Mass Spectrometry Methods for Characterising the Dynamic Behaviour of Proteins and Protein Complexes

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

2015

Rebecca Beveridge

School of Chemistry
Table of Contents

List of Figures .......................................................................................................................... 8
List of Tables ............................................................................................................................. 13
List of Equations ....................................................................................................................... 13
Abbreviations ........................................................................................................................... 15
Abstract .................................................................................................................................. 18
Declaration ................................................................................................................................. 19
Copyright Statement ............................................................................................................... 20
Acknowledgements ............................................................................................................... 21
Preface ....................................................................................................................................... 22

Chapter 1: Introduction ........................................................................................................... 24

1.1 Protein Structure ............................................................................................................... 25
  1.1.1 The Structure – Disorder Continuum ............................................................................ 25

1.2 Intrinsically Disordered Proteins ...................................................................................... 27
  1.2.1 Biophysical Properties .................................................................................................. 28
  1.2.2 Formation of Complexes ............................................................................................... 29
  1.2.3 Disorder in Disorders; the D² Concept .......................................................................... 31
  1.2.4 Primary Sequence Characteristics of IDPs .................................................................. 32
  1.2.5 Relationship Between Charge Distribution and Disorder ......................................... 32
  1.2.6 Regulating the Unfoldome .......................................................................................... 34

1.3 Biophysical Techniques and Computational Methods for the Detection and Analysis of IDPs ................................................................................................................. 35
  1.3.1 Small-angle X-ray Scattering ....................................................................................... 35
  1.3.2 Nuclear Magnetic Resonance ......................................................................................... 36
  1.3.3 Single Molecule Förster Resonance Energy Transfer ............................................... 37
  1.3.4 Other Spectroscopic Techniques .................................................................................. 37
  1.3.5 Disorder Predictors ....................................................................................................... 38
  1.3.6 Computational Approaches .......................................................................................... 38

1.4 Biological Mass Spectrometry ......................................................................................... 39
  1.4.1 Ionisation Methods ....................................................................................................... 40
    1.4.1.1 Electrospray Ionisation .......................................................................................... 40
    1.4.1.2 Nano-Electrospray Ionisation ................................................................................. 40
    1.4.1.3 Matrix Assisted Laser Desorption Ionisation ......................................................... 43
1.4.2 Mass Analysers ........................................................................................................... 44
  1.4.2.1 The Quadrupole Mass Analyser ................................................................. 44
  1.4.2.2 The Time-of-Flight Mass Analyser .............................................................. 45
1.4.3 Detection ................................................................................................................... 47
  1.4.3.1 Photomultiplier and Electron Multiplier Detectors ................................. 47
  1.4.3.2 The Microchannel Plate Detector ............................................................... 48
1.5 Ion Mobility Mass Spectrometry .............................................................................. 48
  1.5.1 Drift Tube Ion Mobility Mass Spectrometry .................................................... 49
  1.5.2 Travelling Wave Ion Mobility ........................................................................ 50
  1.5.3 Field Asymmetric Waveform Ion Mobility Mass Spectrometry ............... 51
1.6 Hydrogen-Deuterium Exchange Mass Spectrometry ........................................... 51
1.7 Computational Methods to Predict Collision Cross Sections ............................. 52
  1.7.1 The Projection Approximation ....................................................................... 53
  1.7.2 Projection Superposition Approximation ....................................................... 53
  1.7.3 The Exact Hard Sphere Scattering Method .................................................... 53
  1.7.4 The Trajectory Method .................................................................................... 54
1.8 References .................................................................................................................. 54

Chapter 2: Mass Spectrometry Methods for Intrinsically Disordered Proteins.. ............... 60
  2.0 Declaration .............................................................................................................. 62
  2.2 Introduction ............................................................................................................ 63
  2.3 ESI and charge state distribution ....................................................................... 66
    α-synuclein - A case study protein for the use of Mass Spectrometry to examine
    IDPs ......................................................................................................................... 70
  2.4 Limited Proteolysis ............................................................................................... 73
  2.5 Mass spectrometry based proteomics ................................................................. 74
  2.6 Hydrogen-deuterium exchange .......................................................................... 76
  2.7 Ion mobility mass spectrometry ......................................................................... 77
  2.8 Conclusion ............................................................................................................. 82
  2.9 Acknowledgements .............................................................................................. 82
  2.10 References ........................................................................................................... 82

Chapter 3: A Mass Spectrometry Based Framework to Define the Extent of Disorder in
Proteins ........................................................................................................................... 86
  3.0 Declaration ............................................................................................................. 87
3.1 Abstract ............................................................................................................ 89
3.2 Introduction ..................................................................................................... 90
3.2 Materials and Methods .................................................................................. 94
  3.2.1 Materials .................................................................................................. 94
  3.2.2 Mass Spectrometry ................................................................................ 94
  3.2.3 Ion Mobility ............................................................................................ 95
  3.2.4 Fitting Procedures .................................................................................. 95
  3.2.5 Charge-Hydropathy Analysis ................................................................. 96
  3.2.6 Theoretical Modeling of Collision Cross Sections .................................. 96
  3.2.7 Determining the Population of Conformations from Charge State
    Distribution (CSD) and Collision Cross Section (CCS) Data ..................... 97
3.3 Results ............................................................................................................. 98
  3.3.1 Nano-electrospray Ionisation (n-ESI) Ion Mobility–Mass Spectrometry
    (IM-MS) Signatures for a Structured Protein versus a Disordered Protein .... 98
  3.3.2 Can the Median Charge (z) and the Width of the Charge State
    Distribution (Δz) Distinguish between Structured and Disordered Proteins? .. 100
  3.3.3 Can the Median Collision Cross Section (CCS) and the Width of the
    CCS (ΔCCS) Distinguish between Structured and Disordered Proteins? .... 104
  3.3.4 A Comparison of Experimental Findings with Theoretical Predictors of
    Disorder ........................................................................................................ 106
  3.3.5 How Much of the Possible Conformational Space Is Occupied in the
    Present Collision Cross Sections? .................................................................. 110
  3.3.6 Can We Quantify the Relative Population of Conformations from
    Compact to Extended? ................................................................................... 113
  3.3.7 The Framework and How To Determine the Structural Preferences of an
    Unknown Protein .......................................................................................... 115
3.4 Conclusions .................................................................................................... 116
3.5 Acknowledgements ......................................................................................... 118
3.6 References ...................................................................................................... 118

Chapter 4: Relating gas phase to solution conformations: Lessons from
disordered proteins .......................................................................................... 121
4.0 Declaration ..................................................................................................... 122
4.1 Abstract ......................................................................................................... 123
4.2 Introduction .................................................................................................... 124
4.3 Materials and methods .................................................................................. 128
  4.3.1 Sample preparation ................................................................................ 128
  4.3.2 Hydrogen-deuterium exchange .............................................................. 129
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.3.3 Mass spectrometry and ion mobility</td>
<td>129</td>
</tr>
<tr>
<td>4.3.4 Mass spectrometry</td>
<td>129</td>
</tr>
<tr>
<td>4.3.5 Ion mobility-mass spectrometry</td>
<td>130</td>
</tr>
<tr>
<td>4.3.6 Wimley-White plot</td>
<td>130</td>
</tr>
<tr>
<td>4.3.7 Modelling of theoretical CCS extremities</td>
<td>131</td>
</tr>
<tr>
<td><strong>4.4 Results and discussion</strong></td>
<td>132</td>
</tr>
<tr>
<td>4.4.1 The use of HDX-MS to examine the solvent accessibility of the proteins α-Syn and ApoC-II</td>
<td>132</td>
</tr>
<tr>
<td>4.4.2 The analysis of charge state distributions to probe the accessibility of chargeable sites in α-Syn and ApoC-II following nESI–MS</td>
<td>133</td>
</tr>
<tr>
<td>4.4.3 The use of IM-MS to probe the conformational spread of the gas phase forms of α-Syn and ApoC-II</td>
<td>136</td>
</tr>
<tr>
<td>4.4.4 The use of a simple model to predict conformational occupancy</td>
<td>138</td>
</tr>
<tr>
<td><strong>4.5 Discussion and implications for the nESI process for these proteins</strong></td>
<td>139</td>
</tr>
<tr>
<td><strong>4.6 Concluding remarks</strong></td>
<td>142</td>
</tr>
<tr>
<td><strong>4.7 Acknowledgements</strong></td>
<td>143</td>
</tr>
<tr>
<td><strong>4.8 References</strong></td>
<td>143</td>
</tr>
<tr>
<td><strong>Chapter 5: How the Distribution of Charges Distributes Conformation in Conformationally Dynamic Proteins</strong></td>
<td>145</td>
</tr>
<tr>
<td><strong>5.0 Declaration</strong></td>
<td>146</td>
</tr>
<tr>
<td><strong>5.1 Abstract</strong></td>
<td>147</td>
</tr>
<tr>
<td><strong>5.2 Introduction</strong></td>
<td>148</td>
</tr>
<tr>
<td><strong>5.3 Results</strong></td>
<td>151</td>
</tr>
<tr>
<td>5.3.1 Distribution of charges is revealed by mass spectrometry of p27-C permutants sprayed from solvents of different ionic strengths</td>
<td>151</td>
</tr>
<tr>
<td>5.3.2 Ion mobility shows the distribution of conformation in each of the charge states of p27-C-WT sprayed from solvents of different ionic strengths</td>
<td>154</td>
</tr>
<tr>
<td>5.3.3 CCS values calculated from Monte Carlo (MC) simulations done with implicit solvent allow a comparison between solution conformation to those in vacuo</td>
<td>158</td>
</tr>
<tr>
<td><strong>5.4 Discussion</strong></td>
<td>159</td>
</tr>
<tr>
<td><strong>5.5 Conclusions</strong></td>
<td>161</td>
</tr>
<tr>
<td><strong>5.6 Materials and Methods</strong></td>
<td>162</td>
</tr>
<tr>
<td>5.6.1 Protein preparation</td>
<td>162</td>
</tr>
<tr>
<td>5.6.2 Nano-electrospray ionisation</td>
<td>162</td>
</tr>
<tr>
<td>5.6.3 Mass spectrometry</td>
<td>162</td>
</tr>
</tbody>
</table>
Appendix 1 ................................................................. 201
Appendix 2 .................................................................. 219
Appendix 3 .................................................................. 221
Appendix 4 .................................................................. 225
Appendix 5 .................................................................. 229

Word count: 60 603
List of Figures

Chapter 1

Figure 1.1 The progression from structured proteins to intrinsically disordered proteins .................................................. 26

Figure 1.2 Structural view of disorder in protein complexes. The figure shows a static complex (a) and dynamic (b-d) complexes in order of increasing disorder.... 31

Figure 1.3 Ensemble-averaged radii of gyration (Rg) for sequence variants of (Glu-Lys)25. Insets show representative conformations for four sequence variants .... .................................................. 33

Figure 1.4 Phosphorylation events of p27........................................... 34

Figure 1.5 Schematic representation of positive mode ESI showing Taylor cone formation, production of multiply charged droplets, droplet fission, desolvation and entry of the ions into the mass spectrometer.................................................. 41

Figure 1.6 Summary of ESI mechanisms; a) the ion ejection model is the main ionisation mechanism for small molecules. b) the charged residue model is the mechanism by which globular proteins are transferred into the gas phase. c) The chain ejection model is the mechanism responsible for the ionisation of unfolded proteins. .................................................. 43

Figure 1.7 Schematic of a quadrupole mass analyser. Opposite rods have the same polarity. .................................................. 45

Figure 1.8 Schematic of a time-of-flight (ToF) analyser.. ........................................... 47

Figure 1.9 Schematic of a cross section of an MCP detector from a side view. An incident ion begins a cascade of electrons which amplifies the signal................. 48

Figure 1.10 The influence of conformation on the arrival time of an ion. Ions are pulsed into the drift tube and are drawn through by the potential gradient. Ions with the same m/z but a different CCS are separated since they will interact differentially with the buffer gas. .................................................. 25

Chapter 2

Figure 2.1 Schematic representations of procedures which involve the use of mass spectrometry. Electrospray ionisation charge state distribution analysis (a),
MS-based proteomics (b), electron capture dissociation top-down sequencing (c) and ion mobility mass spectrometry (d). ................................................................. 67

Figure 2.2 The results of deconvolution of charge state distributions of α-synuclein ions in ESI MS acquired in the pH range of 2.5–8. ............................................. 68

Figure 2.3 The results of deconvolution of charge state distributions of α-synuclein ions in ESI MS in different concentrations of ethanol, which has been proven to increase the amount of β-sheet. ....................................................... 69

Figure 2.4 α-Synuclein – a case study protein for the use of mass spectrometry to examine IDPs. ........................................................................................................ 70

Figure 2.5 Summary of cleavage site frequency of xXPA. ................................ 74

Figure 2.6 The disorder to order transition of ACTR and CBP upon complex formation, as determined by HDX [71]. ................................................................. 77

Figure 2.7 (i) Arrival time distributions for α-synuclein −7, −8 and −9 charge states at injection energies 20 V, 40 V and 90 V. (ii) CCS vs. charge for the dominant peaks in the ATD measurements. ........................................... 81

Chapter 3

Figure 3.1 Graphical abstract ........................................................................ 89

Figure 3.2 (a) Mass spectrum of Cytochrome C. (b) Mass spectrum of β-casein. (c) The rotationally averaged collision cross sections (CCSs) of each charge state of Cytochrome C (d) The rotationally averaged CCSs of each charge state of β-casein ........................................................................................................... 99

Figure 3.3 How the molecular mass of a protein relates to different mass spectrometry and ion mobility measurements. a) median charge state, b) range of charge states, c) median collision cross section and d) range of collision cross sections. Proteins shown in red are known to be unfolded, either due to intrinsic disorder or denaturing conditions. ......................................................... 101

Figure 3.4 (A) Comparison of the mean net charge and the mean scaled hydrophobicity of 10 of the investigated proteins. Comparisons of the charge state range are made to (B) the mean net charge and (C) the mean scaled hydrophobicity. .................................................................................... 107

Figure 3.5 Comparison of the theoretical and experimental collision cross sections: ........................................................................................................ 111
Figure 3.6  Range of theoretical CCSs occupied by a protein ((A) α-synuclein, (B) β-casein, and (C) Cytochrome C) is shown in red. The relative intensity of each charge state is plotted against its CCS. On the right-hand side, the illustration shows a postulated shape of the folding landscape of each protein. .............................. 114

Figure 3.7  Flowchart displaying the procedure for elucidating structural information on an unknown protein. ................................................................. 116

Chapter 4

Figure 4.1  Comparison of the biophysical characteristics of ApoC-II and α-Synuclein. (a) General properties of the proteins. (b) Sequences showing positive and negative residues in green and red, respectively. (c) Wimley–White plots showing the hydrophobicity of the proteins. ................................................................. 126

Figure 4.2  HDX data of three peptides of α-Syn (top) and ApoC-II (bottom). For each of the three peptides, the mass change at 15 s indicates that the maximum number of type II deuteriums that could exchange for hydrogens, have already been exchanged. ................................................................................. 132

Figure 4.3  nESI of α-Syn. In both (a) and (b), the solution conditions are identical (50 mM ammonium acetate pH 6.8) but the CSD differs. In (c) the ionic strength has been increased to 100 mM ammonium acetate and the pH lowered to 3.5. .......... 134

Figure 4.4  Mass spectra of ApoC-II taken under different solution conditions (a-d) and sprayed from different temperatures (e, f). (a) 10 mM ammonium acetate pH 6.8. (b) 100 mM ammonium acetate pH 6.8. (c) 100 mM ammonium acetate pH 2.5. (d) Methanol, (e) 100 mM ammonium acetate, 80˚C, (f) 100 mM ammonium acetate, 14˚C. .......................................................................................................................... 136

Figure 4.5  Ion mobility data of α-Synuclein (a) and ApoC-II (b). Collision cross sections of each charge state (top) and arrival time distributions of selected charge states (bottom).............................................................................................................. 137

Figure 4.6  Three proposed electrospray mechanisms; (a) the charged residue model proposed by Dole [47] and (c) the chain ejection model proposed by Konermann et al. [46]. We also propose an intermediate between the two previously described mechanisms (b). Figure adapted from Ref. [46].]................................. 140
Chapter 5

Figure 5.1  Mass spectra of p27-C permutants sprayed from low and high ionic strength solutions. p27-C-WT (a and b), p27-C-κ0.14 (c and d) and p27-C-κ0.56 (e and f). All permutants were sprayed from ammonium acetate solutions of 10 mM (a, c and e) and 200 mM (b, d and f) ammonium acetate. Peaks annotated with a red asterisk are due to a contaminant peptide with a mass of 4870 Da. .......................... 152

Figure 5.2 p27-C-WT analysed by IM-MS. CCS distributions of each charge state sprayed from 10 mM ammonium acetate (grey) and 200 mM ammonium acetate (purple). .......................................................... 154

Figure 5.3 IM-MS data of p27-C-WT in 10 mM ammonium acetate (a) and 200 mM ammonium acetate (b). The CCS distribution of each charge state is shown. The height of each curve is normalised to that of the MS peak shown in figures 5.1a and 5.1b. .......................................................... 156

Figure 5.4 CCS distributions calculated from molecular coordinates obtained from Monte Carlo simulations of the permutants in implicit solvent. Grey; p27-C-WT. Red; p27-C-κ0.14 and Blue; p27-C-κ0.56. .......................................................... 158

Chapter 6

Figure 6.1  Top; Conformational flexibility of the C-terminal domain of p27 in the p27/Cdk2/cyclin A complex. Superposition of p27/Cdk2/cyclin A structures at varying time intervals during a 12.7-ns MD simulation. Bottom; Schematic of the complex containing CDK2 (blue), Cyclin A (purple) and p27 (black) containing different C-terminal domains; wild-type (black), κ0.14 (red) and κ0.56 (blue). ...... 171

Figure 6.2 MS of Cdk2/ Cyclin A complex (a) and of Cdk2 /Cyclin A/ p27-WT (b), Cdk2/ Cyclin A /p27κ0.14 (c) and Cdk2/ Cyclin A/ p27-κ0.56 (d)................................. 172

Figure 6.3 DTCCS_{the} distributions of each charge state of the Cdk2/ Cyclin A complex (a) and of the Cdk2/ Cyclin A/ p27-WT (b), Cdk2/ Cyclin A /p27-κ0.14 (c) and Cdk2/ Cyclin A /p27-κ0.56 (d). The red lines are the sum of the distributions of the four charge states of each complex (shown by the blue lines). The purple dashed line corresponds to the calculated CCS of the crystal Structure of the Cdk2/ Cyclin A complex [16]. The black dashed line corresponds to the calculated CCS of the crystal Structure of the Cdk2/ Cyclin A/ N-terminal p27 [17]................................. 173
Figure 6.4  Average $^{\text{DT}}$$^\text{CCS}_{\text{He}}$ values for each charge state (filled squares) and overall $^{\text{DT}}$$^\text{CCS}_{\text{He}}$ spread (lines) of C-terminal p27, full length p27, the Cdk2/ Cyclin A complex and the Cdk2/ Cyclin A/ p27 complex. The WT protein is shown in black, the κ0.14 variant is shown in red and the κ0.56 variant is shown in blue. The data points have been offset wrt to the molecular mass which is the same for each set of permutants /permutant complexes.

Chapter 7

Figure 7.1;  Native MS of 5 μM Fdc1 (top) and 10 μM Fdc1$^{\text{UbiX}}$ (bottom) ....... 183

Figure 7.2;  Arrival time distributions of the 3 main charge states of Fdc1 (top) and Fdc1$^{\text{UbiX}}$ with one bound cofactor and two bound cofactors. ............................. 185

Figure 7.3;  Top; the difference in deuterium uptake between Fdc1 and Fdc1$^{\text{UbiX}}$, summed across all time points. The crystal structures have the differences in deuterium uptake coloured according to the key. Secondary structure elements are shown in (a) with the cofactor shown in purple, while (b) and (c) show the surface. (b) and (c) are equivalent, but rotated $180^\circ$. PDB ID 4ZA4. In the sequence, the residues that interact with the cofactor are outlined in purple, while the residues that are involved in Mn$^{2+}$ binding are outlined in orange. ............................. 188
List of Tables

Chapter 3
Table 3.1 Percentage of Disordered Residues in Six Different Proteins, as Predicted by DISOPRED, PONDR-VL3, PONDR-VSL2, and PONDR VLXT. Data extrapolated from Figure S12 in the Supporting Information, appendix 1. ............ 109

Chapter 5
Table 5.1 amino acid sequences of C-terminal p27, and the two permutants. Acidic residues are coloured red and basic residues are coloured green. .......... 150
Table 5.2 CCS parameters calculated from MD simulations of the three p27-C permutants. .............................................................. 158

Chapter 6
Table 6.1 Average $D_TCCS_{He}$ of each complex ........................................ 174

List of Equations

Chapter 1
Equation 1.1 $F = ma$ ......................................................................................... 38
Equation 1.2 $\pm \phi_0 = \pm (U - V \cos \omega t)$ .................................................. 44
Equation 1.3 $E_k = \frac{mv^2}{2} = zEv_z$ .................................................................. 45
Equation 1.4 $t = \frac{L}{v}$ ...................................................................................... 46
Equation 1.5 $t^2 = \frac{m}{z} \left( \frac{L^2}{2eV_v} \right)$ ............................................................. 46
Equation 1.6 $v_d = K \cdot E$ .............................................................................. 50
Equation 1.7 $K = \frac{3 ze}{16N} \left( \frac{2 \pi}{\mu k_B T} \right)^{1/2} \frac{1}{\Omega}$ ........................................... 50
Equation 1.8 $K_0 = K \left( \frac{T_0}{T} \right) \left( \frac{p}{p_0} \right)$ ......................................................... 50
Chapter 3

Equation 3.1 \[ CCS_{lower} \left( \tilde{A}^2 \right) = \pi r^2 = \pi \left( \frac{3V}{4\pi} \right)^{2/3} \] ........................................ 96

Equation 3.2 \[ \bar{V} = \sum_{i}^{\text{amino acids}} V_i n_i \] ........................................ 97

Equation 3.3 \[ CCS_{upper} \left( \tilde{A}^2 \right) = \left( \text{Projection area of cylinder} \right) = \frac{4}{\pi} rl + 2r^2 \] .... 97

Equation 3.4 \[ Z_R = 0.0778\sqrt{m} \] ........................................ 102

Chapter 4

Equation 4.1 \[ K_0 = \frac{3ze}{16N} \left( \frac{2\pi}{\mu k_B T} \right)^{1/2} \frac{1}{\bar{N}} \] ........................................ 125

Equation 4.2 \[ \Omega_{\text{avg}} = \frac{(18\pi)^{1/2}}{16} \left[ \frac{1}{m_b} + \frac{1}{m} \right]^{1/2} \frac{ze}{(k_B T)^{1/2}} \frac{1}{\rho L^2} \bar{t} \] ........................................ 130

Equation 4.3 \[ t_D = t_a - t_0 \] ........................................ 130

Equation 4.4 \[ CCS_{lower} \left( \tilde{A}^2 \right) = \pi r^2 = \pi \left( \frac{3V}{4\pi} \right)^{2/3} \] ........................................ 131

Equation 4.5 \[ \bar{V} = \sum_{i}^{\text{amino acids}} V_i n_i \] ........................................ 131

Equation 4.6 \[ CCS_{upper} \left( \tilde{A}^2 \right) = \left( \text{Projection area of cylinder} \right) = \frac{4}{\pi} rl + 2r^2 \] 131

Chapter 5

Equation 5.1 \[ \Omega_{\text{avg}} = \frac{(18\pi)^{1/2}}{16} \left[ \frac{1}{m_b} + \frac{1}{m} \right]^{1/2} \frac{ze}{(k_B T)^{1/2}} \frac{1}{\rho L^2} \bar{t} \] ........................................ 163

Equation 5.2 \[ t_D = t_a - t_0 \] ........................................ 163

Chapter 7

Equation 7.1 \[ K_0 = \frac{3ze}{16N} \left( \frac{2\pi}{\mu k_B T} \right)^{1/2} \frac{1}{\bar{N}} \] ........................................ 192

Equation 7.2 \[ \Omega_{\text{avg}} = \frac{(18\pi)^{1/2}}{16} \left[ \frac{1}{m_b} + \frac{1}{m} \right]^{1/2} \frac{ze}{(k_B T)^{1/2}} \frac{1}{\rho L^2} \bar{t} \] ........................................ 193

Equation 7.3 \[ t_D = t_a - t_0 \] ........................................ 193
### Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>α-Syn</td>
<td>α-Synuclein</td>
</tr>
<tr>
<td>ACTR</td>
<td>activator of thyroid and retinoid receptors</td>
</tr>
<tr>
<td>AmAm</td>
<td>Ammonium acetate</td>
</tr>
<tr>
<td>ApoC-II</td>
<td>Apolipoprotein C-II</td>
</tr>
<tr>
<td>ATD</td>
<td>Arrival time distribution</td>
</tr>
<tr>
<td>ATP</td>
<td>Adenosine triphosphate</td>
</tr>
<tr>
<td>CBP</td>
<td>CREB binding protein</td>
</tr>
<tr>
<td>CCS</td>
<td>Collision cross section</td>
</tr>
<tr>
<td>CD</td>
<td>Circular dichroism</td>
</tr>
<tr>
<td>Cdk2</td>
<td>Cyclin dependent kinase 2</td>
</tr>
<tr>
<td>CEM</td>
<td>Chain ejection model</td>
</tr>
<tr>
<td>CRM</td>
<td>Charged residue model</td>
</tr>
<tr>
<td>CID</td>
<td>Collision induced dissociation</td>
</tr>
<tr>
<td>CSD</td>
<td>Charge state distribution</td>
</tr>
<tr>
<td>Da</td>
<td>Dalton</td>
</tr>
<tr>
<td>DC</td>
<td>Direct current</td>
</tr>
<tr>
<td>DISOPRED</td>
<td>Disorder predictor server</td>
</tr>
<tr>
<td>DT IM-MS</td>
<td>Drift tube ion mobility- mass spectrometry</td>
</tr>
<tr>
<td>ΔCCS</td>
<td>Collision cross section range</td>
</tr>
<tr>
<td>ΔZ</td>
<td>Charge state range</td>
</tr>
<tr>
<td>EHSS</td>
<td>Exact hard sphere scattering</td>
</tr>
<tr>
<td>ESI</td>
<td>Electrospray ionisation</td>
</tr>
<tr>
<td>FAIMS</td>
<td>Field asymmetric ion mobility spectrometry</td>
</tr>
<tr>
<td>FCR</td>
<td>Fraction of charged residues</td>
</tr>
<tr>
<td>FMN</td>
<td>Flavin mononucleotide</td>
</tr>
<tr>
<td>FRET</td>
<td>Föster resonance energy transfer</td>
</tr>
<tr>
<td>FTIR</td>
<td>Fourier transform infrared spectroscopy</td>
</tr>
<tr>
<td>Glu</td>
<td>Glutamic acid</td>
</tr>
<tr>
<td>HDX</td>
<td>Hydrogen-deuterium exchange</td>
</tr>
<tr>
<td>HFiP</td>
<td>hexafluoroisopropanol</td>
</tr>
<tr>
<td>HMGA</td>
<td>High Mobility Group A</td>
</tr>
<tr>
<td>IAPP</td>
<td>Islet amyloid polypeptide</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Definition</td>
</tr>
<tr>
<td>--------------</td>
<td>------------</td>
</tr>
<tr>
<td>IDP</td>
<td>Intrinsically disordered protein</td>
</tr>
<tr>
<td>IDR</td>
<td>Intrinsically disordered region</td>
</tr>
<tr>
<td>IEM</td>
<td>Ion ejection model</td>
</tr>
<tr>
<td>IM-MS</td>
<td>Ion mobility-mass spectrometry</td>
</tr>
<tr>
<td>IMS</td>
<td>Ion mobility spectrometry</td>
</tr>
<tr>
<td>KID</td>
<td>Kinase inhibitory domain</td>
</tr>
<tr>
<td>LC</td>
<td>Liquid chromatography</td>
</tr>
<tr>
<td>Lys</td>
<td>Lysine</td>
</tr>
<tr>
<td>m/z</td>
<td>Mass-to-charge ratio</td>
</tr>
<tr>
<td>MALDI</td>
<td>Matrix assisted laser desorption ionisation</td>
</tr>
<tr>
<td>MCP</td>
<td>Microchannel plate</td>
</tr>
<tr>
<td>MC</td>
<td>Monte Carlo</td>
</tr>
<tr>
<td>MD</td>
<td>Molecular Dynamics</td>
</tr>
<tr>
<td>MOBCAL</td>
<td>A program to CALculate MOBilities</td>
</tr>
<tr>
<td>MoQToF</td>
<td>Mobility quadrupole time-of-flight mass spectrometer</td>
</tr>
<tr>
<td>MS</td>
<td>Mass spectrometry</td>
</tr>
<tr>
<td>NCPR</td>
<td>Net charge per residue</td>
</tr>
<tr>
<td>nESI</td>
<td>Nano-electrospray</td>
</tr>
<tr>
<td>NMR</td>
<td>Nuclear magnetic resonance</td>
</tr>
<tr>
<td>NOE</td>
<td>Nuclear overhauser effect</td>
</tr>
<tr>
<td>p27-C</td>
<td>C-terminal p27</td>
</tr>
<tr>
<td>PA</td>
<td>Projection approximation</td>
</tr>
<tr>
<td>PD</td>
<td>Parkinson's disease</td>
</tr>
<tr>
<td>PDB</td>
<td>Protein data bank</td>
</tr>
<tr>
<td>pI</td>
<td>Isolectric point</td>
</tr>
<tr>
<td>PONDR</td>
<td>Predictor of naturally disordered regions</td>
</tr>
<tr>
<td>prFMN</td>
<td>Prenylated flavin mononucleotide</td>
</tr>
<tr>
<td>PRE</td>
<td>Paramagnetic resonance enhancement</td>
</tr>
<tr>
<td>PSA</td>
<td>Projection superposition approximation</td>
</tr>
<tr>
<td>PSIPRED</td>
<td>Protein structure prediction server</td>
</tr>
<tr>
<td>PTM</td>
<td>Post-translational modification</td>
</tr>
<tr>
<td>Q-ToF</td>
<td>Quadrupole time-of-flight mass spectrometer</td>
</tr>
<tr>
<td>RF</td>
<td>Radiofrequency</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Definition</td>
</tr>
<tr>
<td>--------------</td>
<td>------------</td>
</tr>
<tr>
<td>ROA</td>
<td>Raman optical activity</td>
</tr>
<tr>
<td>RDC</td>
<td>Residual dipolar coupling</td>
</tr>
<tr>
<td>$R_g$</td>
<td>Radius of gyration</td>
</tr>
<tr>
<td>SARA</td>
<td>Smad anchor for receptor activation</td>
</tr>
<tr>
<td>SAXS</td>
<td>Small angle x-ray scattering</td>
</tr>
<tr>
<td>SBD</td>
<td>Smad binding domain</td>
</tr>
<tr>
<td>smFRET</td>
<td>Single molecule Förster Resonance Energy Transfer</td>
</tr>
<tr>
<td>TM</td>
<td>Trajectory method</td>
</tr>
<tr>
<td>TOF</td>
<td>Time-of-flight</td>
</tr>
<tr>
<td>TWIMS</td>
<td>T-wave ion mobility spectrometry</td>
</tr>
<tr>
<td>UV</td>
<td>Ultraviolet</td>
</tr>
<tr>
<td>wt</td>
<td>Wild-type</td>
</tr>
<tr>
<td>XPA</td>
<td>Xeroderma pigmentosum group A</td>
</tr>
</tbody>
</table>
Abstract

Research into the relationship between the structure and function of proteins has been ongoing now for several decades. More recently, there has been an explosion in the investigation of the dynamic properties of proteins, and how their dynamic propensity relates to their function. This new direction in protein research requires new techniques to analyse protein dynamics, since most traditional techniques are biased towards a fixed tertiary structure. Mass spectrometry (MS) is emerging as a powerful tool to probe protein dynamics since it can provide information on interconverting conformations and has no preference towards the folded state. Furthermore, its low sample consumption, rapid data acquisition and low data processing positions MS as an attractive tool in protein structure research. The hybrid technique of ion mobility-mass spectrometry provides further insight into the range of conformations adopted by proteins and protein complexes, by providing information on the size in terms of rotationally averaged collision cross section. The work presented in this thesis considers proteins with a range of structural characteristics. We use ion mobility mass spectrometry to investigate proteins of different extents of disorder, protein complexes with dynamic entities and a system that undergoes structural rearrangement upon ligand binding.

First, a framework of mass spectrometry experiments is described which allows identification of the extent of structure and disorder within proteins. This framework is tested on a range of different systems throughout the thesis. Differences in the gas-phase properties of two conformationally dynamic proteins which behave similarly in solution are investigated and from this research we postulate a new ionisation mechanism for partially folded proteins. The dynamic propensity of C-terminal p27 is investigated and compared to two permutants which allows us to delineate how the location of charged residues in a primary sequence affects the structure of a protein. We monitor the ‘folding-upon-binding’ behaviour of p27 upon association with its binding partners, and how this differs with the order of charged residues in the linear sequence. Finally, we describe the structural rearrangement of Fdc1 upon the binding of its cofactor; a prenylated FMN molecule.

This thesis demonstrates the suitability of ion mobility-mass spectrometry for the investigation of dynamic properties of proteins and protein complexes.
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 she 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 the 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://documents.manchester.ac.uk/DoculInfo.aspx?DocID=487), in any relevant Thesis restriction declarations deposited in the University Library, The University Library’s regulations (see http://www.manchester.ac.uk/library/aboutus/regulations)
Acknowledgements

My first thanks goes to my supervisor, Professor Perdita Barran. Thank you for your advice and direction, as well as the freedom to investigate what I found most exciting. I’m also grateful to Professor Cait MacPhee for helpful discussions and support throughout my PhD. Dr Milena Quaglia and the rest of the team at LGC have provided great advice on my research and have always made me feel welcome on my visits.

My PhD wouldn’t have been possible without the fantastic collaborators that I’ve had the privilege of working with. I’m grateful to Richard Kriwacki and Rohit Pappu for the p27 samples and insightful discussions, and to Rahul Das for computational input to the research. I also thank Karl Payne and David Leys for involving me in their research into Fdc1, and for supplying samples. Thanks also to Lukasz for the MOBCAL calculations.

I’ve had a great four years spread between Edinburgh and Manchester and this wouldn’t have been possible without the colleagues and friends in both cities that made my time so enjoyable. Thanks to the Barran group: To Jason, Hattie, Yana, Ewa, Sophie and Kamila, thank you for keeping me right in Edinburgh during my early years and for providing me with a solid background in the arts of coffee making, cake eating and mass spectrometry. To Jakub, Ellie, Chris N, Ashley, Bruno, Cunyu, Rosie, Lukasz, Alina and Jacqui, thanks for a great time in Manchester and the many coffees, chats, beers and parties; you’ve all been amazing. Additional people in the MIB who have made my time here so enjoyable include Andrew, Maria, Chris G, Giles, Irma, Reynard, Nik and Cat. Thanks for being so welcoming.

I’m grateful to all my friends, old, new, near and far, who have provided much-needed distraction and kept me (kind of) sane. In particular; Sam, Jonny, Caz, Nath, Adam, Jamie… you are all stars. Thanks also to my brother and sister, Andrew and Karen, for the encouragement, support and reality checks. You both keep my feet on the ground.

Lastly, and absolutely most importantly, I cannot thank my wonderful parents enough. Noel and Brenda Beveridge, I am so grateful for the love and support you’ve given me over the years. I simply could not have done this without you. From the bottom of my heart, thank you.
Preface

This thesis describes my research on the use of ion mobility-mass spectrometry (IM-MS) to investigate the dynamic properties of proteins and protein complexes. The introduction provides a background to protein structure and dynamics, and includes a comprehensive overview of intrinsically disordered proteins (IDPs) which are a key feature of this thesis. IM-MS theory is described in depth, and I outline other associated techniques that are utilized in this thesis. I also include in the introduction a published review which outlines the use of mass spectrometry and associated methods in the analysis of IDPs.

The main body consists of five first author papers, two of which have been published at the time of thesis submission and three of which have been prepared to be submitted for publication. The format of these papers is modified to fit in the template of the thesis, but no modifications are made to the main text. References for the paper are at the end of each manuscript, numbered as in the submitted version. Supporting information for each research article is provided in the appendices.

The first paper in the main body (Chapter Three) is a perspective article that compiles a large amount of MS and IM-MS data of proteins with different structural preferences. Trends observed herein were used to develop a framework of MS-based experiments to elucidate the extent of disorder in proteins. This framework was then tested during the analysis of the IDPs α-Synuclein and Apolipoprotein C-II (Chapter Four) and p27 (Chapters Five and Six). The application of MS methods to delineate changes in conformation and dynamics of an enzyme upon cofactor binding is presented in Chapter Seven. Conclusions and outlook are discussed in Chapter Eight. The supplementary information for chapters 3-7 is found in the appendices.
“I don't pretend we have all the answers.
But the questions are certainly worth thinking about”

Arthur C. Clarke
1

Introduction
1.1 Protein Structure

Proteins are large, complex molecules that play many critical roles in the human body. They are the building blocks for skin and muscles; antibodies and enzymes are proteins; processes such as growth, digestion, reproduction and respiration rely on proteins, and these are just a few of their important roles.

A protein molecule is a string of amino acids, of which there are 20. Each protein has a unique sequence of amino acids. The way in which such polypeptide chains self-organise into specific 3-dimensional (3D) shapes has been the subject of investigation for many years. An early view of protein structure was the simplistic ‘lock and key’ model proposed by Emil Fischer in 1894 [1, 2] in which protein – ligand and protein – protein interactions are dependent upon exact stereochemical recognition between two or more partners. The hypothesis was formulated to explain the specificity of the enzymatic hydrolysis of glucoside multimers by related enzymes; one enzyme could hydrolyse α- but not β- glycosidic bonds, and another could only hydrolyse β- glycosidic bonds. Fischer wrote, as translated by Lemieux and Spohr [2], “To use a picture, I would like to say that enzyme and glucoside have to fit each other like a lock and key in order to exert a chemical effect on each other.” In this analogy the key is the substrate while the lock is the enzyme and the key-hole is the active site of the enzyme. This model was based on the assumption that every protein has one energy-minimised structure, essential for its biological function, and was in agreement with emerging evidence of 3D structures being solved by crystallography [1, 3, 4].

Karush was among the first to reconsider this structure function paradigm [5] and proposed an alternative theory of conformational adaptability, which was referred to independently as the “induced fit theory” by Koshland [6]. As increasing numbers of protein structures were being solved by x-ray crystallography it was noticed that the resolution could vary across the unit cell, and the idea of localised disorder was born [1]. The concept that proteins can be functional and yet lack a resolvable 3D structure has gained momentum in recent years. It is now widely accepted that there is a progression of structure ranging from proteins that are highly structured, through to proteins which are completely disordered (Figure 1.1).

1.1.1 The Structure – Disorder Continuum

Proteins fall onto a structural progression ranging from structured with minimum dynamics, to disordered with minimum structure. Intermediate levels of structure
include folded proteins with localised disorder, folded domains linked by disordered regions, and molten globules.

At the ‘structured’ end are single-domain proteins which were originally thought, as discussed above, to have a fixed, tertiary conformation that is essential for their function. We now recognise that whilst these proteins are relatively rigid, they still fluctuate between closely related conformers. This behaviour facilitates functions such as catalysis and macromolecular associations [7].

Some folded proteins may have areas of localised disorder, perhaps in the form of a disordered tail or loop. Proteins with linked, folded domains have independent regions of structure joined together by flexible linker regions. Such proteins often behave like beads on a string since the linker domains allow for spatial flexibility between the structured regions, enabling binding to many different partners. This ‘binding promiscuity’ will be discussed in section 1.2.2. Molten globules contain some elements of flexible secondary structure but not the fixed tertiary structure or tight packing of sidechains of structured proteins [8-10]. Formation is likely due to hydrophobic collapse and the structure is often en-route to a folded state. Some proteins however exist as stable molten globules, indicating that such partially-
folded states are often functional [11, 12]. Physiological processes associated with molten globules include interaction with chaperones [13], protein insertion into membranes [14] and interactions with ligands [15].

The location of a given protein within this structure continuum appears to be related with function; structured proteins tend to be involved in biosynthesis and transport [16] whereas disordered proteins are often involved in regulation of transcription and translation, cellular signal transduction, protein phosphorylation, storage of small molecules and regulation of the self-assembly of large, multiprotein complexes such as flagella and ribosomes [17-21].

1.2 Intrinsically Disordered Proteins

Intrinsically disordered proteins are polypeptide chains which exist and function without a well-defined 3D conformation under physiological conditions [22]. They are more flexible and dynamic than folded globular proteins and are often classified as having no secondary structure on the timescale of an NMR experiment [16-18, 23, 24]. Intrinsic disorder is highly prevalent in biology. Bioinformatics studies indicate that about 25 – 30% of eukaryotic proteins are mostly disordered [25], that more than half of eukaryotic proteins have long segments of disorder [25], and that more than 70% of signalling proteins have long disordered regions [23, 26-28].

Many different terms have been used to classify the disordered characteristics of IDPs. These include natively denatured [29], natively unfolded [30], intrinsically denatured [29], intrinsically unstructured [18], intrinsically unfolded [30], intrinsically disordered [26], rheomorphic [31], floppy [32], flexible [33], mobile [34], partially folded [35], vulnerable [36], chameleon [37], malleable [38], dancing proteins [39] and protein clouds [40]. Each of these terms has advantages and limitations. The terms intrinsically unstructured and natively unfolded may be suitable for extended random coils but are not appropriate for proteins that form transient or fixed secondary structure. The term disorder is inappropriate due to its negative connotation and its possible confusion with a pathological state. Disorder can also be used, however, for proteins like the molten globule which contain secondary structure but are also dynamic. Intrinsically disordered seems to have been considered within the community the most appropriate term since it is the most widely used.
1.2.1 Biophysical Properties

IDPs exist as dynamic ensembles in which the atom positions within the protein and backbone Ramachandran angles vary greatly with no specific equilibration values, and these ensembles typically undergo non-cooperative conformational changes [41]. IDPs have a shallow conformational landscape, meaning that they can populate many interconverting, coil-like conformations of similar energy [18, 27, 42].

The behaviour of IDPs in response to changes in the environment is very different to that of structured proteins [43]. The behaviour of IDPs is characterised by a partial or complete lack of cooperativity during denaturant-induced unfolding, by the lack of heat absorption peaks which can be measured during the melting of globular proteins, by a gain of structure in response to heat and changes in pH, by the gaining of structure in the presence of specific counter ions and binding partners and by protein-specific responses to macromolecular crowding [43].

IDPs often develop an increase in compaction in response to high temperatures, as opposed to thermally induced denaturation that is typically experienced by globular proteins. This compaction may be due to hydrophobic interactions becoming stronger at higher temperatures, resulting in a stronger hydrophobic driving force for partial folding [20, 44]. A decrease in the pH has been shown to induce compaction of extended IDPs due to a minimisation of their large net charge, therefore decreasing Coulombic repulsion and allowing hydrophobic-driven collapse [20, 44]. These characteristics can be exploited during separation of IDPs from whole cell extracts; the indifference of IDPs to conditions which denature globular proteins allows the isolation of IDPs as these are the only proteins which will ‘survive’ harsh conditions while other proteins precipitate [45].

