Characterising Disordered Proteins of the Cancer Genome using Biophysical Techniques

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

2016

Eleanor R. Dickinson

School of Chemistry
Table of Contents

Table of Figures ........................................................................................................... 9
Table of Equations ....................................................................................................... 11
Table of Tables ............................................................................................................. 11
List of Abbreviations ................................................................................................... 11
Abstract ....................................................................................................................... 15
Declaration .................................................................................................................. 16
Copyright Statement ...................................................................................................... 17
Acknowledgements ....................................................................................................... 18

1 Introduction .................................................................................................................. 20
  1.1 Protein structure and folding ............................................................................... 21
  1.2 (Un)structure: Levels of structure in proteins ................................................... 22
    1.2.1 Intrinsically Disordered Proteins: Definition and Properties ...................... 24
    1.2.2 IDP function and regulation ........................................................................ 25
  1.3 Disorder in disease, the D^2 concept ................................................................... 27
  1.4 p53: “The Death Star” ....................................................................................... 29
    1.4.1 Structure and Function ............................................................................... 29
    1.4.2 Post-translational modifications .................................................................. 33
    1.4.3 Regulation of p53: The p53 – Mdm2 interaction .......................................... 36
    1.4.4 p53 and cancer ............................................................................................. 39
  1.5 Identification and prediction of IDPs or disordered regions .............................. 40
    1.5.1 X-ray crystallography ................................................................................. 41
    1.5.2 NMR spectroscopy ..................................................................................... 41
    1.5.3 Small Angle X-ray Scattering (SAXS) ....................................................... 42
    1.5.4 Circular Dichroism ..................................................................................... 43
    1.5.5 Other Biophysical Techniques ..................................................................... 44
    1.5.6 Disorder Predicting Algorithms ................................................................... 44
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.3</td>
<td>Introduction</td>
<td>93</td>
</tr>
<tr>
<td>2.4</td>
<td>Experimental Methods</td>
<td>96</td>
</tr>
<tr>
<td>2.4.1</td>
<td>Mass Spectrometry and Variable Temperature Ion Mobility Mass Spectrometry</td>
<td>96</td>
</tr>
<tr>
<td>2.4.2</td>
<td>Proteins</td>
<td>97</td>
</tr>
<tr>
<td>2.5</td>
<td>Results and Discussion</td>
<td>97</td>
</tr>
<tr>
<td>2.5.1</td>
<td>cytochrome c</td>
<td>97</td>
</tr>
<tr>
<td>2.5.1.1</td>
<td>VT-MS cytochrome c</td>
<td>97</td>
</tr>
<tr>
<td>2.5.1.2</td>
<td>VT-IM-MS cytochrome c</td>
<td>100</td>
</tr>
<tr>
<td>2.5.1.3</td>
<td>cytochrome c Discussion</td>
<td>100</td>
</tr>
<tr>
<td>2.5.2</td>
<td>p53 DBD</td>
<td>103</td>
</tr>
<tr>
<td>2.5.2.1</td>
<td>VT-MS p53 DBD</td>
<td>103</td>
</tr>
<tr>
<td>2.5.2.2</td>
<td>VT-IM-MS p53 DBD</td>
<td>103</td>
</tr>
<tr>
<td>2.5.2.3</td>
<td>p53 DBD Discussion</td>
<td>105</td>
</tr>
<tr>
<td>2.5.3</td>
<td>N terminal MDM2</td>
<td>105</td>
</tr>
<tr>
<td>2.5.3.1</td>
<td>VT-MS N-terminal MDM2</td>
<td>105</td>
</tr>
<tr>
<td>2.5.3.2</td>
<td>VT IM-MS N-Terminal MDM2</td>
<td>106</td>
</tr>
<tr>
<td>2.5.3.3</td>
<td>N-Terminal MDM2 Discussion</td>
<td>107</td>
</tr>
<tr>
<td>2.6</td>
<td>Conclusions</td>
<td>108</td>
</tr>
<tr>
<td>2.7</td>
<td>References</td>
<td>109</td>
</tr>
<tr>
<td>3</td>
<td>The use of Ion Mobility Mass Spectrometry to probe Modulation of the Structure of p53 and of MDM2 by Small Molecule Inhibitors</td>
<td>112</td>
</tr>
<tr>
<td>3.1</td>
<td>Declaration</td>
<td>113</td>
</tr>
<tr>
<td>3.2</td>
<td>Abstract</td>
<td>115</td>
</tr>
<tr>
<td>3.3</td>
<td>Introduction</td>
<td>115</td>
</tr>
<tr>
<td>3.4</td>
<td>Materials and Methods</td>
<td>118</td>
</tr>
<tr>
<td>3.5</td>
<td>Results</td>
<td>120</td>
</tr>
<tr>
<td>3.5.1</td>
<td>Modulation of N-terminal p53 by RITA</td>
<td>120</td>
</tr>
<tr>
<td>3.5.2</td>
<td>Modulation of N-terminal MDM2 by Nutlin-3</td>
<td>125</td>
</tr>
</tbody>
</table>
4  The Contribution of Proline to the Conformational Landscape of the p53
Transcriptional Activation Domain........................................................................135

4.1  Declaration....................................................................................................136
4.2  Abstract........................................................................................................137
4.3  Introduction ..................................................................................................137
4.4  Methods ........................................................................................................140
  4.4.1  Mass Spectrometry and Ion Mobility Mass Spectrometry ..................140
  4.4.2  Molecular Modelling..............................................................................142
    4.4.2.1  Ensemble generation.........................................................................142
    4.4.2.2  Theoretical collision cross sections..................................................143
    4.4.2.3  Derivation of CCSs from $R_g$..........................................................143
  4.5  p53 constructs exhibit a distinct conformational phenotype....................143
  4.6  Can IM-MS delineate the increased helical content?.................................147
  4.7  Comparisons to computational data ..........................................................149
  4.8  Conclusions ...............................................................................................153
  4.9  References ..................................................................................................154

5  Proline residues mediate promiscuous binding of p53 to Mdm2.................158

5.1  Declaration....................................................................................................159
5.2  Abstract........................................................................................................160
5.3  Introduction ..................................................................................................160
5.4  Materials and Methods..............................................................................164
  5.4.1  Protein preparation..................................................................................164
  5.4.2  Mass Spectrometry and Ion Mobility Mass Spectrometry .................165
5.5  Increase in residual helicity of p53 results in a change in conformation ....167
5.6  p53 residual helicity correlates with Mdm2 binding affinity ....................169
5.7 Residual helicity does not affect complex conformation .............................................. 172
5.8 Mutant p53TAD undergoes distinct conformational changes during Mdm2 binding 173
5.9 Conclusions .................................................................................................................. 178
5.10 References .................................................................................................................... 179
6 Protein-protein Interactions of MAGE-A4 with gankyrin and p53 ......................... 182
   6.1 Declaration ................................................................................................................... 183
   6.2 Abstract ....................................................................................................................... 184
   6.3 Introduction .................................................................................................................. 184
   6.4 Materials and Methods ............................................................................................... 185
      6.4.1 Cloning .................................................................................................................... 185
      6.4.2 Protein Expression and Purification ........................................................................ 186
      6.4.3 Protein Labelling with $^{15}$N for $^1$H-$^{15}$N HSQC NMR ................................... 187
      6.4.4 $^1$H-$^{15}$N HSQC NMR ....................................................................................... 187
      6.4.5 MS and IM-MS for Protein-Protein Interaction Studies ...................................... 188
      6.4.6 MS for Protein-DNA Interaction Studies ............................................................... 189
      6.4.7 EMSA .................................................................................................................... 189
   6.5 Results .......................................................................................................................... 190
      6.5.1 Protein-Protein Interaction between MAGE-A4 and Gankyrin Probed by $^1$H-$^{15}$N HSQC NMR .......................................................... 190
      6.5.2 Protein-Protein Interactions between MAGE-A4 Proteins and Gankyrin Probed by Nanoelectrospray Ionisation Mass Spectrometry (nESI-MS) and Ion Mobility Mass Spectrometry (IM-MS) ........................................................................ 192
      6.5.3 HSQC NMR for the protein-protein interaction between MAGE-A4 and p53 DBD 194
      6.5.4 Protein-protein interaction between MAGE-A4 and p53 probed by nESI-MS and IM-MS ..................................................................................................................................................................................... 197
      6.5.5 Interaction studies of MAGE-A4 with p53 downstream pro-apoptotic and cell cycle arrest DNA promoters using EMSA ................................................................................................................ 198
Appendix E ......................................................................................................................... 253
Supplementary Figures ................................................................................................. 253
Appendix F ....................................................................................................................... 258

Word count: 68432
Table of Figures

Figure 1.1 Proteins to represent classes in the progression of structure .......................... 22
Figure 1.2 The energy landscape model illustrating native globular proteins .................. 24
Figure 1.3 Examples of 14 binding partners of p53 ................................................................................. 30
Figure 1.4 Schematic showing major domains of p53 ............................................................ 32
Figure 1.5 Schematic showing p53 domain structure and post-translational modification ...... 34
Figure 1.6 The p53:MDM2 autoregulatory feedback loop ........................................................... 37
Figure 1.7 The p53:MDM2 interaction .......................................................................................... 38
Figure 1.8 CD spectra showing characteristic signals ............................................................ 43
Figure 1.9 Schematic showing the ESI process in positive mode .............................................. 50
Figure 1.10 Summary of ESI proposed mechanisms ................................................................. 52
Figure 1.11 Schematic of a quadrupole mass analyser ................................................................. 54
Figure 1.12 Schematic of a ToF mass analyser ............................................................................ 57
Figure 1.13 Schematic of the drift tube ion mobility mass spectrometer ............................... 61
Figure 1.14 a) Schematic representation of the drift cell and associated lenses .................. 63
Figure 1.15 The data collection and fitting process ................................................................. 65
Figure 1.16 Schematic of a typical HDX experiment ................................................................. 68
Figure 2.1 Graphical abstract ................................................................................................. 92
Figure 2.2 a) X-ray diffraction crystal structure, protein sequence, Disopred disorder prediction ... 95
Figure 2.3 nESI mass spectra of 50µM cytochrome c sprayed from 3:1 H2O:ACN .................. 99
Figure 2.4 a) b) Stacked plots showing CCS distributions derived from arrival time distributions .... 101
Figure 2.5 CCS distributions derived from arrival time distributions for 50 µM p53 DBD .......... 104
Figure 2.6 a), b and c show CCS distributions at a range of temperatures for NT-MDM2 .......... 107
Figure 3.1 n-ESI mass spectra recorded for a) wild-type Np53 .................................................. 121
Figure 3.2 Collision cross section distributions (\(\text{CCSD}_d\)) .................................................... 122
Figure 3.3 a) c) deuterium uptake mass spectra for Np53 peptide residues .......................... 124
Figure 3.4 nESI mass spectra of a) 50 µM N-terminal MDM2 .................................................. 126
Figure 3.5 Collision cross section distributions (\(\text{CCSD}_w\)) .................................................... 127
Figure 4.1 Composite figure showing a) mass spectrum of WT p53 ........................................ 138
Figure 4.2 Cumulative collision cross section distributions (CCSDs) ..................................... 145
Figure 4.3a) circular dichroism spectra showing each p53TAD construct ............................ 149
Figure 4.4 Experimental and computational CCSDs for WT p53 ............................................. 151
Figure 5.1 Composite figure showing side and top views of X-ray crystal structure ............... 162
Figure 5.2 a) nESI mass spectrum of p53TAD ........................................................................ 168
Figure 5.3 nESI mass spectra for a) WTP53 and N-Mdm2 ......................................................... 170
Figure 5.4 CCS distributions derived from ATDs .................................................................. 174
Figure 5.5 CCS distributions derived from ATDs .................................................................. 176
Figure 6.1 $^{1}$H-15N HSQC NMR.................................................................................................................. 192
Figure 6.2 a) nESI mass spectrum of MAGE-A4 and gankyrin................................................................. 194
Figure 6.3 $^{1}$H-15N HSQC NMR.................................................................................................................. 197
Figure 6.4 a) nESI mass spectrum obtained of MAGE-A4 and p53 DBD.................................................... 198
Figure 6.5 EMSA studies ............................................................................................................................... 200
Figure 6.6 EMSA studies ............................................................................................................................... 201
Figure 6.7 nESI mass spectra ......................................................................................................................... 203
Figure 6.8 nESI mass spectra ......................................................................................................................... 204
Figure 6.9 Proposed model for the biological functions of MAGE-A4. ....................................................... 208

Figure S 1 a) CCSD for cytochrome c [M+5H]$^{5+}$...................................................................................... 222
Figure S 2 nESI mass spectra of 50 µM DBD p53....................................................................................... 223
Figure S 3 nESI mass spectra of 50µM N-terminal MDM2........................................................................ 224
Figure S 4 CCS distributions at increasing buffer gas temperature ............................................................ 225
Figure S 5 Collision cross section distribution (CCSD)............................................................................. 231
Figure S 6 Collision cross section distribution (CCSD)............................................................................. 232
Figure S 7 Circular dichroism spectra of wild-type Np53........................................................................... 232
Figure S 8 nESI mass spectra of a) 50 µM N-MDM2.................................................................................. 233
Figure S 9 Waterfall plot representing in vacuo conformations of N-terminal MDM2............................... 234
Figure S 10 Positive Ca random-coil chemical shift deviations................................................................... 238
Figure S 11 nESI spectra in negative ionisation mode.................................................................................. 238
Figure S 12 Collision cross section distributions ...................................................................................... 239
Figure S 13 Collision cross section distributions...................................................................................... 240
Figure S 14 Collision cross section distributions...................................................................................... 241
Figure S 15 Collision cross section distributions...................................................................................... 242
Figure S 16 Collision cross section distributions...................................................................................... 243
Figure S 17 nESI spectra in positive ionisation mode.................................................................................. 244
Figure S 18 CCS$_{EHS}$ vs R$_g$ for (top left) random, (top right) SAXS weighted ...................................... 244
Figure S 19 CCSD derived from arrival time distributions for all charge states ....................................... 245
Figure S 20 nESI mass spectra ................................................................................................................... 247
Figure S 21 CCSDs derived from ATDs for each charge state of WT p53...................................................... 248
Figure S 22 CCSDs derived from ATDs....................................................................................................... 249
Figure S 23 nESI mass spectra for a) p53$^{12,13A}$ and N-Mdm2 incubated in varying ratios..................... 250
Figure S 24 Overlaid cumulative CCSDs.................................................................................................... 251
Figure S 25 Overlaid CCSDs derived from arrival time distributions for N-Mdm2.................................... 252
Figure S 26 nESI mass spectrum of 10 µM MAGE-A4................................................................................ 253
Figure S 27 nESI mass spectrum for 10 µM gankyrin.................................................................................. 254
Figure S 28 nESI mass spectrum of 10 µM p53 DNA binding domain...................................................... 255
Figure S 29 nESI spectra of a) p53 DBD incubated incubated in increasing ratios................................. 256
Table of Equations

Equation 1.1 \[ \pm \Phi_0 = \pm U - V \cos \omega t \] .......................................................... 55
Equation 1.2 \[ EK = mv^2 = zeV \] .......................................................... 55
Equation 1.3 \[ t = Lv \] ................................................................................. 56
Equation 1.4 \[ t2 = mz = L22eV \] .......................................................... 56
Equation 1.5 \[ vd = KE = LtD \] .......................................................... 60
Equation 1.6 \[ K0 = K0PTP0 \] ................................................................................. 60
Equation 1.7 \[ K0 = 3ze16N2\pi\mu_kBT121\Omega \] .......................................................... 60
Equation 1.8 \[ td = ta - t0 \] ................................................................................. 66
Equation 1.9 \[ \Omega_{avg} = 18\pi1/2161mb + 1m1/2zeKBT1/21ptDVL2 \] ........................................... 67
Equation 5.3 Fraction of Complex ion intensity = \[ \Sigma P \Sigma P + \Sigma PL \] ........................................... 166

Equation S 2 \[ \Omega_{avg} = \pm \Phi_0 = \pm U - V \cos \omega t \] .......................................................... 227

Table of Tables

Table S 1 Centroid CCS values for all conformations assigned ........................................... 237
Table S 2 conformational families assigned to CCSDs ....................................................... 246

List of Abbreviations

AC – alternating current
ATD - arrival time distribution
CBP – CREB-binding protein
CCS - collision cross section
CCSD - collision cross section distribution
CD – circular dichroism
CEM - chain ejection model

CRM - charged residue model

CSD – charge state distribution

DBD – DNA-binding domain

DC – direct current

DT IM-MS - drift tube ion mobility mass spectrometry

$^{12} \text{CCSD}_{\text{he}}$ – drift time collision cross section measured in helium

EHSS - Exact Hard Sphere Scattering

ESI – electrospray ionisation

FCS – fluorescence correlation spectroscopy

FRET - Förster resonance energy transfer

HAT – histone acetyl transferase

HDMS - high definition mass spectrometer

HDX-MS - hydrogen deuterium exchange mass spectrometry

IDP – intrinsically disordered protein

IDR – intrinsically disordered region

IE - injection energy

IEM - ion evaporation model

IM-MS - ion mobility mass spectrometry

IMPACT - Ion Mobility Projection Approximation Calculation Tool

KMT – lysine methyltransferase

m/z – mass-to-charge ratio

MALDI - matrix-assisted laser desorption ionisation

MC – Monte Carlo
MCP - microchannel plates detector
MD – molecular dynamics
MDM2 – Murine Double Minute 2
MHD – Mage homology domain
MOMP – mitochondrial outer membrane permeabilisation
MoQToF - Mobility Quadrupole Time-of-Flight
MoRFs – Molecular Recognition motifs
mRNA – messenger RNA
MS – mass spectrometry
nESI – nanoelectrospray ionisation
N-Mdm2 - N-terminal domain of Mdm2 (residues 17-125)
NMR – Nuclear Magnetic Resonance
Np53 - N-terminal transactivation domain of p53
NT-MDM2 – N-terminal Murine Double Minute2 (residues 1-126)
p53AIP1 – p53-mediated Apoptosis-Inducing Protein 1
p53TAD – p53 transactivation domain
PA - projection approximation
PCA – principle component analysis
PCAF – P300/CBP-associated factor
PDB – Protein Databank
PFG – pulse-field gradient
PLGS - ProteinLynx Global Server
PONDR - Predictor of Natural Disordered Regions
PPI – protein – protein interaction
PRE – paramagnetic relaxation enhancement

PreSMos – pre-structural motifs

PRMT5 - protein arginine N-Methyl Transferase 5

PSA - Projection Superposition Approximation

PTMs – post-translational modifications

QToF – quadrupole time-of-flight

RF – radio frequency

$R_g$ – radius of gyration

RING – Really Interesting New Gene

RITA – Reactivation of p53 and induction of tumour cell apoptosis

$R_s$ – Stokes radius

SAXS – Small Angle X-ray Scattering

TAD – transactivation domain

tATD - total arrival time distribution

TDC - time-to-digital

TIC - total ion count

TIMS - travelling wave ion guide ion mobility

TM - Trajectory Method

ToF - time of flight

UV – ultraviolet

VT IM-MS - variable temperature drift time ion mobility mass spectrometry

WT – wild-type
Abstract
The University of Manchester
Eleanor R. Dickinson
Doctor of Philosophy
Characterising Disordered Proteins of the Cancer Genome using Biophysical Techniques
26 September 2016

Protein function and dysfunction, and their intimate ties to protein structure, has been a core focus of research for several decades. More recently, research into the lack of structure in proteins has reached fever pitch. Intrinsically disordered proteins (IDPs) are proteins (or protein regions) that exist as collapsed or extended, dynamically mobile conformational ensembles, either at secondary or tertiary level, whilst remaining biologically active. The properties of IDPs can impede their study; they are often inherently unstable, are vastly wide-ranging in molecular weight and often difficult to express in large quantities. Mass spectrometry (MS) has evolved into a tool for the study of dynamic systems such as IDPs due to its large dynamic range, high sensitivity, low sample consumption and its lack of bias towards the folded state of a protein. The addition of ion mobility separation to mass spectrometry analysis (IM-MS) provides insight into the conformations adopted by proteins and their complexes, measuring their rotationally averaged collision cross section which can be compared with coordinates from other biophysical techniques such as X-ray crystallography, NMR and to molecular modelling. The work presented in this thesis uses both MS and IM-MS, along with several other biophysical techniques, to interrogate a number of IDPs which are implicated in cancer.

