New Methods for Exploration of the Conformational Space of Intrinsically Disordered Proteins

A thesis submitted to The University of Manchester for the degree of Doctor of Philosophy in the Faculty of Science and Engineering

2021

DALE STUCHFIELD

SCHOOL OF NATURAL SCIENCES

Department of Chemistry

Table of Contents

List of Figures

Introduction

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Appendix 2

Appendix 3

List of Tables

Abbreviations

AIMS – Aspiration Ion Mobility Spectrometry

ATD – Arrival Time Distribution

CCS – Collision Cross Section

CCSD – Collision Cross Section Distribution

CSD – Charge State Distribution

COR – Cold Regulated

DMS – Differential Mobility Spectrometry

DSSP – Define Secondary Structure of Proteins

DTIMS – Drift Time Ion Mobility Spectrometry

EHSS – Exact hard Sphere Scattering

EOM – Ensemble Optimisation Method

ESI – Electrospray Ionisation

FAIMS – Field-asymmetric Waveform Ion Mobility Spectrometry

FCR – Fraction of Charged Residues

GC – Gas Chromatography

IDP – Intrinsically Disordered Protein

IDR – Intrinsically Disordered Region

IM-MS – Ion Mobility Mass Spectrometry

IMS – Ion Mobility Spectrometry

IM – Ion Mobility

KDE – Kernel Density Estimation

LC-MS – Liquid Chromatography Mass Spectrometry

LC – Liquid Chromatography

LEA – Late Embryogenesis Abundance

MD – Molecular Dynamics

MM – Molecular Mechanics

MS – Mass Spectrometry

MW – Molecular Weight

NCPR – Net Charge Per residues

nESI – nano Electrospray Ionisation

NMR – Nuclear Magnetic Resonance

NPT – Number Pressure Temperature

NVT – Number Volume Temperature

o*p*CCSD – Overall *pseudo*-Collision Cross Section Distribution

PA – Projection Approximation

*p*CCS – *pseudo*-Collision Cross Section

*p*CCSD – *pseudo*-Collision Cross Section Distribution

PG – Poly-Glycine

PSA – Proximity Superposition Approximation

PTMS – Post Translational Modifications

RF – Radio Frequency

RT – Room Temperature

SAXS – Small Angle X-ray Scattering

TFE - Trifluoroethanol

TIMS – Trapped Ion Mobility Spectrometry

TWIMS – Travelling Wave Ion Mobility Spectrometry

VT-IM-MS – Variable Temperature Ion Mobility Mass Spectrometry

VT – Variable Temperature

WT – Wild Type

Abstract

The aim of this thesis was to develop our understanding of Intrinsically Disordered Proteins (IDPs). The first chapter of this thesis is a literature review which at the time of writing aimed to highlight much of the recent work carried out using Ion Mobility Mass Spectrometry (IM-MS) to study IDPs in detail. Chapter two is a methodology chapter on the use of IM-MS and Molecular Dynamics (MD) to examine IDPs. The chapter covers sample preparation, purification, optimisation of MS conditions, tuning parameters, data collection and analysis.

The third chapter is the first results chapter which focuses on determining the secondary structural features in the conformational landscape of the cold regulating plant protein COR15A found in *Arabidopsis Thaliana.* Here comparison of MD simulations with IM-MS data is used to explore the conformational landscape of COR15A. The study showed how combining these two techniques can give us an atomistic scale insight into candidate structures for different conformations that exist within the COR15A system. Chapter four continues with analysis of COR15A by looking at how it behaved under cold temperatures. Variable Temperature IM-MS is used to investigate the conformational changes undergone by COR15A in temperatures ranging from 193 K – 298 K. The work demonstrated how VT IM-MS can provide molecular scale detail on the cold denaturation of the COR15A protein and discusses how there are currently few methods that can adequately do this.

The fifth chapter is the conclusion of the thesis and looks back on the work, drawing conclusions and discussing what has been done and what future work can be done to build on what has been accomplished in this thesis.

Declaration

No portion of the work referred to in the thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

Copyright Statement

i. The author of this thesis (including any appendices and/or schedules to this thesis) owns certain copyright or related rights in it (the "Copyright") and s/he has given The University of Manchester certain rights to use such Copyright, including for administrative purposes.

ii. Copies of this thesis, either in full or in extracts and whether in hard or electronic copy, may be made **only** in accordance with the Copyright, Designs and Patents Act 1988 (as amended) and regulations issued under it or, where appropriate, in accordance with licensing agreements which the University has from time to time. This page must form part of any such copies made.

iii. The ownership of certain Copyright, patents, designs, trademarks and other intellectual property (the "Intellectual Property") and any reproductions of copyright works in the thesis, for example graphs and tables ("Reproductions"), which may be described in this thesis, may not be owned by the author and may be owned by third parties. Such Intellectual Property and Reproductions cannot and must not be made available for use without prior written permission of the owner(s) of the relevant Intellectual Property and/or Reproductions.

iv. Further information on the conditions under which disclosure, publication and commercialisation of this thesis, the Copyright and any Intellectual Property and/or Reproductions described in it may take place is available in the University IP Policy (see http://www.manchester.ac.uk/library/aboutus/regulations) and in The University's policy on Presentation of Theses.

Acknowledgements

First, I would like to thank my main supervisor Professor Perdita Barran for giving me the opportunity to experience a unique chapter of my life. She has always been helpful and believed in me when I'd often doubt myself and especially when I was working in Singapore her support and insight was paramount to the work, I did out there.

I would also like to thank Dr Verma Chandra for the chance to work in his group out in Singapore at the A*STAR research institute. I would also like to thank A*STAR for the chance to live in another country on the other side of the world, as not many people get opportunities like that, and it certainly was an experience and changed me significantly as a person and scientist.

I would like to thank all the members of the PBRG group who helped me especially, Dr Aiden France for being the man to teach me all the ins and outs of IM-MS and how to be a PhD student, Dr Lukas Migas for all the fantastic insight into the world of programming, and (soon to be Doctor!) Emma Norgate for all the help and guidance with experimental work upon my return from Singapore. I'd like to thank Jack for all the great banter and laughs, Rachelle for being a reliable source of information whenever I never knew what was going on and being as sarcastic as me. Big thanks to Reynard for being a great support with all instrument related ailments, same to Lenny for having someone to talk computer science with. Everyone else who was there for a chat or advice and to be a friend, Ellie, Rosie, Alina, Bruno, Emily, Caitlin, Xudong, Michael, Alexey, Jo, Kath, Ellen, Ruth, Luca, Kamila, Jakub, Depanjan, Emilia, Niklas, Drupad and everyone else I met through the group thank you.

To the friends I made in Singapore, I'd like to thank Dr Lauren Reid and Dr Alister Boags for being my two closest friends out there and helping me understand the wild world of

molecular dynamics. I'd also like to thank Dr. Stephen Fox for helping me develop my programmers mind by teaching me how to think logically and for just being a lovely guy. I'd also like to thank Kok Mei from HR who I'm certain protected me from a lot of boring form filling and for helping me get set up in Singapore. I'd also like to thank my house mates Dale, Lorena and Jurga who even though I'm sure we drove each other crazy I always knew we could count on one another when it meant the most. I'd like to thank the rest of my A*STAR cohort, Dave, John, Judith, Alessandro, Eilish, Poonam, Ewan, Finn, Max, Jack, Meg, Ebru, Matt, Dagnia, Sonia, Ellie and others for being good friends and lunch buddies day in and day out for two years.

To my best friend James Curran, thanks for being you and getting me through tough times and being there when I needed you.

To all my old friends, Pete, Petroc, Zim, Ennis, Ben, Becky, Aisha, Oscar, Chiz, Aleks, Elliot, Dom, Kate, Jesse, Hamish, Joe and everyone else from the Surf Club even though our lives will most likely take us on different paths I hope we stay in contact. Thanks for being good friends and good people and thanks for being there for me when I needed you all.

To my Mum who was my rock when I was out in Singapore and always helped me to stay positive and stay strong thank you so much. To my dad as well, always checking in and giving words of encouragement and support thank you. To my Stepdad Steve for sending me pictures of the cats, loving and supporting me, and helping me move to the various houses I lived in throughout my PhD thank you. To my Step-mum Jo for all the love and support thank you.

Introduction to Thesis

i. Inspiration and Motivation

The importance and interest in Intrinsically Disordered Proteins (IDPs) cannot be understated, with many often being involved in intracellular regulation processes and other biological mechanisms. By extension they can be signifiers or even instigators of numerous diseases such as cancer or Parkinson's and Alzheimer's disease.[1,2] As such the growing interest in this 'dark proteome', a name coined due to the shortcomings of conventional methods of biophysical analysis to describe them has led researchers to taking new approaches to understanding these phenomenal molecules.[2–5] One such technique is Ion Mobility Mass Spectrometry (IM-MS) coupled with nano-electrospray ionisation (*n*ESI). A simple search on web of knowledge shows the ramping interest in using Mass Spectrometry to study IDPs (Figure 1).

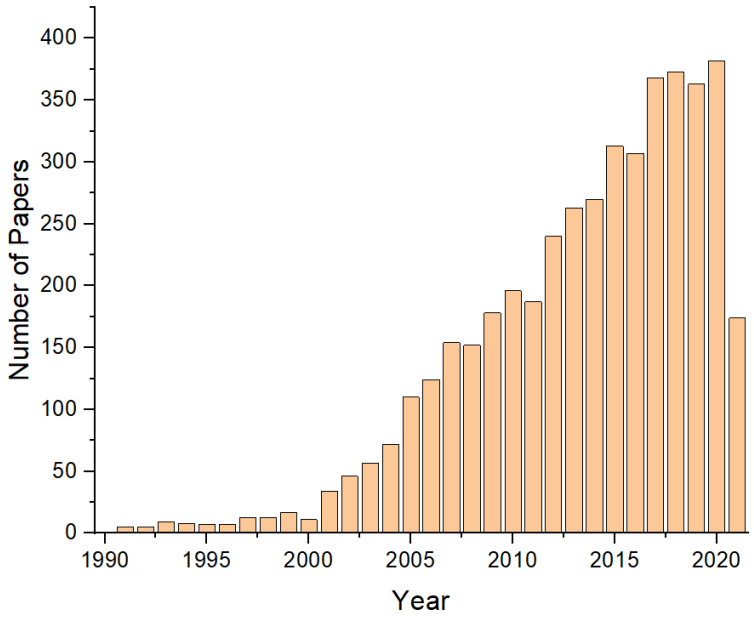

Figure 1 | Results of a Web of Knowledge search using the words disordered proteins and mass spectrometry in the topic field. Data is accurate as of June 2021.

ii. Brief Summary of Chapters

Chapter 1: The Study of Intrinsically Disordered Proteins by Ion Mobility Mass Spectrometry

The first chapter in this thesis aims to highlight much of the recent work carried out using this technique to study IDPs in detail, as well as comment on the challenges in the field and the direction it is heading.

Chapter 2: Methodology – The Use of Mass Spectrometry and Molecular Dynamics to Examine IDPs

Following this, the second chapter is focused on the methodology we apply to our study of IDPs. The cold regulating protein known as COR15A found in *Arabidopsis Thaliana* is used as a vector for this analysis which we compare to the behaviour of Myoglobin. The chapter covers sample preparation, purification, optimisation of mass spectrometry conditions, tuning parameters, data collection and analysis in detail. In addition, discussion of a 'toy' model that can predict the range of conformational space a protein can occupy is also included.

Chapter 3: Comparing MD Simulations with IM-MS Data Determines the Secondary Structural Features in the Conformational Landscape of the Cold Regulating Plant Protein COR15A

Chapter 3 centres on the comparison of Molecular Dynamics (MD) simulations with IM-MS Data to determine the secondary structural features present in COR15A's conformational landscape. The study showed us how combining these techniques can give us atomistic scale candidate structures for different conformations that exist within the COR15A system. Furthermore, it allows us to discriminate against thermodynamically unstable structures that are generated in MD space, narrowing down the amount of conformational space that needs to be explored. Lastly it highlighted how IM-MS gas phase structures are representative of

the dehydrated state of COR15A. The work carried out was also compared to work done by Thalhammer *et al.*[6] using Small Angle X-ray Scattering (SAXS) with Ensemble Optimisation Method (EOM) modelling, which reported similar structural features and regularity.

Chapter 4: Exploring the Conformational Space of the Cold Regulating Plant Protein COR15A at Differing Temperatures

Moving on, the next logical step was to continue with the analysis of COR15A by looking at how it behaved under cold temperatures which is what is covered in Chapter 4. Here we used Variable Temperature IM-MS to investigate the conformational changes to COR15A in temperatures ranging from 193 K – 298 K. The work demonstrated how VT IM-MS can provide molecular scale detail on the cold denaturation of the COR15A protein and discusses how there are currently few methods that can adequately do this. The study showed that overall the protein was in line with known theory on a rigid ions behaviour in cold temperature which was reported by Bowers *et al.*[7] in their work on Fullerene. However, deviation from this behaviour was seen at 270 K which models like the Proximity Superposition Approximation (PSA) do not predict and we proposed a hypothesis for this and comment on the need for further experimentation on cold denaturation to be carried out. Lastly, we highlight how VT IM-MS will play a significant role in future study of cold denaturation.

Chapter 5 – Conclusions, Outlook and Future Work

The fifth and final chapter looks back over the whole thesis, drawing conclusions and discussing what has been done and what future work can be done after this thesis ends.

iii. References

1 J. Ananiev, G. Tchernev, J. W. Patterson, M. Gulubova and G. Ganchev, *Acta Medica Bulg.*, 2011, **38**, 72–82.

2 P. Sormanni, D. Piovesan, G. T. Heller, M. Bonomi, P. Kukic, C. Camilloni, M. Fuxreiter, Z. Dosztanyi, R. V. Pappu, M. M. Babu, S. Longhi, P. Tompa, A. K. Dunker, V. N. Uversky, S. C. E. Tosatto and M. Vendruscolo, *Nat. Chem. Biol.*, 2017, 13, 339–342.

3 C. Navarro-Retamal, A. Bremer, J. Alzate-Morales, J. Caballero, D. K. Hincha, W. González and A. Thalhammer, *Phys. Chem. Chem. Phys.*, 2016, **18**, 25806–25816.

4 P. Tompa, *Curr. Opin. Struct. Biol.*, 2011, 21, 419–425.

5 D. Stuchfield and P. Barran, *Curr. Opin. Chem. Biol.*, 2018, 42, 177–185.

6 K. Shou, A. Bremer, T. Rindfleisch, P. Knox-Brown, M. Hirai, A. Rekas, C. J. Garvey, D. K. Hincha, A. M. Stadler and A. Thalhammer, *Phys. Chem. Chem. Phys.*, 2019, **21**, 18727–18740.

7 T. Wyttenbach, G. Von Helden, J. J. Batka, D. Carlat and M. T. Bowers, *J. Am. Soc. Mass Spectrom.*, 1997, **8**, 275–282.

Chapter 1: The Study of Intrinsically Disordered Proteins by Ion Mobility Mass Spectrometry

1.1 Declaration

This introductory chapter is reproduced from a published literature review I wrote:

D. Stuchfield and P. E. Barran, "Unique Insights to intrinsically disordered proteins provided by ion mobility mass spectrometry", In *Current Opinion in Chemical Biology,* Vol. 42, pp. 177-185.

As the first author of the article, I drafted and edited the manuscript as well as attained permission to reproduce all figures form other people's work. The article in its published form can be found in Appendix A.

Unique Insights to the Structures of Disordered Proteins Provided by Ion Mobility Mass Spectrometry

Dale Stuchfield and Perdita Barran*

Michael Barber Centre for Collaborative Mass Spectrometry, Manchester Institute for Biotechnology, School of Chemistry, University of Manchester, 131 Princess Street, Manchester M1 7DN, United Kingdom.

To Whom Correspondence should be addressed: perdita.barran@manchester.ac.uk

1.2 Abstract

Entire functional proteins as well as large regions of proteins lack structural elements which are resolvable via crystallography. These intrinsically disordered proteins (IDPs) or regions (IDRs) are often involved in cell regulation processes, for example in signalling hubs and as a result aberrant behaviour can cause or be representative of disease. As a consequence, there is a pressing need to develop alternative structural methods for IDPs and the interactions that they may form with other proteins and/or with potential inhibitors of binding. One such method is Ion Mobility Mass Spectrometry (IM-MS) coupled with careful application of electrospray ionisation, which shows great promise as a technique that does not 'care' if a protein is structured or not. We highlight recent work which has employed IM-MS to study conformational heterogeneity in disordered proteins, and discuss the opportunities, as well as the challenges of this approach.

1.3 Introduction

Intrinsically disordered proteins defined as possessing function from a disordered structure can perhaps be considered as sitting on one end of a structure to disorder continuum.[1–3]It is convenient to define five different classifications of protein structure from fully structured to unstructured (Figure 1.1)

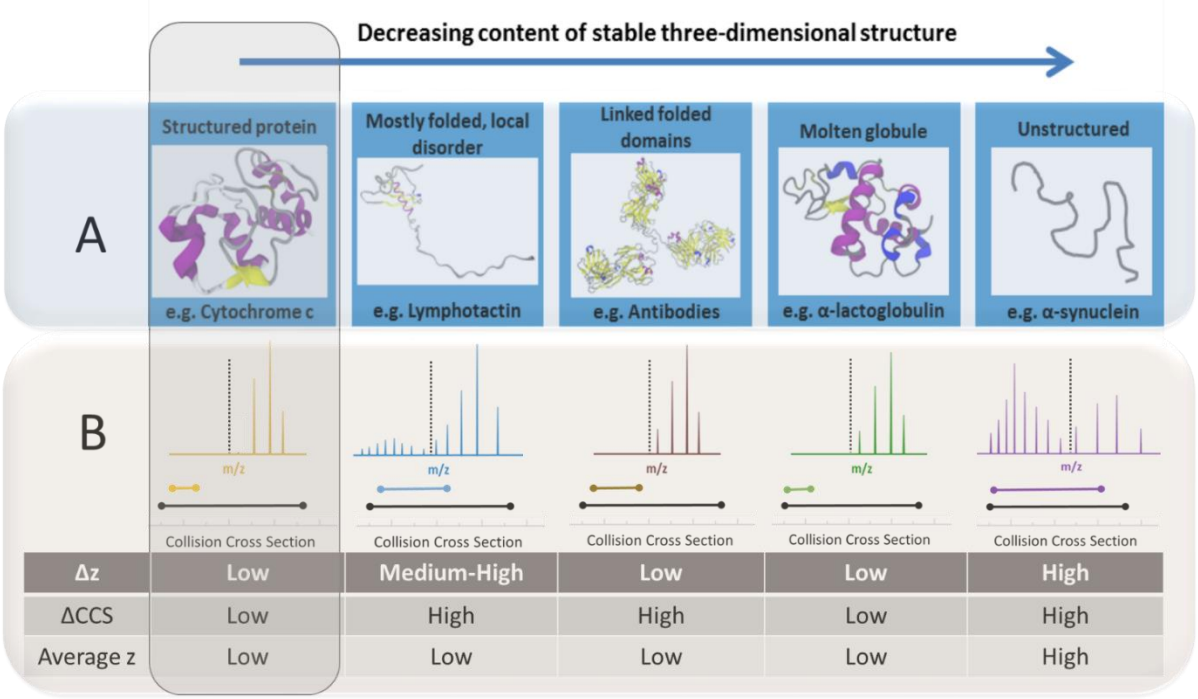

Figure 1.1 | Progression of protein structure from stable three-dimensional structure to unstructured and the corresponding ion mobility mass spectrometry observables. A. We have adapted a representation first shown by Dyson and Wright4 to highlight the continuum of protein structure from ordered to disordered. In our version we illustrate this with proteins that have been examined with mass spectrometry.5–9 From left to right the exemplar proteins are the electron transport protein Cytochrome c (PDB; 3CYT), readily able to form a crystal structure; the metamorphic protein Lymphotactin (PDB; 1J8I) for which NMR provides some information regarding the location and structure of the disordered N and C termini present in both the monomeric and dimeric forms10; a monoclonal human antibody - immunoglobulin G (PDB; 1IGT) where structured regions rich in β-sheet are connected by flexible disordered linking regions8,11]; α-lactoglobulin (PDB; 1HMK) which is a molten globule consisting of a compact form with mobile side chains. This state was first reported for cytochrome c12 highlighting the progression of structure that can be adopted by proteins; and finally α-synuclein

a highly disordered amyloid forming protein associated with the progression of Parkinson's Disease.13,14 B: Schematic to show the read out from mass spectrometry (m/z) following electrospray ionisation under gentle conditions that act to preserve solution structure(s). The charge state distributions for each class of protein are shown, with each peak on the mass spectrum representing a different charge state in descending order going from left to right along the m/z axis. The corresponding spread in collision cross section values (coloured lines) and the predicted possible range of collision cross section values (black lines) based on a toy model described previously, which predicts the most compact and most extended CCS values for a given linear protein sequence, are also shown.15 The black dotted line on the schematic mass spectra shows the upper value for the net charge (z) on the protein that would correspond to a globular/folded form according to de la Mora16, it can be seen for structured proteins and molten globules the mass spectrometry data falls below this upper limit, whereas for partially or completely disordered proteins much of the signal intensity is above this line indicative of extended forms.

With increased recognition of the existence and importance of IDPs[17] there are reports which predict between 30-70% of all Eukaryotic proteins are either IDPs or contain intrinsically disordered regions (IDRs).[18–20] Recent experimental work by Picotti et al. supports the lower end of this spread[21], regardless, for IDPs the central dogma that relates one structure to a given function no longer holds and it is more appropriate to consider IDPs as structural ensembles with promiscuous functions. IDPs have been shown to fold upon binding[22,23] and perhaps more generally to form fuzzy complexes with a myriad of sampled interactions and hence active conformational states.[24,25] Since IDPs are implicated in many diseases, including neurodegenerative diseases and cancer, it is important to gain better understanding of their structure(s) function(s) relationships. The lack of resolvable structure reduces the effectiveness of crystallography for this class of protein, and IDPs are more commonly examined with NMR, SAXS, and other spectroscopies.[26] Ion Mobility Spectrometry with Mass

Spectrometry has also been applied to the study of IDPs and shows some promise with its capability to resolve structure and stoichiometry, and also because it reports the net number of charges on the molecule under examination in the gas phase. Though in cases this may not necessarily be reflective of solution phase charge state.

1.4 Ion Mobility Mass Spectrometry (IM-MS)

An IM-MS instrument carries out five basic operations: sample introduction, compound ionisation, ion mobility separation, mass separation, and ion detection. There are a variety of methods employed to form gas-phase ions for IM-MS analysis[27], but for proteins the preeminent method is electrospray ionisation (ESI)[28] and in order to best preserve non covalent interactions from solution the related technique of nano-ESI[29–31] is most commonly used. Four methods are commonly available for ion mobility separation: aspiration ion mobility spectrometry (AIMS)[32] differential-mobility spectrometry (DMS) also known as field-asymmetric waveform ion mobility spectrometry (FAIMS)[33], drift-tube ion mobility spectrometry (DTIMS)[27] and travelling-wave ion mobility spectrometry (TWIMS).[34] For the purpose of this review we will only consider the latter two since they have had the largest application for IDPs.[15,26,35–38]

1.5 IM-MS to study Amyloid forming IDPs

Over the past 15 years a number of studies have used IM-MS to examine the early stages of the aggregation of peptides which go on to form amyloid fibrils.[39–42] Bowers[41,43–45], and Ashcroft[42,46,47] amongst the most prolific authors. Whilst amyloid forming peptides and proteins are often also categorised as IDPs the focus of these excellent studies has principally been to understand the assembly mechanisms and to develop MS methods that can help to screen for inhibitors of assembly. A number of researchers have focused more on the use of

mass spectrometry to study the intrinsic structure of IDPs for example, Grandori et al.[48] who used experiment and theory to explain the charge state distributions (CSD) obtained following ESI-MS of disordered proteins. They conclude whilst proton transfer does occur during electrospray and/or drift, that structural rearrangements associated with this do not have a substantial effect on structure. From this is it suggested that the CSD does indeed provide a convenient 'read out' of the solution phase form as indicated in Figure 1.1 B.

α-synuclein has been extensively studied with mass spectrometry as well as other techniques due to its significance in the progression of Parkinson's Disease as well as its use as an exemplary IDP for biophysical studies.[43,49,50] In 2015 Phillips et al.[9] used hybrid MS approaches including IM-MS to reveal the conformational flexibility of α-synuclein.[9] These experiments revealed the presence of multiple conformational families. Evidence for this was found in the large range of $^{DT}CCS_{He}$ values (~1800 Å^2) found for the monomeric form of α-synuclein which presents over a correspondingly large range of charge states $\Delta z = 15$ ($5 \leq z \leq 21$). Chemical crosslinking was used in conjunction with IM-MS to capture transient conformational populations, demonstrating how it was possible to trap solution phase species prior to gas phase analysis. This revealed three stable conformational families that made up the α-synuclein ensemble which were assigned as compact (~1200 Å^2), extended (~1500 Å^2) and unfolded (~2350 Å^2). Beveridge et al.[15] compared α-synuclein with Apolipoprotein C-II (ApoC-II), a plasma protein that has a role in cardiovascular disease known to form amyloid fibrils.[51] HDX-MS experiments confirmed that both proteins were indeed conformationally dynamic in solution since high levels of exchange were seen on transfer to the gas phase nESI-IM-MS experiments for ApoC-II yielded $\Delta z = 4$, and a low conformational spread (579 Å^2 compared to α-synuclein 1800 Å^2) irrespective of pH and solution conditions. Use of a toy model[15] showed

that α-synuclein explored approximately 90% of the collision cross section space available to it whilst ApoC-II only explored around 30%. An explanation for this anomalous result is that ApoC-II restructures as it desolvates, perhaps towards a helix rich structure akin to that in which it is found in membranes[52], or more likely to a well self-solvated compact globular state. Unlike structured proteins which can be characterised by wide CSDs when sprayed from solutions at low or high pH, ApoC-II shows little variation, which is also explained by restructuring upon transfer.

Borysik et al. also have demonstrated how IM-MS data of IDPs provides wider conformational distributions than those found in solution measurements.[26] They investigated a series of IDPs by ESI-IM-MS in a systematic comparison of solution and gas-phase structural space. One specific example was ERD10[53] a Dehydrin family protein examined using TW-IM-MS. Their results (Figure 1.2) were compared to SAXS data obtained in solution and led the authors to conclude that whilst the gas phase data for high charge state is representative of the ERD10 ensemble in solution, the lower charge states are indicative of collapsed states which are not evident from solution measurements.

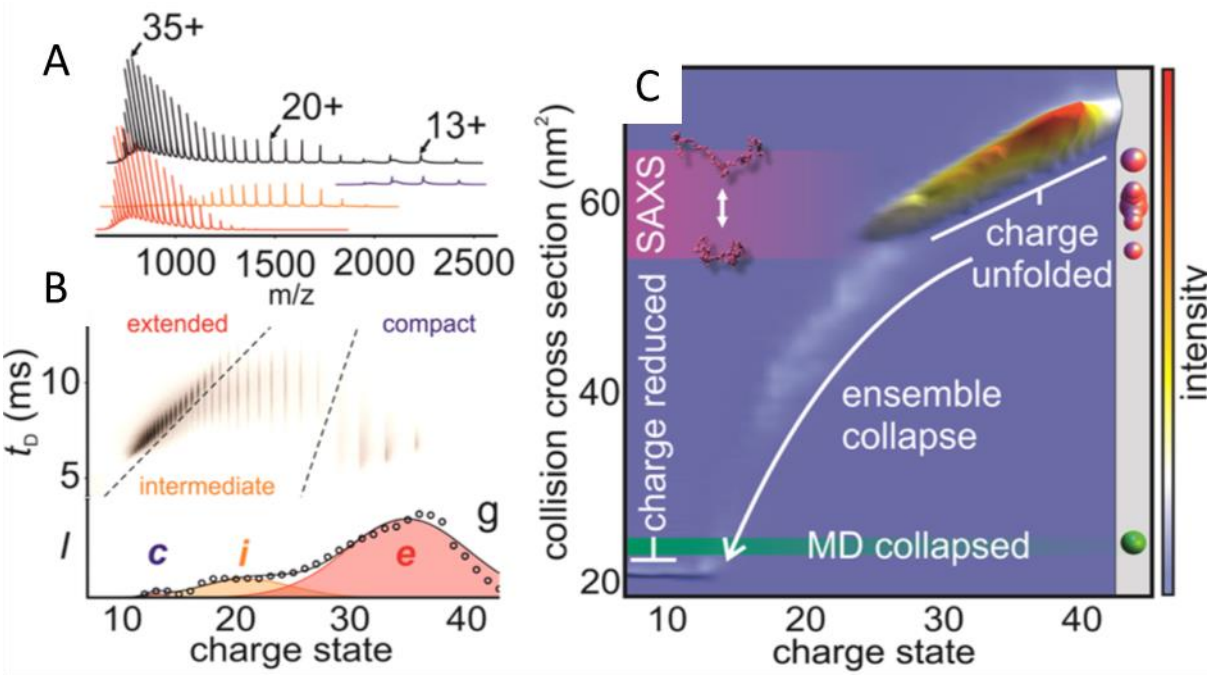

Figure 1.2 | IM-MS investigations of ERD10 adapted from Borysik et al.[26] . A: Mass spectrum of ERD10 (black) with a charge state distribution ranging 12 ≤ z ≤ 43. Extracted mass spectra for the extended (red) intermediate (orange) and compact (blue) subpopulations. B: Upper part of graph shows the mobiligram of ERD10 with drift times of three distinct regions, the lower portion of the graph shows the linear deconvolution of the ERD10 charge state distribution displaying 3 subpopulations of the protein. C: Here CCS of ERD10 obtained from IM-MS and the SAXS ensemble are plotted against charge and intensity. This allowed for a direct comparison of the conformational space of the protein in both the solution gas-phase. Features relating to charge unfolded and charge reduced ions are highlighted. The CCS for each structure obtained from the EOM solution ensemble of ERD10 is showed on the grey strip to the RHS of the figure as a bubble plot with bubble size representing relative weight of each conformation within the solution ensemble, the values range from 55 – 65 nm^2, whereas IMS gas-phase structures range from 20 – 70 nm^2 and are represented by the grey strip. The SAXS derived ensemble is represented by the pink strip and example structures are shown. Lastly the fully collapsed charge neutral species of ERD10 obtained from MD is given in green. The comparison highlights the great extent to which the conformational space of ERD10 expands in the absence of solvent effects. Gas-phase structures of ERD10 were further characterized by employing various strategies of charge reduction and thermal activation. Charge stripping of ERD10 to generate 7+ and 11+ ions of the protein saw no further decrease in CCS indicating the structures at 20 nm^2 were representative of fully collapsed gas-phase structures.

Studies such as those described above confirm that IDPs can be readily examined with IM-MS. As for structured proteins[54,55], the spread CCS values and the corresponding relative weight of conformational populations is not a precise recapitulation to that found in solution. This is no surprise, and perhaps is best understood as being due to the interplay between self-solvation versus coulombic repulsion of highly charged groups. Regardless, IM-MS data from IDPs following electrospray ionisation from Ammonium Acetate at neutral pH is distinctive from that of structured proteins, and hence this method can be used to assay the intrinsic conformational landscape. IM-MS is also useful in determining how such a landscape alters in the presence of binding partners.

1.6 IM-MS to study conformational modes of IDPs in protein:protein complexes

Whilst it is now commonplace to use native mass spectrometry methods to probe stoichiometry and structure of protein complexes, and the technique has become recognised as a tool for structural biology. It has long been noted that the collision cross sections of gas phase proteins readily range from smaller to larger values than those expected if the crystal structure was preserved unaltered into the mass spectrometer.[8,55–58]This observation is also the case for IM-MS investigations of IDPs, where the flexibility of the protein sequence is perhaps more likely to result in the conformation being altered as the protein transfers to the gas phase.

Grandori and co workers[35] used experimental data obtained from ESI-IM-MS measurements to impose constraints on molecular modelling simulations of the polypeptide segment between the ID NTAIL domain and the phosphoprotein X domain, PXD of the measles virus, MeV(NTAIL – PXD). IM-MS data indicated the structural diversity that is inherent to the

complex and along with data from other techniques, informed constrained modelling simulations. The models indicated that most interactions which characterise the conformational ensemble are hydrophobic in nature an observation supported by SASA values for solution and gas phase models for the open form. As with the ERD10 study (Figure 1.2)[26] they conclude that the higher charge states in the MS spectra are better representative of the solution ensemble and that a compact state which is present in relatively high abundance in the IM-MS data may be representative of a subgroup of the solution ensemble or possibly a gas phase artefact following ESI.

1.7 IM-MS to study transcription factors – drugging the undruggable

The conformational heterogeneity of IDPs and the corresponding difficulty in obtaining crystal structures renders them problematic for classical drug discovery workflows, where computational dock and score strategies start from solved structures. NMR is used to good effect to probe details of protein-protein interactions (PPIs) in IDPs, however it is sample hungry and not best suited to high throughput use. Due to the sensitivity of ESI coupled to IM-MS it has the potential for use as a high-throughput screening assay for PPI inhibitors. An example of this was work done by Harvey et al.[59] which focused on the transcription factor c-MYC. Following heterodimerisation with its binding partner MAX, MYC proteins regulate the expression of 10-15% of all genes by the recruitment of co-factors. MYC proteins contain a basic helix-loop-helix leucine zipper (bHLHZip) domain which binds to a similar region on MAX.[60,61]The bHLHZip regions of c-MYC and MAX are intrinsically disordered until they dimerise, when they form a parallel left handed four helix bundle with each monomer composed of two α-helices separated by a loop. Targeting c-Myc to disrupt the formation of

the heterodimer (i.e., to inhibit the protein-protein interaction c-MYC:MAX) is a potential strategy to down-regulate c-MYC transcriptional activity.[62]

Hammoudeh et al. reported several c-MYC inhibitors, of low μM affinity, that they claimed stabilise the disordered state, preventing MAX dimerization. Harvey et al. used IM-MS to examine the binding of a lead inhibitor 10058-F4[59] looking at two peptide models of this PPI one of which included four amino acids purported to form the site of ligand binding (referred to as c-MYC-zipDT) and one without (c-MYC-zipΔDT). Upon incubation with the ligand, no c-Myc:ligand complex could be seen in the mass spectrum of either form of the peptide, but the relative intensity of the +5:+4 charge state of the c-MYC:MAX charge state decreased. Inspection of the arrival time distributions of this charge state of the complex revealed that in the presence of the inhibitor an extended form of the c-MYC:MAX complex is removed (Figure 1.3). The $^{DT}CCSD_{He}$ of this extended form corresponds remarkably well to that from a helix:helix form of the complex obtained from MD simulations starting from an NMR structure 1A93[63], on ligand incubation the only remaining form of the c-MYC:MAX complex corresponded to a disordered and collapsed complex (Figure 1.3 B). This indicated that despite a binding interaction too weak to be preserved into the gas phase, the ligand did indeed modulate the binding interaction of c-MYC:MAX preventing the formation of the structured leucine zipper that is the functional form that subsequently binds to DNA. The similarity of the data between the two forms of c-MYC (Figure 1.3 A versus B) suggests that the postulated drug binding site is not as indicated.

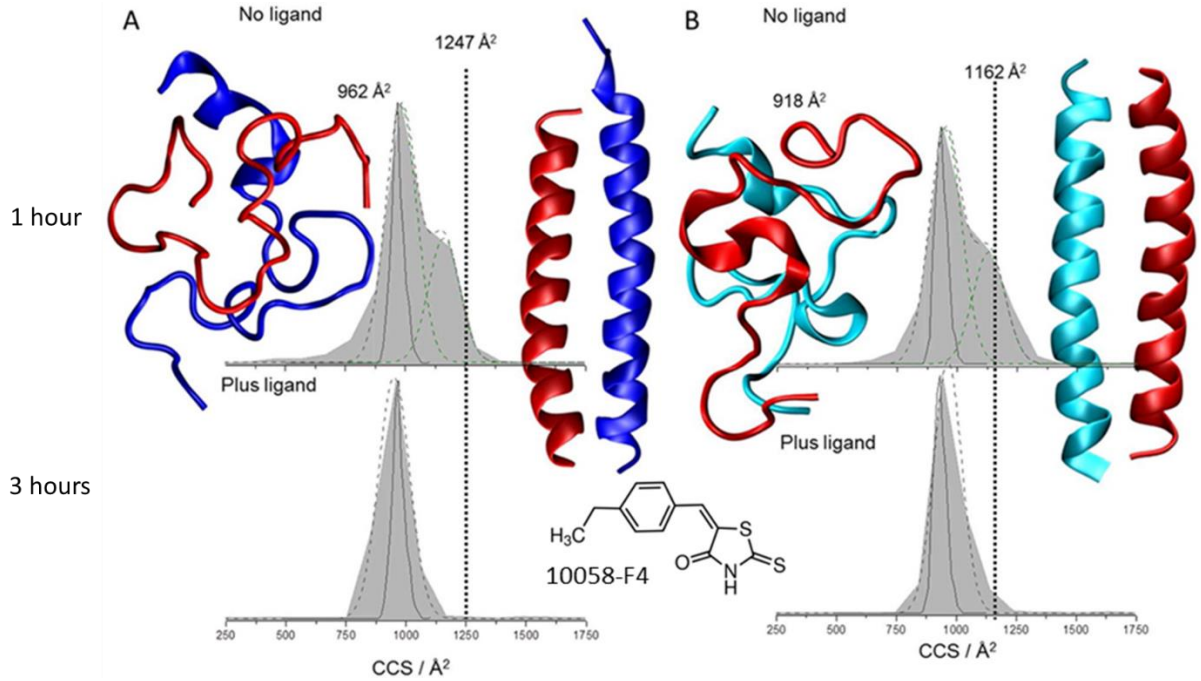

Figure 1.3 | The use of IM-MS to examine conformational rearrangement in the presence of small molecules. Adapted from Harvey et al.[59] $^{DT}CCSD_{He}$ derived from arrival time distributions at a drift voltage of 45 V for the $[M+5H]^{5+}$ ions from c-MYC:MAX complexes are shown in the absence of ligand (after 1 h incubation at 37 °C) and in the presence of the ligand 10058-F4 (after 3 h incubation at 37 °C) at the top and bottom panels, respectively. Profiles for the predicted conformation from MD simulated annealing (solid lines) and from NMR structures (vertical dotted line) are displayed together with experimental fitted curves (dashed lines: black being total CCSD and green charge state specific CCSD). 3D molecular models for the collapsed and extended forms of the c-MYC:MAX complex are also given. (A) c-MYC-zip:MAX-Zip and (B) c-MYC-ZipΔDT:MAX-Zip.

