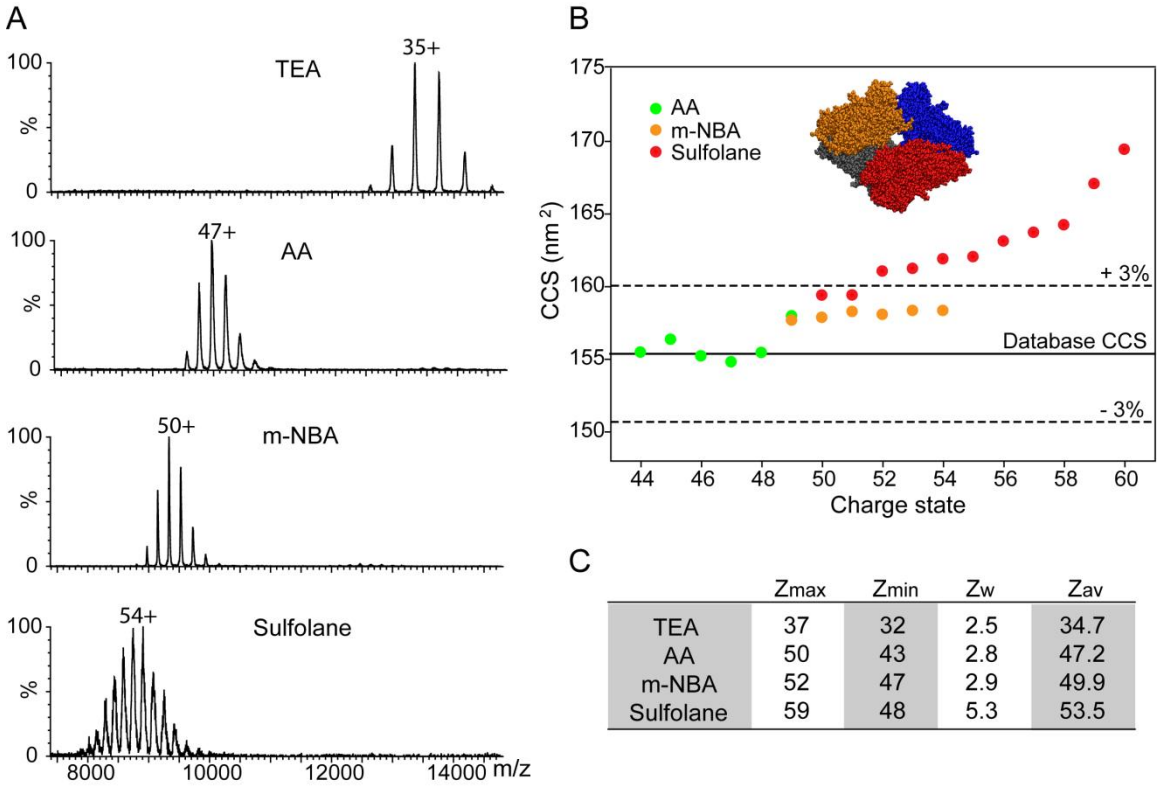
Supercharging and charge reduction: Effects on conformation



C

	Zmax	Zmin	Zw	Zav
TEA	27	23	2.2	25.3
AA	35	30	2.5	33.2
m-NBA	39	32	3.0	35.9
Sulfolane	44	35	4.6	40.4