IDPs in native conditions have different biophysical properties to structured proteins in a denatured state in terms of their hydrodynamic behaviour. Most IDPs do not behave as random coils; they are relatively compact due to transient elements of secondary structure (mainly α-helix) which reduce their hydrodynamic radius [41]. Dynamic tertiary structures can also occur as a result of non-covalent interactions such as hydrogen bonding and electrostatic interactions [41]. These interactions are highly dependent on the environment, and hence the presence of extrinsic species such as salts will have a large effect on the ‘compactness’ of a protein [46].
1.2.2 Formation of Complexes

Many IDPs function by molecular recognition and are hence integral components of protein complexes. An interesting and important characteristic of IDPs is their ‘binding promiscuity’; their ability to bind to many different partners. This allows them to play a key role within cellular signalling networks [47-50] and can result in three different behaviours [47]:

1. One IDP binds to many different structured proteins [48]
2. Multiple IDPs bind to one structured protein [47]
3. Intrinsically disordered regions act as flexible linkers between ordered domains, allowing binding promiscuity [47, 48, 51]

Often, but not always, IDPs are induced to fold upon interaction with a specific binding partner [52-54]. When binding to a structured domain the free energy required for the disorder to order transition is subtracted from the contact free energy, resulting in a highly specific interaction along with a low energy of association. p21 is known to undergo a disorder-to-order transition upon binding to cyclin-Cdk complexes; in sharp contrast to the disordered free state, the N-terminus of p21 adopts an ordered stable conformation when bound to Cdk2 as shown directly by NMR spectroscopy [55]. 2D $^1$H-$^{15}$N HSQC NMR spectra for p21 alone exhibits limited resonance dispersion, typical of disordered proteins. In complex with unlabelled Cdk2 there is a large increase in resonance dispersion, indicative of increased structure. This change in conformational entropy may be associated with increased specificity for this protein-protein interaction at the expense of absolute thermodynamic binding affinity. The implication is that, if p21 existed in a more ordered free state, its affinity for Cdk2 might be greater, albeit at the expense of specificity [55]. These specific yet weak interactions are suitable for signalling cascades as they ensure reversibility which is essential for turning the signal ‘off’, promptly [26].

Another scenario of the association of an IDP to its partner is the formation of a dynamic complex in which the IDP samples various conformations on the surface of the partner [56, 57]. Again, disorder within complexes lies on a continuum, much like isolated proteins; the disordered part of the complex may adopt several alternative conformations, it may remain disordered but connect or neighbour binding regions which have undergone induced structure or, at the extreme, the
bound IDP may remain entirely disordered (Figure 1.2). An example is the interaction between the Smad2 Mad homology domains and the Smad binding domain (SBD) of Smad anchor for receptor activation (SARA). The complex formed between the SBD and Mad homology domain appears to be heterogeneous, with no single region of the SBD making a dominant contribution to the hydrophobic interactions between the proteins [58]. It has been shown that disorder in complexes is often beneficial since it can add adaptability, versatility and reversibility to protein binding and therefore an ease of regulation [57].

Figure 1.2 shows the different levels of structure in protein complexes. Structural heterogeneity may be the result of a number of well-defined conformations which can individually be resolved at the atomic level (Figure 1.2a). These different conformations may result in different functional outcomes, and illustrate how dynamic complexes can be exploited for regulatory functions [59, 60]. In Figure 1.2b folded regions are connected by a linker region which remains disordered in the bound state; the ordered regions serve as clamps, and the linker region limits the conformational freedom between these domains. This ‘clamp’ mode of binding has been suggested to be common due to its ability to provide flexibility and adaptability in recognition [61, 62]. Figure 1.2c shows the binding of an IDP to its partner through short recognition elements within the disordered environment, therefore containing a flanking segment that retains its conformational variability in the complex [63, 64]. The most extreme version of disorder in protein complexes is shown in Figure 1.2d where the whole or majority of the protein remains disordered in the bound state [65, 66]. Overall, disorder in protein-protein interactions can be beneficial in a variety of functional settings; however the bias of experimental methods and approaches towards gaining well-defined structures of complexes has caused this phenomenon to remain largely overlooked in the past.
Figure 1.2 Structural view of disorder in protein complexes. The figure shows a static complex (a) and dynamic (b-d) complexes in order of increasing disorder. (a) shows the polymorphic model in which there is more than one resolvable bound conformation of the IDP, exemplified by the binding of Tcf4 (blue) or Tcf3 (magenta) with β-catenin [67]. (b) shows the clamp model in which a disordered segment links two structured regions, exemplified by the binding of Ste5p (magenta) to MAP kinase Fus3p [68]. (c) shows a flanking segment of the bound protein, not present as part of the solved structure, making a significant contribution to the region directly bound to the partner. This is depicted by binding of CREB KID (magenta) to CBP-KIX [69]. (d) shows the random model where the entire protein remains disordered in the bound state. This is visualised by a hypothetical model of T cell receptor ζ-chains [65]. Image taken from Tompa and Fuxreiter [57].

1.2.3 Disorder in Disorders; the \( D^2 \) Concept

Due to their important role in many biological functions and central positions in protein interaction networks, IDPs are extensively associated with human diseases including cancer, diabetes, cardiovascular diseases, neurodegenerative diseases and amyloidoses [28, 70-72]. Intrinsic disorder is also abundant in proteins from pathogenic microbes and viruses [73-75]. This conclusion is based both on findings from individual case studies and extensive bioinformatics studies investigating the role of IDPs in a variety of pathological conditions.
An analysis of the human disease genome revealed that intrinsic disorder is common in proteins associated with many human genetic conditions [76]. By comparing protein function, with the level of disorder in the family of proteins which carry out this function, it was shown that there is a strong correlation between many diseases and proteins predicted to be disordered [50, 77, 78].

1.2.4 Primary Sequence Characteristics of IDPs

Understanding how disordered regions mediate function requires accurate physical descriptors of sequence-disorder relationships. IDPs have a lower sequence complexity than that of structured proteins. Certain amino acids have been proposed to be more disorder inducing than others. IDPs and IDR$s$ are often enriched in P, E, K, S and Q, yet depleted in W, Y, F, C, I, L and N in comparison to folded, globular proteins [23, 26, 79]. This low hydrophobicity is thought to prevent the formation of a hydrophobic cluster, the high charge is thought to promote intramolecular electrostatic repulsion favouring extended conformations, and proline is known to disrupt secondary structural elements [80]. The net charge per residue (NCPR, defined as $NCPR = f_+ - f_-$ where $f_+$ and $f_-$ are the fractions of positively and negatively charged residues, respectively) is a useful parameter to predict whether a polyelectrolytic IDP will form a globule or a swollen coil. However, most IDPs are polyampholytic (contain both positively and negatively charged residues) rather than polyelectrolytic (contain either positively or negatively charged residues) and the NCPR is therefore inadequate to describe the sequence-ensemble relationships of such proteins.

1.2.5 Relationship Between Charge Distribution and Disorder

Das and Pappu have recently proposed that the linear sequence patterning of oppositely charged residues will influence the plasticity of an IDP [81]. They introduced a patterning parameter $\kappa$ to describe the different sequence variants based on the linear sequence distributions of oppositely charged residues. $\kappa$ values lie between 0 and 1; low values relate to well-mixed sequences of positive and negative residues and at $\kappa=1$ oppositely charged residues are segregated in the linear sequence. Metropolis Monte Carlo simulations were used to demonstrate that low $\kappa$-values give rise to an extended conformation since intrachain electrostatic repulsions and attractions are balanced which results in a random-like chain. As the
κ-value increases the protein adopts a more compacted conformation caused by long range electrostatic attractions.

Figure 1.3 shows the relationship between the $R_g$ of 30 sequence variants of (Glu-Lys)$_{25}$ spanning a range of κ-values ranging from $κ = 0.0009$ in which Glu and Lys alternate, to $κ = 1$ in which Glu and Lys are segregated in the linear sequence. There is a general decrease of $R_g$ with respect to the κ-value of the sequence. Additionally, the $R_g$ values are greater than expected for classical Flory random coils (~18Å), the smallest $R_g$ value which was obtained for $κ = 1$ is 1.6 times greater than that expected of a compact globule (11 Å) [82] and, for well mixed sequences, the $R_g$ values are slightly larger than expected for self-avoiding random walks (~28 Å).

![Figure 1.3](image.png)

**Figure 1.3** Ensemble-averaged radii of gyration ($R_g$) for sequence variants of (Glu-Lys)$_{25}$. Insets show representative conformations for four sequence variants. Side chains of Glu are shown in red while side chains of Lys are shown in blue. The top dashed line shows the predicted $R_g$ value modelled as the EV limit, while the lower dashed line shows that modelled as the Flory random coil. It can be seen that as the κ-value increases, the $R_g$ is reduced. Figure taken from [81].
1.2.6 Regulating the Unfoldome

Signalling mechanisms often involve posttranslational modifications that couple an upstream input to a conformational change, which alters the function of a protein and produces a downstream signal [83]. The extent of conformational change can range from subtle, localised unfolding to full unfolding of entire domains. Regulated unfolding mechanisms occur in cell cycle control and programmed cell death.

The conformational manipulation of the flexibility of p27 via post-translational modifications provides an example of how the (un)folded state of a protein may be regulated in order to control its activity. Here, the progression of the cell from G2 to S-phase is controlled by p27 and its interaction with cyclin-dependent kinase 2 (Cdk2) and Cyclin A in the nucleus [84], which inhibits the enzymatic activity of the kinases. The first step of reactivation of Cdk/ Cyclin is performed by non-receptor tyrosine kinases which phosphorylate Tyr88 of p27 which is a residue that binds to the active site of Cdk2, resulting in partial restoration of kinase activity of Cdk2. p27 is then labelled for degradation by phosphorylation at Thr187 [85], and this phosphorylation is performed by the Cdk2/ Cyclin A complex to which p27 is bound [85, 86]. Degradation of p27 via the proteasome pathway then reinstates full Cdk2 activity and allows progression of the cell to S phase.

Figure 1.4 Step 1; phosphorylation of Y88 of p27 by non-receptor tyrosine kinases resulting in the ejection of Y88 from the active site of Cdk2 and restoration of partial kinase activity. Step 2; phosphorylation of T187 within the flexible C-terminal domain of p27 by Cdk2 creating a signal for ubiquitination of Lys residues within C-terminal p27 (step 3). Step 4; ubiquitinated p27 is degraded leading to the release of the fully active Cdk/cyclin complex and ultimately progression of the cell into S phase. Image taken from [83].
It has been shown that intrinsic flexibility of the C-terminal domain of p27 allows Thr187 to drift into the active site of the active site of Cdk2 and become phosphorylated. This leads to p27 ubiquitination and degradation and hence full activation of Cdk/ cyclin complexes. Here, tyrosine phosphorylation regulates partial unfolding of the inhibitory conformation of p27 results in this signalling cascade that results in cells passing into S phase of the division cycle.

There are many other ways in which regulated unfolding propagates biological signals. A theme of this phenomenon is that through different triggering mechanisms, such as post-translational modifications or ligand binding, regulated unfolding is used to alter the dynamic properties of proteins, or regions of a protein, which results in altering their function. Enhanced sampling of unfolded states in response to a stimulus provides physical mechanisms for proteins to transmit biological signals [87].

1.3 Biophysical Techniques and Computational Methods for the Detection and Analysis of IDPs

Several experimental approaches are sensitive to the dynamic behaviour of proteins and have hence been used as tools to characterise IDPs.

1.3.1 Small-angle X-ray Scattering

Small-angle X-ray scattering (SAXS) is an established method with which to study the low resolution structure of flexible systems, and structural transitions of biological macromolecules in solution [88]. During a typical SAXS experiment a solution of protein molecules in a capillary or cuvette is illuminated by a collimated monochromatic x-ray beam and the intensity of the scattered beam is measured as a function of the scattering angle. The scattering profile of an IDP is the average of all those arising from the conformations that the protein adopts in solution, and is orientation averaged [89].

SAXS data can be interpreted in a number of ways. Kratky plots can be used to identify qualitatively disordered states and distinguish them from globular conformations. The Kratky representation is able to enhance particular features of scattering profiles, therefore allowing easier identification of degrees of
compactness [90]. A strong advantage of the use of Kratky plots is that contributions are presented from structurally distinct regions.

SAXS data can also be used to calculate the radius of gyration ($R_g$) of a protein. This is the most common way to quantify the size of proteins in solution. The experimental $R_g$ is a single value representation of the size of the molecule, which for disordered states is a representation of the average radius over all occupied conformations of the protein. It can be used to detect intrinsic disorder within proteins, as well as monitor environmental changes such as temperature [91, 92], pH [93], ionic strength [94], reducing agents [95, 96], crowding agents [97] and post-translational modifications such as phosphorylation [98, 99]. Structural changes as a result of point mutations have also been monitored by SAXS [100, 101]. Importantly, data collected via SAXS can be used as input for computational procedures for the quantitative description of conformers in which an IDP may reside [102], providing valuable information about the behaviour of the protein [103, 104].

### 1.3.2 Nuclear Magnetic Resonance

Nuclear magnetic resonance (NMR) is perhaps the most widely used analytical technique in the investigation of IDPs [105]. The spectroscopic features produced by IDPs are the same as those produced by small molecules, allowing resonance assignment, even for large IDPs [106]. This makes possible the measurement of ensemble-averaged, conformationally-dependent parameters from nuclei distributed throughout the protein [107]. Many observables can be measured for multiple sites throughout the protein to infer transient organisation at the secondary and tertiary levels including chemical shift, coupling constant, nuclear Overhauser effect (NOE), residual dipolar coupling (RDC), paramagnetic resonance enhancement (PRE) and spin relaxation.

The chemical shift is the most experimentally accessible parameter, which gives information about the chemical and electronic environment of the nucleus, and hence the local geometry of each amino acid [108-110]. The chemical shift, as well as coupling constants and NOEs, can detect local structural propensities, such as $\alpha$-helix and $\beta$-sheet populations.

By artificially introducing a spin label to a protein via a cysteine mutation of a native residue, one can measure long-range dipole-dipole interactions between each
nuclear spin and the unpaired electron on the spin label. Line-broadening due to the PRE can investigate even long-range intramolecular contacts and can be used to describe any transient tertiary organisations. Care must be taken in incorporating PRE constraints into IDP applications however since backbone and probe dynamics both have an effect on the intensity of induced effects.

One key limitation of NMR, however, is that only ensemble-averaged properties can be measured and they must be represented as averaged quantities of a heterogeneous structure ensemble [105]. Another difficulty lies in the poor $^1$H-amide chemical shift dispersion that is a hallmark of IDPs which limits the application of NMR to select IDPs whose spectra display sufficient peak dispersion [111].

### 1.3.3 Single Molecule Förster Resonance Energy Transfer

Single molecule Förster Resonance Energy Transfer (smFRET) is a well-established technique that is suited to the analysis of IDPs. It is capable of resolving and quantifying the properties of individual proteins or subpopulations [112], unlike classical ensemble approaches in which the signal is averaged over many molecules. When combined with FRET, intramolecular distance distributions and conformational dynamics can be assessed [113].

In a typical smFRET experiment a donor dye and an acceptor dye are attached to specific residues of a protein. Upon excitation of the donor dye by a focussed laser beam, energy is transferred to the acceptor [114]. The distance between the fluorophores determines the efficiency of this energy transaction, which can be determined from the rates of detected donor and acceptor photons. The changes in fluorescence intensity from the donor to the acceptor can therefore be used to identify different conformational families of a protein and the rate of interconversion between them.

### 1.3.4 Other Spectroscopic Techniques

Residual levels of secondary structure in partially disordered proteins can be provided by a number of spectroscopic techniques including far-ultraviolet (UV) circular dichroism (CD) [115, 116], Fourier transform infrared spectroscopy (FTIR) [117], Raman optical activity (ROA) [118] and deep UV resonance Raman spectroscopy [119].
1.3.5 Disorder Predictors

A number of computer programmes are available for the prediction of disordered regions from the amino acid sequence of a protein. These include PONDR [120], FoldIndex [121], DisEMBL [122], GLOBPROT2[123] and DISOPRED2 [27]. These predictors can be classified into two categories; those that predict disorder based on the amino acid sequence of the submitted protein, and those which are machine learning algorithms that are trained on existing experimental datasets which indicate disorder, for example, missing electron density in crystallographic data, or regions which have been defined as being disordered by NMR.

1.3.6 Computational Approaches

The behaviour of proteins in solution or in the gas phase can be studied by computer simulations, which provide a detailed representation of a system of interest [124]. The two main methods of simulation are molecular dynamics (MD) and Monte Carlo (MC). Information can be provided on individual atomic motions as a function of time, which allows the exploration of the conformational energy landscape accessible to protein molecules [125]. This provides insights into the mechanisms and driving forces underlying experimental results. The simulations are based on Newton’s second law of motion;

\[ F = ma \]  

Equation 1.1

Where \( F \) is the force exerted on the particle, \( m \) is the mass and \( a \) is the acceleration of the particle. These values can be integrated yielding the trajectory that describes the position, velocities and the accelerations of the particles as they vary with time.

Both MD and MC represent molecules as a collection of atom-centred interaction sites and use classical force fields for the potential energy terms [126]. A force field is a set of parameters that define how the potential energy of a molecule varies with the locations of its component atoms; such parameters may include torsional potentials, free energies of solvation, van der Waals interactions and short-range repulsive and attractive terms.

The difference between MD and MC are in the modes of sampling the conformational space of the protein. During a MD simulation new configurations are generated by applying Newton’s equations of motion to all atoms simultaneously over a small time step to determine the new atomistic positions. This is suitable for
observing the trajectory of a protein folding or unfolding since the conformation develops over time. For MC, a new configuration is generated by selecting a random molecule, translating it, rotating it, and performing any internal structural variations. Therefore, the MC is often considered to be more suitable than MD for IDPs since it permits sampling of more diverse conformations in which an IDP can exist. During MC, acceptance of the new configuration is determined by the Metropolis sampling algorithm; the Metropolis sampling algorithm is a method for obtaining random samples from a probability distribution from which direct sampling is difficult. Application over enough configurations yields properly Boltzmann-weighted averages for structure and thermodynamic properties. Mittal et al. [127] have deployed a series of moves in MC simulations that have been designed to be ergodic to enable the sampling of a broad range of conformations that are thermodynamically relevant for IDPs [128].

1.4 Biological Mass Spectrometry

Mass spectrometry (MS) was originally used for the determination of atomic isotopes [129]. For most of the 20th century the main purpose was to measure the molecular mass of small chemical compounds [130] as well as ascertain structural information via the use of dissociation. Developments of ‘soft’ ionisation methods such as electrospray ionisation (ESI) [131] have allowed MS to increase in popularity as a technique with which to study conformations of intact biological molecules. Whole protein mass spectrometry involves the transfer of a single protein, or protein complex with potentially several components, from solution into the gas phase [132]. Several MS-based studies have addressed the process of folding and unfolding both in the presence and absence of solvent. To investigate protein folding in solution, solvent conditions can be altered and the distribution of charge states can be monitored as a function of this [133, 134]. Gas phase protein folding can be examined by trapping the ions in the mass spectrometer and observing the evolution of conformations over the timescale of the experiment (ms – s) [135-137]. This ability to discern conformational flexibility by the measurement of discrete conformers within a mixture is a key strength of MS over other biophysical techniques which report on an ensemble average [138-140]. Mass spectrometry is now well positioned as an informative biophysical tool, which can be used along with other techniques to gain structural and functional information about biological molecules [141, 142].
1.4.1 Ionisation Methods

Ionisation is the first step of any MS experiment and there is a variety of ionisation techniques available. The internal energy transfer during the process is the most important consideration when choosing a suitable ionisation method. Intact protein MS requires very ‘soft’ ionisation in order to reduce fragmentation and produce ions of the molecular species. Since the development of electrospray ionisation \([131]\) and matrix assisted laser desorption/ionisation (MALDI) \([143]\) in the late 1980s, mass spectrometry has become an established method for the analysis of intact proteins with impact on biochemistry, structural biology and medicine. These ‘soft’ ionisation methods allow proteins to be transferred intact into the gas phase without excessive fragmentation, and in the case of ESI, multiprotein complexes can also be maintained \([144]\).

1.4.1.1 Electrospray Ionisation

Although in 1968 Dole et al. speculated that ESI could be used to produce a molecular beam of large molecules \([145]\), the true potential of the technique was not uncovered until 1988 when Professor Fenn demonstrated the transfer of large molecules with minimal fragmentation \([146, 147]\). ESI is now the most widely used technique, not just for protein MS but for the analysis of all samples in liquid form. This is due to characteristics such as the ease of coupling ESI to chromatographic separation techniques, the low chemical specificity, the stability of the ions and the high efficiency of ionisation \([148]\).

In an ESI experiment the analyte in solution is infused through a conductive capillary with a diameter of ~100 \(\mu m\), a flow rate of 1 – several hundred \(\mu l\ min^{-1}\), at atmospheric pressure \([148]\). A high electric potential is applied to the capillary (2 – 5 kV) which leads to the accumulation of charge at the tip of the capillary, subsequent formation of a Taylor cone and release of solvent droplets containing charged analyte molecules \([149]\). These initial droplets have a radius in the \(\mu m\) range \([150]\). Evaporation of solvent from these droplets causes them to decrease in size and because charge is being retained, the charge per unit volume increases. This causes the droplet to be unstable and deform under the strong electric field; the accumulation of charge causes the droplet to elongate, in a similar fashion to the probe tip, and produce a new Taylor cone, resulting in droplet fission and the production of smaller droplets. Repeated evaporation/ fission events provide the final generation ESI droplets which have a radius of just a few \(nm\). It is from these
highly charged nanodroplets that gaseous ions are produced, which are detected in an MS experiment [151-153]. These nanodroplets are guided into the mass spectrometer via decreasing voltages and decreasing pressure regions.

Figure 1.5  Schematic representation of positive mode ESI showing Taylor cone formation, production of multiply charged droplets, droplet fission, desolvation and entry of the ions into the mass spectrometer.

There is still much debate surrounding the process by which ions are transferred from these nanodroplets into the gas phase. There are three widely accepted molecules for the mechanism (Figure 1.6). The ion ejection model (IEM) is thought to apply to low molecular mass species and small organic molecules [154]. It proposes that the ejection of a charged analyte ion could be ejected from the Rayleigh-charged nanodroplet by the electric field.

The charged residue mechanism (CRM) [145] is widely accepted as the method for the release of globular proteins into the gas phase. Here, Rayleigh-charged nanodroplets which contain a single analyte evaporate to dryness. They lose charge as the droplet shrinks, via the ejection of protons and small ions, which allows the droplet to remain close to the Rayleigh limit as the size decreases. Remaining protons are transferred to the protein during the final stages of desolvation.

It has been shown by MD simulations that unfolded and unstructured proteins are transferred from the droplet into the gas phase via the chain ejection mechanism (CEM) [150]. This is because of the more extended nature of such proteins, with
exposed hydrophobic groups making it more difficult for the protein to reside in the droplet interior. Such proteins therefore migrate to the surface. One terminus gets expelled into the vapour phase, followed by stepwise ejection of the remaining protein and separation from the droplet.

### 1.4.1.2 Nano-Electrospray Ionisation

Nano-electrospray ionisation (nESI) is an advancement of the ESI theme. The capillary tips have an opening of just a couple of $\mu$m rather than 100 $\mu$m for standard ESI. The flow rate is greatly reduced, down to $\sim 10$ nL min$^{-1}$. The benefits of nESI over conventional ESI include reduced sample consumption, lower concentrations, higher sensitivity and higher ionisation efficiency [155-157]. nESI produces smaller droplets [148, 158], resulting in lower non-specific aggregation, and because of fewer fission events, a more gentle desolvation process [159]. Because of the smaller size of the initial droplets from the tip, the final droplets will contain lower amounts of impurities and salts, resulting in a cleaner mass spectrum [159].
1.4.1.3 Matrix Assisted Laser Desorption Ionisation

At around the same time as ESI was introduced by Fenn, Karas and Hillenkamp proposed matrix assisted laser desorption ionisation (MALDI) [161-163] which is another soft ionisation method currently widely employed. Here, the analyte is dissolved in a solvent which contains small organic molecules which are chosen to absorb the laser wavelength. This is called the matrix. The matrix is then dried resulting in a ‘solid solution’ deposit of analyte-doped matrix crystals. A laser is then used to irradiate the sample, causing ablation and desorption of the sample and matrix material. Ionisation of the analyte occurs via protonation or deprotonation in
Chapter 1       Introduction

the hot plume of ablated gases, and the ion is then analysed in the mass spectrometer.

1.4.2 Mass Analysers

Once the ions have been produced they are separated according to their m/z value in a mass analyser, of which there are many different types. These can be used alone or in combination. Experiments within this thesis utilised quadrupole and time-of-flight (ToF) mass analysers, in combination (q-ToF mass spectrometers).

1.4.2.1 The Quadrupole Mass Analyser

The quadrupole mass analyser consists of four perfectly parallel rods which are either of circular (as in Figure 1.7) or hyperbolic section. Both a radio frequency (RF) and a direct current (DC) potential are applied; With respect to their RF field, opposite rods are in-phase, while adjacent rods are 180° out of phase. An oscillating electric field is therefore produced which draws the ions through the quadrupole and is described by Equation 1.2;

\[ \pm \phi_0 = \pm (U - V \cos \omega t) \quad \text{Equation 1.2} \]

Where \( \phi \) is the total potential on the rods, \( U \) is the DC potential, \( V \) is the zero-peak amplitude of the RF potential, \( \omega \) is the angular frequency of the RF potential and \( t \) is time. \( U \) will typically vary from 500 – 2000 and \( V \) from 0 – 3000V (-3000 V to +3000V peak to peak).

A positive ion entering the quadrupole will be attracted to a negative rod. The alternating RF potential causes the rods to continually switch polarity, causing the ion to travel along an oscillating path. This path is dependent upon the values of \( U \), \( V \) and \( \omega \). Scanning the RF amplitude (\( V \)) allows ions of a range of m/z values to be transmitted. However by fixing \( U \), \( V \) and \( \omega \) at values that only provide ions of a specific m/z value with a stable trajectory only the ion of interest will traverse the length of the quadrupole (resonant ions). All ions with differing m/z values will collide with the rods (non-resonant ions). It is therefore possible to use the quadrupole as a mass filter, transmitting only an ion of choice. This is property is often exploited during fragmentation studies to select the desired precursor.
Although the concept of time-of-flight (ToF) mass analysers was described in concept in 1946 [164], it wasn’t until 1955 that the design of a linear ToF instrument was published [165]. This later became the first available commercial instrument. During ToF analysis ion packets are accelerated into a drift tube via a potential difference, or pusher pulse. They then enter a field free region where their \( m/z \) ratio determines their velocity through the vacuum. Therefore, by measuring the time taken for the ion to reach the detector after being pulsed into the ToF, it is possible to measure the \( m/z \) ratio as shown in Equations 1.3 – 1.5 [164, 165].

Ions leave the source with mass \( m \) and total charge \( ze \). They are then accelerated by a voltage \( V_s \). The kinetic energy \( E_k \) of an ion can therefore be measured by Equation 1.3:

\[
E_k = \frac{mv^2}{2} = zeV_s
\]  

Equation 1.3

After initial acceleration the ion travels in a straight line to the detector, at constant velocity \( v \). The time \( t \) taken to traverse distance \( L \) to reach the detector is given by;
\[ t = \frac{L}{v} \] \hspace{1cm} \text{Equation 1.4}

Therefore;

\[ t^2 = \frac{m}{z} \left( \frac{L^2}{2eV_z} \right) \] \hspace{1cm} \text{Equation 1.5}

Equation 1.5 shows how the \( m/z \) value of an ion can be calculated from the measurement \( t^2 \).

Advantages of ToF analysers include high transmission efficiency, no theoretical upper limit of mass range and fast data acquisition. A main drawback of the linear ToF, however, was poor resolution due to several factors that introduced a distribution of flight times among ions with the same \( m/z \) ratio. These distributions are based on the time of the pulse, the volume of the ion packet and differences in the initial kinetic energy of the ions. The latter arises from the proximity of an ion to the acceleration voltage; closer ions will experience a greater potential difference. The use of a reflectron was introduced to overcome this.

A reflectron is a set of equally spaced electrodes which act as an 'ion mirror', reflecting ions back up the flight tube towards the detector (Figure 1.8). It corrects for differences of kinetic energy between ions of the same \( m/z \). Ions with higher kinetic energy and therefore more velocity will penetrate the reflectron further and spend more time there, and will arrive at the detector simultaneously with an ion with the same \( m/z \) that has lower kinetic energy. The resulting increase in resolution is partially due to the increased flight path, but mainly due to correction of differences in initial kinetic energy.
Chapter 1       Introduction

1.4.3 Detection

Once the ions have travelled through the mass spectrometer and have been separated in the analyser they reach a detector which transforms them into a useable signal. From the impact of an ion an electric current is produced that is proportional to their abundance which is then further amplified into a detectable current.

1.4.3.1 Photomultiplier and Electron Multiplier Detectors

Electron multiplier and photomultipliers are often coupled with quadrupole and ion trap analysers; in a q-ToF instrument if only the quadrupole is used the ions are detected by a photomultiplier at the entrance to the ToF. Here, the ions strike a conversion dynode which emits electrons. These electrons collide with a phosphor screen which results in the emission of photons, which pass into a photomultiplier tube where they produce an electric current which is then amplified.

Electron multiplier detectors works in a similar way to the photomultiplier; a series of dynodes are held at increasing potential. Collision of an ion with the first dynode produces electrons which are amplified by a cascade to produce a current.
1.4.3.2 The Microchannel Plate Detector

In ToF mass spectrometers microchannel plate (MCP) detectors are the most common. They are disc-shaped devices which contain thousands of tiny electron multiplier channels (Figure 1.9). Each channel is coated with a semiconductor substance to eject electrons in response to a collision by an ion, resulting in significant signal amplification, of up to $10^8$ if multiple MCPs are used. The cascade of electrons are then measured as current.

![Figure 1.9 Schematic of a cross section of an MCP detector from a side view. An incident ion begins a cascade of electrons which amplifies the signal.](image)

1.5 Ion Mobility Mass Spectrometry

Ion mobility mass spectrometry (IM-MS) is a hybrid method that allows separation of gas phase ions firstly according to their mobility, then on the basis of their $m/z$ ratio. Ion mobility measurements are based on the velocity of ions under a weak electric field on a millisecond timescale while they experience collisions with a buffer gas. The mobility of an ion is dependent upon its shape and charge; IM-MS experiments therefore separate ions according the shape of the ion, in terms of its rotationally averaged collision cross section (CCS). Three main types of ion mobility techniques can be coupled to mass spectrometry; drift time (DT), travelling wave (TWIMS) and field asymmetric ion mobility (FAIMS).


1.5.1 Drift Tube Ion Mobility Mass Spectrometry

Drift tube ion mobility spectrometry (DT-IMS) was the first type of ion mobility to be developed and requires the simplest instrumentation of the three main techniques. The first drift tube was developed by McDaniel et al. [166] which was based on previous investigations carried out at the beginning of the 20th century into the movement of ions in different gases [167, 168]. It is the only type of ion mobility which allows for direct calculation of CCS.

A DT-IMS experiment involves a chamber filled with an inert buffer gas (usually helium), at a known temperature and pressure, across which is applied a weak electric field. Ions are pulsed into the chamber in discrete packets and experience an electrostatic force which draws them through the cell. They are hindered, however, by collisions with the inert buffer gas. Two of the factors which influence how quickly the ions travel are the rotationally averaged collision cross section of the ion, which determines how frequently the collisions occur, and the charge present on it, which determines how quickly the ion is pulled through the drift cell. A larger ion will experience more collisions with the buffer gas and will hence be hindered to a greater extent and have a longer drift time than a smaller ion. Ion intensity is recorded as a function of time in a plot called an arrival time distribution (Figure 1.10). By measuring the drift time of an ion through the chamber, the rotationally averaged CCS can be calculated directly.

Figure 1.10  The influence of conformation on the arrival time of an ion. Ions are pulsed into the drift tube and are drawn through by the potential gradient. Ions with the same m/z but a different CCS are separated since they will interact differentially with the buffer gas.

The behaviour of an ion as it travels through the cell is dependent upon the ratio of the electric field applied across the cell \( E \) to the buffer gas number density \( N \). At high values of \( E/N \) ions align in the electric field and their mobility becomes
dependent upon \( E \). When \( E/N \) is sufficiently low the velocity \( (v_d) \) of ions is proportional to \( E \):

\[ v_d = K \cdot E \]  

**Equation 1.6**

The mobility \( (K) \) of an ion is dependent upon its mass, charge \( (z) \) and CCS \( (\Omega) \) and can be described by Equation 1.7;

\[ K = \frac{3z_e}{16N} \left( \frac{2\pi}{\mu k_B T} \right)^{1/2} \frac{1}{\Omega} \]  

**Equation 1.7**

Where \( \mu \) is the reduced mass of the analyte and buffer gas, \( k_B \) is the Boltzman constant and \( T \) is the temperature (in Kelvin). Low field ion mobility of an ion depends upon the pressure and temperature at which the measurements are recorded. It is therefore often reduced mobility \( (K_0) \) that is reported, since this accounts for temperature and pressure, as shown in Equation 1.8, and enables comparisons to be made.

\[ K_0 = K \left( \frac{T_0}{T} \right) \left( \frac{P_0}{P} \right) \]  

**Equation 1.8**

Normalised pressure \( P_0 \) and temperature \( T_0 \) are 760 Torr and 273.15 K respectively.

Collision cross sections are dependent upon the buffer gas used for the experiment [169, 170]. The resolution of a DT-IM-MS depends on the pressure of the buffer gas and the length of the drift cell since better separation can occur if the ions spend longer traversing the drift cell.

### 1.5.2 Travelling Wave Ion Mobility

The first commercially available integrated IM-MS device was the Synapt G1 [171] which was released in 2006 by Waters MS technologies. More recent versions (Synapt G2 in 2009 and G2S in 2012) have since been released which have provided users with enhanced resolution and sensitivity. Since the introduction of the Synapt, many more research groups have been enabled to use IM-MS.

Travelling wave ion mobility instrumentation utilises three stacked ring ion guides with RF applied to consecutive electrodes. A travelling wave consisting of a series of transient DC voltages is superimposed on top of the RF voltage to propel ions through the device. Upon injection into the cell, ions with the highest mobility are propelled by the travelling wave to a greater extent than ions with lower mobility.
which fall behind waves. The lower the mobility of an ion, the more waves it will fall behind and the longer it will take to reach the end of the cell.

Unlike DT-IM-MS, the direct calculation of a CCS from TWIMS data is not possible. A calibration must therefore be applied via the use of standards for which the CCS values have been measured by DT-IM-MS. While this approach has been effectively in a number of studies it has been reported that for conformational studies certain parameters must be carefully controlled to avoid thermal denaturation of the ions [172].

1.5.3 Field Asymmetric Waveform Ion Mobility Mass Spectrometry

High-field asymmetric waveform ion mobility spectrometry (FAIMS), also known as differential ion mobility, separates ions on the basis of their behaviour in both low-field and high-field electronic gradients, at atmospheric pressure [173, 174]. Ions travel between two parallel electrodes in a buffer gas which is flowing in the same direction as the ions. While one electrode is grounded, to the other an asymmetric waveform is applied, alternating between short high-voltages and long low-voltages of opposite polarity. The asymmetric wave is applied orthogonally to the flow of the gas resulting in the ions following oscillating paths. A compensating DC voltage is also applied which corrects the trajectory of the ions, therefore preventing them from colliding with the electrode. The compensation voltage that allows transmission of an ion depends on its ratio of high field to low field mobility, and it is this that allows separation of ions. The CCS of an ion cannot be determined from FAIMS due to the high field component of the technique. FAIMS devices are commercially available from Thermo Scientific (San Jose, USA), Ab Sciex (Concord, Canada) and Agilent technologies (Santa Clara, USA) in collaboration with Owlstone (Cambridge, UK).

1.6 Hydrogen-Deuterium Exchange Mass Spectrometry

Hydrogen-deuterium exchange-mass spectrometry (HDX-MS) has been used to investigate the dynamic properties of the proteins in solution, in order to relate the solution-phase and gas-phase dynamics. HDX-MS is a sensitive and rapid technique which allows localisation of protected areas within proteins; that is areas which are not accessible to the solvent due to secondary or tertiary structure [175]. HDX-MS was originally developed to study protein folding and unfolding events
where measurements of amide proton exchange rates were used to interpret the extent of hydrogen bonding [176]. Now, applications have been extended to the detection of disordered regions [177, 178], the indication of solvent accessibility-changes in proteins upon during coupled folding and binding [179, 180] and for the characterisation of aggregates and oligomers [181]. A key advantage of HDX-MS is that there is no upper limit on the size of the protein that can be analysed.

During an HDX experiment the protein of interest is diluted into deuterated buffer, and the exchange of solvent deuterium atoms is allowed to proceed for a given amount of time. The reaction is then quenched by the addition of a low pH buffer which reduces the back-exchange of deuterated backbone amides. The protein is digested by an acid-stable protease, in this case pepsin, and the HDX extent is analysed by mass spectrometry of the protein fragments. Because IDPs contain few protected amide hydrogens due to the lack of tertiary structure, peptides from such proteins are expected to be fully saturated with deuterium at the earliest on-exchange time point.

1.7 Computational Methods to Predict Collision Cross Sections

The mobility of an ion is inversely proportional to its CCS, which allows the mobility and buffer gas specific CCSs to be calculated from a set of co-ordinates. Theoretical CCSs can be calculated using molecular co-ordinates obtained from techniques such as NMR, X-ray crystallography and molecular dynamics simulations. Such calculations involve the evaluation of rotationally averaged CCSs with an approximate but appropriate treatment of ion-buffer gas collisions. These calculated CCSs can then be compared with experimentally derived CCSs from IM-MS to match the data with tentative low-energy structures.

MOBCAL [182-184], developed by Shvartsburg and Jarrold, is widely used for CCS calculations of biological molecules, and can do so in three different ways; by using the projection approximation (PA), the exact hard sphere scattering (EHSS) method or the trajectory method (TM). Recently, a new method for theoretical CCS calculation, known as the projection superposition approximation (PSA) method, has also been developed.
1.7.1 The Projection Approximation

The PA equates the CCS with the orientationally averaged projection of the molecule; it essentially measures the average shadow as the ion is rotated through all possible orientations. Each atom within the ion is modelled as a hard sphere and the ion-neutral collisions are modelled as hard sphere collisions. No scattering effect is considered as the method is unable to factor for multiple collisions between the ion and the buffer gas. Additionally, as the 3D ion is converted into 2D surface details are lost since cavities within the molecules are not well represented. In this method the CCS is therefore underestimated, and deviations increase with the size and complexity of the ion. An improved version of the PA implemented in the sigma program uses a 12-6-4 Lennard Jones potential to account for the ion-neutral interaction effects. This method was trained on molecules of 500-1000 Da; it therefore performs well for systems of this size, especially those with rounded, convex shapes. For smaller molecules the CCS will be overestimated by this method, while for large molecules with cavities, CCS will be underestimated since multiple scattering events are important for these systems.

1.7.2 Projection Superposition Approximation

The PSA measures the projection of the ion as in the PA, but it is improved by the inclusion of size and shape effects. These are accounted for by deepening the well of the interaction potential by the increasing number of atoms and the surface curvature of the polyatomic ions. The average He- C60+ interaction is approximately eight times as strong as that of a single He-C+ interaction [183]; this size effect can be described as the effect the size of the molecule has on long-range interactions with the buffer gas. The shape effect is related to the curvature of the molecular geometry; the effect becomes significant for analytes with concave features.

1.7.3 The Exact Hard Sphere Scattering Method

The EHSS method is more accurate than the PA and remains computationally efficient. It assumes that both the individual atoms of the protein and the collision gas behave as hard spheres. It accounts for multiple scattering events, accounts for the surface of the ion and therefore considers any cavities that determine the CCS. This method performs best for large systems, where the effect of the ion-neutral interaction is not as great as multiple scattering. It has also been found that
it may occasionally overestimate the effect of multiple scattering, resulting in larger values than the TM.

### 1.7.4 The Trajectory Method

This is the most accurate method for calculating the CCS of a protein but is very computationally expensive, especially for large biomolecules. Here, the ion itself is treated as a collection of atoms, each of which is represented by a Leonard-Jones (12-6-4) potential. These individual potentials are summed to obtain the effective potential of the ion. Multiple collisions between the neutral buffer gas and the ion are simulated using an empirical potential for the ion-neutral interaction potential. Trajectories are then run at this potential in order to obtain the scattering angle (the angle between the incoming and outgoing buffer gas atom trajectory) and this is averaged over all possible collision geometries.

### 1.8 References

Chapter 1  Introduction

Chapter 1  Introduction

Chapter 1       Introduction

Chapter 1  Introduction

2

Mass Spectrometry Methods for Intrinsically Disordered Proteins
2.0 Declaration


This article has been reproduced in an unchanged format except for minor adjustments to incorporate them in to this thesis.

As first author on this publication I produced the first draft of the manuscript with assistance from QC, and I carried out subsequent editing.
Mass spectrometry methods for intrinsically disordered proteins

Rebecca Beveridge\textsuperscript{a}, Quentin Chappuis\textsuperscript{a}, Cait MacPhee\textsuperscript{b} and Perdita Barran\textsuperscript{*}

\textsuperscript{a}School of Chemistry, University of Edinburgh, West Mains Road, Edinburgh EH9 3JJ, Scotland. E-mail: perdita.barran@ed.ac.uk

\textsuperscript{b}School of Physics and Astronomy, University of Edinburgh, West Mains Road, Edinburgh EH9 3JJ, Scotland

Received 21st May 2012, Accepted 21st September 2012

First published on the web 25th September 2012

2.1 Abstract

In the last ten years mass spectrometry has emerged as a powerful biophysical technique capable of providing unique insights into the structure and dynamics of proteins. Part of this explosion in use involves investigations of the most recently ‘discovered’ subset of proteins: the so-called ‘Intrinsically Disordered’ or ‘Natively Unstructured’ proteins. A key advantage of the use of mass spectrometry to study intrinsically disordered proteins (IDPs) is its ability to test biophysical assertions made about why they differ from structured proteins. For example, from the charge state distribution presented by a protein following nano-electrospray (n-ESI) it is possible to infer the range of conformations present in solution and hence the extent of disorder; n-ESI is highly sensitive to the degree of folding at the moment of transfer from the liquid to the gas phase. The combination of mass spectrometry with ion mobility (IM-MS) provides rotationally averaged collision cross-sections of molecular ions which can be correlated with conformation; this too can be applied to IDPs. Another feature which can be monitored by IM-MS is the tendency of disordered proteins to form amyloid fibrils, the protein aggregates involved in the onset of neurodegenerative diseases such as Parkinson’s and Alzheimer’s. IM-MS provides a useful insight into events that occur during the early stages of aggregation including delineating the structure of the monomer, identifying oligomer
distributions, and revealing mechanistic details of the aggregation process. Here we will review the use of MS and IM-MS to study IDPs using examples from our own and other laboratories.

### 2.2 Introduction

Mass spectrometry was originally applied to elucidate atomic isotopes and for most of the 20th century was principally used to obtain the molecular weight of small chemical compounds, as well as structural information via the use of dissociation. Due to developments in soft ionisation methods [1], it is now an increasingly popular technique with which to investigate conformations of intact biological molecules under a variety of conditions [2]. Whole protein mass spectrometry relies on the transfer of either a single protein or a protein complex, which may often contain several components, into the gas phase from a solution [3]. Several mass spectrometry based studies have addressed the process of protein folding and unfolding; both based on altering solution conditions and observing how this effects the distribution of charge states into the gas phase [4, 5] and also by trapping ions in the gas phase and observing the evolution of conformations over the time scale of mass spectrometry analysis (μs-s) [6, 7]. This ability to discern conformational flexibility, by the measurement of discrete conformers in a mixture, is one of the key strengths of mass spectrometry over other biophysical techniques [8-10]. Mass spectrometry can also be used to detect specific regions of disorder in proteins [11], measure the impact of metal ions on protein conformation [12] and monitor conformational changes that occur upon the formation of protein complexes [3].