The multi domain transcription factor p53 possessing high levels of disorder has also been the subject of IM-MS investigation.[57,64–67] Using similar methods as described above for c-MYC, Dickinson et. al characterized the effect of two inhibitors, RITA and Nutlin-3 on their respective binding partners: p53 and MDM2. RITA binds the N-terminal transactivation domain of p53 (Np53) weakly, preventing direct observation of the complex in the gas phase, but as with c-MYC and 10058-F4, on incubation with RITA, both the CSD and the $^{DT}CCSD_{He}$

alter. This finding supports a hypothesis that RITAs mode of action proceeds via a conformational change in p53.

1.8 Towards a framework to determine the extent of protein disorder from IM-MS experiments

Drawing from the growing number of mass spectrometry studies that have investigated IDPs we sought to consider if ESI-IM-MS can be used to differentiate the conformational character of both ordered and disordered proteins.[15] IM-MS data from 26 proteins, some of which were known to be disordered and some structured were considered in terms of the distribution of charge states (CSD Δz) and the distribution of collision cross sections (ΔCCS). For example, the structured protein Cytochrome c with its narrow CSD has a Δz = 4. They noted that for all structured proteins, when sprayed from Ammonium Acetate at a pH close to the pI of the protein, the charge states lie below the Rayleigh limit introduced by de La Mora (Equation 1) which indicates that they are globular[16] and remarkably Δz is almost invariant with the mass of the protein.

$$Z_R = 0.0778\sqrt{m} \quad (1)$$

where Z_R is the maximum charge and m is the mass of the protein. This contrasted with data from disordered proteins, for example, β-casein with a CSD ranging from $7 \leq z \leq 27$ and a considerable number of these charge states are above the Rayleigh limit indicating they are likely to be extended conformations. IDPs also had large ΔCCS cf. their structured counterparts and it was concluded that IM-MS data is more accurate than hydropathy plots for assessing the extend of disorder within a protein system.[68]

A simple toy model was introduced which predicts the smallest CCS that might be reasonably be expected for a given monomeric protein by an extrapolation from the linear sequence and

density from crystal structure data, and the largest CCS value, here based on extending the entire sequence. It is notable that despite the noted collapse in structure for low charge states, especially for IDPs, for all proteins measured the lowest predicted CCS value is always lower than that measured. As can be seen in Figure 1.4 where relative intensities of the different charge states for three example proteins are plotted with respect to the average CCS value they span and compared to the toy model data (red line). This allows the construction of model energy landscapes (Figure 1.4 RHS). A protein that populates a wide range of CCS values, with similar intensities of ions across the entire range will denote a polypeptide chain sampling many different conformations with little preference for any particular one.

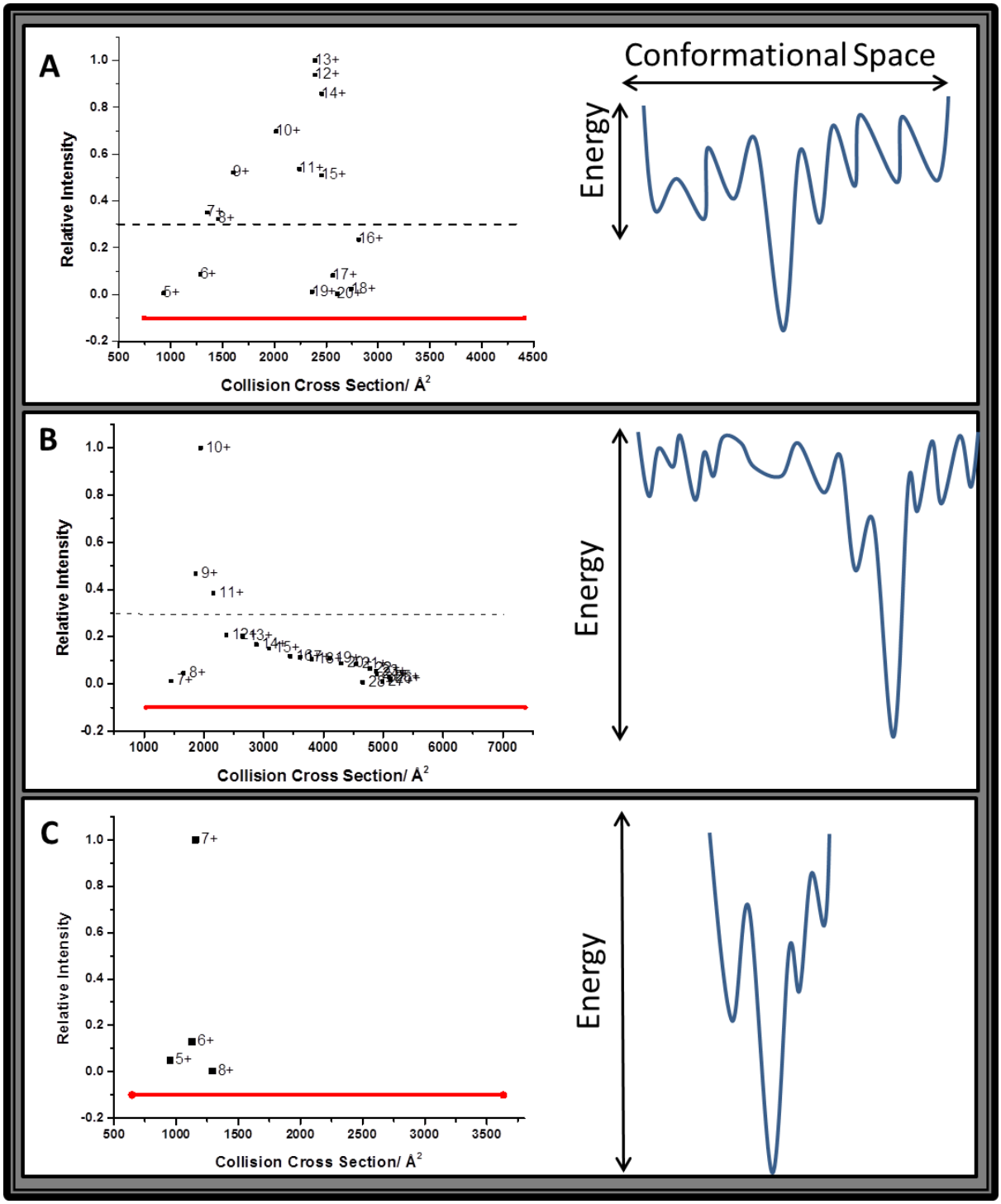

Figure 1.4 | Adapted from [15]. Shown in red is the theoretical CCS space calculated from the toy model that a protein can occupy. The relative intensities of differing charge states are shown above with a dashed line showing 30 % intensity of the base peak. The relative intensities of the differing charge states provide information on the extent of structure and disorder within the system. Plotting the CSD along with CCS data allows us to construct postulated energy landscapes as seen on the RHS of the figure. A: α-synuclein is seen to

sample a wide range of the theoretically possible CCS, with similar intensities of ions across this range (5+ ≤ *z* ≤ 20+). Nine of these charge states have more than 30 % intensity of the base peak (dashed line Figure 5A). This behaviour is indicative of a polypeptide chain sampling many different conformations, with little preference for any particular one. B: β-casein is seen to have the most intense peaks corresponding to the $[M+9H]^+$, $[M+10H]^+$, and $[M+11H]^+$, which are of low charge and CCS (1862 – 2157 $Å^2$) relative to the majority of the peaks. These correspond to the energetically stable form of the protein, represented by the troughs in the postulated energy landscape shown in Figure 5B. The peaks of higher charge and larger CCS are much lower intensity, indicative of a dynamic ensemble of transient species. Only three charge states are above 30 % intensity base peak (dashed line Figure 5B). C: Cytochrome C has a narrow charge state distribution with the dominant species [M+7H]+ having a low CCS value compared to the theoretical CCS values available to the protein chain. This corresponds to the narrow folding funnel with one energetically favourable conformational family observed for Cytochrome C demonstrating its highly structured nature.

1.9 Conclusion and outlook

Working from a growing body of literature that uses mass spectrometry methods to examine IDPs it is apparent that new insights to the intrinsic structure of these important proteins can be found. These have the potential to assist our understanding of how these proteins act in disease as well as support drug discovery. Also of note is the use of IM-MS methods to understand the fundamental aspects of IDP form. The framework model described above has attractive features, but is highly simplistic and does not, for example, explain the data found for Apo-CII.[69] Clearly the ESI process can influence the CSD and perhaps also the range of conformations that are realised in the gas phase for any protein and most markedly for IDPs. There is certainly evidence for a change in the weights of the populations of solution phase ensembles[35,48,55,65] as they are transferred to the gas phase with a dominance of compact states that appear not well sampled in solution. This may simply be because of desolvation on IDPs, which commonly have more charged residues and less hydrophobic residues than

their structured cousins. As solvent and salt is removed coulombic effects are strengthened[6] and whilst this can cause unfolding to increase due to repulsion of proximal charges it can also favour compact states for low charge states as basic and acidic groups seek to pair up.

The relationship between net charge and charge residue location has been shown to influence the conformational landscape of IDPs[70] and given the ability of IM-MS to measure net charge we expect that findings from the gas phase will be instructive here. We have started to investigate how ion-molecule charge reduction reactions between radical anions and protein cations regulate protein conformation.[71,72] This study found that the addition of a single electron can cause drastic changes to a proteins conformation for highly populated charge states when sprayed from salty aqueous solution, analogous to findings with proton transfer reagents.[73] Both approaches have shown that the CCSD of proteins of low charge states can be found by charge reduction from conformers at higher charge states.

Knowledge of the active conformation and interaction partners of proteins and their complexes is paramount to understanding their intercellular function. Many protein systems are dynamic with respect to both conformation and interaction, so during a proteins 'lifespan' it will sample many configurations and bind to several targets. This is particularly true for IDP's where is it is desirable to measure the range of structures an IDP can occupy, and also the effect of selected binding partners on altering this conformational space. The solvent free environment of a mass spectrometer is ideally suited to examine intrinsic interactions and consider how they contribute to structure, and we envisage it will continue to be an attractive route for the study of IDP's.

1.10 Acknowledgements

I would like to acknowledge the hard work and insights given from many former and current Barran group members in our study of IDPs: Jason Kalapothakis, Peter Faull, Harriet Cole, Max Porrini, Ewa Jurnezcko, Sophie Harvey, Jakub Ujma, Rebecca Beveridge, Eleanor Dickinson, Ashley Phillips, Chris Nortcliffe, and Jacquelyn Jhingree. We also gratefully acknowledge our many collaborators in this area without whose beautiful and challenging samples we might only look at Ubiquitin.

1.11 References

1 P. Tompa, *Curr. Opin. Struct. Biol.*, 2011, 21, 419–425.

2 G. W. Daughdrill, L. J. Hanely and F. W. Dahlquist, *Biochemistry*, 1998, **37**, 1076–1082.

3 J. J. Ward, J. S. Sodhi, L. J. McGuffin, B. F. Buxton and D. T. Jones, *J. Mol. Biol.*, 2004, **337**, 635–645.

4 P. E. Wright and H. J. Dyson, *J. Mol. Biol.*, 1999, **293**, 321–331.

5 Y. Mao, J. Woenckhaus, J. Kolafa, M. A. Ratner and M. F. Jarrold, *J. Am. Chem. Soc.*, 1999, **121**, 2712–2721.

6 E. R. Dickinson, E. Jurneczko, K. J. Pacholarz, D. J. Clarke, M. Reeves, K. L. Ball, T. Hupp, D. Campopiano, P. V. Nikolova and P. E. Barran, *Anal. Chem.*, , DOI:10.1021/ac503720v.

7 S. R. Harvey, M. Porrini, A. Konijnenberg, D. J. Clarke, R. C. Tyler, P. R. R. Langridge-Smith, C. E. Macphee, B. F. Volkman and P. E. Barran, *J. Phys. Chem. B*, 2014, **118**, 12348–12359.

8 K. J. Pacholarz, M. Porrini, R. A. Garlish, R. J. Burnley, R. J. Taylor,

A. J. Henry and P. E. Barran, *Angew. Chemie Int. Ed.*, 2014, **53**, 7765–7769.

9 A. S. Phillips, A. F. Gomes, J. M. D. Kalapothakis, J. E. Gillam, J. Gasparavicius, F. C. Gozzo, T. Kunath, C. MacPhee and P. E. Barran, *Analyst*, 2015, **140**, 3070–3081.

10 R. L. Tuinstra, F. C. Peterson, S. Kutlesa, E. S. Elgin, M. A. Kron and B. F. Volkman, *Proc. Natl. Acad. Sci.*, 2008, **105**, 5057–5062.

11 K. J. Pacholarz, S. J. Peters, R. A. Garlish, A. J. Henry, R. J. Taylor, D. P. Humphreys and P. E. Barran, *ChemBioChem*, 2016, **17**, 46–51.

12 M. Ohgushi and A. Wada, *FEBS Lett.*, 1983, **164**, 21–24.

13 A. Iwai, E. Masliah, M. Yoshimoto, N. Ge, L. Flanagan, H. A. Rohan de Silva, A. Kittel and T. Saitoh, *Neuron*, 1995, **14**, 467–475.

14 V. N. Uversky, *J. Biomol. Struct. Dyn.*, 2003, **21**, 211–234.

15 R. Beveridge, S. Covill, K. J. Pacholarz, J. M. D. Kalapothakis, C. E.

Macphee and P. E. Barran, *Anal. Chem.*, 2014, 86, 10979–10991.

16 J. Fernandez De La Mora, *Anal. Chim. Acta*, 2000, **406**, 93–104.

17 V. N. Uversky and A. K. Dunker, *Biochim. Biophys. Acta - Proteins Proteomics*, 2010, 1804, 1231–1264.

18 R. van der Lee, M. Buljan, B. Lang, R. J. Weatheritt, G. W. Daughdrill, A. K. Dunker, M. Fuxreiter, J. Gough, J. Gsponer, D. T. Jones, P. M. Kim, R. W. Kriwacki, C. J. Oldfield, R. V. Pappu, P. Tompa, V. N. Uversky, P. E. Wright and M. M. Babu, *Chem. Rev.*, 2014, **114**, 6589–6631.

19 P. Tompa, *Nat. Chem. Biol.*, 2012, **8**, 597–600.

20 J. Ananiev, G. Tchernev, J. W. Patterson, M. Gulubova and G. Ganchev, *Acta Medica Bulg.*, 2011, **38**, 72–82.

21 P. Leuenberger, S. Ganscha, A. Kahraman, V. Cappelletti, P. J. Boersema, C. von Mering, M. Claassen and P. Picotti, *Science (80-.).*, 2017, **355**, eaai7825.

22 M. Arai, K. Sugase, H. J. Dyson and P. E. Wright, *Proc. Natl. Acad.*

Sci., 2015, **112**, 9614–9619.

23 H. J. Dyson and P. E. Wright, *Curr. Opin. Struct. Biol.*, 2002, 12, 54–60.

24 P. Tompa and M. Fuxreiter, *Trends Biochem. Sci.*, 2008, **33**, 2–8.

25 M. Miskei, C. Antal and M. Fuxreiter, *Nucleic Acids Res.*, 2017, **45**, D228–D235.

26 A. J. Borysik, D. Kovacs, M. Guharoy and P. Tompa, *J. Am. Chem. Soc.*, 2015, **137**, 13807–13817.

27 A. B. Kanu, P. Dwivedi, M. Tam, L. Matz and H. H. Hill, *J. Mass Spectrom.*, 2008, 43, 1–22.

28 S. Nguyen and J. B. Fenn, *Proc. Natl. Acad. Sci. U. S. A.*, 2007, **104**, 1111–1117.

29 M. Wilm and M. Mann, *Anal. Chem.*, 1996, **68**, 1–8.

30 M. S. Wilm and M. Mann, *Int. J. Mass Spectrom. Ion Process.*, 1994, **136**, 167–180.

31 A. Konijnenberg, A. Butterer and F. Sobott, *Biochim. Biophys.*

Acta - Proteins Proteomics, 2013, 1834, 1239–1256.

32 D. Juan and J. Simon, *Rev. Mex. física*, 2012, **77**, 1–26.

33 I. A. Buryakov, E. V. Krylov, E. G. Nazarov and U. K. Rasulev, *Int. J. Mass Spectrom. Ion Process.*, 1993, **128**, 143–148.

34 K. Giles, S. D. Pringle, K. R. Worthington, D. Little, J. L. Wildgoose and R. H. Bateman, *Rapid Commun. Mass Spectrom.*, 2004, **18**, 2401–2414.

35 A. D'Urzo, A. Konijnenberg, G. Rossetti, J. Habchi, J. Li, P. Carloni, F. Sobott, S. Longhi and R. Grandori, *J. Am. Soc. Mass Spectrom.*, 2015, **26**, 472–481.

36 A. Natalello, C. Santambrogio and R. Grandori, *J. Am. Soc. Mass Spectrom.*, 2017, **28**, 21–28.

37 R. Beveridge, Q. Chappuis, C. Macphee and P. Barran, *Analyst*, 2013, 138, 32–42.

38 R. Beveridge, A. S. Phillips, L. Denbigh, H. M. Saleem, C. E. MacPhee and P. E. Barran, *Proteomics*, 2015, **15**, 2872–2883.

39 H. Cole, M. Porrini, R. Morris, T. Smith, J. Kalapothakis, S. Weidt, C. L. Mackay, C. E. MacPhee and P. E. Barran, *Analyst*, 2015, **140**, 7000–7011.

40 H. L. Cole, J. M. D. Kalapothakis, G. Bennett, P. E. Barran and C. E. MacPhee, *Angew. Chemie - Int. Ed.*, 2010, **49**, 9448–9451.

41 C. Bleiholder and M. T. Bowers, *Annu. Rev. Anal. Chem.*, 2017, **10**, 365–386.

42 D. P. Smith, S. E. Radford and A. E. Ashcroft, *Proc. Natl. Acad. Sci.*, 2010, **107**, 6794–6798.

43 M. Grabenauer, S. L. Bernstein, J. C. Lee, T. Wyttenbach, N. F. Dupuis, H. B. Gray, J. R. Winkler and M. T. Bowers, *J. Phys. Chem. B*, 2008, **112**, 11147–11154.

44 T. D. Do, W. M. Kincannon and M. T. Bowers, *J. Am. Chem. Soc.*, 2015, **137**, 10080–10083.

45 C. Bleiholder, N. F. Dupuis, T. Wyttenbach and M. T. Bowers, *Nat. Chem.*, 2011, **3**, 172–177.

46 L. M. Young, J. C. Saunders, R. A. Mahood, C. H. Revill, R. J. Foster, L. Tu, D. P. Raleigh, S. E. Radford and A. E. Ashcroft, *Nat. Chem.*, 2015, **7**, 73–81.

47 L. A. Woods, G. W. Platt, A. L. Hellewell, E. W. Hewitt, S. W. Homans, A. E. Ashcroft and S. E. Radford, *Nat. Chem. Biol.*, 2011, **7**, 730–739.

48 J. Li, W. Lyu, G. Rossetti, A. Konijnenberg, A. Natalello, E. Ippoliti, M. Orozco, F. Sobott, R. Grandori and P. Carloni, *J. Phys. Chem. Lett.*, 2017, **8**, 1105–1112.

49 C. Mensch, A. Konijnenberg, R. Van Elzen, A. M. Lambeir, F. Sobott and C. Johannessen, *J. Raman Spectrosc.*, 2017, **48**, 910–918.

50 A. Konijnenberg, S. Ranica, J. Narkiewicz, G. Legname, R. Grandori, F. Sobott and A. Natalello, *Anal. Chem.*, 2016, **88**, 8468–8475.

51 D. M. Hatters, C. E. MacPhee, L. J. Lawrence, W. H. Sawyer and G. J. Hewlett, *Biochemistry*, 2000, **39**, 8276–8283.

52 Y.-F. Mok, C. L. Teoh, G. J. Howlett and M. D. W. Griffin, in *Bio-nanoimaging*, Elsevier, 2014, pp. 247–254.

53 D. Kovacs, E. Kalmar, Z. Torok and P. Tompa, *PLANT Physiol.*, 2008, **147**, 381–390.

54 E. Jurneczko, F. Cruickshank, M. Porrini, P. Nikolova, I. D. G. Campuzano, M. Morris and P. E. Barran, *Biochem. Soc. Trans.*, , DOI:10.1042/BST20120125.

55 E. Jurneczko and P. E. Barran, *Analyst*, 2011, **136**, 20–28.

56 K. Pagel, E. Natan, Z. Hall, A. R. Fersht and C. V. Robinson, *Angew. Chemie - Int. Ed.*, 2013, **52**, 361–365.

57 E. Jurneczko, F. Cruickshank, M. Porrini, D. J. Clarke, I. D. G. Campuzano, M. Morris, P. V. Nikolova and P. E. Barran, *Angew. Chemie - Int. Ed.*, 2013, **52**, 4370–4374.

58 R. Beveridge, L. G. Migas, K. A. P. Payne, N. S. Scrutton, D. Leys and P. E. Barran, *Nat. Commun.*, 2016, **7**, 12163.

59 S. R. Harvey, M. Porrini, C. Stachl, D. MacMillan, G. Zinzalla and

P. E. Barran, *J. Am. Chem. Soc.*, 2012, **134**, 19384–19392.

60 N. M. Sodir and G. I. Evan, *J. Biol.*, 2009, **8**, 77.

61 L. Soucek, J. Whitfield, C. P. Martins, A. J. Finch, D. J. Murphy, N. M. Sodir, A. N. Karnezis, L. B. Swigart, S. Nasi and G. I. Evan, *Nature*, 2008, **455**, 679–683.

62 E. V. Prochownik and P. K. Vogt, *Genes Cancer*, 2010, **1**, 650–659.

63 P. Lavigne, M. P. Crump, S. M. Gagné, R. S. Hodges, C. M. Kay and B. D. Sykes, *J. Mol. Biol.*, 1998, **281**, 165–181.

64 P. A. Faull, H. V. Florance, C. Q. Schmidt, N. Tomczyk, P. N. Barlow, T. R. Hupp, P. V. Nikolova and P. E. Barran, *Int. J. Mass Spectrom.*, , DOI:10.1016/j.ijms.2010.01.007.

65 K. Pagel, E. Natan, Z. Hall, A. R. Fersht and C. V. Robinson, *Angew. Chemie - Int. Ed.*, 2013, **52**, 361–365.

66 E. R. Dickinson, E. Jurneczko, J. Nicholson, T. R. Hupp, J. Zawacka-Pankau, G. Selivanova and P. E. Barran, *Front. Mol.*

Biosci., 2015, **2**, 39.

67 A. Sinz, C. Arlt, D. Chorev and M. Sharon, *Protein Sci.*, 2015, **24**, 1193–1209.

68 C. J. Oldfield, Y. Cheng, M. S. Cortese, C. J. Brown, V. N. Uversky and A. K. Bunker, *Biochemistry*, 2005, **44**, 1989–2000.

69 R. Beveridge, A. S. Phillips, L. Denbigh, H. M. Saleem, C. E. MacPhee and P. E. Barran, *Proteomics*, 2015, **15**, 2872–2883.

70 R. K. Das and R. V Pappu, *Proc. Natl. Acad. Sci.*, 2013, **110**, 13392–13397.

71 J. R. Jhingree, R. Beveridge, E. R. Dickinson, J. P. Williams, J. M. Brown, B. Bellina and P. E. Barran, *Int. J. Mass Spectrom.*, 2017, **413**, 43–51.

72 J. R. Jhingree, B. Bellina, K. J. Pacholarz and P. E. Barran, *J. Am. Soc. Mass Spectrom.*, 2017, **28**, 1450–1461.

73 K. J. Laszlo, E. B. Munger and M. F. Bush, *J. Am. Chem. Soc.*, 2016, **138**, 9581–9588.

Chapter 2: Methodology – The Use of Mass Spectrometry and Molecular Dynamics to Examine IDPs

2.1 Declaration

Part of this methodology chapter is reproduced from a published book chapter I wrote: D. Stuchfield, A. P. France, L. G. Migas, A. Thalhammer, A. Bremer, B. Bellina and P. E. Barran, "The Use of Mass Spectrometry to Examine IDPs: Unique Insights and Caveats", In *Methods in Enzymology*, Vol. 611, Ch. 15, pp. 459 – 502.

Being first author on the chapter, I drafted and edited the final manuscript as well as produced Figures 2.3 – 2.15. A. P. France produced Figures 2.1 and 2.2 as well as help with editing the manuscript. L. G. Migas, A. Thalhammer, A. Bremer, B. Bellina and P.E. Barran all provided feedback on the drafts that helped with the editing of the final manuscript. The chapter in its published form can be found in Appendix A. The second part of this chapter discusses the molecular dynamics methods used in this thesis to study IDPs.

The Use of Mass Spectrometry to Examine IDPs: Unique Insights and Caveats

Dale Stuchfield, Aiden P. France, Lukasz G. Migas, Anja Thalhammer, Anne Bremer, Bruno Bellina and Perdita E. Barran*

Michael Barber Centre for Collaborative Mass Spectrometry, Manchester Institute for Biotechnology, School of Chemistry, University of Manchester, 131 Princess Street, Manchester M1 7DN, United Kingdom.

To Whom Correspondence should be Addressed: Perdita.barran@manchester.ac.uk

2.1 Abstract

A sizeable proportion of active protein sequences lack structural motifs making them irresolvable by crystallography. Such intrinsically disordered proteins (IDPs) or regions (IDRs) play a major role in biological mechanisms. They are often involved in cell regulation processes, and by extension can be the perpetrator or signifier of disease. Considering their importance and the shortcomings of conventional methods of biophysical analysis to identify them and to describe their conformational variance, IDPs and IDRs have been termed “the dark proteome”. In this chapter we describe the use of Ion Mobility Mass Spectrometry (IM-MS) coupled with electrospray ionisation to analyse the conformational diversity of IDPs. Using the Late Embryogenesis Abundance (LEA) protein COR15A as an exemplar system and contrasting it with the behaviour of myoglobin, we outline the methods for analysing an IDP using nanoelectrospray ionisation (nESI) coupled with IM-MS, covering sample preparation, purification; optimisation of mass spectrometry conditions and tuning parameters; data collection and analysis. Following this, we detail the use of a ‘toy’ model that provides a predictive framework for the study of all proteins with ESI-IM-MS.

2.3 Introduction

2.3.1 Background

Over the past years, the study of Intrinsically Disordered Proteins has gained more and more momentum.[1] The growing interest in these proteins stems largely from their role in proteopathic diseases such as Parkinson's and Alzheimer's disease[2] as well as the desire to develop a thorough understanding of the role of these conformationally diverse protein complexes.[34] A classic example of this is the tumor suppressor protein p53 which is involved in a high proportion of human cancers.[56] IDP's differ significantly from ordered proteins, in that due to their conformational diversity it is not possible to characterise them by a single 'native' structure; instead they are better considered to exist as conformational ensembles.[7] This conjecture helped to disprove the dogma that a proteins function is directly related to a resolvable single structure.[8, 9]

New definitions were developed following the discovery of IDPs that better described the protein paradigm. A 'protein trinity' consisting of three structures: ordered, molten globule and random coil was proposed by Dunker.[10] This trinity was later expanded by Dyson and Wright[11] to include a better description for the ordered state by partitioning it into two subsets of 'mostly folded with localised disorder' and 'linked folded domains'. It became apparent that the study of IDPs by methods such as X-ray crystallography could not resolve valuable information on the inherent nature of the proteins.[7,12,13]. As a result, there arises an opportunity to develop alternative methods for the structural analysis of IDPs and the interactions they partake in with potential co-factors and substrates. One such method, which has grown in use, [14–18][19] is Ion Mobility-Mass Spectrometry (IM-MS) coupled with electrospray ionisation (ESI).

The origins of Mass Spectrometry came from the desire to identify atomic isotopes and it was initially used as a method for determining the molecular weight of small chemical compounds [20]. However with the development of soft ionisation it became possible to transfer whole protein assemblies into the gas phase[21,22–24]. Once in the gas phase, the mass spectrometer can be used to determine the stoichiometry of complexes, as well as the strength of subunit interactions with exquisite detail[25262728].

MS is capable of trapping conformational intermediates, or manipulating structures and storing gas phase ions over a microsecond to hour timescale, facilitating the investigation of protein folding and unfolding as well as the observation of conformational change(s).[29,30,31] This ability to probe the conformational behaviour of isolated complexes positions MS as an attractive biophysical technique for the study of conformationally diverse proteins.[32] In addition, mass spectrometry is an extremely versatile technique which is often coupled to numerous separation methods in tandem, such as Liquid Chromatography (LC) and Gas Chromatography (GC). Following the development of ESI, the coupling of MS directly to the eluent from a chromatographic column was such a successful union that it became a standard procedure when beginning any analysis of proteins or other biomolecules,[33] and had phenomenal power in top down sequencing efforts.[34,35] MS alone lacks the ability to provide further detail regarding structure and dynamics, hence when carrying out a study on a given protein, LC-MS is used primarily for the determination or confirmation of the molecular weight and for the assignment of post translational modifications.

The development of ESI provided an extra dimension to the analysis of proteins through the presentation of the protein in a charge state distribution (CSD). It was observed very early on that when a protein was sprayed using ESI, it would present to the mass spectrometer in a

range of charge states which primarily corresponded to the protein/ protein complex with a neutral mass M, presenting as ions with several m/z values. Each of these m/z values corresponded to ions in positive ionisation mode of the general form $[M+nH]^{nz+}$ (or in negative ionisation mode as $[M-nH]^{nz-}$ where n increments by 1. Such charge state distributions can be deconvoluted to elucidate accurate mass, and interpreted in terms of the conformational state of the protein under study.[36–38,39, 40] The range of n is pertinent, and denatured proteins (for example following LC separation) or even disordered proteins, present with wider CSDs, whilst globular proteins which have retained aspects of a stable 'native' fold present with narrower CSDs. [41]

The term native mass spectrometry has been coined to explain the careful use of electrospray ionisation, primarily from salty aqueous solutions, followed by slow removal of solvent to best preserve the non-covalent interactions present in the native protein or protein complex.[42] In IM-MS this tends to result in the production of a narrow CSD coupled with a narrow spread of conformations.[38] For IDPs the use of native MS methods mostly provides a wide CSD and correspondingly wide range of conformational states, as shall be seen later in this chapter.

Following seminal work in the 1990s, by Clemmer, Jarrold, Bowers, Russell and others,[4344–4647] as well as significant technical developments,[48] IM-MS instruments started to become commercially available in the mid-2000s.[49,50] Adding ion mobility separation to MS adds an extra dimension of analysis to MS experiments as proteins are separated based on their charge and size.[51] As mass spectrometry separates ions according to their movement due to an applied electric (or magnetic) field in a vacuum, ion mobility separates ions due to an applied electric field in the presence of a neutral buffer gas. In the case of drift tube ion mobility, DTIMS, the protein is introduced to a gas filled chamber as a pulse of ions that

traverse the cell due to a weak electric field (5 – 50 V cm^{-1}). The field causes the ions to move through a cell, and collisions with the inert drift gas impede the progress of the ions. Experiments are performed at the so called 'low field limit'[52,53] such that the ions rapidly reach a constant drift velocity (v_d) which is dependent on their charge and mass (as for mass spectrometry), and also on the number and nature of the collisions with the buffer gas. The mobility (K) of a given ion is defined as the constant of proportionality between the drift velocity v_d and the applied electric field (E).

Following data acquisition in an ion mobility-mass spectrometry experiment, an ion of a given *m/z* will have a corresponding arrival time distribution (ATD). This ATD can be used to calculate what is known as a rotationally averaged collisional cross section (CCS), which provides a direct measurement of the proteins' conformational diversity.[54] It is also worth noting that it is possible to computationally calculate a CCS, allowing for the comparison between experiment and theory.[38] IM-MS has been applied to the study of more than 1000 proteins or protein complexes[1] and whilst most of these studies have focussed on folded proteins, there are an increasing number of examples where it has been applied to examine IDPs; such as the tumour suppressor protein p53,[55, 56, 57, 58, 10, 59] and the highly disordered amyloid forming protein, α-synuclein.[60, 61] Some systems such as the family of Cold Regulated (COR) proteins, adopt a disordered or ordered conformation depending on the environment they are in. These COR proteins are vital to the cold acclimation of plants.[62] Specifically they are involved in the stabilisation of chloroplast membranes in the leaf during freezing. COR15A (used as an example in this chapter) and COR15B are cold regulated proteins, found in

[1] On June 1st 2018 a search of the terms 'ion mobility' 'mass spectrometry' and 'protein' on PubMed returned 1113 hits

Arabidopsis thaliana, which are critical to the plants adaptation to cold temperatures and specifically to dehydration.[63–65] The COR15 proteins are found to be relatively disordered in hydrophillic conditions and undergo conformational transition to helical forms upon dehydration.

A mechanism of action for the COR15 proteins was proposed by Thalhammer *et al*[62]. Upon freezing, a cellular dehydration occurs which results in a crowding effect in the cell membrane; this is accompanied by a rise in the expression levels of COR15. As the water content decreases, the COR15 proteins are proposed to fold into an α-helix and then to associate with the inner envelope membrane of the chloroplast. The association further promotes helicity in COR15, and in this form it is found to stabilise the membrane against the formation of hexagonal II (H_{II}) phase lipid domains.

2.3.2 Overview of Practice

In this chapter we will explain our methodology for the analysis of IDPs using IM-MS, this will cover the initial sample preparation and purification procedures, the importance of running a standard when working with an unknown protein, how best to optimise spectrometer settings, the acquisition of consistent data and post-acquisition analysis. Following this we will detail the use of a predictive framework and 'toy model' that provides additional theoretical data regarding a proteins CCS and the degree of disorder within its structure.

2.4 Experimental IM-MS

As well as the drift tube IMS (DTIMS) described above[66], there are currently four other types of methods commonly available for ion mobility separation: Aspiration Ion Mobility Spectrometry (AIMS),[67] Field-Asymmetric Waveform (or Differential) Ion Mobility Spectrometry (FAIMS/DIMS),[68] Travelling Wave Ion Mobility Spectrometry (TWIMS),[49] and

Trapped Ion Mobility Spectrometry (TIMS)[69]. This chapter will only cover DTIMS in detail, but similar approaches can be used for TWIMS and TIMS measurements albeit following calibration to obtain CCS values using well studied known protiens.[6914] The mechanism for how DTIMS separates ions will not be covered here, but the reader can find more detailed explanations in a number of papers.[537071 54]

2.4.1 Sample Preparation and Purification

Sample preparation is the first step in the analysis of proteins by native ion mobility mass spectrometry. This procedure typically involves the formulation of a solution at a pH appropriate for the protein under study, adjustment of the protein concentration (depending on ionisation method) and at times, some purification steps (e.g., desalination). For the majority of native (IM)-MS experiments the low flow rate variant of ESI, namely nano-ESI (nESI) is utilised. This uses borosilicate or quartz capillaries of ID ~ 0.9 mm often home pulled to orifices 0.5-20 µM in diameter, with the ESI potential applied either to an inserted platinum wire or to a gold or palladium exterior. The principal advantage of this is that nESI uses *cf.* 100x less sample than conventional ESI, and therefore is suited to low concentration samples. nESI from home pulled tips also allows far better control of the size of the droplets, resulting in smaller parent droplets during the spraying process which gives less time for solvent evaporation that can influence the conformational equilibration prior to ionization which in turn can influence the measured conformations of the protein.

For a typical nESI-(IM)-MS experiment, 20 µl of a 1-30 µM solution of the protein of interest will last for c. 8 hours of continual use (if aggregation does not occur).

2.4.1.1 Buffers and Solutions

Several salt solutions are deemed suitable for the solvation of proteins which are subjected to native mass spectrometry experiments. These solutions can also act as buffers if the pH needs to be adjusted, with the central aim of being volatile enough to depart the protein upon desolvation. Aqueous ammonium acetate (AmAc) with a pH range of ~6.8 – 7.4 is the 'go to' salt for spraying intact proteins, as it is volatile and readily evaporates during electrospray. Furthermore, high concentrations of ammonium acetate also help to reduce the effects of non-volatile buffer components.[23] That being said, a number of groups have recently shown that it is possible to perform native (IM)-MS from highly salty solutions[72] and with great promise for the future from crude cell lysates.[73–75]

For initial analysis (to check the protein mass) it may be appropriate to first examine the protein under conditions that will not preserve its native structure, for example 50:50 H_2O:MeOH + 0.01 % formic acid solution, which is similar to the make-up flow from an LC experiment.[23]

For certain experiments other solvents can be used, for example 2,2,2-trifluoroethanol (TFE) which is known to induce helical structures in proteins at ambient temperatures.[76]

2.4.1.2 Equipment & Reagents

- pH meter
- Ammonium acetate
- MeOH:H_2O (50/50) + 0.01 % formic acid
- Ammonium hydroxide solution (AmOH) (~ % 20)
- Acetic acid (AcOH)

2.4.1.3 Procedure

1. Dissolve solid ammonium acetate (in the required quantity) in 50 mL ultrapure water to yield a solution of ammonium acetate with the desired ionic strength.

2. Measure the pH of the solution.

3. Raise or lower the pH to the desired level.

- If the pH of the solution is below the desired level add ammonium hydroxide to raise it.

- If the pH of the solution is above the desired level use acetic acid to lower it.

4. Once the desired pH is reached, the solution is ready to be used.

2.4.1.4 Notes

- Generally, not much AmOH or AcOH is needed to alter the pH. It is advisable to start with ~ 5 µL aliquots (however, these amounts will vary significantly depending upon the concentration of the AmOH/AcOH used, the ionic strength of the solution and its target pH).
- Due to the small amount of solvent used in nESI, solution quantities are small. As such, it is advisable to calibrate pipettes every 6 months.

2.4.2 Buffer Exchange & Purification

Proteins often arrive in a solution that has been optimised for other biophysical experiments which are rarely suited to direct infusion MS. Even lyophilised proteins are often 'enriched' in sodium or potassium salts and often both powders and solutions can contain some impurities. As a result, it is necessary to carry out buffer exchange to transfer the protein to an MS compatible solvent (e.g., AmAc mentioned earlier).