Firstly, variable temperature IM-MS is used to probe several proteins of increasing disorder; structured protein cytochrome c, the tumour suppressor protein p53 and the oncoprotein Murine Double Minute 2 (Mdm2), performing IM-MS measurements at a range of temperature from 200 K to 571 K to elucidate the gas-phase unfolding behaviour of each protein. The interaction between p53 and Mdm2 is a current target for cancer drug therapy. Hence MS and IM-MS, alongside circular dichroism and hydrogen-deuterium exchange are next employed to determine the effect of several known small molecule ligands on the conformations adopted by these disordered proteins upon binding to their respective ligands. The significant structuring of both of these disordered proteins upon binding to their respective ligands can be observed using IM-MS, but is not apparent when using other biophysical techniques, highlighting the ability of IM-MS to capture conformational changes occurring in solution on a short timescale. The regulation of disorder in cells is postulated to be mediated by proline residues. I investigate the impact of proline replacement on the populations of conformers presented by p53 using a range of mutants and then go on to study how these mutations impact upon the binding stoichiometry, affinity and conformational preference of p53 for its interaction partner Mdm2. Finally, the disordered melanoma associated antigen 4 MAGE-A4, and its ability to bind to p53 and block its transcriptional activity is probed using MS and IM-MS.
Declaration

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

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

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

iii. The ownership of certain Copyright, patents, designs, trade marks 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 University IP Policy (see http://documents.manchester.ac.uk/display.aspx?DocID=24420), in any relevant Thesis restriction declarations deposited in the University Library, The University Library’s regulations (see http://www.library.manchester.ac.uk/about/regulations/) and in The University’s policy on Presentation of Theses.
Acknowledgements

My first and most important thanks go to my wonderful supervisor, Professor Perdita Barran. Thank you for your guidance, support, great ideas and excellent parties. Thank you for teaching me how to be a scientist.

I have had the opportunity to collaborate with some amazing people during my PhD. Thanks to Ted Hupp and Dave Clarke at the University of Edinburgh, Penka Nikolova and Yoshio Hagiwara at Kings College London, Joanna Zawacka-Pankau and Galina Selivanova at the Karolinska Institutet, Gary Daughdrill and Wade Borcherds at the University of South Florida and to my project students; Leonora, Alexandra and Ellie. Thank you for the proteins, conversations and comments. I have loved working on a range of different problems and projects with you.

My PhD has been spent between Edinburgh and Manchester, and all of the PBRG members from both cities have contributed to making the last four years so much fun. Jason, Hattie, Yana, Ewa, Kamila, Sophie, Bex, Jakub, Ashley, Chris N, Chris G, Reynard, Bruno, Rosie, Alina, Cunyu, Jacky, Bin, Ellie 2, Aidan, Emily and a special mention to Lukasz for helping me so much with everything computational. Thanks for all the pub trips, tea breaks, hill climbs, silliness, cards, hugs, flat parties and for being all-round great.

A huge mention to all of my “non-science” friends spread around the country (and the world) who have kept me sane, both before and during “The PhD”. A special thanks to Kim, Bex, Lizzie, Holly and Jade – my favourites from day one. To James, thanks for your love, support and for not letting my head get too big. A thank you too to my family, especially to Maggie, Alan and my sister Imogen for all your support, hugs and enthusiasm.

Finally, the biggest thanks go to my parents, Lynn and Peter. I am so grateful for your constant love and support, and for giving me so many opportunities over the years. I love you. Thank you.
“I almost wish I hadn’t gone down that rabbit hole – and yet – and yet – it’s rather curious, you know, this sort of life!”

Lewis Carroll, Alice in Wonderland
1

Introduction
1. Introduction

1.1 Protein structure and folding

Proteins are major components within a living cell where they are crucial to maintaining cellular life. Proteins play roles in all cellular processes; cell division, gene transcription, cellular transportation, immune response and apoptosis, to name a few. The primary structure of proteins comprises a sequence of amino acid monomers enzymatically linked to form a polypeptide chain. All eukaryotic proteins are built from L-amino acids, of which there are 20 common amino acids. The chemical composition of each amino acid side chain is varied, with differences in charge, size, shape, reactivity, and hydrogen binding capacity. This allows the function of proteins within the cell to be remarkably wide-ranging. The secondary structure of a protein is a higher order arrangement of the polypeptide backbone, in which amino acid residues close in distance to each other form interactions between their respective sidechains. Secondary structural elements are stabilised by hydrogen bonding between the amino- and carboxy-groups of the peptide bonds. The most common secondary structures are α-helices, β-strand, turns and coils [1]. Tertiary structure includes all secondary structures alongside long-range interactions between amino acids. These are stabilised by both covalent (disulphide bridges) and non-covalent (hydrophobic and electrostatic interactions) bonding. If a protein is comprised of multiple polypeptide chains, these interact to form the quaternary structure.

Historically, there has been a long-standing belief that the tertiary structure of a protein determines the specific role it plays within the cell. Known as the structure-function paradigm, the origins lie in 1894 when a “lock-and-key” mechanism to explain the hydrolysis of different glucoside multimers by specific enzymes was proposed by Emil Fischer [2]. The notion that only the correctly shaped/sized substrate would be able to bind into the active site of a protein was supported by emerging protein structures solved by x-ray diffraction [3, 4]. Since then, the field of structural biology has grown significantly with over 120,000 sets of atomic coordinates for proteins now deposited into the Protein Data Bank (PDB, rcsb.org). More recently, there have been challenges to the classical structure-function paradigm. It has been systematically shown that many proteins which are partially or completely disordered are fully functional [5-7], alongside confirmation that antibody binding depended on unfolded, rather than structured protein [8], giving rise to idea of conformational selection was born. Amongst others, Karush proposed the theory known as conformational adaptability [9, 10] which was later coined the “induced fit theory” in 1958 [11, 12]. As more protein structures were deposited into the PDB, it became clear that not
all proteins are structured throughout their entire lengths. Indeed, many proteins have portions of their sequences missing from the electron density [13, 14] due to imprecise scattering of x-rays from the polypeptide side chains. An increasing number of disordered protein examples have been studied and catalogued [15, 16], and the notion that a protein can have regions of disorder and carry out specific functions is now widely accepted.

1.2 (Un)structure: Levels of structure in proteins

Proteins can range in level of structure, from a fully structured protein with minimal dynamics to mostly disordered with little structure. In between these extremes lie structures such as folded proteins with disordered tails, proteins with folded domains which are linked by disordered loops and molten globule-like proteins.

Figure 1.1 Proteins to represent classes in the progression of structure. A) Cytochrome c (PDB; 3CYT) a stable, structured protein with minimal dynamics. B) Lymphotactin (PDB; 1J8I) a mostly stable structured protein with a disordered C-terminal domain. C) Antibody immunoglobulin G (PDB; 1IGT) with multiple stable folded domains connected by disordered linker regions. D) α-lactoglobulin (PDB; 1HMK) a molten globule-like protein, with some secondary structure but a dynamic tertiary structure. E) α-synuclein, a highly dynamic protein with very little secondary structural elements.

Figure 1.1 shows representative proteins which exhibit varying levels of structure. As a folded protein with stable tertiary structure, cytochrome c follows the structure-function paradigm well. Its 3D structure strongly complements its cellular activity primarily in electron transfer in the respiratory chain [17]. Highly structured proteins are abundant in several protein categories; catalysis, transport, biosynthesis, metabolism and trans-
1. **Introduction**

membrane proteins [18, 19]. Proteins involved in catalytic, biosynthetic and metabolic functions are often enzymes, and are therefore structured for catalysis. Proteins involved in transport are often also classified as membrane proteins. Whilst many structured proteins fit the lock-and-key hypothesis well, it is now widely accepted that even structured proteins will have a degree of flexibility, and form closely-related conformational states. Many proteins possess local regions of disorder, and the length of these regions can vary extensively. Some structures may have a stable hydrophobic core, with C- or N-terminal disordered tails. Such a protein is the metamorphic protein lymphotactin [20] (Figure 1.1 b), a mostly folded protein with an extended disordered C-terminus. The solved structure of proteins with regions of disorder often has no backbone information regarding the disordered region, and in many cases this will be retrospectively mapped on to the PDB structure. Some proteins possess multiple structured domains linked together by disordered loops or bridges. This is often referred to as “beads on a string” [21]. Antibodies represent this type of disordered protein well, as their function is reliant on the spatial flexibility of folded domains within the higher order structure, enabling tight binding to various binding partners [22, 23]. Originally defined as a universal intermediate state in the folding pathway of a structured protein, molten globule proteins are compact structures with a highly disordered tertiary structure. They often have a well-defined secondary structure and some fixed positioning of these secondary structural elements [24-26]. Molten globule proteins can be characterised by their significant increase in protease accessibility [27-29] and their increase in affinity for hydrophobic fluorescence probes [30]. Fully disordered proteins such as α-synuclein (Figure 1.1 Proteins to represent classes in the progression of structure. e) possess little or no secondary or tertiary structure. These proteins may form transient secondary structure upon interaction with binding partners [31].

These proteins of varied structural content will also display an altered protein folding landscape. Many globular proteins have rigid structures in solution and are only functional in this state. This unique conformation is entropically unfavourable as there are numerous restrictions upon its conformational freedom. In contrast, an unfolded polypeptide chain is entropically favourable, forming a large ensemble of conformational states. As such, the ability for a given polypeptide chain to fold is determined by its ability to form intramolecular interactions in order to compensate the free energy increase due to the decrease in the entropy. Conformational energy landscapes can be used to describe the dependence of free energy on all the coordinates determining protein conformation [32].
When approaching their native states, the number of conformations available for globular proteins to adopt is reduced, resulting in a so-called ‘energy funnel’ (Figure 1.2a). IDPs, in contrast, have a relatively flat ‘hilly’ free energy landscape, which describes the large number of conformational states available to such proteins. Each energy plateau is separated by the high energetic barriers, which correspond to transitional states of the protein (Figure 1.2b).

Figure 1.2 The energy landscape model illustrating native globular proteins and intrinsically disordered proteins. A) Globular proteins: local energy minima indicate formation of partially folded intermediates (1 and 2). B) intrinsically disordered proteins: many IDPs can gain ordered structure upon binding to various interaction partners, occurring if the free energy of the complex is lower than the free energies of the IDP and its partner (1, 2 and 3). Figure adapted from [32] and based on the energy funnel model [33, 34].

1.2.1 Intrinsically Disordered Proteins: Definition and Properties

The definition of intrinsic disorder is accepted as the notion that a protein (or protein region) exists as a collapsed or extended, dynamically mobile conformational ensemble, either at secondary or tertiary level, whilst remaining biologically active [35]. This is often further defined as possessing no stable secondary or tertiary structure on the timescale of an NMR experiment [6, 21, 36-38]. Disordered proteins are discussed in the literature under a myriad of different names; pliable [39], floppy [40], flexible [41], natively disordered [42], natively unfolded [43, 44], intrinsically unstructured [6, 37], intrinsically disordered [45], vulnerable [46], malleable [47], dancing proteins [48] and protein clouds [49]. Since the hypothesis was put forward at the turn of the century that disordered proteins represent a broad class of proteins, instead of singular exceptions to the lock-and-key rule [5-7, 37], the term “intrinsically disordered proteins” (IDPs) has become most widely used.

Evidence for disorder in vivo can be shown using in-cell NMR, where results show the retention of the disordered monomer of well-studied IDP α-synuclein [50-52]. Additionally, experiments designed to simulate the effect of macromolecular crowding in cells show that IDPs do not fold (fully) under these conditions [53, 54]. Further studies have provided indirect evidence of disorder in vivo; the chaperone function associated with disorder in late embryogenesis abundant (LEA) proteins can be observed in vivo and in vitro [55].
1. Introduction

Together, this in cell evidence suggests that disorder in maintained in vivo and can be the functional state of these proteins.

The biophysical properties of IDPs have been the subject of discussion since their discovery. Just as the tertiary structure of folded proteins is heavily dependent on its amino acid sequence, the “non-structure” of IDPs is also encoded in the chemical properties of the amino acid side chains. IDPs were first distinguished by their high number of uncompensated charged groups, often negative, resulting in high net charge at neutral pH and their low mean hydrophobicity, which produces a low driving force for protein compaction. This led to the charge-hydropathy approach for defining IDPs, establishing these properties as an important prerequisite for the absence of stable structure in proteins under physiological conditions[7]. IDPs are enriched for amino acids which satisfy these properties, such as polar, disorder promoting Alanine, Arginine, Glycine, Serine, Glutamic acid and Lysine, and the hydrophobic but structure breaking Proline. There is also depletion of “structure-promoting” residues which include bulky, hydrophobic residues Isoleucine, Leucine and Valine, alongside aromatic residues Tryptophan, Tyrosine and Phenylalanine. IDPs are also depleted in Cysteine residues; presumably as these have a crucial role in protein stabilisation by the formation of disulphide bonds [5, 56-59]. Other properties of IDPs provide a good distinction from ordered proteins, these include hydropathy, β-sheet propensity, coordination number, bulkiness and compositions of groups of amino acids such as W+Y+F (aromaticity)[5]. In 2008, over 500 amino acid scales were analysed to produce a novel amino acid scale which discriminated between order and disorder, ranking amino acids from order-promoting to disorder-promoting:


1.2.2 IDP function and regulation

IDPs are highly abundant on the proteome level, especially within Eukaryotes[61], suggesting that although these proteins are unable to adopt a 3D structure under physiological conditions, it is likely they carry out important biological processes [5-7, 21, 37, 43, 62-65]. It is estimated that approximately 25% of all proteins in the PDB are mostly disordered [66], and that over 50 % contain regions of disorder[66, 67]. Whilst prevalent in a variety of biological processes, bioinformatics analyses and numerous studies of
individual proteins have shown that the levels of disorder are very high in proteins which are involved in cell signalling, transcription and translation regulation, with over 70% of all cell signalling related proteins estimated to contain long regions of disorder [18, 61].

As IDPs don’t fit the classical ‘lock-and-key’ hypothesis, they rely on short sequences within disordered regions to act as an “active site” in protein-protein interactions. Often this is in the form of short (10 – 70 residues) linear molecular recognition motif (MoRFs) [66, 68, 69]. These motifs may be fully disordered in the absence of a partner, folding into a specific conformation upon binding. Some protein recognition motifs have also been shown to possess some transient secondary structure in the absence of a binding partner, and this type of MoRFs has recently coined pre-structural motifs (PresMos) [70, 71].

The binding of a MoRF to an interaction partner often involves a disorder-to-order transition of the IDP [7], incurring a penalty of entropy loss due to the free energy required for the transition. As a result, IDPs often bind their interaction partners with low affinity but high specificity. This is especially advantageous in signalling networks, where the requirement for the signal to be “turned off” at the correct time is paramount [21, 37]. Other kinetic advantages include an increased speed of interaction, as flexibility allows the IDP to search through the interaction space and increased surface area available for interaction [31, 37, 45, 72, 73].

IDPs function by interacting with both other disordered proteins and structured proteins. In protein–protein interaction networks, disorder is utilised via two dominant mechanisms. Either a single IDP MoRF will recognize numerous binding partners of differing 3D structure, known as “one-to-many” signalling [57], or many IDPs bind to one, often ordered, interaction partner, known as the “many-to-one” mechanism. In this way, IDPs are able to act as hub proteins, where their flexibility allows many signalling processes to occur without the need for an individual protein for each pathway. Hub proteins can bind sequentially to multiple partners, or act as scaffold proteins whereby they interact with many of their partners at the same time. This can be in order to satisfy temporal or spatial requirements, and can influence the kinetics and specificity of the interactions occurring [31]. Whilst these one-to-many or many-to-one binding mechanisms may result in a complex of stable structure, many bound-states of IDPs retain the intrinsic flexibility of the individual IDP monomers and are deemed “fuzzy complexes” [74]. The retention of dynamics in “fuzzy” complexes can range from local to global disorder and involve both compact and extended states. Complex “fuzziness” also has been shown to play roles in
Introduction

biological processes such as signalling [75], transcription and translation regulation [47, 76-78] and cell cycle regulation [79, 80].

As IDPs perform important functions and possess unusual structural features, they require tight regulation within the cell. Linear motif features are present in a wide range of related protein families with similar binding sequence preferences [81-83], and so cross-binding between interaction partners is likely if their cellular concentration is increased [64, 84, 85]. The availability of IDPs within the cell is controlled in a number of ways, including numerous mechanisms during transcription and translation [86]. For example transcripts encoding proteins with disordered regions have higher messenger RNA (mRNA) decay rates and micro RNA target sites, indicative of lowered gene expression levels [46, 87], and a larger number of predicted ubiquitination sites compared with stable structured proteins [87, 88]. IDPs are also susceptible to degradation by proteases and the proteasome [89-91], with disordered regions proven important for protein turnover via both ubiquitin dependent and independent proteasomal degradation pathways [90-94]. Therefore the majority of IDPs have a short half-life and are present at low basal levels compared with structured proteins. Certain IDPs however, are required at high levels in the cell, some for long periods of time [86]. In this case, so-called “nanny proteins” that bind to the disordered regions that facilitate degradation are present [91, 95], or the protein forms a stable complex or interacts with a binding partner. Due to the numerous interaction partners of many IDPs, often binding at different times [57, 96], it is likely that the stability of the IDP is modulated by their affinity for different binding partners [97].

1.3 Disorder in disease, the D² concept

Due to the prevalence of IDPs in important biological functions, their roles as hub proteins in protein interaction networks and high abundance in signalling, it comes as no surprise that IDPs are implicated in a wide variety of human diseases. Bioinformatics studies have addressed the abundance of IDPs in disease, implicating the class of proteins in cancer, cardiovascular disease, neurodegenerative disease and amyloidoses [64, 98-101]. It has been concluded that the high reliance of many diseases upon signalling pathways has linked IDPs so strongly with pathological roles, leading the D², or the “disorder in disorder”, concept [84]. Contrary to this, there is no strong correlation found between disease-associated proteins and the absence of disorder [19].
IDP related diseases can arise from a number of pathways. Fibrillogenesis occurs whereby a protein fails to adopt its functional conformational state, usually through protein misfolding, often resulting in pathological fibril formation. Misfolding can arise from point mutation, an intrinsic propensity to misfold, impaired post-translational modifications (PTMs), oxidative damage, lost binding partners or exposure to internal or external toxins, amongst others. The largest group of misfolding pathologies, neurodegenerative disorders, arise from proteins adapting from their soluble functional form into highly ordered, stable filamentous structures enriched with β-sheet. Although amyloid fibrils have core structural similarities, their polypeptide constituents are structurally diverse, containing α-helices, β-strands, or intrinsically disordered regions [102]. In the case of disordered proteins which form fibrils, a partially folded conformer is thought to be stabilised, enabling specific intermolecular interactions such as hydrogen bonding, electrostatic interactions and hydrophobic contacts to be formed, a necessary step for oligomerisation and eventual fibrillation. An excellent example of an IDP involved in fibrillogenesis is α-synuclein, a very disordered protein under physiological conditions [44, 103, 104] with a high propensity to form oligomers and fibrils under a wide range of conditions [105]. α-synuclein is implicated in neurodegenerative diseases such as Parkinson’s disease, dementia with Lewy bodies, Alzheimer’s disease, multiple system atrophy and Down’s syndrome.