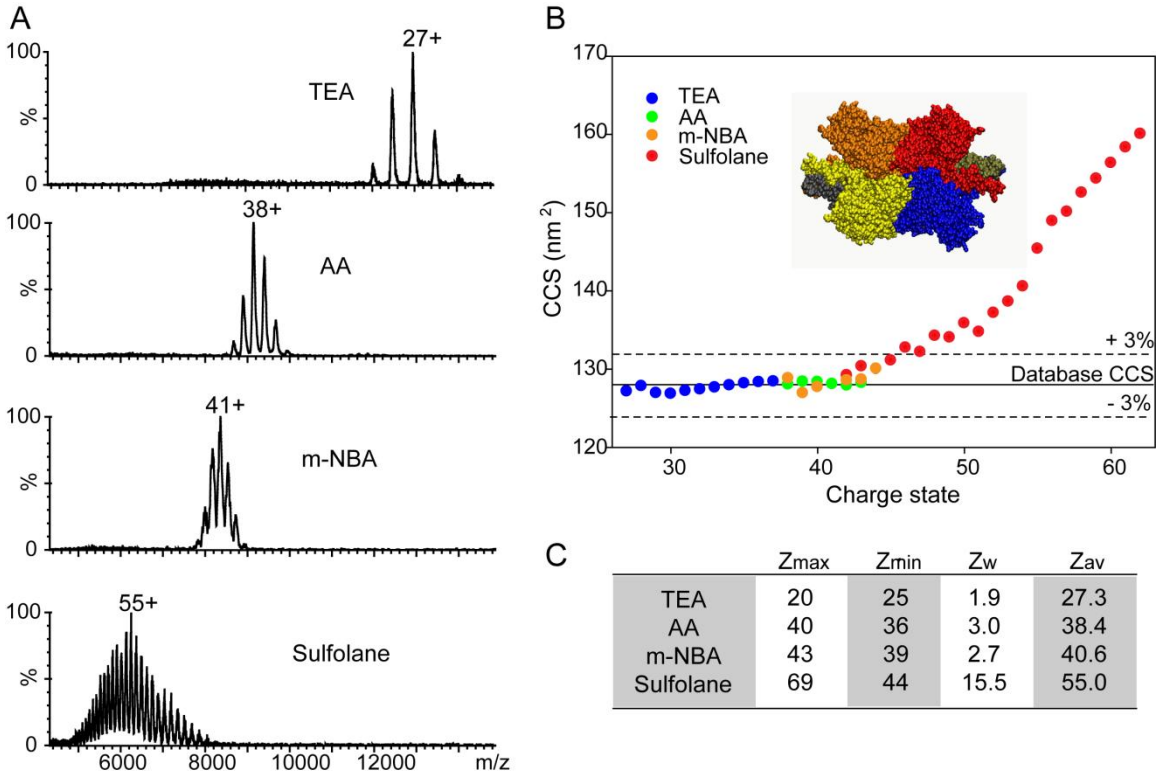
Figure A1. Charge manipulation of pyruvate kinase. **A** MS spectra of pyruvate kinase (232 kDa, 4-mer) in 200 mM AA, or 200mM AA with either 20 mM TEA, 1 % *m*-NBA or 7 % sulfolane. **B** The CCS of pyruvate kinase ions are plotted against charge state for various solution conditions. Average CCS and associated error, as determined in Chapter 3, are shown as solid and dashed lines. **C** The charge state distribution parameters are shown in table form.



C

	Zmax	Zmin	Zw	Zav
TEA	37	32	2.5	34.7
AA	50	43	2.8	47.2
m-NBA	52	47	2.9	49.9
Sulfolane	59	48	5.3	53.5