2.4.2.6 Equipment & Reagents

- Micro Bio-Spin™ 6 Column (Bio-Rad, Hercules, CA, USA))
- Micro Centrifuge Collection Tubes (Bio-Rad, Hercules, CA, USA)
- MSE Micro Centaur Centrifuge
- Buffer solution
- Protein sample

2.4.2.7 Procedure

1. Shake the spin column (in order to re-suspend the gel and remove any air bubbles) and break the bottom off, then place the spin column in the collection tube.
2. Remove the lid on top of the spin column and then centrifuge for 1 minute at 1956 xg in order to remove the packing buffer.
3. Remove any waste buffer collected in the bottom of the collection tube and re-run for another 1 minute at 1956 xg.
4. Remove any waste and if the powder (Bio-Gel P-6 gels) appears wet repeat step 3, else pipette 500 µL of the desired buffer to the spin column.
5. Centrifuge for 1 minute at 1956 xg. This will mark the first column equilibration step in the buffer exchange (BE) process. As such, after this first step, the resulting buffer will contain a 50:50 ratio of desired buffer: packing buffer.

6. Repeat step 5 again until a 99.99:0.01 ratio is reached:
    - Buffer composition after 2x BE cycles - 90:10
    - Buffer composition after 3x BE cycles - 99:1
    - Buffer composition after 4x BE cycles - 99.9:0.1
    - Buffer composition after 5x BE cycles - 99.99:0.01
7. Now place the spin column into a fresh collection tube, pipette between 20 – 75 µL of the sample into the spin column, (larger volumes may affect the efficiency of the buffer exchange process) and centrifuge for 5 minutes at 1956 xg.
8. Collect the filtrate from the bottom of the collection tube and repeat step 7 (using filtrate instead of sample) in a fresh, equilibrated spin column (i.e. 5 x BE). Repeat for the third time and collect your MS compatible sample solution.

2.4.2.8 Notes

- Another method for buffer exchange and/or desalination is dialysis. Typically, dialysis is performed in cassettes, and this is feasible if larger amounts of protein (> 1 mg) are available.
- Buffer exchange utilising spin columns are a good alternative to dialysis when sample quantity is an issue. However, sample loss and dilution during the purification step are not uncommon.
- Typically, 2-3 buffer exchanges are carried out when purifying any given protein; this number should be minimised based on MS analysis.
- This process is also applicable to the purification of a protein obtained as a lyophilised powder, which can be dissolved in the desired buffer beforehand and buffer exchanged as described in step 7.

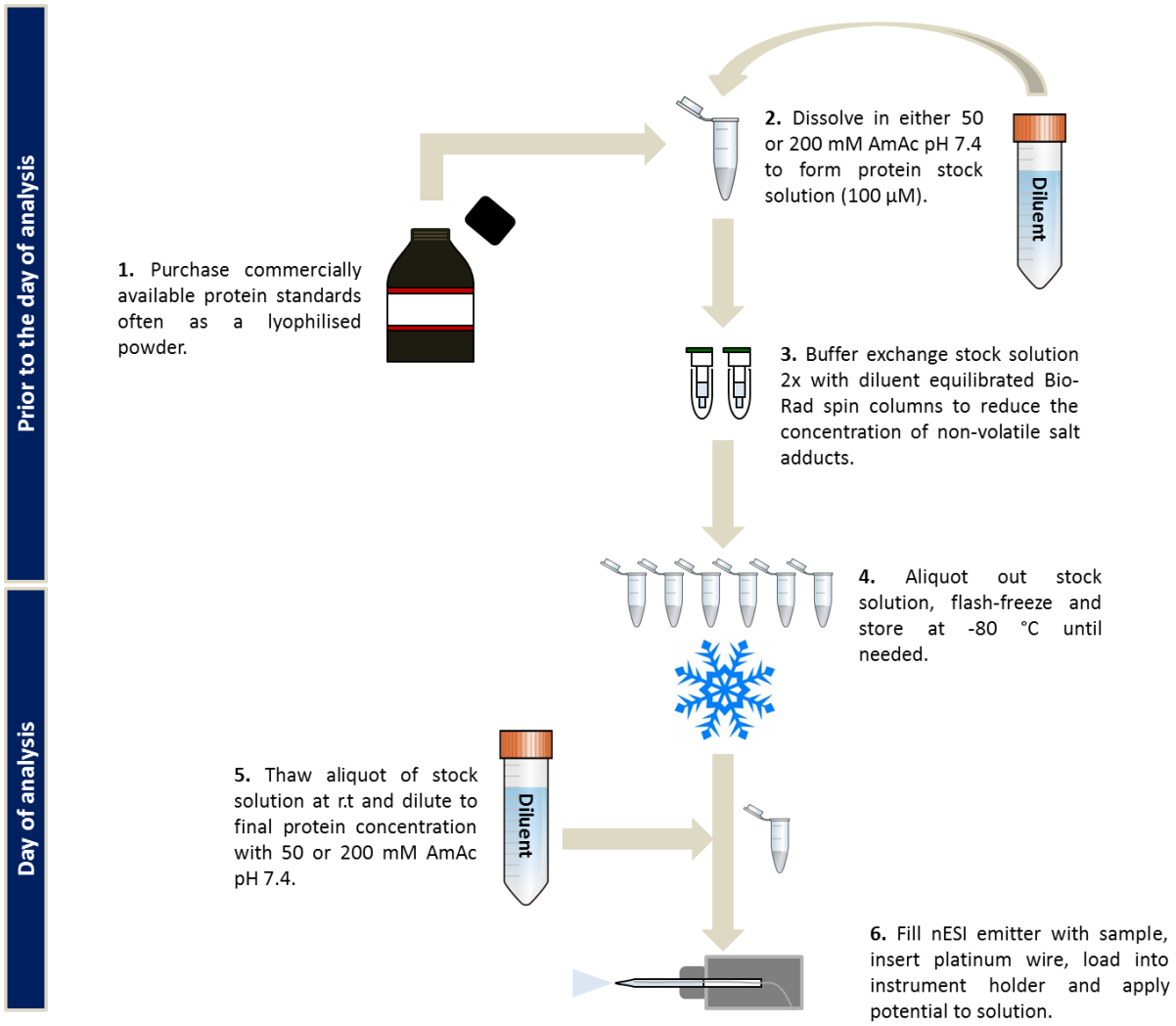

Figure 2.1 | Typical preparative procedure for commercially available protein 'standards' prior to native (IM)-MS analysis. Majority of the 'standard' protein/protein complexes we use for native (IM)-MS experiments are obtained as lyophilised powders (1). The formulation of the diluent, used to dissolve the lyophilised protein 'standard' to form a stock solution (2), is described in Steps 1-4 in '2.4.1.3 Procedure' in the main text. The typical buffer exchange procedure (3) used to remove non-volatile adducts from the protein sample is described in Steps 1-8 in '2.4.2.7 Procedure' in the main text. Following this process, the buffer exchanged stock solution is aliquoted out, flash frozen and stored at -80 °C until the day of analysis (4). On the day of analysis, single aliquots are thawed (as required) at r.t and diluted to their final concentration (5). Following this, the analyte solution is loaded into an 'in-house' pulled nESI emitter (Tip pulling/tuning parameters on our tip puller are described in '2.4.3.2 Tip Pulling/Tuning' in the main text), platinum wire (if the tip is not coated) is inserted into the emitter

so as to make contact with the solution and the instrument holder and a potential is applied to the solution to permit sample ionisation (6).

2.4.3 Ion Mobility and Mass Spectrometry

Each spectrometer is different and requires a certain level of finesse to operate in such a way as to obtain high quality data. Accordingly, it is sensible to run a well characterised, cheap, readily available 'standard' protein prior to analysing an unknown sample. This preliminary analysis is used to check whether the nanospray emitters are optimal and whether the mass spectrometer is working correctly. Furthermore, this process allows for the pre-optimisation of instrument tuning parameters prior to the analysis of the subject protein. Standard proteins should ideally present with a similar *m/z* to the protein under investigation and be selected on the basis that they are of high quality and low batch variability whilst presenting with available mass and CCS literature values. Commonly these standards are globular proteins rather than IDPs, but this does not influence their applicability within the pre-optimisation process. Following emitter and instrument pre-optimisation and buffer exchange of the subject protein, we can proceed to analyse the sample. From a samples mass spectrum, we obtain a charge state distribution (CSD) from which we can infer the analytes experimental mass, which can be compared to its theoretical mass to ascertain if there are any post-translational modifications (PTMS) or uncleaved tags present on the analyte. The CSD gives us valuable information about the proteins structure as mentioned previously (2.3.1).

The next step is to acquire mobility data by carrying out an Ion Mobility experiment. In drift tube instruments which have a quadrupole prior to the IM cell, this can involve mass selection

of specific charge states followed by IM analysis at a range of different drift voltages (typically ≥ 5). Following successful data acquisition, arrival time distributions of specific charge states can be extracted and after further data processing collision cross section distributions (CCSDs) can be generated.

2.4.3.1 Equipment and Materials

- An Ion Mobility-Mass Spectrometer, some examples from our lab include: A Waters Synapt G2 fitted with an RF confining drift cell across which is applied a static potential gradient,[77] an Agilent 6560 DTIMS QToF,[78] and two home-built DT-IM-MS instruments capable of measurements at a range of drift gas temperatures[7170]
- Thin-wall borosilicate glass capillaries (i.d. 0.9 mm, o.d. 1.2 mm, World Precision Instruments, Stevenage, UK) or equivalent
- Model P-1000 Flaming/Brown Micropipette Puller (Sutter Instrument Company, USA) or equivalent
- 10 µL gas tight syringe (Hamilton)
- Spool of platinum wire, 0.125 mm diameter (Goodfellow, Cambridge Ltd, UK)
- Helium/Nitrogen & Argon gas (> 99.9 %, vol/vol)

2.4.3.2 Tip Pulling/Tuning

In line with the majority of native MS studies, we ionise IDPs using nano-ESI. In this study the solution is ionised *via* the application of a potential (which in this case is positive) to a thin platinum wire inserted into the loaded nESI emitter. That being said, emitters can also be coated with gold or palladium and the source potential applied to this surface instead. All emitters utilised within this work were pulled in-house using a model P-1000 Flaming/Brown

micropipette puller. This puller has various settings which can be tuned in order to produce emitters of varying taper and orifice morphology; these settings are listed in Table 2.1.

Table 2. 1| Lists the various settings on the P-1000 tip puller we use. More detail can be found in the operations manual which is available as a PDF online (https://www.sutter.com/manuals/P-1000_OpMan.pdf).

Setting	Range	Description
Heat	0-999	Controls the level of electrical current supplied to the filament.
Pull	0-255	Controls the force of the pull. The higher the PULL the longer the taper and smaller the diameter of the tip.
Velocity	0-255	The adjustable velocity allows for a selection of a precise glass temperature as the trip point for the hard pull.
Time	0-500	Controls the length of time the cooling air is active.
Delay	0-500	Controls the delay time between when the heat turns off and when the hard pull is activated.

2.4.3.3 Mass Spectrometry and Ion Mobility

Our group has a variety of instruments to carry out (IM)-MS experiments, and we have analysed IDPs on several of these[14]. For this chapter, data has been taken on the modified Synapt G2 (with an RF confining drift cell) to highlight the steps undertaken to analyse IDPs by IM-MS.

Samples analysed via nESI are typically sprayed from 10 – 500 mM ammonium acetate (the chosen salt concentration depends on the protein of interest, as non-native self-interactions can occur if ionic strength is too low, and salt clustering can occur at high concentrations) solutions buffered to pH 6.8 – 7.4, in order to best mimic physiological conditions. We often apply capillary voltages of 0.7-1.2 kV, with (on Waters instruments) cone voltages ranging from 10 – 200 V, and source temperatures of 50-80 °C. The pressure in the desolvation region of the source is also an important parameter to adjust since it alters how and when the solvent

and salt departs the protein. This is affected either by throttling the flow on the roughing pump on the source and/or by adding more gas into the source.[79–81]

2.4.3.4 CCS from DTIMS

In order to obtain necessary data to derive CCSs of individual ions, we perform IM experiments at several (≥ 5) drift voltages whilst also recording the temperature and pressure of the buffer drift gas (N_2 or He) at each drift voltage. The ATDs of the entire mass spectrum are recorded by synchronisation of the release of ions into the drift cell with the mass spectral acquisition. All MS and IM-MS data was analysed using MassLynx v4.1 (Waters, Manchester, UK), Origin v8.5 (OriginLab Corporation, USA) software, as well as Microsoft Excel and ORIGAMI.[82]

The CCS values are derived from raw arrival time data using the Mason-Schamp Equation 2.1,[52] shown below:

$$\Omega_{avg} = \frac{(18\pi)^{1/2}}{16}\left[\frac{1}{m_b}+\frac{1}{m}\right]^{1/2}\frac{ze}{(k_B T)^{1/2}}\frac{1}{\rho}\frac{t_d V}{L^2} \qquad \text{Equation 2.1}$$

Where m_b and m represent the buffer gas and ion masses respectively, z is the charge of the ion, e is the elementary charge, k_B is the Boltzmann constant, T is the gas temperature (in Kelvin), ρ is the buffer gas density, L is the drift tube length (in cm), V is the voltage across the drift tube and t_d is the ion drift time. It should be noted that because the raw arrival time output, t_a, includes the time the ion spends outside of the drift cell, t_0, Equation 2.2 must be used to obtain the actual ion drift time, t_d. An ions dead time (t_0) is calculated from the y-intercept of the linear plot of the ion arrival time (t_a) (extracted from the apex of the ATDs

obtained at each Drift Voltage (DV) versus the reciprocal of the effective drift voltages (1/V). Accordingly, we obtain a t_d which can be applied to Equation 2.1 to calculate the CCS for the apex of an ion population.

$$t_d = t_a - t_0 \qquad \text{Equation 2.2}$$

2.4.3.5 DTIMS Procedure

1. Standard Preparation | Begin by running a protein standard. We recommend myoglobin (*holo* = 17.6 kDa, *apo* = 16.9 kDa) or cytochrome c (12.4 kDa) for smaller IDPs. Prepare a solution of the protein standard in 10-50 mM ammonium acetate at ~1-5 µM (use MeOH:H_2O:FA, 50:50:0.01 mixture if analysing under denaturing conditions).

2. Drift Gas | Choose a drift gas for use in IM separation, the majority of experiments on proteins have used either N_2 or He but other gases are viable. In Synapt instruments argon is the default gas employed in the Trap/Transfer regions.

3. Instrument Settings | Example tuning parameters employed for the native DTIM-MS analysis of *holo*-myoglobin on the modified Synapt G2 (with an RF-confining drift cell) are given in Table 2.2 and a schematic diagram of the modified Synapt G2 is shown in Figure 2.2. The parameters outlined in Table 2.2 yielded a good compromise between MS quality and the minimisation of gas-phase structural activation. This was concluded by comparison of our experimentally derived CCSs with previously reported literature values.[83] As such, *holo*-myoglobin, sprayed from a non-denaturing solution (made up as described previously) utilising the instrument tuning parameters described below should yield a similar MS to what is observed in Figure 2.3.

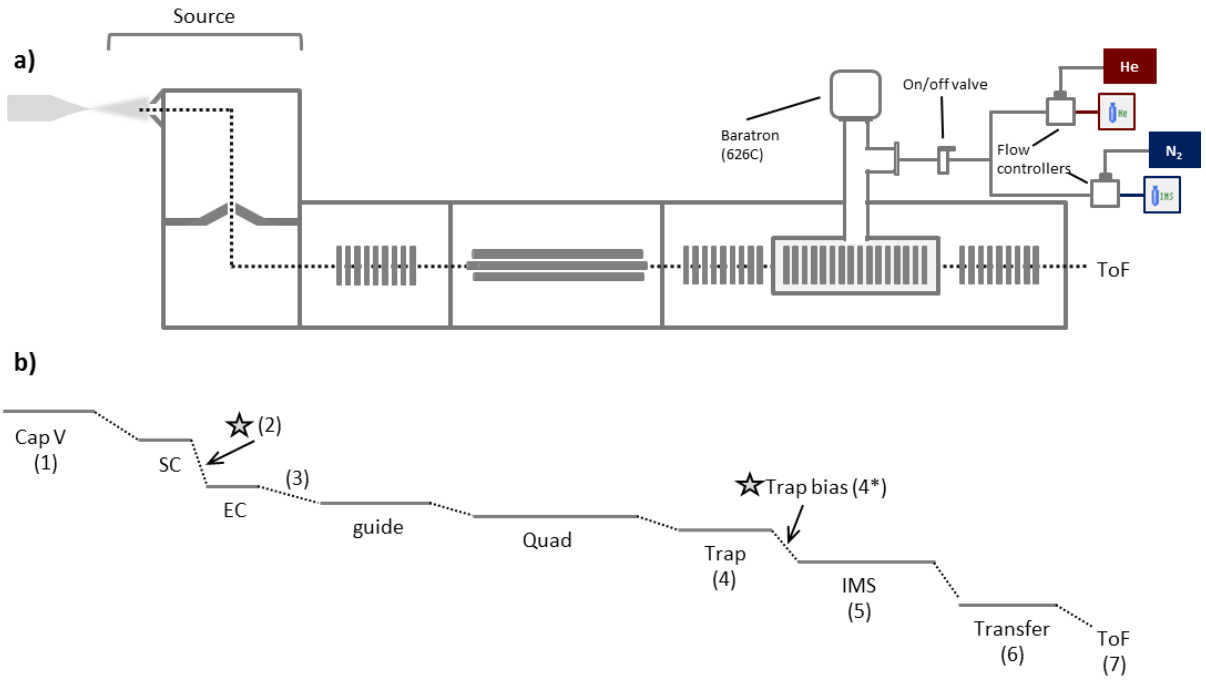

Figure 2.2 | a) Schematic of the modified Synapt G2 (with an RF confining drift cell), in which the black dotted line denotes the ion beam path through the instrument. Drift gases can be switched using the MassLynx software which is connected to a modified gas inlet system. Absolute measurements of the drift gas pressure were enabled by installation of a capacitance manometer (MKS Baratron type 626C) above the centre of the drift cell. b) Schematic of the potential gradient along the instrument, the numbers next to specific regions within the schematic correspond to their respective voltages(s)/gas flows and pressures outlined in Table 2.2. Voltages which impart the greatest energy upon ions, thus promoting their gas-phase activation, are highlighted by grey stars. The voltage gradient from the sample cone (SC) to the extractor cone (EC) can be utilised to promote in-source ion activation. Furthermore, the trap bias voltage, which acts to accelerate ions from a comparatively low-pressure environment (the trap cell) to a higher pressure one (the Helium and IM cell) can also significantly perturb ion structures if not optimised for the system in question.

Table 2.2 | Optimised instrument parameters used to run a sample of 5 µM myoglobin in 50 mM ammonium acetate (pH 7.4) on the modified Synapt G2 (with an RF confining drift cell). Voltages, gas flows and typical pressures utilised and observed within the analysis of myoglobin (in this instance) are linked to regions within the modified Synapt G2 (outlined in Figure 2.2) by the numbers in the left-hand column.

Assignment in Figure 2	Source voltages (V) and gas flows (mL/min)	
1	Capillary voltage (kV)	0.7 – 1.2
Source	Source temperature (°C)	40.0
2	Sample cone	20
3	Extractor cone	1
4	Trap gas flow	2
5	Helium gas flow	100.0
	DC voltages	
4	Trap cell energy (V)	0.0 – 4.0
4	Trap DC entrance	1.0
4*	Trap DC bias	1.0
4	Trap DC	0.0
4	Trap DC exit	1.0
4	Trap height (V)	10-15
5	IMS DC entrance	-6.5
5	Helium cell DC	altered incrementally for DVs
5	Helium cell exit	-40
5	IMS bias	altered incrementally for DVs
5	IMS DC exit	0.0
6	Transfer cell energy (V)	0.0 – 6.0
6	Transfer DC entrance	4.0
6	Transfer DC exit	15.0
	Pressures	
Source	Backing (mbar)	2.7
4	Trap (mbar)	$1.3e^{-2}$
5	IM (Torr)	1.94
6	Transfer (mbar)	$1.4e^{-2}$
7	ToF (mbar)	$7.4e^{-7}$

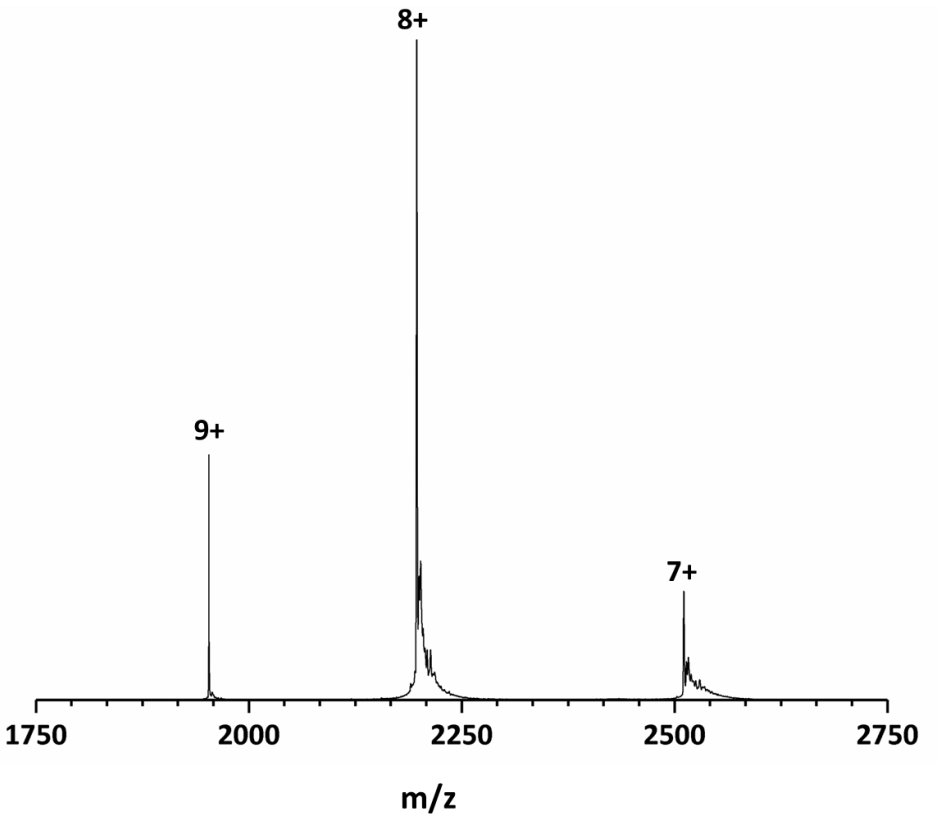

Figure 2.3 | Mass spectrum obtained following nESI of 5 µM Myoglobin in 50 mM ammonium acetate (pH 7.4). The numbers above the peaks denote the charge state *z* for the ion of the form $[M+nH]^{n+}$, where *n* = *z.* This spectrum was acquired on a modified Synapt G2 (with RF confining drift cell).

4. Sample Preparation | Once the standard is spraying stably with good ion transmission (shot-to-shot of > 1E3 from non-denaturing solutions), analysis of the subject protein can commence. Prepare a solution of the subject protein in ammonium acetate at ~5-20 µM (higher concentrations of the subject will be used relative to the standard protein just to observe signal, after which the concentration can be reduced) following the procedures described above (2.1).

5. Instrument Settings | With the same tuning parameters employed for the analysis of the standard, spray the subject protein (these parameters may be the same as in Table 2.2, or

different depending on whether slight alterations were needed to improve signal intensity and/or the likelihood of structural activation).

6. Judging Quality | Mass spectral quality will differ depending upon the samples' solution composition (non-volatile buffer components reduce mass spectral quality), homogeneity and its propensity for salt adduction. As mentioned in Section 2.1 most IDP samples we obtain have been recombinantly expressed and as such, are rarely homogenous in nature. This is why the buffer exchange procedure, outlined previously, is essential prior to MS analysis. An example of a spectrum containing insufficiently high amounts of non-volatile adducts, where the analyte requires further purification, is depicted in Figure 2.4. One way to solve this issue is to repeat the steps covered in Section 2.1.2.

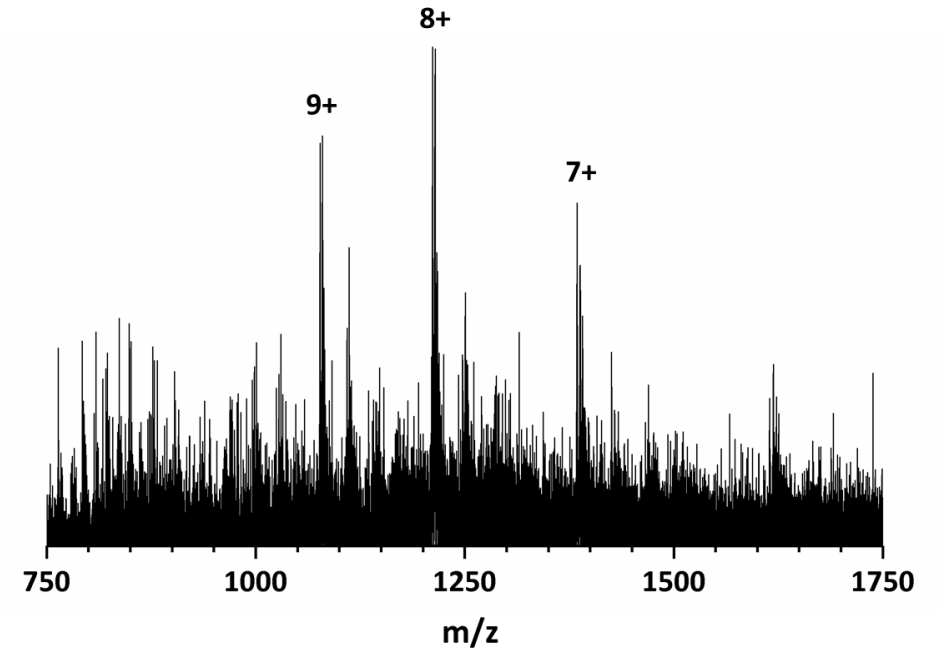

Figure 2.4 | Mass spectrum of COR15a 20 µM in 100 mM TFE solution. The numbers above the peaks denote the charge state *z* for the ion of the form $[M+nH]^{n+}$, where $n = z$. The low signal:noise ratio within the spectrum, as well as the comparatively wide mass spectral peaks within the charge state envelope, indicate that the sample solution still contains a high level of non-volatile salts. This spectrum was acquired on a modified Synapt G2 (with RF confining drift cell).

7. MS Analysis | Improvement of the data quality is achieved, either by further purification or instrument parameter optimisation, whereby source voltages within the instrument can be increased to effectively dissociate non-volatile salt adducts from the analyte ions. After optimising the preparative procedure and tuning parameters to obtain a high quality mass spectrum, post-acquisition analysis can commence. Initially the protein charge states can be found from the m/z peaks and their spacing within the mass spectrum (Equation 3). After which the experimental molecular weight of the analyte can be determined (Equation 4).

$$z = \frac{M_2 - mA}{M_1 - M_2} \qquad \text{Equation 2.3}$$

$$MW = \frac{(M_1 - mA)(M_2 - mA)}{M_1 - M_2} \qquad \text{Equation 2.4}$$

Where *z* is the charge state of M_1, M_1 is the m/z of the lower charge state ion (higher m/z) M_2 is the m/z of the higher charge state ion (lower m/z), *mA* is the mass of a single proton (or any other observed adduct) and MW is the molecular weight of the analyte. Regarding the CSD, a large Δ*z* observed for proteins analysed from near-physiological solution pH, is indicative of at least partial disorder within the protein structure, albeit with some exceptions.[8485] It is worth noting that another method for determining charge on a protein is using the Rayleigh Limit (Z_R) parameter which was developed by de La Mora[86] (Equation 5). It allows the calculation of the maximum number of charges that a globular protein of a specific molecular weight can carry.

$$Z_R = 0.0778\sqrt{m} \qquad \text{Equation 2.5}$$

Where Z_R is the maximum charge and *m* is the mass of the protein. As mentioned previously (1.1) the spread of *z*, *ergo* the CSD provides valuable information about the structure of the protein; wider CSDs generally suggest that the protein occupies a large range of conformers

in solution, thus yielding a multitude of available protonation sites upon ionisation.[41] The nESI mass spectrum from the LEA protein COR15A (Figure 2.5) possesses a broad CSD with Δ*z* = 9, as would be expected for an IDP. Interestingly and symptomatic for many proteins, the lower charge states (z = 5-7) still retain some salt adducts (see insert in Figure 2.5) relative to z ≥ 8, which would indicate that the lower charge states might sample more compact, globular structures with binding cavities which permit increased salt adduction.

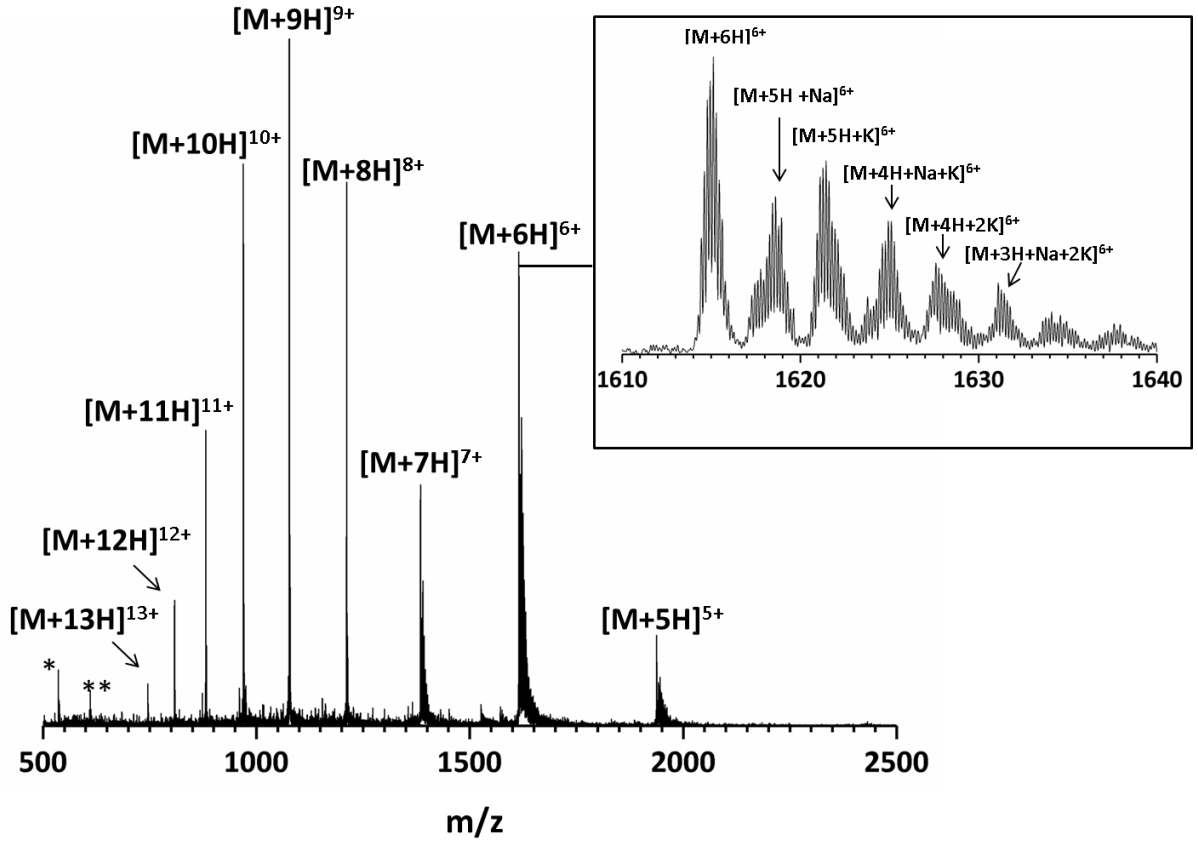

Figure 2.5 | Mass spectrum of COR15A sprayed from 200 mM ammonium acetate solution, pH ~ 7.3.. The experimentally determined MW is 9688.3 Da ± 3.4 and the theoretical MW is 9683 Da. Peaks denoted with ' * ' and ' ** ' are contaminant species with m/z 527 and 610 respectively. This spectrum was acquired on a modified Synapt G2 (with an RF confining drift cell).

8. Acquiring Mobility Data | The next step is to conduct IM experiments in order to obtain CCS data for the subject protein; for linear field instruments, this involves taking

measurements at five or more drift voltages. The measurements taken at each voltage provide an arrival time distribution (ATD) for each charge state. The apex values from these distributions gives a time t_a (Equation 2) at which the ion intensity is at a maximum. For each of the different voltages, these t_a values can be used to generate a plot of t_a vs. $\frac{1}{V}$ with a gradient that is proportional to $\frac{1}{K_0}$ and an intercept that corresponds to the dead time t_0, which is independent of the time spent in the drift cell. Linearity of these data points reflects an adherence to the Mason-Schamp equation (Equation 1) and indicates that all data is obtained in the low field conditions. The slope of the plot can be used to determine the CCS value that corresponds to the apex of the ATD for that charge state. The plot stems from Equation 6 shown below which relates the mobility of an ion to its drift time.

$$KE = v_d = \frac{L}{t_d} \qquad \text{Equation 2.6}$$

Where K is the ion mobility, E is the electric field strength, v_d is the drift velocity, L is the drift tube length and t_d is the drift time. Because low field ion mobility measurements are dependent on pressure and temperature we use the reduced mobility, K_0, where pressure and temperature are normalised as shown in Equation 7 below.

$$K_0 = K\left(\frac{T_0}{T}\right)\left(\frac{P}{P_o}\right) \qquad \text{Equation 2.7}$$

Normalised temperature T_0 and pressure P_0 are 273.15 K and 760 Torr, respectively. With this knowledge we can also rewrite the Mason-Schamp equation to directly relate to an ion's mobility (K_0) as shown below in Equation 2.8.

$$\Omega_{avg} = \frac{(18\pi)^{1/2}}{16}\left[\frac{1}{m_b} + \frac{1}{m}\right]^{1/2}\frac{ze}{(k_B T)^{1/2}}\frac{1}{\rho}\frac{1}{K_0} \qquad \text{Equation 2.8}$$

Mass Select | Tandem mass spectrometers allow any charge state of a given protein to be selected. In some IM-MS instruments (e.g., The Waters Synapt family) this takes place prior to the mobility separation, such that the ATD and the mass spectrum only arise from that parent ion. This prevents contamination of the ATD for a given charge state, from higher charge state ions that have been 'charge stripped' during or at the end of the drift cell, or from multimers that have dissociated in or after the drift cell to yield a product ion of the charge state of the ion of interest. If mass selection prior to IM is not possible, then the analyst must be aware of these confounding factors when interpreting data. Irrespective of whether this is technically possible, it is always possible to extract a given ion from the acquired data, and this is a necessary step in data analysis. A zoomed-in region of the mass spectrum focusing on the $[M+10H]^{10+}$ ion of COR15A and corresponding arrival time distribution is shown in Figure 2.6.

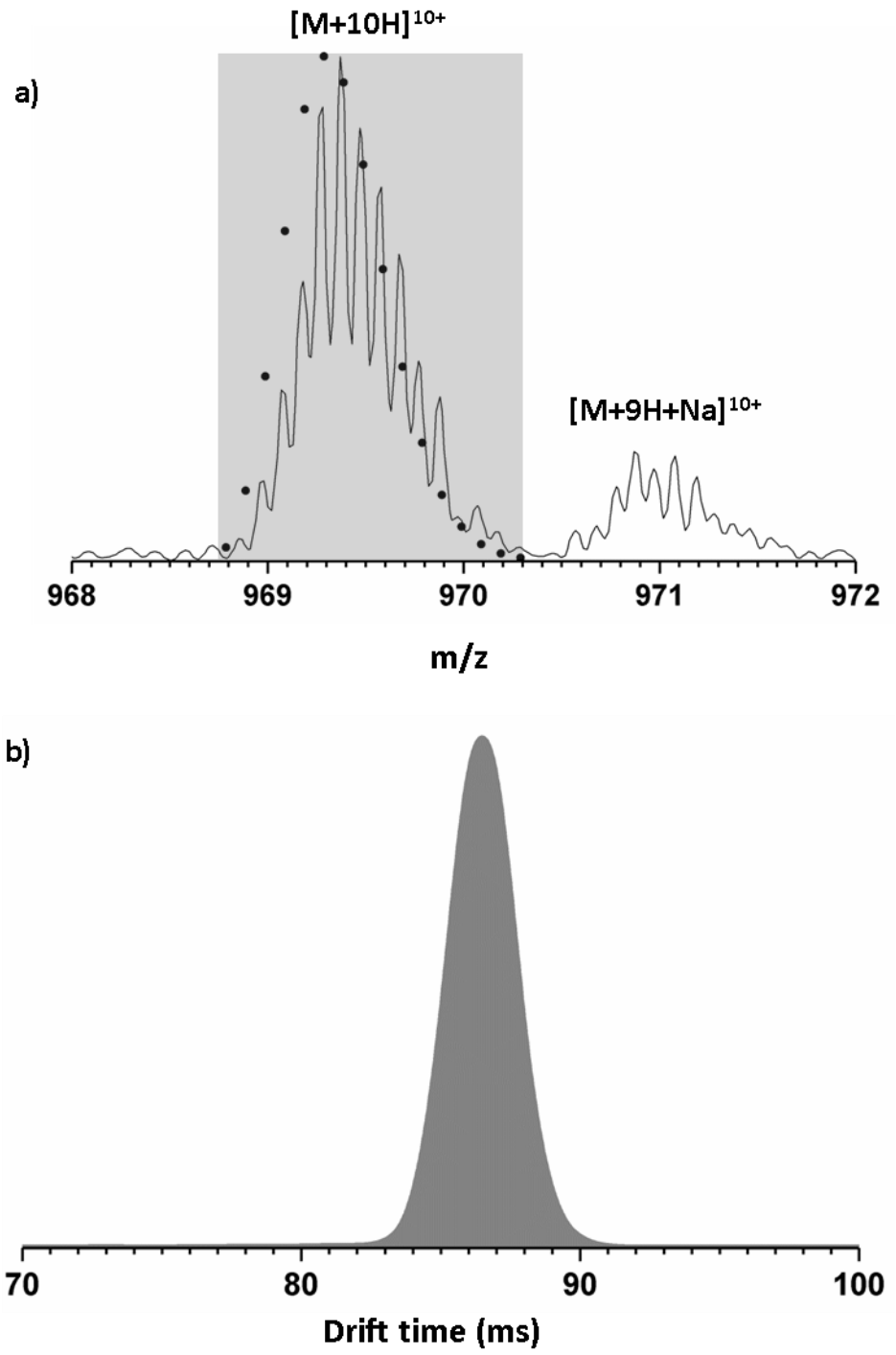

Figure 2.6 | (a) Isotopic distribution of the $[M+10H]^{10+}$ and the salt adduct $[M+9H+Na]^{10+}$ charge state of the COR15A protein. The solid circles represent theoretical isotopic distributions for the $[M+10H]^{10+}$. (b) the arrival time distribution for the COR15A, z = 10 charge state at a drift voltage of 100 V. COR15A was sprayed from 200 mM AmAc, pH ~ 7.3 at a concentration of 20 µM. Data for a) and b) were acquired on a modified Synapt G2 (with an RF confining drift cell).