Mass spectrometry is now positioned as an informative biophysical tool, which can be used along with other techniques to gain structural and functional information about biological molecules. Under particular scrutiny at present is the newly discovered subset of proteins, the intrinsically disordered proteins (IDPs) [13]. IDPs are polypeptide chains which exist and function without a well-defined three-dimensional structure [14]. They have increased flexibility, are more dynamic than folded globular proteins and can populate several heterogeneous conformations of similar energy [15, 16]. It follows that gathering information on the structures of IDP’s, often intractable by crystallography, is proving a great challenge requiring the development of new techniques [17, 18], or at least new ways to interpret data. This review considers the emerging role of mass spectrometry to interrogate IDPs.
The discovery of IDPs strongly contradicted the original protein paradigm which stated that the function of a protein is derived entirely from its folded 3-dimensional structure [19, 20]. Therefore, new definitions have been developed since. Dunker proposed the ‘protein trinity’ [17] which declares that proteins reside in one of the three states: the ordered state, the molten globule and the random coil. Dyson and Wright [21] expand this to a quartet model which is similar to the protein trinity except for the discrimination between two states within the ordered state into ‘mostly folded with localised disorder’ and ‘linked folded domains’ in which independently folded globular domains are separated by flexible linker regions. Two points should be stressed: firstly, while some proteins can be found in each of these different states, not all can. For example, many IDPs do not adopt folded, ordered states. Secondly, these states are not absolute limits and proteins actually populate a continuum containing these different states.

Analysis of the primary structure of IDPs reveals features that enable us to predict disordered regions and also give information about their behaviour. Tertiary folds in structured proteins are maintained by a high proportion of hydrophobic groups that interact favourably with each other inside the fold, away from solvent; by contrast charged and polar residues will most likely be located on the surface of the protein fold where they can favourably interact with solvent [22, 23]. Too many of these hydrophilic groups will destabilise any hydrophobic core and cause the protein to be unstructured or disordered. For these reasons, amino acids considered as ‘ordering’ are Val, Leu, Ile, Met, Phe, Trp, Tyr and those considered as a source of disorder are Gln, Ser, Pro, Glu, Lys, and, on occasion, Gly and Ala [23-25]. This categorisation enables a coarse prediction of disorder from primary sequence information alone. Common features identified across IDPs have been developed into predictors to determine if a given protein is likely to be structured or not, more than 50 of these predictors have so far been developed but they are by no means infallible [26], which suggests that disorder may not be encoded into the primary sequence, rather that it is a combination of intrinsic and extrinsic interactions.

IDPs in native conditions are not to be considered as ordered proteins in the denatured state; their hydrodynamic behaviour often differs. Most IDPs do not behave as random coils. Their lack of structure is not absolute and often they show a high degree of ‘compactness’ [17]. Transient elements of secondary structure (mainly α-helix) are observed and tend to reduce the hydrodynamic radius of IDPs [13]. Other non-covalent interactions, such as hydrogen bonds and electrostatic
interactions, also lead to dynamic tertiary structures [13]. Of course, such interactions are highly dependent on the environment and the presence of specific extrinsic species will have an effect on the ‘compactness’ of a protein [23].

An interesting and important characteristic of IDPs is their ability to bind to multiple partners, and this allows them to play a key role in many cellular signalling networks [27-30]. This can result in three different behaviours [27]:

1. An IDP binds to many different structured proteins [28].
2. Multiple IDPs bind to one structured protein [27].
3. Intrinsically disordered regions act as flexible linkers between ordered domains, allowing binding promiscuity [27, 28, 31].

Often, but not always, IDPs are induced to fold upon interaction with specific binding partners [32-34]. Upon binding to an ordered domain, the free energy required for the disorder to order transition is subtracted from the contact free energy, resulting in a highly specific interaction along with a low energy of association; such highly specific but weak interactions are perfect for signaling cascades as they ensure reversibility [17]. As well as binding to other proteins, some IDPs are also known to bind to small molecules which can have an effect on the structure of the protein.

Furthermore, even under denaturing conditions, the behaviour of IDPs differs from ordered proteins. Indeed, a frequent feature of IDP’s is their significantly low mobility via SDS-PAGE and gel filtration chromatography. For example, the apparent mass on SDS-PAGE of the repair protein xenopus XPA is 40–45 kDa, but its actual mass (measured by mass spectrometry and predicted from the amino acid sequence) is 30 922 Da [35]. This example and the reason for this reduced mobility will be discussed below.

IDPs are flexible and because they are unfolded – or partially unfolded – they have a high solvent accessible surface [13]. These two features make them more sensitive to proteolysis than globular proteins, since proteases require a protein to be unfolded over ten residues or more for cleavage to occur [17]. The amino acids in the inside of a globular protein are not accessible to proteases and it is mostly at external loops or at the termini that a protease can attack a globular protein. For IDPs, none or very few of the amino acids are buried in a protective core so there are no preferential cleavage sites beyond those that a given protease will select for [11].
This review will cover the different ways in which MS can be applied to the study of IDPs, the methods that MS has been coupled to, and examples of IDPs which have been analysed by these techniques.

### 2.3 ESI and charge state distribution

Since the development of electrospray ionisation [1] and matrix assisted laser desorption/ionisation (MALDI) [36] mass spectrometry has become a widely used method for the analysis of intact proteins with impacts on biochemistry, structural biology and medicine. These ‘soft’ ionisation methods allow proteins to be transferred intact into the gas phase without excessive fragmentation, and in the case of ESI, multiprotein complexes can also be maintained [37]. ESI is of particular importance in the study of IDPs because it is highly sensitive to the degree of disorder in solution [38]. Proteins are observed by ESI-MS in a range of charge states. In positive ionisation mode these charge states are due to protonated forms of the protein. Proteins with a lower degree of compactness have a wider charge state distribution since more ionisable sites are exposed to the solvent [39], and hence ESI can be used to distinguish between different conformations of the same protein [40-42] (Figure 2.1a). A noteworthy point is the importance of removal of the His-tag if one is employed for purification of a recombinant protein as this can create a second, higher charge state distribution which mimics that of a disordered protein. In addition we have often observed erroneous binding effects between His tagged proteins and their interaction targets in our laboratory and recommend against the use of His-tagged proteins for MS investigations. Other factors that must be considered when designing ESI-MS experiments to examine proteins are the effects of source and solvent conditions on the observed species and the source of protein material, which can at times be extremely variable – even from commercial sources – in our laboratory we take care to desalt protein samples and to measure the concentration using BCA assays or other spectroscopic methods. The storage of proteins is also important, since all proteins are susceptible to decomposition if subjected to repeated freeze–thaw cycles.
Figure 2.1 Schematic representations of procedures which involve the use of mass spectrometry. Electrospray ionisation charge state distribution analysis (a), MS-based proteomics (b), electron capture dissociation top-down sequencing (c) and ion mobility mass spectrometry (d). Components of image adapted from M. L. Gross et al. [42]

Frimpong et al. [40] used ESI-MS to probe the structure of monomeric α-synuclein, a 140 amino acid neuronal protein implicated in the onset of Parkinson's disease due to its propensity to form intracellular fibrillar aggregates [43]. Although classified as intrinsically disordered, α-synuclein has been found to populate four distinct conformations which coexist in solution, all which have different extents of disorder [40]. The deconvolution of charge state distributions obtained from α-synuclein following ESI from solutions with pH in the ranges 2.5–8 yields 4 basis functions, each of which was selected as a Gaussian curve (Figure 2.2) [40]. The most compact state spans charge states $5 \leq z \leq 10$ (corresponding to 5–10 extra protons), the compact intermediate spans charge states $7 \leq z \leq 13$, the more disordered intermediate spans charge states $10 \leq z \leq 17$ and the most unfolded conformer spans charge states $14 \leq z \leq 24$. The spectra are similar when the pH decreases from 8 through to 4, however at pH 2.5 the presence of the most
disordered state (U) becomes negligible and the presence of the compact state (C) and the α-helix containing intermediate is significantly increased. This shows that conditions which usually would be considered highly denaturing for globular proteins induce compaction of α-synuclein into a solution form(s) with lower surface accessibility. A proposed theory for this is the protonation of negative amino acids which would otherwise repel each other due to electrostatic forces [38]. Whether this hypothesis of collapse at low (or high) pH will extend to other IDPs remains to be investigated, but certainly the work of Kaltashov and co-workers shows how ESI-MS can offer detailed insight into the solution stability of disordered proteins.

Figure 2.2  The results of deconvolution of charge state distributions of α-synuclein ions in ESI-MS acquired in the pH range of 2.5–8. The four basis functions are assigned to the following putative states of the protein: U, unstructured; I₁, helix-rich intermediate; I₂, β-sheet-rich intermediate; and C, highly compact. Image taken from Frimpong et al. [40].

CD spectroscopy has previously shown that the α-helical content of α-synuclein can be increased by lowering the pH. By contrast, the presence of ethanol increases the amount of β-sheet [44]. Frimpong et al. [40] suggest that the two most compact
conformations, which are enhanced at low pH, may be characterised by α-helix structures and the two more unfolded conformations by β-sheets. The study of these semi-folded states is of particular importance because they are thought to be the intermediates in the aggregation which is responsible for amyloid diseases (Figure 2.3).

![Figure 2.3](image)

**Figure 2.3** The results of deconvolution of charge state distributions of α-synuclein ions in ESI MS in different concentrations of ethanol, which has been proven to increase the amount of β-sheet. The four basis functions are assigned to the following putative states of the protein: U, unstructured; I₁, helix-rich intermediate; I₂, β-sheet-rich intermediate; and C, highly compact. Image taken from Frimpong et al. [40]

Work by Natalello et al. [38] presents ESI-MS analysis of the same protein from identical solvent conditions (aqueous and 10 mM ammonium acetate) but in negative ionisation mode and employing nano-ESI rather than ESI. Surprisingly, the spectrum at pH 7.4 shows only three states. This discrepancy shows the limits (or subtleties) when using ESI-MS for conformational study: retention of protein conformation from solution to gas phase during ESI depends on many parameters which are yet not fully understood and can vary between instruments, and certainly as a function of solution, source and the polarity of the ionisation mode chosen (Figure 2.4).
**α-synuclein - A case study protein for the use of Mass Spectrometry to examine IDPs**

α-Synuclein (α-Syn) is a 140-amino acidic acid protein with a pI of 4.7 encoded by a single gene consisting of seven exons located in chromosome 4 [45]. It was first described by Maroteaux et al. in 1988 as a neuron-specific protein localised in the presynaptic nerve terminals and nucleus, giving rise to the name synuclein [46]. Although this protein is at the focus of extensive research efforts, its exact function is still unknown. α-Syn attracted significant interest in 1997 after a mutation in its gene was found to be associated with the familial cases of early-onset Parkinson's disease, [47] and its aggregates were found to be the major components of Lewy bodies, the hallmarks of PD [48]. The natively unfolded nature of α-Syn is characterised by its relatively low hydrophobicity and high net charge. It has been shown that alterations in the protein environment leading to an increase in its hydrophobicity and/or decrease in net charge can induce partial folding [49].

α-Syn belongs to a class of intrinsically disordered amyloid proteins that form fibrils by converting either all or part of the previously unstructured polypeptide into well-defined, β-sheet rich secondary structures. Other examples of these amyloids include islet amyloid polypeptide (IAPP), tau, and Aβ. The atomic structure of cross-β spines in α-Syn fibrils was first determined by X-ray crystallography and by X-ray diffraction of synthetic human synuclein filaments and filaments extracted from DLB and MSA brains [50]. These studies were later confirmed by X-ray diffraction of many short amyloid peptides, including a small segment from α-Syn [51, 52]. Together, these studies revealed that α-Syn fibrils are composed of several protofilaments containing a cross-β structure in which β-strands are arranged 15 in parallel, and the β-sheets are in-register with highly ordered amino acid side chain patterns exposed on the surface of the β-sheets.

In the past 10 years this protein has been the subject of several mass spectrometry based investigations including charge state distribution analysis, ion mobility mass spectrometry and HD exchange as detailed in the main text. The net charge at pH 7 of α-Syn is -9; it possesses 15 basic amino acids and 24 that are acidic. Due to this high number of charged side chains, α-Syn has been examined by mass spectrometry as deprotonated ions of the general form \([M-nH]^{nz}\) in negative ionisation mode [38] and as protonated ions \(([M+nH]^{nz+})\) in positive ionisation mode [40] and IMMS (negative mode) [53]. There are various agreements and some
subtle differences in the conclusions from the research carried out on this bio-
medically important protein.

Charge state distribution analysis of protonated species produced following ESI 
from solutions of pH 6 and 8, provides evidence for four conformational families 
[40]. These are assigned to a compact form (low charge states) two unfolding 
intermediates and an extended form (high charge states) (Figure 2.2). At low pH the 
compact family is favoured whereas the unfolded intermediate family is favoured at 
high pH and with the addition of ethanol. These observations are supported by 
Circular Dichroism (CD) which assigns secondary structure forming propensity to 
each solution condition. Experiments performed in the same way but on 
dero protonated species, in negative ionisation mode produced slightly different results 
[38]. When sprayed from neutral pH just three conformations could be fitted to the 
charge state distribution rather than four as seen in positive mode. Both groups 
explored the effect of adding alcohol to the solution (methanol or ethanol) and 
oberved a similar effect: an increase in abundance of the assigned β-sheet 
containing conformation.

The effects of hexafluoroisopropanol (HFiP) were also studied in negative mode 
[38]. Fluorinated alcohols enhance fibrillation at low concentrations (1-5%) yet exert 
an inhibitory effect at higher concentrations [44]. At 2.5% HFiP most of the protein is 
found as an intermediate, different from both the disordered and compact states that 
represent the two main components of the protein at pH 7.4 in the absence of 
cosolvents. The rest of the protein is found in a highly compact state. At 15% HFiP 
the unimodal, symmetric distribution suggests that soluble protein monomers 
populate a rather homogeneous conformational state. The main partially folded form 
stable by HFiP is distinguishable from the methanol-induced intermediate, 
characterised by lower compactness.

A further dimension was added to the investigations by negative mode IM-MS 
experiments [53]. As above, a very wide CSD at neutral pH and a narrower CSD 
when sprayed from acidic conditions was observed. Although no Gaussian curves 
are fitted to the mass spectra it can be inferred from the presented mass 
spectrometry data that compaction of the protein occurs at low pH. This is borne out 
by evaluation of rotationally averaged collision cross sections from IM analysis. 
These show two distinct conformations; charge states -6 to -8 are very compact 
structures and conformations with charge states of -9 and higher are much more 
extended with an increase in CCS of over 50%. The large difference in collision
cross sections between charges -8 and -9 agrees well with data from mass spectrometry since this is where the two Gaussian curves for the compact and disordered conformations meet.

CCSs of compact conformations are in good agreement with theoretical cross sections of the compact globular structure confirming these charge states are partially collapsed forms of the protein. The experimental cross sections in the -9 and higher charge states are much larger than the globular theoretical structure but smaller than the all-helical structure, consistent with substantially unfolded structures.

Vlad et al. [54] used IM-MS to identify a highly aggregating fragment formed from a cleavage between Val71 and Thr72. The aggregation of this carboxyl-terminal peptide has been shown to aggregate faster than full length α-Syn by ThT fluorescence assays, as well as produce more autoproteolytic fragments as aggregation proceeds. HDX-MS was also performed on both the full length structure and the fragment (α-Syn72-140). Full length α-Syn showed rapid exchange for 115 of the 134 backbone hydrogens, whereas 19 residues remained resistant to exchange for more than 14 days. 19 amino acids were also resistant to HDX in the α-Syn(72-140) fragment.

We can conclude from these investigations that mass spectrometry can be a powerful tool to examine the effect of solution conditions on the conformations adopted by this highly disordered protein, and that the intrinsic structures adopted (in vacuo) may be different depending on the protonation state of the protein. It is also likely that the high level of disorder possessed by α-Syn is indicative of a potential energy surface where the barriers between different conformation families are low and may easily be overcome by very subtle effects. This latter point may account for the discrepancies found in the mass spectrometry based studies performed to date and suggests that further work is necessary to fully characterise this chameleonic protein, in particular when relating data on the monomeric form to its role in neurological disease.

Figure 2.4  α-Synuclein – a case study protein for the use of mass spectrometry to examine IDPs.
2.4 Limited Proteolysis

Resistance to proteolysis correlates strongly with structural stability [55, 56] and limited proteolysis in solution coupled with mass spectrometry [11, 57, 58] has been shown to identify regions of reduced stability, for example domain borders [55] and linker regions [55]. Fragments resulting from partial proteolysis can be analysed by ESI-MS to allow identification of specific proteolytic peptides by comparison with a database of predicted cleavable regions. Because intrinsically disordered regions are more accessible to the protease and are therefore cleaved more frequently, it is possible to deduce from the fragments areas of structural stability and regions of disorder from the analysis of fragmentation patterns [11].

Using time resolved proteolysis coupled with ESI-FTICR, Iakoucheva et al. [11] identified disordered regions of Xeroderma pigmentosum group A (XPA). XPA is a protein involved in nucleotide excision repair; it is able to recognise damaged DNA albeit with the help of other ligands and to trigger the repair process through mechanisms which are yet unclear [59]. Trypsin cleavage, and lack of it, revealed that there are certain preferred trypsin cleavage sites and certain sites that are never cut. Disordered regions were revealed by partial proteolysis to be both termini of XPA, and a core fragment was found to be structurally stable. This core domain possesses 18 possible cleavage sites and no fragments cleaved at those sites were detected, which suggests that this domain is ordered while the two termini are disordered (Figure 2.5). This result is in close agreement with the PONDR disorder prediction, a neural network predictor originally developed from literature searches of intrinsically ordered and disordered regions in proteins [60]. The structured domain approximately corresponds to the minimal binding domain with DNA. We use the word ‘approximately’ because the limited proteolysis experiment was done on xenopus XPA and the minimal binding domain is only known for human XPA, but the comparison is relevant as the sequences of the two share 67% amino acids identity and 82% similarity.
Figure 2.5  Summary of cleavage site frequency. Each of the 48 cleavage sites is indicated on the x axis with the number of unique peptide fragments resulting from limited proteolysis on the y axis [11].

The same group of authors studied the aberrant mobility of XPA on SDS-PAGE and gel filtration chromatography [35]. XPA has an expected molecular weight of 30 922 Da while its apparent molecular weight is 40–45 kDa on SDS-PAGE and 92 kDa via gel filtration. Mass spectrometry measurements show that no post-translational modifications are responsible for those phenomena, and provide an exact mass of 30 922.02 Da which corresponds well with the sequence mass of 30 922.45 Da. Fragments of the protein also show low mobility but their respective deviations are less important. The structured domain is the one which is most mobile, with a deviation on SDS-PAGE of 15% while other fragments have a deviation of 30% and the full length protein has a deviation of 40%. This is consistent with the assertion that disorder results in lower mobility than expected for globular proteins. The authors of this paper concluded that this low mobility can be attributed to highly extended conformation(s) of all forms of xXPA [35].

2.5  Mass spectrometry based proteomics

The study of proteomics involves the determination of gene and cellular function directly at the protein level [61], and at the forefront of techniques available in this field of research as mass spectrometry [61]. Methods have been developed to examine specific subsets of proteins, for example, those containing post-translational modifications or those that change in response to specific stimuli, developmental events or during disease. The latter are of great importance to biomedical science because they may help to elucidate the mechanism of processes and could be extended to act as potential biomarkers.
There are five stages in a typical MS-based proteomics experiment: fractionation, digestion, chromatography, MS and finally MS/MS (Figure 2.1b). In the first step the proteins of the cell or tissue in question are purified, either by affinity selection or biochemical fractionation, and then further separated, often by SDS-PAGE, or liquid chromatography to define a smaller set of proteins for characterisation. These proteins must then be enzymatically digested as intact protein masses provide insufficient information for certain identification by MS. These peptides are then separated by liquid chromatography, the eluent sprayed directly into the mass spectrometer by electrospray ionisation and a spectrum recorded. A prioritised list of peptides for fragmentation can then be generated by the computer, determined by intensity, charge state and/or other sample specific information. Peptide ions are isolated, fragmented by a high-energy collision with gas and a MS/MS spectrum recorded. The spectra of fragmented peptides can then be compared against protein sequence databases for identification, and the proteins in question can also be identified.

Washburn et al. [62] used two-dimensional liquid chromatography (LC) coupled to tandem mass spectrometry (MS/MS) to characterise the proteome of yeast Saccharomyces cerevisiae. The method, which had been developed initially by Link et al., [63] involves filling a pulled microcapillary column with two independent chromatography phases and loading a complex peptide mixture which is eluted from the column directly into the mass spectrometer. The peptides and respective proteins were resolved which resulted in the identification of 1484 proteins from the S. cerevisiae proteome which included those with extremes in pi, molecular weight, abundance and hydrophobicity.

A review by Csizmok et al. [64] outlines several proteomic approaches for the identification of structural disorder in a complex mixture of proteins. Cortese et al. [65] exploit the resistance of IDPs to acid denaturation to enrich cell extracts with unfolded proteins. Although the reduction in the total amount of soluble Escherichia coli proteins was almost 100 000-fold after treatment by 9% PCA, 158 spots were observed on silver-stained 2-D SDS-PAGE gels. It was therefore suggested that resistance to acid denaturation by IDPs, as well as to other denaturation methods such as high temperatures, can be exploited to separate unstructured and structured proteins to study IDPs on a proteomic scale[65].

Galea et al. show that the heat treatment of NIH3T3 mouse fibroblast cell extracts at 98 °C also selects IDPs for proteomic analysis [66]. It is likely that resistance to
thermal aggregation is a result of the low mean hydrophobicity and high net charge characteristic of these proteins. The IDP enriched cell extracts were then separated by 2-D SDS-PAGE, excised from the gel and then digested with trypsin into smaller peptides for analysis by mass spectrometry. These peptides were then identified by comparison against sequence databases and classified according to their known subcellular location (cytoplasm, 38.1%; nucleus, 20.6%; mitochondria, 4.0%; cytoskeleton, 16.7%; extracellular matrix, 4.0%; ER, 7.1%) and further to their reported biological function. It was found that this heat treatment resulted in the enrichment of proteins involved in regulation and maintenance of cell structure and a corresponding depletion of metabolic proteins. Enriched to a lesser extent were proteins involved in cell signalling and protein folding as well as heat shock proteins. This demonstrates that exploiting the biophysical characteristic of IDPs (here their thermal stability) can be used in a MS based workflow.

2.6 Hydrogen-deuterium exchange

Hydrogen–deuterium exchange (HDX) experiments are frequently used to detect regions of disorder in proteins [67]. The technique exploits the increase in the rate of exchange of amide protons (N–H) with solvent protons in areas of reduced stability, which is due to a lack of protection from strong intramolecular H-bonds which occur in structured regions. Thus, the observation of the hydrogen-exchange process can provide valuable information on structural stability at many sites along the polypeptide chain [68]. HDX studies can be carried out in both directions. The labelling of a protonated protein with deuterium ('exchange in') is more widely used than the labelling of a deuterated protein with protons ('exchange out') as it eliminates the need for the initial step of complete deuteration [67]. A distinct benefit of HDX is its ability to report on the entire length of the polypeptide chain because every amino acid (with the exception of proline) has an amide proton. This gives it an advantage over techniques such as fluorescence emission spectroscopy, which only probe the structural environments of a select few chromophores.

Liquid chromatography followed by mass spectrometry of proteolytic fragments [69] or a top-down fragmentation MS approach [70] can be used to measure the extent of HDX in different regions of the polypeptide chain and hence provide information on disordered regions. The former approach has been used by Zhang and Smith [69] to elucidate disordered regions of horse heart cytochrome c, and the latter
approach has been used by Pan et al. [70] to distinguish between helices and loops in horse myoglobin.

Keppel et al. [71] used HDX in combination with pepsin digestion and mass spectrometry to investigate the disorder-to-order transition of IDPs that occurs upon the formation of a protein complex. The intrinsically disordered protein ACTR (activator of thyroid and retinoid receptors, NCOA3_HUMAN, residues 1018–1088) is known to bind to the molten globular protein CBP (the nuclear coactivator binding domain of the CREB binding protein, CBP_MOUSE, residues 2059–2117). The extent of HDX was investigated for each protein alone and in the complex. It was found that deuteration of the proteins in the complex was much slower than that of the individual proteins in isolation, indicating that the formation of the protein–protein complex confers structure to both of the participating polypeptide chains (Figure 2.6).

Figure 2.6 The disorder to order transition of ACTR and CBP upon complex formation, as determined by HDX [71].

2.7 Ion mobility mass spectrometry

Ion mobility is a measure of how quickly a gas phase ion moves through a buffer gas under the influence of an electric field. A typical Drift tube Ion Mobility Mass Spectrometry (DT IM-MS) experiment involves the injection of a pulse of ions into a chamber filled with a known gas at a known pressure across which is applied a weak static electric field (5–50 V cm$^{-1}$). Upon injection into the chamber the ions experience an electrostatic force pulling them through the cell. This force is countered by low energy collisions between the ions and the buffer gas. Two of the factors which influence how quickly the ions travel are the number of collisions which occur with the protein and the buffer gas which is determined by the size of the protein and the charge present on it which determines how quickly the ion is pulled through the drift cell [72, 73]. By measuring the drift time of an ion through the
cell, the rotationally averaged collision cross-section can be measured which provides valuable information on the conformation of a protein (Figure 2.1d) [74].

Ion mobility is frequently coupled with mass spectrometry which allows separation of protein conformers based on their mass-to-charge ratios as well as their interactions with the buffer gas. This has proven to be a powerful analytical tool which has been used in several instances to interrogate IDPs [12, 75-77].

Maurizio et al. [75] have used IM-MS to report on the High Mobility Group A (HMGA) proteins that are involved in an abundance of biological processes from transcription regulation to chromatin remodelling [78]. Through different mechanisms the HMGA proteins are also known to be involved in both benign and malignant neoplasias [79]. Based on the fact that the loss of the highly acidic C-terminal tail increases cell growth [80], ion mobility measurements of wildtype and C-terminal truncated HGMA2 were recorded. Despite the increase in mass of WT with respect to the C-terminal truncated HGMA2, the conformation of the former was more compact, indicating that the presence of the C-terminal tail is here responsible for further compacting HGMA2 and hence reducing oncogenic activity.

IM-MS has also been used to study the structure of the tumour suppressor protein p53. The p53 protein is known to bind to DNA, and this complex is stabilised by a single zinc ion which plays a regulatory role in the folding and DNA binding ability of p53 [81-83]. Removal of this zinc ion disrupts the structure of the DNA-binding domain, resulting in rapid cysteine oxidation and disulphide-linked aggregation. Faull et al. [12] used IM-MS to investigate the conformations of p53 with and without the presence of zinc.

In the presence of zinc, the mass spectrum has a charge state distribution from \(7 \leq z \leq 17\) with highest intensity species assigned to the \([M + 9H]^{9+}\) and \([M + 10H]^{10+}\). Their dominance can be associated with compact conformations being prevalent in solution as there are few solvent-accessible sites available for protonation. Two dimeric species are seen, and low intensity monomeric species between charge states \(11 \leq z \leq 16\) can be attributed to unfolded states in solution. Ion mobility data revealed that for seven of the ten charge states, at least two conformations are present, further reflecting the conformation flexibility of this system. Low charge states adopt compact conformations and as the charge increases, unfolding occurs.

When zinc has been removed, the charge state distribution alters, with strong signal from \(7 \leq z \leq 12\), and a dominant peak for \([M + 10H]^{10+}\), indicating that p53 has not
denatured extensively in the absence of zinc. For protein conformations that carry a low number of positive charges ($8 \leq z \leq 11$) ion mobility shows that the cross-sections are smaller than those for the equivalent charge for zinc present by 12.3\% for the large $[M + 10H]^{10+}$ conformer. The collision cross-section is small at low charge, with a large increase between $[M + 11H]^{11+}$ and $[M + 13H]^{13+}$ which corresponds to an unfolding transition. For $[M + 12H]^{12+}$ the arrival time distribution was very wide, indicative of a large number of conformations at the unfolding transition. The intensity of this $[M + 12H]^{12+}$ species in the mass spectrum is low, indicative of an ion that is not as stable as those either side of it.

The protein α-synuclein has also been examined using IM-MS, illustrating how different mass spectrometry based techniques can provide complementary and at times conflicting information. Bernstein et al. [53] used negative-ion IM-MS to decipher how the size of conformations of α-synuclein differed with charge state at initial solutions of neutral and acidic pH. Negative-ion mode was used because in pH 7 solution α-synuclein has an overall charge of −9. As shown by Frimpong et al. [40] the mass spectrum of the protein sprayed from a pH 2.5 solution shows a narrow charge state distribution at low charge states corresponding to a tightly folded protein, whereas when sprayed from pH 7 there is a wide charge state distribution with a maximum intensity peak at a higher charge. Arrival time distributions (ATDs) were reported for the $[M − 7H]^{7−}$ $[M − 8H]^{8−}$ and $[M − 9H]^{9−}$ species obtained from pH 2.5 solutions at several injection energies. The ATD of $[M − 7H]^{7−}$ at an injection voltage of 20 V has a narrow distribution characteristic of a single conformer, and a short arrival time characteristic of a compact structure. As the injection energy increases to 40 V the ATD becomes broader, indicative of some less compact isoforms, and by 100 V a narrow distribution at longer time is observed, representative of the annealing of the compact structures to a more extended conformation via substantial internal excitation which is more stable in the gas phase. This effect of thermally induced structural reorganisation has also been reported for structured proteins [84], but it may be that disordered proteins will respond differently to increased injection energy. This would be analogous to the observation in solution assays of sharp unfolding transitions for ordered proteins versus gradual unfolding for disordered proteins.

The $[M − 8H]^{8−}$ ion acts much like the $[M − 7H]^{7−}$, however the $[M − 9H]^{9−}$ charge state has only extended structures regardless of the injection energy, indicating that an extended structure is being sprayed from solution. Figure 2.7 shows the collision
cross-sections (CCSs) for species where $-6 \leq z \leq -11$ compared to those calculated by molecular modelling. The structures of charge states $-6 \leq z \leq -8$ are very compact, whereas for ions where $z > -9$ the structures are elongated. Between $-8$ and $-9$ a conformational rearrangement occurs which increases the CCS by over 50%. As the charge increases above $-9$ the CCSs also continue to increase, indicating that as more charges are added the structure continues to elongate. It can be seen that the CCSs of charge states $-6$, $-7$ and $-8$ are in good agreement with the theoretical cross-section of the compact globular structure confirming that these charge states are collapsed. The experimental CCSs of charge states $-9$ and higher lie between the globular and all-helical theoretical structures, in agreement with the hypothesis of substantially unfolded structures.
Figure 2.7  (i) Arrival time distributions for α-synuclein −7, −8 and −9 charge states at injection energies 20 V, 40 V and 90 V. (ii) CCS vs. charge for the dominant peaks in the ATD measurements. Theoretical CCSs are represented for globular and helical structures by dotted lines. Images taken from Bernstein et al. [53].

Vlad et al. also studied α-synuclein with IM-MS and identified a highly aggregating fragment formed from a cleavage between Val71 and Thr72. The aggregation of this carboxyl-terminal peptide has been shown to occur faster than full length α-synuclein by ThT fluorescence assays, as well as produce more autoproteolytic fragments as aggregation proceeds. HDX-MS was performed on both the full length structure and the fragment (α-Syn72-140). Full length α-Syn showed rapid
exchange for 115 of the 134 backbone hydrogens, with 19 residues remaining resistant to exchange for more than 14 days. 19 amino acids were also resistant to HDX in the α-Syn(72-140) fragment.

2.8 Conclusion

This review has covered many of the mass spectrometry-based techniques used in researching intrinsically disordered proteins. New developments in this research area are continuing to emerge and are likely to be of great importance given the role of IDPs in cell signalling and regulation, errors in which can result in the onset of cancer [85, 86]. Mass spectrometry is a useful tool in the elucidation of IDP structure, providing detail on the conformational spread of a given protein at the single conformer level rather than averaged data for all structures adopted by the polypeptide. Undoubtedly, mass spectrometry will be developed further and will continue to yield results in the elucidation of protein structure–function relationships.

2.9 Acknowledgements

The authors acknowledge the Schools of Chemistry and Physics at the University of Edinburgh for continuing support of their interdisciplinary work, and allowing them the space to develop new lines of research.

2.10 References

3

A Mass Spectrometry Based Framework to Define the Extent of Disorder in Proteins
Chapter 3; An MS-based framework to define disorder in proteins

3.0 Declaration


This article has been reproduced in an unchanged format except for minor adjustments to incorporate them in to this thesis.

As first author on this publication I compiled the data for figures 3.3 and 3.4, and I drafted and edited the manuscript. SC and JMDK set the initial scripts for the calculations of the CCSs of most extended and compact conformations of the protein which I then refined with PEB for the final submission. KJP collected the data for TTR, Avidin, Concanavalin A, SAP and IgG which was used in figure 3.3, allowing trends to be observed in the high mass range.
Chapter 3; An MS-based framework to define disorder in proteins

*Anal. Chem., 2014, 86 (22), pp 10979–10991. DOI: 10.1021/ac5027435*

**A Mass Spectrometry Based Framework to Define the Extent of Disorder in Proteins**

Rebecca Beveridge,¹ Sam Covill,² Kamila J. Pacholarz,² Jason M.D. Kalapothakis,²,³ Cait E. MacPhee,³ Perdita E. Barran¹

¹ Manchester Institute of Biotechnology, Michael Barber Centre for Collaborative Mass Spectrometry, 131 Princess Street, University of Manchester, Manchester. M1 7DN.

E-mail: perdita.barran@ed.ac.uk

² School of Chemistry, Joseph Black Building, West Mains Road, Edinburgh. EH9 3JJ

³ School of Physics and Astronomy, University of Edinburgh, James Clerk Maxwell Building, Mayfield Road, Edinburgh. EH9 3JZ.

Received 23rd July 2014, Accepted 3rd October 2014
First published on the web 3rd October 2014

**Supporting Information**

Protein sequences, mass spectra, arrival time distributions and tables of experimental collision cross sections obtained from DT IM-MS experiments, and an extension of the Δz versus molecular mass relationship for other proteins can be found in the Supporting Information. This material is available free of charge via the Internet at [http://pubs.acs.org](http://pubs.acs.org) and in appendix 1 of this thesis.

The authors declare no competing financial interest.
3.1 Abstract

Figure 3.1 Graphical abstract

In the past decade, mass spectrometry (MS) coupled with electrospray ionisation (ESI) has been applied extensively to the study of intact proteins and their complexes, often without the requirement of labels. Solvent conditions (for example, pH, ionic strength, and concentration) affect the observed desolvated species; the ease of altering such extrinsic factors renders ESI-MS an appropriate method by which to consider the range of conformational states that proteins may occupy, including natively folded, disordered and amyloid. Rotationally averaged collision cross sections of the ionised forms of proteins, provided by the combination of mass spectrometry and ion mobility (IM-MS), are also instructive in exploring conformational landscapes in the absence of solvent. Here, we ask the following question: “If the only technique you had was ESI-IM-MS, what information would it provide on the structural preferences of an unknown protein?” We have selected 20 different proteins, both monomeric and multimeric, ranging in mass from 2846 Da (melittin) to 150 kDa (Immunoglobulin G), and we consider how they are presented to a mass spectrometer under different solvent conditions. Mass spectrometry allows us to distinguish which of these proteins are structured (melittin, human beta defensin 1, truncated human lymphotactin, Cytochrome C, holo haemoglobin-α, ovalbumin, human transthyretin, avidin, bovine serum albumin, concanavalin, human serum amyloid protein, and Immunoglobulin G) from those that contain at least some regions of disorder (human lymphotactin, N-terminal p53, α-Synuclein, N-terminal MDM2, and p53 DNA binding domain) or denatured due to solvent conditions (ubiquitin, apo haemoglobin-α, apo haemoglobin-β) by considering two experimental parameters: the range of charge states occupied by the protein (Δz) and the range of collision cross sections in which the protein is observed (ΔCCS). We also provide a simple model to predict the difference between the collision cross sections of the most compact and the most extended
form of a given protein, based on the volume of the amino acids it contains. We compare these calculated parameters with experimental values. In addition, we consider the occupancy of conformations based on the intensities of ions in the mass spectra. This allows us to predict the potential energy landscape of each protein qualitatively. Our empirical approach to assess order or disorder is shown to be more accurate than the use of charge hydropathy plots, which are frequently used to predict disorder, and could provide an initial route to characterisation. Finally, we present an ESI-IM-MS methodology to determine if a given protein is structured or disordered.

3.2 Introduction

The organisation of polypeptide chains into three-dimensional (3D) shapes has been investigated for many years. An early view of protein structure, described as the lock and key model, was proposed by Emil Fischer [1, 2] in which exact stereochemical recognition was deemed essential for protein–protein or protein–ligand interactions. This purported that every protein possessed one energy-minimised structure, which was essential for its biological function. This model agreed with emerging evidence of organised 3D structures from crystallography [1, 3, 4]. Karush was among the earliest to critique this structure function paradigm [5], proposing instead a theory of configurational adaptability, which was also supported by Koshland [6], who called the concept “the induced-fit theory” [1, 6].

As the number of protein structures solved by X-ray crystallography increased, it became apparent that the resolution could vary across the unit cell [1], and this was attributed to localised disorder. The concept that proteins can be functional and yet lack a resolvable structure has gained momentum in recent years. It is now accepted that there is a progression of structure in functional proteins from those that are highly structured, perhaps possessing few flexible regions, to those that are completely disordered. Here, a protein, or region of a protein, is considered to be disordered if it lacks a stable structure, on the time scale of a nuclear magnetic resonance (NMR) experiment [7-11], and displays flexible, interconverting, coil-like conformations [11]. The location of a given protein in this structure progression appears to be associated with function; highly structured proteins are found to be involved in biosynthesis and transport [10], whereas disordered proteins are involved in regulation of transcription and translation, cellular signal transduction, protein phosphorylation, storage of small molecules, and regulation of the self-
assembly of large multiprotein complexes such as flagella and ribosomes [8, 11-14]. It has been shown that proteins arising from chromosomal translocations are enriched in higher levels of disorder, which allows the fused proteins to avoid removal by cellular surveillance mechanisms, and such fusion proteins are, in turn, associated with disease [9] and implicated in many neurodegenerative conditions [15, 16].

Certain amino acids have been suggested to be more disorder-inducing than others; intrinsically disordered proteins (IDPs) and intrinsically disordered regions (IDRs) are enriched in P, E, K, S and Q, and depleted in W, Y, F, C, I, L and N, compared to the average folded protein [7, 17, 18]. This low hydrophobicity is thought to deter the formation of a hydrophobic cluster; the high net charge, in turn, induces electrostatic repulsion, favouring extended conformations, while proline is a known breaker of secondary structural elements [19].

Different methods must be used to purify and characterise IDPs than traditionally applied to more structured systems. Unfolded or disordered proteins can be aggregation prone, structurally dynamic [20] and susceptible to proteolysis [21] and, as already stated, they are incompatible with high resolution in crystallography. In NMR experiments, full assignment of unfolded and partially folded proteins is possible [22, 23]. NOEs can be used to characterise the structure and dynamics of proteins, while coupling constraints and amide temperature coefficients can show regions of secondary structure [22]. Regions of disorder are signified in HSQC spectra by low dispersion in the proton dimension, while retaining high dispersion in the $^{15}$N dimension, because of the sequence dependence of the nitrogen chemical shift [24, 25]. However, NMR is limited by the molecular mass of a protein, it has a low tolerance to paramagnetic ligands and the high protein concentration that is required can lead to aggregation [26] as well as requiring milligram-level amounts of protein, which may be hard to obtain.

Since the development of electrospray ionisation by Fenn and co-workers [27], by which large proteins can be transferred from solution into the gas phase as multiply charged ions, mass spectrometry has been used to monitor protein structure [28-31]. The technique that transfers proteins as gently as possible from aqueous solutions in the presence of additives that stabilise structures (for example, volatile salts) to the gas phase for analysis is known as native MS. It is fast and sensitive and can provide mass and compositional information on several species in one spectrum simultaneously [32]. Native MS tends to utilise nanoelectrospray ionisation
Chapter 3; An MS-based framework to define disorder in proteins

(n-ESI) with low sample consumption (flow rates on the order of nanoliters per minute (nL/min)). The n-ESI spray droplets are smaller than those obtained with ESI, and the methodology requires lower spray potentials and source temperatures, which help to maintain native-like interactions, along with collisional cooling as the proteins enter the mass spectrometer, although the latter also applies to standard ESI [33, 34].

A plethora of studies provides compelling evidence that the stoichiometry of protein–protein or protein–ligand complexes can be preserved from solution. This supports the view that aspects of solution topology can be retained following ionisation [35, 36]; although there is also evidence for collapsed and elongated states, these, of course, are not mutually exclusive. ESI presents proteins to the gas phase with a distribution of protons, and MS separates these because they have different charges. Desolvation gives rise to charge-state distributions which indicates different solution conformations [37, 38] each with a somewhat different presentation of chargeable sites during solvent and salt removal. Protons will preferentially remain on or migrate to solvent-accessible sites of high proton affinity, and they will be removed from sites of low proton affinity [35]. It is for this reason that ordered proteins tend to occupy lower charge states; their tertiary fold will include a hydrophobic core, which can remain stable following ionisation and, hence, less residues are solvent-accessible. Conversely, disordered proteins are less hydrophobic and often have an excess of charged sites [35]; hence, they exist in solution in extended forms, resulting in more solvent-accessible protonation/deprotonation sites.

A crystallisable form of a protein that will diffract is required for X-ray crystallographic studies and NMR is unable to resolve multiple conformers in a solution; in contrast, ESI-MS is able to distinguish all adopted conformations of a protein, perhaps with a bias to those that are present with higher charged states [39]. Although seemingly simple information, the extent of charging of monomeric protein ions holds information about its conformation in solution [26]. Through comparison to other biophysical techniques, it is known that solution conditions that preserve structure in a structured protein will tend to show only low charge-state values (high m/z values) with a relatively narrow distribution of such peaks, whereas intrinsically disordered, or denatured, proteins present with a wider range of charge states distributed about the mode [40, 41]. This was demonstrated by Kaltashov et al. for the protein myoglobin by fitting Gaussian functions to the envelopes of the charge states as a linear addition of basis functions [38]. They observed that the
basis functions (specific to each protein) do not change greatly under different solvent conditions, except in intensity. They concluded that each Gaussian envelope corresponds to one main conformation and the width of these envelopes reflects the degree of heterogeneity of that conformation, with wider envelopes corresponding to greater heterogeneity. It still remains to be seen whether any physical parameters concerning the proteins in solution can be derived from such Gaussian fitting functions, but, at a minimum, it provides an intuitive relationship between charge-state distributions (CSDs) in the gas phase and the conformations present in solution.

Both intrinsic order and disorder will influence the CSDs of proteins, as will induced order and disorder through changing solvent conditions. Acids and methanol will tend to unfold and denature proteins, allowing increased charging during ionisation. Ethanol can induce beta-sheets [42] and its addition to aqueous solutions of α-Synuclein prior to ESI alters the ensuing CSD [40]. The choice of counterions and ionic strength will also have an effect, although many traditional biological buffers are not compatible with MS analysis. A salt used commonly to prepare aqueous solutions for ESI-MS of proteins is ammonium acetate, which, as well as being volatile, has a tendency to produce a more compact CSD, attributed to a “salt-crowding” effect [43] (similar to “salting in”).