Due to the large correlation between IDPs, signalling networks and disease, many IDPs have been linked with cancer. Over 70% of cancer associated proteins, or oncoproteins, have been shown to have regions of disorder of 30 residues or longer[64]. Alterations to proteins via mutation, deregulation or over expression contribute to cancer, with recent studies showing that when overexpressed, IDPs, especially oncoproteins, are prone to initiating promiscuous interactions. This suggests that over-expression driven molecular interactions may be frequent causes of cancer [85, 106]. IDPs heavily implicated in cancer include p53, Mdm2, c-MYC, Bcl-XL, BRCA-2, α-Fetoprotein and p21. c-MYC is a transcription factor with roles in regulating cell development, differentiation and growth [107], and is thought to regulate 10 - 15 % of all genes[108]. c-MYC expression is tightly regulated, and deregulation via a number of mechanisms can lead to genome instability, uncontrolled cell proliferation and alterations to the apoptotic pathway[109]. As such, c-MYC is heavily implicated in a large number of human cancers.
1. Introduction

1.4 p53: “The Death Star”

p53, dubbed the ‘guardian of the genome’[110] is a multi-domain, multi-functional transcription factor encoded by gene TP53. p53 is placed at a critical nodal point of converging pathways from diverse cellular stresses which elicit coordinated responses. Simply, p53, in response to a wide range of cellular inputs binds DNA and initiates transcription of many target genes to provide an appropriate cellular response to the detected stress. In reality, the roles of p53 are vast and still being unravelled. Since its discovery by Levine et al. [111] research into p53, its cellular roles, interaction partners, regulation, and mutation etc. has made it one of the most extensively studied proteins in the genome. To date, almost 74,000 papers in Pubmed [http://www.ncbi.nlm.nih.gov/pubmed] contain ‘p53’ either in title or abstract. Despite this, p53 remains somewhat an enigma, its functional complexity still being poorly understood. Much of the work in this thesis is carried out on domains of p53. The following sections will give an overview of the p53 network, however excellent reviews can be found in references [112-116].

1.4.1 Structure and Function

As with many transcription factors, p53 is a modular protein of 393 amino acids with several distinct regions, or domains, with different but inter-dependent functions [117, 118]. p53 is predicted to be ca. 50 % intrinsically disordered [119]. p53 consists of the N-terminal transactivation domain (TAD), comprised of two transactivation domains; TAD1 and TAD2 (residues 1-62); a proline rich domain (residues 64-92); a large central DNA-binding domain (residues 93-293); a tetramerisation domain (residues 325-356) connected to the DBD via a flexible linker; and a basic C-terminal regulatory domain (367-393).

Through extensive study, p53 has emerged as a hub protein with protein-protein interactions (PPIs) between disordered regions of p53 and ordered partners, between ordered regions of p53 and disordered partners, or between two disordered regions of p53 (Figure 1.3). So far, 993 unique interactors of p53 have been discovered (BioGRID database), with the N-terminal TAD and C-terminal domain the most multi-functional binding sites. With such a vast array of binding partners, p53 is the epitome of the one-to-
many binding mechanism described in Section 1.2.2, with some example interactions shown in Figure 1.3.

Figure 1.3 Examples of 14 binding partners of p53 showing one-to-many binding mechanism. Disorder predictor plot in centre shows the predicted level of disorder along the polypeptide chain from PONDR (black line) and IUPred (red line). p53 bound structures are shown with p53 colour-coded dependent on sequence binding. Starting at the top left (magenta protein, royal blue DNA) is the p53-DNA complex (PDB 1tsr), p53-53BP1 (PDB, 1gzh), p53-gcn5 (PDB, 1q2d), p53 tetramerisation domain (PDB, 3sak), p53-set9 (PDB, 1xqh), p53-cyclinA (PDB, 1h26), p53-sirtiun (PDB, 1ma3), p53-CBP bromodomain (PDB, 1jsp), p53-s100ββ (PDB, 1dt7), p53-sv40 Large T antigen (PDB, 2h1l), p53-53BP2 (PDB, 1ycs), p53-PH (PDB, 2gs0), p53-MDM2 (PDB, 1ycr) and p53-rpa70 (PDB, 2b3g). Image adapted from [35].

The domains of p53 are shown in Figure 1.4. Free cellular p53TAD exists in equilibrium between disordered and partially helical [70, 119-121], with several PreSMos; one helical (residues 18-26) and two short turns (residues 40-44 and 48-53) [70]. p53 TAD allows p53 protein, in the context of specific DNA binding, to recruit interaction partners that form the basal transcription machinery required for transcribing new mRNA. p53TAD has been
shown to interact with MDM2, TFIID, TFIIH, RPA, CBP/p300 and CSN4/Jab1 amongst many other binding partners. As a disordered domain, the TAD undergoes folding upon binding when interacting with many of its partners [122]. When undergoing this disorder-to-order transition, p53 TAD often forms an α-helix with specific amino acids inserted deep into the binding interface of the partner protein. It is thought that these local PreSMos may represent specificity determinants important for transcriptional activity [123]. Due to its high level of disorder, the structure of N-terminal p53 has only been solved when in complex with binding partners [124] where it forms a stable helix.

The central DNA-binding domain (DBD, residues 93-293) comprises the largest and most highly conserved domain of p53. p53 DBD is responsible for its sequence specific binding to double stranded DNA in order to mediate its downstream actions, with a consensus binding site consisting of two decameric elements (half-sites) of general sequence 5′-PuPuPuC(A/T)(T/A)PyPyPy-3′ (Pu = A/G, Py = T/C) that can be separated by 0-13 base pairs [125]. In solution and in the absence of DNA, p53 DBD exists mainly as free monomer, with some weak monomer-monomer interactions. When incubated with DNA, the DBD possesses remarkable organising ability [126], forming a stable tetrameric bound structure of two dimers bound to a single stretch of double-stranded DNA [127, 128]. p53 DBD is inherently unstable, with a low melting temperature of ~42-44 °C [129] and a high propensity for mutagenesis. Despite this instability, X-ray crystallography and NMR structures of p53 DBD have been solved, both free in solution [130] and in complex with DNA [131, 132], showing a sandwich of two antiparallel β-sheets with four and five β-strands and multiple disordered loop regions stabilised in part by coordination with a zinc atom. When bound to DNA, the structure of the DBD does not alter, and a loop-sheet-helix motif of a large loop is used to bind DNA. Recognition of the DNA consensus site requires both major and minor grooves, with the helix and loop of p53 DBD making contact with the major groove, and an arginine of a large loop fitting into the minor groove [131].
1. Introduction

Figure 1.4 Schematic showing major domains of p53 and corresponding PDB structures. Transactivation domain encompassing TAD1 and TAD 2 shown in red, central DNA binding domain (DBD) shown in blue, C-terminal oligomerisation and tetramerisation domains shown in green and brown, respectively. PDB structures of domains shown underneath. Left; computational structure of p53TAD. Middle; PDB structure showing p53 DNA binding domain in complex with double-stranded DNA (PDB; 1TUP). Right; tetrameric C-terminal structure (PDB; 2J0Z [133]).

The C-terminal region of p53 is composed of primarily basic residues and has key roles in p53 regulation, interacting with proteins such as GSK3β, PARP-1, TAF1, TRRAP, hGcn5, TAF, 14-3- and S100ββ to perform its cellular functions. p53 C-terminus is also partially disordered and folds-upon-binding in a partner dependent manner, similar to the TAD [134-137]. Interestingly, the same residues are often involved in binding interactions, suggesting that the amino acid “code” is read differently by each binding partner [138]. The function of the C-terminal domain is poorly understood, and some controversy in the literature about its in vivo roles is still ongoing.

Together, these domains regulate p53 function as a stress-activated, sequence-specific DNA-binding protein and transcription factor. The biologically active quaternary structure of full-length free solution p53 is that of a loosely bound tetramer, composed of a dimer of dimers [139-141], which is able to access multiple conformational states [142] and form weak interactions between the DBD and extended N- and C-terminal domains [121, 134, 143]. In contrast, when in complex with DNA tetrameric p53 folds around the DNA helix forming a more stable, rigid structure [144]. The TAD remains extended in both free and DNA-bound p53 structures, allowing multiple regulatory interactions [144, 145].
p53 is a hub protein at the centre of numerous signalling pathways; as a transcription factor, p53 functions by binding to specific DNA promoter sequences to initiate transcription of downstream genes in response to signals emanating from a wide range of cellular stresses. As such, p53 is fundamental for genomic stability. Historically, much of the literature has focused on p53 activation by DNA damage. Double stranded DNA breaks lead to the transcription of target genes which elicit several responses; cell cycle arrest in G1 or G2/M [146], mediated by proteins p21 or GADD45, and/or initiation of apoptosis, mediated by proteins such as Bax or PUMA [114, 147], or cellular senescence [148], mediated by proteins such as PAI-1 or the p21-Rb-E2F pathway. More recently, it was shown that different stresses evoke altered dynamic patterns of p53 cellular levels [149]; for example DNA breaks caused by γ-irradiation results in pulses in p53 cellular levels with a fixed amplitude and frequency [150]. On the other hand, ultraviolet (UV) radiation resulted in a single p53 pulse, the amplitude and duration of which was dose dependent [150]. The downstream effects of different p53 dynamics are still poorly understood. Whilst γ-irradiation and UV irradiation result in different cellular outcomes, whether this is in response to the dynamic profiles of p53 or due to p53-independent events is unknown.

Whilst major p53 roles include control of cell-cycle checkpoints and apoptotic cell death to prevent the emergence of transformed cells, it is now known that p53 is implicated in a vast range of transactivation-dependent and independent cellular processes. Nuclear transactivation-dependent functions include DNA repair, metabolism and antioxidant response [148]. Additionally, p53 has transactivation-independent nuclear roles such as transrepression, DNA replication and homologous recombination [151]. Further, cytoplasmic effects include centrosome duplication, apoptosis initiation via induction of mitochondrial outer membrane permeabilisation (MOMP) [152], and inhibition of autophagy [153].

1.4.2 Post-translational modifications

The primary sequence of p53, especially throughout the N-terminal transactivation domain, contains a complex and diverse array of post-translational modifications (PTMs) with crucial roles in regulating p53 interactions [154]. PTM involves the covalent addition of a functional group after protein translation. For p53, these PTMs include phosphorylation of serine or threonine residues, acetylation, sumoylation, neddylation and ubiquitination and
1. Introduction

methylation of lysine residues (Figure 1.5). Interestingly, a high proportion of PTMs are located in regions of p53 which lack stable structure, with 90 % and 86 % of phosphorylation and acetylation modifications, respectively, found in intrinsically disordered regions (IDRs) [155]. The cellular response to PTMs is dependent upon the patterning and number of modifications, which are themselves dependent on the cell-type in which the protein resides and the nature of the stress stimuli. Clearly, the regulation of p53 by PTMs is a very complicated picture.

![Figure 1.5 Schematic showing p53 domain structure and post-translational modification sites. From left to right; N-terminal transactivation domain, proline rich domain, central DNA-binding domain, C-terminal oligomerisation and tetramerisation domains. The major sites for p53 phosphorylation, ubiquitination, acetylation and methylation are shown. Image adapted from [156].](image)

Phosphorylation is the most commonly reported PTM occurring in mammalian cells. A reversible addition of phosphates, phosphorylation is fundamental in the regulation of biological activity for hundreds of proteins. p53 phosphorylation sites are numerous and span the entire protein, however are particularly concentrated in the TAD and C-terminal domain (Figure 1.5). Most of these sites are phosphorylated by kinases in response to cellular stresses and serve to stabilise p53, although there are some which are
Introduction

phosphorylated in unstressed cells, and become dephosphorylated in response to a stressor [157, 158]. So far, 18 p53 phosphorylation/dephosphorylation sites have been found in human cells. Of those, all but Thr55, Ser376 and Ser378 are phosphorylated as a result of DNA damage. Many phosphorylation sites at the N-terminus can be phosphorylated by multiple kinases, and single kinase can phosphorylate multiple residues [159], which is not surprising considering the functional complexity of p53 with many roles requiring a unique PTM pattern for functionality. Phosphorylation of Ser46 is shown to give p53 distinct DNA-binding characteristics [160] and is critical for induction of pro-apoptotic genes such as p53-mediated Apoptosis-Inducing Protein 1 (p53AIP1) [161], but not required for cell-cycle arrest [162, 163]. The full spectrum of roles of phosphorylation in p53 is beyond the scope of this thesis, other p53 regulation reviews covering further examples can be found in references [156, 157, 159].

Acetylation is another common PTM of p53, mainly occurring in the C-terminal domain, with roles in p53 stabilisation by blocking ubiquitination on the same lysine residue, inhibiting the formation of repressive complexes on target genes and recruiting cofactors for activation of p53 transcriptional activity [156]. Whilst the mutation of several acetylation sites can be compensated for by acetylation at other sites, it is indispensable for p53 function; if all acetylation sites are mutated, p53 is rendered inert [164]. Histone Acetyl Transferases (HATs) are responsible for the addition of acetyl groups, including p53 HATs such as p300, CREB-binding protein (CBP), P300/CBP-associated factor (PCAF) and the MYST family (consisting of MOZ, Ybf2/Sas3, Sas2 and Tip60) [165]. Most acetylation sites are modified by CBP/p300. Two acetylation sites have been found in the DBD of p53, K120 and K164 (Figure 1.5). K120 acetylation is required for activation of target genes resulting in apoptosis of the cell, but not cell cycle arrest [166, 167]. In contrast, K164 acetylation is important for the activation of most target genes.

Ubiquitination is the covalent conjugation of either one or a number of ubiquitin molecules to a lysine residue and requires three enzymes; an E1 ubiquitin-activator, and E2 ubiquitin-conjugator and an E3 ubiquitin-ligase. Mono-ubiquitination of p53 activates cytoplasmic translocation, whereas poly-ubiquitination of a protein targets it for degradation by the 20s proteasome [168, 169]. Both mechanisms maintain low basal levels of p53 in unstressed cells.

Methylation of p53 occurs only in the C-terminal regulatory domain; lysine residues are monomethylated by three lysine methyltransferases (KMTs) and dimethylated by two KMTs.
Introduction

Arginine residues are methylated by only one known methyltransferase; protein arginine N-Methyl Transferase 5 (PRMT5) [171]. Methylation of p53 residues can either have an activating or repressive effect on p53 function, dependent on the location and number of methyl groups added.

1.4.3 Regulation of p53: The p53 – Mdm2 interaction

Whilst p53 function is controlled by PTMs, its cellular levels must also be tightly controlled. This is mainly carried out by Murine Double Minute 2 (MDM2; HDM2 in humans, henceforth denoted MDM2), an E3 ubiquitin ligase and negative cellular regulator of p53 which functions via two main mechanisms. Firstly, the direct interaction between MDM2 and p53 via their respective N-terminal domains inhibits the transcriptional functions of p53. Secondly, MDM2 exports p53 protein into the cytoplasm and/or targets p53 for proteasomal degradation by addition of several ubiquitin molecules. Initial studies presented in Chapter 6 indicate that p53 activity may also be controlled by MAGE-A4, by showing MAGE-A4-dependent displacement of p53 protein from target promoter genes p21, Bax and PUMA. The mechanism of this regulation is yet to be understood, and details regarding the binding between p53 and MAGE-A4 remain to be determined. Additional mechanisms to regulate p53 signalling are provided by numerous proteins, the full discussion of which is beyond the scope of this thesis.

MDM2 is a multi-domain protein of 491 amino acids which primarily functions in cells as an ubiquitin E3 ligase but with additional roles as a molecular chaperone [172] and in translational control. MDM2 is expressed in the nucleus but translocates to the cytoplasm where it interacts with multiple binding partners to mediate degradation by the proteasome [173, 174]. The N-terminal domain is partially flexible [175], and contains a hydrophobic binding pocket which drives MDM2 protein: protein interactions. The first 23 amino acids of MDM2 N-terminus form a highly flexible disordered ‘lid’ region, which interacts with the hydrophobic pocket as a pseudo-substrate and regulates its conformation [176]. The central region contains an acidic domain and a nuclear localisation signal. The acidic domain is structured and confers MDM2 binding to a number of protein including signalling enzymes that contain a general consensus sequence SxxLxGxxxF [177, 178]. The highly-conserved zinc-finger domain is also centrally located, containing a $3_{10}$ helix and four $\beta$-strands and incorporating a single tetrahedrally coordinated zinc ion [179].
1. Introduction

The C-terminal domain contains the Really Interesting New Gene (RING) finger domain, which forms a C$_2$H$_2$C$_4$ dimeric protein: protein interaction domain [180] and contains an ATP motif that regulates MDM2 molecular chaperone function [172].

MAGE-A4 is a ~35 kDa protein from the Melanoma Antigens Genes (MAGE) family, which is divided into two subfamilies dependent on their tissue pattern expression. MAGE-A4 belongs to the MAGE-I subfamily, and is a tumour specific protein belonging to the group of Cancer Testis Antigens. All MAGE proteins possess a conserved Mage homology domain (MHD) consisting of approximately 200 amino acids. The N-terminus is hypothesised to be highly disordered. MAGE-A4 has been shown to bind with MIZ-1 in order to repress the p21 promoter [181, 182], and the oncoprotein Gankyrin [183].

![Diagram of p53:MDM2 autoregulatory feedback loop](image-url)

**Figure 1.6** The p53:MDM2 autoregulatory feedback loop. p53 binds to MDM2 promoter sequence and induces MDM2 transcription. Increased MDM2 levels results in targeting of p53 for degradation in the proteasome, lowering cellular p53 levels. Low p53 levels prevents p53-induced transcription of MDM2, thus lowering MDM2 levels. MAGE-A4 is postulated to disrupt binding of p53 to target promoter sequences Bax, p21 and PUMA.

The direct targeting of p53 by MDM2 occurs in an elegant autoregulatory feedback loop. In unstressed cells, p53 concentration above basal level causes increases MDM2 expression due to p53 binding to the MDM2 promoter. The increased cellular MDM2 levels result in targeting of p53 for degradation by the proteasome, exporting p53 from the nucleus, preventing p53 interaction with transcriptional coactivators etc. Together, this reduces the
levels of p53, and as a result reduces the transcription of MDM2 promoter sequence (Figure 1.6). p53 primarily binds to MDM2 through a conserved peptide motif in the N-terminal domain, BOX-1 (residues 14-27). BOX-1 possesses some pre-structured tendencies [70], folding-upon-binding to form a short α-helix upon insertion into the N-terminal hydrophobic binding cleft of MDM2 [124]. p53 residues Phe19, Trp23 and Leu26 make direct contact with MDM2 [124]. The structure of a p53 peptide inserted into the MDM2-binding cleft (Figure 1.7; PDB 1YCR; rscb.org) is one of the only solved structures of p53 N-terminal domain produced to date, all which show p53 in complex with a binding partner (2GS0, 2B3G, 1Q2F, 1T4F, 2GV2 and 3DAC [184-189].