Optimise | Following mass selection we look at the mobiligram to judge the quality, usually a degree of tuning of the voltages and optimisation of the gas flow in the drift cell is needed in order to improve the data quality (again consult Table 2.2 for good 'start-point' settings; specifically, the Trap DCs and IMS DCs). Figure 2.7a and 2.7b show examples of an acceptable arrival time distribution (ATD) and an ATD where concentration dependent non-native dimers are seen; it is worth noting that these dimers can be removed by serial dilution. Figure 2.7c demonstrates the effect poor voltage optimisation has on the mobiligram. In these steps, which adjust the passage of the ion into the drift cell, it is important to consider that we do not want to disrupt the structure of the ion. If a protein ion is injected into the drift tube with high acceleration voltages it is likely to deform either by first compacting and then (or directly) unfolding, thus distorting the subsequent interpretation of the ATD. With higher injection energies the transmission of ions through the drift cell becomes more favourable resulting in higher signal intensity, so it is important that the analyst tunes the signal whilst monitoring the ATD to prevent such distortion. At very high injection energies, the ion will penetrate far into the drift cell before it reaches the constant drift velocity that satisfies the Mason Schamp equation (1) and this will effectively shorten the length of the drift region, causing errors in the CCS calculation. Conversely, if an ion is not given sufficient energy, it will not enter the cell against the egressing drift gas, resulting in a loss of signal. When a pulse/packet of ions is introduced to the drift tube they begin to separate, therefore the measurement of the packet must be completed before the second packet is introduced otherwise overlap, known as 'wrap around' occurs.

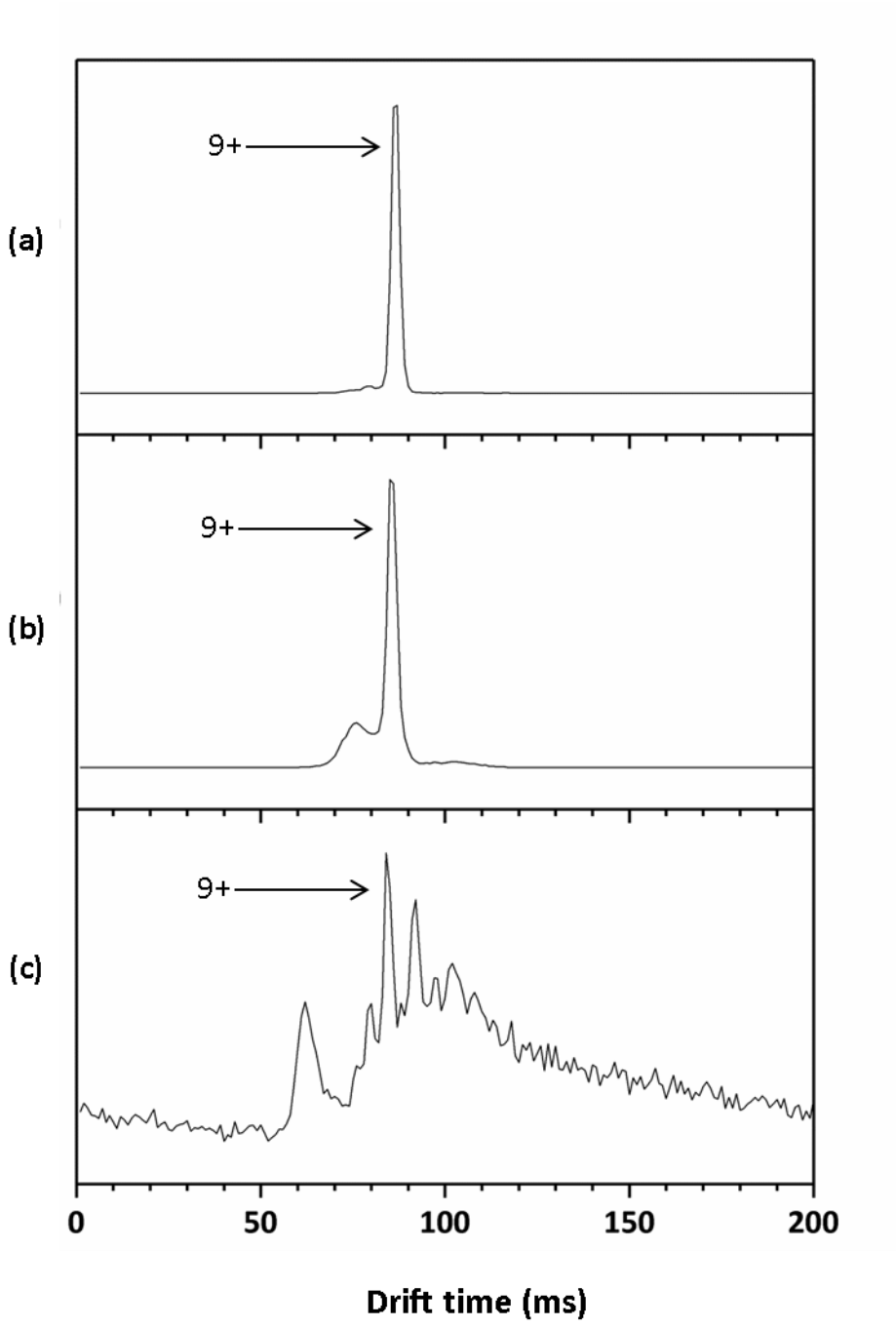

Figure 2.7 | (a) Shown is an example of a good ATD of the 9+ charge state of the COR15A protein. (b) Shown is an ATD of the 9+ charge state of COR15A, the shoulder seen to the left of the main peak likely corresponds to the 18+ dimer of COR15A which has the same nominal m/z and so will be selected along with the monomeric 9+ species. (c) Here is an example of poor quality data where the clean peaks we saw previously are obscured by noise, this stems from poor voltage optimisation. All ATDs were obtained on a modified Synapt G2 (with an RF confining drift cell).

Acquisition | Once a good quality ATD is obtained, experimental data can be acquired. The analyst should take note of the drift cell DC, IMS bias, He cell exit and transfer DC entrance (as these are used in combination to establish the drift voltage across the cell), as well as the temperature and pressure within the drift cell. After acquisition of the measurement at a given voltage (for example 100 V), the drift voltage is then incrementally increased by c. 25 V to begin the second measurement until we have six different drift voltage measurements. By the end of the experiment, we have six ATDs at six different voltages as shown in Figure 2.8 below. Using this drift data, we can calculate the CCS of the ion.

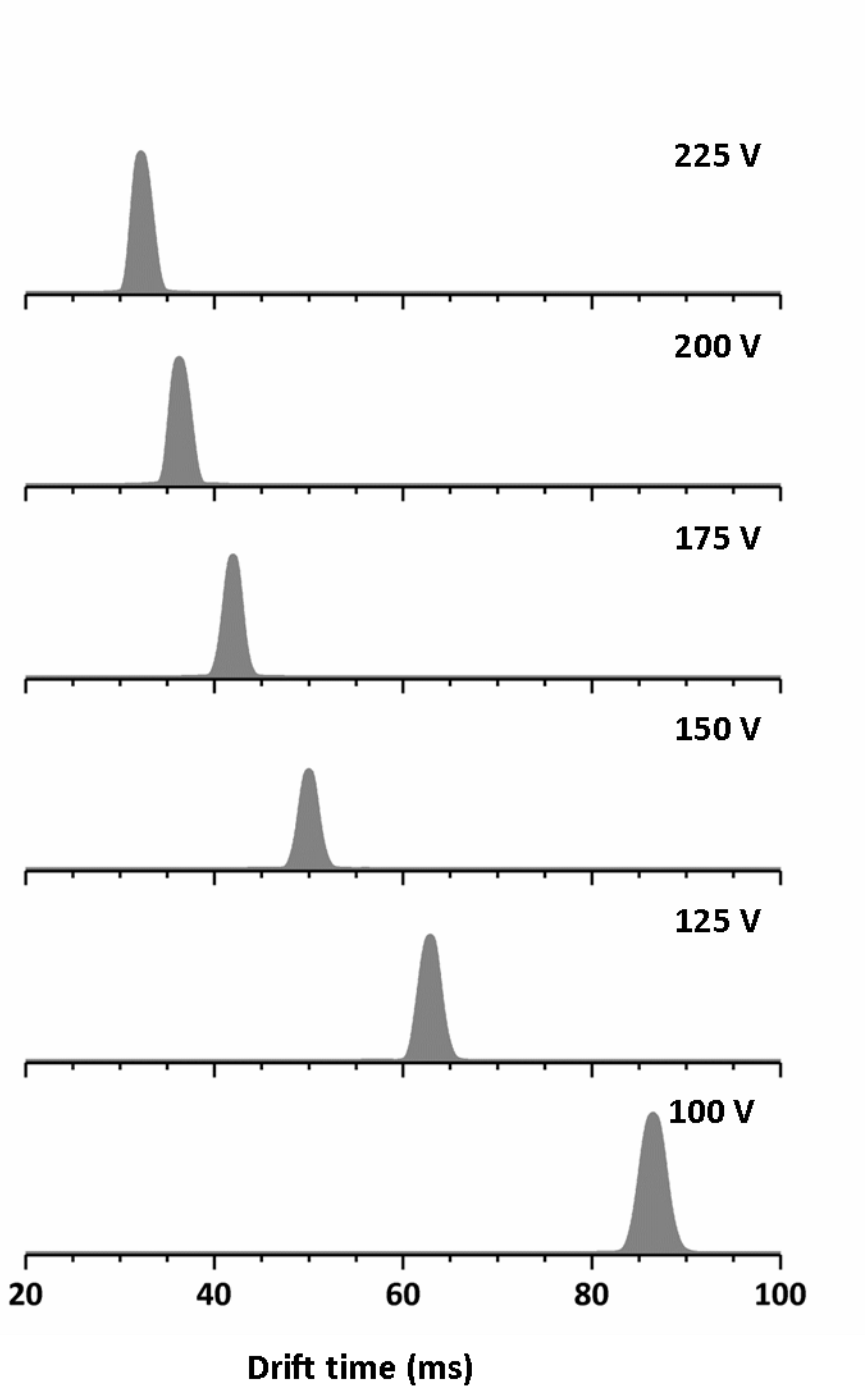

Figure 2.8 | Shown are six selected ion ATDs taken at six different drift voltages for the $[M+10H]^{10+}$ ion of COR15A protein. As the drift voltage is increased, the ions drift velocity also increases, resulting in shorter residency times in the drift cell. For the modified Synapt G2 (with an RF confining drift cell) used here, each ATD was acquired over *c.* 2 minutes.

2.5 Data Analysis

2.5.1 CCS direct from Ion Mobility

The measurements taken at different drift potentials are used to generate a plot of t_d vs. $\frac{1}{V}$ with a gradient that is proportional to $\frac{1}{K_0}$, which can be used to determine an ions CCS. An example is shown in Figure 2.9.

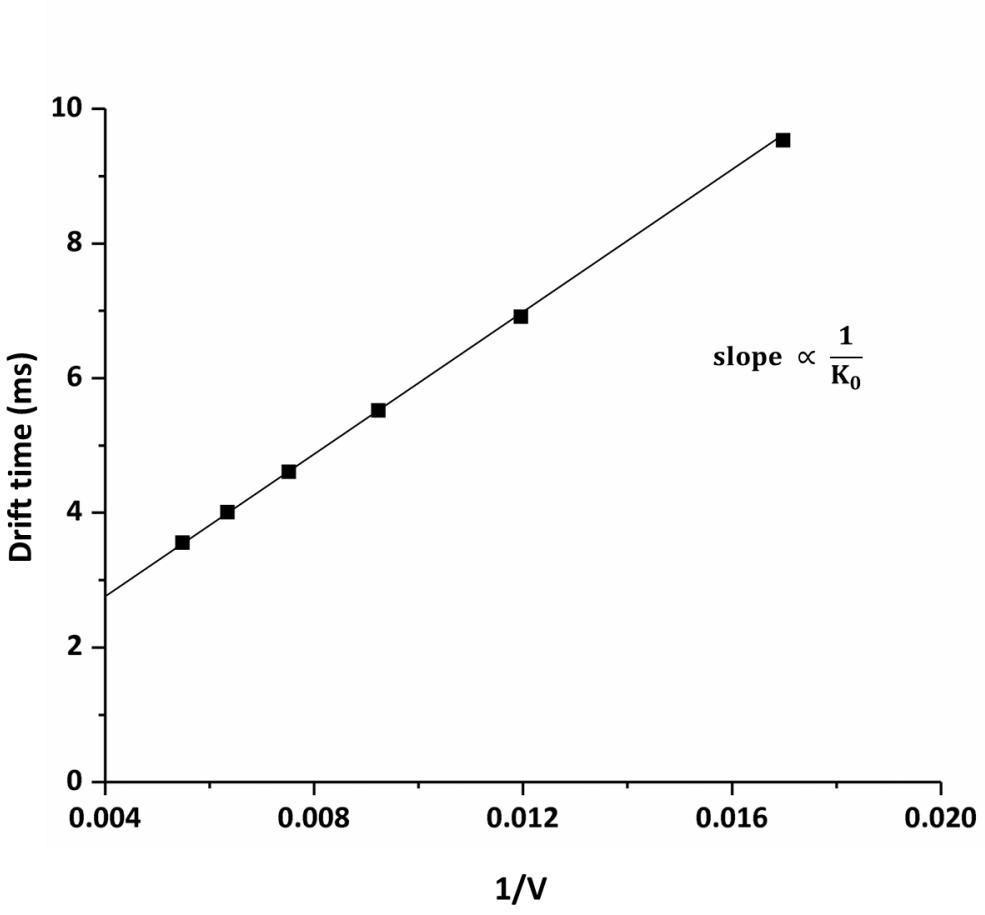

Figure 2.9 | Shown is a plot of t_D vs $^{1}/_{V}$ obtained using the six ATDs for COR15A as seen in Figure 2.8. The slope of the plot is proportional to the inverse of the reduced mobility of the ion. Using this we can calculate the CCS of a specific ion. A similar approach is automated in the Agilent Software for use of their 6560 IM-QToF.[78]

The plot stems from Equation 2.6 shown earlier. Because low field ion mobility measurements are dependent on pressure and temperature we use the reduced mobility, K_0, where pressure

and temperature are normalised as shown previously in Equation 2.7. In terms of specific analysis, we multiply the gradient by the instrument's drift length squared (L^2) and use that as our ion mobility in the Mason-Schamp equation (Equation 2.8).

2.5.2 Converting ATDs to CCSDs

In Section 1.1 we discussed that the most useful structural depiction of an IDP defines them in terms of conformational ensembles. Considering this, it is advantageous to think of the CCS as a distribution rather than a discreet value. We can derive a CCS distribution from the raw arrival time data using Equation 1, which directly relates drift time to the CCS. The process of doing this is as follows:

1 | We obtain an arrival time distribution of a single charge state. In our case, we convert scan time (bins) to time (s) using Equation 8 shown below:

$$Arrival\ Time\ (s) = scan\ number\ \times pusher\ time\ (s) \quad \text{Equation 8}$$

2 | For each charge state of our protein and at every drift voltage we must determine the apex of the major conformation in the arrival time distribution and use the plot shown in Figure 2.9 to elucidate the intercept (dead time, t_0) and the slope ($1/K_0$) which is subsequently used to calculate the collision cross section using the Mason-Schamp equation (Equation 1).

3 | Application of the derived parameters for a single drift time value (i.e. apex of the distribution) gives us a single $^{DT}CCS_{He}$ value, however as mentioned earlier on, IDPs are better explained using CCSDs which show the entire conformational ensemble of the ion. In order to obtain the CCSD, we use the Mason-Schamp equation for each drift time value from the arrival time distribution, as shown in Figure 2.10.

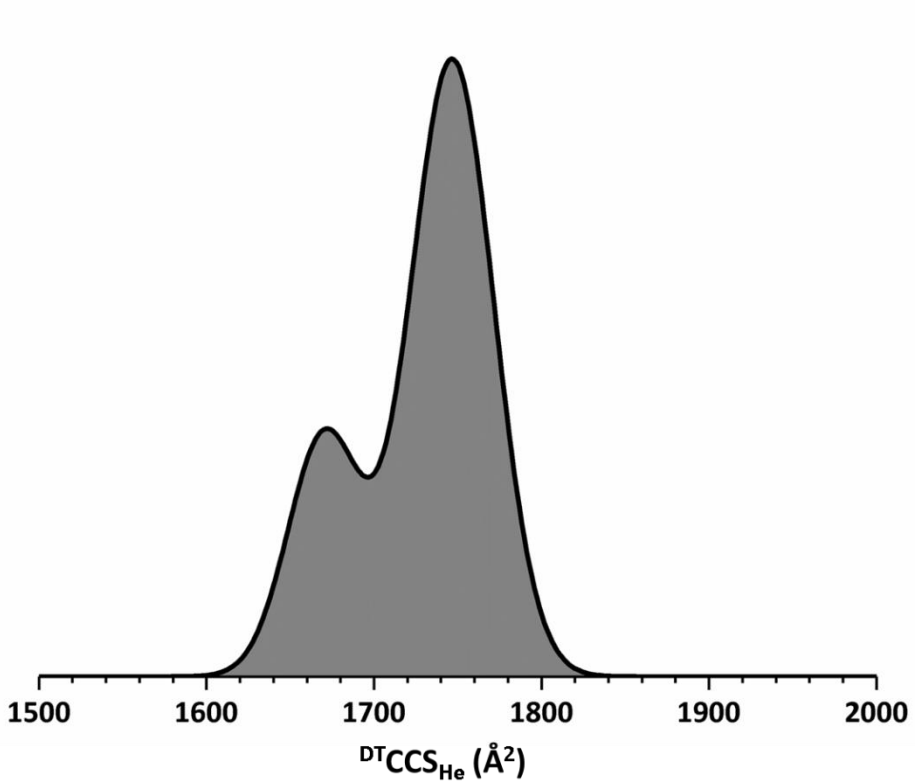

Figure 2.10 | Shown is the $^{DT}CCSD_{He}$ for the $[M+9H]^{9+}$ ion of COR15A. The CCSD is composed of two peaks, with the apex of the conformational families centred at 1670 and 1760 $Å^2$. The lack of baseline resolution may be instrumental, or due to interconversion between these two conformers on the timescale of the experiment. COR15A was sprayed from 200 mM AmAc, pH ~ 7.3 at a concentration of 20 µM. This data was acquired on a modified Synapt G2 (with an RF confining drift cell).

When presenting IM-MS results, we tend to use the CCSD obtained from the lowest drift voltage as it tends to result in better separation of species with multiple conformations as the gradient in the drift tube is less steep and ions have more time to separate. It is worth noting however, that as the ion velocity decreases (due to decreased electric field), the signal-to-noise of the ions of interest can decrease too meaning a higher drift voltage might be more appropriate.

2.5.3 Uncovering Conformational Diversity

The practical steps outlined above can be considered as guidelines for IM-MS analysis of proteins regardless of their inherent order. It is now pertinent to consider the difference seen in data obtained from ordered and disordered proteins. As well as CCS distributions for a specific charge state of a given protein (Figure 2.10), it is possible to plot a total CCS distribution for a protein (Figure 2.11). The process we use to obtain a total CCSD is given below:

Replicate Standardisation| The first step is to fit Gaussian peaks so as to best reflect the experimental data. The cumulative fit of these peaks will have the same range and number of data points across each CCSD replicate for a given ion we do this in order to standardise the CCS bins across replicates. This is because the replicate CCS axes will be slightly different, as the dead time and drift time will vary across experiments due to variables such as temperature and pressure.

Normalisation | We need to normalise the intensities of the different charge states to one another. This is done as follows:

- For each charge state, we average the intensities of the replicates.
- For each charge state, we determine the area under the CCSD.
- For each charge state, we determine the peak height in the mass spectrum.
- We sum the areas for each charge state together.
- We sum the MS peak heights for each charge state together.
- We determine the fraction (area and MS) for each charge state using Equation 2.10:

$$z^{fraction} = \left(\frac{A_z}{\sum_{z_{min}}^{z_{max}} A_z}\right)\left(\frac{h_z}{\sum_{z_{min}}^{z_{max}} h_z}\right) \qquad \text{Equation 2.10}$$

Where, A_z is the area under the CCSD for charge state z, h_z is the peak height for charge state z.

- We then normalise to the highest fraction using Equation 2.11:

$$z^{normalised} = \frac{z^{fraction}}{z_{max}^{fraction}} \qquad \text{Equation 2.11}$$

- Following this, we multiply the averaged intensities by $z^{normalised}$ to obtain normalised intensities for each charge state.
- Lastly we sum together the normalised intensities of each charge state to obtain the intensity for the total CCSD.

Following electrospray ionisation from salty solutions, structured globular proteins present ATDs and corresponding CCSDs that are Gaussian-like and often represent single conformational families (Figure 2.11b) [3884,87]. However, it is not unusual for each charge state to have a slightly different CCSD, with higher values of charge corresponding to higher average CCS values.[88–90] By contrast for IDPs, multiple resolved or partially resolved distributions are observed in the ATD even for single charge states.[9138,84] Under denaturing or partially denaturing conditions, structured proteins present with a greater Δz, often with higher charge states yielding larger CCSs. [44,91]

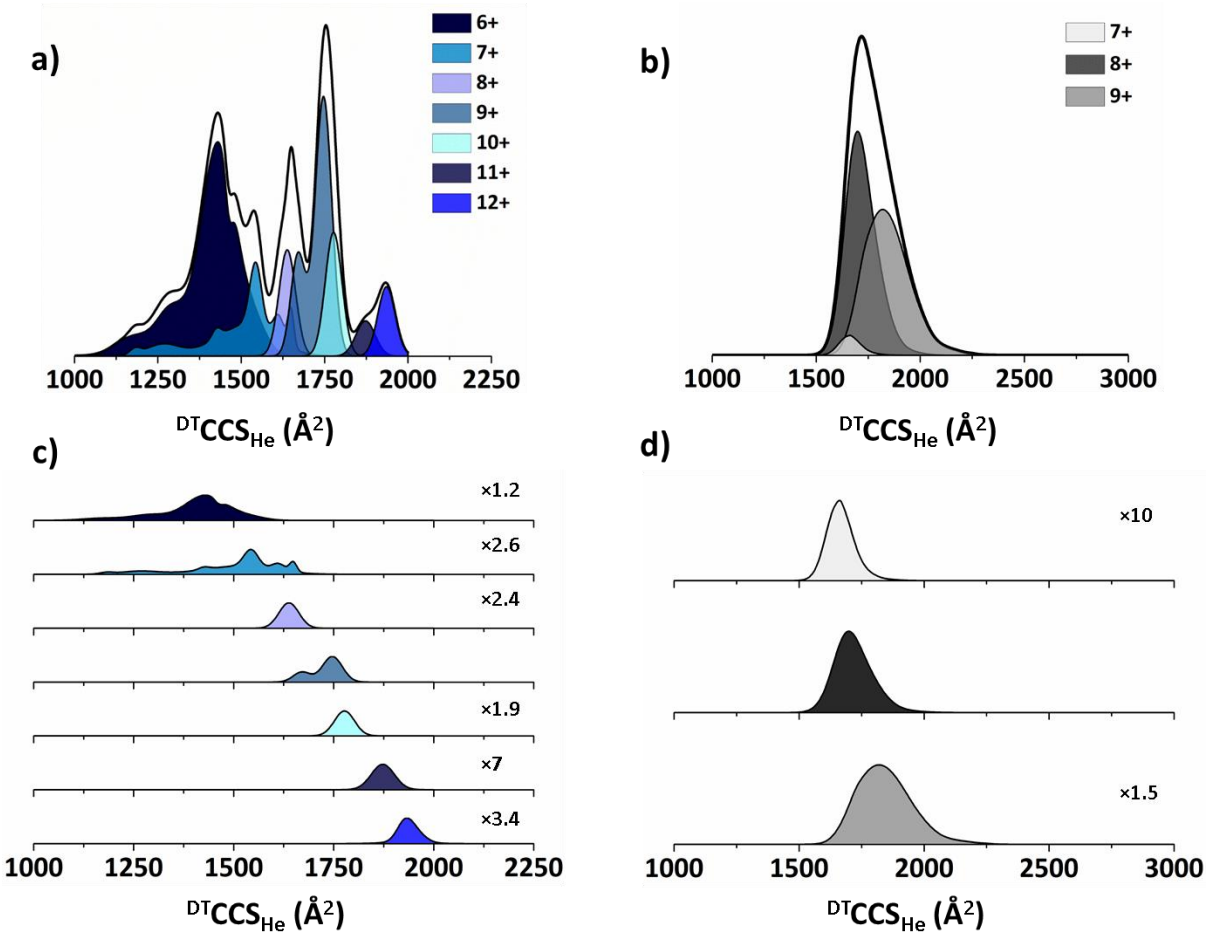

Figure 2.11 | (a) A total CCSD plot of the 7 charge states of the COR15A protein, showing a typical spread in CCS values within and across charge states. The two lowest charge states, z = 6, and z = 7 have distinctively different distributions to the higher charge states; they are much broader with a number of partially resolved conformational ensembles. For *z* = 8 only one conformational family can be observed. Interestingly the *z* = 9 charge state appears to represent an intermediate species which is comprised of two partially resolved peaks each representing a distinct conformation. Regarding the higher charge states (*z* = 10-12), the CCSDs are much narrower than *z* = 6 & 7 and consist of a single observable conformational family. (b) A total CCSD plot of the 3 charge states of myoglobin, each charge state is seen to have a singular Gaussian-like distribution which is due to only one resolvable conformational family being present for each. (c) A stack plot of the CCSD for each of the 7 charge states of COR15A. (d) A Stack plot of the CCSD for the 3 charge states of myoglobin. Magnification values relative to the most abundant charge state are represented in the right-hand corner of each row within the stacked plots.

For Myoglobin, Δ*z* = 3 and ΔCCS = ~400 Å^2 (ΔCCS = apex of the largest CCS – apex of the smallest CCS). The Gaussian-like distribution of the peaks indicates the presence of a single conformer for each charge state, although it can be noted that as the charge state increases so does the average CCS value. The low Δ*z* and ΔCCS as well as the single Gaussian-like distribution of each charge state peak indicates that myoglobin occupies a narrow conformational landscape with few closely related conformational families, as might be expected for a well ordered protein following ionisation from native-like solution conditions.[38] The total CCSD plot for COR15A, with Δ*z* =7 and ΔCCS of ~1000 Å^2, supports the assertion made from the broad CSD (Figure 2.5) that this protein presents to the mass spectrometer in numerous conformations. Additionally, for several charge states (*z* = 6, 7 and 9) there are multiple conformational families in the CCSD, indicative that COR15A occupies a conformational landscape which has many low energy conformations, which are close in energy to one another, but distinct in structure.

The charge states of *z* = 6, 7 and 9 of COR15A (Figure 2.11a) give rise to complex ATDs with numerous partially resolved Gaussian-like signatures attributable to the presence of multiple conformers for the same net charge (Figure 2.11c). In comparison, for myoglobin a Gaussian-like distribution is found for each charge state (Figure 2.11d) indicative of single closely related conformational families. Comparison between the IM-MS data for myoglobin and COR15A (Figure 2.11) which was taken on the same instrument with very similar conditions, reveals how dramatic the difference in structures between the two proteins are. This work highlights how we can use IM-MS to discern the conformational heterogeneity of any given protein/ protein complex.

The total CCSD also allows us to define potential conformational families that exist within the conformational landscape of the IDP in question; this is shown in Figure 2.12.

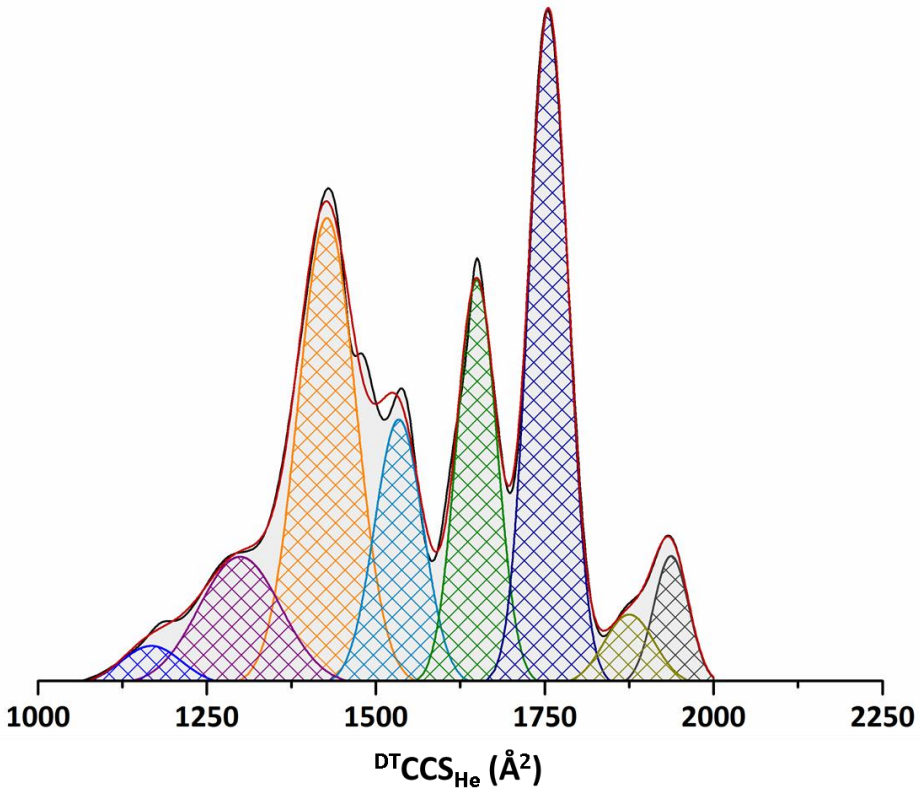

Figure 2.12 | Gaussian peaks were manually fitted to the total CCSD of COR15A and represent possible conformers present within the gas-phase conformational landscape of the protein. A total of eight peaks were fitted to the total CCSD and likely correspond to 8 different conformational populations of the COR15A protein. Fitting was done using the Gaussian fit function in the OriginPro software, the top of each candidate peak is selected manually and are then used as the peak maxima for Gaussian fit function.

By fitting Gaussian peaks to the total CCSD of COR15A we can identify possible conformational families within the overall gas phase landscape of the protein. This shows that in comparison to a globular protein, like myoglobin, COR15A most likely adopts numerous conformations in solution which can be sampled *in vacuo*.

2.5.4 A Predictive Framework for Disorder

We now understand that CSDs and CCSDs can be used to provide structural information on the conformational diversity and level of disorder within a protein. As we mentioned previously (Section 3.3), proteins have distinct features in their CSDs and CCSDs that can be used to categorise them as principally ordered or disordered when sprayed from certain solution conditions. We have developed a predictive framework that uses a proteins CSD and CCSD to predict the level of disorder within it and a 'toy model' that calculates a minimum and maximum boundary CCS for the protein. The framework and toy model have been written in python code and are available on GitHub for free use (https://github.com/ModronMan/Origami-Analyst).

2.5.4.1 Trends in collected IM-MS data – 'Beveridge-Barran' Plots

Using data acquired from many proteins we have developed an empirical predictive framework that can be used to evaluate the level of disorder in an unknown protein.[38] By plotting the range of CCS values (ΔCCS) and observed charge states (Δ*z*) for proteins against their molecular weight, Beveridge and co-workers[38] showed data plots and trends of the form below (Figure 2.13). These plots indicate that some proteins have a narrow charge state and CCS range which hardly varies with molecular weight, whilst some have a far broader range of charge states and a correspondingly large range of CCSs. The former category is primarily made up of proteins that are known to be structured following careful native mass spectrometry analysis and sample preparation as shown above, whereas the latter set, with large Δ*z* and ΔCCS consist either of conformationally dynamic proteins, sprayed from pseudo-native solutions, or from structured proteins sprayed from denaturing solutions.

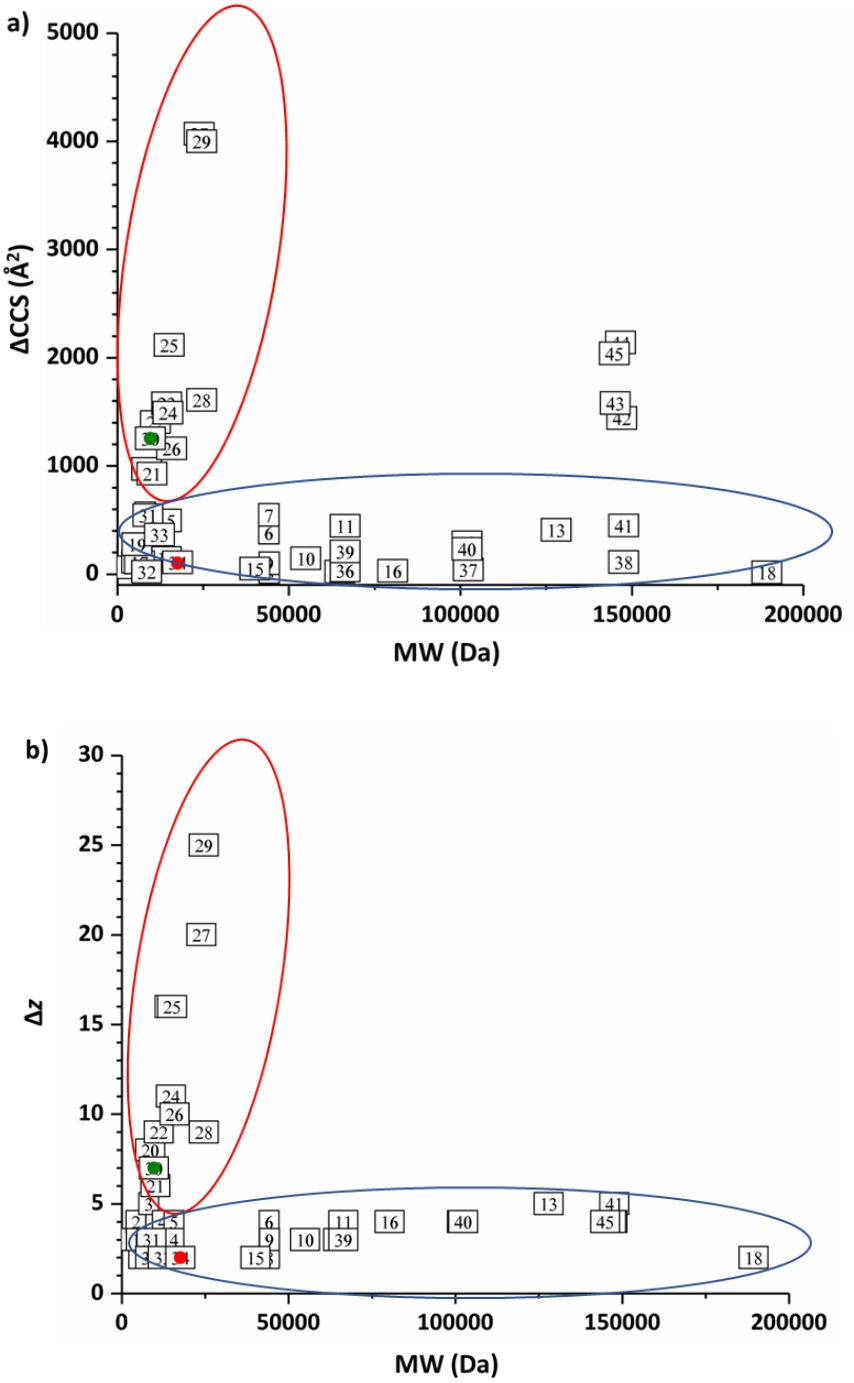

Figure 2.13 | Beveridge-Barran plots for a data set of proteins which can be found in Appendix 1. Each number corresponds to a protein, COR15A (green dot) and myoglobin (red dot) are included in the set. Numbers within the red oval represent proteins which exhibit more dynamic structures, while numbers within the blue oval represent proteins with more ordered structures. ΔCCS and Δz represent the difference in lowest and highest value of CCS and z respectively. (a) A plot of ΔCCS vs. molecular weight for a data set of proteins, including COR15a and myoglobin. Proteins with larger ΔCCS, meaning they occupy a wider conformational spread, tend

to be more structurally dynamic whereas proteins with a small ΔCCS, occupying a narrower conformational spread, tend to be more structured. The numbers 42 – 45 represent antibodies, which are seen to fall outside of the two regions of structured and dynamic proteins, highlighted by the coloured ovals. (b) A plot of Δz vs. molecular weight for a data set of proteins, including COR15a and myoglobin. Proteins with a larger Δz, tend to be more structurally dynamic whereas proteins with a smaller Δz exhibit more structure.

Following IM-MS analysis, the analyst can obtain and plot ΔCCS and Δ*z* vs. molecular weight and estimate the structural propensity for a given analyte depending upon which region it resides in. For example, following the procedures described above, COR15A falls within the 'disordered' region of the graphs, whereas myoglobin is found in the 'ordered' region. This demonstrates how these plots can be used to give valuable insight into the structure of an unknown protein.

2.5.4.2 A Toy Model for Predicting CCS Boundaries

In addition to the Beveridge-Barran plots, we have developed a 'toy model' which calculates a theoretical upper and lower boundary of a proteins CCS based off its amino acid sequence.[38] The lower CCS boundary is calculated by assuming that a folded proteins' shape approximates to a sphere. This approach is representative of a globular highly compact form of a protein. We start by using a proteins molecular weight to calculate its volume *via* Equation 2.12 shown below:

$$V = {}^{M_w}/_{\rho} \qquad \text{Equation 2.12}$$

Where M_w is molecular weight of the protein, and ρ is the density of the protein; the value of protein density used is ρ = 0.904 Da Å^{-3}.[92] This volume can then be used to calculate the radius of a sphere as shown in Equation 2.13:

$$r = \left(\frac{3V}{4\pi}\right)^{1/3} \qquad \text{Equation 2.13}$$

The CCS of a sphere of this radius is then given by Equation 13:

$$CCS_{lower}(\text{Å}^2) = \pi r^2 = \pi \left(\frac{3V}{4\pi}\right)^{2/3} \qquad \text{Equation 2.14}$$

This equation gives the geometric size of the sphere of a compact protein. The value obtained by this equation however is exceeded by CCS numbers obtained in IM-MS using helium as a buffer gas. We therefore use a scaling parameter of 1.19 developed by De la Mora [86] to predict the smallest possible measured CCS. The scaling parameter is also applied to the upper boundary calculation.