Ion mobility is complementary to mass spectrometry, and they are both employed here as the hybrid technique (IM-MS). This provides information on the shape of the protein, in the form of rotationally averaged collision cross sections (CCSs), in addition to m/z ratios [44]. A typical mobility experiment involves pulsing ions into a drift tube filled with a buffer gas, at a known pressure and temperature, across which a weak electric field is applied. As the ions are pulsed into the chamber, they are drawn through by the electric field. They are hindered by collisions with the buffer gas, which slow the ions. Larger ions will collide more frequently with the buffer gas; hence, they will be slowed to a greater degree than more-compact ions. The arrival time of the protein ions will therefore depend on the size of the ion, and the charge present upon it: proteins with a higher charge will be drawn through the cell faster by the electric field. For any given charge state there may be multiple conformers which can be separated by their arrival times. Smaller arrival times correspond to more-compact conformers and larger arrival times correspond to more-extended conformers.
Mass spectrometry and ion mobility experiments can provide us with a wealth of information on protein structure. Whether a protein exists in one structural family or several interconverting transient conformations, IM-MS is a valuable technique that can be utilised as a route to structural characterisation, and it is applicable to proteins up to the megadalton (MDa) range [45]. Here, we present a framework derived from mass spectrometry experiments that can be applied to any protein to gain information on the degree of folding and dynamics, as an initial route to characterisation. During this process, we also fit parameters to be used during data analysis. Although it is widely assumed that IDPs present a wide CSD and structured proteins present a narrow CSD, and, similarly, that IDPs will present a wide range of CCSs, compared to structured proteins, no comprehensive dataset has yet been tested to validate this assumption. We set out to do this here.

### 3.2 Materials and Methods

#### 3.2.1 Materials

The sequences of all proteins used are found in the Supporting Information Table S1 (appendix 1), along with the ordering information (where possible). Human Cytochrome C (C3483), bovine β-casein (C6905), bovine serum albumin (A9418), hen egg white Lysozyme (L6876), hen egg white ovalbumin (A7641), egg white avidin (A9275), Canavalia ensiformis concanavalin A (C2010), and equine heart myoglobin (M1882) were purchased from Sigma–Aldrich. Human transthyretin was purchased from SCIPAC, U.K., human serum amyloid P component was purchased from CalBioChem Germany. Human lymphotactin residues 1–72 and full-length human lymphotactin were provided as a kind gift from Brian Volkmann, Medical College of Wisconsin [46]; N-terminal-MDM2 was a gift from Ted Hupp, University of Edinburgh; human α-Synuclein was a gift from Tilo Kunath, University of Edinburgh; N-terminal human p53 was a gift from Galina Selivanova, The Karolinska Institute; and Immunoglobulin G was a gift from UCB Pharma [47]. Ammonium acetate, liquid chromatography-mass spectrometry (LC-MS)-grade water, and LC-MS-grade methanol were all purchased from Fisher Scientific.

#### 3.2.2 Mass Spectrometry

Mass spectrometry (MS) experiments were carried out on a Q-ToF Ultima instrument, unless otherwise stated. The details of each experiment are found in
previously published work [48-53]. Data were analysed using modified Masslynx software (version 4.1), Microsoft excel 2010 and Originlab 9.0. Ions were produced by positive nanoelectrospray ionisation (Z-spray source) within a spray voltage range of 1.6–1.7 kV, a cone voltage of 45–80 V, and a source temperature of 80 °C. Nanospray tips were prepared in-house from borosilicate glass capillaries (Kwik-Fil, World Precision Instruments, Inc., Sarasota, FL, USA) using a Flaming/Brown micropipette puller (Model P-97, Sutter Instrument Co., Novato, CA, USA). Tips were filled with 10–15 μL of sample using gel loading tips (Eppendorf, Hamburg, Germany).

3.2.3 Ion Mobility

Instrument operation and measurement acquisition methods for each system examined have been described in detail elsewhere [48-53]. Briefly, ion mobility experiments were performed at a helium pressure of ~3.5–4.0 Torr and a cell temperature of ~300 K. These were measured carefully during each experiment. Drift times were recorded at six drift voltages between 60 V and 15 V. All mobility data were obtained from plots of arrival time versus drift voltage over this range. Experiments were performed in triplicate on separate days. Nanoelectrospray ionisation produces a constant beam of ions that must be trapped to allow discrete ion packets to be pulsed into the drift cell. This is achieved by raising the DC voltage on the top hat lens element directly before the entrance elements to the cell, trapping the ions in the precell hexapole. The trapping DC is then lowered for 40 μs through the use of a pulse generator (Stanford Research Systems, Sunnyvale, CA, USA) allowing a pulse of ions into the drift cell. Ions are focused into the drift cell using a three-element Einzel lens, separated by mobility within the cell and then accelerated through the quadrupole and ToF mass analysers to a MCP detector. A minimum of 2000 scans was collected for each drift voltage. Data were processed according to published procedures [54] to determine CCSs from measured arrival times over a range of drift voltages.

3.2.4 Fitting Procedures

The data of Figure 3.3a (structured proteins), presented later in this work, were fitted by a power function \( y = ax^b \), with \( b \) fixed to a value of 0.5. The rest of the data were subject to linear fitting. For fits to \( \Delta z \) and \( \Delta CCS \) for structured proteins, the slope was fixed to zero (0).
3.2.5 Charge-Hydropathy Analysis

Predictions of disorder in six proteins were analysed using four different prediction methods. Protein sequences (found in the Supporting Information, appendix 1) were submitted to the PONDR and DISOPRED Web servers. For PONDR, they were analysed using PONDR-VLXT, PONDR-VSL2, and PONDR-VL3. Access to PONDR was provided by Molecular Kinetics (Indianapolis, IN, USA; E-mail: main@molecularkinetics.com). [Note that VL-XT is copyrighted by the WSU Research Foundation (1999, all rights reserved). PONDR is copyrighted by Molecular Kinetics (2004, all rights reserved). For DISOPRED, we used DISOPRED3 available from the PSIPRED Protein Sequence Analysis Workbench at University College London (U.K.), http://bioinf.cs.ucl.ac.uk/psipred/ [55]]

We compare the charge and hydropathy indices with experimental data from charge state and collision cross section analysis, respectively.

3.2.6 Theoretical Modelling of Collision Cross Sections

The lower bound of the collision cross section of a protein was calculated by assuming that, when folded, the protein approximates a sphere in shape. This would describe a globular highly compact form of a protein. The value of protein density used here was \( \rho = 0.904 \text{ Da/Å}^3 \), which has been suggested as being most relevant for small globular proteins [56]. Using the molecular weight \( M_w \) of the protein, the volume of the protein sphere can be calculated via \( V = \frac{M_w}{\rho} \). The radius of the sphere is therefore \( r = \left(\frac{3V}{4\pi}\right)^{1/3} \). The collision cross section of a sphere of this radius is given by Equation 3.1:

\[
\text{CCS}_{\text{lower}} \left(\text{Å}^2\right) = \pi r^2 = \pi \left(\frac{3V}{4\pi}\right)^{2/3} \tag{Equation 3.1}
\]

This gives the geometric size of the sphere of the protein. However, the CCS of the molecules inferred from their electrical mobility in helium always exceeds this value [57, 58]. Work by De la Mora on a series of monomeric proteins demonstrates that the experimental CCS is approximately 1.19 times that of the theoretical CCS based on mass [59]. We therefore multiply the answer from equation 1 by the scaling factor 1.19 to predict the smallest possible measured collision cross section.

Conversely, the upper bound on the collision cross section of a protein can be calculated knowing that the furthest distance between \( \alpha \)-carbons in a protein chain is 3.63 Å [60]. Therefore for a polypeptide of \( n \) residues, the maximum linear
Chapter 3; An MS-based framework to define disorder in proteins

dimension is $n(3.63) \ \text{Å}$ [60]. A completely unfolded protein which has been stretched out from end to end can be modelled as a cylinder of length, $l = 3.63n \ \text{Å}$, and radius given by the geometric average of the sum of the radii of the amino acids contained in the protein’s sequence. The average volume of an amino acid in a protein’s sequence is given by Equation 3.2:

$$V = \frac{\sum_{i \text{ amino acids}} V_i N_i}{n}$$  \hspace{1cm} \text{Equation 3.2}

Here, the sum is over all amino acids $i$, $V_i$ is the volume of the $i$th amino acid, and $N_i$ is the number of amino acids of type $i$ in the protein sequence. The average amino acid radius is then approximated by $r = \left(\frac{V_l}{\pi h}\right)^{1/2}$, where $h$ is the height ($h = 3.63 \ \text{Å}$). This forms the radius of the fully extended protein cylinder. The CCS is then given by the rotationally averaged CCS of this cylinder, since the protein “tumbles” in the drift tube of the apparatus, because of the low electric field.

The rotationally averaged CCS is given by the projection area of the cylinder:

$$CCS_{\text{upper}}(\text{Å}^2) = (\text{Projection area of cylinder}) = \frac{4}{\pi} rl + 2r^2$$  \hspace{1cm} \text{Equation 3.3}

This figure is then multiplied by the helium scaling factor of 1.19 to give the final answer of the largest possible CCS.

These theoretical values are highly approximate and do not take into consideration disulphide bridges, proline residues, or other noncovalent interactions or restrictions, but they do serve as upper and lower boundaries with which to compare experimental data.

3.2.7 Determining the Population of Conformations from Charge State Distribution (CSD) and Collision Cross Section (CCS) Data

For any given CSD, the relative intensities of ions in the mass spectrum have been normalised against the base peak. These data are then plotted against the average CCS value for each charge state to provide the gas-phase conformational occupancy for each protein studied. The response factors of the mass spectrometer for any given charge and conformational state are considered equivalent.
3.3 Results

3.3.1 Nano-electrospray Ionisation (n-ESI) Ion Mobility–Mass Spectrometry (IM-MS) Signatures for a Structured Protein versus a Disordered Protein

We first illustrate the behaviour of two exemplar proteins under native ESI-MS solution conditions where (i) Cytochrome C (12 229 Da) is known to be folded [61] and (ii) β-casein (23 980 Da) is known to be unstructured [62]. Figure 3.2a shows the mass spectrum of Cytochrome C, which presents a narrow CSD where Δz = 4, with most of the intensity in the [M+7H]⁺ ion. This narrow CSD is typical of a structured protein [38, 48]. The dashed black line in Figure 3.2a and Figure 3.2b, labelled Z_R, corresponds to the Rayleigh limit [59] which predicts the highest charge that a sphere of a given molecular mass can accommodate (see below). Because all of the charge states for Cytochrome C are lower than the Rayleigh limit, it can be assumed that all of the ionised proteins are roughly spherical in shape, although it is possible that a low charged anisotropic conformation may also fulfil this criterion.

From ion mobility experiments, the rotationally averaged CCS for each charge state has been calculated (Figure 3.2c), and ranges from 952 Å² for the [M+5H]⁵⁺ ion to 1326 Å² for the larger [M+8H]⁸⁺ conformation, giving a ΔCCS of 374 Å². The collision cross section calculated from the crystal structure is shown by the dashed black line, which is larger than the cross section populated by the gas-phase molecules. This indicates a compaction of the protein structure in the gas phase, which is to be expected upon the removal of solvent [63]. By contrast, the mass spectrum of β-casein shows a very wide CSD from [M+7H]⁺ to [M+27H]²⁷⁺; Δz = 21 (Figure 3.2b). This appears typical of an unstructured protein [40]. The higher intensities for the lower charge states (z = 9–12) suggest that β-casein may exist predominantly in a compact state in solution, possessing less solvent-accessible area. (This point will be revisited later in the paper.) There are 16 peaks which correspond to ions with a charge state above the Rayleigh limit; we assume that these conformations are extended and therefore not spherical in shape. A very wide range of CCSs are found; from 1450 Å² for [M+8H]⁸⁺ to 5519 Å² for [M+26H]²⁶⁺, providing a ΔCCS value of 4069 Å². Interestingly, there is slight compaction between [M+26H]²⁶⁺ and [M+27H]²⁷⁺. This could either be due to neutralisation, by protonation, of two repulsing groups or perhaps due to the very highly charged species aligning in the drift tube. No crystal structure has been solved for the
disordered protein β-casein; hence, we cannot reference our CCS measurements to data from an alternative structural technique as for Cytochrome C.

Figure 3.2  (a) Mass spectrum of Cytochrome C. (b) Mass spectrum of β-casein. Both spectra were recorded under similar conditions: 50 mM ammonium acetate, pH 6.8. (c) The rotationally averaged collision cross sections (CCSs) of each charge state of Cytochrome C (see Table S2 in the Supporting Information, appendix 1); the dashed line represents the CCS of the crystal structure of Cytochrome C, calculated by the trajectory method [63]. (d) The rotationally averaged CCSs of each charge state of β-casein (see Table S3 in the Supporting Information, appendix 1).
3.3.2 Can the Median Charge (z) and the Width of the Charge State Distribution (Δz) Distinguish between Structured and Disordered Proteins?

As exemplified by the test proteins Cytochrome C and β-casein above (Figure 3.2), structured proteins appear to occupy a small number of charge states (low Δz) with a low median value of z, whereas intrinsically disordered proteins occupy a wider range of charge states (high Δz) with a higher value of median z. IM-MS analysis was performed on 18 other proteins under a variety of solution conditions to test the generality of this phenomenon (see Figure 3.3). Figure 3.3a correlates the median charge state presented by a protein (the midpoint between the highest and lowest charge state observed) against the molecular mass. For structured proteins (denoted by black letters in the figure), there is a strong positive correlation between molecular mass and median charge state, which is fitted to a curve where \( y = 0.6118x^{3/2} \). This correlation is expected, since the size of a protein increases with the surface area on which it can accommodate protons during ionisation. The median charge states range from 4 for melittin (2.8 kDa) to 24 for Immunoglobulin G (150 kDa). The curve for the natively structured proteins initially is steep for proteins below 20 kDa, and then begins to level off at higher molecular weights.
Figure 3.3  How the molecular mass of a protein relates to different mass spectrometry and ion mobility measurements. a) Median charge state, b) range of charge states, c) median collision cross section and d) range of collision cross sections. Proteins shown in red are known to be unfolded, either due to intrinsic disorder or denaturing conditions. A; Melittin [50], B; Human beta-defensin 2 [64], C; Lymphotactin (1-72) [53], D; Ubiquitin (denatured) [65], E; Lymphotactin [53], F; N-terminal p53 (Figure S1, Table S4), G; cytochrome c (Figure 3.2, Table S2), H; α-Synuclein (Figure S2, Table S5), I; N-terminal MDM2 (Figure S3, Table S6), J; haemoglobin α apo [48], K; haemoglobin α holo [48], L; haemoglobin β apo [48], M; β-Casein (Figure 3.2, Table S3) N; p53 DNA binding domain [66], O; p53 DNA binding domain pH 1.5 [66], P; Hen egg white ovalbumin (reduced, conformation 1), (Figure S4, Table S8), Q; Hen egg white ovalbumin (reduced, conformation 2) (Figure S4, Table S8), R; Hen egg white ovalbumin (intact, conformation 1) (Figure S4, Table S7) , S; Hen egg white ovalbumin (intact, conformation 2) (Figure S4, Table S7), T; TTR (tetramer) (Figure S5, table S9), U; Avidin (tetramer) (Figure S6 Table S10), V; BSA (conformation 1) (Figure S7, Table S11), W; BSA (conformation 2) (Figure S7, Table S11), X; Concanavalin A (tetramer) (Figure S8, Table S12), Y; SAP (pentamer) (Figure S9, Table S13), Z; Immunoglobulin G4 A (Figure S9, Table S11, [67]) . The solid black line on Figure 3.3a refers to
the De la Mora relation (equation 3.4) [59]. The dashed black line refers to the equation \( y = 0.06118 \sqrt{x} \) which is fitted to our data with an \( R^2 \) value of 0.97334. Supporting information can be found in appendix 1 of this thesis.

The proteins that are expected to present more extended conformations, because of either intrinsic disorder or denaturing conditions (depicted in Figure 3.3a by red letters), have a distinctly higher median charge than folded proteins of a similar molecular mass, with just a small amount of overlap in the low \( M_w \) region. This dataset was best-fitted with a straight line function corresponding to \( y = (5.8 \times 10^{-4})x + 3.21 \), with a standard deviation to the slope of 1.5 \( \times 10^{-3} \). The significantly higher median charge for disordered or solvent-disrupted proteins versus molecular mass can be attributed to a higher number of solvent-accessible ionisable sites available for protonation in these proteins, compared with natively folded proteins, which have less ionisable residues available for protonation. The p53 DNA-binding domain studied here is the wild-type protein, rather than the superstable quadruple mutant designed by Nikolova et al [68].

The black line in Figure 3.3a is the solution of an empirical relationship derived by De la Mora [59] over the mass range of our test set of proteins, which describes the limit to the maximum charge that a spherical protein can hold (Equation 3.4) [59]:

\[
Z_R = 0.0778 \sqrt{m}
\]

Equation 3.4

where \( Z_R \) is the maximum charge and \( m \) is the mass of the protein. The median charge states of folded proteins under physiological-like conditions lie under this line, whereas the majority of the unstructured proteins lie above it (see Figure 3.3a). The De La Mora relationship, which is more commonly used to justify folded/compact low-charge-state species, can also be seen to provide a boundary between predominantly or completely folded/globular proteins and unfolded/extended proteins. Although the de la Mora equation lies substantially above the median charge state of the structured proteins, it correlates well with the maximum charge states of these proteins, as originally described [59]; we find a better fit to our median charge state data with a constant of 0.6118. A similar approach has been previously reported by Testa et al. [30], although they did not consider any proteins larger than haemoglobin.

Figure 3.3b plots the number of charge states populated by a protein following n-ESI, (\( \Delta z \)). Here, the results are remarkable. Over the entire molecular mass range
investigated, from 2846 Da to 150 kDa, folded proteins are present, with $\Delta z \leq 7$. By contrast, for disordered or denatured proteins, $\Delta z \geq 7$. It is surprising that (i) the charge state range of the natively folded proteins appears to be independent of molecular mass, and, in particular, (ii) even large proteins display a narrow charge state range following n-ESI. We might predict that for large proteins, protonation on one part of the molecule would be independent of protonation sites on other areas. However, Figure 3.3b demonstrates the following, when characterising an unknown protein of molecular mass $\leq 100$ kDa, using ESI-MS: (1) if it is present with more than 7 charge states, it can be assumed to have at least some disordered regions, or be denatured under the chosen solution conditions; or (2) if it presents with less than 7 charge states, it is more likely structurally stable. Again, we have fit the data for both the structured proteins and the disordered set; the fitting functions and standard deviations are found on the figure.

The comparison between charge state range and molecular mass was extended into the MDa range. (See Figure S10 in the Supporting Information, appendix 1.) It is apparent that, for protein complexes above 150 kDa, $\Delta z$ increases slightly with molecular mass. Hepatitis B Virus Capsid $T = 4$ has a value of $\Delta z = 25$ [69], which allows extrapolation of a straight line fit from the data for Cytochrome C where $\Delta z = 4$, to the data for the 4MDa capsid; this gives a gradient of $5 \times 10^{-6}$ (data not shown), as opposed to the gradient of zero used in the fit to our data (Figure 3.3b), showing how weak the correlation between mass and $\Delta z$ is for ordered proteins. We can only speculate here on the reasons for the small $\Delta z$: it is probable that, during the last stages of desolvation, protonated side chains on the surface of a quasi-spherical protein will seek to pair with any proximal deprotonated acidic groups. Given the lowered dielectric of the vacuum compared with bulk solution, the range of electrostatic attraction between these groups will increase as the solvent is removed. As these groups pair, the charge on each is neutralised. The location of each chargeable group on the surface of the protein or protein complex will ultimately limit the ability of a given side chain to find a partner, and there will be side chains left unpaired. The net difference in the total number of unpaired protonated side chains (for positive ionisation mode, as shown here) is apparently small for structured proteins, resulting in the low $\Delta z$ that we observe. The energy in the pairing interactions will be stronger than that found in a solution salt bridge, and this will result in a caging effect to the surface of the protein. We might anticipate that such an effect would tighten the conformation of the protein compared to that found in bulk solution, and indeed that is what we find (see, for example, Figure 3.2.
above for Cytochrome C and ref [63]). This apparent ability of structured proteins to accommodate a discrete range of charges on their surface may be considered to be a partial solution to the Thomson problem [70].

### 3.3.3 Can the Median Collision Cross Section (CCS) and the Width of the CCS (ΔCCS) Distinguish between Structured and Disordered Proteins?

Plots similar to those in Figure 3.3a and b were constructed using ion mobility data. Again, the best fits to the data and the corresponding standard deviations are shown in the figure. The median CCS displayed by a protein, plotted against the molecular mass (Figure 3.3c), highlights a positive correlation between molecular mass and median CCS. Unfolded proteins have a tendency to have a higher median CCS, although there is some overlap for low masses as for the median charge (Figure 3.3a).

Figure 3.3d shows the range of collision cross sections for each protein (ΔCCS) plotted against molecular mass (M), and it is immediately obvious that this parameter shows very little dependence on molecular mass. With the exception of IgG (“Z” on plot), the presumed folded proteins occupy a consistently very narrow CCS range; the difference between the largest and smallest conformation is generally <550 Å², and the ΔCCS values scatter, with respect to M. This result is remarkable. Even SAP, a 128 kDa pentamer with a median CCS of over 6000 Å², has a ΔCCS value of just 408 Å². Conversely, unfolded proteins occupy an extremely wide range of collision cross sections, indicating conformational ensembles ranging from compact to extended states, as shown previously for β-caesin (Figure 3.2b). This supports the assertion made above in the CSD analysis, that a large Δz is found along with a large ΔCCS, and that both are signatures of disordered or denatured proteins. The narrow CCS range found for the protein complexes also supports the electrostatic caging effect that we have attributed to the small Δz found for structured proteins.

IgG, “Z” on the plot, has a ΔCCS value of 2140 Å², which is larger than that found for the other structured proteins. We attribute this to inherent conformational flexibility between different subunits [67]. Data obtained from static measurements via CryoEM and TEM have also indicated that IgGs are flexible and dynamic, capable of subunit waving and rotation, as well as bending and wagging[71]. The CCS range is even larger than shown in Figure 3.3d, since the arrival time
distribution (ATD) for each charge state for IgG is very wide, but the CCS shown here is taken from the apex of the ATD [67].

Hen egg white ovalbumin and bovine serum albumin (BSA) each show two distinct conformations within one charge state (see Tables S7, S8, and S11 in the Supporting Information, appendix 1). Here, the two conformations have been treated separately. Both conformations of BSA are present in [M+14H]^{14+} to [M+17H]^{17+}, spanning four charge states (Figure 3.3, letters “V” and “W”). The smaller conformation has a median cross section of 3615 Å² and a range of 345 Å² (V), while the larger conformation has a median cross section of 4390 Å² and a range of 446 Å² (W). Ovalbumin was analysed in its reduced (P and Q) and oxidised state (R and S), since reduced ovalbumin has been previously reported to possess a similar structure to intact ovalbumin, albeit with significantly reduced conformational stability [72, 73]. The conformations of the reduced form have median CCSs of 2565 Å² (Q) and 2964 Å² (P) and ranges of 679 Å² and 382 Å², respectively. The larger conformation (P) spans four charge states, while the smaller conformation spans three (Q) (see Figure S4, as well as Tables S7 and S8, in the Supporting Information, appendix 1). The smaller conformation of intact ovalbumin (R) spans charge states [M+10H]^{10+} and [M+11H]^{11+}, has a median CCS of 2235 Å², and a ΔCCS value of 19 Å². The larger conformation (S) spans charge states [M+11H]^{11+} to [M+13H]^{13+}, has a median CCS of 2752 Å², and a ΔCCS value of 99 Å². This agrees with previous work by Tatsumi [72] and Takahashi [73], who proposed that reduction of the disulphide bond lowers conformational stability; the reduced form has a wider ΔCCS value in both conformations, although the median CCSs are similar to the oxidised form.

The CCS values of the proteins shown in Figure 3.3c and d are taken from the peak of the ATD for each resolvable conformation, or conformational family. The width of the ATDs may also be used to differentiate structured proteins from unstructured proteins (see Figure S11 in the Supporting Information, appendix 1), and this will form the subject of further study.

To summarise, we have shown that structured proteins and complexes below 130 kDa have Δz ≤ 5 and the difference between the largest and smallest CCS is <550 Å². Proteins that are unfolded, either due to intrinsic disorder or to denaturing solvent conditions, have Δz ≥ 7 and present with a range of conformations with a difference in CCS of >750 Å².
3.3.4 A Comparison of Experimental Findings with Theoretical Predictors of Disorder

Proteins that possess relatively high numbers of charged groups in solution have been linked with a tendency to be disordered. It is common to consider the net charge of the amino acids in the protein at a given solution pH versus their hydropathy as a predictor to distinguish ordered and disordered proteins [74]. One such method, based on the research by Uversky et al. [74], proposes that a combination of high net charge and low hydrophobicity is an important contributing factor to intrinsic disorder in proteins. Here, we compare 10 of the investigated proteins, 5 structured and 5 disordered, with the charge-hydropathy scale (provided by PONDR, Figure 3.4A). We also compare the calculated parameters (mean net charge and mean scaled hydrophobicity), separately, with the Δz values that we derive experimentally.
Figure 3.4  (A) Comparison of the mean net charge and the mean scaled hydrophobicity of 10 of the investigated proteins. Comparisons of the charge state range are made to (B) the mean net charge and (C) the mean scaled hydrophobicity. Legend: (1) α-synuclein, (2) bovine serum albumin, (3) Cytochrome C, (4) hemoglobin-α chain, (5) lymphotactin (1–72), (6) WT lymphotactin, (7) N-terminal MDM2, (8) N-terminal p53, (9) hen egg white ovalbumin, and (10) p53 DNA-binding domain. Image adapted from Uversky et al. [74] (Copyright 2000, Wiley–Blackwell). Proteins shown in red are known to be unfolded, either due to intrinsic disorder or denaturing conditions. Proteins shown in black have been previously shown to be structured in solution.
Two proteins were assigned by charge-hydropathy analysis as being disordered; N-terminal p53 and Cytochrome C. N-terminal p53 has also been shown to be devoid of tertiary structure by several biophysical techniques [75, 76], and here we have shown that it presents a wide range of Δz values (Figure 3.3). Native Cytochrome C is folded into a compact conformation with a covalently bound haem group, required for the folding of Cytochrome C. Without this haem group the protein loses the majority of its secondary and tertiary structure [77, 78], and this haem group is not taken into account in the charge-hydropathy analysis, which can explain the discrepancy between mass spectrometry and charge-hydropathy results. Lymphotactin WT and lymphotactin (1–72) both sit on the boundary between ordered and disordered. Lymphotactin (1–72) lacks the disordered tail that the WT contains [79], so a bigger difference would be expected between the two forms of the protein, as is shown by the IM-MS outputs of Δz and ΔCCS (see Figure 3.4B and C).

α-Synuclein, p53 DBD, and NT-MDM2 are all classified by this charge-hydropathy analysis as being ordered. This is inconsistent with our MS results, as well as the perceived wisdom from a range of other biophysical techniques [42, 76, 80-82]. Bovine serum albumin, hemoglobin-α, and ovalbumin are assigned as structured by charge-hydropathy analysis, and this is in agreement with the IM-MS data.

Figure 3.4B and C both show that there appears to be little differentiation of net charge or hydropathy with respect to whether the protein is folded or intrinsically disordered. Our experimental results, however, illustrate the strength of experiment compared with this theoretical approach; there is clear distinction of folded and disordered proteins, with respect to the experimental parameters Δz which is compared to charge, and ΔCCS which is compared to hydropathy.

We also used two more methods of predicting disorder from the primary sequence that are more sophisticated than the charge-hydropathy analysis, namely, other versions of PONDR (PONDR-VLXT PONDR-VSL2 and PONDR-VL3) and DISOPRED [55]. PONDR uses the primary sequence to evaluate attributes such as the sequence complexity and hydrophobicity, which are then used as inputs to make predictions. The predictors have been trained on sets of disordered and ordered sequences, which allows them to generalise new sequences. There are then several outputs that differ in the method of characterisation of the training data, such as X-ray crystallography or NMR [83]. DISOPRED3, on the other hand, was trained on a set of sequences with high-resolution X-ray structures. Disorder is
identified by residues that are in the sequence but have coordinates missing from the crystal structure [55]. We carried out PONDOR and DISOPRED analysis on a set of six proteins: three structured proteins (Cytochrome C, bovine serum albumin, and ovalbumin) and three disordered proteins (α-synuclein, N-terminal MDM2, and p53 DNA-binding domain). Table 1 shows a summary of the analysis, which is displayed in Figure S11 in the Supporting Information, appendix 1.

<table>
<thead>
<tr>
<th>Protein</th>
<th>% disordered DISOPRED</th>
<th>% disordered PONDOR VL3</th>
<th>% disordered PONDOR VSL2</th>
<th>% disordered PONDOR VLXT</th>
</tr>
</thead>
<tbody>
<tr>
<td>α-Synuclein</td>
<td>56</td>
<td>100</td>
<td>91</td>
<td>52</td>
</tr>
<tr>
<td>N-terminal MDM2</td>
<td>31</td>
<td>26</td>
<td>35</td>
<td>36</td>
</tr>
<tr>
<td>P53 DNA-binding domain</td>
<td>11</td>
<td>37</td>
<td>42</td>
<td>26</td>
</tr>
<tr>
<td>Cytochrome c</td>
<td>8</td>
<td>50</td>
<td>54</td>
<td>9</td>
</tr>
<tr>
<td>Ovalbumin</td>
<td>0.2</td>
<td>5</td>
<td>18</td>
<td>24</td>
</tr>
<tr>
<td>Bovine serum albumin</td>
<td>9</td>
<td>38</td>
<td>32</td>
<td>18</td>
</tr>
</tbody>
</table>

Table 3.1 Percentage of Disordered Residues in Six Different Proteins, as Predicted by DISOPRED, PONDOR-VL3, PONDOR-VSL2, and PONDOR VLXT. Data extrapolated from Figure S12 in the Supporting Information, appendix 1.

α-Synuclein is classified as being over 52% disordered by all of the prediction methods, which fits well with both our analysis and that from other research techniques [42, 84]. Disorder predictions of MDM2 range from 26% to 36% disordered, and all predictors display a structured region between amino acids 42 and 111 (Figure S11 in the Supporting Information, appendix 1). p53-DBD is predicted to be disordered to an extent of 11%–42%, which is lower than we would expect from such a wide range of Δz and ΔCCS values. Surprisingly, the prediction values of this IDP are similar to that of BSA (9%–38%), which is characterised as being structured by a variety of techniques. The predictions of the extent of disorder in ovalbumin are very low, as expected, with values ranging from 0.2% to 24%. Cytochrome C has a wide range of predictions, from 8% to 54%, and one reason for these discrepancies could be due to the covalently bound haem group, which is known to stabilise the tertiary structure of the protein but will not be taken into
consideration by the information in the primary sequence, as described above. Unlike the rather binary output provided by our ion mobility mass spectrometry analysis (recall Figure 3.3a-d), where we classify proteins as either ordered or disordered, these disorder predictors seek to capture the progression of structure.

### 3.3.5 How Much of the Possible Conformational Space Is Occupied in the Presented Collision Cross Sections?

In order to put the ion mobility results into context, and to start to consider the progression of structure from ordered to disordered, we apply a simple model to consider the most extended and most compact possible conformations. The largest and smallest cross sections are calculated based on the volume of the amino acids in the polypeptide, arranged in either a spherical (compact) or cylindrical (extended) configuration. These calculated cross sections are compared to the maximum and minimum CCSs seen in ion mobility experiments (see Figure 3.5).
Chapter 3; An MS-based framework to define disorder in proteins

Figure 3.5  Comparison of the theoretical and experimental collision cross sections: The black bars show the theoretical CCS range. Cytochrome C, lysozyme, and myoglobin are examples of structured proteins. Denatured lysozyme is constrained with disulfide bonds; β-casein and α-synuclein are examples of IDPs. (See Table S14 in the Supporting Information, appendix 1.)

This comparison (Figure 3.5) reveals several interesting features. The ordered proteins (Cytochrome C, myoglobin, and lysozyme), have experimentally measured ΔCCS under native conditions that are significantly narrower than the limits from the
model. This can be rationalised by considering that these proteins possess high degrees of tertiary structure under the experimental solution conditions, which prevent the proteins from unfolding. The CCS values measured for these structured proteins have a tendency to be toward the lower end of the allowed CCS range, which implies that they prefer to adopt compact conformations; although Cytochrome C is theoretically able to occupy cross sections ranging from 759 Å² to 3632 Å², the experimental results show that, under native conditions, the cross sections only range from 952 Å² to 1326 Å².

When lysozyme is sprayed from more denaturing conditions (water: methanol: formic acid ratio = 49:50:1), the smallest CCS measured increases from 1016 Å² (with ammonium acetate) to 1222 Å², although the upper end remains very similar (1906 Å², compared to 1895 Å²³). This suggests that the salt provides conditions that stabilise the most compact conformations, but the more hydrophobic solvent environment and decreased pH does not allow lysozyme to present extended conformations (as for Cytochrome C), because of the fact that it is constrained by four intramolecular disulphide bonds. Under reducing conditions, the experimental upper CCS value increases considerably, to 2989 Å², as the four disulphide bridges are cleaved.

Reduced lysozyme, denatured Cytochrome C, and denatured myoglobin show good agreement between their experimentally measured CCSs and their respective theoretically permitted ranges. This can be attributed to solution conditions, which destabilise their tertiary structure, inducing unfolding and, consequently, allowing the protein to adopt a range of conformations, from very compact to very extended.

The disordered proteins, β-casein and α-synuclein, behave as predicted. They explore most of their predicted range of CCS, even when sprayed from high ionic strength solution conditions, because they are intrinsically disordered with little stabilising secondary or tertiary structure. They present to the gas phase with many different conformations, ranging from compact to extended, unhindered by energetic constraints. They are qualitatively very similar to structured proteins sprayed from denaturing conditions.
3.3.6 Can We Quantify the Relative Population of Conformations from Compact to Extended?

To provide more detail on the nature of the observed states of a given protein, and the extent of structure or disorder, the relative intensities of the different charge states (as shown in a CSD) will be more informative than simply considering the width of the CSD alone. Plotting CSD along with CSS data allows us to construct postulated energy landscapes. A protein that populates a wide range of CCSs, 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. This is exemplified by α-synuclein (Figure 3.6A), which samples almost the entire range of theoretically possible conformations. Nine charge states ([M+7H]^{7+} to [M+15H]^{15+}) out of the 16 ([M+5H]^{5+} to [M+20H]^{20+}) have more than 30% intensity of the base peak (shown by the dashed line in Figure 3.6 A and B).

It has been reported previously that β-casein contains areas of structure and disorder [62, 85]. Livney et al. described the structure as having several definable parts, with others that are more flexible and dynamic, allowing transformation between several energetically favourable protein conformations. This is reflected in the intensities of the different charge states and corresponding CCSs of the protein. The most intense peaks correspond to [M+9H]^{9+}, [M+10H]^{10+}, and [M+11H]^{11+}, which are of low charge and low CCS (1862–2157 Å^{2}), relative to the majority of the peaks. These correspond to the protein in an energetically stable conformation, demonstrated by the troughs in the postulated energy landscape shown in Figure 3.6B. The peaks of higher charge and larger CCS are of much lower intensity, suggesting a dynamic ensemble of transiently populated conformations. Just three charge states out of 26 ([M+9H]^{9+}, [M+10H]^{10+}, and [M+11H]^{11+}) have more than 30% intensity of the base peak.
Figure 3.6 Range of theoretical CCSs occupied by a protein ((A) α-synuclein, (B) β-casein, and (C) Cytochrome C) is shown in red. The relative intensity of each charge state is plotted against its CCS. On the right-hand side, the illustration shows a postulated shape of the folding landscape of each protein. The horizontal lines in parts a and b correspond to a relative intensity of 0.3, and are there to allow comparison.

A structured protein can be considered to possess a landscape with one low energy minima, in which most of the protein molecules will be situated [86]. The corresponding mass spectrum of a protein with this funnel-shaped landscape will display a narrow charge state distribution of five peaks or less, corresponding to one
An MS-based framework to define disorder in proteins

single conformational family with minimum dynamics. This is demonstrated by Cytochrome C in Figure 3.6C. Cytochrome C has been reported to possess a narrow folding funnel with one energetically favourable conformational family [61]. This is demonstrated by the narrow CSD, with the dominant species being the [M+7H]^{7+} ion with low CCS values compared to the theoretical CCS values available to the polypeptide chain.

Here, we have demonstrated how the intensities of different charge states of a protein can be used to make structural predictions of proteins with varying extents of disorder.

**3.3.7 The Framework and How To Determine the Structural Preferences of an Unknown Protein**

The objective of this investigation was to determine what structural information could be obtained about an unknown protein if mass spectrometry was the only method available (see Figure 3.7). The first step is to perform an ESI-MS experiment at low pH. From this, it is facile to determine the mass of the protein. Through the addition of a reducing agent and a thiol capping group, one is able to determine whether intramolecular disulphide bonds are present within the protein. This is because the protein conformation—and, therefore, the CSD—will be altered if disulphide bonds are broken. The number of capping groups seen in the mass shift will reflect how many free cysteine residues are present. The CSD will normally be wide, because of the protein being denatured. Returning to the oxidised form, the next step would be to conduct an ESI-MS experiment under high ionic strength conditions, and at a range of pHs. If the protein has any regions of structure under these physiological-like conditions, the CSD will narrow. If the CSD remains the same as that under low pH, this indicates an intrinsically disordered protein with no structure to be disrupted by the low pH. Significantly more populated peaks at lower charge states under physiological conditions indicates a protein that has regions of order (compactness) and disorder (extended states), whereas a natively unstructured IDP will display roughly equal intensity across the entire range of charge states (see Figure 3.7).
Figure 3.7  Flowchart displaying the procedure for elucidating structural information on an unknown protein.

Information obtained from mass spectrometry can be complemented by ion mobility experiments, which provide further information on conformational occupancy of the protein, in the form of a rotationally averaged CCS for each charge state, which can be compared to data obtained from other methods. A structured protein will present a narrow range of CCSs, a disordered protein will present a wide range of CCSs, and a protein with both structured and disordered regions will present a large range of CCSs with higher intensity of more-compact conformations.

### 3.4 Conclusions

The objective of this research was to provide a framework for mass spectrometry (MS) experiments that could be performed as an initial characterisation technique to determine the dynamic properties of a protein for which no structural information is known. To observe trends in ion mobility–mass spectrometry (IM-MS) results with respect to structure and disorder, MS and IM-MS results were compared for a set of 20 proteins, spanning a molecular mass range of over 147 kDa with varying extents
Chapter 3; An MS-based framework to define disorder in proteins

of secondary tertiary and quaternary structure (see Figure 3.3). It was noted that, while structured proteins have a median charge lower than the De la Mora/ Rayleigh limit, disordered proteins have a higher median charge and, more importantly, display a wider charge state distribution than structured proteins. For proteins with a mass of M < 150 kDa, structured proteins present with z ≤ 7 states or less, whereas disordered proteins will present with z ≥ 7. The low Δz for structured proteins is only very weakly dependent on mass, and, here, we have suggested that it correlates to a polyelectrostatic view of charge pairing on desolvation.

Comparison of collision cross sections (CCSs) measured by IM-MS show that, although there is little distinction in terms of median CCS between structured and disordered systems, the range of CCSs that a given protein presents reveals its conformational flexibility; structured proteins will have a CCS range of <750 Å², whereas disordered systems will have a CCS range of >750 Å².

Comparisons were then made between measured CCSs and theoretically calculated CCSs for a set of five proteins, three of which are structured and two of which are disordered (see Figure 3.5). When analysed under native-like conditions, the structured proteins display a much narrower CCS range than is theoretically available and, moreover, these are at the smaller end of the allowed CCS range. When denatured and, in the case of proteins containing disulphide bonds, reduced, the measured cross-section range increases to almost all of the space that is theoretically available. In the case of IDPs, however, the CCS range is wide, even under native-like conditions, because they are free to access most of their conformations. When the shape of the charge state distribution (CSD) was considered along with the CCS range, it is apparent that IDPs will have a wide CSD with similar intensities across the charge states, structured proteins will have a narrow CSD with most of the intensity in just one or two charge states, and proteins with regions of both structure and disorder, or a tendency to disorder, will have a wide CSD, with most of the intensity in the lower charge states (see Figure 3.6). From this information, we have built a framework of exemplar experiments, shown in Figure 3.7. This manuscript allows us to propose a workflow for the use of MS and IM-MS as initial characterisation techniques to obtain information about the extent of structure or disorder in unknown proteins. In this initial investigation we have primarily selected systems that are either distinctively structured or unstructured, future work will focus on more conformationally diverse proteins.
3.5 Acknowledgements

The Schools of Chemistry and Physics and Astronomy at the University of Edinburgh, and the School of Chemistry at the University of Manchester, are thanked for an award of a Biotechnology and Biological Research Council (BBSRC) case studentship to R.B., and for their continuing support of our research. LGC, Ltd. is thanked for providing R.B. with a CASE top up. The BBSRC are also thanked for funding the following projects, which, in part, supported the work herein: Nos. BB/L002655/1, BB/L002655/1, and BB/L015048/1. We thank the following for supplying us materials: Brian Volkmann, Medical College of Wisconsin; Ted Hupp, University of Edinburgh; Tilo Kunath, University of Edinburgh; Galina Selivanova, The Karolinska Institute; and UCB Pharma and all Barran group members past and present who have contributed to these studies. We thank Frédéric Cazals (Inria Research Centre, France) for suggesting that structured proteins may “solve” the Thomson problem and Jim Warwicker (MIB, Manchester, U.K.) for helpful discussions on polyelectrostatics.

3.6 References

Chapter 3;  An MS-based framework to define disorder in proteins

Chapter 3; An MS-based framework to define disorder in proteins

64. Cecco, M.D., Biophysical studies to elucidate structure-activity relationships in beta defensins, in School of Chemistry. 2011, University of Edinburgh: Edinburgh.
4

Relating gas phase to solution conformations: Lessons from disordered proteins
Chapter 4; Relating gas phase to solution conformations

4.0 Declaration


This article has been reproduced in an unchanged format except for minor adjustments to incorporate them into this thesis.

As first author on this publication I expressed and purified ApoC-II, I collected and analysed the MS and IM-MS data for ApoC-II, I analysed the MS and IM-MS data for α-Synuclein and I drafted and edited the manuscript. ASP expressed and purified α-Synuclein, and collected the HDX, MS and IM-MS data for α-Synuclein. LD collected and analysed the HDX data for ApoC-II. HMS analysed the α-Synuclein HDX data. RB drafted and edited the manuscript with assistance from PEB and CEM.
Relating gas phase to solution conformations: Lessons from disordered proteins

Rebecca Beveridge,1 Ashley S. Phillips,1 Laetitia Denbigh,2 Hassan M. Saleem,1 Cait E. MacPhee,3 Perdita E. Barran 1*

1 The Michael Barber Centre for Collaborative Mass Spectrometry, The School of Chemistry, Manchester Institute for Biotechnology, University of Manchester, Manchester, UK. E-mail: perdita.barran@ed.ac.uk

2 Waters Corporation, Wilmslow, UK.

3 School of Physics and Astronomy, University of Edinburgh, Edinburgh, UK.

Received December 17th 2014, Accepted April 24 2015.

Article first published online: 5th June 2015

Supporting Information

Tables of experimental collision cross sections obtained from DT IM-MS experiments and the deuterium uptake for ApoC-II and α-Syn can be found in the Supporting Information. This material is available free of charge via the Internet at http://wiley.com and in appendix 2 of this thesis.

The authors declare no competing financial interest.