Figure 1.7 The p53:MDM2 interaction left) side view right) top view. p53 inserts into a hydrophobic binding cavity on the surface of MDM2. Residues Phe19, Trp23 and Leu26 are shown making contact and labelled. Image created from PDB structure 1YCR [124].

The N-terminal disordered ‘lid’ of MDM2 acts as a pseudo-substrate, interacting with the hydrophobic pocket in the absence of p53, and is displaced upon p53 binding in response to the phosphorylation state of residue Ser17 [190-192]. This N-terminal interaction is heavily studied and numerous small molecule agonists have been developed, however further p53:MDM2 interactions are required for ubiquitination. p53 BOX-1 binding to N-terminal
MDM2 allosterically alters the conformation of MDM2 [178] such that the RING domain is exposed to the p53 DBD and makes weak interaction with a second interaction site (p53 BOX-V). This interaction is required for the MDM2 RING domain to carry out the E2 interaction and for ubiquitination of p53 [193], and has additionally been shown to affect the binding affinity of the MDM2 N-terminal hydrophobic binding pocket [194].

1.4.4 p53 and cancer

As a hub protein with hundreds of individual, vital cellular roles, the most impressive contribution of p53 is its function as a powerful tumour suppressor. Its most extensively studied roles in cell-cycle arrest, apoptosis and senescence as a result of cellular stressors, discussed in Section 1.4.1, means a plethora of information is available regarding the mechanisms by which p53 selective destroys abnormal cells and prevents tumourigenesis. Conversely, impairment of p53 function plays an invaluable role in tumour development, allowing continued proliferation via evasion of p53-dependent pathways. Unsurprisingly therefore, p53 has been implicated in over 50% of human cancers [195], where it is inactivated by either mutation or through abrogation of partner proteins or signalling pathways that regulate p53 activity.

p53-defective tumours where TP53 is mutated are strongly linked to lung, colon and stomach cancers, where it is the most frequently mutated gene [196]. In these tumours, over 95% of mutations lie in the central DBD, and 75% of these mutations are missense mutations [197], resulting in the p53 protein being full-length but with a single amino acid change. Thus, p53 mutations are diverse in their sequence, position and structural impact making identification of patterns in tumour type and aetiology difficult. Mutations in the DBD fall into categories of either structural or DNA contact; p53 DNA contact mutants possess little change to the tertiary structure or effect on protein folding, whereas structural mutations can either cause local structure disruptions or destabilisation of the entire protein [198]. In each case, mutation of p53 often confers dominant-negative function, and the mutant protein can form heterodimers with wild-type p53 [199-201]. Additionally, it has been shown that mutant p53 may experience ‘gain-of-function’ [202] abilities, exhibiting oncogenic properties or dominant-negative activities [195], such as binding to p53 homologues p73 or p63, reducing their transcriptional activity [203, 204].
In over half of p53-defective tumours, p53 is not mutated but p53 signalling pathways are abrogated as a result of changes to p53 effector molecules. The most common of these p53-independent modifications is the overexpression of negative regulator MDM2. As discussed in Section 1.4.3, MDM2 maintains low basal levels and transcriptional activity of p53 in unstressed cells primarily via direct interaction between their respective N-terminal domains. Mechanisms which alter the negative feedback loop between p53, MDM2 and other regulator proteins result in increased expression of MDM2, a decreased cellular level and therefore inactivation of p53. Therapeutic strategies that reactivate p53 by targeting MDM2 itself or inhibiting the p53:MDM2 interaction are therefore attractive cancer drug therapy targets, and proof-of-concept experiments have proved effective in vitro. Blocking MDM2 expression through several gene silencing techniques has resulted in reactivation of p53 pathways in cancer cells [205-207]. Direct inhibition of the p53:MDM2 interaction is difficult due to the inherent flexibility of both proteins, nonetheless small molecule ligands such as cis-imidazolines [208, 209], “stapled” peptides [210, 211], terphenyls [212], oligobenzamides [213], spiro-oxindoles [214], chromenotriazolopyrimidine [215] and Benzodiazepininedione [187] have been shown to block the binding between p53 and MDM2. The cis-imidazoline Nutlin-3 is composed of enantiomers a and b, of which enantiomer a is 150 times more potent, and binds MDM2 in the p53 peptide groove, mimicking the three p53 residues responsible for the bulk of binding interactions [208]. Nutlin-3 is effective in numerous cell lines, and is able to arrest or induce apoptosis in proliferating cancer cells with micromolar concentrations [216]. The drug candidate RITA (reactivation of p53 and induction of tumour cell apoptosis, NSC 652287) was shown to restore wild-type p53 function in tumour cells by preventing the p53:MDM2 interaction [217], although recent studies shows its effect to be via DNA damage and not interaction with p53 [218].

1.5 Identification and prediction of IDPs or disordered regions

Disorder in proteins can be determined both experimentally and via disorder predictor algorithms. There are numerous experimental methods which are sensitive to disordered regions and can give useful structural information. Characterisation of the diverse properties and conformational states of IDPs cannot be based on a single technique. Often
Introduction

multiple experimental and computational techniques are combined to give detailed insights into IDP structure and function, the most commonly used techniques are summarised here.

1.5.1 X-ray crystallography

X-ray crystallography is a powerful biophysical technique which gives atomistic level structural information regarding protein structure through X-ray beam diffraction of a protein in a single crystal, and is often considered the gold standard in structural biology. Due to the inherent dynamic nature of IDPs, X-ray diffraction beams relating to these sequence stretches often become incoherently scattered and are unobserved. This leads to regions of the electron density to be missing, a feature which can correlate with disordered regions. Difficulty arises in the determination of disordered regions, as structured regions of the protein fluctuating around a disordered linker may also not be observed. Long stretches of missing data may not necessarily be disordered [31].

1.5.2 NMR spectroscopy

Heteronuclear multidimensional Nuclear Magnetic Resonance (NMR) spectroscopy is perhaps the most readily suited technique for the study of IDPs or IDR, giving information regarding short and long-range structural parameters. NMR measures the chemical shift of a specific nucleus, giving information regarding the local environment of each nucleus of the protein and as such, local structural propensities such as α-helix and β-strand populations. It can also report directly on long-range interactions using paramagnetic relaxation enhancement (PRE) probes, even when these are weakly populated or transient. NMR parameters such as heteronuclear spin relaxation can report on dynamic timescales of both local and long range reconfigurations, and conformational changes such as disorder-to-order transitions [219]. Overall, NMR gives access to ensemble and time-averaged conformationally dependent parameters at atomistic resolution [220, 221]. IDP signals in NMR experiments possess spectroscopic features of small molecules, and as such resonance assignments are possible even for very large IDPs [222]. The complete assignment of many IDPs [37, 223, 224], along with direct mobility measurements of backbone motions of disordered regions has been carried out by NMR. As secondary
Introduction

Structural elements within disordered regions are often transient, NMR sets itself apart from techniques which measure overall secondary structural content, as it details specific chemical shift deviations away from the expected value of a random coil state. Through these random coil shifts, both the location and population of transient secondary structure can be accessed. In some cases, low populations of these structural conformations can create signal ambiguity, requiring shifts from several different nuclei to be studied.

1.5.3 Small Angle X-ray Scattering (SAXS)

Small angle X-ray scattering (SAXS) is a solution based technique allowing low resolution structural characterisation of macromolecules. It provides information regarding the size, shape and oligomeric state of proteins along with quantitative analysis of flexible proteins [225]. Soluble protein samples are exposed to an X-ray beam and the intensity of beam scatter is detected as a function of scattering angle. Data output from SAXS can be interpreted in a number of ways; the shape of a Kratky plot is sensitive to the conformational state of the studied protein [225, 226]. This representation produces qualitative information regarding protein globularity and can define structural contributions from distinct regions of a protein, however falls down when studying completely unfolded structures due to the requirement for a protein-specific extension maximum [31]. The radius of gyration (R_g) of a protein can also be calculated from SAXS measurements, giving a single value for the size of the molecule in solution. For IDPs, multiple interconverting conformations produce an ensemble of R_g values, which are often displayed as a distribution. The R_g value allows comparison to well defined fully disordered structures or computationally modelled structures of varying conformation. Long range interactions within disordered regions can also be probed using R_g measurements, as the protein radius is expected to decrease compared with a fully extended disordered sequence. Changes in the R_g upon environmental changes, such as pH, temperature, ionic strength and PTMs can also be monitored using SAXS. Other techniques measuring hydrodynamic radius or R_g include size-exclusion chromatography, analytical centrifugation, pulse-field gradient (PFG) NMR and fluorescence correlation spectroscopy (FCS).
1.5.4 Circular Dichroism

Far-UV circular dichroism (CD) is a solution phase technique which measures overall secondary structural content of a protein. CD spectra can be used to estimate the proportion of secondary structural elements giving rise to characteristic signals shown in Figure 1.8, and can measure the percentage α-helix, β-sheet, β-turn and random coil. As CD measures the global structural content of a protein, it provides no location-based information for secondary structure but can be useful in following secondary structural elements under various stepwise condition alterations such as temperature and denaturant curves.

![CD spectra showing characteristic signals for protein secondary structural elements.](image)

Figure 1.8 CD spectra showing characteristic signals for protein secondary structural elements. Spectrum 1, in black, shows a typical signal for an α-helical protein. Spectrum 2, in red, shows a typical spectrum for a protein containing antiparallel β-sheets. Spectrum 3, in green, shows a typical spectrum for a disordered protein. Spectrum 4, in navy, shows an example spectrum for a triple-helix containing protein. Spectrum 5, in cyan, shows a typical spectrum for a denatured protein. X-axis denotes wavelength (nm) and y-axis denotes residue ellipticity, although can also be measured as the change in the photomultiplier tube dynode voltage as a function of wavelength. Image adapted from [227].
1.5.5 Other Biophysical Techniques

Further techniques employed to study IDPs, but beyond the scope of this thesis, include techniques determining fluorescence characteristics including Förster resonance energy transfer (FRET), fluorescence anisotropy and lifetime[31] and single molecule FRET [228], which can be used to track folding and unfolding events of IDPs on a single molecule level [229-232]. For example, the coupled folding and binding of fully disordered α-synuclein has been monitored to reveal an intricate energy landscape [233, 234]. Other powerful experimental techniques include electron microscopy [235, 236], proteolytic degradation [27, 28], immunochemical methods [237, 238], calorimetric melting curve transitions [239], urea- or GdmCl-induced unfolding curve transitions [240, 241] and electrophoretic mobility features of SDS-PAGE [242].

1.5.6 Disorder Predicting Algorithms

The first predictor of disorder was developed in 1997 and named Predictor of Natural Disordered Regions (PONDR) [http://www.pondr.com]. PONDR operates from primary sequence information alone, using 11 sequence attributes of IDPs. Datasets of ordered and disordered protein sequences were used to train neural networks, with accuracies in the range of 70-84 % [243]. More recently, the prediction accuracy of PONDR has been increased by expansion of the database of disordered proteins and improved computational methods [244, 245]. There are now over 50 disorder predictors [246], including FoldIndex [247], GlobPlot [248], DisEMBL[249], DISOPRED [250]and DISOPRED2[61, 250-252], DRIPRED[253], IUPred[254, 255], FoldUnfold [256, 257] etc. The most commonly used are summarised here. GlobPlot uses the relative propensity of a given amino acid to be in an ordered or disordered state, using a scale based on the probability of the amino acid to be either random coil or in a classic secondary structure. Fast and simple, GlobPlot is useful for finding globular domains [246]. DisEMBL uses three artificial neural networks to predict three types of disordered structures; residues involved in loops/coils, hot loops (loops with high B-factors), or those with missing X-ray structures. DISOPRED uses feed-forward neural networks in order to predict disordered regions, and DISOPRED2 takes
Introduction

this further by using support vector machines to increase the speed of prediction. Different from the other predictors, DISOPRED and DISOPRED2 train on the complete sequence of the protein, rather than amino acid composition, biophysical properties such as hydropathy, or sequence complexity. IUPred is based on the idea that a large number of proteins rely on inter-residue interactions, stabilising the protein, and IDPs do not have sufficient numbers of these interactions. IUPred estimates the interaction energy between each pair of amino acids and combines this with amino acid composition. These are then used to calculate the energy boundary between ordered and disordered proteins as a function of length.

Whilst there are benefits to all disorder predictors there are still limitations to the accuracy of these algorithms. DisProt, the databank for disordered proteins now contains over 700 IDPs, a significant increase from release 4.5 (July 2008) which included 520 IDPs. As it is established that a large proportion of Eukaryotic proteins contain regions of disorder, the number of proteins experimentally characterised is a fraction of the total number of IDPs likely present in the genome, and even fewer are deposited into databases. A larger number of annotated IDPs are required for disorder predictors to improve. Similarly, current disorder predictors rely mainly on attributes of IDP primary sequence such as biophysical properties of amino acids and sequence compositional content. As the molecular mechanisms behind disordered structure formation become better understood, the constraints used by disorder predictors will become more extensive and as such prediction rates will increase.

1.5.7 Computational Methods

The addition of computational approaches alongside experimental techniques is vital in progressing our understanding of IDP structure and function. Whilst information, often very detailed, regarding IDP secondary structure, transient long-range interactions and dynamics can be provided by experiments, describing the conformational ensemble of a disordered system remains a challenge. Increasingly, computational methods are well placed to visualise experimentally observed constraints, alongside generating de novo simulations and structure predictions which can explore the energy landscape accessible to proteins.
Simulations and structure prediction methods fall under two main categories; molecular dynamics (MD) and Monte Carlo (MC) simulations. Both represent molecules as collections of atom-centered interaction sites, implement periodic boundary conditions [258] and are based on the use of molecular mechanics force fields [259-262] to control the total energy (for MC) and forces (for MD).

MD is a gradient based technique able to provide detail concerning the motion of individual atoms as a function of time, generating new molecular configurations by applying Newton’s equations of motion to all atoms simultaneously over a small time step. This allows new atomic positions and velocities to be determined. Programmes such as CHARMM [263] and its descendants AMBER [264] and GROMOS [265] are widely used, although the field continues to grow. This method is useful for observing protein folding or unfolding or conformational changes under the influence of environmental factors such as temperature.

MC simulation methodology follows the general rule of randomly sampling a molecular structure, perturbing it via translation, rotation, or performing internal structural variations to form a new configuration. With enough sampling, MC can provide Boltzmann-weighted structural averages and thermodynamic properties [258]. A stochastic process, MC simulations are able to access length scales inaccessible to MD when studying complex protein structures such as aggregation and is often deemed more suitable for simulating IDP structures, as it permits sampling of a wider conformational space available to the IDP.

Computational techniques can also be used to assign ensemble structures to a large pool of input structures using experimental values as conformational restraints. Such programmes include ENSEMBLE [266, 267], which incorporates restraints from NOEs, PRE data and secondary chemical shifts from NMR along with solvent-accessible surface area. Whilst ENSEMBLE does not directly create the structures, it can be an iterative process whereby the conformers of the ensemble are directly modified using experimental restraints and subject to another ensemble population assignment.

### 1.6 Biological Mass Spectrometry

A starting point in the history of mass spectrometry (MS) can be considered to be when Thomson and Wien published their investigations on cathode rays in the late 19th century [268, 269] and Aston went on to determine the isotopes of atoms [270]. For the next ~70
years MS was primarily utilised for mass determination of small compounds, and expansion into the study of larger organic and biological molecules was restricted due to the difficulties in ionisation of sample. Ionisation techniques such as electron ionisation, photon ionisation or chemical ionisation were extremely harsh, and caused extensive decomposition of sample. Ionisation methods such as plasma desorption, fast atom bombardment and laser desorption and field desorption were developed but it wasn’t until the 1980’s, with the advent of “soft” ionisation technique such as electrospray ionisation (ESI) [271], that MS evolved significantly as a biophysical technique.

In short, MS is a gas-phase technique which determines the mass of a molecule. The time taken for a charged molecule to travel through an electric field is measured as a mass to charge ratio (m/z). Gaseous ions of positive or negative charge are produced and guided into the vacuum of the mass spectrometer, where they are influenced by electric and/or magnetic fields. Ions are sorted by their mass to charge ratio (m/z) and the abundance of each m/z value is detected. The output is a mass spectrum, which gives information regarding all ions across a range of m/z values determined by the mass range of the instrument. There are many types of mass spectrometer, however in this thesis only quadrupole time-of-flight (QToF) mass spectrometers are utilised. Details regarding other types of mass spectrometer can be found in references [272, 273].

Intact protein mass spectrometry relies on the transfer of a protein monomer or complex from solution into the gas phase whilst retaining non-covalent interactions. Whilst non-covalent interactions were known to be preserved in the vacuum of the mass spectrometer early after ESI development [274-276], the preservation of specific non-covalent interactions was not noted until later. Protein-protein complexes [277, 278], protein-ligand complexes [279-281] and protein aggregates [282-284] have all been preserved in the gas phase using ESI or its counterpart, nano-electrospray ionisation (nESI).

Mass spectrometry is now a well-established technique in structural biology. Macromolecules and their complexes of a vast molecular weight range can be transported intact into the gas-phase environment [285-287]. There exists now a strong body of evidence demonstrating that the secondary and tertiary structure (or the lack thereof) of a protein in solution can be retained upon transfer into the gas-phase environment [288-290]. Structural changes to proteins, such as folding and unfolding as a function of environmental changes such as pH [291], temperature [292-294], along with instrumental changes such as trapping [295, 296] or dissociation of ions can be readily probed by MS. Charge state
distribution alterations attributed to changes in the solution conditions can be readily observed. For example, the charge state range displayed by globular folded proteins will often widen upon the addition of acid, indicating loss of structure [297]. In contrast, the addition of acid to solutions containing IDPs results in minimal changes to the observed charge state distribution [297], indicating that the flexibility of the protein is retained in the gas phase. The complementary technique of ion mobility mass spectrometry (IM-MS), discussed further in Section 1.7, has allowed extensive computational studies to be performed in order to relate gas phase structures to those observed using solution phase techniques [298, 299], giving good agreement. Additionally, soft landing experiments have been performed whereby the activity of a protein is tested once it has traversed the mass spectrometer. These showed retention of the functional capabilities of proteins after traversing the instrument [300]. From a more practical viewpoint, the time scale of a mass spectrometry experiment (on the order of 15 ms) is short enough that it is unlikely the protein will be unable to undergo significant structural alterations.

### 1.6.1 Studying IDPs with Mass Spectrometry

Observing the conformational space occupied by IDPs, their active conformation and their specific interaction with partners is key to understanding their cellular function. Study of IDPs using biophysical techniques has advanced significantly in the past 20 years, however as a class of proteins information regarding IDP structure and function is still significantly lagging behind that of ordered proteins [84].