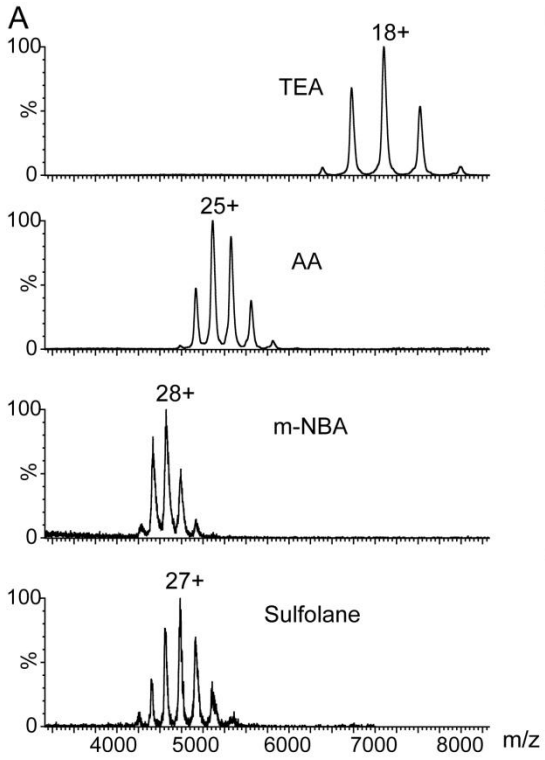
Figure A2. Charge manipulation of *β*-galactosidase. **A** MS spectra of *β*-galactosidase (465 kDa, 4-mer) in 200 mM AA, or 200mM AA with either 20 mM TEA, 1 % *m*-NBA or 7 % sulfolane. **B** The CCS of *β*-galactosidase ions are plotted against charge state for various solution conditions. CCS measurements from TEA buffered solutions were not possible due to instability in the electrospray. Average CCS and associated error, as determined in Chapter 3, are shown as solid and dashed lines. **C** The charge state distribution parameters are shown in table form.

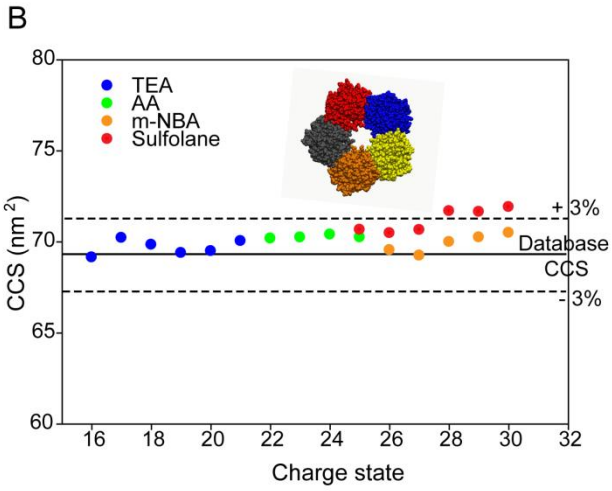


C

	Zmax	Zmin	Zw	Zav
TEA	20	25	1.9	27.3
AA	40	36	3.0	38.4
m-NBA	43	39	2.7	40.6
Sulfolane	69	44	15.5	55.0

Figure A3. Charge manipulation of glutamate dehydrogenase. **A** MS spectra of glutamate dehydrogenase (336 kDa, 6-mer) in 200 mM AA, or 200mM AA with either 20 mM TEA, 1 % *m*-NBA or 7 % sulfolane. **B** The CCS of glutamate dehydrogenase ions are plotted against charge state for various solution conditions. Average CCS and associated error, as determined in Chapter 3, are shown as solid and dashed lines. **C** The charge state distribution parameters are shown in table form.





C

	Zmax	Zmin	Zw	Zav
TEA	20	16	2.2	18.1
AA	27	22	2.6	24.6
m-NBA	30	25	2.5	28.2
Sulfolane	30	24	3.1	27.1

Figure A4. Charge manipulation of SAP. **A** MS spectra of SAP (125 kDa, 5-mer) in 200 mM AA, or 200mM AA with either 20 mM TEA, 1 % *m*-NBA or 7 % sulfolane. **B** The CCS of SAP ions are plotted against charge state for various solution conditions. Average CCS and associated error, as determined in Chapter 3, are shown as solid and dashed lines. **C** The charge state distribution parameters are shown in table form.