4.1 Abstract

In recent years both mass spectrometry (MS) and ion mobility mass spectrometry (IM-MS) have been developed as techniques with which to study proteins that lack a fixed tertiary structure but may contain regions that form secondary structure elements transiently, namely intrinsically disordered proteins (IDPs). IM-MS is a suitable method for the study of IDPs which provides an insight to conformations that are present in solution, potentially enabling the analysis of lowly populated structural forms. Here, we describe the IM-MS data of two IDPs; α-Synuclein (α-
Syn) which is implicated in Parkinson's disease, and Apolipoprotein C-II (ApoC-II) which is involved in cardiovascular diseases. We report an apparent discrepancy in the way that ApoC-II behaves in the gas phase. While most IDPs, including α-Syn, present in many charge states and a wide range of rotationally averaged collision cross sections (CCSs), ApoC-II presents in just four charge states and a very narrow range of CCSs, independent of solution conditions. Here, we compare MS and IM-MS data of both proteins, and rationalise the differences between the proteins in terms of different ionisation processes to which they may adhere.

4.2 Introduction

The recent years have seen an explosion of research into intrinsically disordered proteins (IDPs) [1]. This subset of proteins are flexible and dynamic compared to globular proteins; they populate many interconverting conformations of similar energy and are classified as having no secondary structure on the timescale of an NMR experiment [2-4]. The lack of three-dimensional (3D) structure in IDPs allows them to bind to multiple partners, enabling them to play a key role in many cellular signalling networks [5-8]. It is for this reason that they are frequently implicated in cancers, since a disruption in the function or regulation of IDPs often results in a breakdown of cell division control leading to uncontrolled cell proliferation.

It is convenient to consider that IDPs have similar characteristics to structured proteins that have been denatured by the solution conditions, but this is an oversimplification. The dynamic properties of IDPs under physiological conditions differ from those of denatured structured proteins due to differences in their hydrodynamic behaviour. Rather than behaving as random coils, IDPs are often relatively compact compared to denatured globular proteins; transient elements of secondary structure reduce the hydrodynamic radius, giving rise to conformations of differing compactness [1, 9]. The extent of this compactness differs between IDPs due to differing levels of intramolecular non-covalent interactions such as hydrogen bonds and electrostatic interactions, as well as being affected by the solution conditions [10].

The lack of fixed structure in IDPs causes challenges when gathering structural information; they do not readily crystallise and NMR is unable to provide information on interconverting populations. While disorder can be detected by NMR spectroscopy by the chemical shifts of disordered residues, it is unable to report on more specific residual information since several conformations are interconverting
Chapter 4; Relating gas phase to solution conformations

on a timescale that is faster than that over which the NMR experiment takes place. Electrospray ionisation–mass spectrometry (ESI–MS) has provided valuable information on dynamic ensembles of IDPs; it is sensitive to the degree of disorder and can represent the full conformational range of an IDP [11]. In ESI-MS, proteins are observed in a range of charge states which, in positive ionisation mode, are commonly due to protonated forms of the protein. Structured proteins have a limited number of solvent-accessible ionisable sites and hence display in a narrow range of charge states upon ionisation [12]. Disordered proteins, however, exist in a range of conformations, from compact to extended, with differing numbers of protonatable sites at the surface of the proteins. IDPs, therefore, display a wide charge state range which reflects their dynamic behaviour and allows distinction between different conformations of the same protein [13, 14].

Ion mobility (IM) is a gas phase electrophoretic technique which can be used to give an extra experimental dimension to MS data. During an ion mobility experiment, a packet of ions is pulsed into a drift tube, across which is applied a weak electric field. The ions are drawn though the cell, but are hindered by collisions with an inert buffer gas (in this case helium) which is at a known temperature and pressure. The velocity of a given ion is influenced by two factors: the shape and the charge, both of which will determine the number of collisions with the buffer gas and how quickly the ion is pulled through the drift cell [15]. In a typical IM–MS experiment we record the drift time of m/z separated species and from this measurement we can calculate the buffer gas dependent rotationally averaged collision cross section (Ω) which correlates to the available conformation(s) of any given protein [16]. IM–MS has a particular application in the study of IDPs [17-20].

The mobility of an ion is determined as the ratio of the drift velocity (v_d) and applied electric field (E). It is then possible to determine the CCS on the basis of Equation 4.1:

$$K_0 = \frac{3ze}{16N} \left( \frac{2\pi}{\mu k_B T} \right)^{1/2} \frac{1}{\bar{B}}$$

Equation 4.1

where $z$ is the ion charge state, $e$ is the elementary charge, $N$ is the gas number density, $\mu$ is the reduced mass of the ion-neutral pair, $k_B$ is the Boltzman constant, $T$ is the gas temperature and $K_0$ is the reduced mobility (the measured mobility $K$ standardised for pressure and temperature to 273.15 K and 760 Torr).

Hydrogen–deuterium exchange coupled with mass spectrometry (HDX–MS) is a sensitive and rapid technique with which to investigate both structure and dynamics
of proteins in solution. HDX-MS can localise protected areas within proteins; that is areas which are not accessible to the solvent due to secondary or tertiary structure [21]. During a typical HDX-MS workflow, the protein of interest is diluted into deuterated buffer, and the exchange of solvent deuterium atoms is allowed to proceed for a given amount of time. The reaction is then quenched by the addition of a low pH buffer which reduces the back-exchange of deuterated backbone amides. The protein is digested by an acid-stable protease, in this case pepsin, and the HDX extent is analysed by MS of the protein fragments. Because IDPs contain few protected amide hydrogens due to their lack of tertiary structure, peptides from such proteins are expected to be fully saturated with deuterium at the earliest on-exchange time point [22]. Here, we use HDX-MS and ESI-IMS-MS to study two proteins which have been shown by many techniques to be intrinsically disordered; α-Syn and ApoC-II (Figure 4.1)

<table>
<thead>
<tr>
<th>Protein</th>
<th>Mass (Da)</th>
<th>Number of acidic residues (D,E)</th>
<th>Number of basic residues (R, K, H)</th>
<th>Net charge</th>
<th>pI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Apolipoprotein C-II</td>
<td>8959</td>
<td>11</td>
<td>7</td>
<td>-4</td>
<td>5.01</td>
</tr>
<tr>
<td>α-Synuclein</td>
<td>14451</td>
<td>24</td>
<td>16</td>
<td>-8</td>
<td>4.67</td>
</tr>
</tbody>
</table>

**Figure 4.1** Comparison of the biophysical characteristics of ApoC-II and α-Synuclein. (a) General properties of the proteins. (b) Sequences showing positive and negative residues in green and red, respectively. (c) Wimley–White plots showing the hydrophobicity of the proteins.
α-Syn is a protein of 140 amino acids (14,460 Da) which is highly expressed in the brain and is the primary component of the Lewy body deposits found in dopaminergic neurons that characterise Parkinson's disease (PD) and other neurodegenerative diseases. Despite being the focus of much research, its exact function is still unknown, but it is perhaps one of the most thoroughly investigated IDPs due to its role in PD. While α-Syn tends to be natively disordered, it has been shown that alterations to the solution environment, either an increase in hydrophobicity or a decrease in pH, can induce partial folding [23]. ApoC-II is a plasma protein containing 79 amino acids (8,959 Da) and is involved in lipid transport and metabolism. ApoC-II is a protein activator of lipoprotein lipase which is an enzyme that hydrolyses triacylglycerol during the metabolism of chylomicrons and very low-density lipoproteins [24].

Both ApoC-II and α-Syn are classified as being IDPs [25, 26]. They both have a net negative charge; -4 and -8 for ApoC-II and α-Syn, respectively, and no hydrophobic regions. The Wimley–White hydrophobicity scale is a combination of two different scales; one which considers the enthalpy change upon transfer of unfolded domains from water into the lipid bilayer, and one which considers the enthalpy change upon transfer of folded chains into the hydrocarbon interior. These two scales are then combined to create a whole residue hydropathy plot which accounts not only for the side chains, but also the peptide bonds, both of which are important to consider [27]. Neither of the proteins examined here have any regions which would preferably exist in a hydrophobic environment than an aqueous one, so it can be assumed that hydrophobicity for both the proteins is very low; the more positive the value, the more hydrophobic are the amino acids in that segment of the protein. Here, all the values for both the proteins are negative (Figure 4.1c), indicating a predominantly hydrophilic amino acid composition. It is worth noting that this scale assesses the likelihood of a protein segment forming a transmembrane helix and does not account for binding to the surface of membranes, which is discussed below.

ApoC-II and α-Syn have both been investigated in the presence of lipids. ApoC-II plays a role in plasma lipid metabolism; it binds reversibly to the polar lipid surface of plasma lipoprotein particles in vivo and also to a range of synthetic and natural lipid surfaces in vitro [28] with a corresponding change in the secondary structure characteristics. MacRaild et al. have demonstrated the propensity of ApoC-II to fold into α-helices both in the presence of sodium dodecyl sulphate [29] and dodecyl phosphocholine [30]. It has also been shown that α-Syn forms helices in the
Chapter 4; Relating gas phase to solution conformations

presence of lipids [31, 32]. Moreover, it has been documented that both proteins form amphipathic α-helices when bound to a synthetic membrane. When the residues of α-Syn are plotted onto a helical wheel [33], several regions have polar and non-polar residues distributed on opposite sides of the helix [31], agreeing very closely to amphipathic α-helix found in the lipid binding domains of the exchangeable lipoproteins, a family of which ApoC-II is a member.

As described above, the biophysical characteristics of ApoC-II and α-Synuclein are similar. They have similar primary structures in terms of lots of charged residues, net negative charge and low hydrophobicity. They have a similar pH and they both bind to lipid membranes in a similar fashion which includes the formation of amphipathic α-helices. Because of all these similarities it could be hypothesised that the two proteins under scrutiny will display in a similar fashion in the gas phase. However, we have found major differences in the presentation of these two proteins to the gas phase from solution, as shown by MS and IM–MS data which we attribute to differences in the electrospray ionisation mechanism. This is supported by the findings from HDX–MS which reports on the solvated structure of a given protein, and both proteins are indicated to have a conformation which is completely solvent accessible even on a short timescale (15s).

4.3 Materials and methods

4.3.1 Sample preparation

α-Synuclein was expressed recombinantly, from a pT7-7 vector containing human α-Synuclein gene, kindly provided by Professor Chris Rochet, Purdue University and purified as described previously [34]; a Resource Q column (GE Healthcare Life Sciences, UK) was used. α-Synuclein was concentrated using Vivaspin 6, MWCO 10 kDa centrifugal sample concentrators (GE Healthcare Life Sciences, UK) and applied to a HiPrep 26/10 desalting column (GE Healthcare Life Sciences, UK), pre-equilibrated with 100 mM ammonium acetate (Fisher Scientific, UK), flow rate 10 mL/min. The eluent was lyophilised and stored at −80°C. Prior to use, α-Synuclein was resuspended in 50 mM ammonium acetate. ApoC-II was bacterially expressed and purified as previously described [35]. The pET11a/human ApoC-II construct was kindly provided by Associate Professor Geoff Howlett, University of Melbourne. The purified protein was buffer exchanged into the relevant MS-compatible solutions using Bio-Rad micro Bio-Spin P-6 Columns.
4.3.2 *Hydrogen-deuterium exchange*

Protein solutions were prepared of 6 μM ApoC-II and 20 μM α-Syn. HDX labelling and quenching procedures were automatically performed using the CTC PAL sample manager (LEAP Technologies, Carrboro, NC, USA). The samples were diluted with 10 mM phosphate in 99.99% deuterium oxide, pH 6.6 (pD 7.0) 30-fold and 20-fold for ApoC-II and α-Syn, respectively, and incubated for 0, 15 and 60 s at 20°C. Labelled samples were then quenched with an equal volume of pre-chilled 100 mM phosphate pH 2.5. All labelling time-points were analysed in triplicate. 50 μL of sample was injected on a nanoACQUITY UPLC™ system with HDX technology (Waters). Online pepsin digestion for ApoC-II was performed in 0.1% formic acid for 2 min at 20°C on a Porosyme immobilised pepsin cartridge (Applied biosystems). For α-Syn, this was performed for 1 min at 20°C on a Waters Enzymate™ immobilised BEH pepsin column (2.1 × 30 mm). The peptides were separated on a UPLC BEH C18 column (Waters) at 0°C. ApoC-II peptides were separated with a 6 min linear acetonitrile gradient (8–40%) containing 0.1% formic acid at 40 μL/min. α-Syn peptides were separated with a 7 min linear acetonitrile gradient (8–35%) containing 0.1% formic acid at 40 μL/min. Mass spectra were acquired on a SYNAPT G2-Si HDMS in MS5 mode over the m/z range of 50–2000. Non-deuterated peptides were identified using ProteinLynx Global Server software 3.1 (Waters). DynamX 2.0 software (Waters) was used to filter the peptides, to generate deuterium uptake plot for each of them and to visualise the deuteration on the protein sequence.

4.3.3 *Mass spectrometry and ion mobility*

Nano-electrospray ionisation (nESI) was used for all MS and IM–MS experiments. Solutions were ionised through a positive potential applied to a thin platinum wire inserted into a thin-walled glass capillary (inner diameter 0.9 mm, outer diameter 1.2 mm, World Precision Instruments, Stevenage, UK) that was pulled to an nESI tip in house with a Flaming/Brown micropipette puller (Sutter Instrument Co., Novato, CA).

4.3.4 *Mass spectrometry*

MS experiments were performed on a Q-ToF Ultima (Waters, Manchester, UK). α-Syn samples (concentration 70 μM) were sprayed from solutions of 50 mM ammonium acetate pH 6.8 or pH 3.5. ApoC-II samples (concentration 30 μM) were sprayed from solutions of 10 mM ammonium acetate pH 6.8, 100 mM ammonium
acetate pH 6.8, 100 mM ammonium acetate pH 2.5 or 100% MeOH, as outlined in the text/Fig. 4. Capillary voltage 1.6–1.9 kV, cone voltage 60–100 V, source temperature 80°C, collision energy 5.

4.3.5 Ion mobility-mass spectrometry

IM–MS experiments were carried out on a Waters Q-ToF I instrument that was modified in house to include a 5.1 cm drift tube which has been described elsewhere [15]. The temperature and pressure of helium in the drift cell were approximately 28°C and 3.7 Torr, respectively. Measurements were made at six different drift voltages from 60 to 20 V. The precise pressure and temperature were recorded for every drift voltage and used in the calculations of CCSs. Each experiment was performed in triplicate. Data were analysed using MassLynx v4.1 software (Waters, Manchester, UK), Origin v8.5 (OriginLab Corporation, USA) and Microsoft Excel. Ion arrival time distributions were recorded by synchronisation of the release of ions into the drift cell with mass spectral acquisition. The CCS distribution plots are derived from raw arrival time data using Equation 4.2, below [35].

$$\Omega_{avg} = \frac{(18\pi)^{1/2}}{16} \left[ \frac{1}{m_b} + \frac{1}{m} \right]^{1/2} \frac{z e}{(K_B T)^{1/2}} \frac{1}{\rho} \frac{t_d V}{L^2}$$

Equation 4.2

where $m$ and $m_b$ are the masses of the ion and buffer gas, respectively; $z$ is the ion charge state; $e$ is the elementary charge; $K_B$ is the Boltzmann constant; $T$ is the gas temperature; $\rho$ is the buffer gas density; $L$ is the drift tube length; $V$ is the voltage across the drift tube and $t_d$ is the drift time.

The raw arrival time output ($t_a$) includes time the ions spend outside of the drift cell but within the mass spectrometer, known as the dead time ($t_0$). The value for $t_0$ is calculated by taking an average value of the intercept from a linear plot of average arrival time versus pressure/temperature and was subtracted from the arrival time to calculate drift time [15] ($t_d$):

$$t_d = t_a - t_0$$

Equation 4.3

4.3.6 Wimley-White plot

Hydrophobicity scales were calculated using the MPEX software found at http://blanco.biomol.uci.edu/mpex/ [27]. Sequences can be found in Figure 4.1b.
4.3.7 Modelling of theoretical CCS extremities

This procedure has been described elsewhere [36]. Briefly, the lower boundary was calculated by assuming that the globular form of the protein approximates a spherical shape with a density of $\rho = 0.904\text{ Da/Å}^3$. Using the molecular weight $M_w$ of the protein, the volume of the protein sphere can be calculated via $V = M_w/\rho$. The radius of the sphere is therefore $r = \left(\frac{3V}{4\pi}\right)^{1/3}$. The collision cross section of a sphere of this radius is given by Equation 4.4:

$$CCS_{lower}(Å^2) = \pi r^2 = \pi \left(\frac{3V}{4\pi}\right)^{2/3}$$  \hspace{1cm} \text{Equation 4.4}

The upper boundary can be calculated by assuming that the protein is cylindrical in shape; the furthest distance between $\alpha$-carbons in a protein chain is 3.63 Å [37]. Therefore for a polypeptide of $n$ residues, the maximum linear dimension is $n(3.63) Å$. The radius is given by the geometric average of the sum of the radii of the amino acids contained in the protein’s sequence. The average volume of an amino acid in a protein’s sequence is given by Equation 4.5:

$$\bar{V} = \frac{\sum_i V_i N_i}{n}$$  \hspace{1cm} \text{Equation 4.5}

Here the sum is over all amino acids $i$, $V_i$ is the volume of the $i^{th}$ amino acid, $N_i$ is the number of amino acids of type $i$ in the protein sequence. The average amino acid radius is then approximated by $r = (V/\pi h)^{1/2}$, where $h$ (height) $= 3.63\text{Å}$. This forms the radius of the fully extended protein cylinder. The collision cross section is then given by the rotationally averaged collision cross section of this cylinder since the protein ‘tumbles’ in the drift tube of the apparatus due to the low electric field.

The rotationally averaged collision cross section is given by the projection area of the cylinder:

$$CCS_{upper}(Å^2) = (\text{Projection area of cylinder}) = \frac{4}{\pi} rl + 2r^2$$  \hspace{1cm} \text{Equation 4.6}

These values are then multiplied by a scaling factor of 1.19 to convert from geometric size to CCS in helium as outlined in [38]. These theoretical values are highly approximate and do not take into consideration disulphide bridges, proline residues or other non-covalent interactions or restrictions, but serve as upper and lower boundaries with which to compare experimental data.
4.4 Results and discussion

4.4.1 The use of HDX–MS to examine the solvent accessibility of the proteins α-Syn and ApoC-II

HDX–MS was used to confirm previous results that α-Syn and ApoC-II behave in a similar fashion in solution; as disordered proteins with little secondary structure. Such proteins are expected to undergo hydrogen–deuterium exchange rapidly since there is no protection of the backbone amide protons from the solvent due to the adoption of structure by the protein. This is indeed what is observed for both proteins. Figure 4.2 illustrates this, it shows MS data over a time course of exposure to a deuterated solution for three peptides from each protein. We have selected one from each terminal region and one roughly in the middle, but these are typical for all peptides found from each protein. Three time points are shown for each peptide; t = 0 (non-deuterated), t = 15 s and t = 60 s. In each case, no more deuteration is observed at t = 60 s than t = 15 s. This demonstrates that at just 15 s, all the deuteration that is possible has already occurred and there has been no prevention by secondary structure of the protein. This confirms that both proteins are flexible and dynamic under these solution conditions.

Figure 4.2 HDX data of three peptides of α-Syn (a) and ApoC-II (b). For each of the three peptides, the mass change at 15 s indicates that the maximum number of type II deuteriums that could exchange for hydrogens, have already been exchanged. For further details, see SI Table 3)
4.4.2 The analysis of charge state distributions to probe the accessibility of chargeable sites in α-Syn and ApoC-II following nESI–MS

nESI of α-Syn from 50 mM ammonium acetate (Figure 4.3a) produces ions ranging in charge state from [M+5H]^{5+} to [M+20H]^{20+}. This wide charge state distribution (CSD) is very typical of IDPs [39]. Because of the lack of secondary structure the protein is free to adopt many conformations, ranging from compact to extended, hindered only by weak energetic constraints. This is reflected in the CSD of the protein. The compact conformations contribute the lower charge states seen since there are few ionisable sites that are accessible to the solvent for protonation. As conformations become more extended there are more solvent accessible protonable sites which gain charges and contribute the higher charge states.

α-Syn produces a very variable CSD, as seen in Figure 4.3 a and b, which were taken under the exact same conditions and instrumental parameters. The fact that the CSD displayed by α-Syn is so variable further highlights the plasticity of the protein. The folding landscapes of structured proteins have one energy minima in which all or most of the molecules reside. This is characterised by a narrow CSD of five or fewer charge states. Some IDPs, for example β-casein, have structured regions that are joined together by regions of disorder allowing transitions between several low energy minima [40]. This is displayed by a wide CSD of more than five charge states, but with higher intensity peaks for the lower charge states [41]. Even most IDPs will have a shallow well in the folding landscape which represents a preferential conformation; proteins with a wide CSD often have a particular profile which is reproducible. The fact that α-Syn has such a wide CSD with variation in the most intense peaks suggests that the folding landscape is shallow and there are very few energetic constraints preventing access to certain shapes. This information from ESI–MS characterisation is in agreement with the HDX data shown in Figure 4.2.

When the pH is reduced to pH 3.5 there is a compaction of the protein conformation (Figure 4.3c). The most intense charge state is [M+7H]^{7+} which is lower than in either of the spectra taken at pH 6.8. This compaction has been observed by other biophysical techniques and is frequent in IDPs [42].
Figure 4.3  nESI of α-Syn. In both (a) and (b), the solution conditions are identical (50 mM ammonium acetate pH 6.8) but the CSD differs. In (c) the ionic strength has been increased to 100 mM ammonium acetate and the pH lowered to 3.5. The measured mass of the protein is 14 453 Da cf. the theoretical average mass, 14 460 Da.
nESI of ApoC-II from 10 mM ammonium acetate produces monomers ranging in charge state from \([M+4H]^{4+}\) to \([M+7H]^{7+}\), with most of the ions in the \([M+5H]^{5+}\) charge state (Figure 4.4a). This narrow CSD is unusual for an IDP, and contrasts greatly with the signature shown by α-Syn in Figure 4.3. Increasing the ionic strength, by raising the concentration of ammonium acetate to 100 mM, often causes a compaction of the protein conformation and a resulting shift to lower charge states; here it has very little effect on the CSD (Figure 4.4 a and b). Although reducing the pH of the solution conditions tends to have a denaturing effect on structured proteins, for IDPs there is little secondary structure to disrupt so the results are often minimal. Here, although there is a slight increase in the intensity of the \([M+6H]^{6+}\) charge state, the overall CSD is unchanged and remains narrow (Figure 4.4c). Adding a hydrophobic solvent often has the same denaturing effect as low pH, and it is highly unusual that spraying the protein from 100% methanol produces a narrower CSD than buffered conditions; the \([M+7H]^{7+}\) ion is totally depleted (Figure 4.4 d). After showing that the CSD cannot be altered by changing the solvent environment, it was investigated whether physical changes could induce any effect. The temperature of the capillary from which the protein solution was sprayed was altered, and spectra were recorded at 14 and at 80˚C (Figure 4.4 e and f). Even altering the temperature by such a large extent produces no change in the CSD. Although HDX experiments suggest an extended conformation of ApoC-II in solution, MS experiments indicate a compact conformation in the gas phase, regardless of solution conditions or temperature.
Figure 4.4 Mass spectra of ApoC-II taken under different solution conditions (a–d) and sprayed from different temperatures (e, f). (a) 10 mM ammonium acetate pH 6.8. (b) 100 mM ammonium acetate pH 6.8. (c) 100 mM ammonium acetate pH 2.5. (d) Methanol, (e) 100 mM ammonium acetate, 80°C, (f) 100 mM ammonium acetate, 14°C. Measured mass = 8 959 Da.

4.4.3 The use of IM-MS to probe the conformational spread of the gas phase forms of α-Syn and ApoC-II

IM analysis provides additional and complementary information to the MS results shown earlier. CCSs have been measured for each charge state (Figure 4.5 and Table S1 and S2 in the supporting information). The apex of the CCS distributions found for α-Syn range from 1043 Å² for the [M+5H]5+ ion to 2742 Å² for the [M+18H]18+ ion. Two distinct conformational families are observed for the ions carrying between 8 and 16 positive charges, with just one conformational family for the lower and higher charge states. The increase in CCS with respect to charge begins to level off at the [M+16H]16+, presumably as the addition of extra charges cannot induce any further coulombic unfolding as the protein is already in an almost fully extended state. This is typical behaviour of an IDP and can be observed on other systems [39].
Chapter 4; Relating gas phase to solution conformations

Figure 4.5  Ion mobility data of α-Synuclein (a) and ApoC-II (b). Collision cross sections of each charge state (top) and arrival time distributions of selected charge states (bottom).

The collision cross sections measured for ApoC-II are starkly different from that of α-Syn. For ApoC-II, the CCSs range from 945 Å² for the smaller conformation of the [M+5H]⁵⁺ ion to 1524 Å² for [M+7H]⁷⁺. Apart from [M+4H]⁴⁺ all charge states are present in two conformational families, with the more extended conformational family being more intense.

There is a wealth of information in the width of the CCS distributions of the proteins. The widths of the ApoC-II CCS distributions are ~700, 750, 900 and 950 Å² for the increasing charge states. For α-Syn, the widths are 1250, 1850, 2100 and 1700 Å² for [M+7H]⁷⁺, [M+10H]¹⁰⁺, [M+14H]¹⁴⁺ and [M+17H]¹⁷⁺, respectively. There is much more variance of CCS in each charge state of α-Syn. This could either be due to the presence of multiple conformers which we are unable to resolve, or interconversion of conformations on the timescale of the experiment. One way to discern which is occurring would be to carry out the experiments under reduced temperatures to see if it is possible to freeze out specific conformers. Either way, the wider CCS distributions for the α-Syn charge states indicates more conformational disparity.
4.4.4 The use of a simple model to predict conformational occupancy

We have developed a simple approach to predict the CCS of the most compact and most extended forms of both proteins, as described above in the method section and in [39]. A comparison of measured CCS values with those predicted for any protein sequence reveals how much of the possible conformational space that might be occupied by a given protein actually is. Using this model, we have previously demonstrated [39] that proteins that are known to be structured under given solution conditions present with a narrow range of CCS compared to that which they could occupy. By contrast, disordered proteins or proteins that are sprayed from denaturing conditions give an experimental ΔCCS that covers much of the predicted allowed space. Using this approach, α-Syn behaves as predicted for an IDP. The extremities of CCSs are predicted to be 743 Å² and 3380 Å² (ΔCCS 2637 Å²) for the smallest and the largest, respectively. The smallest measured CCS is 870 Å² and the largest is 3249 Å². The collision cross section ranges are therefore 2637 and 2379 Å² for the theoretical and measured values, respectively. This means that the calculated:measured CCS range ratio of α-Syn is 0.9, or rather that 90% of the allowed space is occupied.

For ApoC-II the calculated collision cross section range is 873–2786 Å² (ΔCCS = 1913 Å²) while the measured range is 945–1524 Å² (ΔCCS = 579 Å²). This means that in contrast to α-Syn, ApoC-II has a calculated:measured CCS range ratio of just 0.3. In addition, the largest conformation is just 55% of the size of the largest possible cross section that was calculated, and this is due to the low intensity [M+7H]⁺ ion; most of the ions are in the [M+5H]⁵⁺ charge state which provides two conformational families of just 945 and 1132 Å², the most intense of which is the larger. According to the mass spectrum, 86% of ions are present at [M+5H]⁵⁺. The ΔCCS of this charge state alone is 187 Å², which gives a calculated:measured CCS range of 0.10. The calculated:measured CCS range ratio for cytochrome c is 0.13, for lysozyme is 0.24 and myoglobin is 0.19. These are all proteins with a very specific three-dimensional structure that is required for their function. The occupancy for ApoC-II is not very different from these structured proteins when the full CSD is considered and is much smaller when percentage occupancy is taken into account, even though it is thought to exist in a dynamic ensemble of conformations in solution.
Although ApoC-II is 5 kDa smaller than α-Syn this does not account for the differences in observed gas phase conformations. A similar sized protein to ApoC-II is ubiquitin which has been widely investigated in MS and IM–MS studies. Ubiquitin sprayed from denaturing conditions has been reported to have a $\Delta z = 8$ and a $\Delta CCS = 998$ [43], with a calculated:measured CCS range of 0.62; much larger than that of ApoC-II. This is the conformational spread which can be expected in IM–MS experiments of a disordered form of ApoC-II.

### 4.5 Discussion and implications for the nESI process for these proteins

The variance in the recorded CSDs and the CCS distributions for these two examples of protein that in solution are conformationally dynamic, raises interesting questions regarding what happens as they desolvate. Figure 4.6 shows three possible electrospray processes to which a protein undergoing desolvation can adhere. Figure 4.6a shows the charge residue mechanism (CRM) which is a well-established mechanism for the desolvation of large and relatively lowly charged macromolecular species which are ‘structured’ [44, 45] and Figure 4.6c shows the chain ejection mechanism (CEM), recently proposed by Konermann et al. [46]. During the CRM process, it has been hypothesised that Rayleigh-charged nanodroplets containing a single analyte evaporate to dryness. They lose charge as the droplet shrinks, via the ejection of protons and small ions, which allows the droplet to remain close to the Rayleigh limit [38] as the size decreases. Remaining protons are transferred to the protein during the final stages of desolvation.
Figure 4.6 Three proposed electrospray mechanisms; (a) the charged residue model proposed by Dole [47] and (c) the chain ejection model proposed by Konermann et al. [46]. We also propose an intermediate between the two previously described mechanisms (b). Figure adapted from Ref. [46].

Konermann et al. [46] have used MD simulations to demonstrate that unfolded and unstructured proteins are transferred from the droplet into the gas phase via the CEM. This is attributed to their more extended nature, with exposed hydrophobic groups making it more difficult for the protein to reside in the droplet interior. Such proteins therefore readily migrate to the surface. One terminus gets expelled into the vapour phase, followed by stepwise ejection of the remaining protein and separation from the droplet. The prevailing view is that the CRM is adhered to when
proteins are present in solution in compact, and/or globular forms, whereas extended conformations will undergo a CEM. Therefore, structured proteins with a fixed tertiary conformation will commonly follow the CRM to produce ions with a median charge that is low, and satisfies the Rayleigh relationship \[38\].

Disordered proteins are capable of existing in a range of conformations, ranging from compact to extended, unhindered by energetic constraints. They also contain forms where a given region is compact and another is extended. These three types of conformation must all be accounted for in the transfer of protein to the gas phase. If, in the nanodroplet, an IDP happens to be in a compact conformation, it enters the gas phase via the CRM, resulting in the lower charge state region of the mass spectrum. The higher charges will be contributed by proteins undergoing the CEM which will take place if the protein in the nanodroplet is in an extended conformation. This leaves the intermediate charge states for which to rationalise a mechanism. Are they comprised of ions that may have desolvated via either (or both) of the above mechanisms? To answer this we propose an intermediate mechanism which will be followed by conformations with regions that are compact and regions that are extended. During this mechanism the extended part of the protein will be ejected from the droplet as in the CEM, but this will be an unfavourable way to remove compact region of the protein from the droplet which is more likely to undergo the CRM. The differing amount of the polypeptide which is in an extended or compact state will therefore give rise to intermediate charge states, and also multiple conformations within the same charge state.

The wide CSD of α-Syn demonstrates that it has multiple conformations in solution which give rise to a wide CSD due to different conformations undergoing different ionisation mechanisms. Some of these conformations will be capable of binding to membranes, however the mechanism for this remains elusive. ApoC-II is also disordered in solution, so we hypothesised that a wide CSD would be observed for this proteins also, however as described in this paper this was not the case. A narrow CSD centred on low charge states was observed from all tested solution conditions. This implies that conformational switching occurs during the very last stages of desolvation to provide a compact form that always presents to the gas phase as having undergone a CRM.

Ogorzalek Loo et al. have recently also described an intermediate region in the electrospray process where the desolvating ions have properties lying between those of solution-phase and gas-phase \[48\]. They contact the solvent transiently in a high electric field, facilitating proton redistribution to confer the most stability to the
emerging gas-phase ion. We speculate that extended forms of gas-phase ApoC-II are extremely unfavourable during this intermediate phase, resulting in charge redistribution and conformational switching to a compact, favourable conformation. This may be because it is a membrane interacting protein, stabilised in a lower solution dielectric. A similar effect was previously reported from IM–MS and HDX–MS studies of melittin [49], which was seen to rearrange to a helical structure upon desolvation, irrespective of solution conformation.

4.6 Concluding Remarks

Here, we have analysed two proteins by several MS methods; HDX–MS, nESI–MS and nESI-IM–MS. The proteins, namely α-Syn and ApoC-II, have both previously been assigned as IDPs and share many biophysical characteristics, as outlined in Figure 4.1. Firstly, the proteins were analysed via HDX–MS which confirmed that both proteins are disordered in solution, in agreement with previous studies [26, 50]. We then analysed the proteins via nESI–MS which shows stark differences in terms of CSD. α-Syn, as expected for an IDP, displays a very wide range of charges in fitting with a lack of structure [39]. Disparity is observed between different spectra taken under the same conditions, further highlighting the plasticity of the protein. ApoC-II however is present in no more than four charge states, irrelevant of the solution conditions from which it is sprayed, indicating a small number of gas-phase conformations. IM–MS experiments support the nESI–MS experiments of both proteins. A wide range of CCSs are measured for α-Syn. CCS tends to increase with charge, and there are also multiple conformations within each charge state. ApoC-II is present in a narrow range of CCSs, but also has more than one conformational family for most charge states. Comparison with a theoretical model which predicts the CCSs of the most extended and most compact conformations of a protein shows that α-Syn explores most of the conformational space that is available to it, whereas ApoC-II is only present in a small amount of available conformational space. Finally, we rationalise the range of conformations displayed in the gas phase by each protein by the ESI mechanisms that they both adhere to. Our research into disordered systems is forcing us to think more deeply about the processes that occur during the transfer of proteins to the gas phase, and teaching us more about electrospray.
4.7 Acknowledgements

The authors thank the support staff of the University of Manchester and the University of Edinburgh, whose efforts makes ours possible. We acknowledge Dr. Dominic Campopiano, Dr. John White, Dr. Martin Wear and the Edinburgh Centre for Translational and Chemical Biology (http://ctcb.bio.ed.ac.uk/) for assistance with protein expression. We gratefully acknowledge Dr. Tilo Kunath (University of Edinburgh) for his extensive knowledge of alpha synuclein. BBSRC and LGC ltd. are thanked for the award of a BBSRC funded studentship to R.B. RASOR Interdisciplinary Research Centre is thanked for the award of an EPSRC funded studentship to A.S.P. We also thank the British Mass Spectrometry Society for a grant that allowed us to purchase our nanospray tip pipette puller still going strong after 12 years of pulling.

4.8 References

Chapter 4; Relating gas phase to solution conformations

How the Distribution of Charges Distributes Conformation in Conformationally Dynamic Proteins
5.0 Declaration

This chapter consists of one research article drafted for publication: Beveridge, R., Migas, L., Das, R., Huang, Y., Pappu, R. V., Kriwacki, R. and Barran, P. E.. How the distribution of charges distributes conformation in conformationally dynamic proteins.

As first author on this publication I performed the MS and IM-MS research and drafted the manuscript. LM calculated the theoretical CCSs of the model structures. RD performed the Monte Carlo calculations. JH expressed and purified the protein samples.
How the distribution of charges distributes conformation in conformationally dynamic proteins.

Rebecca Beveridge,\(^1\) Lukasz Migas,\(^1\) Rahul Das,\(^2\) Yongqi Huang\(^3\), Rohit V. Pappu,\(^2\) Richard Kriwacki,\(^3\) Perdita E. Barran\(^1\)*

\(^1\)The Michael Barber Centre for Collaborative Mass Spectrometry, The School of Chemistry, Manchester Institute for Biotechnology, University of Manchester, Manchester, UK. E-mail: perdita.barran@ed.ac.uk

\(^2\) Washington University in St. Louis School of Engineering & Applied Science, Department of Biomedical Engineering, Whitaker Hall, Campus Box 1097, One Brookings Drive, St. Louis, Missouri 63130.

\(^3\) Structural Biology, MS 311, Room D1024F, St. Jude Children's Research Hospital, 262 Danny Thomas Place, Memphis, TN 38105-3678.

Supporting Information

Protein sequences, materials and methods, and additional mass spectra and arrival time distributions can be found in the Supporting Information. This material is available in appendix 4 of this thesis.

5.1 Abstract

The effect of charge and the location of charged residues on the conformational preferences of a given polypeptide chain is assessed using in vacuo and in silico approaches. The 110 amino acid C terminus of the protein p27 is highly flexible, and in vacuo presents a conformational distribution that stretches from a compact state centred on 1000 Å\(^2\) and more extended families. Two permutants of this protein are also examined that have charged residues distributed differently in the primary sequence; oppositely charged residues are more evenly distributed in one permutant which results in conformational heterogeneity resulting in a broad distribution of molecular shapes. In the other, oppositely charged residues are segregated in the linear sequence which results in more narrowly distributed, compact conformations. These trends were observed in both theoretical and experimental results. In vacuo the permutants populate a wider conformational
Chapter 5       How the distribution of charges distributes conformation

landscape than in solution. This arises due to the absence of ionic or solvent screening, causing electrostatic interactions to dominate and revealing differences in the conformational features of permutants with different charge distributions.

5.2 Introduction

Intrinsically disordered proteins (IDPs) are highly dynamic and are classified as having little resolvable secondary structure on the timescale of an NMR experiment. They have been the subject of extensive research due to their involvement in a range of diseases, particularly cancers and neurodegenerative disorders, and also due to their mesomorphic biophysical characteristics. IDPs exist in a dynamic ensemble of conformations that can overcome the weak energy barriers separating compact and extended states, allowing interconversion on a very fast timescale. It is important to consider that IDPs in physiological-like conditions behave differently to structured proteins in denaturing conditions; they tend to have a smaller hydrodynamic radius than denatured proteins due to transiently-populated secondary structural elements, and other non-covalent interactions, both of which can lead to fluctuating tertiary structures. The primary sequence of an IDP encodes a very complex ensemble that includes a multitude of elements resulting in potentially foldable, partially foldable, differently foldable or not foldable at all protein segments [1]. This allows one of the most important functional characteristics of IDPs; to be able to bind to multiple partners in the cell in complex cell signaling networks.

The way in which intrinsic disorder is encoded into the primary sequence of proteins has been the subject of much consideration. The net charge per residue (NCPR, defined as $NCPR = f_+ - f_-$ where $f_+$ and $f_-$ are the fractions of positively and negatively charged residues, respectively) is a useful parameter to predict whether a polyelectrolytic IDP will form a globule or a swollen coil [2]. However, most IDPs are polyampholytic [3] (contain both positively and negatively charged residues) rather than polyelectrolytic (contain either positively or negatively charged residues) and the NCPR is therefore inadequate to describe the sequence-ensemble relationships of such proteins.

Das and Pappu [4] have proposed that both the fraction of charged residues (FCR, defined as $FCR = (f_+ + f_-)$) and the linear sequence patterning of oppositely charged residues (the $\kappa$-value) will influence the plasticity of an IDP. This patterning
parameter, $\kappa$, was introduced to describe the different sequence variants based on the linear sequence distributions of oppositely charged residues. $\kappa$ values lie between 0 and 1; low values relate to well-mixed sequences of positive and negative residues and at $\kappa=1$ oppositely charged residues are segregated in the linear sequence. Metropolis Monte Carlo simulations were used to demonstrate that low $\kappa$-values give rise to a protein with high inherent flexibility: Intrachain electrostatic repulsions and attractions are balanced resulting in a random-like chain. An increase in $\kappa$-value corresponds with a reduction in conformational flexibility of a protein, a result of long range electrostatic attractions between oppositely charged blocks of same-charge residues resulting in hairpin-like conformations.

Mass spectrometry is an extremely useful technique in the structural investigation of IDPs [5]. Electrospray ionisation (ESI) desolvates proteins from solution to the gas phase leaving them with differing net charge. The ensuing distribution of charged forms of the solvent free protein provides insight into the different conformations that are present in solution [6]. Compact proteins will carry a low number of charges since many of the chargeable residues will be buried within the protein which will protect from protonation by the solvent. Extended structures, however, have a much larger solvent-accessible solvent area on which protonation can take place so will give rise to high charge states in the mass spectrum. Because IDPs generally exist in both compact and extended conformations they tend to give rise to a wide charge state distribution (CSD). MS is therefore a useful tool which provides information on the range of transiently populated conformations a disordered protein exists in, since no time or spatial averaging occurs. Experiments in this paper utilise nano-ESI (nESI) which requires smaller sample quantities, produces smaller droplets than those obtained with ESI and requires lower spray potentials and source temperatures which help to maintain native-like interactions that are present in solution. Working against this are the increased Coulombic forces that are 80 times stronger in a vacuum than in pure water. This renders highly charged ions prone to extend due to repulsive interactions between proximal like charges no longer screened by interdispersed salt ions and solution.

Ion mobility (IM) is a complementary technique to MS and these are frequently coupled as a hybrid technique (IM-MS). Ion mobility allows separation of ions on the basis of their charge and size. In a typical experiment ions are pulsed into a drift tube, across which is applied a weak electric field which draws the ions through the tube at a velocity which is proportional to the net charge on the ion. The ions are
hindered by collisions with an inert buffer gas which is held at a known pressure and
temperature, and the ions traverse the cell at a speed which is determined by their
mobility. Two parameters which contribute most to the mobility of an ion are the
rotationally averaged collision cross section (CCS) and the charge present upon it: a
higher charge on an ion means that it will be drawn through the drift tube faster,
while ions with a larger CCS will be hindered due to more collisions with the buffer
gas which slow the ions down. By measuring the time that the ions take to traverse
the cell, and subsequently the charge that is present on the ion via mass
spectrometry, it is possible to calculate the mobility of the ion, and from this one can
directly calculate the rotationally averaged CCS.

Here, IM-MS data are presented on C-terminal p27 (p27-C) ionised from different
ionic strength solutions; the conformations we observe in the gas phase are strongly
dependent upon the solution conditions from which they are sprayed. We also
compare IM-MS results of two permutants (Table 5.1); one that has a lower κ-value
than the WT and is therefore expected to be less compact, and another that has a
higher κ-value than the WT and is therefore expected to be more compact. We
report that our IM-MS experiments are able to effectively differentiate the
permutants, and that our results follow the trend that has been observed in solution-
phase experiments. CCSs were calculated from structures obtained during MC
simulations; here the trend for the permutants also applies. Comparison of
experimental and calculated CCS values provides evidence for a collapse of the gas
phase molecules; we observe consistently smaller conformations in vacuo than are
seen in solution or calculated by MC simulations in implicit solvent.