ESI as a tool can be used to show the degree of conformational flexibility within a protein. ESI or nESI produces ions with multiple charges directly from solution. These multiply charged ions are as a result of either proton gain to available basic residues (positive ionisation) or the loss of protons to form negatively charged groups (negative ionisation), giving a charge state distribution (CSD). The availability of a residue for proton gain/loss is dependent on its solvent accessibility in solution; in general, a more compact protein has fewer sites available for protonation/deprotonation compared with an extended protein, and as such will produce a smaller range of charge states. In this way, the resulting mass spectrum will give information regarding the solution conformation of the protein [301].
Conformational states of disordered proteins have been probed using MS under soft ionisation techniques, with populations of conformations observed in the range and intensity of charge states produced. Frimpong et al., using ESI-MS, showed the conformational heterogeneity of α-synuclein, a highly disordered protein [302]. Through deconvolution of the CSD for α-synuclein under varied pH, they were able to track four distinct conformational states, ranging from compact to random coil [302]. Interestingly, ESI-MS analysis of the same protein under negative ionisation mode produced only three of the previous observed conformational families [303]. This suggests that alongside solution conditions, ionisation and instrumental parameters can affect the retention of protein conformation when transferred from solution to gas-phase, with further studies confirming this notion [294, 304].

The use of mass spectrometry for the study of IDPs has several significant advantages over other biophysical techniques. The mass spectrometer provides extremely high sensitivity, allowing detection of species at very low abundance. This can be especially useful when studying proteins which are partially disordered, as the disordered state may not be the most abundant, or when studying IDPs which can only be produced in small quantities or at low concentrations, as is often the case for IDPs. The unparalleled selectivity of MS is also highly advantageous; the use of mass analysers allows defined species to be selected with ease for further experiments such as fragmentation. The speed of the experiment and the ability to measure multiple species present in a complex mixture simultaneously also support MS as a technique for the study of IDPs. In particular, mass spectrometry lacks a bias towards the folded state of a protein [305], allowing observation of all conformational states present in solution. Together, these advantages position MS well as a tool for probing the unstructured nature of IDPs.

1.6.2 Ionisation Methods

In the first stage of any mass spectrometry experiment, ionisation transfers charge to the sample allowing control of the ion through the mass spectrometer and a signal to be created at the detector. For the study of macromolecules, ionisation techniques which do not result in the fragmentation of the sample are paramount. The advent of electrospray ionisation [306-308] allowed for the first time molecules of a vast range in mass to be transported, intact, into the mass spectrometer. Development of the technique by Fenn et
1. Introduction

_{al. [271, 309]}_ in producing multiply charged ions of macromolecules and their complexes, opened the door for mass spectrometry to become a powerful biophysical technique for the study of proteins and peptides. Matrix-assisted laser desorption ionisation (MALDI) [310, 311] is also considered a ‘soft’ ionisation technique, producing ions with little or no fragmentation of the sample. Both ESI and MALDI produce ions of charge \([M+zH]^+\), however MALDI commonly produces singly charged species \(z = 1\), whereas ESI produces multiply charged ions. Multiple charging of a sample is beneficial for macromolecules as it allows the detection of large molecules on mass spectrometers which only have a limited \(m/z\) range. MALDI requires the co-crystallisation of sample with organic solvents, resulting in low pH levels incompatible with protein native structure retention, and as such electrospray ionisation techniques were used for all studies reported in this thesis.

During ESI, a capillary at atmospheric pressure is filled with the sample in solution. A high electric potential, \(\sim \pm 2.6 \text{kV}\), is then applied to the sample. In positive mode, a positive potential is applied and in negative ionisation mode a negative potential is applied, producing positively and negatively charged ions respectively.

![ESI Process Schematic](image)

**Figure 1.9 Schematic showing the ESI process in positive mode.** As shown in Figure 1.9, applying an electric current to the sample in the capillary creates an electric field between the capillary and the entrance to the mass spectrometer. When done in positive ESI mode, this creates an abundance of positive charges at the end of the capillary, causing the liquid surface to distort out and form a Taylor cone. This liquid surface then breaks to form solution droplets which contain a number of these positive charges, often aided by a coaxial gas flow (not depicted in Figure 1.9) [312]. The droplets emitted from the Taylor cone are approximately several \(\mu\text{m}\) in diameter and undergo rapid solvent evaporation, a process
Introduction

which is often aided by a raised temperature. This reduction in droplet surface area results in an increase in the charge density on the surface of the droplet, up to the point where the surface tension is balanced by Coulombic repulsion of the like charges. This point is known as the Rayleigh limit \cite{313}. Droplets at the Rayleigh limit produce smaller offspring droplets via jet fission \cite{314}, before ongoing evaporation/fission events produce the final droplets with nm radii. These droplets contain the analyte which will be measured by the mass spectrometer. Investigation into the mechanism of release of these ions is still debated in the literature, however three main mechanisms have been proposed; the ion evaporation model (IEM), charged residue model (CRM) and the chain ejection model (CEM).
1. Introduction

Figure 1.10 Summary of ESI proposed mechanisms. Left) Ion ejection model (IEM); an ion is ejected from a charged nanodroplet. Centre) Charged residue model (CRM). Right) Chain ejection model (CEM). Image adapted from [314]

The IEM is thought to occur with low molecular weight species [315] and small inorganic ions. The mechanism works on the basis that the electric field which occurs from a droplet at the Rayleigh limit is a great enough force to cause ejection of small ions from the surface. The ejecting ion retains contact with the droplet as it leaves by a string of solvent molecules which ruptures as the ion leaves the droplet [316]. This phenomenon is shown to occur more with solvated droplets, as opposed to purely aqueous droplets [316]. The ion is ejected with a small solvent cluster surrounding it, which is then removed by collisions with
gas molecules [317]. Due to this, comparisons can be drawn between IEM and droplet fission occurring during the ESI process.

The charge residue model (CRM, Figure 1.10 centre) is the widely accepted mechanism by which large globular species, such as folded proteins, enter into the gas-phase [318, 319]. Nanodroplets at the Rayleigh limit which contain a single analyte evaporate until no solvent molecules remain, and as the final shell is removed, the charge it carries is transferred onto the analyte [318, 319]. As the nanodroplet is already at the Rayleigh limit, it is suggested that the droplet loses charge as it shrinks, to remain close to the Rayleigh limit throughout the CRM process. Usually this would occur via ejection of small ions or solvated protons [320]. In support of this mechanism is experimental data, which shows that ions of composition near to \([M+Z_R H]^\pm\), where \(Z_R\) is the Rayleigh charge of protein sized water droplets, occur during ESI of globular proteins [318, 321-323].

Finally, the charge ejection model (CEM; Figure 1.10, right) shown to be the mechanism followed by unfolded proteins [324, 325]. Unfolded proteins have nonpolar residues, which would otherwise be confined to a hydrophobic core, exposed to solvent. This gives the protein extended/hydrophobic properties, and makes it unfavourable for the analyte to reside within the nanodroplet interior [314]. MD simulations reveal that the protein migrates to the edge of the droplet, where one terminus of the polypeptide chain is ejected followed by stepwise sequential expulsion of the remaining polypeptide chain [314].

Nano-electrospray ionisation (nESI) occurs via the same mechanisms as ESI described above, but with a smaller tip orifice (~1-2μm radii, compared with ~100μm for ESI) and a lower flow rates of approximately <10 nL min\(^{-1}\) [314]. As such, nESI requires a much lower sample volume, allowing analysis of precious samples. nESI also enhances ionisation efficiency and increases sensitivity [312, 326, 327].

### 1.6.3 Mass Analysers

Once ions are produced they need to be sorted by their mass to charge ratio \((m/z)\). This is carried out by mass analysers, which can be used in combination or individually. The work in this thesis makes use of both quadrupole and time of flight (ToF) mass analysers in hybrid mass spectrometers.
1.6.3.1 Quadrupole Mass Analyser

The quadrupole mass analyser consists of four parallel cylindrical rods which are arranged in two pairs (Figure 1.11). A fixed direct current (DC) is applied to the rods, such that voltages of opposite polarity are applied to adjacent rods. Superimposed on this DC current, an alternating radio frequency (RF) is applied, such that the polarity of the rods constantly changes; when one pair is in phase (i.e. positive), the other is out of phase (i.e. negative) with respect to RF potential and vice versa. This creates an electric field which oscillates between the rods and drives ions through the space between.

![Schematic of a quadrupole mass analyser](image)

**Figure 1.11** Schematic of a quadrupole mass analyser. Rods of opposite polarity are next to each other. Resonant ions with stable trajectory (red) are transmitted through the centre of the rods to the detector. Ions with an unstable trajectory (grey) are discharged onto the rods upon collision and do not pass through to the detector.

When an ion enters the space between the rods, it is drawn to the rod of opposite polarity. If the RF potential causes a rod to change polarity before the ion collides with it, this will result in a trajectory change of the ion. As such, ions traverse the quadrupole in an...
Introduction

oscillating path until surviving ions exit the analyser, the path of which is determined by values of U, V and ω, in Equation 1.1:

\[ \pm \Phi_0 = \pm (U - V \cos \omega t) \]

Where \( \Phi_0 \) is the potential applied to the rods, U is the DC voltage, V is the ‘zero-to-peak’ amplitude of the RF potential, \( \omega \) is the angular frequency of the RF potential, and \( t \) is time.

Quadrupoles can be used in several modes during an MS experiment. If the RF amplitude (V) is scanned, ions of different m/z values are able to be transmitted through the length of the analyser. This is the most common mode in MS experiments. If ions of specific m/z values are required to be selected, specific values for parameters U, V and \( \omega \) can be set to allow stable trajectory of the chosen ion. Ions of differing m/z values will not be able to travel through the analyser and will collide with the rods (Figure 1.11). In this way, the quadrupole analyser can act as a mass filter, allowing mass selection prior to further MS experiments such as fragmentation studies.

1.6.3.2 Time-of-Flight Mass Analyser

Time-of-flight (ToF) mass analysers have evolved significantly since their first implementation in a commercial instrument in 1955 [328]. The initial linear arrangement (Figure 1.12a) meant ions entered the ToF analyser at one end and were detected at the other. Pulses of ions are injected in and accelerated by a potential V, before entering a field-free region where their velocity is dependent on their m/z. The time spent by ions in this field-free region is measured and the mass-to-charge ratio of the ions calculated using Equation 1.2:

\[ E_K = \frac{mv^2}{2} = zeV \]
1. Introduction

Where $E_k$ is the kinetic energy of the ion, $m$ is the mass of the ion, $v$ is the velocity of the ions, $ze$ is the total charge of the ion and $V$ is the accelerating potential.

If the time ($t$) taken be ions to travel the analyser length ($L$) with a velocity $v$ between entering the field-free region and arriving at the detector is:

**Equation 1.3** \[ t = \frac{L}{v} \]

Then by combining Equation 1.2 and Equation 1.3, Equation 1.4 demonstrates how the $m/z$ is related to the time taken for an ion to traverse the field-free region of the mass analyser; the lower the mass of an ion and the higher its charge, the faster it will travel.

**Equation 1.4** \[ t^2 = \frac{m}{z} = \left( \frac{L^2}{2eV} \right) \]
Introduction

Figure 1.12 Schematic of a ToF mass analyser. a) Linear ToF analyser. Ions with lower \( m/z \) values are represented by smaller spheres. Kinetic energy dispersion is noted by unequal travel of ions of equal \( m/z \) to the detector. b) Reflectron ToF analyser. Ions with lower kinetic energy (light blue spheres) and ions with high kinetic energy (dark blue spheres) represented.

Whilst the linear technology does give \( m/z \) values, it results in poor mass resolution as ions of the same \( m/z \) become distributed along the linear analyser. This occurs due to (i) the volume of ions in each packet; (ii) the length of each pulse of ions injected; (iii) the initial kinetic energy distribution during the injection. The mass resolution can be increased by increasing the length of the field-free region, or decreasing the acceleration voltage of the ions. Reducing the acceleration voltage would result in a loss of sensitivity, whilst increasing the flight path would be a practical issue.

A proposal which circumvented these drawbacks was proposed by Mamyrin [329] in 1973 who placed a reflectron at the far end of the ToF tube; a series of ring electrodes and grids which act as an ion ‘mirror’, deflecting ions 180° before they hit a detector near the ion source (Figure 1.12b). This increased the length of the field-free region two-fold without a large increase in instrument size, correcting the ion packet dispersion and thus increasing mass resolution. The dispersion is minimised by allowing ions with a greater kinetic energy to penetrate into the reflectron to a greater extent than low energy ions, allowing ions of the same \( m/z \) but at opposite ends of the kinetic energy distribution to reach the detector simultaneously.

Quadrupole and ToF mass analysers are often coupled together in hybrid mass spectrometers to form a quadruple time-of-flight mass (QToF) spectrometer. There are many commercially available QToF instruments, many of which have been coupled with ‘soft’ ionisation techniques to produce powerful instrumentation for the analysis of macromolecules.

1.6.4 Detectors

Once the ions have passed through the mass analyser(s) the last stage of the MS experiment is ion detection, where a useable signal is created. Whilst there are many types
of detector, this thesis will focus on Photomultiplier and Electron Multiplier detectors and Microchannel Plate Detectors.

1.6.4.1 Photomultiplier and Electron Multiplier Detectors

Photomultiplier and Electron Multiplier detectors are often coupled with quadrupole MS instruments. When ions travel through the quadrupole mass analyser only, they will be detected on the photomultiplier or electron multiplier located at the entrance to the ToF analyser.

A photomultiplier assembly consists of two conversion dynodes, one for each polarity, a phosphorescent screen and a photomultiplier. A photomultiplier can detect both positive and negative ions; ions are accelerated towards a dynode of the opposite polarity, and upon impact secondary electrons are emitted. These electrons are then accelerated towards the phosphorescent screen, where they are converted to photons. These photons then pass into the photomultiplier, where they produce an electric current which is amplified, with a typical amplification range of $10^4$ to $10^6$.

1.6.4.2 Microchannel plates detector

The microchannel plates detector (MCP) is most commonly used in ToF MS instruments. The MCP assembly consists of thousands of tiny electron multiplier channels, conductive glass capillaries (2 to 25 μm in diameter) which are combined into a thin disc. Each of these channels works independently as a continuous secondary-electron multiplier. Combined, these channels form an array, connected by deposition of a metallic semiconducting substance with a voltage applied across both the input and output side electrodes. When an ion collides with the inner wall of the channel, electrons are ejected and cascade through the channel, with each collision ejecting more electrons. In this way significant signal amplification occurs. Typically, the channels are positioned at a small angle (0 to $19^\circ$) to increase the chances of an incident particle striking the inner wall of the channel. Other MCP setups consist of multiple plates working in tandem, which further amplifies the signal.
Introduction

with gains ranging from $10^4$ to $10^7$ depending on the MCP arrangement. Once the cascade of electrons passes through the channel, the signal is measured as current.

1.7 Ion Mobility Mass Spectrometry

Ion mobility mass spectrometry (IM-MS) is a gas-phase electrophoretic technique which separates ions based on their mass ($m$), charge ($z$) and also their mobility ($K$) through a given buffer gas. The mobility of an ion can be related to its size and shape, thus giving conformational information regarding an analyte.

1.7.1 Drift Tube Ion Mobility Mass Spectrometry

The simplest setup of IM-MS is drift tube IM-MS (DT IM-MS), which has been in development since the early 20\textsuperscript{th} century, when both theoretical [330] and experimental [331, 332] investigations into the movement of ions through gases were carried out. In the 1960s, McDaniel et al. [333, 334] developed the first low-field drift tube instrument, through which gaseous ions could drift the length of the tube under the influence of a weak electric field. The concept of DT IM-MS is that an ion of given mass ($m$) and charge ($z$) will move at a velocity through a drift region under the influence of an electric field, whilst being impeded by collisions with gas molecules. Once an equilibrium between these two processes is reached, the ion will attain a constant average velocity along the electric field gradient. From measurement of the ions velocity, it is possible to directly determine the rotationally averaged collision cross section (CCS; $\Omega$) of an analyte, giving insight into its size and shape. More compact ions will undergo fewer collisions with buffer gas molecules than more extended ones, thus traversing the drift region faster. Similarly, ions carrying higher charge will travel faster than those with lower charge. As DT IM-MS relies on a packet of ions being pulsed into the drift region, the technique is well positioned for coupling to ToF mass analysers; as its pulse triggered release of packets of ions into the analyser allows its duty cycle to be synchronised with mobility separation events.

1.7.1.1 DT IM-MS Theory
DT IM-MS separates species based on their velocity in the presence of an inert buffer gas, under the influence of weak linear electric field. The behaviour of the ion within the electric field \((E)\) is dependent on the ratio of field strength to buffer gas number \((N)\). At high values of \(E/N\) (high field) ions align in the field and their motion becomes dependent on \(E\). At low values of \(E/N\) (low field) the motion of the ions is independent of \(E\), and ions possess low velocities \((v_d)\) proportional to \(E\) [335]. As such, the mobility \((K)\) of an ion therefore is determined by its velocity \((v_d)\) attained under the electric field \(E\) as it traverses a drift tube of length \(L\) over time \(t_d\):

\[ v_d = KE = \frac{L}{t_d} \]

Since the mobility of an ion \((K)\) is dependent on buffer gas number \((N)\), normalised standard pressure and temperature values decouple this relationship from experimental variation in the parameters. Normalised temperature \((T_0; 273.15 K)\) and pressure \((P_0; 760 Torr)\) allow the calculation of reduced mobility:

\[ K_0 = K \frac{T_0P}{TP_0} \]

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

\[ K_0 = \frac{3ze}{16N} \left( \frac{2\pi}{\mu k_B T} \right)^{\frac{1}{2}} \frac{1}{\Omega} \]

Where \(z\) is the ion charge, \(e\) \([C]\) is the elementary charge, \(\mu\) \([kg]\) is the reduced mass of the analyte and buffer gas, \(k_B\) is the Boltzmann constant, \(T\) \([K]\) is the temperature and \(N\) \([m^{-3}]\) is the buffer gas number density at standard conditions.
1.7.1.2 DT IM-MS: The MoQToF

All of the IM-MS work in this thesis was carried out on the MoQToF (Mobility QToF). This instrument is an in-house modification of the commercial QToF mass spectrometer (Waters, Manchester, UK) enabling IM-MS separation. The main modification to the QToF instrument was the inclusion of a drift cell chamber containing a drift cell of length 5.1 cm, along with the associated ion optics, gas lines, electrical connectors, etc. [336]. A schematic diagram of the instrument is shown in Figure 1.13. The drift cell is filled with buffer gas, in this work helium, to a pressure of 3.5-4.2 Torr (4.66 – 5.6 mbar). This additional gas pressure in the vacuum pumped mass spectrometer is handled by a 500 L/s Pfeiffer TMHS20 turbomolecular pump (Pfeiffer Vacuum Ltd, Newport Pagnell, UK) which is backed by an Edwards two-stage E2M40 rotary vane pump (Edwards Vacuum, Crawley, UK) alongside a mechanical booster pump which is utilised when the drift cell is filled with gas.

Figure 1.13 Schematic of the drift tube ion mobility mass spectrometer ‘MoQToF’. Drift chamber highlighted in purple and drift cell highlighted in blue.

The temperature of the buffer gas in the drift tube can be altered, ranging from ~80 K to ~550 K. Multiple temperature settings are used in this work. Cell heating is implemented via a series of tungsten wire wound ceramic heaters embedded onto both the drift cell body (8
1. Introduction

heaters) and end caps (2 heaters) and controlled with two alternating current (AC) generators. Cooling of the drift cell is carried out by passing a flow of nitrogen gas cooled with liquid nitrogen around the channels of the cell body and end caps.