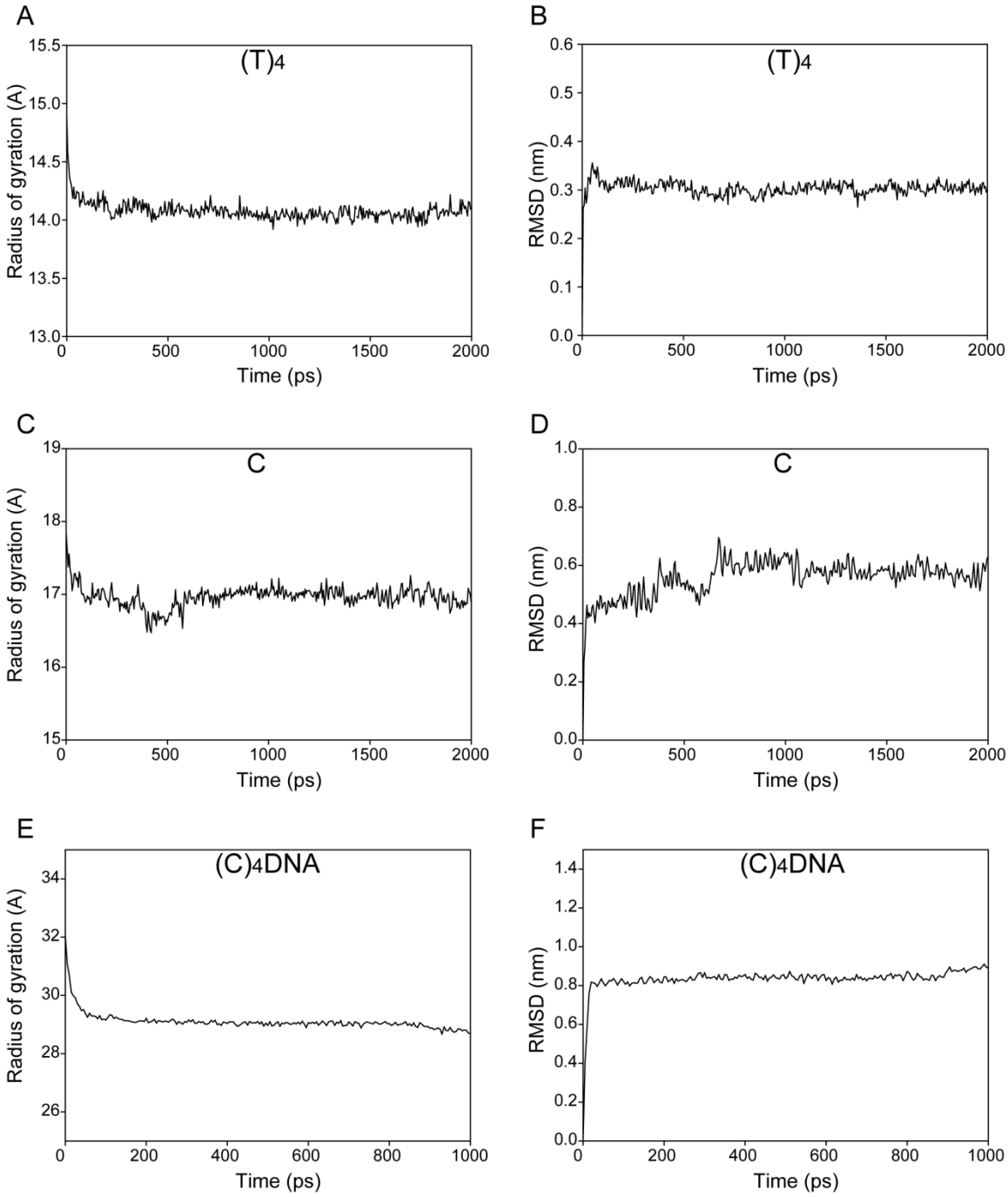


Figure A5. *In vacuo* MD simulations (300 K, 1 or 2 ns) were carried out for isolated p53 domains. The radius of gyration (left hand side) and backbone RMSD from the starting structure at 0 ns (right hand side) is plotted against simulation time in ps for $(T)_4$ (**A** and **B**), C (**C** and **D**) and $(C)_4$DNA (**E** and **F**). The average CCS for atomic coordinates from 500 ps to the end of the simulation was calculated.