<table>
<thead>
<tr>
<th>WT</th>
<th>GSHMKGACKV PAQESQDVSG SRPAAPLIGA PANSEDTHLV DPKTDPSDSQ TGLAEQQCAGI RKRPATDDSS TQNKRANRT E NVSDGSPNA GSVQETPKQP GLRRRTQ</th>
</tr>
</thead>
<tbody>
<tr>
<td>κ0.14</td>
<td>GSHMKGACKS SSPPSNDQGR PGDEPDQVIDK TEVERTQDTS NIQEQTQSNR SGPDKPSRC D LAVSGVAAA A LPGANIST A RDLTRDEEE GSVQETPKQP GLRRRTQ</td>
</tr>
<tr>
<td>κ0.56</td>
<td>GSHMKGACS SVLGTGNPR QARVSDTSL EDDDEQCODST PDEVVSACTI VASALDINAA TRSPKPASPK RRKRRQSTAP AQGNEPPGNA GSVQETPKQP GLRRRTQ</td>
</tr>
</tbody>
</table>

Table 5.1 Amino acid sequences of C-terminal p27, and the two
permutants. Acidic residues are coloured red and basic residues are coloured
green.
5.3 Results

5.3.1 Distribution of charges is revealed by mass spectrometry of p27-C permutants sprayed from solvents of different ionic strengths

Figure 5.1 shows MS data of p27-C-WT, p27-C-κ0.14 and p27-C-κ0.56. All proteins were sprayed from solutions of both low and high ionic strength (10 mM and 200 mM ammonium acetate respectively). Differences were observed in the MS of all permutants and solution conditions. p27-C-WT produces ions ranging from [M+6H]^{6+} to [M+17H]^{17+} when sprayed from 10 mM ammonium acetate, giving a median charge state of 11.5 and a charge state range (Δz) of 12 (Figure 5.1a). As shown in figure S1 (appendix 3) there is a trimodal charge state distribution with Gaussian distributions centred at z=6.5, z=9 and z=11.6. The highest intensity peaks correspond to the higher charge states, with the Gaussian curve centred on z=11.6 occupying 76% of the total curve area (Table S1, appendix 3). A higher ionic strength starting solution (200 mM ammonium acetate) does not alter the median z or Δz of p27-C-WT (Figure 5.1b), but stark differences are observed in the shape of the charge state distribution. The lowest charge states increase dramatically in relative intensity; the distribution centred at z=6.5 increases from 11% to 62% of the total area (Table S1, appendix 3).
Figure 5.1  Mass spectra of p27-C permutants sprayed from low and high ionic strength solutions. p27-C-WT (a and b), p27-C-κ0.14 (c and d) and p27-C-κ0.56 (e and f). All permutants were sprayed from ammonium acetate solutions of 10 mM (a, c and e) and 200 mM (b, d and f) ammonium acetate. Peaks annotated with a red asterisk are due to a contaminant peptide with a mass of 4870 Da.

p27-C-κ0.14 has a κ-value of 0.14 which is lower than that of the WT (κ=0.31). This lower κ-value means that oppositely charged residues are more evenly distributed in the linear sequence. p27-C-κ0.14, like the WT, has a median charge state of 11.5 and a Δz of 12 but there is now a tetramodal charge state distribution with curves centred around z=6.5, z=8.9, z=9.9 and z=11.3 (Figure S1, Table S1). Changes in
Chapter 5    How the distribution of charges distributes conformation

the charge state distribution as a result of differing ionic strength of the starting solution are minimal. There is a slight relative increase in the area of the curves of the two lowest charge distributions when sprayed from 200 mM ammonium acetate but the largest change corresponds to less than 5% of the total curve area.

The p27-C-κ0.56 permutant has a κ-value of 0.56, meaning that residues of the same charge are mainly located adjacent to one another in the linear sequence. This protein displays a very different charge state distribution to the other permutants; when sprayed from low ionic strength the median charge is reduced to 9.5 and the Δz reduced to 8 with just very low intensity in the high charge states. The CSD is trimodal with the Gaussian curves centred at z=6, z=8.1 and z=11.2 with relative areas of 30%, 66% and 4%, respectively. In contrast to the previously discussed permutants an increase in ionic strength shifts the CSD in favour of higher charge states compared to the lower ionic strength; the Gaussian curve at z=6 decreases from 30% to 3% of the total area, while the curve centred at z= 8.1 increases from 66% to 93% of the total area.
5.3.2 Ion mobility shows the distribution of conformation in each of the charge states of p27-C-WT sprayed from solvents of different ionic strengths

![Graph showing CCS distribution for different charge states of p27-C-WT sprayed from 10 mM and 200 mM ammonium acetate solutions.]

Figure 5.2 p27-C-WT analysed by IM-MS. CCS distributions of each charge state sprayed from 10 mM ammonium acetate (grey) and 200 mM ammonium acetate (purple).

Ion mobility data of p27-C provides an extra dimension to the MS experiments; it provides the rotationally averaged CCS distribution for each charge state. The IM-MS results of p27-C-WT sprayed from 10 mM ammonium acetate will first be discussed (grey curves, figure 5.2). The lowest charge state, [M+6H]$^6^+$, is present in a single conformational family centred around 1100 Å$^2$. [M+7H]$^7^+$ contains two conformational families; one at around 1100 Å$^2$ similar to [M+6H]$^6^+$, and a larger one centred around ~1400 Å$^2$. All charge states between [M+7H]$^{7^+}$ and [M+13H]$^{13^+}$ correspond to two conformational families, which have increasing CCSs as the charge increases. The more extended conformational family is dominant in all cases. The smaller conformation of each charge state is roughly the same size as the larger conformation of the charge state below it, indicating that there are several discrete conformational families in the gas phase. This implies that the disorder is
not absolute and that weak intramolecular interactions provide low energy barriers in the conformational landscape of the protein. For all charge states from $[M+14H]^{14+}$ upwards just one conformational family is present which is around $\sim 2500 \text{ Å}^2$; the CCSs cease to increase with charge. We propose that at $[M+14H]^{14+}$ the protein is in its most extended conformation and any addition of protons beyond this has a negligible effect on the structure in terms of Coulombic repulsion.

There are dramatic differences in the CCS distributions of ions produced from 200 mM ammonium acetate (purple curves) compared to 10 mM. For example, while at 10 mM ammonium acetate the $[M+7H]^{7+}$ ion has a CCS distribution spreading from 870-1820 Å$^2$ consisting of two conformational families, at 200 mM ammonium acetate this is reduced to one conformational family with a spread of 850-1430 Å$^2$. The larger conformation is no longer present. This narrowing of CCS distributions can be seen for every charge state up to $[M+14H]^{14+}$. Again, just one conformational family is present for charge states $[M+15H]^{15+}$ upwards, which is around 2500 Å$^2$.

IM-MS data for p27-C-κ0.14 can be seen in Figure S2 (Supplementary information, Appendix 3). At 10 mM ammonium acetate the CCS ranges for each charge state are similar to the WT under equivalent conditions. However, the shape of the distributions for p27-C-κ0.14 is less well defined into separate conformational families indicating that movement between conformations is even less energetically constrained than the wild-type. For example, $[M+9H]^{9+}$ has a CCS range from $\sim 1100$-2100 Å$^2$ for both the WT and κ0.14 when sprayed from 10 mM ammonium acetate. This is divided into two conformational families for the WT centred around 1300 Å$^2$ and 1750 Å$^2$. For κ0.14 this is one large unresolved conformational family. As for the WT protein, the CCSs cease to increase above $[M+14H]^{14+}$. The conformations, however, reach a larger CCS than the WT; 2700 Å$^2$ as opposed to 2500 Å$^2$.

When p27-C-κ0.14 is sprayed from higher ionic strength, some of the charge states have a CCS distribution that is constricted in width indicating less conformational disparity, but the differences are less extreme than for the WT protein. For example, $[M+12H]^{12+}$ narrows from a range of 1700-2800 Å$^2$ to 1700-2500 Å$^2$. Interestingly, the $[M+9H]^{9+}$ charge state is resolved back into two conformations, centred around 1350 and 1700 Å$^2$, similar to that of the WT. The $[M+10H]^{10+}$ ions produced from a solution of high ionic strength are in fact larger in CCS than those from low ionic strength. Reasons for the alteration of response to solution conditions will be discussed later in the paper.
An interesting feature of κ0.14 is the similarity of CCSs of \([M+6H]^{6+}\), \([M+7H]^{7+}\) and \([M+8H]^{8+}\) charge states, with a large jump to the CCSs of \([M+9H]^{9+}\). This conformational switch between \([M+8H]^{8+}\) and \([M+9H]^{9+}\) is also reflected in the mass spectra; there is a large change in intensity between these two charge states with the \([M+9H]^{9+}\) being much more intense. This could either be due to a conformation resulting in the \([M+9H]^{9+}\) ion being more highly populated in solution, or the ionisation of this conformation being more favourable in the electrospray process.

IM-MS results of p27-C-κ0.56 are very different to that of the WT or κ0.14; while the overall CCS range is much narrower for κ0.56 (900 – 2500 Å² when sprayed from 10 mM ammonium acetate) the CCS ranges for each charge state are incredibly wide, and the increases in CCS with respect to charge state are very small. There is very little difference between CCS distributions when κ0.56 is sprayed from solutions of different ionic strength; just a very slight increase at \([M+10H]^{10+}\) and \([M+11H]^{11+}\).

Figure 5.3 IM-MS data of p27-C-WT in 10 mM ammonium acetate (a) and 200 mM ammonium acetate (b). The CCS distribution of each charge state is shown. The height or each curve is normalised to that of the MS peak shown in figures 5.1a and 5.1b.
Figure 5.3 shows the IM-MS data of p27-C-WT in a waterfall plot. The height of the CCS distribution for each charge state is normalised to the relative intensity of the MS peak of the corresponding spectra in Figure 5.1. This allows us to view with ease the distribution of ions according to their CSD, and their CCS distribution within each charge state. We can also compare the distribution of ions easily as a function of the solution from which they were sprayed. When the starting solution of the protein is 10 mM ammonium acetate the most populated region is that corresponding to ions carrying between 10 and 13 positive charges with CCSs of 1500 – 2500 Å². Multiple conformational families can be observed for most charge states, especially when \(7 \leq z \leq 9\). There is also a gradual increase in CCS along with charge. When p27-C-WT is sprayed from 200 mM ammonium acetate, starkly different features are observed. While there is still the gradual increase of CCS values with increasing charge (a general feature of such experiments attributable to Coulombic repulsion in the gas phase as well as the fact that more extended forms in solution will have more protonatable sites) the most populated area is that corresponding to charge states \(6 \leq z \leq 7\) that have a CCSs of 800 – 1500 Å². As mentioned regarding Figure 5.2, the CCS distributions are narrower for each charge state. So, not only are the lower charge states more populated at high salt concentrations, they are also more compact.

Such a plot is an effective way to visualise the distribution of ions across the CCS landscape. These were also constructed for p27-C-κ0.14 and p27-C-κ0.56 as seen in figure S4. p27-C-κ0.14 can be thought of as a very diffuse system with wide CCS distributions for each charge state and no resolvable conformational families. Some small differences occur upon changing the ionic strength of the starting solution, mainly around the charge states \(7 \leq z \leq 10\) where there is quite a jump in intensity between the \([M+8H]^{8+}\) and the \([M+9H]^{9+}\) ions. p27-C-κ0.56 is a highly unusual system; there are fewer charge states than for the other permutants, but each charge state covers a wider CCS distribution. In contrast to the wildtype, increasing the ionic strength of the starting solution causes an increase in the population of higher charge states and higher CCSs, although the rise in CCS with charge state is less so than for WT. Possible reasons for this will be discussed in the later sections.
5.3.3 CCS values calculated from Monte Carlo (MC) simulations done with implicit solvent allow a comparison between solution conformation to those in vacuo.

CCS values that were calculated from MC simulations of all three permutants are shown in Figure 5.4 and Table 5.2. As well as clear differences in terms of the tabulated data, it can be seen by eye that the histogram for κ0.56 is much narrower than for the other permutants and that most of the intensity corresponds to a lower CCS value. The differences between the WT and κ0.14 are less stark; the apex is at a higher CCS for κ0.14 but the shape is narrower than the WT.

![CCS distributions calculated from molecular coordinates obtained from Monte Carlo simulations of the permutants in implicit solvent. Grey; p27-C-WT. Red; p27-C-κ0.14 and Blue; p27-C-κ0.56.](image)

<table>
<thead>
<tr>
<th>Construct</th>
<th>Apex/ Å²</th>
<th>Lower CCS/ Å²</th>
<th>Higher CCS/ Å²</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT</td>
<td>2190</td>
<td>1600</td>
<td>2850</td>
</tr>
<tr>
<td>κ0.14</td>
<td>2200</td>
<td>1700</td>
<td>3000</td>
</tr>
<tr>
<td>κ0.56</td>
<td>2100</td>
<td>1700</td>
<td>2700</td>
</tr>
</tbody>
</table>

Table 5.2 CCS parameters calculated from MD simulations of the three p27-C permutants.
5.4 Discussion

A great deal of information can be provided on the dynamic properties of a protein by nESI-MS alone (Figure 5.1). We have previously reported that structured proteins up to 150 kDa tend to present in a maximum of seven charge states, whereas disordered proteins, or proteins containing regions of disorder, tend to present in seven charge states or more [6]. Additionally, a wide CSD with significantly higher intensity peaks at lower charge states indicates a protein that has both regions of structure and disorder. Using this analysis, it can be inferred from MS data that p27-C-WT is very dynamic in 10 mM ammonium acetate (Figure 5.1). It is present in 12 charge states that can be fitted by a trimodal Gaussian distribution, with the highest charge envelope being the most intense (Figure S1). Under identical conditions, p27-C-κ0.14 produces a similar charge state range (Δz=13), but with a tetramodal Gaussian distribution rather than the trimodal distribution that is fitted to the WT. This indicates that this construct exists in a more disperse range of conformations and can be classed as being ‘more disordered’ than the WT. The CSD of p27-C-κ0.56 is much narrower than the two other proteins (Δz=7) indicating less conformational flexibility; the high charge states that would represent the elongated shapes are not present in the spectrum. The CCS distributions of the permutants also reflect their extent of disorder. p27-C-WT has a ΔCCS = 1544 Å², p27-C-κ0.14 has a ΔCCS=1680 Å² and p27-C-κ0.56 has a ΔCCS=776 Å².

As the κ-value of the protein increases (the κ-value for the WT is 0.31), both MS and IM-MS results suggest that conformational disparity of the protein decreases. This corresponds with theoretical predictions made by Das et al. [4] who postulated that proteins with alternating positive and negative charges (low κ-value) would behave similarly to a self-avoiding random walk distribution due to the counterbalancing of intrachain electrostatic interactions and repulsions. As the distribution of charges becomes more asymmetric (higher κ-value), oppositely charged blocks of residues act as counterion clouds for each other, leading to electrostatic interactions resulting in reduced dynamics of the protein conformations.

All of the constructs behave differently upon an increase in ionic strength of the initial solution, and these results require more careful consideration. p27-C-κ0.56 behaves as expected for a protein with blocks of same-charge residues, undergoing an expansion at 200 mM ammonium acetate, c.f. 10 mM. This can be attributed to shielding of the long-range intrachain electrostatic interactions that cause the
collapse at lower salt concentration. In contrast p27-C-κ0.14 which is the well-mixed sequence shows negligible salt dependence which is explained by the intrachain electrostatic attractions and repulsions being counterbalanced. Shielding of charges therefore has minimal effect.

The behaviour of p27-C-WT when sprayed from solutions of differing ionic strength is both stark and unexpected. Upon increasing the ionic strength the intensity moves to the lowest charge states. Such a large degree of collapse is very much unexpected and is in disagreement with predictions from previous MC simulations for a protein of this κ-value. Whilst a small compaction in the solution-phase conformation may be expected [7], this cannot account for such a large shift and to explain this we must consider the ionisation mechanisms that give rise to the low and high charge states of unfolded proteins.

The charge residue mechanism (CRM) is a well-established mechanism for the desolvation of large and relatively lowly charged macromolecular species which are ‘structured’. Here, Rayleigh-charged nanodroplets that contain a single analyte evaporate to dryness, losing charge as the droplet shrinks via fission events. Remaining protons are transferred to the protein during the final stages of desolvation, resulting in a compact gas-phase ion with a low number of charges. This is the ionisation mechanism responsible for the lowly charged ions. The chain ejection mechanism (CEM) was recently proposed by Konermann et al. [8] as the ionisation mechanism for polypeptides that are unfolded in solution. This involves expulsion of one terminus of an unfolded protein from the droplet into the vapor phase, followed by stepwise ejection of the remaining chain and separation from the droplet. This results in a highly-charged protein ion with an extended gas-phase conformation. We also proposed an intermediate process to account for protein conformations containing regions of both structure and disorder [9]. Here, an extended part of the protein will be ejected from the droplet as in the CEM, but this will be an unfavourable way to remove a compact region of the protein from the droplet which is more likely to undergo the CRM. The differing amount of the polypeptide which is in an extended or compact state will therefore give rise to intermediate charge states, and also multiple conformations within the same charge state.

We propose that the increase in abundance of lower charge states of p27-C-WT at 200 mM ammonium acetate is predominantly due to the high salt concentration in the nanodroplet causing the CEM to be unfavorable for this particular protein,
resulting in preferential ionisation via the CRM and therefore a high intensity of lower charge states. The maximum number of charges that a spherical conformation of a protein this size can hold is 8.2 (as determined by De la Mora’s interpretation of the Rayleigh limit [10]) so it can be assumed that the cut-off point between CRM and CEM is at charge state z=8. Interestingly, all spectra for p27-C-WT and p27-C-κ0.14 display a large jump in intensity between z=8 and z=9, and we attribute this to the different ionisation mechanisms that the differently charged ions are a result of. For p27-C-κ0.56 the cut-off point appears to be lower, between z=7 and z=8 according to the switch in intensity. So, it would appear that a high ionic strength solution causes p27-C-WT to favour the CRM over the CEM. This could be in part due to an alteration of conformation of the protein in the droplet towards a compact state, and the salt environment of the droplet is also likely to contribute to this effect. A question is why this only happens to the WT protein, and we postulate that it could be due to evolution of controlled flexibility, allowing modulation of dynamic behaviour either by salt concentration, or another effect which is mimicked by ionic strength alterations. Indeed, C-terminal p27 is phosphorylated by the Cdk-cyclin complex in the cell which alters its dynamic behaviour, and this modulation of flexibility could be an inherent property of the wild-type protein that we observe in these experiments.

From comparison of IM-MS results with calculated CCSs from MC simulations and solution-phase SAXS data it is apparent that the proteins undergo a gas-phase compaction within the mass spectrometer. The lowest CCSs observed via MC for any of the permutants is 1600 Å², while that of IM-MS results is below 1000 Å². This collapse is not surprising since the simulations were run in the presence of implicit solvent which has a dielectric constant of 78, which will weaken any long range interactions compared to the vacuum which has a dielectric constant of 1. The extent of collapse is different for all permutants, and is dependent upon the solution from which they are sprayed. For the wildtype protein it is the most extreme, especially at 200 mM ammonium acetate. For some reason at these conditions we are favouring the ionisation of the collapsed conformations.

5.5 Conclusions

In this paper we have demonstrated the suitability of IM-MS to compare qualitatively the conformational diversity of a set of proteins that differ only in the location of charged residues in the primary sequence. The differences we observe in the
permutants by IM-MS are clearer than are shown by MC simulations or solution-phase data. This research has allowed us to assess the extent that the solution-phase conformations are transferred into the gas phase.

5.6 Materials and Methods

5.6.1 Protein preparation

Proteins were provided by Richard Kriwacki, St. Jude Children's Research Hospital, Memphis, TN. p27-C constructs were generated by insertion of synthetic DNA sequences (Integrated DNA Technologies) into a pET28a vector (Novagen). All variants were generated using the QuickChange II XL Site-Directed Mutagenesis Kit (Stratagene). p27 variants were expressed in E. coli, purified by Ni²⁺ affinity chromatography, His-tags removed by cleavage with thrombin or TEV, and further purified by reverse phase HPLC. p27-C-κ0.56 has an internal thrombin site and therefore the His-tag cleavage site was mutated to a TEV site. The purified proteins were buffer exchanged into the relevant concentration of ammonium acetate using Bio-Rad micro Bio-Spin P-6 Columns.

5.6.2 Nano-electrospray ionisation

n-ESI was employed for all MS and IM-MS experiments. 30 μM protein solutions, in either 10 mM or 200 mM ammonium acetate as outlined in the text, were ionised from a thin-walled glass capillary (inner diameter 0.9mm, outer diameter 1.2mm, World Precision Instruments, Stevenage, UK) that was pulled to a nESI tip in house with a Flaming/Brown micropipette puller (Sutter Instrument Co., Novato, CA). A positive potential of 1.6 kV was applied to the solutions via a thin platinum wire.

5.6.3 Mass Spectrometry

All MS experiments were performed on a Q-ToF Global (Waters, Manchester, UK). The following instrument parameters were utilised; Cone voltage; 60V. Source pressure; 2.7 mbar. Source temperature; 80°C. Collision cell pressure; 2.3e-3 mbar. Collision energy; 5V.

5.6.4 Ion mobility-mass spectrometry

IM-MS experiments were carried out on a Waters Q-ToF I instrument that was modified in house to include a 5.1cm drift tube which has been described elsewhere.
The temperature and pressure of helium in the drift cell were approximately 28˚C and 4 Torr respectively. Measurements were made at 6 different drift voltages from 60V to 20V. The precise pressure and temperature was recorded for every drift voltage and used in the calculations of CCSs. Each experiment was performed in triplicate. Ion arrival time distributions were recorded by synchronisation of the release of ions into the drift cell with mass spectral acquisition. The CCS distribution plots are derived from raw arrival time data using Equation 5.1 below [12].

$$\Omega_{avg} = \frac{(18\pi)^{1/2}}{16} \left[ \frac{1}{m_b} + \frac{1}{m} \right]^{1/2} \frac{ze}{(K_BT)^{1/2} \rho L^2} \frac{1}{t_d V}$$  \hspace{1cm} \text{Equation 5.1}

Where \(m\) and \(m_b\) are the masses of the ion and buffer gas, respectively; \(z\) is the ion charge state; \(e\) is the elementary charge; \(K_b\) is the Boltzmann constant; \(T\) is the gas temperature; \(\rho\) is the buffer gas density; \(L\) is the drift tube length; \(V\) is the voltage across the drift tube; and \(t_d\) is the drift time.

The raw arrival time output \(t_a\) includes time the ions spend outside of the drift cell but within the mass spectrometer, known as the dead time \(t_0\). The value for \(t_0\) is calculated by taking an average value of the intercept from a linear plot of average arrival time versus pressure/temperature and was subtracted from the arrival time to calculate drift time \(t_d\):

$$t_0 = t_a - t_d$$  \hspace{1cm} \text{Equation 5.2}

All MS and IM-MS data were analysed using Masslynx v4.1 software (Waters, Manchester, UK), Origin v8.5 (Originlab Corporation, USA) and Microsoft Excel.

5.6.5 Monte Carlo Simulations

All simulations and analyses were performed using the CAMPARI molecular modeling software suite (http://campari.sourceforge.net) using the ABSINTH implicit solvation model and force-field paradigm [13]. The simulations were carried out using spherical boundary conditions. In each simulation, the system comprised the polypeptide chain, neutralizing Na\(^+\), Cl\(^-\) ions plus 49 excess ion pairs to mimic 10 mM NaCl enclosed within a spherical droplet of radius 125 Å. The choice for the droplet radius was justified using the end-to-end distance distributions for all sequences simulated in the excluded volume (EV) limit. Metropolis Monte Carlo simulations were performed at a simulation temperature of 298 K.
5.6.6 Calculation of CCSs from models

Theoretical CCSs of models obtained from MC sampling were calculated in MOBCAL [14] using the exact hard sphere scattering method [15]. 700 representative models were selected of each permutant for the CCS calculations.

5.7 References

The Distribution of Charged Residues in the C-terminal Domain of p27 Modulates the Overall Conformation of the Cyclin A/ Cdk2/ p27 Complex.
6.0 Declaration

This chapter consists of one draft paper awaiting submission: Rebecca Beveridge, Rahul Das, Yongqi Huang, Rohit V. Pappu, Richard Kriwacki and Perdita E. Barran. The distribution of charged residues in the C-terminal domain of p27 modulates the overall conformation of the Cyclin A/ Cdk2/ p27 complex.

As first author on this publication I performed the MS and IM-MS research, contributed to the calculations the CCSs of the crystal structures and drafted the manuscript. JH expressed and purified the protein samples.
Chapter 6  Cdk2/ Cyclin A/ p27 complex

Draft awaiting submission

The distribution of charged residues in the C-terminal domain of p27 modulates the overall conformation of the Cdk2/ Cyclin A/ p27 complex.

Rebecca Beveridge¹, Rahul Das², Rohit V. Pappu², Richard Kriwacki³ and Perdita E. Barran*.

¹ Manchester Institute of Biotechnology, Michael Barber Centre for Collaborative Mass Spectrometry, 131 Princess Street, University of Manchester, Manchester. M1 7DN.

E-mail: perdita.barran@ed.ac.uk

² Washington University in St. Louis School of Engineering & Applied Science, Department of Biomedical Engineering, Whitaker Hall, Campus Box 1097, One Brookings Drive, St. Louis, Missouri 63130.

³ Structural Biology, MS 311, Room D1024F, St. Jude Children's Research Hospital, 262 Danny Thomas Place, Memphis, TN 38105-3678.

Supporting Information

Protein sequences, materials and methods, and additional mass spectra and arrival time distributions can be found in the Supporting Information. This material is available in appendix 4 of this thesis.
6.1 Abstract

A common structure function relationship that has been postulated for intrinsically disordered proteins is their ability to undergo a ‘disorder to order’ transition, in whole or in part, upon binding to their partners in the cell. One such example of that is the binding of p27 to the Cdk2/ Cyclin A complex which is accompanied by partial folding of p27 with the retention of some dynamic behaviour, particularly in the C-terminal domain. Here we use native mass spectrometry coupled with ion mobility to investigate how the inherent dynamic properties of isolated C-terminal p27 (p27-C) modulate the conformational behaviour of the Cdk2/ Cyclin A/ p27 complex. We examine the change in collision cross section of the Cdk2/Cyclin A/ p27 complex formed with wildtype (WT) p27 compared with two C-terminal permutants to test a ‘folding upon binding’ hypothesis. The two permutants have been modulated to form a more (k=0.14) and a less (k=0.56) conformationally dynamic form of the WT protein (k=0.31). It is observed that the extended mutant contributes to a more extended trimeric complex, and that the WT p27 folds upon incorporation into the complex into smaller conformations. It is demonstrated that the dynamic behaviour of the 11 kDa C-terminal p27 domain is able to modulate the behaviour of the 87 kDa Cdk2/ Cyclin A/ p27 complex, and that p27 transfers its conformational dynamics to the intact protein complex.

6.2 Main Text

Intrinsically disordered proteins (IDPs) are a group of proteins that have a flexible structure that allows them to adopt many conformations under physiological conditions. They are dynamic and can interconvert between compact and extended geometries on a nanosecond timescale [1]. An interesting and important characteristic of IDPs is their ‘binding promiscuity’; their ability to bind to many different partners. The ability of IDPs to be molded by their binding partner ensures that binding to multiple targets will proceed selectively and with optimised binding rates [2]. This allows them to play a key role within cellular signaling networks [3-6].

The extent of disorder can differ between IDPs, and this has been shown to depend on the pattern of charged amino acids in the linear sequence [7]. Alternating positive and negative residues results in a very flexible and dynamic protein since the electrostatic repulsion between the same charges will roughly equate to the electrostatic attraction between opposite charges, resulting in a chain that has
similar properties to that of a random coil [7]. Proteins containing blocks of same-charge residues will be influenced by short-range repulsion and long-range attractions between oppositely charged regions, resulting in a hairpin-like conformation. The linear sequence patterning of oppositely charged residues can be described by the \( \kappa \)-value. This patterning parameter, \( \kappa \), was introduced to describe the different sequence variants based on the linear sequence distributions of oppositely charged residues. \( \kappa \) values lie between 0 and 1; low values relate to well-mixed sequences of positive and negative residues and at \( \kappa = 1 \) oppositely charged residues are segregated in the linear sequence. Initially, Metropolis Monte Carlo simulations were used to demonstrate that low \( \kappa \)-values give rise to a protein with high inherent flexibility, while an increase in \( \kappa \)-value corresponds to a reduction in conformational flexibility of a protein [7].

Experimental evidence for the trend between the \( \kappa \)-value of a protein and its conformational preferences was provided in the investigation of C-terminal p27 (p27-C) along with two permutants; the permutants contain the same amino acids as the wild-type protein, but they are arranged in a different order to modulate the \( \kappa \)-value. The wild-type protein has a \( \kappa \)-value of 0.31, and the two permutants have \( \kappa \)-values of 0.14 and 0.56 so are predicted to show increased and decreased dynamic behavior compared to the wild-type, respectively. Indeed, native mass spectrometry (MS) experiments [8-10] indicated this trend when the proteins were sprayed from 10 mM ammonium acetate; the charge state distribution (CSD) presented by the \( \kappa 0.14 \) permutant depicted a more disordered protein than the wild-type, while that of the \( \kappa 0.56 \) permutant suggested a lower dynamic propensity than the other two constructs. The rotationally averaged collision cross section (\( ^{DT}CCS_{He} \)) distribution of each charge state, measured in ion mobility- mass spectrometry (IM-MS) experiments, provides information on the conformational distribution of each protein. This gave further evidence of the relationship between the \( \kappa \)-value and the conformational preferences of each construct. The \( \kappa 0.14 \) permutant has the widest \( ^{DT}CCS_{He} \) range (\( \Delta CCS = 1680 \text{ Å}^2 \)), indicative of the most conformationally diverse system of the three [9]. The wild-type protein has a slightly narrower \( \Delta CCS \) of 1544 \( \text{Å}^2 \), and that of the \( \kappa 0.56 \) permutant is narrower still at \( \Delta CCS = 796 \text{ Å}^2 \), indicating reduced conformational dynamics with respect to the other two permutants. Interestingly, all permutants respond differently to an increase in the ionic strength of the initial solution condition. When sprayed from a high ionic strength solution \( \kappa 0.14 \) remains the most dynamic; the high ionic strength has little effect on the CSD or \( ^{DT}CCS_{He} \) distribution. At high ionic strength the wild-type protein displays a
preference for the smaller conformations whilst retaining access to the extended shapes at low relative intensity. Most of the κ0.56 protein molecules now exist in a conformational family that is larger than that of the wild-type, but this permutant is totally unable to access the extended geometries, so for all it is bigger, it is less conformationally dynamic than the wild-type. The reasons for the differential response to higher ionic strengths are discussed in detail in Chapter 5.

Here, we extend the methodological approach to investigate how the location of charge in p27-C affects the conformation of the Cdk2/ Cyclin A/ p27 complex. p27 is an IDP that blocks the transition of a cell from G1 to S-phase by binding to the Cdk2/ Cyclin A complex via its N-terminal domain, thereby inhibiting the kinase activity of Cdk2 [11, 12]. Within the complex, the C-terminal domain of p27 retains its flexibility which has been found necessary to facilitate phosphorylation, resulting in its ubiquitin-dependent proteolysis and subsequent reactivation of the Cdk2/ Cyclin A complex and cell cycle progression [13]. The different sequences of C-terminal p27 are incorporated into the full length protein (Figure 6.1), and we performed native IM-MS experiments on the Cdk2/ Cyclin A/ p27 complexes to investigate how the different C-terminal p27 sequences influence the dynamic and structural properties of the trimeric complex.
Figure 6.1  Top; Conformational flexibility of the C-terminal domain of p27 in the p27/Cdk2/cyclin A complex. Superposition of p27/Cdk2/cyclin A structures at varying time intervals during a 12.7-ns MD simulation. Bottom; Schematic of the complex containing CDK2 (blue), Cyclin A (purple) and p27 (black) containing different C-terminal domains; wild-type (black), κ0.14 (red) and κ0.56 (blue).

MS data of Cdk2/ Cyclin A alone (64 456 Da), and in complex with each of the different permutants (86 796 Da), is shown in Figure 6.2. The experiments were performed in 200 mM ammonium acetate since a lower buffer strength was incompatible with MS analysis of the Cdk2/ Cyclin A complex. The dimeric complex of Cdk2/ Cyclin A presents in three charge states z=15 – 17, and the trimeric complex of Cdk2/ Cyclin A/ p27 is observed in four charge states, z=18 – 21. In the presence of p27, no Cdk2/ Cyclin A dimer remains, but some Cdk2 is displaced by p27 to form a Cyclin A/p27 complex. This provides insight into the mechanism of complex formation which complements previous results: Lacy et al. [14] discovered that binding of p27-KID to cyclin A occurs at a much higher rate than binding Cdk2 alone. Furthermore, binding of p27 to the Cdk2/ Cyclin A complex occurs at a rate similar to that observed for binding to Cyclin A alone. The rate of association (kₐ) of
the p27-KID + Cdk2 interaction is $5.12 \times 10^3$ M$^{-1}$ s$^{-1}$, the $k_a$ of p27-KID + Cyclin A interaction is $2.86 \times 10^6$ M$^{-1}$ s$^{-1}$ and that of the p27-KID + Cdk2/ Cyclin A interaction is $1.57 \times 10^6$ M$^{-1}$ s$^{-1}$. They attribute the slow kinase binding to p27-induced remodelling of Cdk2 that occurs upon association.

The DTCCS$_{He}$ distributions of each charge state of the Cdk2/ Cyclin A/ p27 complexes are shown in Figure 6.3. The DTCCS$_{He}$ distributions can be viewed separately in the supporting information (Figure S2, appendix 4). Gaussian curves are fitted to each DTCCS$_{He}$ distribution (blue lines), and the sum of these Gaussian curves is shown by the red line. The apex of this red line for each complex is reported in table 6.1. The Cdk2/ Cyclin A complex shows a very narrow DTCCS$_{He}$ distribution, ranging from 3100 – 4600 Å$^2$ ($\Delta$CCS=1500 Å$^2$) with only a minimal
increase in average $^{DT}\text{CCS}_{\text{He}}$ with increasing charge state. This corresponds to a very rigid tertiary structure with minimum dynamics [9]. The average $^{DT}\text{CCS}_{\text{He}}$ of this Cdk2/ Cyclin A complex (taken by the apex of the sum of the three curves, red line) is 3758 Å$^2$. Upon addition of p27-WT, the $^{DT}\text{CCS}_{\text{He}}$ increases in size to an average of 4614 Å$^2$. The $^{DT}\text{CCS}_{\text{He}}$ range also increases, now spanning 3500-6250 Å$^2$ from baseline values ($\Delta\text{CCS}=2750$ Å$^2$). The increase in $\Delta\text{CCS}$ is ~85% which is large for a mass increase of 35%. In comparison, the production of a Hepatitis B capsid protein homodimer (which was shown to be disordered) has a mass increase of 100% and an increase in the CCS range of just 67% [15]. These results reflect the retention of dynamic behaviour of p27 when it is part of the complex.

Figure 6.3 $^{DT}\text{CCS}_{\text{He}}$ distributions of each charge state of the Cdk2/ Cyclin A complex (a) and of the Cdk2/ Cyclin A/ p27-WT (b), Cdk2/ Cyclin A/ p27-κ0.14 (c) and Cdk2/ Cyclin A/ p27-κ0.56 (d). The red lines are the sum of the distributions of the four charge states of each complex (shown by the blue lines). The purple dashed line corresponds to the calculated CCS of the crystal Structure of the Cdk2/ Cyclin A complex [16]. The black dashed line corresponds to the calculated CCS of the crystal Structure of the Cdk2/ Cyclin A/ N-terminal p27 [17]. All samples were sprayed from 200 mM ammonium acetate, pH 6.8.
Table 6.1 Average $^{DT}\text{CSC}_{\text{He}}$ of each complex.

<table>
<thead>
<tr>
<th>Complex</th>
<th>Average $^{DT}\text{CSC}_{\text{He}}$/ Å$^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cdk2/ Cyclin A</td>
<td>3758</td>
</tr>
<tr>
<td>Cdk2/ Cyclin A/ p27wt</td>
<td>4614</td>
</tr>
<tr>
<td>Cdk2/ Cyclin A/ p27κ0.14</td>
<td>4876</td>
</tr>
<tr>
<td>Cdk2/ Cyclin A/ p27κ0.56</td>
<td>4716</td>
</tr>
</tbody>
</table>

The complex containing WT p27 has the smallest average $^{DT}\text{CSC}_{\text{He}}$. This is in agreement with p27-C sprayed from 200 mM ammonium acetate, where most of the molecules are in conformations that are smaller than those of other permutants. It is the endogenous binding partner of Cdk2/ Cyclin A, so it is conceivable that it will interact with the protein in a more specific way than the other permutants, even if just transiently. The average $^{DT}\text{CSC}_{\text{He}}$ is larger for the complex containing the κ0.14 construct than either of the other two permutants, reflecting the higher conformational flexibility of C-terminal p27-κ0.14. The complex containing κ0.56 has an average $^{DT}\text{CSC}_{\text{He}}$ distribution in between the other permutants, reflecting the C-terminal dynamics previously observed; most of the molecules have a larger $^{DT}\text{CSC}_{\text{He}}$ than the wild-type, and a narrower $^{DT}\text{CSC}_{\text{He}}$ distribution than the κ-0.14.

The crystal structure of the 69 amino acid N-terminal ‘kinase inhibitory domain’ (KID) of p27 (residues 25-93 of full length p27) bound to the Cdk2/ Cyclin A complex was solved by Russo et al. [17]. Often in crystallography experiments the more dynamic regions of proteins are omitted since they are unable to form crystals. The CCS of this crystal structure, calculated by the trajectory method following minimisation, is 4628 Å$^2$, as shown by the black line on Figure 6.3. The calculated CCS of a protein from crystal structure coordinates tends to be larger than the measured CCS, attributed to the loss of solvent allowing partial collapse of the structure preserved in the crystalline form [18]. A comparison of the measured $^{DT}\text{CSC}_{\text{He}}$ with the calculated CCS of the crystal structure can be made for the Cdk2/ Cyclin A complex [16]; the average measured $^{DT}\text{CSC}_{\text{He}}$ is 3758 Å$^2$ while the calculated CCS is 4365 Å$^2$; a 23% increase for the crystal structure. If we assume that the calculated CCS value Cdk2/ Cyclin A/ p27-KID is overestimated by a similar amount, we can estimate that the gas-phase CCS is ~4000 Å$^2$; much smaller than the Cdk2/ Cyclin A/ p27. This suggests that most of the increase in $^{DT}\text{CSC}_{\text{He}}$ is due to the flexibility of the C-terminal domain. This could be tested via hydrogen-deuterium exchange experiments, in which the C-terminal domain would be predicted to have a higher uptake of deuterium due to higher solvent accessibility.
DT\textsubscript{CCSHe} distributions for p27-C, p27, the Cdk2/ Cyclin A complex and the Cdk2/ Cyclin A/ p27 complex are compared in Figure 6.4. MS and IM-MS data of p27 can be found in the supplementary information (figure S1, appendix 4). p27 has a wider DT\textsubscript{CCSHe} range than p27-C; the number of amino acids in the protein increases from 107 to 201, and the DT\textsubscript{CCSHe} range increases from 850-3250 Å\textsuperscript{2} (ΔCCS = 2400Å\textsuperscript{2}) to 1300-5900Å\textsuperscript{2} (ΔCCS = 4600Å\textsuperscript{2}). This indicates that p27 is dynamic along the full length of the protein in the unbound form; if the N-terminal domain was folded, only a small increase in ΔCCS with respect to the increase in mass would be observed.

Upon binding of p27 to Cdk2/ Cyclin A, the DT\textsubscript{CCSHe} range is reduced to 3450-6500 Å\textsuperscript{2} (ΔCCS = 3050 Å\textsuperscript{2}) despite the molecular mass of the trimeric complex being almost three times that of the isolated p27. This reflects the ‘folding-upon-binding’ effect where the contact free energy compensates the free energy required for the disorder-to-order transition [19]. The resulting trimeric structure shows a lower dynamic propensity than p27 alone, but higher than the Cdk2/ Cyclin A complex due to a degree of conformational diversity retained by p27. For comparison with a similar sized protein, the Concanavalin A tetramer is 100 kDa and has a DT\textsubscript{CCSHe} spread of 4000-6000 Å\textsuperscript{2} [20], narrower than that of Cdk2/ Cyclin A/ p27 despite the 10% increase in size. The figure also shows a comparison between the different permutants. It can be seen that the κ0.14 permutant accesses more extended conformations at the C-terminal level, the difference is mitigated at the full-length level but returns when bound to the Cdk2/ Cyclin A complex. The DT\textsubscript{CCSHe} spread remains the same as the wild-type, but the apexes of the peaks shift to higher CCSs. Potential explanations for this are that i) that the C-terminal domain of p27 tends to exist in a more extended conformation than the wild-type or ii) the C-terminal domain does not interact with the Cdk2/ Cyclin A complex as closely as the wild-type. The κ0.56 permutant has a much lower DT\textsubscript{CCSHe} spread than the WT at the C-terminal level, but has a very similar DT\textsubscript{CCSHe} spread to the WT at the full length and the trimeric complex. This further suggests that the wild-type protein is encoded to fold upon binding to the Cdk2/ Cyclin A, since the dynamics are reduced upon binding to a level similar to the κ0.56 permutant.
Figure 6.4  Average $^{ DT}_{\text{CCS}}$ values for each charge state (filled squares) and overall $^{ DT}_{\text{CCS}}$ spread (lines) of C-terminal p27, full length p27, the Cdk2/ Cyclin A complex and the Cdk2/ Cyclin A/ p27 complex. The WT protein is shown in black, the $\kappa_{0.14}$ variant is shown in red and the $\kappa_{0.56}$ variant is shown in blue. The data points have been offset with respect to the molecular mass which is the same for each set of permutants/ permutant complexes.

In this IM-MS study we show how the charged residue patterns of the 11.2 kDa p27-C affects the conformational behaviour of the 86.8 kDa Cdk2/ Cyclin A/ p27 complex. Native MS results confirm the stoichiometry of the protein complexes; Cdk2/ Cyclin A alone exists only as a dimer, and when p27 is added the trimeric complex is the predominant species, with some lower-intensity peaks corresponding to the Cyclin A/p27 complex. IM-MS detects a vast reduction in the $^{ DT}_{\text{CCS}}$ distributions of p27 as it binds to the binary complex indicating a reduction in dynamic propensity, in agreement with the ‘folding-upon-binding’ process that has been shown previously. Subtle changes are detected in the gas-phase conformations by IM-MS of the complexes containing the different permutants; the average $^{ DT}_{\text{CCS}}$ of the complex containing $\kappa_{0.14}$ is higher than that of the wild-type protein, whilst the complex containing $\kappa_{0.56}$ is similar to that of the wild-type. Overall we are able to monitor dynamic behavior of proteins and protein complexes. Native mass spectrometry (MS) coupled with ion mobility (IM-MS) is highly
appropriate to examine dynamic properties of proteins, in part since there is no apparent bias towards a folded structure [10]. IM-MS allows the visualisation of the shape of a given protein or protein complex via a $\text{DT}_{\text{CCS}}$ distribution, which provides direct information of the size and conformational variability of a given system. With IM-MS it is possible to separate multiple conformational states [11] for example to observe how individual conformers are affected by ligand binding or mutations [12]. This positions it as highly complementary to x-ray crystallography; it cannot provide atomistic detail but it can report on conformationally dynamic systems and heterogeneic stoichiometries all in a single experiment, which does not rely on successful crystal formation.

### 6.3 References

7
Mass Spectrometry Reveals how Fdc1 Dynamics are Influenced by Prenylated FMN Binding
This chapter consists of one research article drafted for publication: Beveridge, R., Payne, K. A. P., Scrutton, N. S., Leys, D. and Barran, P. E.. Mass Spectrometry reveals how Fdc1 dynamics are influenced by prenylated FMN binding.

As first author on this publication I performed the MS, IM-MS and HDX-MS research and drafted the manuscript. KAPP expressed and purified the protein samples.
Draft awaiting submission

Mass Spectrometry reveals how Fdc1 dynamics are influenced by prenylated FMN binding

Rebecca Beveridge, Karl A. P. Payne, Nigel S. Scrutton, David Leys, Perdita E. Barran

1The Michael Barber Centre for Collaborative Mass Spectrometry, The School of Chemistry, Manchester Institute for Biotechnology, University of Manchester, Manchester, UK

E-mail: perdita.barran@ed.ac.uk

Supporting Information

Additional mass spectra, arrival time distributions and HDX data can be found in the Supporting Information. This material is available in appendix 5 of this thesis.