1.7.1.3 The MoQToF: operation and typical settings

The MoQToF can be operated in MS mode only or IM-MS mode. In both cases helium gas is added in to the drift cell. In MS mode, ions are generated in the source and enter into the first vacuum region in a z-shaped path due to the orthogonal positioning of the capillary and the first two apertures; the sample and extraction cones. In a standard experiment the source block is heated to 80 °C and this, alongside the z-spray setup improves sensitivity, removing neutral species and aiding desolvation. The ions are then focused through the hexapole RF lens into the second vacuum region, the quadrupole analyser. The last vacuum region is the ToF analyser, which is held at the highest vacuum to ensure minimal background ion interference and better detection. Ions enter this region and are both focussed and orthogonally accelerated by a pulsed pusher, sending the ions down the drift tube towards the reflectron. At the reflectron, the ions undergo a 180° change in flight path and return to the microchannel plates. A voltage of ~1800 – 2000 V is applied to the MCPs and the ions are detected via electron cascade. This signal is converted to a time of arrival for each pulse of ions in a time-to-digital (TDC) converter at a rate of 1 GHz. This is read by the software MassLynx 4.1 (Waters, Manchester, UK) and output as a mass spectrum.
In IM-MS mode the ions undergo a very similar flight path to that in MS mode; however an additional pulsed event occurs in the drift cell to enable mobility separation. The ions are trapped in the pre-cell hexapole by raising the voltage on a lens located at the end of the hexapole, known as ‘top-hat’ lens 1 (TH₁) (Figure 1.13, Figure 1.14). The trapping voltage is lowered for 40 μs allowing the ions to enter the drift cell due to a potential difference between the end of the hexapole and the entrance of the drift cell. This potential difference is termed the injection energy (IE) and is set at 34-38. Too low an IE, and the packet of ions will not completely enter the drift cell, too high and they will travel too far into the drift cell before they begin to drift, lowering the effective separation length. The ions drift down a uniform linear electric field set across the cell. A typical mobility separation involves performing a number of mobility events at different drift voltages applied across the cell, further discussed in Section 1.7.1.4. Once the ions exit the drift cell they are refocused by passing through a short hexapole, before travelling along the rest of the instrument for detection at the MCPs.

The pulse of ions entering the drift cell is synchronised with the ToF pulse, to allow one mobility separation to be sampled 200 times (i.e. 200 pushes by the pusher) by the ToF. This synchronisation is carried out by setting the frequency at which ions are pulsed into the drift cell, which is dependent on the TOF pusher period. The ToF pusher period is itself dependent on the m/z range chosen and as such, the protein sample being analysed. The frequency (f) of the drift cell pulse is calculated by \( f = \frac{1}{t} \), where \( t \) is the ToF pulse time (or pusher cycle time) multiplied by the number of ToF pulses per ion mobility separation (200). As an example, for a ToF pusher time of 120 μs, \( f = \frac{1}{(120*200)} = 41.7 \) Hz. This value is set and applied using a DG 535 digital delay/pulse generator (Stanford research systems, Sunnyvale, CA, USA).

Tuning of the MoQToF is carried out in two ways; mass spectrometer settings are controlled via MassLynx (Waters, Manchester, UK) and the drift cell and associated ion optics are controlled by an external power supply. A schematic of the ion optics is shown in Figure 1.14a. In short, the drift voltage across the cell is determined by the difference...
between C₁ and C₂, and is altered throughout an IM-MS experiment. Voltages applied to both pre- and post-cell lenses are tuned for maximum transmission of ions whilst retaining native structure. Lenses C₁, C₂, H₂ and TH₂ float upon the hexapole collision cell voltage which is applied in MassLynx. Pre-cell lenses H₁, TH₁ and L₁₃ are referenced to the cell body (C₁) and post-cell lens L₄ is referenced to the end-cap of the drift cell C₂.

1.7.1.4 The experimental workflow: Acyl Carrier Protein, an example

Once nESI conditions have been optimised and drift cell optics are tuned in MS mode, a reference spectra is recorded. Then, in IM-MS mode, the drift cell voltage is set at 60 V and is varied by lowering the voltage on lens C₁ in steps of 10 or 15 V. IM-MS separations are taken at a minimum of five distinct drift voltages with temperature and pressure recorded at each drift voltage.

The output of a mobility experiment is the total ion count (TIC) which is a peak (or closely related peaks) termed the total arrival time distribution (tATD) (Figure 1.15b). At each drift voltage value a number of these peaks is acquired (i.e. a number of mobility events per experiment) which are summed to yield greater signal intensity (Figure 1.15a). The number of tATDs recorded per experiment can be varied to account for weak signals. To enable output of both MS and IM-MS experiments together, we use a modified version of MassLynx 4.1 (Waters, Manchester, UK) which accommodates the ion mobility capabilities of the instrument and allows the user to view the full mass spectrum of all ions arriving at the detector following an IM-MS separation (Figure 1.15c) along with the tATD. MassLynx also allows extracted ion arrival time distributions (ATDs) to be obtained for any given region of the m/z range (Figure 1.15c, red selected m/z range), allowing ATDs for individual species to be viewed (Figure 1.15d). In this way, the CCS (Ω) for ions of a given m/z can be calculated.
Figure 1.15 The data collection and fitting process. a) TIC peaks for each mobility separation. Multiple peaks can be summed to increase total signal. b) total arrival time distribution (tATD) for all ions arriving at the detector following all mobility separations, example from drift voltage 35 V. c) mass spectrum for apo-Acyl carrier protein with highlighted charge state. d) extracted ATD of apo-acyl carrier protein. Gaussian peaks fitted in Origin 9.1 shown as green and blue, with cumulative fit shown in red. E) plot of arrival time vs P/V, with average arrival time plotted for each conformer. X-axis intercept is equal to t₀.

An extracted ion ATD can contain a number of peaks. These may be well separated but could also be poorly resolved, resulting in shoulders or tails adjacent to the main peak.
These occur under a number of conditions; species of the same m/z but of different oligomeric order will be separated by IM-MS, for example a monomeric species of charge [M+3H]^{3+} will be m/z coincident with a dimeric species of [D+6H]^{6+}. Multiple peaks could also be due to the analyte presenting in multiple conformations, the more extended of which will have a longer arrival time and the more compact a shorter arrival time. In the example of the acyl carrier protein, the protein presents in multiple monomeric conformations, giving rise to multiple peaks in each extracted ATD (Figure 1.15d).

The extracted ATDs for a chosen m/z value are summed for each drift voltage using Microsoft Excel (Microsoft, Redmond, WA, USA), for example the acyl carrier protein [M+7H]^{7+} at drift voltage 35 V in Figure 1.15c. At each drift voltage, the average arrival time for each species present in the ATD is calculated. For ATDs with a single distribution, the average arrival time is equal to the peak centre. For ATDs with multiple distributions, the Peak Analyser tool in Origin 9.1 (OriginLab Corporation, USA) is used to fit multiple Gaussian distributions that satisfy the ATD (Figure 1.15d), calculating the centre of each Gaussian peak and using this as the average arrival time. The raw arrival time output (t_a) of an IM-MS experiment includes the 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 can be calculated by taking the average value of the intercept from a linear plot of average arrival time versus pressure/temperature (Figure 1.15e). This can be subtracted from the arrival time to calculate the drift time of the ions (t_D) using Equation 1.8:

\[
\text{Equation 1.8} \quad t_D = t_a - t_0
\]

For analytes with multiple, poorly resolved conformations, the minimum number of Gaussian distributions required to satisfy total ATD spread for all charge states were fitted in Origin (OriginLab Corporation, USA). Due to the presence of closely related conformers, individual CCS values calculated from arrival times for each conformer are not reported. Instead, the ATD for each m/z species is converted to a collision cross section distribution (CCSD). Due to the direct relationship between drift time (t_D) and CCS (Ω) shown in Equation 1.6 and Equation 1.7, this conversion can be done directly using Equation 1.9 [292]:

\[
\text{Equation 1.9} \quad \Omega = \frac{t_D}{t_a}
\]
Introduction

1. Equation 1.9

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

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.

For CCSDs presented in Chapters 3, 4 and 5 conformational families are fitted to the raw CCSD data to allow simple description of the changes to the conformational landscape. In each case, the minimum number of conformational families required to satisfy total CCSD spread for all charge states were determined using the Peak Fitter tool in Origin 9.1 (OriginLab Corporation, USA) and the centroid value for each conformational family held constant across charge states and sample, as applicable. For the CCSDs presented in Chapters 4 and 5, CCSDs for every charge state were summed to produce a total CCSD, prior to the fitting of conformational families.

1.7.2 Travelling Wave Ion Guide Mobility Mass Spectrometry

Travelling wave ion guide ion mobility (TWIMS) is a technique developed by Waters (Waters, Manchester, UK), who introduced the first commercially available IM-MS instrument in 2006 known as the Synapt G1 high definition mass spectrometer (HDMS). Since then, they have expanded their instrument range to include the Synapt G2, Synapt G2S and the Synapt G2Si. In a TWIMS instrument, ions are stored in the trap region, before being injected into the mobility cell where they are separated. After exiting the mobility cell, the ions travel through an additional ion guide, the transfer cell before they enter the ToF region. Ion mobility separation in a Synapt occurs due to a travelling voltage wave created by a series of transient DV voltages. The ions are radially confined by an RF field applied to consecutive electrodes, and are propelled forwards by the DC wave. The mobility cell is filled with nitrogen buffer gas, with which the ions collide. Ions with the highest mobility are able to travel on the crest of the DC wave, whereas ions with low mobility fall behind wave fronts and take longer to reach the end of the ion guide. As such, ions of differing mobilities are separated.
1. Introduction

Unlike DT IM-MS, the time taken for ions to pass through the TWIMS mobility cell cannot be directly related to its CCS. Instead, an ions CCS is proportional to $t_0^X$, where X is an experimentally determined parameter which is dependent on individual instrument settings. As such, in order to calculate the CCS of a species of interest using TWIMS, calibration using a set of standards of known CCS must be carried out under identical instrumental conditions. An excellent review of TWIMS calibration and CCS calculation can be found in reference [277].

1.8 Hydrogen Deuterium Exchange Mass Spectrometry

Hydrogen deuterium exchange mass spectrometry (HDX-MS) is a solution-phase technique which allows labelling of solvent accessible sites on a protein, effectively mapping protected areas of the protein backbone. HDX was first pioneered in the 1950’s [337], before its coupling to ESI-MS allowed protein conformation to be probed [338]. Since then HDX-MS has been widely used for the study of protein conformation [339], mapping protein:protein [340] and protein:ligand [280] interactions, and characterisation of oligomeric species and aggregates [341].

Figure 1.16 Schematic of a typical HDX experiment. A protein is incubated with D₂O for numerous time points. Exposed residues (blue) exchange more readily and become red. The reaction is quenched and the sample digested before being subject to mass spectrometry analysis.

A typical HDX experiment begins with the labelling of protein via incubation with excess deuterium atoms in D₂O. Exposed backbone amide hydrogens or those involved in weak hydrogen bonds can exchange readily with deuterium, whereas those buried in the hydrophobic core of the protein or involved in strong hydrogen bonds exchange more slowly. As shown in Figure 1.16, protein samples are incubated with D₂O for a series of time
1. Introduction

points, before the reaction is quickly quenched with a low pH buffer (pH 2.5) at low temperatures (0-1 °C). The careful pH and temperature control minimises back-exchange of incorporated deuterium atoms. The protein is then digested by a protease, usually a non-specific protease which operates at low pH such as Pepsin [342], to produce many overlapping peptides of varying mass. These peptides are subject to liquid chromatography separation before being analysed by mass spectrometry.

The excess of deuterium ensures exchange for hydrogen atoms occurs readily [343], and the time points chosen can be optimised between seconds and hours, depending on the timescale of protein dynamics or binding events. The mass difference of 1 Da between hydrogen and deuterium atoms results in a mass shift of 1 Da per deuterium atom incorporated. Thus, shift of the isotopic distribution of each peptide gives insight into the solvent exposure of each region of the protein.

HDX-MS experiments presented in this thesis were carried out on a fully automated LEAP autosampler system (HTS PAL, Leap Technologies, Carrboro, NC, USA) previously described [344, 345] and an online Acquity UPLC M-class HDX System (Waters Inc, Manchester, UK). Protein was diluted to 10 µM, before protein solution was incubated with excess D₂O and incubated at 18 °C for a range of time points. Following deuterium on-exchange, 50 µl of the labelled protein solution was quenched by adding 50 µl of quench buffer (pH 2.66) at 1 °C, and samples were passed across an immobilized pepsin column (Enzymate BEH pepsin column, Waters Inc, Manchester, UK) at 100 µL min⁻¹ (H₂O + 0.1% Formic acid, 20 °C). The resulting peptides were trapped on a UPLC BEH C₁₈ Van-Guard Pre-column (Waters Inc, Manchester, UK) and then gradient eluted (1 minute loading time, 8% - 85% ACN + 0.1% formic acid gradient, 40µl min⁻¹, 1°C) across a UPLC BEH C₁₈ column (Waters Inc, Manchester, UK) before undergoing electrospray ionization and analysis using a Synapt G2Si mass spectrometer (Waters Inc, Manchester, UK). Data was analysed using ProteinLynx Global Server (PLGS) (Waters, Manchester, UK), Dynamx v1.0 (Waters, Manchester, UK) and Origin v9.0 (OriginLab Corporation, USA). Results were taken in triplicate and all peptides selected were present in all three zero second incubation time points, conditions were optimised to ensure maximum peptide sequence coverage with good intensity.

1.9 Computational Methods to Predict Collision Cross Section
A major strength of IM-MS for the study of protein conformation is the ability to compare CCS ($\Omega; \text{Å}^2$) to molecular coordinates from X-ray crystallography and NMR experiments and molecular dynamics simulations. Most frequently the ion coordinates are taken from the protein databank (PDB) or molecular modelling outputs and are input into MOBCAL [346-348], developed in the Jarrold group, to determine the rotationally averaged CCS for a molecule. The main types of CCS calculation methods available are the Projection Approximation (PA), the Exact Hard Sphere Scattering (EHSS) method, or the Trajectory Method (TM) and the more recently developed Projection Superposition Approximation (PSA) and the Ion Mobility Projection Approximation Calculation Tool (IMPACT).

### 1.9.1 The Projection Approximation Method

The PA is the simplest of the CCS calculation models. When using PA, the ion is modelled as a collection of overlapping hard spheres with radii equal to hard sphere collision distances. The CCS is calculated by averaging the geometric cross section over all possible collision geometries, accounting for buffer gas collisions. As a result, this method is accurate if molecular surfaces are convex, whereas it is well known that biological systems contain cavities which are either shielded from buffer gas collisions or undergo multiple ion-molecule internal collisions due to reflections. CCS values calculated by PA can be significantly underestimated; however the method is computationally cheap, very fast and has shown good correlation to other methods for macromolecules [349].

### 1.9.2 Projection Superposition Approximation

The PSA [349, 350] models the ion in the same way as the PA, measuring the projection of the ion but with inclusion of size and shape effects. A large molecule with many atoms will have an increased attractive potential effect on long-range interactions the buffer gas molecules. PA accounts for this effect by deepening the potential energy well and increasing the surface curvature for molecules with increasing the number of atoms.
1.9.3 The Exact Hard Sphere Scattering Method

The exact hard sphere scattering (EHSS) method is similar to the PA model, and models ions by assuming an infinite hard wall potential between each atom of the ion and gas molecule. Unlike the PA, the EHSS model accounts for multiple collisions and scattering events occurring during an ion-molecule collision, therefore considering rough surfaces and cavities. This method works best for large macromolecular structures, where the effect of multiple scattering is greater than the ion-neutral interaction. The EHSS has been shown to overestimate the contributions of multiple scattering, sometimes overestimating CCS values by ~20% [346, 351].

1.9.4 The Trajectory Method

The trajectory method (TM) is widely accepted as the most accurate of the CCS calculation methods; however it is very computationally expensive and can therefore be slow, particularly when used for CCS estimations of macromolecules. TM treats ions as a collection of atoms, each one represented by a potential [347]. The effective potential for a given molecule is obtained by summing over multiple ion-buffer gas collisions and atomic contributions, and trajectories are then run at this potential to obtain the scattering angle of a buffer gas atom path. All possible collision geometries are averaged to find the CCS.

1.9.5 Ion Mobility Projection Approximation Calculation Tool

IMPACT is an algorithm developed to consider structural proteomics. A more recent addition to the computational technique repertoire, IMPACT has been shown to have excellent correlation with the TM, with an error of ~1%, whilst maintaining very fast calculations[352]. IMPACT functions by separating atoms of a given molecule into arranged cuboidal subsections, each of which can be further subdivided. When monitoring the collision events of a gas molecule, IMPACT first considers whether this collision has occurred in the largest subsection of the molecule. If so, the algorithm then interrogates smaller and smaller cuboidal subsections of the molecule surface, until the smallest subsection containing the location of the gas molecule collision event is reached. As such,
IMPACT is able to retain accuracy whilst maintaining significant speed, calculating the CCS values for the entire PDB databank (rscb.org) in several hours[352].

1.10 References

1. Introduction


1. Introduction


Introduction

1. Introduction

1. Introduction


1. Introduction


1. Introduction


1. Introduction


1. Introduction

1. Introduction


1. Introduction


1. Introduction


1. Introduction


1. Introduction


1. Introduction


1. Introduction


2
Insights into the conformations of three structurally diverse proteins: Cytochrome c, p53 and MDM2, provided by Variable Temperature Ion Mobility Mass Spectrometry
2.1 Declaration

This chapter consists of one published perspective article:


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

As a first author on this publication I expressed and purified p53DBD with the aid of DC. I carried out all experiments in Figure 2.5 and Figure 2.6 and compiled the data for all figures. I also drafted and edited the manuscript. TH and PVN provided samples. KJP and MR collected the data for Figure 2.3 and Figure 2.4 to allow comparison of disordered proteins with a structured protein.
Insights into the conformations of three structurally diverse proteins: Cytochrome c, p53 and MDM2, provided by Variable Temperature Ion Mobility Mass Spectrometry.

Eleanor R. Dickinson¹, Ewa Jurneczko², Kamila J. Pacholarz², David J. Clarke², Matthew Reeves³, Kathryn L. Ball³, Ted Hupp³, Dominic Campopiano², Penka V. Nikolova⁴, Perdita E Barran*¹

¹Manchester Institute of Biotechnology, University of Manchester, M1 7DN, U.K.
²School of Chemistry, University of Edinburgh, Edinburgh, EH9 3JJ, U.K.
³Institute of Genetics and Molecular Medicine, CRUK Cancer Research Centre, University of Edinburgh, Edinburgh EH4 2XR, U.K.
⁴School of Biomedical Science, Institute of Pharmaceutical Sciences, King’s College London, 150 Stamford Street, London, SE1 9NH, U.K.