For the upper boundary we model a completely unfolded protein stretched out from end-to-end, as a cylinder as shown below in Figure 2.14, where each amino acids volume is treated as a cube that when rotationally averaged forms a cylindrical projection.

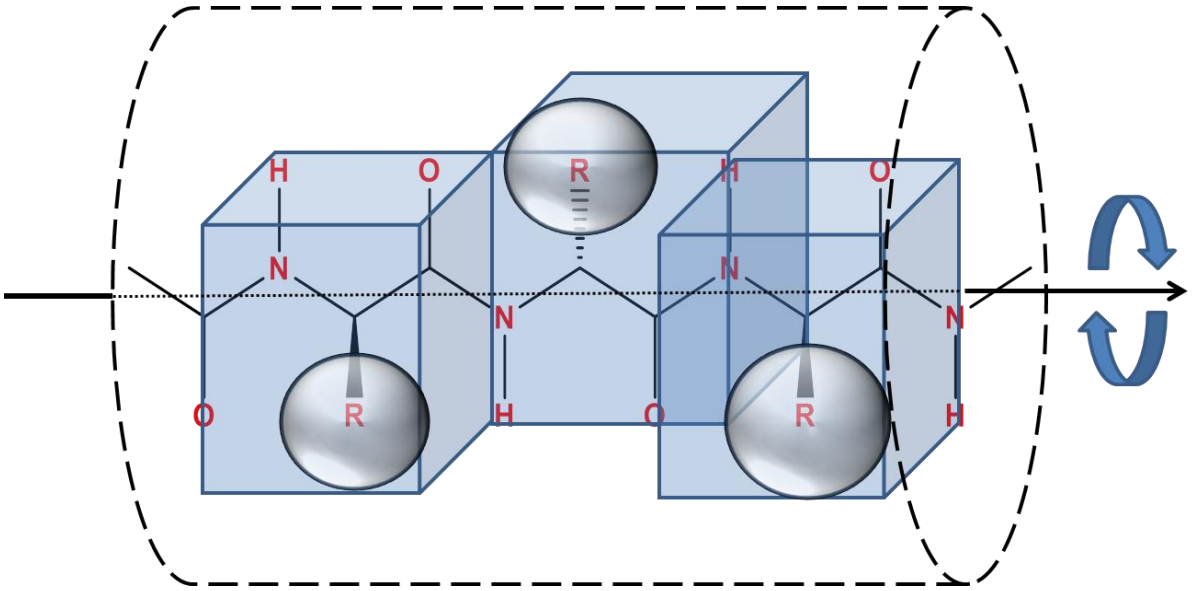

Figure 2.14 | The rotationally average cylindrical projection of a peptide chain, where each amino acid's volume is treated as a cube.

The first step is to use Equation 14 to determine the average volume of an amino acid in a protein's sequence:

$$\bar{V} = \frac{\sum_i^n V_i N_i}{n}$$ Equation 2.15

Where V_i is the volume of a specific amino acid, N_i is the corresponding number of a specific amino acid and n is the total number of amino acids in the sequence. Next we treat the radius of the cylinder to be equal to the length of the cube, as shown in Figure 2.15 below.

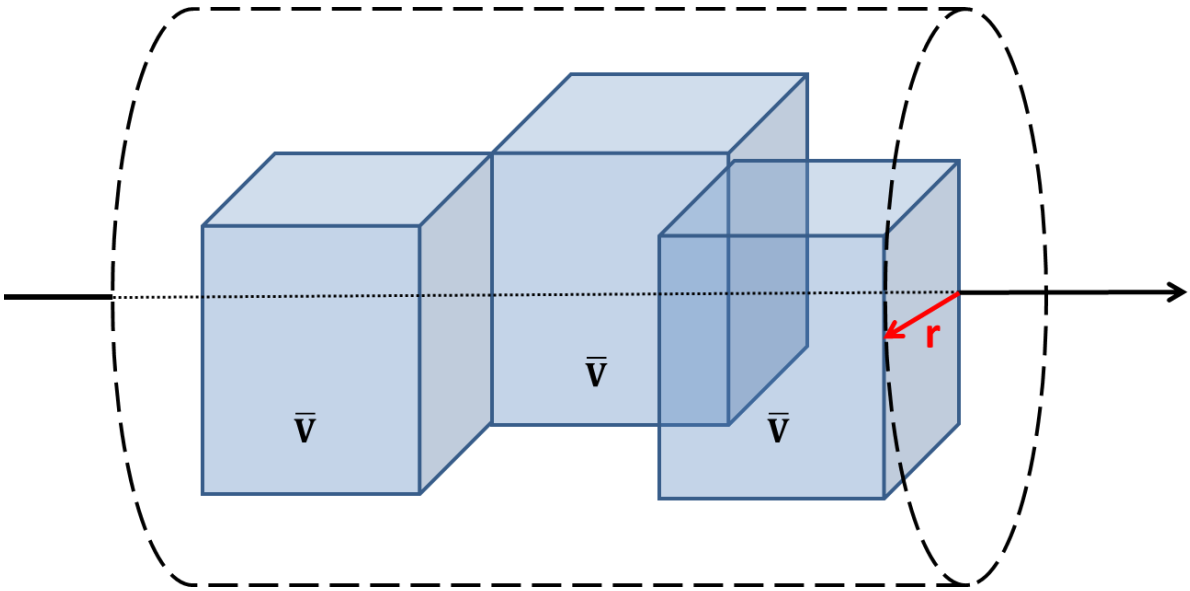

Figure 2.15 | The length of the cube is taken to be the radius of the cylinder.

Therefore, the radius of the cylinder is the cube root of the average amino acid volume. The CCS is then given by Equation 15, since the protein 'tumbles' in the drift tube:

$$CCS_{upper}(\text{Å}^2) = \left(\frac{4}{\pi}\right) rl + 2r^2$$ Equation 2.16

Where length, l, comes from knowing that the furthest distance between α-carbons in a protein chain is 3.63 Å [93]. Therefore, the maximum length of a linear polypeptide chain with n residues is given by $l = n(3.63)$.

2.6 Summary and Conclusion

In summary, our approach to the analysis of IDPs using native IM-MS can be simplified into three major steps: sample preparation and purification (2.1), ion mobility and mass spectrometry experimental procedures (2.2), and data analysis (3).

In the first step, sample preparation and purification, we have tried to highlight the importance of preparing a solution at a pH appropriate for the subject protein. A number of salts are commonly used within these solutions which can act as buffers upon pH adjustment. Furthermore, these salts need to be volatile enough to leave the protein upon desolvation. It is important to note that proteins are often dissolved in solutions that are not suitable for direct infusion MS experiments whilst even lyophilised samples are often saturated with salts and contain impurities. Considering this we buffer exchange sample solutions to transfer the protein to an MS compatible solvent, this process also helps to purify the protein. We use spin columns, as they are compatible with small sample volumes (20 – 75 µL).

Once sample preparation is complete, we move onto the second step which is the running of a sample on an (IM)-MS instrument. We find it is important to initially run a 'standard' protein sample on the instrument before tackling the IDP sample. This allows the analyst to check that the tip and instrument are both working correctly whilst permitting optimisation of instrument settings (Table 2.2) to attain a good starting point for the analysis of the IDP. When the standard is spraying stably with good ion transmission, analysis of the subject IDP can commence. The optimised parameters used for the standard will become the starting parameters for the IDP; depending upon the quality of the mass spectrum and ATDs (which we can observe in real time), alterations may be needed to improve the signal intensity/instrument 'softness'. On the other hand, poor spectral quality observed at this

stage may be due to the sample; therefore, further purification may be required to resolve the issue. Once the IDP is running stably and at good quality, acquisition of the mass spectrum can be carried out. With the acquired mass spectrum, we can use Equation 3 and 4 to determine the molecular weight and CSD of the subject protein, where for an IDP a broad CSD is likely observed and is representative of the numerous conformations occupied. Following this we move on to IM analysis of each charge state seen in the mass spectrum. After mass selecting the charge state of interest, we can acquire IM data for the ion at ≥ 5 different drift voltages.

Upon acquisition of the ATDs we can begin the final step of data analysis; this involves using the ATDs alongside Equation 1 to generate CCSDs for each charge state. In addition, we can also produce a total CCSD that encompasses the whole conformational landscape of the IDP *in vacuo*. The total CCSD of IDPs are observed to be very different to globular proteins which exhibit Gaussian-like distributions (Figure 2.11b) adhering to a small conformational landscape. Fitting Gaussian peaks to the IDPs total CCSD (Figure 2.12) shows us that in comparison to globular proteins, more structures are observed within the conformational landscape of the IDP.

Lastly, we have designed a predictive framework (3.4) to help define the level of disorder in the proteins we analyse. By plotting the ΔCCS and Δz of the subject protein against its molecular weight on a graph, with a large data set of other proteins (Figure 2.13, Appendix 1), we can infer the level of disorder within the protein under study. Furthermore, we have developed a 'toy model' (3.4.2) that can calculate the maximum and minimum boundaries of a given proteins CCS which provides an additional dimension to this predictive framework.

2.7 Molecular Modelling

2.7.1 Introduction

Computational chemistry is a branch of science where chemistry and computing meet. It involves the analysis and simulation of chemical systems using computers. The advent and development of electronic computers over the last century has vastly improved the speed at which theoretical calculations and mathematical modelling can be done at. Computational chemistry can be split into three main areas of study, the first includes the study of electronic structure using quantum mechanical (QM) calculations, large and complex biological systems can also be studied using classical mechanics and lastly the physicochemical properties of a system can be understood using cheminformatics.[94]

Though it is the most fundamental modelling method QM calculations can only be solved for very simple systems. Approximations can be made to allow more complex systems to be analysed, one very important example being the Born-Oppenheimer approximation that decouples the nuclear and electronic motions, furthermore, semi-empirical methods can also be applied. These methods use both chemical theory and empirical data derived from experiment to help reduce the computational cost of these calculations.[94]

In spite of all this however, the computational power and cost required to run large biological systems at this level of theory is immense and impractical. Molecular Mechanics (MM) steps in here as a technique that can overcome this problem. It uses classical descriptions of chemical systems with the Born-Oppenheimer approximation to allow the energy of the system to be described as a function of the nuclear coordinates.[94,95]

2.7.2 Molecular Mechanics

Molecular Mechanics (MM) represents atoms as hard spheres and models the interactions that occur between them using additive classical functions with parameters derived from *ab initio* calculations and experimental data. The potential energy of the system is modelled by a collection of equations and parameters defined as the force field, which has the general form:

$$E_{total} = E_{bonds} + E_{angles} + E_{dihedrals} + E_{vdW} + E_{electrostatics} \qquad \text{Equation 2.17}$$

Where E_{total} represents the total potential energy of the system, E_{bonds}, E_{angles}, $E_{dihedrals}$, E_{vdW}, and $E_{electrostatics}$ represent the energies associated with bonds, angles, dihedrals, van der Waals forces and electrostatics respectively.[94,96]

2.7.2.1 CHARMM Force Fields

The Chemistry at Harvard Macromolecular Mechanics (CHARMM) series is a biologically exclusive force field used in the simulations carried out in this thesis.[97–100] It was developed in the 1980s in association with a Molecular Dynamics (MD) software package of the same name by the Martin Karplus group at Harvard University.[101,102] Its initial form, CHARMM19, used a united atom representation with special treatment for polar and non-polar hydrogen atoms treating them as explicit and heavy atoms respectively.[103] The following versions of CHARMM used all-atom representations with the first being developed by Mackerell *et al.*[97] in 1998, called CHARMM22 with its energy potential taking the following form:

$$E_{total} = \sum_{bonds} k_r(r - r_0)^2 + \sum_{UB} k_{UB}\left(r_{1,3} - r_{1,3}^0\right) + \sum_{angles} k_\theta(\theta - \theta_0)^2 +$$

$$\sum_{dihedrals} k_\phi(1 + \cos(n\phi - \gamma)) + \sum_{improper} k_{imp}(\chi - \chi_0)^2 +$$

$$\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\left(4\epsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] + \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}}\right) \qquad \text{Equation 2.18}$$

Where:

- The first term is a summation of all bonds modelled as harmonic oscillators with bond length, r, equilibrium bond length, r_0, and force constant, k_r.
- The second term is a summation of all Urey-Bradley (UB) interactions. It models angles using a harmonic spring between atoms 1 and 3 with bond length, $r_{1,3}$, equilibrium length, $r_{1,3}^0$, and force constant k_{UB}.
- The third term is a summation of all angles, θ, modelled as harmonic oscillators with equilibrium angle, θ_0, and force constant k_θ.
- The fourth term is a summation of all dihedral angles, ϕ, modelled using a cosine term. Where n, is the torsional multiplicity, γ, is the phase factor, and k_ϕ, is the force constant.
- The fifth term consists of summing all the improper dihedrals, modelled using a harmonic potential with an out-of-plane angle, χ, an average angle, χ_0, and force constant, k_{imp}.
- The last term models the van der Waals and electrostatic interactions between all possible atomic pairs, i and j, using a Lennard-Jones and Coulombic potential. Where σ_{ij}, represents the collision diameter, ϵ_{ij}, is the well depth, r_{ij}, the interatomic distance, and q_i and q_j represent point charges.

The CHARMM22 force field was continually improved over the years through reparameterisation, corrections and new additions. Of interest to this thesis was the improvements added to the CHARMM36 and CHARMM36m versions of the force field. CHARMM36 saw the addition of a refined protein backbone CMAP potential as well as amino acid side chain dihedral parameters fitted to NMR data of IDPs. This lead to improved sampling of conformational changes in proteins. CHARMM36m reweighted the backbone

CMAP potential, enhancing the sampling of IDPs, it also reduced the over-stabilisation of α-helices which CHARMM was known for.[100,104,105]

2.7.2.2 Solvation Methods

Numerous solvation methods exist for simulating biological molecules in solution. In regard to the simulations carried out in this thesis an explicit water model was used where each individual water molecule is rendered and parameterised. There also exists implicit water models where the solvent is modelled as a continuum with a defined dielectric constant.[106] It is important to use the water model that is most compatible with the force field being used. In this case a modified TIP3P (mTIP3P) that includes Lennard-Jones parameters on the hydrogen atoms was used in conjunction with the CHARMM36m force field.[105]

2.7.3 Molecular Dynamics

The simulations done in this thesis are Molecular Dynamic (MD) simulations, which model the dynamic behaviour of atomic positions and velocities using Newton's 2nd law of motion:

$$F_{r_i} = m_i \frac{d^2 r_i}{dt^2} = m_i \left(\frac{d^2 x_i}{dt^2} + \frac{d^2 y_i}{dt^2} + \frac{d^2 z_i}{dt^2} \right) = m_i \frac{d_{v_i}}{dt} = m_i a_i \qquad \text{Equation 2.19}$$

Where F_{r_i}, represents the force exerted on particle i, m_i is the particle mass, r_i, is the particle position vector (in Cartesian coordinates it is x_i, y_i and z_i), v_i represents the particle velocity, a_i is particle acceleration and t is time.

Using the MM principles discussed previously, we can convert the potential energy provided by a force field (CHARMM36m) into the forces felt by each atom in the system using:

$$F_{r_i} = -\nabla \mathrm{E}_{r_i} \qquad \text{Equation 2.20}$$

These forces are then used to calculate new atomic positions and velocities using integrator algorithms that break down the integration into small time steps (δt) in order to calculate the new positions and velocities of the system as it evolves.[94] The Leap-Frog algorithm was used for all simulations in this thesis and takes the following form:

$$\boldsymbol{r}(t + \delta t) = \boldsymbol{r}(t) + \delta t \boldsymbol{v}\left(t + \frac{\delta t}{2}\right) \qquad \text{Equation 2.1}$$

$$\boldsymbol{v}\left(t + \frac{\delta t}{2}\right) = \boldsymbol{v}\left(t - \frac{\delta t}{2}\right) + \delta t \boldsymbol{a}(t) \qquad \text{Equation 2.22}$$

$$\boldsymbol{v}(t) = \frac{1}{2}\left[\boldsymbol{v}\left(t - \frac{\delta t}{2}\right) + \boldsymbol{v}\left(t - \frac{\delta t}{2}\right)\right] \qquad \text{Equation 2.23}$$

In the leap-frog algorithm the positions at the next step, $\boldsymbol{r}(t + \delta t)$, are calculated from the velocities at the next half time step, $\boldsymbol{v}\left(t + \frac{\delta t}{2}\right)$. The acceleration, $\boldsymbol{a}(t)$, can be calculated from

Equation 2.19 and $\boldsymbol{v}\left(t-\frac{\delta t}{2}\right)$ is known from the previous step. Lastly $\boldsymbol{v}(t)$ is calculated from Equation 2.23.[94]

2.7.4 Summary

To summarise we have discussed in detail the theory behind the computational methodologies we have employed in this thesis. MM is our modelling technique that treats atoms as hard spheres and bonds as harmonic springs and is defined by a general potential energy equation (Equation 2.17). We use a force field (CHARMM) to provide us with the potential energy equation (Equation 2.18) that includes parameters for amino acids and lipids as well as a modified backbone CMAP which improves the sampling of IDPs. MD was used to simulate the dynamics of the systems we studied by calculating the forces felt between each atom (Equation 2.19 and 2.20) and the leap-frog algorithm was employed to calculate the new velocities and positions as the system evolved (Equation 2.21 – 2.23).

2.8 Acknowledgements

We acknowledge funding from The University of Manchester and the A*STAR program for the PhD. studentship of DS and from EPSRC and the MRC for the studentships to LM and AF respectively. We also acknowledge the BBSRC for grants BB/L002655/1 and BB/M017702/1 which support our work, as well as thanking Waters Corp. and Agilent for support and help with our instrumentation.

2.9 References

1. Uversky, V. N. & Dunker, A. K. Understanding protein non-folding. *Biochimica et Biophysica Acta - Proteins and Proteomics* vol. 1804 1231–1264 (2010).

2. Tompa, P. On the supertertiary structure of proteins. *Nat. Chem. Biol.* **8**, 597–600 (2012).

3. Wu, H. & Fuxreiter, M. The Structure and Dynamics of Higher-Order Assemblies: Amyloids, Signalosomes, and Granules. *Cell* vol. 165 1055–1066 (2016).

4. Uversky, V. N., Oldfield, C. J. & Dunker, a K. Intrinsically Disordered Proteins in Human Diseases: Introducing the D 2 Concept. *Annu. Rev. Biophys.* **37**, 215–246 (2008).

5. Ananiev, J., Tchernev, G., Patterson, J. W., Gulubova, M. & Ganchev, G. p53 - 'The Guardian of Genome'. *Acta Medica Bulg.* **38**, 72–82 (2011).

6. Sormanni, P. *et al.* Simultaneous quantification of protein order and disorder. *Nature Chemical Biology* vol. 13 339–342 (2017).

7. Tompa, P. Unstructural biology coming of age. *Current Opinion in Structural Biology* vol. 21 419–425 (2011).

8. Wright, P. E. & Dyson, H. J. Intrinsically unstructured proteins: re-

assessing the protein structure-function paradigm. *J. Mol. Biol.* **293**, 321–331 (1999).

9. Tompa, P. Intrinsically unstructured proteins. *Trends Biochem. Sci.* **27**, 527–533 (2002).

10. Dunker, A. K. *et al.* Intrinsically disordered protein. *J. Mol. Graph. Model.* **19**, 26–59 (2001).

11. Dyson, H. J. & Wright, P. E. Intrinsically unstructured proteins and their functions. *Nature Reviews Molecular Cell Biology* vol. 6 197–208 (2005).

12. Daughdrill, G. W., Hanely, L. J. & Dahlquist, F. W. The C-terminal half of the anti-sigma factor FlgM contains a dynamic equilibrium solution structure favoring helical conformations. *Biochemistry* **37**, 1076–1082 (1998).

13. Ward, J. J., Sodhi, J. S., McGuffin, L. J., Buxton, B. F. & Jones, D. T. Prediction and Functional Analysis of Native Disorder in Proteins from the Three Kingdoms of Life. *J. Mol. Biol.* **337**, 635–645 (2004).

14. Stuchfield, D. & Barran, P. Unique insights to intrinsically disordered proteins provided by ion mobility mass spectrometry. *Current Opinion in Chemical Biology* vol. 42 177–185 (2018).

15. Faull, P. A. *et al.* 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**, (2010).

16. D'Urzo, A. *et al.* Molecular basis for structural heterogeneity of an intrinsically disordered protein bound to a partner by combined ESI-IM-MS and modeling. *J. Am. Soc. Mass Spectrom.* **26**, 472–481 (2015).

17. Jurneczko, E. *et al.* Intrinsic disorder in proteins: a challenge for (un)structural biology met by ion mobility-mass spectrometry. *Biochem. Soc. Trans.* **40**, 1021–6 (2012).

18. Beveridge, R., Chappuis, Q., Macphee, C. & Barran, P. Mass spectrometry methods for intrinsically disordered proteins. *Analyst* vol. 138 32–42 (2013).

19. Borysik, A. J., Kovacs, D., Guharoy, M. & Tompa, P. Ensemble Methods Enable a New Definition for the Solution to Gas-Phase Transfer of Intrinsically Disordered Proteins. *J. Am. Chem. Soc.* **137**, 13807–13817 (2015).

20. Beveridge, R., Chappuis, Q., Macphee, C. & Barran, P. Mass spectrometry methods for intrinsically disordered proteins. *Analyst* **138**, 32–42 (2013).

21. Fenn, J. B., Mann, M., Meng, C. K. A. I., Wong, S. F. & Whitehouse, C. M. Electrospray Ionisation for Mass Spectrometry of Large Biomolecules. *Science (80-.).* **246**, 64–71 (1989).

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

23. Hernández, H. & Robinson, C. V. Determining the stoichiometry and interactions of macromolecular assemblies from mass spectrometry. *Nat. Protoc.* **2**, 715–726 (2007).

24. Uetrecht, C. *et al.* Stability and shape of hepatitis B virus capsids in vacuo. *Angew. Chemie - Int. Ed.* **47**, 6247–6251 (2008).

25. Zhou, M., Dagan, S. & Wysocki, V. H. Protein Subunits Released by Surface Collisions of Noncovalent Complexes: Nativelike Compact Structures Revealed by Ion Mobility Mass Spectrometry. *Angew. Chemie Int. Ed.* **51**, 4336–4339 (2012).

26. Sahasrabuddhe, A. *et al.* Confirmation of intersubunit connectivity and topology of designed protein complexes by native MS. *Proc. Natl. Acad. Sci. U. S. A.* **115**, 1268–1273 (2018).

27. Ruotolo, B. T., Benesch, J. L. P., Sandercock, A. M., Hyung, S.-J. &

Robinson, C. V. Ion mobility–mass spectrometry analysis of large protein complexes. *Nat. Protoc.* **3**, 1139–1152 (2008).

28. Young, G. *et al.* Quantitative mass imaging of single biological macromolecules. *Science* **360**, 423–427 (2018).

29. Breuker, K., Brüschweiler, S. & Tollinger, M. Electrostatic Stabilization of a Native Protein Structure in the Gas Phase. *Angew. Chemie Int. Ed.* **50**, 873–877 (2011).

30. Badman, E. R., Myung, S. & Clemmer, D. E. Evidence for unfolding and refolding of gas-phase cytochrome c ions in a Paul trap. *J. Am. Soc. Mass Spectrom.* **16**, 1493–1497 (2005).

31. Freitas, M. A., Hendrickson, C. L., Emmett, M. R. & Marshall, A. G. Gas-phase bovine ubiquitin cation conformations resolved by gas-phase hydrogen/deuterium exchange rate and extent. *Int. J. Mass Spectrom.* **185**–**187**, 565–575 (1999).

32. Uetrecht, C., Rose, R. J., van Duijn, E., Lorenzen, K. & Heck, A. J. R. Ion mobility mass spectrometry of proteins and proteinassemblies. *Chem. Soc. Rev.* **39**, 1633–1655 (2010).

33. Chen, G. & Pramanik, B. N. LC-MS for protein characterization: Current capabilities and future trends. *Expert Review of Proteomics* vol. 5 435–

444 (2008).

34. Zubarev, R. A. *et al.* Electron capture dissociation for structural characterization of multiply charged protein cations. *Anal. Chem.* **72**, 563–73 (2000).

35. Vorontsov, E. A., Rensen, E., Prangishvili, D., Krupovic, M. & Chamot-Rooke, J. Abundant Lysine Methylation and N-Terminal Acetylation in Sulfolobus islandicus Revealed by Bottom-Up and Top-Down Proteomics. *Mol. Cell. Proteomics* **15**, 3388–3404 (2016).

36. and, A. D. & Kaltashov*, I. A. Detection of Multiple Protein Conformational Ensembles in Solution via Deconvolution of Charge-State Distributions in ESI MS. (2001) doi:10.1021/AC010713F.

37. Testa, L. *et al.* Extracting structural information from charge-state distributions of intrinsically disordered proteins by non-denaturing electrospray-ionization mass spectrometry. *Intrinsically Disord. proteins* **1**, e25068 (2013).

38. Beveridge, R. *et al.* A mass-spectrometry-based framework to define the extent of disorder in proteins. *Analytical Chemistry* vol. 86 10979–10991 (2014).

39. Frimpong, A. K., Abzalimov, R. R., Uversky, V. N. & Kaltashov, I. A.

Characterizationof intrinsically disordered proteins with electrospray inization mass spectrometry: Conformationl heterogeneity of ??-synuciein. *Proteins Struct. Funct. Bioinforma.* **78**, 714–722 (2010).

40. Konermann, L., Collings, B. A. & Douglas, D. J. Cytochrome c folding kinetics studied by time-resolved electrospray ionization mass spectrometry. *Biochemistry* **36**, 5554–5559 (1997).

41. Konermann, L. A minimalist model for exploring conformational effects on the electrospray charge state distribution of proteins. *J. Phys. Chem. B* **111**, 6534–6543 (2007).

42. Leney, A. C. & Heck, A. J. R. Native Mass Spectrometry: What is in the Name? *J. Am. Soc. Mass Spectrom.* (2017) doi:10.1007/s13361-016-1545-3.

43. von Helden, G., Wyttenbach, T. & Bowers, M. T. Inclusion of a MALDI ion source in the ion chromatography technique: conformational information on polymer and biomolecular ions. *Int. J. Mass Spectrom. Ion Process.* **146**–**147**, 349–364 (1995).

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

45. Mao, Y., Woenckhaus, J., Kolafa, J., Ratner, M. A. & Jarrold, M. F. Thermal unfolding of unsolvated cytochrome c: Experiment and molecular dynamics simulations. *J. Am. Chem. Soc.* **121**, 2712–2721 (1999).

46. Wyttenbach, T., Von Helden, G. & Bowers, M. T. Gas-phase conformation of biological molecules: Bradykinin. *J. Am. Chem. Soc.* **118**, 8355–8364 (1996).

47. Kent J. Gillig *et al.* Coupling High-Pressure MALDI with Ion Mobility/Orthogonal Time-of-Flight Mass Spectrometry. (2000) doi:10.1021/AC0005619.

48. Gerlich, D. *Inhomogeneous RF Fields: A Versatile Tool for the Study of Processes with Slow Ions*. *Advances in Chemical Physics* (1992). doi:10.1002/9780470141397.ch1.

49. Giles, K. *et al.* Applications of a travelling wave-based radio-frequency-only stacked ring ion guide. *Rapid Commun. Mass Spectrom.* **18**, 2401–2414 (2004).

50. Pringle, S. D. *et al.* An investigation of the mobility separation of some peptide and protein ions using a new hybrid quadrupole/travelling wave IMS/oa-ToF instrument. *Int. J. Mass Spectrom.* **261**, 1–12 (2007).

51. E. A. Mason, E. W. M. *Transport Properties of Ions in Gases*. (John WIley &

Sons Inc., 1988).

52. Revercomb, H. E. & Mason, E. A. Theory of Plasma Chromatography/Gaseous Electrophoresis. A Review. *Anal. Chem.* **47**, 970–983 (1975).

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

54. Lanucara, F., Holman, S. W., Gray, C. J. & Eyers, C. E. The power of ion mobility-mass spectrometry for structural characterization and the study of conformational dynamics. *Nat. Chem.* **6**, 281–294 (2014).

55. Faull, P. A. *et al.* 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 (2010).

56. Pagel, K., Natan, E., Hall, Z., Fersht, A. R. & Robinson, C. V. Intrinsically disordered p53 and its complexes populate compact conformations in the gas phase. *Angew. Chemie - Int. Ed.* **52**, 361–365 (2013).

57. Jurneczko, E. *et al.* Probing the conformational diversity of cancer-

associated mutations in p53 with ion-mobility mass spectrometry. *Angew. Chemie - Int. Ed.* **52**, 4370–4374 (2013).

58. Dickinson, E. R. *et al.* The use of ion mobility mass spectrometry to probe modulation of the structure of p53 and of MDM2 by small molecule inhibitors. *Front. Mol. Biosci.* **2**, 39 (2015).

59. Sinz, A., Arlt, C., Chorev, D. & Sharon, M. Chemical cross-linking and native mass spectrometry: A fruitful combination for structural biology. *Protein Sci.* **24**, 1193–1209 (2015).

60. Phillips, A. S. *et al.* Conformational dynamics of α-synuclein: insights from mass spectrometry. *Analyst* **140**, 3070–3081 (2015).

61. Bernstein, S. L. *et al.* α-Synuclein: Stable compact and extended monomeric structures and pH dependence of dimer formation. *J. Am. Soc. Mass Spectrom.* **15**, 1435–1443 (2004).

62. Thalhammer, A. & Hincha, D. K. A mechanistic model of COR15 protein function in plant freezing tolerance: Integration of structural and functional characteristics. *Plant Signal. Behav.* **9**, e977722–e977723 (2014).

63. Candat, A. *et al.* Experimental determination of organelle targeting-peptide cleavage sites using transient expression of green fluorescent

protein translational fusions. *Anal. Biochem.* **434**, 44–51 (2013).

64. Lin, C. & Thomashow, M. F. DNA Sequence Analysis of a Complementary DNA for Cold-Regulated Arabidopsis Gene cor15 and Characterization of the COR 15 Polypeptide. *Plant Physiol.* **99**, 519–525 (1992).

65. Nakayama, K. *et al.* Arabidopsis Cor15am Is a Chloroplast Stromal Protein That Has Cryoprotective Activity and Forms Oligomers. *PLANT Physiol.* **144**, 513–523 (2007).

66. Kanu, A. B., Dwivedi, P., Tam, M., Matz, L. & Hill, H. H. Ion mobility-mass spectrometry. *Journal of Mass Spectrometry* vol. 43 1–22 (2008).

67. Juan, D. & Simon, J. Revista Mexicana De. *Rev. Mex. física* **77**, 1–26 (2012).

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

69. Ridgeway, M. E., Lubeck, M., Jordens, J., Mann, M. & Park, M. A. Trapped ion mobility spectrometry: A short review. *Int. J. Mass Spectrom.* **425**, 22–35 (2018).

70. Ujma, J., Giles, K., Morris, M. & Barran, P. E. New High Resolution Ion Mobility Mass Spectrometer Capable of Measurements of Collision Cross Sections from 150 to 520 K. *Anal. Chem.* **88**, 9469–9478 (2016).

71. McCullough, B. J. *et al.* Development of an ion mobility quadrupole time of flight mass spectrometer. *Anal. Chem.* **80**, 6336–6344 (2008).

72. Mortensen, D. N. & Williams, E. R. Electrothermal supercharging of proteins in native MS: Effects of protein isoelectric point, buffer, and nanoESI-emitter tip size. *Analyst* **141**, 5598–5606 (2016).

73. Raijmakers, R., Kraiczek, K., De Jong, A. P., Mohammed, S. & Heck, A. J. R. Exploring the human leukocyte phosphoproteome using a microfluidic reversed-phase-TiO2reversed-phase high-performance liquid chromatography phosphochip coupled to a quadrupole time-of-flight mass spectrometer. *Anal. Chem.* **82**, 824–832 (2010).

74. Visser, N. F. C., Scholten, A., Van Den Heuvel, R. H. H. & Heck, A. J. R. Surface-plasmon-resonance-based chemical proteomics: Efficient specific extraction and semiquantitative identification of cyclic nucleotide-binding proteins from cellular lysates by using a combination of surface plasmon resonance, sequential elution and . *ChemBioChem* **8**, 298–305 (2007).

75. Gan, J. *et al.* Native Mass Spectrometry of Recombinant Proteins from

Crude Cell Lysates. *Anal. Chem.* **89**, 4398–4404 (2017).

76. Mohanta, D. & Jana, M. Can 2,2,2-trifluoroethanol be an efficient protein denaturant than methanol and ethanol under thermal stress? *Phys. Chem. Chem. Phys.* **20**, 9886–9896 (2018).

77. Bush, M. F. *et al.* Collision cross sections of proteins and their complexes: A calibration framework and database for gas-phase structural biology. *Anal. Chem.* **82**, 9557–9565 (2010).

78. Kurulugama, R. T., Darland, E., Kuhlmann, F., Stafford, G. & Fjeldsted, J. Evaluation of drift gas selection in complex sample analyses using a high performance drift tube ion mobility-QTOF mass spectrometer. *Analyst* **140**, 6834–6844 (2015).

79. Sobott, F. & Robinson, C. V. Protein complexes gain momentum. *Current Opinion in Structural Biology* vol. 12 729–734 (2002).

80. Sobott, F., Hernández, H., McCammon, M. G., Tito, M. A. & Robinson, C. V. A tandem mass spectrometer for improved transmission and analysis of large macromolecular assemblies. *Anal. Chem.* **74**, 1402–1407 (2002).

81. Van Den Heuvel, R. H. H. *et al.* Improving the performance of a quadrupole time-of-flight instrument for macromolecular mass spectrometry. *Anal. Chem.* **78**, 7473–7483 (2006).

82. Migas, L. G., France, A. P., Bellina, B. & Barran, P. E. ORIGAMI: A software suite for activated ion mobility mass spectrometry (aIM-MS) applied to multimeric protein assemblies. *International Journal of Mass Spectrometry* vol. 427 20–28 (2017).

83. Jurneczko, E., Kalapothakis, J., Campuzano, I. D. G., Morris, M. & Barran, P. E. 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**, (2012).

84. Beveridge, R. *et al.* Relating gas phase to solution conformations: Lessons from disordered proteins. *Proteomics* **15**, 2872–2883 (2015).

85. Natalello, A., Santambrogio, C. & Grandori, R. Are Charge-State Distributions a Reliable Tool Describing Molecular Ensembles of Intrinsically Disordered Proteins by Native MS? *J. Am. Soc. Mass Spectrom.* **28**, 21–28 (2017).

86. Fernandez De La Mora, J. Electrospray ionization of large multiply charged species proceeds via Dole's charged residue mechanism. *Anal. Chim. Acta* **406**, 93–104 (2000).

87. Jurneczko, E. & Barran, P. E. 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 (2011).

88. Pacholarz, K. J. & Barran, P. E. Distinguishing Loss of Structure from Subunit Dissociation for Protein Complexes with Variable Temperature Ion Mobility Mass Spectrometry. *Anal. Chem.* **87**, (2015).

89. Pacholarz, K. J. *et al.* Hybrid Mass Spectrometry Approaches to Determine How L-Histidine Feedback Regulates the Enzyzme MtATP-Phosphoribosyltransferase. *Structure* **25**, (2017).

90. Pacholarz, K. J. *et al.* Dynamics of Intact Immunoglobulin G Explored by Drift-Tube Ion-Mobility Mass Spectrometry and Molecular Modeling. *Angew. Chemie Int. Ed.* **53**, 7765–7769 (2014).

91. Dickinson, E. R. *et al.* Insights into the Conformations of Three Structurally Diverse Proteins: Cytochrome c, p53, and MDM2, Provided by Variable-Temperature Ion Mobility Mass Spectrometry. *Anal. Chem.* **87**, (2015).

92. Fischer, H., Polikarpov, I. & Craievich, A. F. Average protein density is a molecular-weight-dependent function. *Protein Sci.* **13**, 2825–2828 (2009).

93. Heck, A. J. R. Native mass spectrometry: A bridge between interactomics and structural biology. *Nat. Methods* **5**, 927–933 (2008).

94. Leach, A. *Advanced ab initio methods, Density Functional Theory and Solid-state Quantum Mechanics*. *Molecular Modelling Principles and Applications* (Pearson Prentice Hall, 2009).

95. Understanding Molecular Simulation. *Underst. Mol. Simul.* (2002) doi:10.1016/B978-0-12-267351-1.X5000-7.

96. Cornell, W. D. *et al.* A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J. Am. Chem. Soc.* **117**, 5179–5197 (2002).

97. A. D. MacKerell, J. *et al.* All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins †. *J. Phys. Chem. B* **102**, 3586–3616 (1998).

98. MacKerell, A. D., Nilesh Banavali, J. & Foloppe, N. Development and Current Status of the CHARMM Force Field for Nucleic Acids. (2001) doi:10.1002/1097-0282.

99. Mackerell, A. D., Feig, M. & Brooks, C. L. Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulation. *J. Comput. Chem.* **25**, 1400–1415 (2004).

100. Huang, J. & MacKerell, A. D. CHARMM36 all-atom additive protein force field: Validation based on comparison to NMR data. *J. Comput. Chem.* **34**, 2135–2145 (2013).

101. Brooks, B. R. *et al.* CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* **4**, 187–217 (1983).

102. BR, B. *et al.* CHARMM: the biomolecular simulation program. *J. Comput. Chem.* **30**, 1545–1614 (2009).

103. Neria, E., Fischer, S. & Karplus, M. Simulation of activation free energies in molecular systems. *J. Chem. Phys.* **105**, 1902 (1998).

104. Best, R. B. *et al.* Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone ϕ, ψ and side-chain χ1 and χ2 dihedral angles. *J. Chem. Theory Comput.* **8**, 3257 (2012).

105. Huang, J. *et al.* CHARMM36m: An improved force field for folded and intrinsically disordered proteins. *Nat. Methods* **14**, 71–73 (2016).

106. Anandakrishnan, R., Drozdetski, A., Walker, R. C. & Onufriev, A. V. Speed of Conformational Change: Comparing Explicit and Implicit Solvent Molecular Dynamics Simulations. *Biophys. J.* **108**, 1153–1164 (2015).

Chapter 3: Mapping the Conformational Landscape of the Cold-Regulated Protein COR15A Using Molecular Dynamics and Ion Mobility Mass Spectrometry

3.1 Declaration

This chapter is reproduced from a completed manuscript I wrote that is set to be published:

D. Stuchfield, A. Thalhammer, and P. E. Barran, “Comparing MD simulations with Ion Mobility Mass Spectrometry Data Determines the Secondary Structural Features in the Conformational Landscape of the Cold-Regulated Protein COR15A”.