7.1 Abstract

Fdc1 is a decarboxylase enzyme that requires the novel cofactor prenylated FMN for activity. Here we use top down and bottom up mass-spectrometry approaches to investigate conformational differences between apo- and holo-Fdc1. Both native Ion Mobility Mass Spectrometry (IM-MS) and Hydrogen Deuterium Exchange coupled to mass spectrometry (HDX-MS) show that the cofactor confers structural stability to the enzyme. HDX-MS reveals that while the dominant structural changes occur in a region proximal to the co-factor binding site, rearrangements on co-factor binding are evident throughout the protein predominantly attributable to conformational tightening, which is also shown in IM-MS data.

7.2 Introduction

Decarboxylation reactions are common in nature, despite the fact they are difficult to achieve under ambient conditions. The reaction is made possible by decarboxylases, often making use of cofactors including either organic molecules such as flavins, pyridoxal phosphate or thiamine pyrophosphate, and/or metal ions, for example Mg$^{2+}$, Fe$^{2+}$ or Mn$^{2+}$ [1]. The recently discovered prenylated flavin
cofactor that features in the Pad1/Fdc1 and UbiX/UbiD decarboxylase systems represents a new addition to this list [2, 3]. It has been previously demonstrated that both the fdc1 and pad1 genes are essential for the decarboxylation of phenylacrylic acids by spoilage yeasts and moulds such as S. cerevisiae and Aspergillus niger, however the precise role of each gene had remained enigmatic [4, 5]. Recently, Payne et al. have shown that Fdc1 is in fact the enzyme responsible for decarboxylation, however active recombinant protein can only be achieved by co-expressing it with either pad1 or the corresponding E. coli homologue ubiX [2]. The co-expressed Fdc1 protein (denoted Fdc1\textsuperscript{ubiX}) possesses distinct features in the UV-vis spectrum that are absent in single expressed Fdc1. Determination of the crystal structure of the A. niger Fdc1\textsuperscript{ubiX} revealed the presence of a modified flavin mononucleotide (FMN) cofactor which is in turn bound to the protein in complex with Mn\textsuperscript{2+} and K\textsuperscript{+}. This modified cofactor (prenylated FMN or prFMN) results from addition of a prenyl group to the N5-C6 atoms of FMN to form a fourth, non-aromatic ring. This prFMN cofactor supports decarboxylation of substrate by dipolar 1,3 cycloaddition, an unprecedented biological reaction that is reminiscent of click chemistry. Subsequent studies on UbiX/Pad have confirmed that these are responsible for prFMN synthesis [3]. While Fdc1\textsuperscript{ubiX} could be readily crystallised, no crystals were attainable from Fdc1 expressed in the absence of ubiX, despite many attempts. It is hypothesised that this is due to an increase in conformational freedom of the apo-protein, increasing the sampled conformational heterogeneity and decreasing the likelihood of crystallisation. In this work we present differences in the conformational dynamics of Fdc1 upon cofactor binding. In the absence of any crystallographic reference structure for the apo form we use complementary mass spectrometry-based approaches to determine the effect of prFMN binding; globally conformational change is assessed with ion mobility mass spectrometry (IM-MS) whereas hydrogen-deuterium exchange-MS (HDX-MS) allows the changes to be localised to regions of the Fdc1 dimer. With the use of nano-electrospray ionisation (nESI) [6] protein complexes can retain their native topology and stoichiometry upon transfer into the gas phase [7], an approach termed ‘native mass spectrometry’ [8]. Following desolvation from aqueous solution, the ensuing charge state distribution (CSD) provides mass and stoichiometric information and can be used to infer some conformational preference for the protein or complex [9]. Native mass spectrometry (MS) coupled with ion mobility (IM-MS) is highly appropriate to examine dynamic properties of proteins, in
part since there is no apparent bias towards a folded structure [10]. IM-MS allows the visualisation of the shape of a given protein or protein complex in a form known as a collision cross section distribution (CCSD), which provides direct information of the size and conformational variability of a given system. With IM-MS it is possible to separate multiple conformational states [11] for example to observe how individual conformers are affected by ligand binding [12]. This positions it as highly complementary to x-ray crystallography; it cannot provide atomistic detail but it can report on conformationally dynamic systems and heterogeneous stoichiometries all in a single experiment, which does not rely on successful crystal formation.

HDX-MS is complementary to IM-MS, it can also probe protein dynamics, allowing comparison between conformational changes observed in vacuo to those in vitro, and when coupled with enzymatic digestion of the protein post deuteration, provides structural information at a more localised level. A recent study carried out by Alverdi et al. [13] used a combination of native MS, IM-MS and HDX-MS to characterise further the conformational rearrangement of PKG in response to cGMP-dependent activation. While no structure had been solved for the enzyme, small angle X-ray scattering revealed an overall expansion of the protein structure upon cGMP binding, and Fourier transform infrared spectroscopy indicated that the conformational change is due to a topographical rearrangement of structural domains rather than changes of secondary structure. A combination of further low-resolution techniques had also shown previously that PKG activation is associated with protein elongation and a corresponding increase in net surface charge. In agreement with these results, native MS data showed a shift in the charge state distribution to higher charge states upon binding of cGMP, indicative of more solvent accessible residues available for protonation during the electrospray process and hence a more elongated structure. Ion mobility results show a longer drift time of the holo form with respect to the apo form of the protein, again suggesting that cGMP binding to PKG promotes a more extended structure. HDX experiments allowed the localisation of conformational change to the different domains; upon cGMP binding both the substrate binding region and part of the ATP binding domain became significantly more solvent exposed, while the cGMP binding domains underwent significant protection from the solvent. Data obtained from a combination of these MS techniques allowed the authors to propose a new structural model for the cGMP-induced activation of PKG.

By a combination of MS, IM-MS and HDX-MS we have sought to give insight into the differences in dynamic properties of Fdc1 compared with Fdc1\textsuperscript{UbiX} that are a
result of the binding of prFMN. This provides insight into the structural changes that can be a result of cofactor binding, and will outline potential reasons for the lack of crystallisation in the absence of the cofactor. The study emphasises the utility of combined MS approaches in dynamic structural science by identifying how prFMN imprints new dynamic properties in the Fdc1 protein. Both the approach and observations are of general significance within the context of the structure-defines-function paradigm.

7.3 Results

7.3.1 Native mass spectrometry reveals Fdc1 is primarily dimeric and confirms dimeric stoichiometry for cofactor-bound Fdc1\textsuperscript{UbiX}

![Native MS spectra of Fdc1 and Fdc1\textsuperscript{UbiX}](image)

Figure 7.1: Native MS of 5 μM Fdc1 (top) and 10 μM Fdc1\textsuperscript{UbiX} (bottom). Both samples were sprayed from 100 mM ammonium acetate, pH 6.8. Right hand spectrum; an enlarged view of [M+21]\textsuperscript{21+}. Measured mass of Fdc1 dimer is 112 265 Da (expected 112 270 Da). The predicted masses are shown by dashed lines. Figures adapted from Payne et al. [2]

Native mass spectra are shown in Figure 7.1 of Fdc1 (top) and Fdc1\textsuperscript{UbiX} (bottom), sprayed from physiological-like conditions (100 mM ammonium acetate, pH 6.8). Fdc1, present only in the apo form, presents in charge states [M+19H]\textsuperscript{19+} to [M+23H]\textsuperscript{23+}, with [M+21H]\textsuperscript{21+} being the most intense. The native mass spectrum of Fdc1\textsuperscript{UbiX} shows a narrowing in the charge state distribution; although the most
abundant ion is still [M+21H]^{21+}, the highest charge state is now [M+22H]^{22+}. This indicates a more compact form of the protein, with statistically less surface area available to provide an additional protonatable site. Further evidence for the stabilisation of the dimer by the cofactor is demonstrated in Figure S1 (appendix 5), which shows monomer peaks in the Fdc1 spectrum while all of the Fdc1^{UbiX} is present as a dimer. Whether this reflects the solution-phase properties or is due to some dissociation of the complex in the mass spectrometer, the fact that the monomer is only present in the Fdc1 spectrum indicates strongly that this dimer is less stable than that of the Fdc1^{UbiX}.

The Fdc1^{UbiX} spectrum in Figure 7.1 shows that most of the ions carry two non-covalently bound cofactors. Approximately one third are seen to be carrying one non-covalently bound cofactor and just a small amount of protein is present in the apo form. This might be due to incomplete cofactor occupancy in solution or to loss of the cofactor upon desolvation (vide infra). The measured mass of the Fdc1 dimer (112 265 Da) is close to that predicted (i.e. 112 270 Da: 2x the monomer mass of 56135 Da, without the initiator Met). The measured mass for the apo form of Fdc1^{UbiX} is 112 345 Da, slightly higher than for Fdc1, which is attributed to an increased retention of salt. The increase in mass of 80 Da corresponds to two potassium adducts. When one cofactor is bound the difference between the expected (112 870 Da) and measured masses is higher again: ~178 Da increase above what is predicted when the cofactor is in complex with the protein. This could be attributed to the addition of 2 Mn^{2+} and 2 K^+ ions. The addition of two Mn^{2+} ions would be in agreement with previous crystallography and EPR experiments: electron density could be attributed to manganese, and a characteristic Mn^{2+} signal was observed in EPR spectra of Fdc1^{UbiX} [2].

The replete Fdc1^{UbiX} bound to two cofactors species, which is dominant in the native mass spectrum, (Figure 7.1) has a measured mass of 113 583 Da which is 268 Da larger than expected. Importantly, for all species, the predicted mass (with no retention of salt) corresponds to the left hand side of the peak (dashed lines, Figure 7.1), while the apex of the peak corresponds to the adducted forms. These spectra indicate that while the Fdc1 dimer carries no cofactors and retains very little salt, Fdc1^{UbiX} dimers carry either one or two cofactors of 525 Da along with a variety of unresolved adducts. An increase in salt retention occurs with bound cofactors, indicating that the addition of the cofactor favors the retention of counter ions, some are evidenced in the crystal structure and the ability of the holo form to retain salt may be related to its ability to form a stable unit cell for crystallography.
7.3.2 Ion Mobility Mass Spectrometry shows increased population of smaller conformers for Fdc1\textsuperscript{UbiX} cf. Fdc1

**Figure 7.2** Arrival time distributions of the three main charge states of Fdc1 (top) and Fdc1\textsuperscript{UbiX} with one bound cofactor and two bound cofactors. The blue lines are to guide the eye.

\[^{DT}CCS\text{He}\] distributions are shown of the three main charge states (Figure 7.2) of Fdc1 (top) and of the different forms of Fdc1\textsuperscript{UbiX} (middle and bottom). The cross sections of the [M+20H]\textsuperscript{20+} ions range from ~4000 – 6400 Å\textsuperscript{2} for Fdc1, or ~4000-6000 Å\textsuperscript{2} for Fdc1\textsuperscript{UbiX}, and the apex of the distribution is just under 5000 Å\textsuperscript{2} for both forms of the protein; for this charge state there are only small differences in the \[^{DT}CCS\text{He}\] distributions of Fdc1 and different forms of Fdc1\textsuperscript{UbiX}, and these indicate a compaction in the ligand-bound form, the higher \[^{DT}CCS\text{He}\] distributions are not populated. For [M+21H]\textsuperscript{21+} the lower value for the ions is approximately 4200 Å\textsuperscript{2} for all species, however there is a difference between the upper value of Fdc1 (7250 Å\textsuperscript{2}) and Fdc1\textsuperscript{UbiX} (6750 Å\textsuperscript{2}). The apex of the peak is also different; for Fdc1 it is around 5250Å\textsuperscript{2} whereas for Fdc1\textsuperscript{UbiX} it is around 4800 Å\textsuperscript{2}. The differences in \[^{DT}CCS\text{He}\] distribution are most prominent for the [M+22H]\textsuperscript{22+} ion: Fdc1\textsuperscript{UbiX} displays two conformational families centered around 4900 and 5600 Å\textsuperscript{2}, both at roughly equal intensity; for Fdc1, however, the smaller conformational family is greatly reduced in intensity in comparison to the extended state, indicating a destabilisation of the compact conformation in the absence of the cofactor. These IM-MS experiments are in agreement with native MS data presented above, both indicating a more compact structure in the presence of the cofactor.
A key question is why we observe a proportion of Fdc1\textsuperscript{Ubix} carrying less than two cofactors. The possible reasons are that i) this reflects the cofactor occupancy in solution or ii) some cofactor dissociates from the protein during desolvation. The crystal structures solved by Payne and coworkers [2] show 100% cofactor occupancy, but whether this reflects the solution-phase character, or is due to preferential crystallisation of the dimer with two cofactors, remains unknown. Our IM-MS results show that the different forms of Fdc1\textsuperscript{Ubix} have similarly shaped DT\textsubscript{CCS}\textsubscript{He} distributions for the same charge states, regardless of the number of non-covalently bound cofactors. An explanation for this is that some cofactor is lost upon desolvation but the protein retains the conformation adopted when in complex with the cofactor, a templating effect we have reported previously [14] due to the timescale for conformational change in the protein complex being significantly longer than the experimental timescale of the IM-MS process. This is supported by the fact that in our analysis of Fdc1\textsuperscript{Ubix} we are always able to resolve a species at m/z 525 that corresponds exactly to the mass of the cofactor even under the most gentle desolvation conditions. Further evidence that the cofactor is lost during desolvation is found when on modulating the harshness of the ionisation process we can alter the proportion of single-bound and double-bound cofactor (Figure S2, appendix 5). By increasing the voltage applied to the sample cone we cause the protein to experience a larger potential difference during desolvation, which results in harsher ionisation conditions. When the cone is increased to 200V the ratio of the intensities of singly-bound to doubly bound is approximately 0.75, up from approximately 0.66 at a cone voltage of 60V.

This native IM-MS data indicates that the holo form is dominant and conformationally more rigid than the apo form but also that prFMN is somewhat labile and can dissociate from the protein. Under similar nESI-MS conditions the haem co-factor will not dissociate from myoglobin or haemoglobin [15].


7.3.3 HDX-MS shows that Fdc1<sup>UbiX</sup> undergoes a change in conformation around the cofactor binding site along with overall conformational stabilisation in the presence of the cofactor

The difference in deuterium uptake between Fdc1 and Fdc1<sup>UbiX</sup> is reported in Figure 7.3. There is an overall increase in mass per peptide from Fdc1 than from Fdc1<sup>UbiX</sup>, indicating a more dynamic structure for the protein complex in agreement with both MS and IM-MS experiments. In particular, there is a higher uptake for the apo form in the regions containing residues 187-198, 223-234 and 443-450. X-ray crystallography has shown that H191 is involved in the binding of Mn<sup>2+</sup> and Q190 is involved in the binding of the cofactor, which correlates with the HDX data.

The fact that the cofactor is not present and that we have strong evidence from mass spectrometry for no Mn<sup>2+</sup> binding in Fdc1 is a possible explanation for this apparent destabilisation/ increased solvent exposure of the apo protein around residues 187-198. E233 is also involved in Mn<sup>2+</sup> binding which could account for the increased dynamics of residues 223-234. Other residues involved in Mn<sup>2+</sup> binding are N168 and K391, both of which are within regions of increased uptake for Fdc1 compared with Fdc1<sup>UbiX</sup>, albeit at a lower level. E282, E277 and R173 are other residues known to be associated with the bound cofactor and these are also within regions of higher HDX in the absence of the cofactor. Interestingly, there are some areas within Fdc1<sup>UbiX</sup> that have an increase in deuterium uptake in comparison with Fdc1: residues 109-123 and 460-465. The latter areas are in close proximity to regions that have an increased deuterium uptake for Fdc1. This could imply a switch in conformation between the two species; exposure of one area to the solvent is coupled to protection of another area upon removal of the cofactor.
Figure 7.3; Top; the difference in deuterium uptake between Fdc1 and Fdc1\textsuperscript{UbiX}, summed across all time points. The crystal structures have the differences in deuterium uptake coloured according to the key. Secondary structure elements are shown in (a) with the cofactor shown in purple, while (b) and (c) show the surface. (b) and (c) are equivalent, but rotated 180˚. PDB ID 4ZA4. In the sequence, the residues that interact with the cofactor are outlined in purple, while the residues that are involved in Mn\textsuperscript{2+} binding are outlined in orange. Relative uptake plots can be found in figure S5, appendix 5.
Differences in deuterium uptake were visualised by plotting regions of changed mass at the peptide level onto the crystal structure of Fdc1\textsubscript{UbiX}. Differences in uptake were plotted onto the PDB (ID 4ZA4) showing the secondary structure (Figure 7.3a), significant variations in behaviour are found in close proximity to the cofactor binding site. This tends to be at the end of α-helices, in regions that are leading to disordered loops, indicating a change in stability of secondary structure elements upon binding of the cofactor. For example, the predominantly hydrophobic patch over residues 227-238 have an uptake of >5%, where 227-232 form a disordered loop and 233-238 are part of an α-helix that extends to the glycine (242). It is feasible that in the apo form the disordered loop extends at least to the glycine at residue 238 and the helix is correspondingly shorter. The methionine at residue 239 is actually more solvent accessible in the holo-form, suggesting a significant structural rearrangement in this region; residues 240-243 are more solvent exposed in the apo form at longer timescales, suggesting an allosteric rearrangement in this region of the helix. All these changes in uptake around this area from 227-243 suggest a rearrangement/destabilisation of this α-helix in the absence of the cofactor. Another region of destabilisation corresponds to the N-terminal helix of Fdc1\textsubscript{UbiX} which spans residues 5-20. There is an uptake in deuterium of >5% from residues 7-16, suggesting a reduction of secondary structure for this N-terminal region. It is important to consider that while the most appropriate way to display the difference in deuterium uptake is to plot it onto the Fdc\textsubscript{UbiX} structure, this will not correspond exactly to that of apo-Fdc. It is likely that apo-Fdc has reduced overall secondary structure content compared to the holo form, due to the location of the uptake differences at the ends of α-helices suggesting that these are more dynamic and less configured in the absence of prFMN.

Uptake increases at short time points for Fdc1 (red) and Fdc1\textsubscript{UbiX} (yellow) are in close proximity to one another, either in the linear sequence or the position of the areas. Visualisation of the surface of the crystal structure (Figure 7.3 b and c) illustrates differences in uptake both close to, and distant from, the cofactor binding region. This indicates not only a conformational change in the areas that bind the cofactor, but also an allosteric change in conformation in areas distant from the cofactor. This insight to the effect of prFMN binding is possible since HDX-MS is an approach that allows equal visualisation of the apo and holo states of this protein, in contrast to crystallographic data.
7.4 Conclusions and Outlook

Here we have used MS-based techniques to delineate how the structure and dynamics of Fdc1 are affected by the binding of prenylated FMN. Native MS results confirm the stoichiometry of the protein complex; when Fdc1 is expressed alone, that is in the absence of the UbiX gene and hence is in the apo form, it presents mainly as a dimer with a charge state range (Δz) of 5, with some low intensity monomer peaks. When Fdc is co-expressed with UbiX the monomer peaks are eliminated and the Δz is reduced to 4; both of these changes indicate a more stable structure in the presence of the ligand since the dimer interaction is stronger and there is a lower surface area on which the complex can accommodate charge. IM-MS detects subtle changes in the conformations of the differently expressed Fdc1 in the gas phase; there is a preference for the higher charge states to adopt smaller cross sections in the presence of the cofactor. This agrees with assertions made from the MS data that the protein is more compact when the cofactor is bound. With HDX we have i) observed overall conformational tightening in the presence of the cofactor in solution and ii) been able to localise regions of the protein where the alterations in dynamics occur. There are changes in deuterium uptake close to the cofactor binding and Mn\textsuperscript{2+} binding regions suggesting conformational differences in these regions. The protein also appears to undergo allostery changes in the presence of the cofactor, shown by differential deuterium uptake in regions distant from the cofactor binding site. There is very little change in deuterium uptake on the dimer interface between the apo and holo forms of the protein, suggesting that the monomer species observed via MS are not transferred from solution, and likely arise from dissociation during the ionisation process. Previous to the research presented in this paper, it was unknown how the residues moved in response to the cofactor binding. HDX-MS experiments, however, provide information on the change in solvent accessibility of localised regions of the protein.

Overall we have observed differences in the apo and holo forms of the Fdc1 protein; specifically the protein is more dynamic in the apo form and has less rigid tertiary and quaternary structure. This could not be observed via x-ray crystallography since no crystals could be obtained of apo-Fdc1. This lack of crystallisation may be due to the increased flexibility that we have observed in our experiments. Here we have a comparison between apo- and holo-Fdc1 that is equally weighted for both forms of the protein; the data are not preferentially reporting on the cofactor-bound form, as for other methods. We demonstrate the complementarity of MS to crystallographic approaches and in particular highlight its benefits in its use to study conformationally
dynamic protein complexes and conformational dynamics in any protein cofactor complex.

### 7.5 Methods

#### 7.5.1 Expression and purification of Fdc1 and Fdc1\textsuperscript{UbiX}

The \textit{A. niger fdc1} gene was codon optimised and synthesised (Genscript). \textit{A. niger fdc1} was cloned into the \textit{NdeI} and \textit{XhoI} sites of pET30a and \textit{E. coli ubiX} cloned into the \textit{NdeI} and \textit{XhoI} sites of pET21b. \textit{A. niger fdc1} pET30a was transformed into \textit{E. coli BL21(DE3)} with or without \textit{ubiX} pET21b.

Protein was expressed in BL21(DE3) grown at 37 °C/180 rpm in LB broth supplemented with 50 µg/ml kanamycin (single expressed) or with both 50 µg/ml kanamycin and 50 µg/ml ampicillin (co-expressed). At mid-log phase cells were induced with 0.25 mM IPTG and supplemented with 1 mM MnCl\textsubscript{2} grown overnight at 15 °C/180 rpm and then harvested by centrifugation (4 °C, 7000 g for 10 minutes). Cell pellets were resuspended in buffer A (200 mM NaCl, 1 mM MnCl\textsubscript{2}, 50 mM Tris pH 7.5) supplemented with DNase, RNase, lysozyme (Sigma) and Complete EDTA-free protease inhibitor cocktail (Roche). Cells were lysed using a French press at 20,000 psi and the lysate clarified by centrifugation at 125,000 g for 90 minutes. The supernatant was applied to a Ni-NTA agarose column (Qiagen). The column washed with 3 column volumes of buffer A supplemented with 10 mM imidazole and protein eluted in 1 ml fractions with buffer A supplemented with 250 mM imidazole. Samples were subjected to SDS-PAGE analysis and fractions found to contain the purified protein were pooled. Imidazole was removed using a 10-DG desalting column (Bio-Rad) equilibrated 100 mM NaCl, 1 mM MnCl\textsubscript{2}, 25 mM Tris pH 7.5. Protein was aliquoted and flash frozen until required.

UV-Vis absorbance spectra were recorded with a Cary UV–Vis spectrophotometer. The protein concentration \textit{A. niger} Fdc1 was estimated using $\varepsilon_{280} = 68870$ M\textsuperscript{-1} cm\textsuperscript{-1} (calculated from the primary amino acid sequence using the ProtParam program on the ExPASy proteomics server).

#### 7.5.2 Mass Spectrometry experimental conditions

Proteins were buffer exchanged into 100 mM ammonium acetate using a 10-DG desalting column (BioRad). Native mass spectrometry experiments were carried out on a Synapt G2 instrument (Waters, Manchester, UK) with a nanoelectrospray
ionisation (nESI) source. Mass calibration was performed by a separate infusion of NaI cluster ions. Solutions were ionised through a positive potential applied to a platinum wire of thickness 0.125 mm (Goodfellow) inserted into a thin-walled glass capillary (inner diameter 0.9 mm, outer diameter 1.2 mm, World Precision Instruments, Stevenage, UK) that was pulled to a nESI tip in house with a Flaming/Brown micropipette puller (Sutter Instrument Co., Novato, CA). Fdc1 samples (5–10μM) were sprayed from 100 mM ammonium acetate pH, 6.8. Capillary voltage 1.6 kV, sample cone voltage 80-90V, extractor cone voltage 3V, backing pressure 5 mbar, source temperature 50˚C, Trap gas flow 0.4-5 ml/min. Data were processed using Masslynx V4.1 software and Originlab 9.0.

7.5.3 Ion Mobility-Mass Spectrometry

Ion mobility mass spectrometry is a gas phase electrophoretic technique which is frequently coupled to mass spectrometry. A typical ion mobility experiment involves ions being pulsed into a drift tube, across which is applied a weak electric field which draws the ions through the cell. The drift tube is filled with an inert gas, in this case helium, which collides with the ions, hindering their passage through the drift tube. Larger ions will experience more frequent collisions with the buffer gas and will hence be slowed down to a greater extent. By measuring the arrival time of the ion and the charge present upon it via subsequent MS, the rotationally averaged collision cross section can be directly calculated.

The mobility of an ion ($K_0$) is determined as the ratio of the drift velocity ($v_d$) and applied electric field ($E$). It is then possible to determine the $^{DT}CCS_{He}$ on the basis of Equation 7.1:

$$
K_0 = \frac{3ze}{16N} \left(\frac{2\pi}{\mu k_B T}\right)^{1/2} \frac{1}{\bar{n}}
$$

Equation 7.1

where $z$ is the ion charge state, $e$ is the elementary charge, $N$ is the gas number density, $\mu$ is the reduced mass of the ion-neutral pair, $k_B$ is the Boltzman constant, $T$ is the gas temperature and $K_0$ is the reduced mobility (the measured mobility $K$ standardised for pressure and temperature to 273.15 K and 760 Torr).

IM-MS experiments were carried out on a Waters Q-ToF I instrument that was modified in house to include a 5.1cm drift tube which has been described elsewhere [10]. Ions were produced by nESI with a capillary voltage of 1.6kV and a source temperature of 80˚C. Tips were prepared as above. The temperature and pressure of helium in the drift cell were approximately 30˚C and 4 Torr respectively. Measurements were made at 6 different drift
voltages from 60V to 20V. The precise pressure and temperature was recorded for every drift voltage and used in the calculations of $\Delta T_C S_{He}$ values. Each experiment was performed in triplicate. Data was analysed using MassLynx v4.1 software (Waters, Manchester, UK), Origin v8.5 (OriginLab Corporation, USA) and Microsoft Excel. Ion arrival time distributions were recorded by synchronisation of the release of ions into the drift cell with mass spectral acquisition.

The $\Delta T_C S_{He}$ distribution plots are derived from raw arrival time data using Equation 7.2 below [16].

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

Equation 7.2

Where $m$ and $m_b$ are the masses of the ion and buffer gas, respectively; $z$ is the ion charge state; $e$ is the elementary charge; $K_B$ is the Boltzmann constant; $T$ is the gas temperature; $\rho$ is the buffer gas density; $L$ is the drift tube length; $V$ is the voltage across the drift tube; and $t_d$ is the drift time.

The raw arrival time output ($t_a$) includes time the ions spend outside of the drift cell but within the mass spectrometer, known as the dead time ($t_0$). The value for $t_0$ is calculated by taking an average value of the intercept from a linear plot of average arrival time versus pressure/temperature and was subtracted from the arrival time to calculate drift time ($t_d$):

$$t_0 = t_a - t_0$$

Equation 7.3

### 7.5.4 HDX-MS experimental conditions;

Fdc1 and Fdc1UbiX solutions were prepared at 20 µM. HDX labelling and quenching procedures were automatically performed using the CTC PAL sample manager (LEAP Technologies, Carrboro, NC, USA). The samples were diluted 20-fold with 10 mM phosphate in 99.99% deuterium oxide, pH 6.6 (pD 7.0), and incubated for 0, 0.25, 5, 60 and 240 minutes at 20 °C. Labelled samples were then quenched with an equal volume of pre-chilled 100 mM phosphate pH 2.5. All labelling time-points were analysed in triplicate. 50 µL of sample was injected on a nanoACQUITY UPLC™ system with HDX technology (Waters). Online pepsin digestion was performed for 1 min at 20°C on a Waters Enzymate™ immobilised BEH pepsin column (2.1x30 mm). The peptides were separated on a UPLC BEH C18 column (Waters) at 0 °C. Peptides were separated with a 7 min linear acetonitrile gradient (8-35%) containing 0.1% formic acid at 40 μL/min. Mass spectra were acquired on a SYNAPT G2-Si HDMS in MS$^E$ mode over the m/z range of 50-2000. Non deuterated
peptides were identified using ProteinLynx Global Server software 3.1 (Waters). DynamX 2.0 software (Waters) was used to filter the peptides and to generate deuterium uptake data. Originlab 9.0 was used to create the plots.

7.6 References

8
Conclusions
8. Conclusions

The importance of protein dynamics has always been recognised, and this has increased recently, with significant growth in the characterisation of IDPs. Mass spectrometry coupled with ion mobility has been positioned as an attractive technique with which to study protein dynamics; it provides information on the range of conformations in which a protein exists, rather than average data on a dynamic ensemble of interconverting structures. From the charge state distribution presented by a protein following nano-electrospray ionisation, it is possible to infer the range of conformational ensembles present in solution and hence the extent of disorder. From a combination of ion mobility and mass spectrometry, the size of these conformations can be measured in terms of their rotationally averaged collision cross section. Although the information provided is low resolution in comparison to NMR and X-ray crystallography, CCSs can be compared to models determined by computational techniques, which gives insight into the conformations to which they correspond.

The objective of the work presented in this thesis was to use ion mobility-mass spectrometry to probe protein conformations and their dynamic propensity in the gas phase, and relate this to their solution-phase behaviour. I began by building a framework of mass-spectrometry based experiments to identify both structure and disorder in proteins. I tested this framework on a variety of proteins with different structural characteristics including proteins of different extents of disorder, protein complexes with dynamic entities and a system that undergoes structural rearrangement upon ligand binding.

In Chapter 3 I developed a framework of mass spectrometry-based experiments to provide information on the dynamic properties of a protein for which no structural information is known. I sought to answer the following question: “If the only technique available was ion mobility-mass spectrometry, what information would it provide on the structural preferences of an unknown protein?” MS and IM-MS results were compared for a set of 20 proteins, spanning a molecular mass range of over 147 kDa with varying extents of secondary, tertiary and quaternary structure. Structured proteins present with $z \leq 7$ states or less and have a CCS range of $<750 \text{ Å}^2$, disordered proteins will present with $z \geq 7$ and have a CCS range of $>750 \text{ Å}^2$, and proteins with regions of both structure and disorder, or a tendency to disorder, will have a wide CSD with most of the intensity in the lower charge states. Comparisons were then made between measured CCSs and theoretically
calculated CCSs; structured proteins display a much narrower CCS range than is theoretically available and these are at the smaller end of the allowed CCS range. IDPs have a wide CCS range under native-like conditions because they are free to access most of their conformations. From this information we have built a framework of exemplar experiments. This manuscript describes a workflow for the use of MS and IM-MS as initial characterisation techniques to obtain information about the extent of structure or disorder in unknown proteins.

Chapter 4 contains the analysis of the IDPs α-Syn and ApoC-II by mass spectrometry methods described in Chapter 3. The proteins were first analysed via HDX-MS which confirmed a high degree of disorder in solution, in agreement with previous studies. The MS and IM-MS results are starkly different. Whilst α-Syn displays the expected wide CSD and CCS range of an IDP, ApoC-II presents a narrow CSD and CCS range, irrelevant of the solution conditions from which it is sprayed. Comparison with the theoretical model described in chapter three that predicts the CCSs of the most extended and most compact conformations of a protein shows that α-Syn explores most of the conformational space that is available to it, whereas ApoC-II is only present in a small amount of available conformational space. We rationalise the range of conformations displayed in the gas phase by each protein by the ESI mechanisms to which they both adhere.

Chapter 5 presents IM-MS data on C-terminal p27 (p27-C) when it is ionised from solutions of altered ionic strength. The conformations observed in the gas phase are strongly dependent upon the solution conditions from which they are sprayed. Two permutants are also investigated; one has a lower κ-value than the WT and is hence expected to be more conformationally flexible, and the other has a higher κ-value than the WT and is hence expected to be less conformationally flexible. Our IM-MS experiments are able to differentiate effectively the permutants, and our results follow the trend that has been observed in solution-phase experiments, with much more differentiation. CCSs were calculated from structures obtained during Monte Carlo simulations; here the trend for the permutants also applies. A comparison of experimental and calculated CCS values provides evidence for a collapse of the gas phase molecules; we observe consistently smaller conformations in vacuo than are seen in solution or calculated by MC simulations in implicit solvent. This research has allowed us to assess the extent that the solution-phase conformations are transferred into the gas phase by comparison with in vitro and in silico approaches.
The analysis of the Cdk2/ Cyclin A/ p27 complex is presented in Chapter 6, along with data on the complex containing the different C-terminal domains. We show a vast reduction in the dynamic propensity of p27 as it binds to the binary complex, reflecting the ‘folding upon binding’ behaviour. We also investigate how the charged residue patterns of the 11.2 kDa p27-C affects the conformational behaviour of the 86.8 kDa Cdk2/ Cyclin A/ p27 complex. The incorporation of the more extended C-terminal domain results in a complex with higher CCS values for each charge state, whilst incorporation of the less dynamic p27-C results in conformations similar to the wild-type.

In Chapter 7 the conformational and dynamic properties of Fdc1 and Fdc1\textsuperscript{UbiX} are compared. Fdc1\textsuperscript{UbiX} carries non-covalently bound prFMN cofactors due to co-expression with the \textit{UbiX} gene, whilst Fdc1 is the apo form. MS reveals the binding stoichiometry of prFMN to Fdc1; most of the Fdc1\textsuperscript{UbiX} ions, approximately two thirds, carry two non-covalently bound cofactors. Approximately one third carry one non-covalently bound cofactor and just a small amount of protein is present in the apo form. There is a reduction in populated charge states upon cofactor binding, and IM-MS data suggests that the cofactor has a stabilising effect on the protein conformation. HDX of both forms of the protein revealed an overall increase in solvent accessibility of Fdc1 c.f. Fdc1\textsuperscript{UbiX}, indicating a more dynamic structure in agreement with both MS and IM-MS experiments. In particular, there is a higher increase in deuterium uptake by Fdc1 in the cofactor-binding regions, and in places where α-helices lead into disordered loops indicating a change in stability of secondary structure elements upon binding of the cofactor. This research exemplifies how experiments and analyses from the MS framework can be used to detect subtle changes in dynamics within structured proteins, and has uses beyond the identification of complete unfolding of a protein.

I envisage that native IM-MS could become a widely used initial characterisation technique for proteins and protein complexes; in a single experiment one can obtain information on the structural characteristics of a protein which will indicate how to further investigate the behaviour of the protein. If the IM-MS results show the protein is folded with minimum dynamics then crystallisation may be possible. If the protein is very dynamic then NMR may be a good subsequent method for the investigation of the protein. A way to facilitate the use of IM-MS for the wider community would be to develop the technique onto more accessible, cost effective systems. Currently experiments are carried out on instruments that cost in excess of £500 000, or on MS systems that have been modified in house to be able to perform IM-MS.
analysis. This could involve coupling the Owlstone FAIMS chip to a low cost commercially available mass spectrometer such as the Waters QDA.

It is also feasible that IM-MS could be used for high-throughput, low resolution investigations into the effect of drug binding to the target protein. Commercially available IM-MS instruments such as the Synapt instruments (Waters, UK) along with automated electrospray devices such as the Nanomate (Advion, NY, USA) allow experiments on a relatively high-throughput timescale. An advantage of IM-MS over other techniques is that it allows observation of low-lying protein conformations which may have implications in diseases and are therefore important to monitor.

The combination of MS and IM-MS has proven to be a powerful approach for investigating protein structure and dynamics in a solvent-free environment. Information can be gained on the conformational diversity of individual proteins and multi-protein complexes. This technique allows the analysis of coexisting conformers of a single protein or complex, and also allows different species to be simultaneously studied, providing insight into important biological systems.
Appendices
Appendix 1

Supplementary information for Chapter 3

Table S1  Sequences of proteins used in the study. ............................................................... 202
Table S2  Collision cross sections of cytochrome c ............................................................. 203
Table S3  Collision cross sections of β-casein ................................................................. 204
Figure S1  Mass spectrum of N-terminal p53 ................................................................. 205
Table S4  Collision cross sections of N-terminal p53 ...................................................... 205
Table S5  Collision cross sections of α-synuclein ............................................................. 206
Table S6  Collision cross sections of N-terminal MDM2 .................................................. 207
Table S7  Collision cross sections of ovalbumin ................................................................. 208
Table S8  Collision cross sections of ovalbumin (intact) .................................................... 208
Table S9  Collision cross sections of ovalbumin (reduced) ................................................ 208
Figure S5  Mass spectrum of human TTR tetramer ........................................................... 209
Table S10 Collision cross sections of Avidin tetramer ........................................................ 210
Figure S7  Mass spectrum of bovine serum albumin ......................................................... 211
Table S12 Collision cross sections of bovine serum albumin ........................................... 211
Figure S8  Mass spectrum of Concanavalin A tetramer ..................................................... 212
Table S13 Collision cross sections of Concanavalin A tetramer ....................................... 212
Figure S9  Mass spectrum of Serum Amyloid P component pentamer ............................ 213
Table S14 Collision cross sections of Serum Amyloid P component pentamer ............... 213
Figure S10 Charge state range vs. mass for proteins over 150kDa. .................................. 214
Figure S11 Arrival time distributions for the most intense charge states of lysozyme, cytochrome c, β-casein and α-Synuclein ................................................................. 214
Table S15 Theoretical and experimental maximum and minimum collision cross sections. .................................................................................................................................................. 215
Figure S12 Mass spectrum of lysozyme ........................................................................... 215
Table S16 Collision cross sections of lysozyme in 50mM ammonium acetate ................... 215
Figure S13 Mass spectrum of 70µM lysozyme in a 49.5:49.5 ratio of methanol to water, with 1% formic acid. ................................................................. 216
Table S17 Average collision cross sections and relative intensities of α-synuclein .......... 217
Table S18 Average collision cross sections and relative intensities of β-casein ............... 218
Table S19 Average collision cross sections and relative intensities of cytochrome c .......... 218
## Appendix 1

<table>
<thead>
<tr>
<th>Protein</th>
<th>Sequence</th>
</tr>
</thead>
</table>
| **Human**
  cytochrome c               | GDVEKGKKIF IMKCSQCHTV EGGKHKTGTP NLHLFGIRKT RADIAYLKK AINE               |
|                             | RELEEENVPG EIVESLSSSE ESITRINKKI EKFQSEEQQQ                             |
|                             | TEDELQDKIH PFAPTSQSLV EFPGIPINSL PQNIPPIQQT                             |
|                             | PVVVPFPQLP EMVGVSKVE AMAPKHKEMP FPKYPVEFFP                              |
|                             | ESQSLTLDV ENLHPLPLL QSWMHPHQHP LPPTVMFPPQ                                |
|                             | SVLSSLQKV LPVPFQKAVY PQRDMPQIAF LLYQEPVLGP                               |
| **Bovine**
  β-casein                  | DTHKSEIAHR FKDLEGEEHKF GLVLLAFSQQY LQQCPFDEHV                              |
|                             | KLVMELTEFA KTCVADESHA GCEKSLHTLF GDELCKVSL                                |
|                             | RETYGDMAOC CEKQEPERNNE CFLSHKDDSP DLPKLKDPFN                               |
|                             | TLCDIFGREL KAFWKGKLYE IARKHYYFA PELLYYANKY                                |
|                             | NGVFQECCQA EDFGACLPLL IEMRNEKVL SAQRQLRCA                                 |
|                             | SIQKFEGERAL KAWSVARLSQ KFFKAEFVEV TKLVTDLTKQ                               |
|                             | HKECCCGRDL EACDADDRALA KYICDQNQDTI SSKLKECCDK                              |
|                             | PLEKSHCIA VEKDAEFEN LFP disturbances KDKVDVCKYQ                           |
|                             | AKDAFLGSFP YEYDSRHPFY AVSVLLRLAK YEATLLECC                                |
|                             | AKDPPHDCAF TVFDDLHLV DEPFNQLKQN CDQFEKLEYQ                                |
|                             | GFQNALVRY TRKVQVSTP TLVEVSRLG KVTRCCTKP                                  |
|                             | SERMECTPD YLSLILNRCL VHEKTVPLVE KVTKCCTLESL                               |
|                             | VNRPPFCSAL TPDETYVPKA FDEKLFTHFA DICLTPFDEK                               |
|                             | QINKQTALVE LIIKHFPKATE EQLKTVMENF VAFVDDCA                                 |
|                             | DDKEACFAVE GPKLVSTQAL ALA                                                 |
| **Bovine**
  serum albumin             | KVFGRCELAA AMKRHLGDN RYGSLNWMVC AAKFESNFNT                                |
|                             | QATNRRNTDG TTDYLQINS RWNCDGTRP GSNRCLNIPC                                 |
|                             | SALLSSDTIA SVNACKIVS DGNMNAWV WNRNCRTVD                                  |
| **Hen**
  egg white Lysozyme       | KVFGRCELAA AMKRHLGDN RYGSLNWMVC AAKFESNFNT                                |
|                             | QATNRRNTDG TTDYLQINS RWNCDGTRP GSNRCLNIPC                                 |
|                             | SALLSSDTIA SVNACKIVS DGNMNAWV WNRNCRTVD                                  |
| **Hen**
  egg white ovalbumin      | MGSIGAASME FCFDFVKELK VHNNENIFY CPIAIMSALA                                 |
|                             | MYLGLAKDS RTQINKVVRF DKLPGFDSI EAQCGTSNV                                 |
|                             | HSLRLDDLQ ITKPDVNYSF SLASRLYAEE RYPILPEYQ                                 |
|                             | CVKELYRGDL EPINFQTAAD QARELINNSV ESQINGIIIN                               |
|                             | VLPSSVDSQ TAMVLNVAIF FKGLWEKAF DEDTQAMPF                                  |
|                             | VTEQEPKQFP MYQIGILRFV ASMAQEMKI LELFFASGTM                                |
|                             | SMLVLPLDEV SGLEQLESII NEFLEKETWNS SNVEERKIK                               |
|                             | VYLPFRMKME KYNITLVSMA MGITDVFFSS ANLGLSSIA                                 |
|                             | SLKISQAVHA AHAINEAERG EVVGSAGAVG DAAVSEIIFR                               |
|                             | ADHFLFECQF HIATNAVLFLF GRCVSP                                             |
| **Egg white avidin**       | ARKCSLTGKW TNDLSNNTM GAVNRSCEF T GYTIVATAT                                |
|                             | SNEIKESPLH QTONTINKRT QPTFQFTVNW KFSESTTVFT                               |
|                             | QGCFIDRNGK EVLKTWMLR SSVNDIGDDW KATRVCINIF                                 |
| **C. ensiformis**
  concanavalin A            | ADTIVAVELD TYPNTDIGDP SYPHIGIDIK SVRSKKTAKW                               |
|                             | NMQNGKVGTA HIHYNSVGRK LSAVSYFPNG DASATVSYDVD                               |
|                             | LDVNLPEWVR VGLSASTGLY KETNTLSWS FTSDKLNASN                                 |
|                             | HETNLALHF MQFSKDQKDL ILQDADTTG DNLEGRTLSR                                 |
|                             | SNGSPQGNSV GRALFYAPVH INESSAVAVS FDATFFLKL                                |
|                             | SFSDPHADGI AFFISNIDSS IPSGSTGRLL GLFPPADN                                 |
| **Equine**
  heart myoglobin          | GLSDGQEQQV LNVWGGKEAD IAGHQLVEQL RLFGTHFETL                                |
|                             | EKFDKFKHL TEAEKMASED LKHKTVVHT ALGGLLKKK                                  |
|                             | HHEALKEPLA QSHATKHIP IKYLEFISDA IIVHLHSKHP                                 |
|                             | GDFGADAQQA MTKALELFRN DIAARKEKLG FQG                                     |
| **Human**
  transthyretin             | GPTGTFEGSKC PLMVKVLDAV RGSPIAVN HVRKAAADDT                                |
|                             | WEFFASCTKS EGSELHLGTT EEEFVEGYK VEIDTKSYNK                                |
|                             | ALGISPFHEH AEUVFTANDS GPRRTYIAL LSPYSSSTA                                 |
### Table S1 Sequences of proteins used in the study.