E-mail: Perdita.barran@manchester.ac.uk

Publication Date (Web): January 28, 2015

Supporting Information

More details on the proteins studied in this paper along with some background material about DT-IM-MS is contained in supporting information. This material is available free of charge via the Internet at http://pubs.acs.org and in Appendix A of this thesis.
2. Insights into the conformations of three structurally diverse proteins: Cytochrome c, p53 and MDM2, provided by Variable Temperature Ion Mobility Mass Spectrometry

2.2 Abstract

Thermally induced conformational transitions of three proteins of increasing intrinsic disorder, cytochrome c, the tumour suppressor protein p53 DNA binding domain (p53 DBD) and the N-terminus of the oncoprotein Murine Double Minute2 (NT-MDM2), have been studied using near-native mass spectrometry and variable temperature drift time ion mobility mass spectrometry (VT-DT IM-MS). Ion mobility measurements were carried out at temperatures ranging from 200 K to 571 K. Multiple conformations are observable over several charge states for all three monomeric proteins, and for cytochrome c dimers of significant intensity are also observed. Cytochrome c [M+5H]^{5+} ions present in one conformer of CCS ~1200 Å², undergoing compaction in line with the...
Insights into the conformations of three structurally diverse proteins: Cytochrome c, p53 and MDM2, provided by Variable Temperature Ion Mobility Mass Spectrometry

reported T_m = 360.15 K before slight unfolding at 571 K. The more extended [M+7H]^{7+} cytochrome c monomer presents as two conformers undergoing similar compaction and structural rearrangements, prior to thermally induced unfolding. The [D+11H]^{11+} dimer presents as two conformers, which undergo slight structural compaction or annealing before dissociation. p53 DBD follows a trend of structural collapse before an increase in the observed CCS, akin to that observed for cytochrome c but proceeding more smoothly. At 300 K the monomeric charge states present in two conformational families, which compact to one conformer of CCS ~1750 Å² at 365 K in line with the low solution T_m = 315-317 K. The protein then extends to produce either a broad unresolved CCS distribution, or for z > 9, two conformers. NT-MDM2 exhibits a greater number of structural rearrangements, displaying charge state dependent unfolding pathways. DT IM-MS experiments at 200 K resolve multiple conformers. Low charge states species of NT-MDM2, present as a single compact conformational family centered on CCS ~1250 Å² at 300 K. This undergoes conformational tightening in line with the solution T_m = 348 K before unfolding at the highest temperatures. The more extended charge states present in two or more conformers at room temperature, undergoing thermally induced unfolding before significant structural collapse or annealing at high temperatures. Variable temperature IM-MS is here shown to be an exciting approach to discern protein unfolding pathways for conformationally diverse proteins.

2.3 Introduction

Experimental and theoretical approaches over the last twenty years have supported a hypothesis that that biomolecular folding pathways involve interplay between random coils, partially folded intermediate structures and ‘native’ folded proteins. It has been postulated that up to 33% of all eukaryotic proteins have regions of intrinsic disorder [1], lacking stable secondary or tertiary structure under physiological conditions in vitro [2]. This has led to the quartet model of protein folding to describe structures held by proteins in the cell[3]. Intrinsically disordered proteins (IDPs) are implicated in all cell processes but are especially abundant in signalling pathways where it has been reported that up to 75% of proteins contain large regions of intrinsic disorder[4]. These proteins exhibit a “one-to-many” binding strategy[5], undergoing disorder-to-order transitions to bind to multiple different binding partners with low affinity but high specificity. Study of these disorder-to-order (or reverse) transitions during protein folding has become vital in the understanding of protein function. A wide range of biophysical techniques such as small-angle x-ray
Insights into the conformations of three structurally diverse proteins: Cytochrome c, p53 and MDM2, provided by Variable Temperature Ion Mobility Mass Spectrometry (SAXS) [6, 7], NMR spectroscopy [8, 9], hydrogen deuterium exchange (HDX)[10], mass spectrometry [11, 12] and circular dichroism [13] have been employed to study protein folding and unfolding pathways. A range of methods such as thermally induced [6-9, 12], radiation induced[14, 15], acid induced [8, 10, 16, 17] and mechanical [18] unfolding have also been employed to follow unfolding pathways.

In rapid development since the 1970’s, the hybrid technique ion mobility-mass spectrometry (IM-MS) has evolved as a powerful and sensitive gas-phase technique able to elucidate information regarding protein conformational changes. IM-MS separates analyte ions by their charge, mass and shape, giving insight into the surface area in the form of a rotationally averaged collision cross section (CCS), commonly expressed in Å².

In some home built IM-MS instruments it is possible to vary the temperature of the drift gas, facilitating gas phase experiments where the effect of temperature on the structure of isolated ions may be observed[19-22]. This approach was taken by both Jarrold, and Bowers, who used collisional activation provided by increasing injection energies [21] or temperature [11, 12, 23, 24] to probe the unfolding transitions of various proteins by IM-MS. More recently, protein unfolding was modulated by addition of salt molecules and investigated by IM-MS as a function of temperature[25]. Whilst most studies regarding the effect of temperature on protein structure are performed in the solution phase, where the protein is stabilised by both intramolecular interactions and interactions between the protein and its surrounding solvent, gas phase techniques utilise solvent free charged analyte ions, thus removing all hydration effects and allowing intramolecular interactions upon unfolding to be probed. For some proteins the effect of temperature on structure is difficult to study in a solution phase assay since thermally induced structural transitions are coupled to aggregation. In vacuo, we are able to measure the conformational space occupied by the protein along with the impact of removing all stabilising interactions in the gas phase – including those that might arise from self-assembly. Whilst an unusual environment to study protein structure, use of nano-electrospray ionisation (nESI) allows preservation of protein tertiary and quaternary structure upon transfer into the gas phase. VT DT IM-MS allows isolated ions to be thermally activated, which in turn induces unfolding transitions that can be captured on a timescale of approximately 1 ms[19, 20]. Further advantages include low required sample volume and cost effectiveness compared with a more traditional ITC workflow.
2.
Insights into the conformations of three structurally diverse proteins: Cytochrome c, p53 and MDM2, provided by Variable Temperature Ion Mobility Mass Spectrometry

Obtaining the CCS of a protein conformer enables comparison with data obtained from other biophysical techniques such as NMR or X-ray crystallography or structural coordinates gained from computational studies.[26-29] A handful of studies have exploited homemade IM-MS instrumentation with variable temperature capabilities to report on protein unfolding[11, 12, 24], thermodynamic measurements [30] and hydration[12, 31].

![protein structures](image)

**Figure 2.2** a) X-ray diffraction crystal structure, protein sequence, Disopred disorder prediction and theoretical molecular weight for a) cytochrome c (PDB ID: 1HRC) residues 1-105 b) p53 DNA binding domain (DBD) (PDB ID: 2OCJ) residues 94-312 and c) N-terminal MDM2 (PDB ID: 1YCQ) residues 17-26. All structures are created using PyMol. A crystal structure for MDM2 residues 1-17 is unavailable due its high level of disorder.

We here use the simplest set-up of IM-MS, drift time IM-MS (DT IM-MS) with variable temperature capabilities (VT DT IM-MS) to track conformational changes occurring at a range of temperatures, probing the thermally induced unfolding transitions of three biomolecules with varying levels of disorder: globular protein cytochrome c, p53 DNA binding domain (DBD), shown to be 21% disordered [32] and Murine Double Minute 2 N-terminal domain (NT-MDM2), which is 29%
Insights into the conformations of three structurally diverse proteins: Cytochrome c, p53 and MDM2, provided by Variable Temperature Ion Mobility Mass Spectrometry

disordered (Figure 2.2). The Disopred graphs in Figure 2.2 show the probability of a given amino acid being disordered. A confidence score of over 0.5 indicates disorder, however the calculations do not account for groups such as the Heme group of cytochrome c (Figure 2.2a) or the presence of high number of protein loops in p53 DBD (Figure 2.2b). Such values and figures should therefore be treated with caution and used as a guide only.

2.4 Experimental Methods

2.4.1 Mass Spectrometry and Variable Temperature Ion Mobility Mass Spectrometry

All mass spectra and VT DT IM-MS measurements were acquired on an in-house modified quadrupole time-of-flight mass spectrometer (Waters, Manchester, UK) [19] containing a copper coated drift cell of length 5.1cm. Ions were produced by positive nESI with a spray voltage of 1.48-1.65 kV. IM-MS experiments were performed with helium as the buffer gas, its pressure measured using a baratron (MKS Instruments, UK). The buffer gas is heated by a series of tungsten wire wound ceramic heaters embedded onto both the drift cell body (8 heaters) and the end cap (2 heaters) and controlled with two AC generators. Cooling of the drift cell was carried out by passing a flow of nitrogen gas cooled with liquid nitrogen around the channels of the cell body and end caps. The drift voltage across the cell was varied by decreasing the cell body potential from 60 V to 15 V, with arrival time measurements taken at a minimum of five distinct voltages. Temperature and pressure readings were taken for each drift voltage and used in analysis of drift time measurements. nESI tips were prepared in-house with a micropipette puller (Fleming/Brown model P-97, Sutter Instruments Co., USA) using 4” 1.2 mm thin wall glass capillaries (World Precision Instruments, Inc., USA) and filled with 10-20μL of sample at the specified concentration.

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 CCS distribution plots are derived from raw arrival time data using Equation 1.9 below [12].

\[
\Omega_{\text{avg}} = \frac{(\Omega_0)^{1/2}}{16} \left[ \frac{1}{m_n} + \frac{1}{m} \right]^{1/2} \frac{ze}{(K_B T)^{1/2}} \frac{1}{p \sqrt{L}} \frac{V}{L^2}
\]
Insights into the conformations of three structurally diverse proteins: Cytochrome c, p53 and MDM2, provided by Variable Temperature Ion Mobility Mass Spectrometry

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; \( p \) 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 (\( t_d \)):

\[
\text{Equation 1.8} \quad t_0 = t_a - t_0
\]

Further information on the technique of DT IM-MS can be found in Appendix A.

2.4.2 Proteins

Equine cytochrome c, purchased from Sigma-Aldrich Company Ltd (Dorset, UK), was dissolved in a 3:1 Water:Acetonitrile (LC-MS grade) mixture to a concentration of 50μM. Details of the expression and preparation of p53 DBD and N-terminal MDM2 are found in Appendix A. Both proteins were sprayed via nESI from 50uM solutions with 50 mM ammonium acetate (pH 6.8).

2.5 Results and Discussion

2.5.1 Cytochrome c

2.5.1.1 VT-MS cytochrome c

50 μM cytochrome c was sprayed from 3:1 \( \text{H}_2\text{O}:\text{ACN} \) at drift cell temperatures varying from 300 K to 571 K (Figure 2.3). At 300 K the protein presents as a charge state distribution (CSD) spanning 5 monomeric charge states \( 4 \leq z \leq 8 \), with \([\text{M}+6\text{H}]^{6+}\) and \([\text{M}+7\text{H}]^{7+}\) most dominant at comparable
2. Insights into the conformations of three structurally diverse proteins: Cytochrome c, p53 and MDM2, provided by Variable Temperature Ion Mobility Mass Spectrometry

intensity. At this concentration and with careful control of source conditions, we also observe dimeric species \([\text{D+9H}]^{9+}\) and \([\text{D+11H}]^{11+}\) and trimeric species \([\text{T+13H}]^{13+}\) and \([\text{T+14H}]^{14+}\) as reported previously.[33] We see no evidence for the \([\text{D+13H}]^{13+}\) species. At 422 K there is loss of the \([\text{D+9H}]^{9+}\) species along with a decrease in the intensity of the remaining dimeric and trimeric species as well as \([\text{M+4H}]^{4+}\), \([\text{M+5H}]^{5+}\) and \([\text{M+6H}]^{6+}\) monomeric species. At 472 K and above, the CSD widens to include the monomeric \([\text{M+9H}]^{9+}\) and \([\text{M+10H}]^{10+}\) species, accompanied by a notable reduction in the intensity of the lower charge states and of \([\text{D+11H}]^{11+}\) and both trimeric species. We infer from this that the higher order species are thermally dissociated on entry to the heated drift cell. This wider charge state range persists at high temperatures. At 521 K the trimeric species are lost, and at 571 K, the \([\text{D+11H}]^{11+}\) species is no longer observed. Whilst the \([\text{D+11H}]^{12+}\) dimer has a unique m/z value, it is likely that there are also \([\text{D+10H}]^{10+}\) and \([\text{D+12H}]^{12+}\) dimers under the \([\text{M+5H}]^{5+}\) and \([\text{M+6H}]^{6+}\) monomeric peaks. As we observe no \([\text{D+13H}]^{13+}\) dimer, we assume the \([\text{M+7H}]^{7+}\) peak comprises solely monomer.
Insights into the conformations of three structurally diverse proteins: Cytochrome c, p53 and MDM2, provided by Variable Temperature Ion Mobility Mass Spectrometry.

Figure 2.3 nESI mass spectra of 50µM cytochrome c sprayed from 3:1 H2O:ACN as a function of buffer gas temperature. Single grey spheres indicate monomeric species, double grey spheres indicate dimeric species and triple grey spheres denote trimeric species. Grey bars highlight peaks at increasing temperature.
Insights into the conformations of three structurally diverse proteins: Cytochrome c, p53 and MDM2, provided by Variable Temperature Ion Mobility Mass Spectrometry

2.5.1.2 VT-IM-MS cytochrome c

Monomeric [M+7H]7+ charge state presents as a collision cross section distribution (CCSD) with two conformational families at 300 K, centered at a more populated ~1000 Å² and ~1600 Å² (Figure 2.4a). Both conformations remain present at 323 K. At 374 K the distribution changes, with most analyte ions presenting in the more extended conformational family. When the drift cell temperature is increased to 422 K, a compaction of the protein is observed yielding a conformational family at ~1200 Å² and a lower intensity family at ~900 Å² along with loss of the more extended conformational family. At 472 K we observe a single conformation at an increased CCS of ~1700 Å².

The CCSD for the dimeric [D+11H]11+ species (Figure 2.4b) shows two conformational families centered at ~2000 Å² and ~2900 Å². These conformers persist when the temperature is raised to 323 K, before slight compaction at 374 K. At 422 K we see a loss of the more compact conformer, leaving a single conformational family centered at 2900 Å². This extended conformer remains at 521 K. At the highest temperature of 571 K there was no [D+11H]11+ present.

From the CCSD the compact [M+5H]5+ charge state of cytochrome c (Appendix A, Figure S1a) can be seen to present a broad ATD, in contrast to that reported by Jarrold, M [12] who saw one conformer at all temperatures (Appendix A, Figure S1). We attribute this difference to the presence of dimeric [D+10H]10+ cytochrome c and to some other monomeric conformers. As no [D+9H]9+ or [D+11H]11+ is present in the spectra at 571 K, we assume the CCSD for [M+5H]5+ at 571 K comprises monomer only. This is corroborated by the similar CCS of the [M+5H]5+ observed at ~1200 Å² by Jarrold.

The remaining conformational families of monomer and/or dimer compact at 323 K and present larger conformations at higher temperatures. At 422 K we are able to observe some of the dimeric species in a more extended conformer, as seen for [D+11H]11+ (Figure 2.4b). Upon temperature increase to 472 K the compact conformer is lost. The extended conformer reduces in intensity at 521 K and is lost at 571 K.

2.5.1.3 cytochrome c Discussion
2. Insights into the conformations of three structurally diverse proteins: Cytochrome c, p53 and MDM2, provided by Variable Temperature Ion Mobility Mass Spectrometry

The trend exhibited by \([\text{M+7H}]^{7+}\) is one of compaction prior to unfolding (Figure 2.4a), largely corresponding to the results observed by Jarrold, M.[12] (Figure 2.4c). Most previous work using IM-MS to examine cytochrome c has reported a single resolvable conformer for the \([\text{M+7H}]^{7+}\) species [11, 12, 21, 34, 35] although our DT IM-MS of this ion has shown the protein presents in two conformers centered at \(\sim 1100\ \text{Å}^2\) and \(\sim 1900\ \text{Å}^2\) [25], at both 260K and at 300K when sprayed from an aqueous solution in the presence of sodium iodide, in line with our results here at 300K. An explanation for the difference between these two IM-MS data sets is that our source is significantly softer resulting in the preservation of a compact monomer family (also found for \([\text{M+5H}]^{5+}\) (Appendix A, Figure S 1)) even at this charge state. An alternative explanation is that this compact ion is due to dissociation from multimers in source.

With an increase in temperature to 374 K and above, we begin to observe structural rearrangements in the CCSD, in line with the solution \(T_{\text{melt}} = 360.15\ \text{K}\) [10]. At 422 K we observe previously unobserved unfolding intermediates at lower CCS (Figure 2.4 a vs. c) nonetheless, Jarrold shows

---

**Figure 2.4 a) b) Stacked plots showing CCS distributions derived from arrival time distributions for 50µM cytochrome c as a function of buffer gas temperature for charge states \([\text{M+7H}]^{7+}\) and \([\text{D+11H}]^{11+}\) respectively. c) comparison plot showing \([\text{M+7H}]^{7+}\) taken by Jarrold.[12] CCSD for \([\text{D+11H}]^{11+}\) at 521 K has been magnified x 10 to improve visibility.**
Insights into the conformations of three structurally diverse proteins: Cytochrome c, p53 and MDM2, provided by Variable Temperature Ion Mobility Mass Spectrometry

Structural rearrangements at this temperature, with the protein exhibiting multiple resolved conformers. The [M+7H]⁺ species undergo significant extension at higher temperatures, akin to that shown by Jarrold, although here the CCSD is broader, suggesting more flexibility, or more overlapping conformational families. From the CCSD of the [D+11H]¹¹⁺ species in Figure 2.4b, we attribute the loss of the compact conformational family at 422 K to annealing and/or to symmetric dissociation of the dimer. The latter is supported by the decrease in the intensity of [D+11H]¹¹⁺ and corresponding increase in the intensity of the [M+5H]⁵⁺ and [M+6H]⁶⁺ species (Figure 2.3). Upon temperature increase to 472 K and above, we see evidence for asymmetric dissociation of the [D+11H]¹¹⁺ extended conformer, with highly charged monomeric species 9 ≤ z ≤ 11 observable in the mass spectra, a process that has been previously observed at room temperature following collisional activation in a range of instruments [33, 36]. We attribute the very low intensity shoulder on the CCSD at 521 K to trimer dissociation, corresponding to their loss from the mass spectrum at this temperature (Figure 2.3). A trend of thermally induced symmetric dissociation of the compact conformational family prior to asymmetric dissociation of the unfolded conformational family is therefore observed for the dimeric species of cytochrome c. In line with the total loss of [M+11H]¹¹⁺ dimer from the mass spectrum, there is no CCSD available at 571 K.

The CCSD trend for the monomeric cytochrome c [M+5H]⁵⁺, shown as hatched highlighted Gaussian distributions in Appendix A Figure S 1, is comparable to that observed in earlier work carried out by Mao, Y et al. [11] (shown in Appendix A Figure S 1b). Mao, et al. observe slight compaction at temperatures up to 472 K, before unfolding to ~1250 Å² at 472 K and above, whereas we observe slight compaction at temperatures up to 521 K before unfolding at 571 K only. Mao et al. attribute this compaction to high injection energies; but this is not applicable to our low injection energy workflow. We suggest this conformational tightening and subsequent CCS changes are due to structural rearrangements or annealing of cytochrome c, which remarkably start to be observed at buffer gas temperatures close to the solution T_melt = 360.15 K [10]. We observe no large increase in CCS, suggesting this monomeric conformational family of this charge state of the protein remains compact still possessing some of its folded regions at high temperatures, a result previously corroborated by molecular dynamic calculations [11].