As the first author of the article, I drafted and edited the manuscript as well as designed and rendered all the figures.

Mapping the Conformational Landscape of the Cold-Regulated Protein COR15A Using Molecular Dynamics and Ion Mobility Mass Spectrometry

D. Stuchfield, A. Thalhammer and P. E. Barran*

Michael Barber Centre for Collaborative Mass Spectrometry, Manchester Institute for Biotechnology, School of Chemistry, University of Manchester, 131 Princess Street, Manchester M1 7DN, United Kingdom.

To Whom Correspondence should be addressed: perdita.barran@manchester.ac.uk

3.2 Abstract

COR15A is a Cold Regulated (COR) protein found in *Arabidopsis thaliana*. It plays a vital role in the stabilisation of chloroplast membranes in the leaf which have particular importance in low temperature and high salinity environments. Ion Mobility Mass Spectrometry analysis of COR15A yields wide bimodal Charge State and Collision Cross Section (CCS) distributions indicative of two dominant coexisting conformational families for this highly flexible protein. Molecular Dynamics (MD) was used to provide candidate geometries with which to compare with IM-MS data and elucidate dominant secondary structural features. The helical dry state of COR15A, named as such due to it being the structure the protein takes in the absence of water, provided the starting point for simulations with explicit water. Using elevated temperature for enhanced sampling we observe a decrease in the average CCS as the structure evolves from a predominantly helical structure to a more disordered and globular form. Evaluation of the CCS of COR15A throughout the trajectory yields a *pseudo*-CCS distribution (*p*CCS) which is similar to that obtained experimentally in the gas phase. Following this approach and after 30 ns of simulation time, the conformational landscape for solvated COR15A replicates what was found experimentally providing compelling evidence that the experimental gas phase data correspond to natively accessed states. We conclude that the use of IM-MS allows observation of the disordered globular forms which are kinetically trapped following desolvation as well as forms that are more akin to the dehydrated state.

3.3 Introduction

In recent years there has been increased awareness of the prevalence and functional role of intrinsically disordered proteins (IDP) with predictions that between 30 % and 70 % of all eukaryotic proteins contain some degree of intrinsic disorder. [1] [2–4] The recognition of the importance of such proteins initially stemmed from their role in proteopathic diseases such as Parkinson's and Alzheimer's disease[3] as well as functional roles in many biochemical networks and signalling and regulatory cascades[5] for example in tumour formation.[4,6]

The form of IDPs cannot be 'solved' in the same way as ordered proteins. No single structure can be used to describe an IDP instead they are better described as conformational ensembles,[7] wherein the primary sequence of an IDP may adopt transient secondary structures.[5] Furthermore, due to the conformational variability inherent in IDPs different experimental approaches to study them are required in comparison to their ordered kin. For example X-ray crystallography is limited by its experimental timescales which cannot necessarily capture details of the conformational dynamics of an IDP, and as with all bulk phase methods will display an averaged structure of the conformational ensemble sampled.[5,8]

Ion Mobility Mass Spectrometry offers unique advantages in the conformational analysis of proteins as well as for other biological and non-biological macromolecules.[8] Following transfer into the gas phase an IM-MS instrument separates molecules based on their mass, charge and shape providing stoichiometry along with conformational information.[8] In Drift Time Ion Mobility Spectrometry (DTIMS), a given protein ion is introduced in a short pulse to a chamber containing an inert buffer gas such as N_2 or He (drift gas). The packet of protein ions travel along the chamber under the influence of a weak electric field (5 – 50 Vcm^{-1}). As the ions drift along the tube, they collide with the drift gas and are separated depending on

their size and shape. An ion of a given *m/z* will have a corresponding arrival time distribution (ATD). Using the ATD we can calculate a Collision Cross Section Distribution that provides a direct measurement of a protein's conformational diversity.[9] It is possible to estimate the CCS of candidate geometries of any given molecule from input coordinate files using theoretical methods that model the interaction of the buffer gas with the molecular ion.[10] Force field derived structures and molecular dynamics approaches are often used in the analysis of large biological molecules, including IDPs, accessing biologically relevant timescales[5] and are able to explore candidate conformations. The CCS of these can then be calculated for comparison with experiment. One of the challenges in examining any protein structure in the gas phase, is to preserve solution phase conformation. IM-MS has previously proven to do this as the Δ*z* provides information on structure in the solution phase.[11] In addition the importance of charge on an IDP cannot be understated. Work carried out by Pappu *et al*.[12] defined a parameter (κ) to determine the distribution of charges in any polypetide sequence. They reported that for systems with well mixed charges (κ → 0) conformations involving self-avoiding random walks and coils were observed yet in comparison systems with well separated charge (κ → 1) exhibited hairpin-like folding. MD derived structures, together with IM-MS experiments are used here in the analysis of the intrinsically disordered Cold Regulated (COR) protein COR15A from *Arabidopsis thaliana*.

COR15A belongs to the Late Embryogenesis Abundant (LEA) proteins first characterised over 30 years ago.[5] Specifically, the protein is involved in the stabilisation of chloroplast and plasma membranes in the leaf during freezing. In dilute solution the protein is almost completely disordered and undergoes a conformational transition to an α-helical form upon dehydration. Expression levels of COR15A strongly increase in the process of cold acclimation, where the

plant acquires a higher freezing tolerance by exposure to low but non-freezing temperatures.[13] A mechanism of action for the protein was proposed by Thalhammer *et al.*[14], wherein they assert that upon freezing cellular dehydration occurs due to ice crystal formation in the apoplast causing water content to decrease. This induces the protein to fold into an α-helical structure and then associate with the inner envelope membrane of the chloroplast. The association further promotes α-helicity in COR15A, and in this form it is found to stabilise the membrane against the formation of hexagonal II (HII) phase lipid domains. This mechanism is shown visually in the supplementary Figure 3.1.

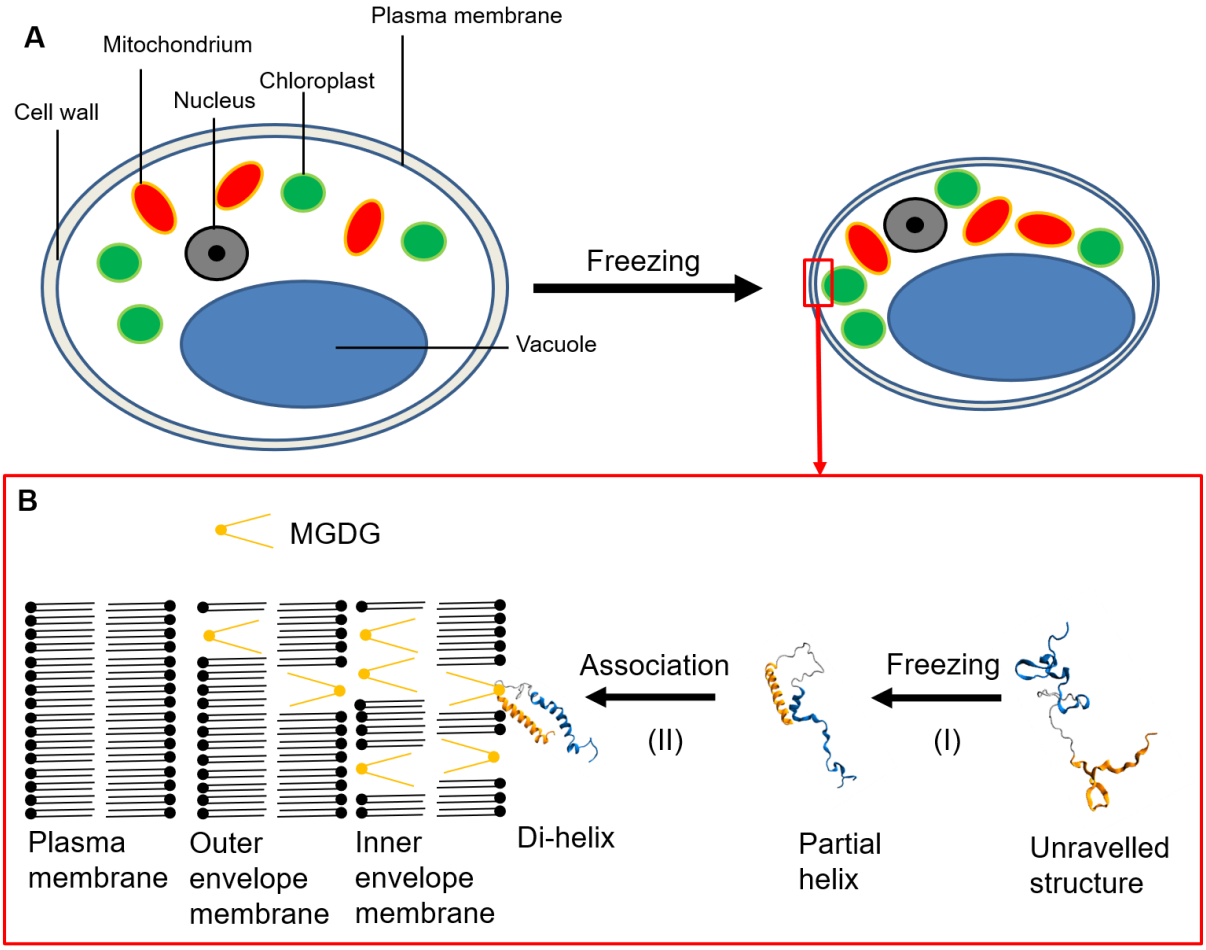

Figure 3.1 | COR15A mechanism of action adapted from Figure 1 in work published by Thalhammer *et al.*[1]. (A) Freezing causes cellular dehydration due to the formation of ice crystals in the extracellular spaces. This triggers a partial folding of COR15A as well as its association to the inner envelope membrane of the chloroplast enriched

with the non-bilayer lipid monogalactosyldiacylglycerol (MGDG). This association further COR15A promotes α-helical structure. (B) The inner envelope membrane is stabilised by protein association which prevents the formation of Hexagonal II (H_{II}) phase lipid domains within the inner envelope membrane and with neighbouring membranes such as the outer envelope membrane of the chloroplast and the plasma membrane. This results in the overall integrity of the cellular membrane being preserved upon freezing.

COR15A is reported to be highly unstructured when in hydrophilic environments and upon dehydration folds into an amphipathic α-helix. As such it represents an exciting candidate with which to consider the conformational transitions that occur to proteins as they enter the solvent free environment of the mass spectrometer. Here we investigate the conformational transitions that occur during the hydration process of COR15A using IM-MS and MD simulations.

3.4 Experimental

3.4.1 Ion Mobility Mass Spectrometry

The protocol used to carry out the experiments has been previously reported.[15] Briefly, all measurements were taken on a Waters Synapt G2 IM-MS modified to possess a linear drift field within the RF confining cell in the triwave assembly[16], using He as the drift gas.

Samples were analysed using nano Electrospray Ionisation (nESI) and sprayed from borosilicate tips (World Precision Instruments) with orifice diameter 3-6 µM pulled on a Sutter Instrument p1000. The COR15A protein was synthesised by the Thalhammer group and supplied as a lyophilised powder.[5] The sample was diluted in 200 mM Ammonium Acetate solution (Merck) with pH ~ 7.3 (using ammonium hydroxide) to a final concentration of 20 µM. Voltages of 0.7 – 1.2 kV were applied to the spray solution via an inserted platinum wire (Goodfellow), the source temperature was optimised between 50 – 80 ºC and cone voltages were applied ranging from 10 to 200 V for different experiments.

3.4.2 Molecular Dynamics Simulations

The starting structure for simulations was obtained from the Thalhammer group who used comparative modelling as previously described[5], the structure of COR15A in vacuum was obtained with the Internal Coordinate Mechanics (ICM) software using the amino acid sequence from the TAIR database. The simulation architecture consisted of a solvent filled box with the protein placed in the centre approximately 10 Å from the box edges. The CHARMM36m[17] force field was used accompanied with the TIP3P water model. [18] The system was initially minimised, followed by a standard equilibration procedure which consisted of two steps. The first equilibration step was a 1 ns simulation using an NVT ensemble where the number of particles, volume and temperature is kept constant. Following this was another nanosecond simulation using the NPT ensemble with a constant number of particles, pressure and temperature. During this period the protein was constrained to prevent any unfolding. The actual MD production run was carried out in an NPT ensemble. Five 100 ns replicates were run all at differing temperatures ranging from 300 – 600 K to speed up sampling. Standard periodic boundary conditions were applied with a two femtosecond (fs) time-step. The short range electrostatic and Van der Waals cut-offs were set to 12 Å and the Van der Waals switching function was set to 10 Å. All simulations were done in Gromacs[19] version 2016.4.

3.4.3 CCS Calculations

Using the trajectories of the MD simulations, CCS calculations were performed on a per frame basis using EHSSRot[10], an accelerated optimized software implementing the Projection Approximation (PA)[20] and Exact Hard Spheres Scattering (EHSS)[21] models, developed by Shvartsberg *et al.*[10] which they state is approximately 20 times faster than MOBCAL. All

calculations were performed in helium to be comparable to experimental data. All data presented is taken from values derived by the EHSS model.[21]

3.5 Results and Discussion

3.5.1 IM-MS Data

Ion Mobility Mass Spectrometry data, of COR15A was acquired on the Waters Synapt G2 IM-MS instrument and is shown in Figure 3.2A and supplementary Figure 3.3. Overall, the experiment provided insight into the conformational landscape of COR15A, reporting a broad Charge State Distribution (CSD) for the system ranging from z = 5 – 12, indicating a wide conformational spread (Supplementary Figure 3.3).[8] Furthermore, the CSD is bi-modal, made up of two sub-distributions one ranging z = 5 – 7, and the other ranging z = 8 – 13. We hypothesize that this arises due to the presence of two distinct structural families. The lower sub-distribution of charge states all exhibited significant amounts of salt adduction which suggests they may be molten globular in nature. This is due to the more compact nature of the structure trapping salt ions within it. The additional information provided by the Collision Cross Section Distribution (CCSD) for charge states ranging z = 6 – 12 in Figure 3.2A and supplementary Figure 3.3, also suggests that the lower charge states, Z = 6 – 7, are disordered globular systems, while the higher charge states, Z = 8 - 12 represent more structured helical systems; higher charge states are seen to have narrower Gaussian-like CCSDs which indicate the presence of just one conformer. The charge states z = 6 – 7, are broad and low in intensity as seen in Figure 3.2A and the supplementary (Figure 3.3) and are likely due to the presence of molten globular structures. Using a 'Toy' model we can theoretically calculate the largest and smallest CCS for a protein which we term the upper and lower CCS boundaries respectively. For COR15A the lower and upper boundaries were 617 and 3377 $Å^2$ respectively,

showing that these charge states are still well within our models[11] predicted range. Additionally, R_g values (Figure 3.2D) lie in similar ranges to SAXS data suggesting that these charge states are made up of globular monomers. The IM data shows abundance for *z* = 6 – 7 is lower experimentally compared to higher charge states. This lower abundance is also seen for more globular MD structures supporting that *z* = 6 – 7 are globular (Figure 3.2B).

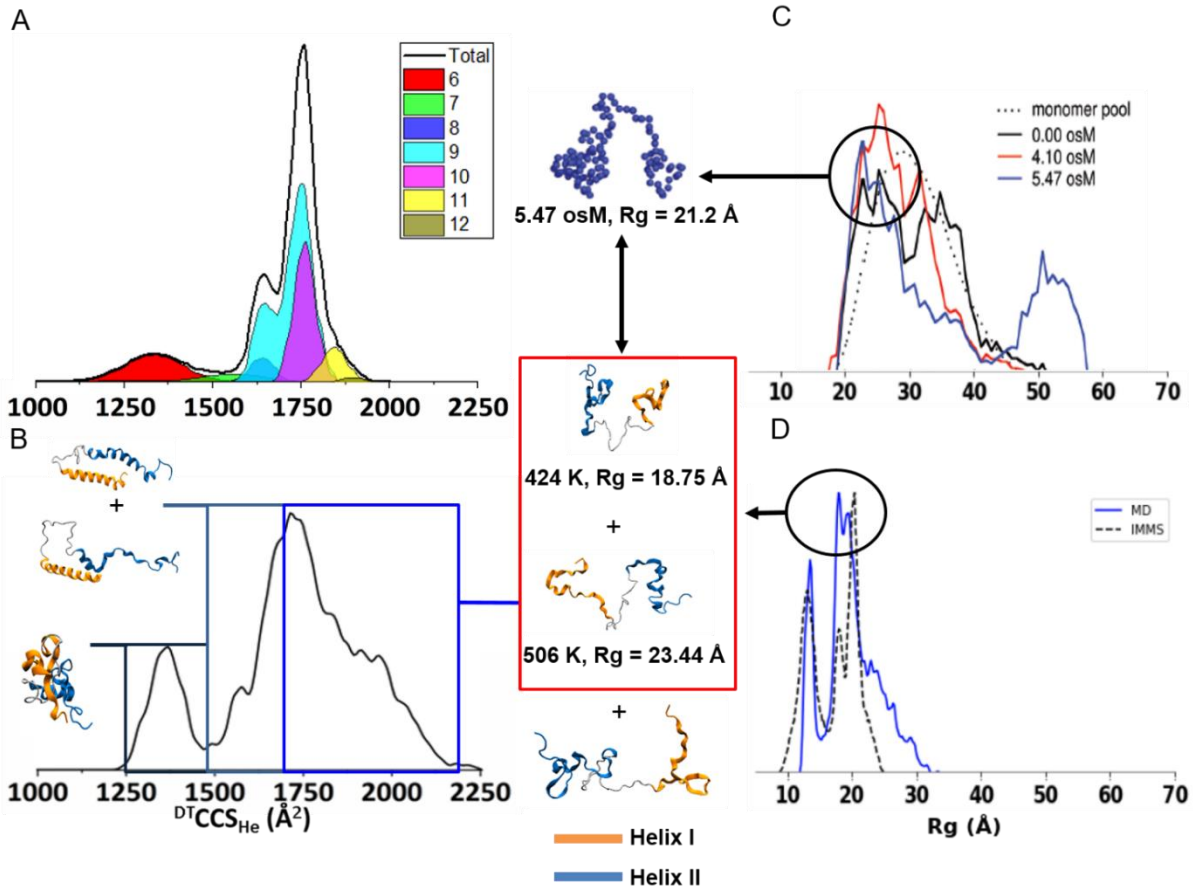

Figure 3.2 | (A) Shown is the total CCS distribution of COR15A obtained from experiment. We see the large conformational space occupied by the protein ranging from ~1100 – 2000 Å^2. The lower charge states, *z* = 6 and 7 are seen to be very broad while all proceeding charge states are narrower and gaussian-like, with *z* = 9 being made up of two unresolved gaussian-like peaks. (B) The *p*CCS distribution from MD simulations obtained using KDE is shown here. The boxed region ranging 1250 – 1500 Å^2 is represented by collapsed molten globular conformers, the boxed region ranging 1500 – 1700 Å^2 is made up of dihelix and partially unravelled helical structures, and lastly the largest boxed region ranging 1700 – 2200 Å^2 is made up of unravelled systems, as well

as expanded and highly disordered conformers. (C) SAXS and EOM modelling derived R_G distribution profiles at representative glycerol concentrations from work done by Thalhammer *et al.*[22]. The dotted black line indicates the R_G distribution of the initial EOM modelling pool, each coloured line represents a different glycerol concentration for which a R_G profile from SAXS was determined. Additionally, a crude bead model from SAXS representing structures in the black circled region is shown. This structure is noted to be like those obtained from MD simulations by the black double ended arrow below. (D) IM-MS and MD derived R_G distribution profiles. The blue line is representative of an R_G profile obtained from the Molecular Dynamics simulation ranging from ~12 – 34 Å, while the dashed black line is representative of R_G obtained from IM-MS Data ranging from ~8 – 24 Å.

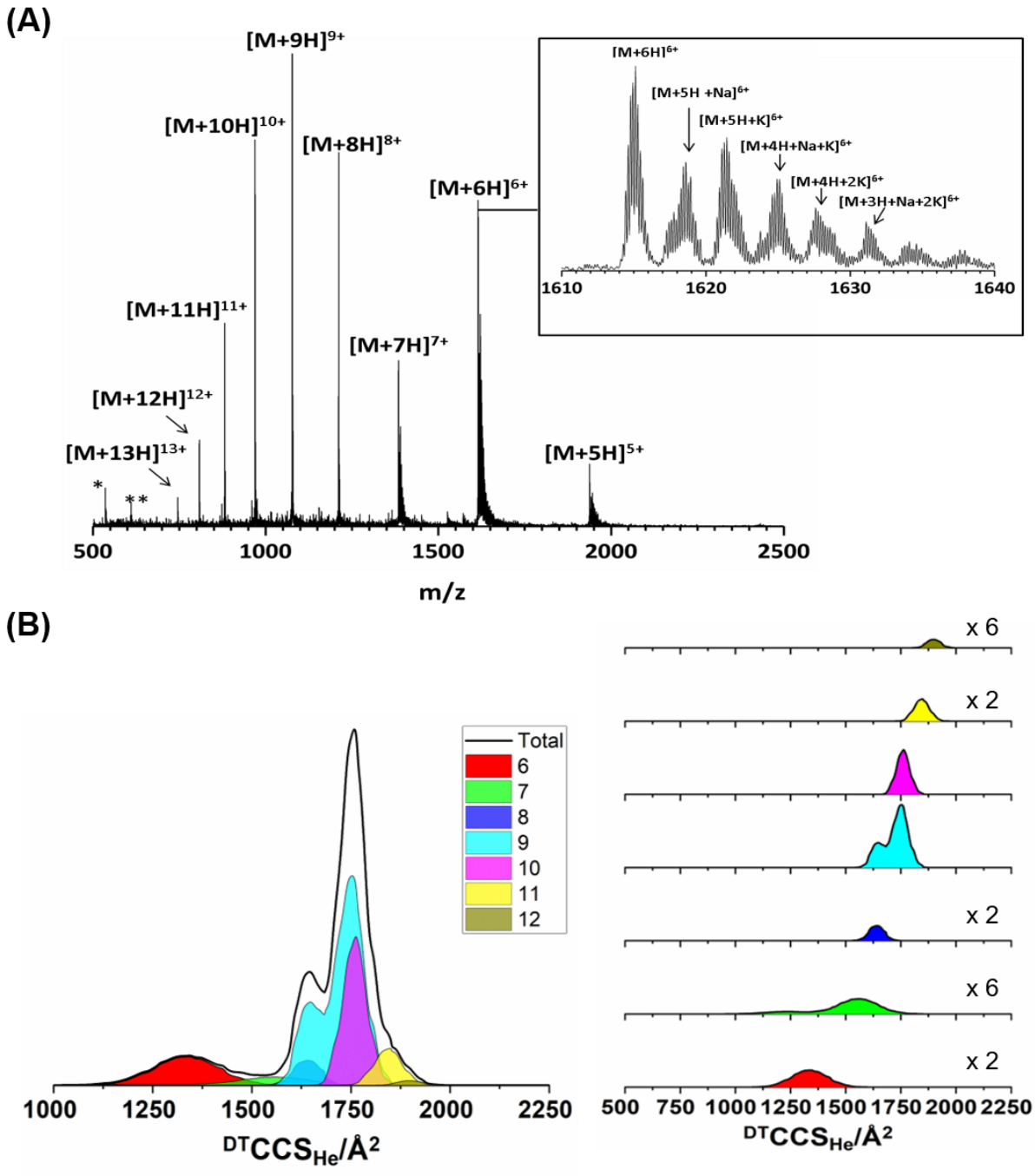

Figure 3.3 | (A) Shown is the Mass Spectrum of COR15a, sprayed from a 200 mM ammonium acetate solution at pH 7.3 using nESI. The molecular weight was determined to be 9688.3 Da ± 3.4 experimentally, and 9683 Da theoretically. Contaminant peaks centred at *m/z* 590 and 620 are denoted with '*' and '**' respectively. Salt adduction was seen for the charge states *z* = 5 – 7, this is highlighted with the insert for the 6+ charge state. (B) Shown is the total CCSD plot of COR15a ranging from ~ 1100 – 2000 Å^2. From this plot can be seen the large conformational family present within the system. The lower charge states *z* = 6 and 7 are broad due to these

states occupying a wide conformational space. For all the proceeding charge states, which are slimmer and Gaussian-like, only one conformation is observed for each with the *z* = 9 state being an exception. The *z* = 9 charge state is made up of two partially resolved peaks indicating the presence of two stable conformers.

3.5.2 MD Simulations

We sought to determine if this CCS distribution could be replicated from structures obtained from Molecular Dynamics (MD) using Temperature as an enhanced sampling technique that would allow us to investigate the relative abundance of differing structural conformers. The starting structure for the simulation was taken from a structure provided by the Thalhammer group, who used the comparative modelling method.[5] This structure was used in previous hydration and conformational analysis work done by Thalhammer *et al.*[5] therefore we believed it to be a suitable starting model for the simulations done here. A large variety of conformational space was explored by these simulations and can be seen below in Figure 3.4.

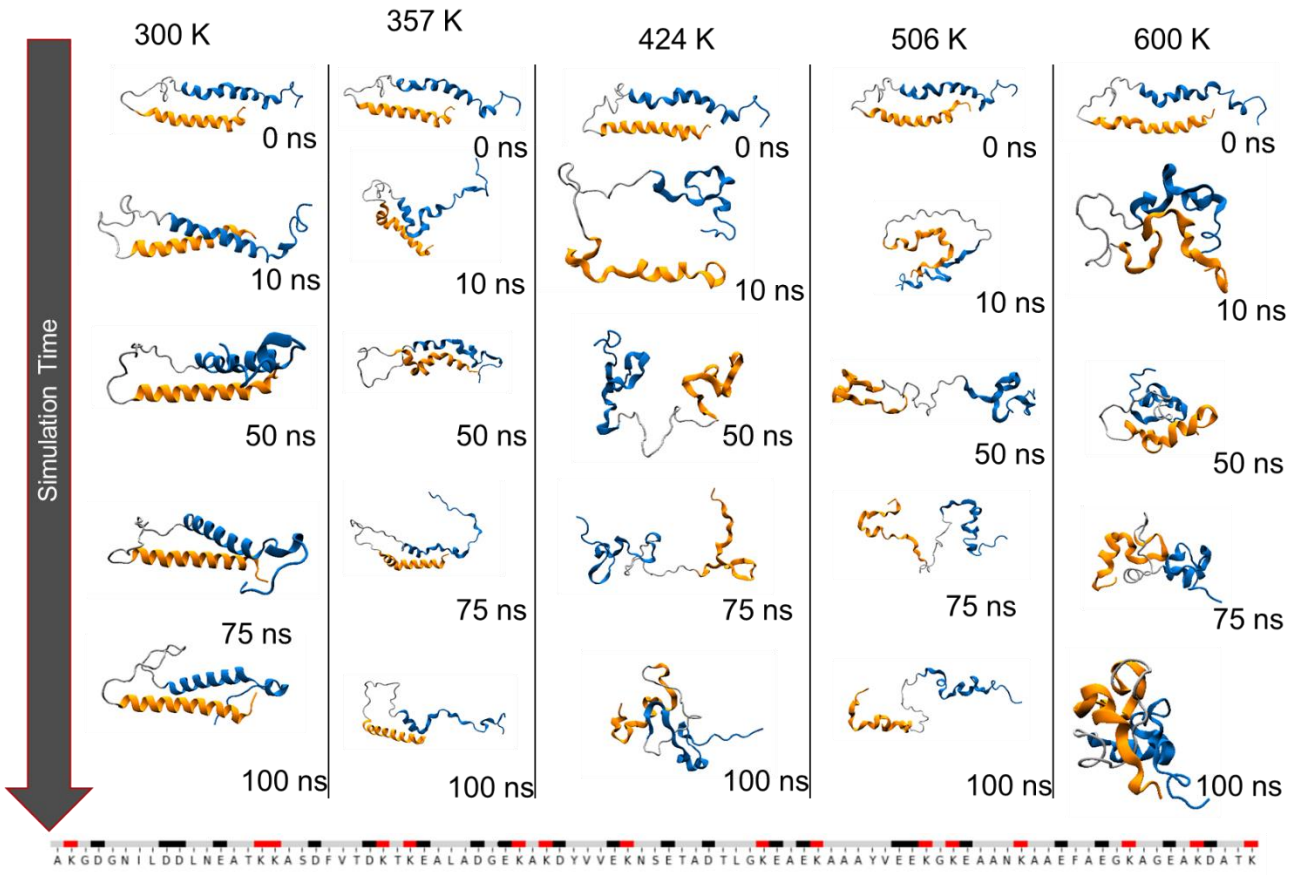

Figure 3.4 | Shown are five trajectories for MD simulations run for 100 ns at temperatures between 300 and 600K. The simulations started with COR15A in the α-helical structure, the colours orange and blue were used to differentiate Helix I and Helix II respectively. Along the bottom of the graph is the amino acid sequence with acidic residues in red, basic residues in black and neutral residues in grey.

For the 300 K simulation little change is seen upon solvation here the structure remains predominantly helical. After 10ns of the simulation at 357 K perturbations begin to appear in the helical network followed by some unravelling of Helix II (blue). At 424 K the helices separate and unravel, and by 100 ns have collapsed into a disordered globule. Similar behaviour is observed at 506 K however the two helices do not collapse to form a globule but remain in an open disordered state. Lastly at 600 K the system undergoes a rapid collapse into

a disordered globule by 50 ns and remains in this state. Shown in Figure 3.5 below are the RMSDs for the different temperature simulations.

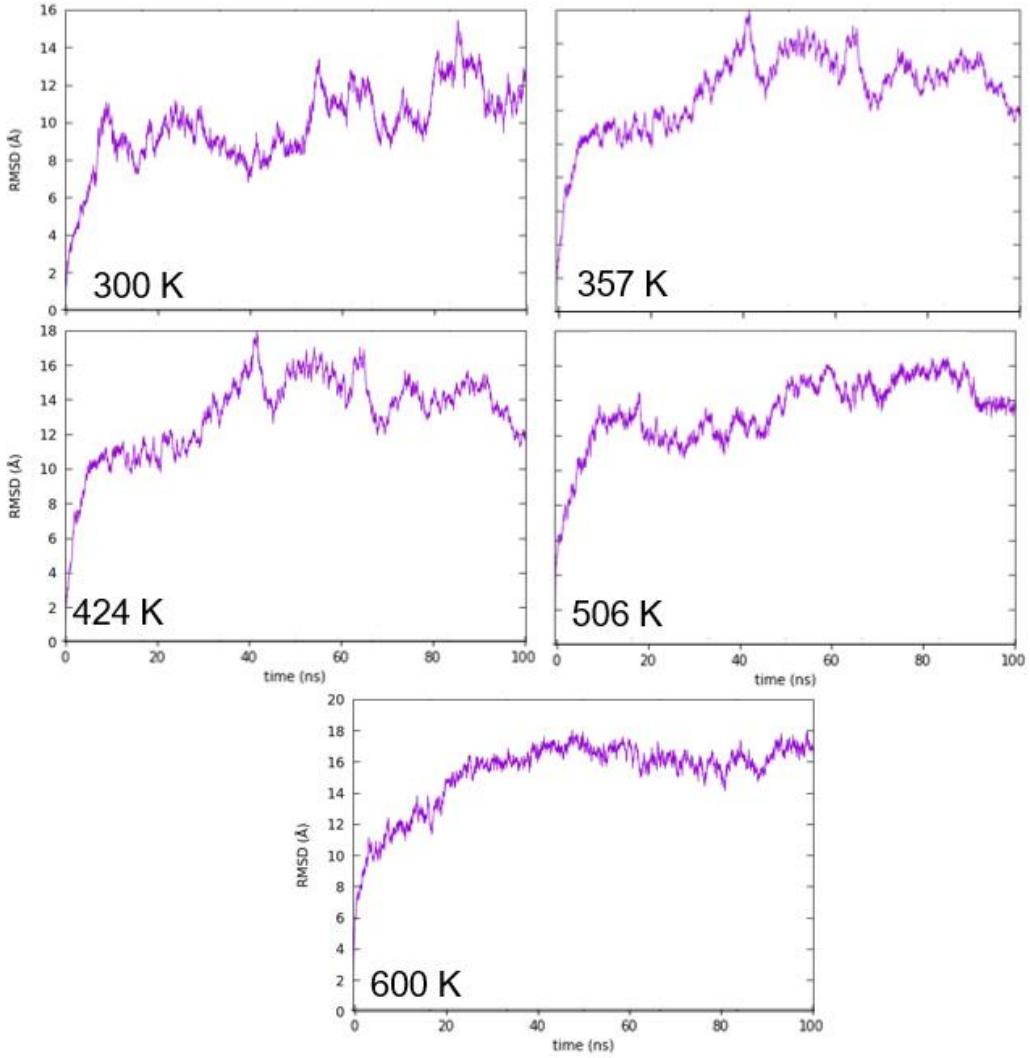

Figure 3.5 | Shown are the RMSDs for the simulations ran at five different Temperatures.

3.5.3 Comparing IM-MS/MD and SAXS/EOM Data

Using computationally calculated CCSs as our variable we applied Kernel Density Estimation (KDE) as a non-parametric way to determine the probability density function of the conformational space explored in our simulations. KDE gives us information about the underlying distribution of the CCS data. It is similar to histogram binning but much better for our CCS dataset as it works well with continuous data which histogram binning does not. In

principle it is quite simple, we start by adding kernels centred around each data point in our data set. Following this the kernels are added together to form the probability density for the data. The result is a series of *pseudo*-CCS distributions (*p*CCS) for each simulation as well as an overarching *p*CCS that included the CCS data from all temperatures. The experimental CCSD (Figure 3.2A) as well as the overall *p*CCSD (o*p*CCSD, Figure 3.2B) are shown in Figure 3.2 for comparison. Figure 3.2A shows the large conformational space occupied by the protein ranging from ~ 1100 – 2000 Å^2. The lower charge states *z* = 6 and 7 are very broad due to these states being made up of a wide structural ensemble. For all the proceeding charge states, which are thinner and Gaussian-like, only one conformation is observed with the *z* = 9 state being an exception. The *z* = 9 charge state is made up of two partially resolved peaks indicating the presence of two stable conformers. Figure 3.2B allows us to compare both the computationally generated CCSD with experiment. The *p*CCS distribution also possesses a bimodal character very similar to experiment with the minima seen at ~1500 Å^2 and can be seen clearly in Figure 3.2B. The lower peak ranging from ~1250 – 1450 Å^2 contains globular structures, while the other feature is the large wedge-like region of unresolved peaks ranging from ~1500 – 2200 Å^2. This region is made up of many different conformations, with different secondary structural content including some structured helices, partially unravelled systems, expanded and highly disordered conformers and large globules (Figure 3.4). The lower subsection of this region, at ~1500 – 1750 Å^2, is dominated by di-helix and partially unravelled Helix II structures, which agrees with what we see in the IM-MS data. Interestingly though a globular conformer observed at 424 K rests within this region at roughly ~1600 Å^2 as well. The larger CCS values < ~2200 Å^2 correspond to more extended disordered structures observed at 424 and 506 K. It is important to note that the larger of these structures are not observed

in experiment with their corresponding CCS being greater than 2000 Å^2. Therefore, it is difficult to say whether they are thermodynamically stable structures. However, those values within range of the experimental CCS distribution indicates that COR15A has the capacity to adopt highly disordered open structures as well as collapsed globules. Furthermore, it may be due to conditions used in experiment that these expanded structures were not seen. Conditions here were used to try and preserve the solution phase structures, i.e they are softer. One may find CIU experiments promote formation of these more expanded structures.

The starting point for the simulations was the di-helix structure of COR15A corresponding the fully folded conformation. Five solvation trajectories at differing temperatures were obtained and can be seen in Figure 3.4. Across the differing temperatures a vast amount of conformational landscape is explored. The room temperature simulation at 300 K remained predominately α-helical, the disulphide bridges along the helices were also generally conserved (Supplementary Figure 3A). When the system was at 357 K significant conformational changes were observed in Helix II which unravelled starting at 10 ns. Moving up in temperature, the 424 K simulation saw a wide separation and unravelling of the two helices starting at 10 ns opening into a disordered structure, by the 100 ns mark however this structure collapsed into a smaller disordered globule. A similar scenario was seen at 506 K which saw separation and unravelling of the helices at 10 ns followed by disordered expansion however no collapse was seen at the 100 ns mark. At 600 K the helical system was seen to collapse rapidly into a disordered globule by 10 ns and remained in this state for the rest of the simulation.

R_G distribution comparisons can be made with Figure 3.2, Analysis of COR15A using EOM modelling of SAXS data was previously carried out by Thalhammer *et al*. [22] (Figure 3.2C). It

provided insight into the conformational variation and more specifically highlights the importance of the di-helix structure and its connecting unstructured linker. Crude bead models from SAXS analysis showed both a compact (R_G = 21.2 Å) and an elongated (R_G = 31.9 Å) form. R_G distributions from IM-MS and MD data were generated for comparison to those presented by Thalhammer *et al.* (Figure 3.2D). We see with SAXS/EOM data that there are some similar structures to those reported in the IM-MS/MD dataset and that, when considering abundances, distributions with bimodal maxima are observed for all methods and the R_G values are slightly similar. There is however a shift seen for SAXS/EOM data (range 15 – 58 Å) compared to IM-MS/MD (range 8 – 35 Å), this is likely due to the differences in methods as EOM derived R_G are reported to be slightly higher than those determined by Guinier analysis.[22] Furthermore certain forcefields have been known to underestimate the R_G of unstructured proteins.[23] What is most interesting is that for all three methods compact structures with similar R_G are seen indicating the importance of these conformers as well as the comparability and accuracy of the methods in investigating the structural landscape of COR15A, this has been highlighted in Figure 3.2.

Secondary structure analysis was carried out using the DSSP algorithm, which stands for Define Secondary Structure of Proteins and is the standard method for assigning secondary structure to the amino acids of a protein. It evaluates protein atomic coordinates to identify the secondary structure per amino acid using an electrostatic potential to identify intramolecular backbone hydrogen bonds. Once the hydrogen bonds are identified, they are used to assign each amino acid to a secondary structure descriptor. The different structural descriptors are denoted on the right-hand side of Figure 3.6. This was done to see how structural helicity changed over the course of simulation time, the results are shown in Figure

3.6. Helicity was seen to be conserved in the 300 K trajectory, in line with what can be seen in Figure 3.4 for the corresponding temperature. For the 357 K trajectory the helicity that dominated residues 50 – 89 at 300 K was significantly reduced. These residues form Helix II mentioned earlier which is seen to partially unravel in Figure 3.4. The systems at 424 and 507 K both saw significant reduction in helicity almost straight away as well as many irregular elements being seen to dominate the secondary structure which became highly disordered as can be seen in Figure 3.4. The 600 K trajectory has an interesting narrative as it shows unravelling of the initial α-helix architecture but starting at approximately 10 ns many new small helices were seen and conserved for the rest of the simulation. This indicates convergence of the system to the collapsed globule structure seen in Figure 3.4, with various small, folded regions perturbed throughout it. Furthermore, for the temperatures 424 – 600 K helicity can be seen to switch on and off in differing regions as the overall structure changes.