<table>
<thead>
<tr>
<th>Species</th>
<th>m/z</th>
<th>Average cross section from 3 repeats/ Å²</th>
<th>Standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>[M+5H]⁺</td>
<td>2446.81</td>
<td>952</td>
<td>1.5</td>
</tr>
<tr>
<td>[M+6H]⁺</td>
<td>2039.18</td>
<td>1128</td>
<td>7.1</td>
</tr>
<tr>
<td>[M+7H]⁺</td>
<td>1748.01</td>
<td>1156</td>
<td>5</td>
</tr>
<tr>
<td>[M+8H]⁺</td>
<td>1529.64</td>
<td>1326</td>
<td>65.0</td>
</tr>
<tr>
<td>[M+8H]⁺</td>
<td>1529.64</td>
<td>1254</td>
<td>58.5</td>
</tr>
</tbody>
</table>

### Table S2 Collision cross sections of 20 µM cytochrome c (12 229 Da) in 50 mM ammonium acetate, pH 6.8
<table>
<thead>
<tr>
<th>Species</th>
<th>m/z</th>
<th>Average Collision Cross Section from 2 repeats/ Å²</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>[M+7H]⁺</td>
<td>3426.75</td>
<td>1450</td>
<td>75</td>
</tr>
<tr>
<td>[M+8H]²⁺</td>
<td>2998.54</td>
<td>1658.5</td>
<td>11</td>
</tr>
<tr>
<td>[M+9H]³⁺</td>
<td>2665.48</td>
<td>1862</td>
<td>13</td>
</tr>
<tr>
<td>[M+10H]⁴⁺</td>
<td>2399.03</td>
<td>2087.5</td>
<td>153</td>
</tr>
<tr>
<td>[M+10H]⁴⁺</td>
<td>2399.03</td>
<td>1800.5</td>
<td>37</td>
</tr>
<tr>
<td>[M+11H]⁵⁺</td>
<td>2181.03</td>
<td>1924.5</td>
<td>21</td>
</tr>
<tr>
<td>[M+11H]⁵⁺</td>
<td>2181.03</td>
<td>2390</td>
<td>20</td>
</tr>
<tr>
<td>[M+12H]⁶⁺</td>
<td>1999.36</td>
<td>2154</td>
<td>17</td>
</tr>
<tr>
<td>[M+12H]⁶⁺</td>
<td>1999.36</td>
<td>2596.5</td>
<td>156</td>
</tr>
<tr>
<td>[M+13H]⁷⁺</td>
<td>1845.64</td>
<td>2317.5</td>
<td>39</td>
</tr>
<tr>
<td>[M+13H]⁷⁺</td>
<td>1845.64</td>
<td>2975</td>
<td>21</td>
</tr>
<tr>
<td>[M+14H]⁸⁺</td>
<td>1713.88</td>
<td>2556.5</td>
<td>25</td>
</tr>
<tr>
<td>[M+14H]⁸⁺</td>
<td>1713.88</td>
<td>3197.5</td>
<td>151</td>
</tr>
<tr>
<td>[M+15H]⁹⁺</td>
<td>1599.69</td>
<td>3481</td>
<td>163</td>
</tr>
<tr>
<td>[M+15H]⁹⁺</td>
<td>1599.69</td>
<td>2696</td>
<td>3</td>
</tr>
<tr>
<td>[M+16H]¹⁰⁺</td>
<td>1499.77</td>
<td>3248.5</td>
<td>86</td>
</tr>
<tr>
<td>[M+16H]¹⁰⁺</td>
<td>1499.77</td>
<td>3637.5</td>
<td>76</td>
</tr>
<tr>
<td>[M+17H]¹¹⁺</td>
<td>1411.61</td>
<td>3419.5</td>
<td>45</td>
</tr>
<tr>
<td>[M+17H]¹¹⁺</td>
<td>1411.61</td>
<td>3793</td>
<td>47</td>
</tr>
<tr>
<td>[M+18H]¹²⁺</td>
<td>1333.24</td>
<td>3588.5</td>
<td>36</td>
</tr>
<tr>
<td>[M+18H]¹²⁺</td>
<td>1333.24</td>
<td>4024</td>
<td>16</td>
</tr>
<tr>
<td>[M+19H]¹³⁺</td>
<td>1263.13</td>
<td>3882.5</td>
<td>95</td>
</tr>
<tr>
<td>[M+19H]¹³⁺</td>
<td>1263.13</td>
<td>4324.5</td>
<td>87</td>
</tr>
<tr>
<td>[M+20H]¹⁴⁺</td>
<td>1200.02</td>
<td>4040.5</td>
<td>155</td>
</tr>
<tr>
<td>[M+20H]¹⁴⁺</td>
<td>1200.02</td>
<td>4559.5</td>
<td>13</td>
</tr>
<tr>
<td>[M+21H]¹⁵⁺</td>
<td>1142.92</td>
<td>4360.5</td>
<td>129</td>
</tr>
<tr>
<td>[M+21H]¹⁵⁺</td>
<td>1142.92</td>
<td>4737</td>
<td>34</td>
</tr>
<tr>
<td>[M+22H]¹⁶⁺</td>
<td>1091.02</td>
<td>4611.5</td>
<td>107</td>
</tr>
<tr>
<td>[M+22H]¹⁶⁺</td>
<td>1091.02</td>
<td>4939.5</td>
<td>37</td>
</tr>
<tr>
<td>[M+23H]¹⁷⁺</td>
<td>1043.63</td>
<td>4643.5</td>
<td>23</td>
</tr>
<tr>
<td>[M+23H]¹⁷⁺</td>
<td>1043.63</td>
<td>5112</td>
<td>119</td>
</tr>
<tr>
<td>[M+24H]¹⁸⁺</td>
<td>1000.19</td>
<td>4682.5</td>
<td>179</td>
</tr>
<tr>
<td>[M+24H]¹⁸⁺</td>
<td>1000.19</td>
<td>5109</td>
<td>51</td>
</tr>
<tr>
<td>[M+25H]¹⁹⁺</td>
<td>960.22</td>
<td>4802</td>
<td>31</td>
</tr>
<tr>
<td>[M+25H]¹⁹⁺</td>
<td>960.22</td>
<td>5392.5</td>
<td>77</td>
</tr>
<tr>
<td>[M+26H]²⁰⁺</td>
<td>923.33</td>
<td>4712.5</td>
<td>47</td>
</tr>
<tr>
<td>[M+26H]²⁰⁺</td>
<td>923.33</td>
<td>5519</td>
<td>72</td>
</tr>
<tr>
<td>[M+27H]²¹⁺</td>
<td>889.17</td>
<td>4663.5</td>
<td>42</td>
</tr>
<tr>
<td>[M+27H]²¹⁺</td>
<td>889.17</td>
<td>5311</td>
<td>249</td>
</tr>
<tr>
<td>[M+28H]²²⁺</td>
<td>857.45</td>
<td>4655.5</td>
<td>91</td>
</tr>
</tbody>
</table>

Table S3  Collision cross sections of 40 µM β-casein (23 980 Da) in 50 mM ammonium acetate, pH 6.8.
Figure S1. Mass spectrum of 200 µM N-terminal p53 (11162 Da) in 50 mM ammonium acetate, pH 6.8

<table>
<thead>
<tr>
<th>Species</th>
<th>m/z</th>
<th>Average collision cross section (no. of repeats)/Å²</th>
<th>Standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>[M+5H]⁵⁺</td>
<td>2233</td>
<td>1091 (3)</td>
<td>50</td>
</tr>
<tr>
<td>[M+6H]⁶⁺</td>
<td>1861</td>
<td>1127 (3)</td>
<td>47</td>
</tr>
<tr>
<td>[M+6H]⁶⁺</td>
<td>1861</td>
<td>1639 (2)</td>
<td>24</td>
</tr>
<tr>
<td>[M+7H]⁷⁺</td>
<td>1595</td>
<td>1693 (3)</td>
<td>16</td>
</tr>
<tr>
<td>[M+8H]⁸⁺</td>
<td>1396</td>
<td>1824 (3)</td>
<td>13</td>
</tr>
<tr>
<td>[M+9H]⁹⁺</td>
<td>1241</td>
<td>1879 (3)</td>
<td>29</td>
</tr>
<tr>
<td>[M+9H]⁹⁺</td>
<td>1241</td>
<td>2166 (1)</td>
<td>-</td>
</tr>
<tr>
<td>[M+10H]¹⁰⁺</td>
<td>1117</td>
<td>1646 (3)</td>
<td>32</td>
</tr>
<tr>
<td>[M+10H]¹⁰⁺</td>
<td>1117</td>
<td>2249 (3)</td>
<td>19</td>
</tr>
<tr>
<td>[M+11H]¹¹⁺</td>
<td>1016</td>
<td>1722 (3)</td>
<td>84</td>
</tr>
<tr>
<td>[M+11H]¹¹⁺</td>
<td>1016</td>
<td>2318 (3)</td>
<td>16</td>
</tr>
<tr>
<td>[M+12H]¹²⁺</td>
<td>931</td>
<td>1707 (3)</td>
<td>41</td>
</tr>
<tr>
<td>[M+12H]¹²⁺</td>
<td>931</td>
<td>2423 (3)</td>
<td>68</td>
</tr>
</tbody>
</table>

Table S4 Collision cross sections of 200 µM N-terminal p53 in 50 mM ammonium acetate
Figure S2  Mass spectrum of 20 µM α-synuclein (14 460 Da) in 50 mM ammonium acetate

Table S5  Collision cross sections of 50 µM α-synuclein in 50 mM ammonium acetate.
### Figure S3
Mass spectrum of 50 µM N-terminal MDM2 (14 790 Da) in 50 mM ammonium acetate

![Mass spectrum](image)

<table>
<thead>
<tr>
<th>Species</th>
<th>m/z</th>
<th>Collision cross section / Å²</th>
</tr>
</thead>
<tbody>
<tr>
<td>[M+5H]⁺</td>
<td>2959</td>
<td>1047</td>
</tr>
<tr>
<td>[M+5H]⁶⁺</td>
<td>2959</td>
<td>1493</td>
</tr>
<tr>
<td>[M+6H]⁺</td>
<td>2466</td>
<td>1103</td>
</tr>
<tr>
<td>[M+6H]⁶⁺</td>
<td>2466</td>
<td>1642</td>
</tr>
<tr>
<td>[M+7H]⁷⁺</td>
<td>2114</td>
<td>1335</td>
</tr>
<tr>
<td>[M+7H]⁷⁺</td>
<td>2114</td>
<td>1839</td>
</tr>
<tr>
<td>[M+8H]⁺</td>
<td>1850</td>
<td>1405</td>
</tr>
<tr>
<td>[M+8H]⁶⁺</td>
<td>1850</td>
<td>2014</td>
</tr>
<tr>
<td>[M+9H]⁸⁺</td>
<td>1644</td>
<td>1575</td>
</tr>
<tr>
<td>[M+10H]¹⁰⁺</td>
<td>1480</td>
<td>1587</td>
</tr>
<tr>
<td>[M+11H]¹¹⁺</td>
<td>1346</td>
<td>1642</td>
</tr>
<tr>
<td>[M+11H]¹¹⁺</td>
<td>1346</td>
<td>2107</td>
</tr>
<tr>
<td>[M+12H]¹²⁺</td>
<td>1234</td>
<td>1723</td>
</tr>
<tr>
<td>[M+12H]¹²⁺</td>
<td>1234</td>
<td>2042</td>
</tr>
<tr>
<td>[M+13H]¹³⁺</td>
<td>1139</td>
<td>1715</td>
</tr>
<tr>
<td>[M+13H]¹³⁺</td>
<td>1139</td>
<td>2245</td>
</tr>
<tr>
<td>[M+14H]¹⁴⁺</td>
<td>1057</td>
<td>1958</td>
</tr>
<tr>
<td>[M+14H]¹⁴⁺</td>
<td>1057</td>
<td>2472</td>
</tr>
<tr>
<td>[M+15H]¹⁵⁺</td>
<td>987</td>
<td>2003</td>
</tr>
<tr>
<td>[M+15H]¹⁵⁺</td>
<td>987</td>
<td>2436</td>
</tr>
</tbody>
</table>

Table S6 Collision cross sections of 50 µM N-terminal MDM2 in 50 mM ammonium acetate
Figure S4  Mass spectrum of 50 μM ovalbumin (44 200 Da) in 10 mM ammonium acetate.

<table>
<thead>
<tr>
<th>Species</th>
<th>m/z</th>
<th>Collision Cross Section/ Å²</th>
</tr>
</thead>
<tbody>
<tr>
<td>[M+10H]^{10+}</td>
<td>4421</td>
<td>2246*</td>
</tr>
<tr>
<td>[M+11H]^{11+}</td>
<td>4019</td>
<td>2227*</td>
</tr>
<tr>
<td>[M+11H]^{11+}</td>
<td>4019</td>
<td>2703Δ</td>
</tr>
<tr>
<td>[M+12H]^{12+}</td>
<td>3684</td>
<td>2775Δ</td>
</tr>
<tr>
<td>[M+13H]^{13+}</td>
<td>3401</td>
<td>2801Δ</td>
</tr>
</tbody>
</table>

Table S7  Collision cross sections of ovalbumin (intact). Conformation 1 cross sections are denoted as * while conformation 2 cross sections are denoted as Δ.

<table>
<thead>
<tr>
<th>Species</th>
<th>m/z</th>
<th>Collision Cross Section/ Å²</th>
</tr>
</thead>
<tbody>
<tr>
<td>[M+10H]^{10+}</td>
<td>4421</td>
<td>2227*</td>
</tr>
<tr>
<td>[M+11H]^{11+}</td>
<td>4019</td>
<td>2358*</td>
</tr>
<tr>
<td>[M+11H]^{11+}</td>
<td>4019</td>
<td>2773Δ</td>
</tr>
<tr>
<td>[M+12H]^{12+}</td>
<td>3684</td>
<td>2840 Δ</td>
</tr>
<tr>
<td>[M+12H]^{12+}</td>
<td>3684</td>
<td>2558Δ</td>
</tr>
<tr>
<td>[M+13H]^{13+}</td>
<td>3401</td>
<td>2753*</td>
</tr>
<tr>
<td>[M+13H]^{13+}</td>
<td>3401</td>
<td>2905Δ</td>
</tr>
<tr>
<td>[M+13H]^{13+}</td>
<td>3401</td>
<td>3115Δ</td>
</tr>
</tbody>
</table>

Table S8  Collision cross sections of ovalbumin (reduced). Conformation 1 cross sections are denoted as * while conformation 2 cross sections are denoted as Δ.
Appendix 1

Figure S5  Mass spectrum of 30 µM human TTR tetramer (64 kDa) in 200 mM ammonium acetate

<table>
<thead>
<tr>
<th>Species</th>
<th>m/z</th>
<th>Average CCS from 2 repeats/ Å²</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>[M+13H]^{13+}</td>
<td>4232</td>
<td>3128</td>
<td>15</td>
</tr>
<tr>
<td>[M+14H]^{14+}</td>
<td>3930</td>
<td>3202</td>
<td>52</td>
</tr>
<tr>
<td>[M+15H]^{15+}</td>
<td>3668</td>
<td>3350</td>
<td>130</td>
</tr>
</tbody>
</table>

Table S9  Collision cross sections of 30 µM human TTR in 200mM ammonium acetate

209
Appendix 1

Figure S6  Mass spectrum of 30 µM Avidin tetramer (55 kDa) in 200 mM ammonium acetate

<table>
<thead>
<tr>
<th>Species</th>
<th>m/z</th>
<th>Average CCS (no. of repeats) / Å²</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>[M+15H]^{15+}</td>
<td>4232</td>
<td>3347 (2)</td>
<td>98</td>
</tr>
<tr>
<td>[M+16H]^{16+}</td>
<td>3930</td>
<td>3434 (2)</td>
<td>60</td>
</tr>
<tr>
<td>[M+17H]^{17+}</td>
<td>3236</td>
<td>3446 (1)</td>
<td>-</td>
</tr>
</tbody>
</table>

Table S10  Collision cross sections of 30 µM Avidin tetramer in 200 mM ammonium acetate
Figure S7  Mass spectrum of 60 µM bovine serum albumin (69 kDa) in 100 mM ammonium acetate

<table>
<thead>
<tr>
<th>Species</th>
<th>m/z</th>
<th>Average collision cross section (no. of repeats)/ Å²</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>[M+14H]^{14+}</td>
<td>4746</td>
<td>4167 (1)</td>
<td>-</td>
</tr>
<tr>
<td>[M+14H]^{14+}</td>
<td>4746</td>
<td>3433 (2)</td>
<td>66</td>
</tr>
<tr>
<td>[M+15H]^{15+}</td>
<td>4430</td>
<td>3653 (3)</td>
<td>56</td>
</tr>
<tr>
<td>[M+15H]^{15+}</td>
<td>4430</td>
<td>4419 (2)</td>
<td>457</td>
</tr>
<tr>
<td>[M+16H]^{16+}</td>
<td>4153</td>
<td>4535 (3)</td>
<td>167</td>
</tr>
<tr>
<td>[M+16H]^{16+}</td>
<td>4153</td>
<td>3788 (3)</td>
<td>358</td>
</tr>
<tr>
<td>[M+17H]^{17+}</td>
<td>3909</td>
<td>4613 (3)</td>
<td>22</td>
</tr>
<tr>
<td>[M+17H]^{17+}</td>
<td>3909</td>
<td>3766 (3)</td>
<td>254</td>
</tr>
</tbody>
</table>

Table S11  Collision cross sections of 60µM bovine serum albumin in 100 mM ammonium acetate
Figure S8  Mass spectrum of 30 µM Concanavalin A tetramer (102 kDa) in 200 mM ammonium acetate

<table>
<thead>
<tr>
<th>Species</th>
<th>m/z</th>
<th>Average CCS (no. of repeats) / Å²</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>[M+19H]^{19+}</td>
<td>5422</td>
<td>4681 (2)</td>
<td>93</td>
</tr>
<tr>
<td>[M+20H]^{20+}</td>
<td>5151</td>
<td>4713 (2)</td>
<td>120</td>
</tr>
<tr>
<td>[M+21H]^{21+}</td>
<td>4906</td>
<td>4804 (3)</td>
<td>17</td>
</tr>
<tr>
<td>[M+22H]^{22+}</td>
<td>4683</td>
<td>4977 (3)</td>
<td>81</td>
</tr>
</tbody>
</table>

Table S12  Collision cross sections of 30 µM Concanavalin A tetramer in 200 mM ammonium acetate
Figure S9  Mass spectrum of 30 µM Serum Amyloid P component pentamer (128 kDa) in 200 mM ammonium acetate

<table>
<thead>
<tr>
<th>Species</th>
<th>m/z</th>
<th>Average CCS of two repeats / Å²</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>[M+22H]^{22+}</td>
<td>5819</td>
<td>5908</td>
<td>101</td>
</tr>
<tr>
<td>[M+23H]^{23+}</td>
<td>5566</td>
<td>6141</td>
<td>86</td>
</tr>
<tr>
<td>[M+24H]^{24+}</td>
<td>5334</td>
<td>6286</td>
<td>76</td>
</tr>
<tr>
<td>[M+25H]^{25+}</td>
<td>5121</td>
<td>6326</td>
<td>102</td>
</tr>
<tr>
<td>[M+26H]^{26+}</td>
<td>4924</td>
<td>6367</td>
<td>60</td>
</tr>
</tbody>
</table>

Table S13  Collision cross sections of 30 µM Serum amyloid P component pentamer in 200 mM ammonium acetate
Figure S10  Charge state range vs. mass for proteins over 150 kDa. Mass spectra of the Hepatitis B virus capsids have been obtained by Uetrecht et al. [1] and P22 complexes and GroEL were taken from Lorenzen et al. [2]

Figure S11  Arrival time distributions for the most intense charge states of lysozyme, cytochrome c, β-casein and α-Synuclein
### Table S14  
Theoretical and experimental maximum and minimum collision cross sections.

<table>
<thead>
<tr>
<th>Protein</th>
<th>Calculated Max.</th>
<th>Calculated Min.</th>
<th>Native Max.</th>
<th>Native Min.</th>
<th>Denatured Max.</th>
<th>Denatured Min.</th>
<th>Reduced Max.</th>
<th>Reduced Min.</th>
</tr>
</thead>
<tbody>
<tr>
<td>α-syn</td>
<td>743</td>
<td>3880</td>
<td>3249</td>
<td>870</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>β-casein</td>
<td>102</td>
<td>5992</td>
<td>5392</td>
<td>1450</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Myoglobin [3]</td>
<td>830</td>
<td>4389</td>
<td>2106</td>
<td>1227</td>
<td>3644</td>
<td>1562</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Lysozyme</td>
<td>738</td>
<td>3669</td>
<td>1906</td>
<td>1061</td>
<td>1895</td>
<td>1222</td>
<td>2989</td>
<td>1313</td>
</tr>
<tr>
<td>Cytochrome c</td>
<td>645</td>
<td>3010</td>
<td>1326</td>
<td>952</td>
<td>2920</td>
<td>1139</td>
<td>1139</td>
<td>1139</td>
</tr>
</tbody>
</table>

Figure S12  
Mass spectrum of 70 µM lysozyme (14 307 Da) in 50 mM ammonium acetate
<table>
<thead>
<tr>
<th>Species</th>
<th>m/z</th>
<th>Average Collision Cross Section (2 repeats)/ Å²</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>[M+4H]^{4+}</td>
<td>3577.76</td>
<td>1061</td>
<td>22.6</td>
</tr>
<tr>
<td>[M+5H]^{5+}</td>
<td>2862.41</td>
<td>1311</td>
<td>43.1</td>
</tr>
<tr>
<td>[M+6H]^{6+}</td>
<td>2385.51</td>
<td>1276</td>
<td>19.8</td>
</tr>
<tr>
<td>[M+7H]^{7+}</td>
<td>2044.87</td>
<td>1276</td>
<td>26.8</td>
</tr>
<tr>
<td>[M+8H]^{8+}</td>
<td>1789.39</td>
<td>1897</td>
<td>10.6</td>
</tr>
<tr>
<td>[M+9H]^{9+}</td>
<td>1590.68</td>
<td>1906</td>
<td>14.1</td>
</tr>
</tbody>
</table>

Table S15  Collision cross sections of 70 µM lysozyme in 50 mM ammonium acetate

Figure S13  Mass spectrum of 70 µM lysozyme in a 49.5:49.5 ratio of methanol to water, with 1% formic acid.
### Table S16  Collision cross sections of 70 µM lysozyme in a 49.5:49.5 ratio of methanol to water, with 1% formic acid.

<table>
<thead>
<tr>
<th>Species</th>
<th>m/z</th>
<th>Average Collision Cross Section from 2 repeats /Å²</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>[M+5H]⁵⁺</td>
<td>2862.41</td>
<td>1222</td>
<td>-</td>
</tr>
<tr>
<td>[M+6H]⁶⁺</td>
<td>2385.51</td>
<td>1246</td>
<td>38.2</td>
</tr>
<tr>
<td>[M+7H]⁷⁺</td>
<td>2044.87</td>
<td>1243</td>
<td>17.0</td>
</tr>
<tr>
<td>[M+8H]⁸⁺</td>
<td>1789.39</td>
<td>1819.5</td>
<td>152.0</td>
</tr>
<tr>
<td>[M+9H]⁹⁺</td>
<td>1590.68</td>
<td>1857</td>
<td>111.7</td>
</tr>
<tr>
<td>[M+10H]¹⁰⁺</td>
<td>1431.71</td>
<td>1823</td>
<td>56.6</td>
</tr>
<tr>
<td>[M+11H]¹¹⁺</td>
<td>1301.65</td>
<td>1895.5</td>
<td>19.1</td>
</tr>
</tbody>
</table>

### Table S17  Average collision cross sections and relative intensities of α-synuclein

<table>
<thead>
<tr>
<th>Species</th>
<th>Average collision cross section of each charge state /Å²</th>
<th>Relative Intensity</th>
</tr>
</thead>
<tbody>
<tr>
<td>[M+5H]⁵⁺</td>
<td>933</td>
<td>0.0063</td>
</tr>
<tr>
<td>[M+6H]⁶⁺</td>
<td>1286</td>
<td>0.085</td>
</tr>
<tr>
<td>[M+7H]⁷⁺</td>
<td>1353</td>
<td>0.3501</td>
</tr>
<tr>
<td>[M+8H]⁸⁺</td>
<td>1457</td>
<td>0.322</td>
</tr>
<tr>
<td>[M+9H]⁹⁺</td>
<td>1606</td>
<td>0.521</td>
</tr>
<tr>
<td>[M+10H]¹⁰⁺</td>
<td>2020</td>
<td>0.6975</td>
</tr>
<tr>
<td>[M+11H]¹¹⁺</td>
<td>2243</td>
<td>0.5364</td>
</tr>
<tr>
<td>[M+12H]¹²⁺</td>
<td>2393</td>
<td>0.9382</td>
</tr>
<tr>
<td>[M+13H]¹³⁺</td>
<td>2392</td>
<td>1</td>
</tr>
<tr>
<td>[M+14H]¹⁴⁺</td>
<td>2457</td>
<td>0.858</td>
</tr>
<tr>
<td>[M+15H]¹⁵⁺</td>
<td>2457</td>
<td>0.5097</td>
</tr>
<tr>
<td>[M+16H]¹⁶⁺</td>
<td>2814</td>
<td>0.2345</td>
</tr>
<tr>
<td>[M+17H]¹⁷⁺</td>
<td>2565</td>
<td>0.082</td>
</tr>
<tr>
<td>[M+18H]¹⁸⁺</td>
<td>2741</td>
<td>0.02369</td>
</tr>
<tr>
<td>[M+19H]¹⁹⁺</td>
<td>2565</td>
<td>0.0111</td>
</tr>
<tr>
<td>[M+20H]²⁰⁺</td>
<td>2607</td>
<td>0.0034</td>
</tr>
</tbody>
</table>
### Appendix 1

<table>
<thead>
<tr>
<th>Species</th>
<th>Average collision cross section of each charge state /Å²</th>
<th>Relative Intensity</th>
</tr>
</thead>
<tbody>
<tr>
<td>[M+7H]^7+</td>
<td>1450</td>
<td>0.0124</td>
</tr>
<tr>
<td>[M+8H]^8+</td>
<td>1658</td>
<td>0.0454</td>
</tr>
<tr>
<td>[M+9H]^9+</td>
<td>1862</td>
<td>0.4678</td>
</tr>
<tr>
<td>[M+10H]^10+</td>
<td>1944</td>
<td>1</td>
</tr>
<tr>
<td>[M+11H]^11+</td>
<td>2157</td>
<td>0.385</td>
</tr>
<tr>
<td>[M+12H]^12+</td>
<td>2375</td>
<td>0.208</td>
</tr>
<tr>
<td>[M+13H]^13+</td>
<td>2646</td>
<td>0.203</td>
</tr>
<tr>
<td>[M+14H]^14+</td>
<td>2877</td>
<td>0.167</td>
</tr>
<tr>
<td>[M+15H]^15+</td>
<td>3088</td>
<td>0.15</td>
</tr>
<tr>
<td>[M+16H]^16+</td>
<td>3443</td>
<td>0.117</td>
</tr>
<tr>
<td>[M+17H]^17+</td>
<td>3606</td>
<td>0.114</td>
</tr>
<tr>
<td>[M+18H]^18+</td>
<td>3806</td>
<td>0.106</td>
</tr>
<tr>
<td>[M+19H]^19+</td>
<td>4103</td>
<td>0.109</td>
</tr>
<tr>
<td>[M+20H]^20+</td>
<td>4300</td>
<td>0.0867</td>
</tr>
<tr>
<td>[M+21H]^21+</td>
<td>4549</td>
<td>0.0843</td>
</tr>
<tr>
<td>[M+22H]^22+</td>
<td>4775</td>
<td>0.0649</td>
</tr>
<tr>
<td>[M+23H]^23+</td>
<td>4877</td>
<td>0.0544</td>
</tr>
<tr>
<td>[M+24H]^24+</td>
<td>4895</td>
<td>0.0399</td>
</tr>
<tr>
<td>[M+25H]^25+</td>
<td>5097</td>
<td>0.03</td>
</tr>
<tr>
<td>[M+26H]^26+</td>
<td>5116</td>
<td>0.022</td>
</tr>
<tr>
<td>[M+27H]^27+</td>
<td>4987</td>
<td>0.0114</td>
</tr>
<tr>
<td>[M+28H]^28+</td>
<td>4655</td>
<td>0.0074</td>
</tr>
</tbody>
</table>

**Table S18**  Average collision cross sections and relative intensities of β-casein

<table>
<thead>
<tr>
<th>Species</th>
<th>Average collision cross section of each charge state /Å²</th>
<th>Relative Intensity</th>
</tr>
</thead>
<tbody>
<tr>
<td>[M+5H]^5+</td>
<td>952</td>
<td>0.049</td>
</tr>
<tr>
<td>[M+6H]^6+</td>
<td>1128</td>
<td>0.13</td>
</tr>
<tr>
<td>[M+7H]^7+</td>
<td>1156</td>
<td>1</td>
</tr>
<tr>
<td>[M+8H]^8+</td>
<td>1290</td>
<td>0.0033</td>
</tr>
</tbody>
</table>

**Table S19**  Average collision cross sections and relative intensities of cytochrome c

### References for Appendix 1

Appendix 2

Supplementary information for Chapter 4

<table>
<thead>
<tr>
<th>Charge State</th>
<th>Rotationally Averaged Collision Cross Section/ Å²</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>[M+4H]^{4+}</td>
<td>982</td>
<td>38</td>
</tr>
<tr>
<td>[M+5H]^{5+}</td>
<td>945</td>
<td>46</td>
</tr>
<tr>
<td>[M+5H]^{5+}</td>
<td>1132</td>
<td>13</td>
</tr>
<tr>
<td>[M+6H]^{6+}</td>
<td>1156</td>
<td>130</td>
</tr>
<tr>
<td>[M+6H]^{6+}</td>
<td>1420</td>
<td>43</td>
</tr>
<tr>
<td>[M+7H]^{7+}</td>
<td>1424</td>
<td>119</td>
</tr>
<tr>
<td>[M+7H]^{7+}</td>
<td>1524</td>
<td>107</td>
</tr>
</tbody>
</table>

Table S1 Rotationally averaged collision cross sections of ApoC-II.

<table>
<thead>
<tr>
<th>Species</th>
<th>m/z</th>
<th>Average collision cross section (no. of repeats)/ Å²</th>
<th>Standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>[M+5H]^{5+}</td>
<td>2893</td>
<td>1043 (2)</td>
<td>133</td>
</tr>
<tr>
<td>[M+6H]^{6+}</td>
<td>2411</td>
<td>1217 (3)</td>
<td>147</td>
</tr>
<tr>
<td>[M+7H]^{7+}</td>
<td>2067</td>
<td>1252 (3)</td>
<td>147</td>
</tr>
<tr>
<td>[M+8H]^{8+}</td>
<td>1809</td>
<td>1333 (3)</td>
<td>181</td>
</tr>
<tr>
<td>[M+9H]^{9+}</td>
<td>1608</td>
<td>1506 (3)</td>
<td>248</td>
</tr>
<tr>
<td>[M+10H]^{10+}</td>
<td>1447</td>
<td>1951 (3)</td>
<td>155</td>
</tr>
<tr>
<td>[M+11H]^{11+}</td>
<td>1316</td>
<td>2161 (3)</td>
<td>173</td>
</tr>
<tr>
<td>[M+12H]^{12+}</td>
<td>1206</td>
<td>2311 (3)</td>
<td>243</td>
</tr>
<tr>
<td>[M+13H]^{13+}</td>
<td>1113</td>
<td>2150 (3)</td>
<td>398</td>
</tr>
<tr>
<td>[M+14H]^{14+}</td>
<td>1034</td>
<td>2446 (3)</td>
<td>136</td>
</tr>
<tr>
<td>[M+15H]^{15+}</td>
<td>965</td>
<td>2476 (3)</td>
<td>161</td>
</tr>
<tr>
<td>[M+16H]^{16+}</td>
<td>904</td>
<td>2560 (3)</td>
<td>229</td>
</tr>
<tr>
<td>[M+17H]^{17+}</td>
<td>851</td>
<td>2617 (3)</td>
<td>153</td>
</tr>
<tr>
<td>[M+18H]^{18+}</td>
<td>804</td>
<td>2742 (2)</td>
<td>11</td>
</tr>
<tr>
<td>[M+19H]^{19+}</td>
<td>762</td>
<td>2605 (3)</td>
<td>183</td>
</tr>
<tr>
<td>[M+20H]^{20+}</td>
<td>724</td>
<td>2620 (1)</td>
<td>-</td>
</tr>
</tbody>
</table>

Table S2 Collision cross sections of 50 µM α-synuclein in 50 mM ammonium acetate
Appendix 2

<table>
<thead>
<tr>
<th>Peptide</th>
<th>Theoretical maximum deuterium uptake</th>
<th>Mass increase at apex of isotopic distribution / Da</th>
<th>Maximum mass increase/ Da</th>
</tr>
</thead>
<tbody>
<tr>
<td>α-Syn residues 5-17</td>
<td>11</td>
<td>5</td>
<td>10</td>
</tr>
<tr>
<td>α-Syn residues 55-69</td>
<td>14</td>
<td>9</td>
<td>14</td>
</tr>
<tr>
<td>α-Syn residues 106-115</td>
<td>9</td>
<td>6</td>
<td>9</td>
</tr>
<tr>
<td>ApoC-II residues 7-15</td>
<td>8</td>
<td>4</td>
<td>8</td>
</tr>
<tr>
<td>ApoC-II residues 37-52</td>
<td>15</td>
<td>8</td>
<td>15</td>
</tr>
<tr>
<td>ApoC-II residues 61-67</td>
<td>6</td>
<td>5</td>
<td>6</td>
</tr>
</tbody>
</table>

Table S3  Deuterium uptake of ApoC-II and α-Syn fragments. The theoretical maximum uptake is n-1, where n is the number of amino acids in the peptide [1].

Figure S1: Deconvoluted mass spectra of p27-C-WT (a, b), p27-C-κ0.14 (c, d) and p27-C-κ0.56 (f, g). Starting solutions were 10 mM ammonium acetate (a, c, f) or 200 mM ammonium acetate (b, d, g).
<table>
<thead>
<tr>
<th>Construct</th>
<th>Salt Concentration</th>
<th>Gaussian Curve</th>
<th>Curve maximum</th>
<th>% of total area</th>
</tr>
</thead>
<tbody>
<tr>
<td>p27-C-WT</td>
<td>10 mM</td>
<td>WT&lt;sub&gt;1&lt;/sub&gt;</td>
<td>6.5</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td></td>
<td>WT&lt;sub&gt;2&lt;/sub&gt;</td>
<td>9</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td></td>
<td>WT&lt;sub&gt;3&lt;/sub&gt;</td>
<td>11.6</td>
<td>76</td>
</tr>
<tr>
<td></td>
<td>200 mM</td>
<td>WT&lt;sub&gt;1&lt;/sub&gt;</td>
<td>6.4</td>
<td>62</td>
</tr>
<tr>
<td></td>
<td></td>
<td>WT&lt;sub&gt;2&lt;/sub&gt;</td>
<td>9.1</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>WT&lt;sub&gt;3&lt;/sub&gt;</td>
<td>12.3</td>
<td>33</td>
</tr>
<tr>
<td>p27-C-κ0.14</td>
<td>10 mM</td>
<td>κ14&lt;sub&gt;1&lt;/sub&gt;</td>
<td>6.5</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td></td>
<td>κ14&lt;sub&gt;2&lt;/sub&gt;</td>
<td>8.9</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td></td>
<td>κ14&lt;sub&gt;3&lt;/sub&gt;</td>
<td>9.9</td>
<td>43</td>
</tr>
<tr>
<td></td>
<td></td>
<td>κ14&lt;sub&gt;4&lt;/sub&gt;</td>
<td>11.3</td>
<td>24</td>
</tr>
<tr>
<td></td>
<td>200 mM</td>
<td>κ14&lt;sub&gt;1&lt;/sub&gt;</td>
<td>6.4</td>
<td>21</td>
</tr>
<tr>
<td></td>
<td></td>
<td>κ14&lt;sub&gt;2&lt;/sub&gt;</td>
<td>9</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td></td>
<td>κ14&lt;sub&gt;3&lt;/sub&gt;</td>
<td>10.6</td>
<td>40</td>
</tr>
<tr>
<td></td>
<td></td>
<td>κ14&lt;sub&gt;4&lt;/sub&gt;</td>
<td>12.8</td>
<td>21</td>
</tr>
<tr>
<td>p27-C-κ0.56</td>
<td>10 mM</td>
<td>κ56&lt;sub&gt;1&lt;/sub&gt;</td>
<td>6</td>
<td>30</td>
</tr>
<tr>
<td></td>
<td></td>
<td>κ56&lt;sub&gt;2&lt;/sub&gt;</td>
<td>8.1</td>
<td>66</td>
</tr>
<tr>
<td></td>
<td></td>
<td>κ56&lt;sub&gt;3&lt;/sub&gt;</td>
<td>11.2</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>200 mM</td>
<td>κ56&lt;sub&gt;1&lt;/sub&gt;</td>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>κ56&lt;sub&gt;2&lt;/sub&gt;</td>
<td>9</td>
<td>93</td>
</tr>
<tr>
<td></td>
<td></td>
<td>κ56&lt;sub&gt;3&lt;/sub&gt;</td>
<td>11</td>
<td>4</td>
</tr>
</tbody>
</table>

Table S1  Parameters for the Gaussian curves in figure S1.
Figure S2   p27-C-κ0.14 analysed by IM-MS. Arrival time distributions of each charge state sprayed from 10 mM ammonium acetate (grey) and 200 mM ammonium acetate (purple).

Figure S3   p27-C-κ0.56 analysed by IM-MS. Arrival time distributions of each charge state sprayed from 10 mM ammonium acetate (grey) and 200 mM ammonium acetate (purple).
Figure S4  IM-MS data of p27-C-WT (a, b), p27-C-κ0.14 (c, d) and p27-C-κ0.56 (f, g). Starting solutions were 10 mM ammonium acetate (a, c, f) or 200 mM ammonium acetate (b, d, g). The CCS distribution of each charge state is shown. The height is normalised to that of the MS peak shown in figure 1.
Appendix 4

Supplementary information for Chapter 6

Materials and Methods

Proteins were provided by Richard Kriwacki, St. Jude Children's Research Hospital, Memphis, TN. The purified proteins were buffer exchanged into 200mM ammonium acetate using Bio-Rad micro Bio-Spin P-6 Columns. The Cdk2/ Cyclin A was provided as a dimer. This was incubated with p27 for 30 minutes at room temperature at a 1:1 ratio for experiments on the trimeric complex.

Nano-electrospray ionisation (nESI); n-ESI was employed for all MS and IM-MS experiments. 15 μM protein solutions in 200mM ammonium acetate were ionised from a thin-walled glass capillary (inner diameter 0.9mm, outer diameter 1.2mm, World Precision Instruments, Stevenage, UK) that was pulled to a nESI tip in house with a Flaming/Brown micropipette puller (Sutter Instrument Co., Novato, CA). A positive potential of 1.6 kV was applied to the solutions via a thin platinum wire.

Mass Spectrometry; All MS experiments were performed on a Q-ToF Global (Waters, Manchester, UK). The following instrument parameters were utilised; Cone voltage; 60V. Source pressure; 3.3 mbar. Source temperature; 100˚C. Collision cell pressure; 6e-3 mbar. Collision energy; 5V.

Ion mobility-mass spectrometry; IM-MS experiments were carried out on a Waters Q-ToF I instrument that was modified in house to include a 5.1cm drift tube which has been described elsewhere [1]. The temperature and pressure of helium in the drift cell were approximately 28˚C and 4 Torr respectively. Measurements were made at 6 different drift voltages from 60V to 20V. The precise pressure and temperature was recorded for every drift voltage and used in the calculations of CCSs. Each experiment was performed in triplicate. Ion arrival time distributions were recorded by synchronisation of the release of ions into the drift cell with mass.
Appendix 4

spectral acquisition. The CCS distribution plots are derived from raw arrival time data using Equation 2 below [2].

\[ \Omega_{\text{avg}} = \frac{(18\pi)^{1/2}}{16} \left( \frac{1}{m} + \frac{1}{m_b} \right)^{1/2} \frac{z e}{(K_B T)^{1/2} \rho} \frac{1}{L^2} t_d V \]

Equation S1

Where \( m \) and \( m_b \) are the masses of the ion and buffer gas, respectively; \( z \) is the ion charge state; \( e \) is the elementary charge; \( K_B \) is the Boltzmann constant; \( T \) is the gas temperature; \( \rho \) is the buffer gas density; \( L \) is the drift tube length; \( V \) is the voltage across the drift tube; and \( t_d \) is the drift time.

The raw arrival time output (\( t_a \)) includes time the ions spend outside of the drift cell but within the mass spectrometer, known as the dead time (\( t_0 \)). The value for \( t_0 \) is calculated by taking an average value of the intercept from a linear plot of average arrival time versus pressure/temperature and was subtracted from the arrival time to calculate drift time (\( t_D \)):

\[ t_D = t_a - t_0 \]

Equation S2

All MS and IM-MS data were analysed using Masslynx v4.1 software (Waters, Manchester, UK), Origin v8.5 (Originlab Corporation, USA) and Microsoft Excel.

Calculation of CCSs from crystal structure: Published x-ray crystallography measurements were taken from the RSCB protein data bank (PDB) and hydrogens were added using Avogadro [3]: an open-source molecular builder and visualization tool. Version 1.1.1. http://avogadro.openmolecules.net/. The structure was minimised and the CCS was calculated using the trajectory method in MOBCAL [4].
Variations in the charge state distributions occurred on a day-to-day basis, when the experiments were performed under the exact same conditions and instrumental parameters. This variation in CSD is a result of the plasticity of the proteins. Because of this, the experiments were repeated on three consecutive days following identical procedures, using freshly defrosted aliquots each day. The height of the MS peaks were then all normalised to 1 and the average height of each charge state was calculated across the three different days. The error bars represent the standard deviation. \( \kappa 0.14 \) always has the highest charge state envelope as the most
Appendix 4

intense as shown by the low error bars in this region. WT and k0.56 both have a higher abundance of lower charged ions than k0.14. The CCS distributions are all incredibly broad, it is very difficult to infer any information because all the permutants are so flexible.

![CCS distributions](image)

**Figure S2**  CCS distributions of each charge state of the Cdk2/ Cyclin A/ p27 complexes.

Appendix 5

Supplementary information for Chapter 7

Figure S1  MS spectra of Fdc1 (top) and Fdc1UbiX (bottom) showing some dissociation of the Fdc1 dimer into monomer, while all of the Fdc1UbiX is present as a dimer.
Figure S2  Fdc1UbiX with harsh ionization conditions (cone voltage= 200V, top) and softer ionization conditions (cone voltage= 60V, bottom). The ratio of doubly bound to singly bound cofactor can be modulated by altering the cone voltage.

Figure S3  CCS distributions of Fdc1 and Fdc1UbiX containing two bound cofactors, one bound cofactors and no bound cofactors.
Figure S4; The relative fractional uptake of deuterium of Fdc1UbiX (top) and Fdc1 (bottom) at $t=15s$ (green), $t=5m$ (blue), $t=60m$ (red) and $t=240m$ (black). Areas of the sequence for which no coverage could be achieved are denoted as uptake=0.
Figure S5. Uptake plots of selected peptides of Fdc1 (blue) and Fdc1\textsuperscript{UbiX} (red).