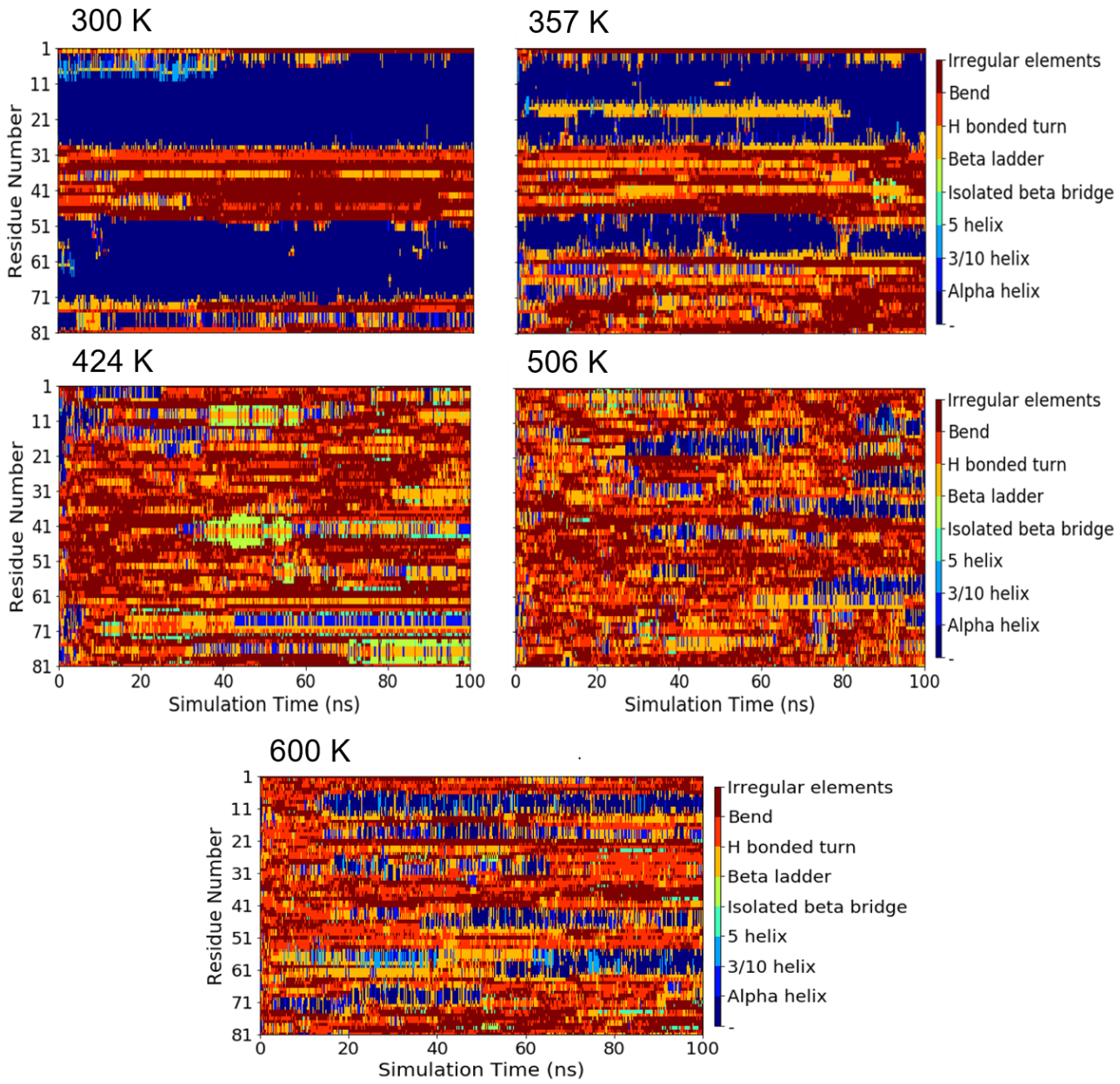

Figure 3.6 | Shown are secondary structure plots of COR15A for the five simulations. General α-helical structure is conserved throughout the 300 K simulation. At 357 K the helicity that dominated residues ~ 60 – 89, corresponding to Helix II on the protein, is reduced. At 424 K and 506 K a dramatic loss in helicity can be seen with an increasing spread of irregular elements arising from the disordered structure present in these trajectories. The 600 K trajectory sees loss of the initial helical regions, followed by the formation of many small helical regions dotted throughout the structure at approximately 10 ns.

The data provided on helicity in Figure 3.6 is further supported by helical probability per residue plots (Supplementary data Figure S3.5). The probability plots tell us which residues were involved in helix formation and to what extent depending on their probability value. Starting at 300 K two distinct and strong helical regions can be seen which correspond to Helix I and Helix II. When the temperature is raised to 357 K the residues 61 – 81 involved with Helix II are seen to lose their helical character in line with partial unravelling reported in Figure 3.4. At 424 K the helical character of the protein is all but vanished, followed by a small increase in various residues at 506 K. The increase is then accompanied by a further increase among numerous small regions at 600 K.

Studies have previously shown that simulations of IDPs tend to collapse and become more compact with increasing temperature, this statement is in line with what we see occurring at 600 K.[25,26] Several theories have been proposed in regards to the nature of these temperature induced collapses, such as strengthened hydrophobic interactions, weakened solvation and reduced thermal fluctuations. [26] A study carried out by Zerze *et al.*[25] found, that as temperature increased, IDPs which were dominated by hydrophobic groups as well as those with a large fraction of negative hydrophilic groups exhibited rapid collapse. In regards to COR15A we observe a loss of protein-water hydrogen bonds and gain of intra-protein hydrogen bonds, along with formation of a hydrophobic core at 600 K which is absent in the helical form of COR15A[5]. We propose that the combination of absence of a hydrophobic core, increasing intra-protein hydrogen bonding (Figure S3.3) and strengthening hydrophobic interactions (Figure S3.6) promotes the di-helix collapse in water. This results in the burying of hydrophobic residues to form a hydrophobic core while the hydrophilic charged residues remain on the surface of the protein (Figure S3.7).

Overall, a semi quantitative correlation between the data is seen demonstrating that when using a combination of MD and CCS calculations we can generate a o*p*CCSD possessing the same features that are present in experimentally determined CCSD. We can thus use the experimental CCSD to apply constraints to our MD derived o*p*CCSD highlighting which structures are likely to be thermodynamically stable while scrutinising structures that lie outside the experimental CCSD boundaries. This shows how MD can be used to highlight the important structural features in an IDPs conformational landscape just like in experiment additionally it also goes beyond experiment giving atomistic models for likely conformations which can then be analysed by comparison with experimental data.

3.6 Conclusion

The work done here provides valuable computational and experimental insight into the structural feature of COR15A. It demonstrates how MD simulations with enhanced sampling can be used to explore the large conformational space of an intrinsically disordered protein providing atomistic detail. Furthermore, it shows how solvated states can be kinetically trapped using IM-MS, which leads to the main point. It highlights the comparability of the (gas-phase) experimental and (solution phase) computational conformers, the CCSDs of which both contain the same structural features of a bimodal distribution of two differing conformational families. The good comparability between experiment and computation means that further refinement of computational analysis can be done. By using experimental data as a constraining factor for the exploration of the conformational landscape one can comment on the thermodynamic and kinetic validity of discreet conformers. This will lead to more rigorous computational analysis as well as provide atomistic level detailed conformers which are absent from experiment.

3.7 Acknowledgements

We acknowledge funding from The University of Manchester and the A*STAR program for the PhD. studentship of DS, as well as thanking Waters Corp. and Agilent for support and help with our instrumentation. We also thank A. Thalhammer for contributions of samples and support in this project.

3.8 References

1 V. N. Uversky and A. K. Dunker, *Biochim. Biophys. Acta - Proteins Proteomics*, 2010, 1804, 1231–1264.

2 R. van der Lee, M. Buljan, B. Lang, R. J. Weatheritt, G. W. Daughdrill, A. K. Dunker, M. Fuxreiter, J. Gough, J. Gsponer, D. T. Jones, P. M. Kim, R. W. Kriwacki, C. J. Oldfield, R. V. Pappu, P. Tompa, V. N. Uversky, P. E. Wright and M. M. Babu, *Chem. Rev.*, 2014, **114**, 6589–6631.

3 P. Tompa, *Nat. Chem. Biol.*, 2012, **8**, 597–600.

4 J. Ananiev, G. Tchernev, J. W. Patterson, M. Gulubova and G. Ganchev, *Acta Medica Bulg.*, 2011, **38**, 72–82.

5 C. Navarro-Retamal, A. Bremer, J. Alzate-Morales, J. Caballero, D. K. Hincha, W. González and A. Thalhammer, *Phys. Chem. Chem. Phys.*, 2016, **18**, 25806–25816.

6 P. Sormanni, D. Piovesan, G. T. Heller, M. Bonomi, P. Kukic, C. Camilloni, M. Fuxreiter, Z. Dosztanyi, R. V. Pappu, M. M. Babu, S. Longhi, P. Tompa, A. K. Dunker, V. N. Uversky, S. C. E. Tosatto and M. Vendruscolo, *Nat. Chem. Biol.*, 2017, 13, 339–342.

7 P. Tompa, *Curr. Opin. Struct. Biol.*, 2011, 21, 419–425.

8 D. Stuchfield and P. Barran, *Curr. Opin. Chem. Biol.*, 2018, 42, 177–185.

9 F. Lanucara, S. W. Holman, C. J. Gray and C. E. Eyers, *Nat. Chem.*, 2014, **6**, 281–294.

10 A. A. Shvartsburg, S. V. Mashkevich, E. S. Baker and R. D. Smith, *J. Phys. Chem. A*, 2007, **111**, 2002–2010.

11 R. Beveridge, S. Covill, K. J. Pacholarz, J. M. D. Kalapothakis, C. E. Macphee and P. E. Barran, *Anal. Chem.*, 2014, 86, 10979–10991.

12 R. K. Das and R. V Pappu, *Proc. Natl. Acad. Sci.*, 2013, **110**, 13392–13397.

13 A. Thalhammer, G. Bryant, R. Sulpice and D. K. Hincha, *Plant Physiol.*, 2014, **166**, 190–201.

14 A. Thalhammer and D. K. Hincha, *Plant Signal. Behav.*, 2014, **9**, e977722–e977723.

15 D. Stuchfield, A. P. France, L. G. Migas, A. Thalhammer, A. Bremer, B. Bellina and P. E. Barran, in *Methods in Enzymology*, Academic Press, 2018, vol. 611, pp. 459–502.

16 M. F. Bush, Z. Hall, K. Giles, J. Hoyes, C. V. Robinson and B. T. Ruotolo, *Anal. Chem.*, 2010, **82**, 9557–9565.

17 J. Huang, S. Rauscher, G. Nawrocki, T. Ran, M. Feig, B. L. De Groot, H. Grubmüller and A. D. MacKerell, *Nat. Methods*, 2016, **14**, 71–73.

18 P. Mark and L. Nilsson, *J. Phys. Chem. A*, 2001, **105**, 9954–9960.

19 B. Hess, C. Kutzner, D. Van Der Spoel and E. Lindahl, *J. Chem. Theory Comput.*, 2008, **4**, 435–447.

20 G. Von Helden, M.-T. Hsu, N. Gotts and M. T. Bowers, *Carbon Cluster Cations with up to 84 Atoms: Structures, Formation Mechanism, and Reactivity*, 1993, vol. 97.

21 A. A. Shvartsburg and M. F. Jarrold, *Chem. Phys. Lett.*, 1996, **261**, 86–91.

22 K. Shou, A. Bremer, T. Rindfleisch, P. Knox-Brown, M. Hirai, A. Rekas, C. J. Garvey, D. K. Hincha, A. M. Stadler and A. Thalhammer, *Phys. Chem. Chem. Phys.*, 2019, **21**, 18727–18740.

23 K. Kasahara, H. Terazawa, T. Takahashi and J. Higo, *Comput. Struct. Biotechnol. J.*, 2019, 17, 712–720.

24 A. Thalhammer, G. Bryant, R. Sulpice and D. K. Hincha, *Plant Physiol.*, 2014, **166**, 190–201.

25 G. H. Zerze, R. B. Best and J. Mittal, *J. Phys. Chem. B*, 2015, **119**, 14622–14630.

26 A. Hicks and H. X. Zhou, *J. Chem. Phys.*, 2018, **149**, 72313.

Chapter 4: Exploring the Conformational Stability of the Cold-Regulating Plant Protein COR15A with Variable Temperature IM-MS

4.1 Declaration

This chapter is reproduced from a completed manuscript I wrote that is set to be published:

D. Stuchfield, A. Thalhammer, and P. E. Barran, “Exploring the Conformational Stability of the Cold-Regulating Plant Protein COR15A with Variable Temperature IM-MS”.

As the first author of the article, I drafted and edited the manuscript as well as designed and rendered all the figures.

Exploring the Conformational Stability of the Cold-Regulating Plant Protein COR15A with Variable Temperature IM-MS

D. Stuchfield, A. Thalhammer and P. E. Barran*

Michael Barber Centre for Collaborative Mass Spectrometry, Manchester Institute for Biotechnology, School of Chemistry, University of Manchester, 131 Princess Street, Manchester M1 7DN, United Kingdom.

To Whom Correspondence should be addressed: perdita.barran@manchester.ac.uk

4.2 Abstract

The effect of temperature on the stability of protein fold has been extensively studied at high temperatures, yet very few methods exist which can examine this below the freezing point of water. Variable Temperature Ion Mobility Mass Spectrometry (VT-IM-MS) can address this issue by studying isolated molecules in the gas-phase at sub-ambient temperatures. The Cold Regulating plant protein COR15A has been studied with IM-MS at ambient temperatures and shown to exist in two major conformations, here we examine the effect of lowering the drift gas temperature (over the range 193 – 298 K) on these conformers. Overall, we see that the CCS increases (between ~7 – 12 % going from 298 K to 193 K) as the temperature decreases, although by more than predicted by current theoretical models. At 270 K anomalous behaviour is seen, with a larger increase in the CCS of all conformers. This is attributed to significant alterations in the strength of the stabilising noncovalent interactions that occur upon cooling to this temperature resulting in substantial restructuring of the protein which is akin to denaturation. Taken together these findings indicate cold denaturation can occur to proteins in the absence of solvent and that current theoretical models are inaccurate with respect to the effect of temperature on CCS.

4.3 Introduction

The study of intrinsically disordered proteins (IDPs) has been gaining momentum in recent years. It is predicted that somewhere between 30 - 70 % of all eukaryotic cell proteins can be classified as having some degree of disorder.[1,2–4] The motivation for studying these proteins initially stemmed from their role in Parkinson's and Alzheimer's disease[3] as well as their role in tumour formation.[4,5] Due to their dynamic behaviour IDPs illude contemporary forms of analysis used to structurally classify proteins. NMR and SAXS are both extensively used to study IDPs but are limited by their experimental timescales.[5–8]

Ion Mobility Mass Spectrometry (IM-MS) has now been extensively applied to the study of IDPs.[6] In previous work (Chapter 3) we showed its use to examine the Cold-Regulated (COR15A) plant protein found in *Arabidopsis thaliana* which is involved in the stabilisation of chloroplast membranes in the leaf when the plant undergoes cellular freezing. This study was carried out at ambient temperature and demonstrated how IM-MS can explore conformational landscapes of IDPs in detail as well as be used to constrain Molecular Dynamic (MD) simulations. We reported good comparability between both methods, observing that the Collision Cross Section Distribution (CCSD) generated by both techniques contained the same structural features of a bimodal distribution with two differing conformational families. The good comparability between IM-MS and MD meant that we were able to explore the large conformational space of an IDP at atomistic detail.

COR15A belongs to the Late Embryogenesis Abundant (LEA) protein family that were first characterised three decades ago.[7] The protein itself is involved in the process of freeze tolerance found in plants native to colder climates termed cold acclimation and studies

showed constitutive expression of the COR15A protein in non-acclimated *A. Thaliana* was associated with increased freezing tolerance.[9] In a hydrophilic environment (such as the cell membrane) COR15A is reported to be an intrinsically disordered protein that upon dehydration, commonly caused by ice formation in the cell due to freezing, adopts a predominantly α-helical form.[9] A protein function mechanism was proposed by Thalhammer *et al*.[10], where they assert cellular freezing causes ice formation in apoplast transitioning the COR15A protein from a hydrophilic to hydrophobic environment. As a result, the protein folds into an α-helical structure and associates with the inner envelope membrane of the chloroplast stabilising it against the formation of hexagonal II (HII) phase lipid domains. COR15A is reported to undergo structural changes from an intrinsically disordered protein to an α-helical structure due to cold induced dehydration therefore the next logical step to is to study COR15A under cold temperatures.

Protein stability is an important issue for understanding a wide variety of biological and biochemical problems. Thermal stability is one of the most well-known parameters that effects protein stability[11], Specifically, when heating of a protein solution causes the structure to unfold this is referred to as “thermal denaturation” and it is well known in the wider scientific community.[11] Yet protein stability is governed by a parabolic curve and as such crosses the zero point of free energy at two points. One being above room temperature where “thermal denaturation” occurs and the other below room temperature where “cold denaturation” occurs. Exemplary work done by Privalov *et al.*[12] established cold denaturation as a concept, however when compared to thermal denaturation it is not as extensively studied.[11] The effect of lowered temperature on the structure and function of proteins is of fundamental interest and also relevant to many fields of research including biocatalysts that

are effective at low temperatures[13–17], transportation and storage of pharmaceuticals[18]. Of critical importance, amidst the climate crisis, is the need to understand the effects of fluctuations in temperature for both plants and animals[19,20] and is the motivation for these investigations on COR15A.

In solution the fold is maintained partly due to interactions between the solvent and polar hydrophilic groups on the exterior of the protein.[21] Polar solvents also interact with hydrophobic groups on the protein driving them into its interior. As the temperature of the solvent is lowered solvent-solvent interactions begin to dominate; this can lead to ice formation in aqueous solutions and has an overall effect of weakening the influence of solvent on the proteins structure. As a result, hydrophobic and hydrophilic groups are no longer confined to the interior and exterior of the protein respectively, thought this is not necessarily true of an IDP. The free energy difference between the folded and unfolded forms is relatively small and can be easily reversed by these effects leading to protein denaturation.[12,22]

The effect of elevated temperatures to induce protein unfolding has been extensively studied[11,23], although there have been fewer studies that examine this for IDPs. It is a challenge to experimentally investigate the destabilisation of protein fold and function with reduced temperature, even though its importance in the bulk phase is well known. Prior experimental studies on the effects of cold temperature on proteins have involved the use of denaturants[24] to destabilise the protein or high pressures to lower the freezing point of water.[25] This inevitably leads to deviations from native structures to occur before cooling takes place blurring the line between native and unfolded structures.[26]

Variable Temperature (VT) IM-MS can be used to address this issue. IM-MS at ambient temperatures has proven to be a valuable technique for analysing the conformational landscapes of proteins[6,27]. The IM-MS instrument in this study is capable of measurements ranging from 120 – 600 K (Figure 4.1).[28] The collision cross section (CCS) measurement taken by IM-MS is dependent on temperature[29], meaning the VT instrument can measure the effect temperature has on the protein molecule and CCS. In this study we look at the COR15A protein at low temperatures ranging from 193 – 298 K using an in-house built VT-IM-MS instrument (Figure 4.1).[30]

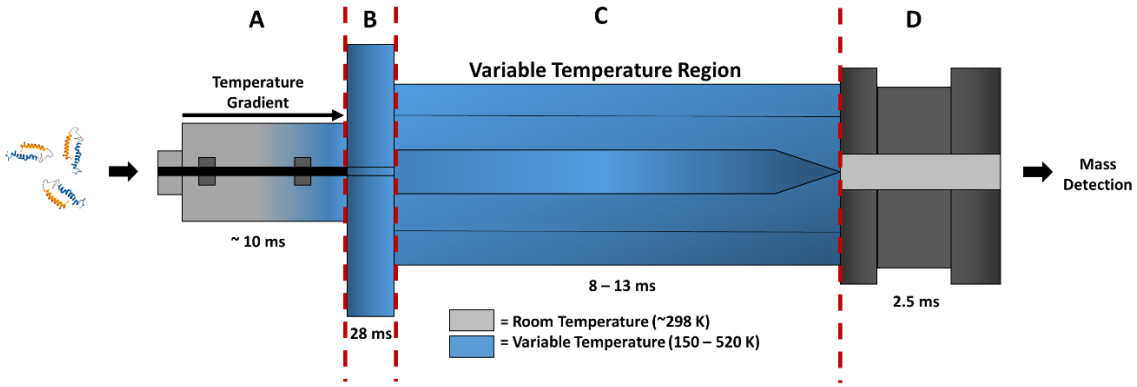

Figure 4.1 | Shown is a simplified schematic of the VT-IM-MS instrument's main separation stages. (A) Here the protein sample enters the instrument *via* *n*ESI, desolvation takes approximately ~10 ms. Just after this a temperature gradient cools the activated ions. (B) Ions are transferred to this region where they are trapped in the ion buncher in the presence of a cold buffer gas. This lasts for 28 ms and allows the ions to form 'packets' as well as equilibrate to the buffer gas temperature. (C) The ion packets are released into the drift cell where they begin to separate based on their ion mobility, this process lasts for 8 – 13 ms. (D) The ions are transferred to the ToF mass analyser detection region at room temperature where an arrival time distribution (ATD) is acquired that can be later used to derive a CCS distribution.

The instrument used in this study has already been extensively applied to numerous other studies looking at thermal stability of mAbs[31], intermediates of ubiquitin and lysozyme[32], as

well as insights into conformations of cytochrome *c*, p53, and MDM2[33] and VT analysis of concanavalin A.[30]

4.4 Experimental

The protocol used to carry out these experiments has been reported previously.[27] In summary, all experiments were carried out on an in house built VT IM-MS instrument as described in Figure 4.1.[28] Samples were analysed by *n*ESI using borosilicate tips (World Precision Instruments) with orifice diameter 3-6 µM pulled on a Sutter Instrument p1000. The COR15A protein was synthesised by the Thalhammer group and supplied in a lyophilised form.[7] The sample was diluted in 200 mM Ammonium Acetate solution (Merck) the pH adjusted to 7.3 (using ammonium hydroxide) to a final concentration of 20 µM. Platinum wire (Goodfellow) was inserted into the tip to apply voltages ranging from 1.0 – 1.4 kV to initiate the spray, cone voltage 6 V, and source temperature was set to 50 °C. Following desolvation, the ions are guided towards the ion buncher by two ion tunnels. The ions are accumulated in the ion buncher for 28 ms at high pressure, ~2 Torr, at the same temperature as the drift gas allowing the ions time to equilibrate. Following this they are released into the drift cell which is 50.5 cm in length. The drift cell is within an external stainless-steel chamber that covers an interior insulating glass tube which houses the electrode stack and buffer gas. Copper tubing is situated between the two layers and is arranged in a longitudinal direction. The copper tubing is connected to an external N_2 gas-supply which passes through liquid nitrogen and then through the cooling lines around the cell. The flow rate of the coolant surrounding the drift cell can be changed by tuning the N_2 flow rate allowing the temperature of the drift gas in the cell to be maintained within ± 1 K. ATD measurements were taken over a range of six voltages ranging from 250 – 300 V in intervals of 10 V and were then used to generate CCSD.

4.5 Results & Discussion

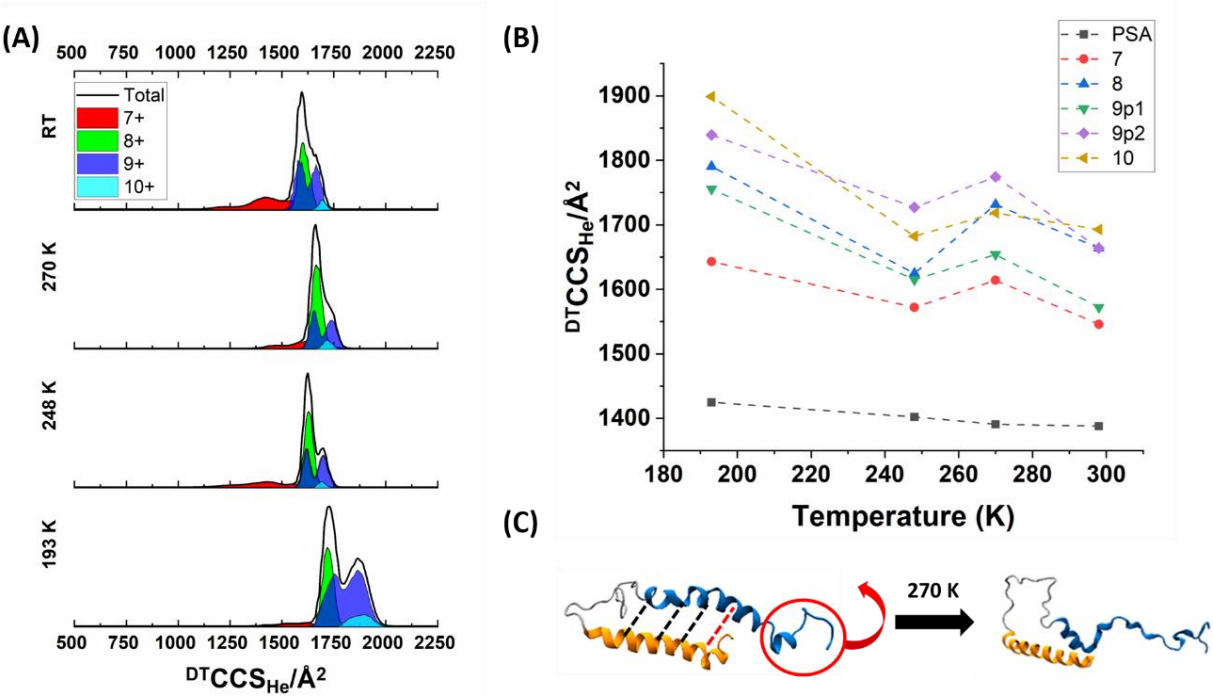

Figure 4.2 | (A) Shown are the total CCSD of COR15A at 298 K, 270 K, 248 K and 193 K. (B) Apex CCS values for each charge state as well as values determined using Projection Superposition Approximation (PSA) are shown here. (C) Proposed 'mechanism' for di-helix opening at 270 K, structures are taken from Molecular Dynamic simulations discussed in Chapter 3.

All VT-IM-MS data was acquired at the following temperatures: 298 K, 270 K, 248 K and 193 K. Figure 4.2 shows the total CCSD as well as apex CCS vs Temperature for each charge state of COR15A. At each temperature four charge states were detected ranging z = 7 – 10, the mass spectra of which are given in Figure 4.3 below.

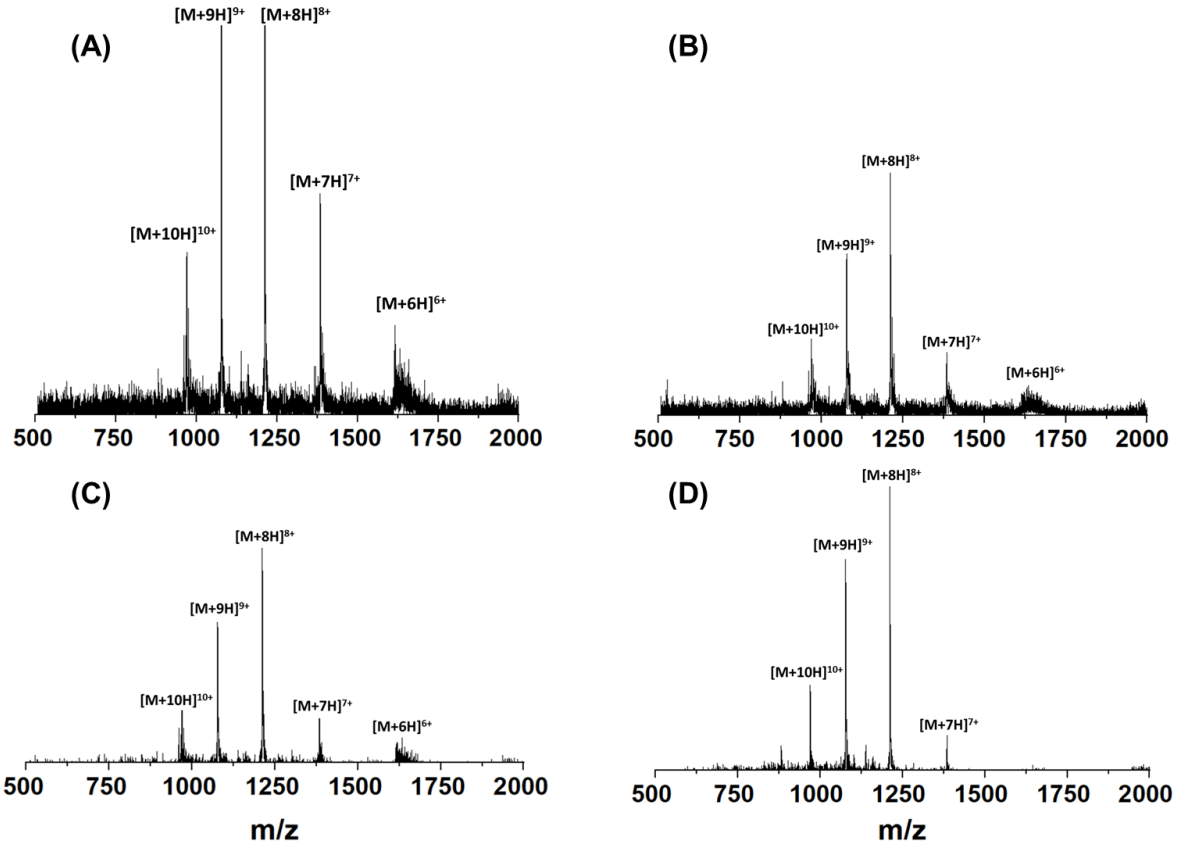

Figure 4.3 | Shown are the mass spectra of COR15A at 298K, 270 K, 248 K, 193 K sprayed on the home-built VT-IM-MS instrument. Charge states ranging from z = 6 – 10 are seen for all temperatures except 193 K which only observed z = 7 – 10. CSDs for this dataset are narrower than those obtained in the previous chapters (2 & 3). These measurements were taken years apart, on different instruments, with different samples, subject to daily temperature and pressure changes. Furthermore, the VT-IM-MS instrument has a harsher ionisation procedure meaning the boundary charge state structures seen in the previous chapters are no longer being preserved. (A) The mass spectrum of COR15A, sprayed from a 200 mM ammonium acetate solution at pH 7.27 using *n*ESI at 298 K on the home-built VT-IM-MS instrument. The molecular weight was determined to be 9688.3 ± 3.4 experimentally and 9683 Da theoretically. Salt adduction was seen for the charge state *z* = 6. (B) The mass spectrum of COR15A, sprayed from a 200 mM ammonium acetate solution at pH 7.27 using *n*ESI at 270 K on the home-built VT-IM-MS instrument. The molecular weight was determined to be 9688.3 ± 3.4 experimentally and 9683 Da theoretically. Salt adduction was seen for the charge state *z* = 6. (C) The mass spectrum of COR15A, sprayed from a 200 mM ammonium acetate solution at pH 7.27 using *n*ESI at 248 K on the home-built VT-IM-MS instrument. The molecular weight was determined to be 9688.3 ± 3.4 experimentally and 9683 Da

theoretically. Salt adduction was seen for the charge state *z* = 6. (D) The mass spectrum of COR15A, sprayed from a 200 mM ammonium acetate solution at pH 7.27 using *n*ESI at 193 K on the home-built VT-IM-MS instrument. The molecular weight was determined to be 9688.3 ± 3.4 experimentally and 9683 Da theoretically.

The ions with *z* = +7 present with a very broad, multimodal ATD covering a range of ~ 500 Å^2, this characteristic was also observed in previous work done at room temperature on a different instrument (Chapter 3) and was attributed to the charge state being made up of a wide structural ensemble. These structures are believed to be disordered and globular in nature due to high salt adduction and CCS values corresponding to collapsed globules generated in MD simulations. The 8+ charge state is the highest in intensity with a CCS of 1664 Å^2 at room temperature. The 9+ charge state is made up of two conformations that both appear Gaussian-like and partially resolved. The apex of both rest at 1572 and 1664 Å^2 respectively at room temperature. Lastly is the 10+ charge state which is much smaller in intensity than both its predecessors, its room temperature apex rests at 1693 Å^2. All of these charge states follow the same general trend of having the CCS increase in magnitude as the temperature is lowered. This behaviour was predicted and follows in-line with the theory that due to the increased influence of the long-range interactions at low temperatures, the CCS of a rigid ion will increase as the temperature is lowered but its structure will remain the same. Bowers reported this with C_{60} experiments they carried out showed the same trend.[34] Temperature dependent computational calculations for the CCS were also carried out using the Projection Superposition Approximation (PSA) model webserver.[35–37] PSA was trained to reflect the fact that the CCS will increase as T decreases[36], these calculations showed an increase in CCS as temperature was lowered. Looking in more detail however when the PSA values are compared to experimental, we can see that the increase in CCS at 270 K is much greater than expected and is not predicted by theory. The PSA model predicts an increase of

0.23 % going from 298 K to 270 K, while experimental data reports an increase of 4.40 %, 4.07 %, 5.20 %, 6.63 % and 1.52 % for the 7+, 8+, 9+ (both peaks) and 10+ charge states respectively. This break in trend can also be seen clearly in the apex plots in Figure 4.2B, this could suggest that the training set used for PSA is not capturing this effect.

These observations can be explained by considering the cumulative effects of the cold induced H-bond lengthening and residual dynamics that would still be present within the system. We propose that at this temperature the lengthening of the H-bonds within the protein is complementary with residual dynamics present in the C-terminus tail. As a result, the tail rotates outwards, and pops open the di-helix making it a more elongated structure and by extension have a larger CCS. A visual representation of this is shown in Figure 4.2C.

The theory for the transport of ions in a gas it is stated that the resolution will increase as temperature is lowered ($R_{max} \propto \frac{1}{\sqrt{T}}$), this leads to narrower CCSD at cryogenic temperatures. This is not observed for charge states 9+ and 10+ at 193 K, instead what is seen is a broadening of the CCSD. A possible hypothesis for this was the partial resolving of newer conformers. This theory was tested by running the COR15A sample on a cyclic IMS instrument which can be used to separate unresolved conformers.[38] The analysis did not resolve any new conformers for these charge states when run for more than one cycle as can be seen in the supplementary Figure S4.1, suggesting this hypothesis to not be the case. The likely cause of this broadening is due to a 'Drag' effect in the VT-IMS instrument reported by Ujma *et al.*[28], where in the region between the ion funnel and cone aperture exists a voltage offset between 'Funnel DC low' and 'Cone Aperture' (supplementary Figure S4.2). If this offset is too small the ions may

experience trapping/drag with this effect being more prominent at low temperatures or when using N_2 as a drift gas.

4.6 Conclusion

In conclusion, we have shown how VT IM-MS can be used to provide molecular level detail into the cold denaturation of the COR15A plant protein. The overall trend seen for each charge state's CCS as temperature is lowered follows the theory that long range interactions dominate at lower temperatures causing CCS of a rigid ion to increase but its structure to remain the same. We see larger than expected increases in CCS at 270 K, this can be explained by the cumulative effects of protein-protein H-bond lengthening and residual dynamics in the C-terminus. We proposed a mechanism where the di-helix is popped open by rotation of the C-terminus tail and lengthening of protein-protein H-bonds causing partial unravelling of the C-terminal helix leading to a more elongated structure with a larger CCS. Furthermore, it is evidence of Cold Denaturation occurring in COR15A in the absence of solvent at 270 K.

At temperatures below 270 K, the conformer occupancies are similar to room temperature, yet the distribution appears narrower suggesting the protein ions are 'snap frozen' and no structural rearrangement can occur. Broadening observed for the +9 and +10 charge states can be explained by the 'Drag' effect seen in the VT-IMS instrument due to the offset between the ion funnel and cone aperture being set too low. When the offset is too low ions can experience trapping/drag in this region with the effect being more prominent at lower temperatures, as a result distributions become broader.

The evidence provided for cold denaturation in the absence of solvent implies the importance of factors inherent to the proteins fold. Though the importance of solvent-solute interactions on the fold has been previously reported it is clear that interactions within the protein also

play a significant role in cold denaturation and this is not well understood or represented in theoretical models like PSA. Therefore, further work needs to be done on understanding cold denaturation and VT IM-MS will play a significant role in carrying out this research.

4.7 Acknowledgements

We acknowledge funding from The University of Manchester and the A*STAR program for the PhD. studentship of DS, as well as thanking Waters Corp. and Agilent for support and help with our instrumentation. We also thank A. Thalhammer for contributions of samples and support in this project.

4.8 References

1 V. N. Uversky and A. K. Dunker, *Biochim. Biophys. Acta - Proteins Proteomics*, 2010, 1804, 1231–1264.

2 R. van der Lee, M. Buljan, B. Lang, R. J. Weatheritt, G. W. Daughdrill, A. K. Dunker, M. Fuxreiter, J. Gough, J. Gsponer, D. T. Jones, P. M. Kim, R. W. Kriwacki, C. J. Oldfield, R. V. Pappu, P. Tompa, V. N. Uversky, P. E. Wright and M. M. Babu, *Chem. Rev.*, 2014, **114**, 6589–6631.

3 P. Tompa, *Nat. Chem. Biol.*, 2012, **8**, 597–600.

4 J. Ananiev, G. Tchernev, J. W. Patterson, M. Gulubova and G. Ganchev, *Acta Medica Bulg.*, 2011, **38**, 72–82.

5 P. Sormanni, D. Piovesan, G. T. Heller, M. Bonomi, P. Kukic, C. Camilloni, M. Fuxreiter, Z. Dosztanyi, R. V. Pappu, M. M. Babu, S. Longhi, P. Tompa, A. K. Dunker, V. N. Uversky, S. C. E. Tosatto and M. Vendruscolo, *Nat. Chem. Biol.*, 2017, 13, 339–342.

6 D. Stuchfield and P. Barran, *Curr. Opin. Chem. Biol.*, 2018, 42, 177–185.

7 C. Navarro-Retamal, A. Bremer, J. Alzate-Morales, J. Caballero, D. K. Hincha, W. González and A. Thalhammer, *Phys. Chem. Chem. Phys.*, 2016, **18**, 25806–25816.

8 P. Tompa, *Curr. Opin. Struct. Biol.*, 2011, 21, 419–425.

9 A. Thalhammer, G. Bryant, R. Sulpice and D. K. Hincha, *Plant Physiol.*, 2014, **166**, 190–201.

10 A. Thalhammer and D. K. Hincha, *Plant Signal. Behav.*, 2014, **9**, e977722–e977723.

11 D. Sanfelice and P. A. Temussi, *Biophys. Chem.*, 2016, **208**, 4–8.

12 P. L. Privalov, *Crit. Rev. Biochem. Mol. Biol.*, 1990, **25**, 281–306.

13 D. Georlette, V. Blaise, T. Collins, S. D'Amico, E. Gratia, A. Hoyoux, J. C. Marx, G. Sonan, G. Feller and C. Gerday, *FEMS Microbiol. Rev.*, 2004, 28, 25–42.

14 K. S. Siddiqui and R. Cavicchioli, *Annu. Rev. Biochem.*, 2006, 75, 403–433.

15 C. Gerday, M. Aittaleb, M. Bentahir, J. P. Chessa, P. Claverie, T. Collins, S. D'Amico, J. Dumont, G. Garsoux, D. Georlette, A. Hoyoux, T. Lonhienne, M. A. Meuwis and G. Feller, *Trends Biotechnol.*, 2000, 18, 103–107.

16 J. Furhan, *J. Genet. Eng. Biotechnol.*, 2020, 18.

17 G. Feller, *Scientifica (Cairo).*, 2013, **2013**, 1–28.

18 C. Ammann, *AAPS PharmSciTech*, 2011, 12, 1264–1275.

19 K. Cascella, D. Jollivet, C. Papot, N. Léger, E. Corre, J. Ravaux, M. S. Clark and J. Y. Toullec, *PLoS One*, , DOI:10.1371/journal.pone.0121642.

20 R. A. Zerebecki and C. J. B. Sorte, *PLoS One*, 2011, **6**, 14806.

21 D. Sanfelice, E. Morandi, A. Pastore, N. Niccolai and P. A. Temussi, *ChemPhysChem*,

2015, **16**, 3599–3602.

22 C. Camilloni, D. Bonetti, A. Morrone, R. Giri, C. M. Dobson, M. Brunori, S. Gianni and M. Vendruscolo, *Sci. Rep.*, , DOI:10.1038/srep28285.

23 R. Yan, P. DeLos Rios, A. Pastore and P. A. Temussi, *Commun. Chem.*, , DOI:10.1038/s42004-018-0015-1.

24 V. Griko Yu., P. L. Privalov and S. Venyaminov Yu., *Biofizika*, 1989, **34**, 940–945.

25 R. Kitahara, A. Okuno, M. Kato, Y. Taniguchi, S. Yokoyama and K. Akasaka, *Magn. Reson. Chem.*, 2006, **44**, S108–S113.

26 D. Liu, T. Wyttenbach, C. J. Carpenter and M. T. Bowers, *J. Am. Chem. Soc.*, 2004, **126**, 3261–3270.

27 D. Stuchfield, A. P. France, L. G. Migas, A. Thalhammer, A. Bremer, B. Bellina and P. E. Barran, in *Methods in Enzymology*, Academic Press, 2018, vol. 611, pp. 459–502.

28 J. Ujma, K. Giles, M. Morris and P. E. Barran, *Anal. Chem.*, 2016, **88**, 9469–9478.

29 E. W. M. E. A. Mason, *Transport Properties of Ions in Gases*, John WIley & Sons Inc., New York, 1988.

30 J. Ujma, K. Giles, M. Morris and P. E. Barran, *Anal. Chem.*, 2016, **88**, 9469–9478.

31 K. J. Pacholarz, S. J. Peters, R. A. Garlish, A. J. Henry, R. J. Taylor, D. P. Humphreys and P. E. Barran, *ChemBioChem*, 2016, **17**, 46–51.

32 J. Ujma, J. Jhingree, R. Upton, F. Benoit, B. Bellina and P. Barran, , DOI:10.26434/CHEMRXIV.11888958.V2.

33 E. R. Dickinson, E. Jurneczko, K. J. Pacholarz, D. J. Clarke, M. Reeves, K. L. Ball, T. Hupp, D. Campopiano, P. V. Nikolova and P. E. Barran, *Anal. Chem.*, 2015, **87**, 3231–3238.

34 T. Wyttenbach, G. Von Helden, J. J. Batka, D. Carlat and M. T. Bowers, *J. Am. Soc. Mass Spectrom.*, 1997, **8**, 275–282.

35 C. Bleiholder, N. F. Dupuis, T. Wyttenbach and M. T. Bowers, *Nat. Chem.*, 2011, **3**, 172–177.

36 C. Bleiholder, S. Contreras, T. D. Do and M. T. Bowers, *Int. J. Mass Spectrom.*, 2013, **345**–**347**, 89–96.

37 S. E. Anderson, C. Bleiholder, E. R. Brocker, P. J. Stang and M. T. Bowers, *Int. J. Mass Spectrom.*, 2012, **330**–**332**, 78–84.

38 K. Giles, J. Ujma, J. Wildgoose, S. Pringle, K. Richardson, D. Langridge and M. Green, *Anal. Chem.*, 2019, **91**, 8564–8573.

Chapter 5: Conclusions and Future Work

In this thesis, ion mobility mass spectrometry (IM-MS) as well as molecular dynamics (MD) has been used to investigate the structural landscapes of intrinsically disordered proteins (IDPs). These investigations build upon developments in native IM-MS and MD over the last years for IDPs, showing how together they can be used to describe these structurally dynamic systems. The motivation for these experiments arose from IDPs lacking any significant structure that could be resolved via crystallography or NMR. In addition to their role in many proteopathic diseases such as Parkinson's and Alzheimer's disease there was a pressing need to develop alternative structural methods to help understand this 'dark proteome'. We used COR15A as an exemplar system for more in-depth investigations due to its ability to adopt an ordered and disordered form depending on its solution environment, and the impact that this has on its function.

In Chapter 1, we looked at much of the (at the time) recent work done using IM-MS to study IDPs. It became apparent that new insights into the intrinsic structure of these proteins can still be found and that these have the potential to improve our fundamental understanding of protein behaviour as well as the role of IDPs in disease and drug discovery. Chapter 2 focused on the methodology we employed for the investigations performed in the subsequent chapters. Using COR15A and myoglobin as model systems we compared the differences between a dynamic system and a highly ordered system covering sample preparation, purification, instrument optimisation and data analysis.

In Chapter 3, IM-MS yielded a bimodal charge state and collision cross section distribution for the COR15A plant protein. It provided information on the protein's conformational landscape, indicating the presence of two distinct conformational families, with the lower charge states (z = 6 and 7) likely corresponding to more collapsed globular forms and the higher charge

states (z= 8 – 12) likely represented more ordered and helical structures. These helical forms were likely to correspond to the desolvated states of the protein, indicating that gas phase data corresponded to natively accessible states. Following this the experimental IM-MS data was compared to MD simulations which were used to generate candidate geometries and identify dominant secondary structural features. Evaluation of the simulations showed significant similarities and agreement with the IM-MS data, providing compelling evidence that the experimental gas phase data corresponded to natively accessible states.

The investigation highlighted the comparability between the gas phase experimental and solution phase computational conformers, both of which contained the same structural features. This means that IM-MS data does represent native conformers and can be used as a constraining factor for the exploration of conformational landscapes by MD allowing discrimination against thermodynamically and kinetically unstable conformers. This results in MD providing atomistic scale detail on conformers detected in experiment that would otherwise be unavailable.

In Chapter 4, VT-IM-MS was used to study COR15A at sub-ambient temperatures. The next step for analysis of the cold regulating protein was to see how this protein behaved in sub-ambient temperatures, the investigation provided data on conformational changes at temperatures ranging from 193 K – 298 K. The overall trend reported an increase in the protein's CCS as predicted by VT-IM theory that for a rigid ion in cold temperature the CCS will increase due to strengthening of the long-range interaction potential, but its apparent structure will remain the same. However, anomalous behaviour was observed at 270 K, where the increase in CCS was much larger than expected. This increase was attributed to changes in the intramolecular interactions of the protein, where protein-protein H-bond lengthening

and residual dynamics in the C-terminus has a cumulative effect on the dihelix structure causing it to be popped open giving it a larger CCS. Temperature dependent theoretical calculations (Proximity Superposition Approximation) also modelled an overall increase in CCS at lower temperatures but did not report the same anomalous behaviour seen at 270 K. The behaviour reported at 270 K is evidence of cold denaturation in the absence of solvent occurring which has not been extensively studied. It highlights the importance intramolecular interactions have on the protein fold and that these are not well represented in theoretical models like PSA, and furthermore that VT-IM-MS will play a significant role in developing our understanding of cold denaturation.

Looking forward, preliminary work on two mutants given to us by the Thalhammer group has been started. One is a poly-glycine mutation, and the other is a four times glycine-alanine substitution, the sequences of both can be seen in Figure 5.1. Mutation work is of interest because it will give us information about the effect changing residues will have on the protein's structure and charge states. Alongside IM-MS analysis of these mutants it would be interesting to run MD simulations, starting with the WT structure one could substitute the necessary residues to replicate the mutations and then follow the simulation protocols discussed in Chapters 2 and 3. Following these experiments VT-IM-MS analysis on the mutants would also yield interesting results showing the effects residue make up will have on thermal stability of the protein.

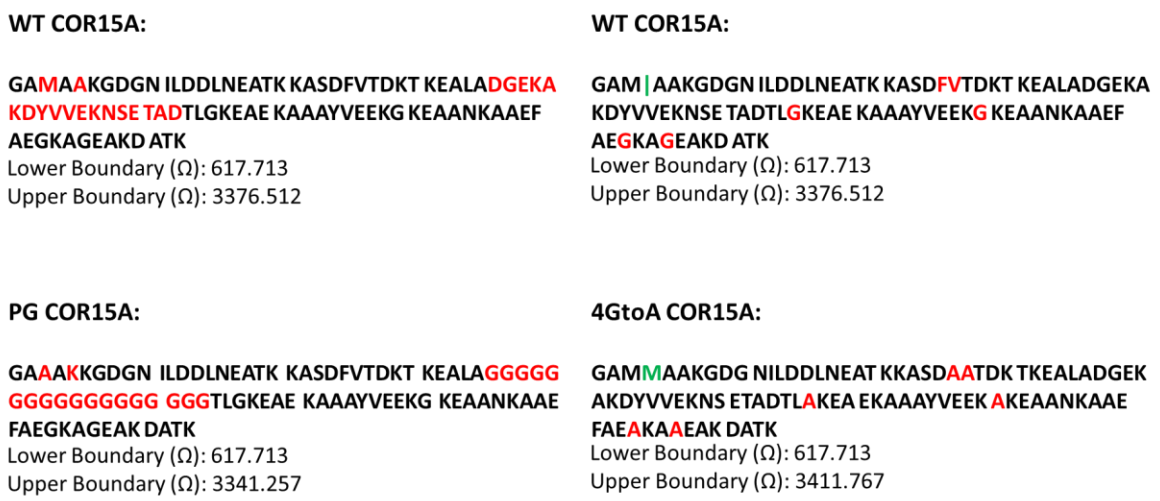

Figure 5.1 | Shown is the amino acid sequence for the poly-glycine (PG) and G to A (4GtoA) substitution mutations of COR15A, the wild type (WT) is also shown for comparison. The PG Substitution has residues 36 – 53 substituted for Glycine (G). The 4GtoA has the Glycine residues at 57, 71, 84 and 87 replaced with an Alanine (A). Upper and lower CCS boundaries (calculated using the toy model discussed in Chapter 2) for the wild type and mutants are also given. The drive for studying these variants is the information they may provide on the effect changing residues will have on the protein's structure and charge states.

Another avenue of interest is in protein charge positioning and regulation, exemplary work done by Pappu *et al.*[1] highlighted the importance of the overall charge content and patterning of charged residues in an IDP. These can be altered by effects such as charge screening where solution ions and conformational fluctuations can alter the strengths of intra- and intermolecular electrostatic interactions between charged residues of an IDP. Charge renormalisation also has an effect, referring to the altering or inversion of charge profiles resulting from accumulation of solution ions and charge regulation which relates to parameters such as fraction of charged residues (FCR), net charge per residues (NCPR), and charge patterning. Current MD simulations have a limitation on how they can explore charge as it is common practice to fix the charges of ionisable groups, meaning these simulations (including those discussed in Chapter 3) do not consider the previously mentioned effects.

In order to address this issue *q*-canonical monte carlo sampling developed by Pappu *et al*.[1] can be used to compute pH-dependent weights of different charge microstates and their conformational ensembles. If done prior to the MD protocol used in Chapter 3 we can generate charge dependent ensembles which can be used as starting structures for MD production runs and *p*CCSD generation with the aim of having a deeper understanding of the role charge has on protein structure.

Lastly, work done by Clemmer *et al*.[2] investigated the use of MD to simulate droplet desolvation similar to the desolvation process occurring in electrospray ionisation. Their work found that for an 11-residue polypeptide conformational preference was seen to significantly change during the desolvation process. They reported highly solvated structures when fully immersed in water droplets and that during droplet shrinkage inter- and intramolecular charge solvation of charge sites had an increasingly strong effect on conformational preference. As such carrying out desolvation simulations on COR15A would certainly prove interesting in investigating which residue sites on the protein have significant impact on the desolvation process.

IDPs are fascinating molecules and each one is unique; many techniques are needed to help us understand their inherent nature. IM-MS paired with MD here has shown its potential as a technique for characterising IDP structure, together the methods complement one another allowing for the generation of atomistic scale candidate structures that are thermodynamically and kinetically stable. As both fields continue to develop more and more studies will be done where both methods are used together demonstrating the excellent work that can be done when theory and experiment are both employed to understand a topic.

References

1 M. J. Fossat and R. V. Pappu, *J. Phys. Chem. B*, 2019, **123**, 6952–6967.

2 D. Kim, N. Wagner, K. Wooding, D. E. Clemmer and D. H. Russell, *J. Am. Chem. Soc.*, 2017, **139**, 2981–2988.

Appendix 1

Table containing the proteins used in Beveridge-Barran plots in Figure 2.13. The numbers correspond to those used in Figures 2.13 a) and b).

Number	Name	molecular weight (Da)	dZ	dCCS (Å^2)
1	Melittin	2846	3	61
2	Human beta-defensin 2	4328	4	113
3	Lymphotactin 1-72	8111	5	560
4	Cytochrome C denatured	12229	4	392
5	Haemoglobin-alpha holo monomer	15743	4	493
6	Ovalbumin reduced conformation 1	44200	4	382
7	Ovalbumin reduced conformation 2	44200	3	547
8	Ovalbumin intact conformation 1	44287	2	19
9	Ovalbumin intact conformation 2	44287	3	99
10	TTR tetramer	55000	3	144
11	BSA conformation 2	66430	4	446
12	Concanavalin A tetramer	102000	4	292
13	SAP Pentamer	128000	5	408
14	Lysozyme	14314	3	148
15	Core UVR8 Monomer	40150	2	48
16	Core UVR8 Dimer	80300	4	29
17	Bovine Pancreatic Trypsin Inhibitor (BPTI)	6531	2	107
18	MtATP-phosphoribosyltransferase (MtATP-PRT) he	189374	2	14
19	Insulin	5730	3	274
20	Ubiquitin denatured	8556	8	970
21	Lymphotactin	10175	6	928
22	N-terminal p53	11162	9	1404
23	alpha-synuclein	14460	16	1577
24	N-terminal MDM2	14790	11	1489
25	Haemoglobin-alpha apo monomer	15126	16	2117
26	Haemoglobin-beta apo monomer	15876	10	1161
27	beta-casein	23944	20	4069
28	p53 DNA binding domain	24615	9	1609
29	p53 DNA binding domain pH 1.5	24616	25	4000
30	COR15a	9683	10	1254
31	Apolipoprotein C-II (ApoC-II)	8959	3	542
32	Ubiquitin	8566	2	21.66667
33	Cytochrome C	12358	2	363.6667
34	Myoglobin	17567	2	107.6667
35	Avidin Tetramer	64933	3	20
36	BSA	66433	3	32
37	Concanavalin A	102393	4	42
38	Alcohol Dehydrogenase	147659	5	111
mAbs				
39	IgG1 A	147106.78	4	1440
40	IgG1 B	145246.13	4	1580
41	IgG4 A	146776.18	4	2140
42	IgG4 B	144915.53	4	2040

Appendix 2

Appendix 2 consists of all the supplementary figures mentioned in chapter 3

(A)

Disulfide Bridge	Temperature (K)				
	300	357	424	506	600
ASP20-LYS16	95.4	23.78	-	-	5.69
GLU28-LYS25	48.15	37.66	-	3.4	-
GLU56-LYS53	47.75	26.17	-	-	-
GLU68-LYS65	43.66	-	-	-	-
ASP10-LYS2	38.26	37.66	-	-	-
GLU64-LYS67	37.76	-	-	2.7	-
ASP24-LYS25	-	-	5.19	3.1	6.19

(B)

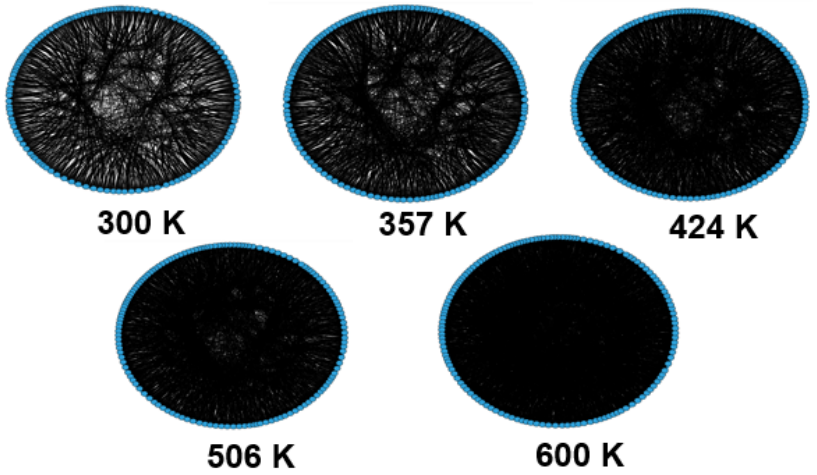

Figure S3.1 | (A) Shown here are the disulphide bridge occupancies along the two helices of COR15A, it can be seen that at 300 and 357 K where helicity is most preserved the presence of disulphide bridges is conserved. At the higher temperatures starting at 424 K, these disulphide bridges do not form, indicative of their role in helix stability. (B) Chord diagrams showing the H-bonding across the protein (residue to residue), as temperature is increased and the amount of intramolecular H-bonding increases.

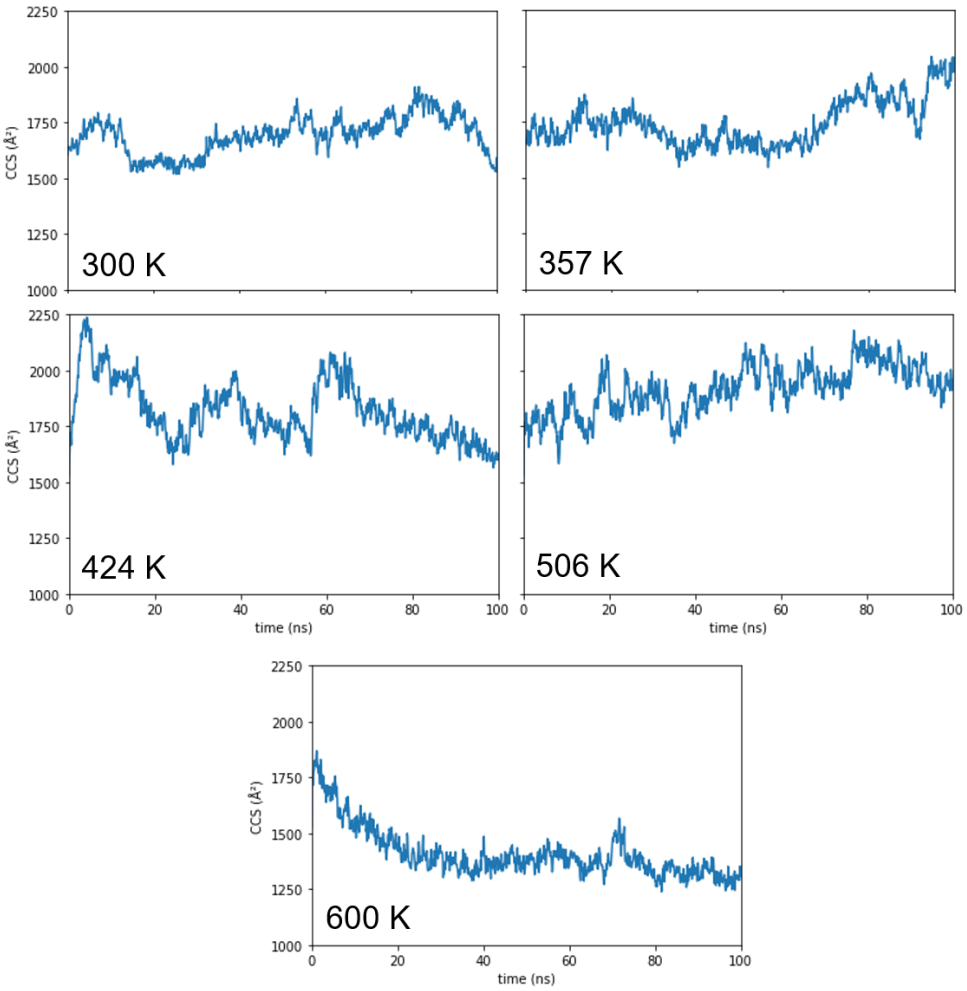

Figure S3.2 | The change in CCS over time for each temperature trajectory is shown here. At 300 K fluctuation in the CCS is seen throughout the simulation however it remains fairly constant and no significant increase or decrease is seem in comparison to the other temperatures. When temperature is raised to 357 K a gradual increase in gradient for the CCS is seen. Moving to 424 K significant change in CCS is seen as sharp increases followed by steady decreases are seen throughout them simulation. The 506 K trajectory sees an overall steady increase in CCS while at 600 K the opposite is true where a steady decrease can be seen that remains constant at approximately 30 ns.

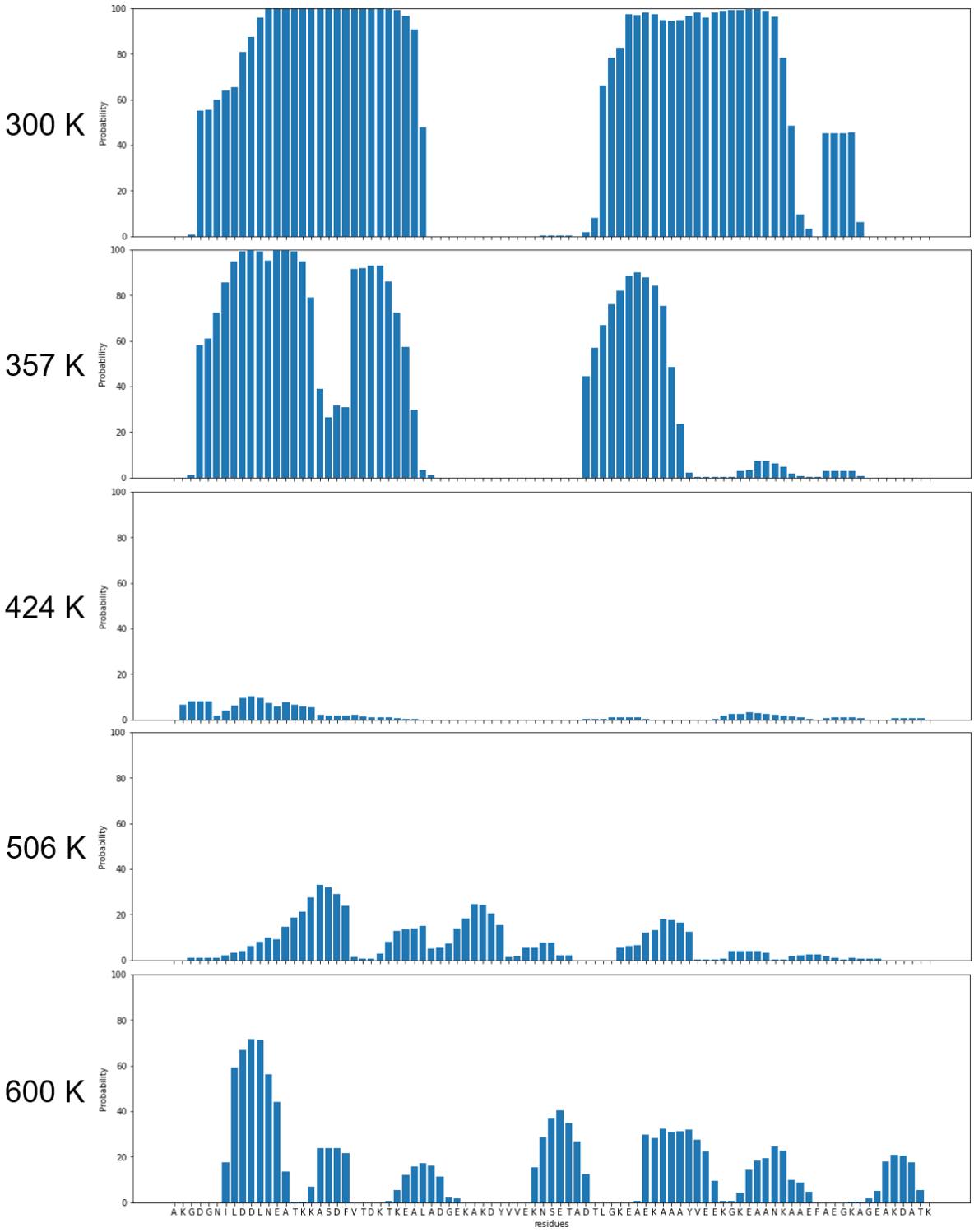

Figure S3.3 | Helical probability per residue plots are shown here for each of the differing temperature trajectories. The 300 K plot shows the residues involved in Helix I and II, following this it can be seen the residues involved in Helix II at 357 K are significantly reducing their helical character. Moving to 424 K the domination of

helicity is all but absent, helicity slowly begins to return at 506 K however it appears in new residue regions that are not seen for the lower temperatures. Again, further increase in small helical regions is seen at 600 K.

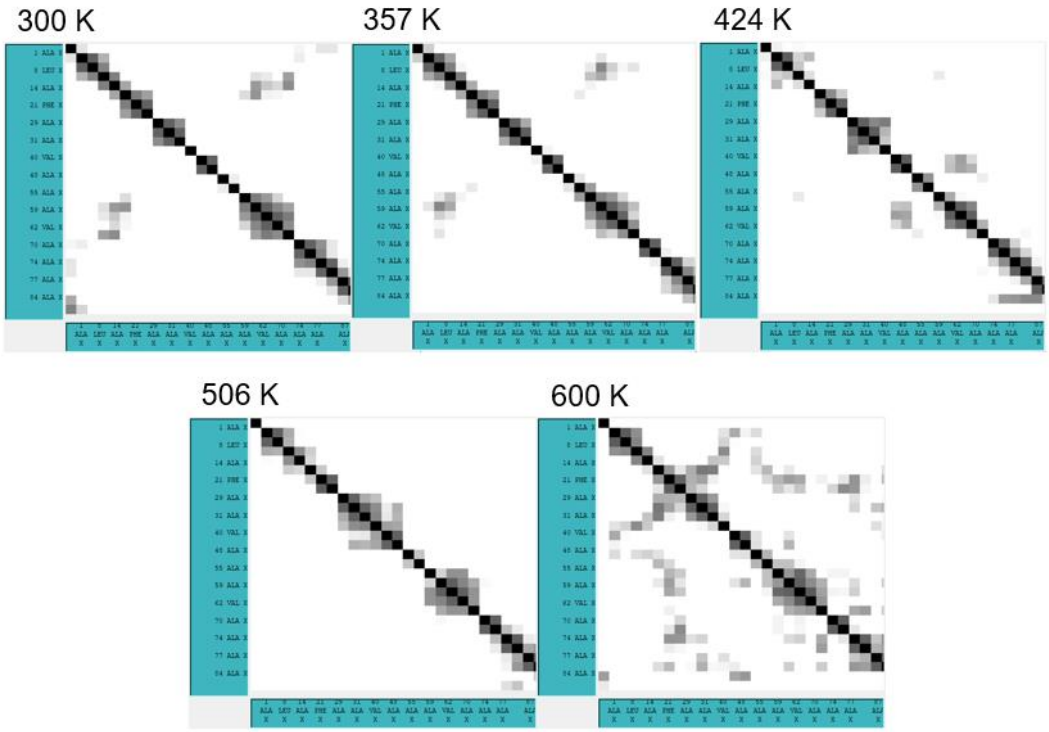

Figure S3.4 | Shown here are contact maps generated using trajectory data in VMD for the hydrophobic residues on COR15A. Very little contact is seen between the hydrophobic residues at temperatures 300 – 506 K, however at 600 K where they form a hydrophobic core a significant increase in contact between hydrophobic residues is seen.

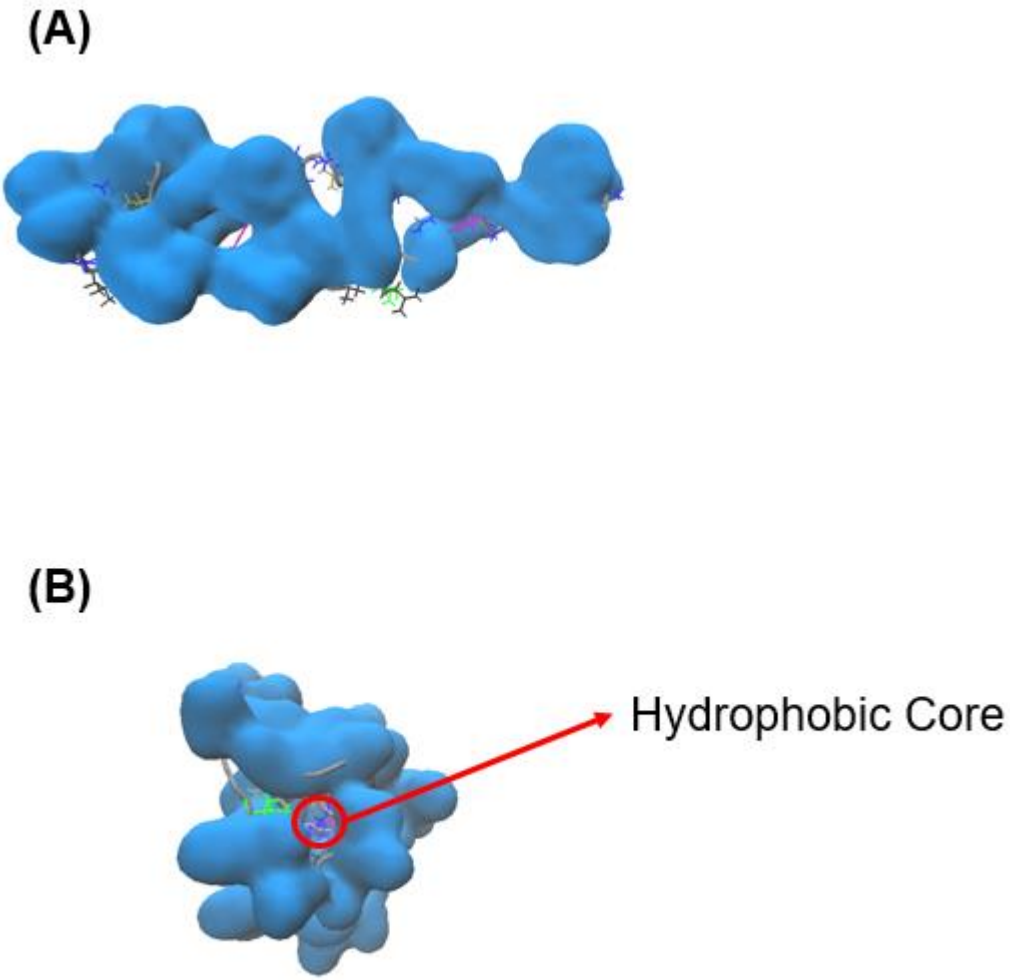

Figure S3.5 | (A) Shows the helical structure of COR15A where the blue shows hydrophilic charged residues surface area. (B) Shows the collapsed globule of COR15A acquired from the 600K simulation with the hydrophobic core circled in red.

References

1 A. Thalhammer and D. K. Hincha, *Plant Signal. Behav.*, 2014, **9**, e977722–e977723.

Appendix 3

Appendix 3 contains all the supplementary material mentioned in Chapter 4.

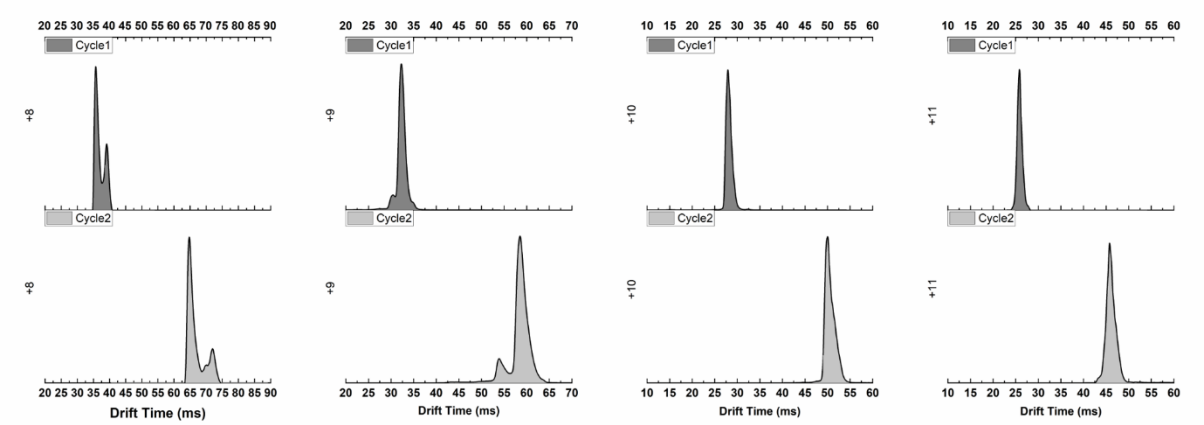

Figure S4.1 | Shown are the ATDs for the first and second cycles of the charge states z = 8 – 11 for the COR15A protein, sprayed from a 200 mM ammonium acetate solution at pH 7.27 using nESI. The data was acquired on a prototype Waters Synapt G2-Si IM-MS instrument with its IM separation region modified to accept a cyclic ion mobility (cIM) device.[1]

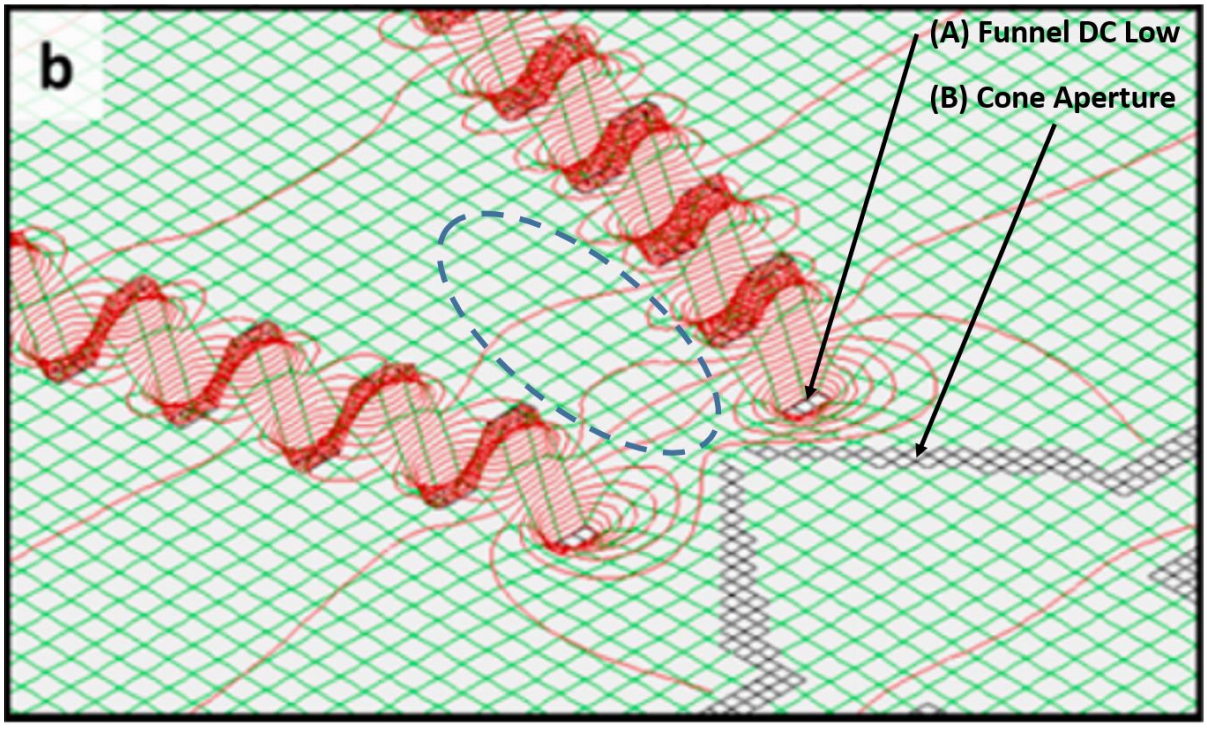

Figure S4.2 | Shown is a SIMION rendering of the ion funnel and cone aperture region of the instrument. Here ions may experience trapping/drag (resulting in broadened CCSDs) due to a voltage offset between 'Funnel DC low' and 'Cone Aperture' being set too low. This may be more prominent at lower temperatures or when using

N_2 drift gas because ion velocity will be lower. This image was taken from the publication by Ujma *et al.*[2] and is used with permission from the authors.

References

1 K. Giles, J. Ujma, J. Wildgoose, S. Pringle, K. Richardson, D. Langridge and M. Green, *Anal. Chem.*, 2019, **91**, 8564–8573.

2 J. Ujma, K. Giles, M. Morris and P. E. Barran, *Anal. Chem.*, 2016, **88**, 9469–9478.