It may also be true that the [D+10H]¹⁰⁺ dimer (Appendix A Figure S 1) follows a similar trend to the [D+11H]¹¹⁺ dimer in Figure 2.4b, with symmetric dissociation of the compact conformer prior to asymmetric dissociation of the extended conformer at a higher temperature, but the presence of some more dynamic monomer species prevent a conclusive argument here.
Insights into the conformations of three structurally diverse proteins: Cytochrome c, p53 and MDM2, provided by Variable Temperature Ion Mobility Mass Spectrometry

The use of VT IM-MS on this structured protein sprayed from aqueous solution reveals the difference in the stability of each observed charge state. All forms exhibit some compaction during IM measurement at a drift gas temperature close to the solution T_{melt}. The [M+5H]^{5+} ion remains relatively compact at elevated temperatures, by contrast the [M+7H]^{7+} ion extends following compaction, and there is evidence for a sharp unfolding transition as the temperature is raised above 450 K. The dimeric forms as well as higher order multimers dissociate as the temperature is elevated. The dimers present a wide CCSD at room temperature, with more than one conformational family present. The cytochrome c dimers also compact below 400 K, and at higher temperatures lose the compact families prior to the loss of more extended conformers.

2.5.2 p53 DBD

2.5.2.1 VT-MS p53 DBD

The mass spectra of p53 DBD show the protein presents as a zinc-bound monomer at 300 K (Appendix A Figure S 2) of charge state range $7 \leq z \leq 12$. There is no evidence for dimeric species. The vast majority of the protein presents in charge states [M+9H]^{9+} and [M+10H]^{10+} as shown previously.[37] With a drift gas temperature of 365 K, the signal intensity remains centered on the [M+9H]^{9+} and [M+10H]^{10+} charge states. Upon further temperature increase to 400 K, we see an increase in the intensity of [M+11H]^{11+} and a corresponding decrease in the intensity of the [M+9H]^{9+} ion with the [M+10H]^{10+} charge state remaining most intense. As the temperature is raised the peaks also narrow, which can be attributed to a loss of salts and/or associated water.

2.5.2.2 VT-IM-MS p53 DBD
Insights into the conformations of three structurally diverse proteins: Cytochrome c, p53 and MDM2, provided by Variable Temperature Ion Mobility Mass Spectrometry

The CCSD for p53 DBD [M+9H]$^9+$ species is shown in Figure 2.5a. At 300 K two conformational families are seen, with the protein ions presenting largely in one conformer centered at $\sim$1800 Å$^2$ with a small proportion of the protein centered at CCS $\sim$2500 Å$^2$. When the temperature is raised to 365 K, the extended conformer is no longer observed, leaving the compact conformer centered at $\sim$1800 Å$^2$. Upon temperature increase to 400 K this conformational family remains centered on CCS $\sim$1800 Å$^2$, although a widening of this distribution is observed.

![Figure 2.5 CCS distributions derived from arrival time distributions for 50 µM p53 DBD as a function of buffer gas temperature for charge states a) [M+9H]$^9+$, b) [M+10H]$^{10+}$ and c) [M+11H]$^{11+}$. X-axis denotes the rotationally averaged CCS (Å$^2$).](image)

For [M+10H]$^{10+}$ the two conformational families at $\sim$1800 Å$^2$ and $\sim$2500 Å$^2$ are again observed at 300 K. The compact conformer is most populated once more, however a greater proportions of ions present in the larger conformer compared with that observed for [M+9H]$^9+$. When the temperature is raised to 365 K the extended conformer is lost, leaving the single conformational family at $\sim$1800 Å$^2$. At 400 K, two conformational families are observable, the more compact at $\sim$1800 Å$^2$ the most intense, with a lower intensity family at $\sim$2250 Å$^2$.

For the more extended charge state of [M+11H]$^{11+}$ there are again the two conformational families centered at $\sim$1750 Å$^2$ and $\sim$2500 Å$^2$. We observe a different CCSD compared with charge states.
2. Insights into the conformations of three structurally diverse proteins: Cytochrome c, p53 and MDM2, provided by Variable Temperature Ion Mobility Mass Spectrometry

[M+9H]$^9+$ and [M+10H]$^{10+}$, with the majority of the ions presenting as the larger conformer. The more extended conformer is lost at 365 K, leaving a conformational family at ~1800 Å$^2$ as previously observed. At 400 K the larger conformer is again observed, with the two families showing comparable intensities.

2.5.2.3 p53 DBD Discussion

The mass spectra for p53 DBD exhibits a narrow charge state range indicative of a stable globular protein, arising from solution conditions where the high salt content preserves a compact form of the protein. At 400 K, the intensity of the [M+9H]$^9+$ charge state lowers and [M+11H]$^{11+}$ increases. This apparent gain of average charge held on the protein is attributed to a loss of negatively charged counter ions, as supported by the narrowing of the peak widths in the mass spectrum. The CCSDs show evidence for several conformations across this charge state range as reported previously [37]. For the [M+9H]$^9+$ charge state at room temperature most of the ions are in a compact conformation whereas for [M+11H]$^{11+}$ the extended family is more populated. In line with the protein solution $T_{melt}$ = 315-317 K we see evidence of structural compaction and loss of the more extended conformer at 365 K for all charge states, producing a single narrow CCSD. Following this compaction we see unfolding of the protein to a higher CCS, with the appearance of a larger conformational families for [M+10H]$^{10+}$ and [M+11H]$^{11+}$ and a widening of the CCSD for the [M+9H]$^9+$ ions. The broader CCSD suggests flexibility in the protein at the higher temperature, resulting in unresolved conformations. We can assign this to an unfolding pathway similar to that seen for monomeric [M+7H]$^7+$ cytochrome c in Figure 2.4a, with compaction of the protein around the $T_{melt}$ prior to protein unfolding. In contrast to cytochrome c, p53 does not exhibit a large unfolding transition, although it is possible that this may occur at elevated temperatures we were unable to reach here.

2.5.3 N terminal MDM2

2.5.3.1 VT-MS N-terminal MDM2
2. Insights into the conformations of three structurally diverse proteins: Cytochrome c, p53 and MDM2, provided by Variable Temperature Ion Mobility Mass Spectrometry

nESI mass spectra obtained from 50 µM NT-MDM2 sprayed from high salt conditions (50 mM Ammonium Acetate) at a range of buffer gas temperatures (Appendix A Figure S 3) exhibits a wide charge state range $5 \leq z \leq 14$. At 200 K the $[M+7H]^7+$ charge state is the most intense, with low intensity $[D+11H]^{11+}$, and $[D+13H]^{13+}$ dimers present. At 300 K the CSD is again centred on $[M+7H]^7+$, upon temperature increase to 350 K, the CSD shifts towards higher charge states and $[M+10H]^{10+}$ becomes the most intense species. The $[M+10H]^{10+}$ charge state remains the most intense as the temperature is raised to 500K, however at 500 K we also observe a charge state distribution shift towards lower charge states, and an increased intensity in the $[M+7H]^7+$ species.

2.5.3.2 VT IM-MS N-Terminal MDM2

The compact $[M+7H]^7+$ charge state (Figure 2.6a) is the most intense species at 300 K where it is present in one distinct conformational family of CCS $\sim 1250 \text{ Å}^2$. The CCSD at the lowered temperature of 200 K shows this conformer at $\sim 1250 \text{ Å}^2$ and a more intense, more compact conformer at $\sim 950 \text{ Å}^2$. Upon temperature increase to 350 K, these two conformers are again observable at $\sim 950 \text{ Å}^2$ and $\sim 1250 \text{ Å}^2$, whereas at 400 K the compact conformer is lost, and only the extended conformer of CCS $\sim 1250 \text{ Å}^2$ is observed. Further extension of this conformer is seen at 500 K, to a CCS of $\sim 1500 \text{ Å}^2$. The $[M+8H]^{8+}$ (Figure 2.6b) charge state shows a similar distribution with two, well separated conformers observed at 200 K, at CCS’s of $\sim 1200 \text{ Å}^2$ and $\sim 1500 \text{ Å}^2$. 
Insights into the conformations of three structurally diverse proteins: Cytochrome c, p53 and MDM2, provided by Variable Temperature Ion Mobility Mass Spectrometry

Figure 2.6 a, b and c show CCS distributions at a range of temperatures for NT-MDM2 at charge states [M+7H]^{7+}, [M+8H]^{8+} and [M+12H]^{12+}, respectively.

At 300K the protein is again observable in a single conformational family at ~1250 Å^2. At 350 K, the wide ATD for this charge state suggests a rearrangement of the protein back into at least two conformations, although these are not well resolved. When the temperature is raised to 400 K, we see loss of the more compact conformer and an increase in the CCS of the more extended conformer to ~1500 Å^2, although multiple indistinct conformational families may be present. At 500 K, one main conformational family is present at ~1600 Å^2, with a very small proportion of the protein ions existing in a more compact conformer at ~1300 Å^2. The extended [M+12H]^{12+} charge state depicted in Figure 2.6c again shows the protein existing in multiple conformations when the drift gas temperature is 200K, with two widely distributed conformational families centered at ~1500 Å^2 and ~2000 Å^2. At 300 K, these two families are observed in a narrower distribution centered at ~1500 Å^2 and ~1900 Å^2, suggesting that the low temperature data shows conformations that are frozen, but that interconvert at the higher temperature, before loss of the more compact conformer at 350 K. Further extension of the larger conformer to ~2300 Å^2 is observed at 400 K, however upon temperature increase to 500 K we see a significant compaction event resulting in a single conformer centered at ~1700 Å^2.

2.5.3.3 N-Terminal MDM2 Discussion

The mass spectra of NT-MDM2 exhibits two “foldamers”, one compact and centered around charge state [M+7H]^{7+} and one more extended and centered around charge state [M+10H]^{10+}. This is indicative of a partially disordered protein, for example a structured domain with an extended disordered tail as seen for NT-MDM2 [38]. At 200 K the peak for each species appears badly resolved suggesting that it has retained significant amounts of salt and/or water. At 300 K, the spectra are better resolved suggesting that collisions with the drift gas at room temperature (and above) remove these excipients and effectively enhance the desolvation. The injection energy is the same in all experiments. This effect may also account for the change in the CSD between 200 and 300 K since salts may be lost as charged species. A general trend of temperature dependent extension before
Insights into the conformations of three structurally diverse proteins: Cytochrome c, p53 and MDM2, provided by Variable Temperature Ion Mobility Mass Spectrometry

compaction at very high temperatures is observable in the mass spectra. A decrease in the intensity of the compact species (charge states $5 \leq z \leq 8$) and an increase in the extended species ($z \geq 8$) with temperature increase from 200 K to 400 K is seen, indicating protein extension. We then see a decrease in the intensity of the high charge species ($z \geq 8$) at 500 K, suggesting protein collapse or melting. For the compact $[M+7H]^{7+}$ and $[M+8H]^{8+}$ charge states this trend is also observed in the CCSD (Figure 2.6 a, b). The single conformer at room temperature exhibits some compaction at 350 K consistent with the $T_{\text{melt}} = 348$ K [39]. At temperatures above 350 K protein extension is observed with a CCS increase to $\sim 1500 \text{ Å}^2$ at 500 K for both charge states. Loss of protein structure is seen alongside unfolding resulting in a narrow CCS distribution at 500 K, indicative one conformational family with little dynamics. The more extended charge state $[M+12H]^{12+}$ exhibits similar unfolding transitions, although no compaction of the protein close to the solution $T_{\text{melt}}$ is observed. Instead, smooth thermal unfolding is observed from 300 K to 400 K with loss of the more compact conformational family. At 500 K a pronounced compaction in the protein occurs producing a single conformer at low CCS. We hypothesise that this species could be molten globule like in structure and is observable at all extended charge states (Appendix A Figure S4). By lowering the drift cell temperature to 200 K we are able to ‘freeze out’ conformations which would not otherwise be observable. At this temperature, we report a wider CCS distribution for all charge states and an increased CCS for $z < 8$, a phenomenon discussed elsewhere [23, 25]. Compared with the trends reported for cytochrome c and p53 DBD, for MDM2 we observe charge state dependent unfolding pathways. It seems thermal activation of this disordered and flexible protein produces multiple conformational changes as the protein unfolds, many of which cannot be resolved on our experimental timescale.

2.6 Conclusions

Variable temperature drift time ion mobility-mass spectrometry has been implemented to study the unfolding pathways of three proteins of different structural content. In-house modified instrumentation enabled the gas–phase collision cross section distributions at increasing temperature to be elucidated. The difference in the behaviour of the $[M+5H]^{5+}$ and $[M+7H]^{7+}$ species of cytochrome c indicates different unfolding pathways of these two charge states which differ only by two protons. The $[D+11H]^{11+}$ species presents in two conformers, and undergoes compaction before the compact conformer undergoes symmetric dissociation at 422 K, followed by the
Insights into the conformations of three structurally diverse proteins: Cytochrome c, p53 and MDM2, provided by Variable Temperature Ion Mobility Mass Spectrometry

extended conformer which asymmetrically dissociates at 521 K. p53 DBD follows similar compaction and extension events that are observed for monomeric cytochrome c. The compact charge state [M+9H]$^9+$ exists in one broad conformational family at 300 K, exhibiting a narrowing of the CCSD at 365 K before a widening at 400 K. The more extended charge states [M+10H]$^{10+}$ and [M+11H]$^{11+}$ present as two conformers at 300 K, losing the more extended conformer at 365 K before extending once more to two conformers at 400 K. The low $T_{\text{melt}} = 315-317$ K suggests the compaction events may be collapse of protein structure. Aggregation of the protein at high temperatures prevented data above 400 K from being taken. Highly disordered NT-MDM2 undergoes multiple charge state dependent unfolding stages in the gas phase. IM-MS analysis at 200 K ‘freezes out’ a wide range of conformations, presumably as the barrier to interconversion is lowered [23]. Structural rearrangements at 350 K; compaction at the low charge states and extension at the higher charge states, correlate well with $T_{\text{melt}} = 348$ K before high temperature unfolding of a single, potentially collapsed globular structure is observed. DT IM-MS captures the intermediate unfolding stages of highly dynamic proteins in the gas phase which may not be observable in solution.

2.7 References

Insights into the conformations of three structurally diverse proteins: Cytochrome c, p53 and MDM2, provided by Variable Temperature Ion Mobility Mass Spectrometry

Insights into the conformations of three structurally diverse proteins: Cytochrome c, p53 and MDM2, provided by Variable Temperature Ion Mobility Mass Spectrometry


3

The use of Ion Mobility Mass Spectrometry to probe Modulation of the Structure of p53 and of MDM2 by Small Molecule Inhibitors
3.1 Declaration

This chapter consists of one published perspective article:


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

As a first author on this publication I carried out all experiments, compiled the data for all figures, drafted and edited the manuscript. TRH, JZP and GS provided Np53 and NT-MDM2 samples.
The use of Ion Mobility Mass Spectrometry to probe Modulation of the Structure of p53 and of MDM2 by Small Molecule Inhibitors

Eleanor R. Dickinson¹, Ewa Jurneczko², Judith Nicolson²-³, Ted R Hupp³, Joanna Zawacka-Panaku⁴, Galina Selivanova⁴ and Perdita E. Barran¹

¹ The Michael Barber Centre for Collaborative Mass Spectrometry Manchester Institute of Biotechnology, University of Manchester, Manchester, M1 7DN, U.K.
² School of Chemistry, University of Edinburgh, Edinburgh, EH9 3JJ, U.K.
³ Institute of Genetics and Molecular Medicine, CRUK Cancer Research Centre, University of Edinburgh, Edinburgh EH4 2XR, U.K.
⁴ Department of Microbiology, Tumor and Cell Biology, Karolinska Institutet, Stockholm, Sweden

E-mail: Perdita.barran@manchester.ac.uk

Publication Date (Web): July 10, 2015

Supporting Information

More details about DT-IM-MS and HDX-MS along with supplementary figures are contained in supporting information. This material is available free of charge via the Internet at http://journal.frontiersin.org/article/10.3389/fmolb.2015.00039 and in Appendix B of this thesis.
3. The use of Ion Mobility Mass Spectrometry to probe Modulation of the Structure of p53 and of MDM2 by Small Molecule Inhibitors

3.2 Abstract

Developing drug-like molecules to inhibit the interactions formed by disordered proteins is desirable due to the high correlation of disorder with protein implicated in disease, but is challenging due in part to the lack of atomistically resolved and resolvable structures from conformationally dynamic systems. Ion Mobility Mass Spectrometry (IM-MS) is well positioned to assess protein ligand interactions along with the effect of a given inhibitor on conformation. Here we demonstrate the use of IM-MS to characterize the effect of two inhibitors RITA and Nutlin-3 on their respective binding partners: p53 and MDM2. RITA binds N-terminal transactivation domain of p53 (Np53) weakly, preventing direct observation of the complex in the gas phase. Nonetheless, upon incubation with RITA, we observe an alteration in the charge state distribution and in the conformational distributions adopted by Np53 in the gas phase. This finding supports the hypothesis that RITAs mode of action proceeds via a conformational change in p53. Circular dichroism corroborates our gas phase findings, showing a slight increase in secondary structure content on ligand incubation, and HDX-MS experiments also highlight the dynamic properties of this protein. Using the same approach we present data to show the effect of Nutlin-3 binding to MDM2. MDM2 presents as two conformational families in the absence of Nutlin-3. Upon Nutlin-3 binding, the protein undergoes a compaction event similar to that exhibited by RITA on Np53. This multi-technique approach highlights the inherent disorder in these systems; and in particular exemplifies the power of IM-MS as a technique to study transient interactions between small molecule inhibitors and intrinsically disordered proteins.

3.3 Introduction

The transcription factor p53, dubbed the Death Star [1], is a multi-domain, intrinsically disordered protein (IDP) [2, 3]. The protein comprises the disordered N-terminal domain (Np53)[4] containing the transactivation domain (residues 1-61) and the proline-rich domain (residues 62-94), the central DNA binding domain (residues 94-292), the tetramerization domain (residues 325-355) and the C-terminal regulatory domain (residues 363-393). It is strongly implicated in tumour suppression pathways, where it functions to
The use of Ion Mobility Mass Spectrometry to probe Modulation of the Structure of p53 and of MDM2 by Small Molecule Inhibitors

block tumour development by triggering cellular senescence or apoptosis upon signals indicating DNA damage, oncogene activation or telomere erosion [5]. Under non-stressed conditions, low p53 levels are tightly maintained by MDM2 (murine double minute 2). MDM2 is a ~55 KDa IDP with roles as an Ubiquitin E3 ligase, as a molecular chaperone and also in translational control. MDM2 comprises the disordered “lid” mini-domain (residues 1-24) [6], the N-terminal domain (residues 25-109), the disordered central acidic domain (residues 221-276), the zinc finger domain (residues 299-331) and the C-terminal RING (really interesting new gene) domain (residues 430-480). MDM2 down regulates p53 activity in a negative autoregulatory feedback loop via three mechanisms; firstly, MDM2 blocks the transcription ability of p53 by direct binding through their respective N-terminal domains [7, 8]. Secondly, MDM2 exports p53 from the nucleus and thirdly, targets p53 by Ubiquitination for degradation via the proteasome [9, 10], p53 N-terminal domain binds into the MDM2 N-terminal domain hydrophobic pocket as an amphipathic helix, with residues Ph19, Trp23 and Leu26 comprising a triad of required contacts which insert into the MDM2 binding cleft [11].

Alteration of the p53 pathway is an almost universal hallmark of human cancers, with 22 million cancer patients living with abrogation of the p53 pathway, half of which display suppressed p53 function [12] and half of which exhibit p53 mutations. Cellular overexpression of MDM2 effectively abolishes p53 function, allowing unregulated cell cycle events in tumour cells. Inhibition of the p53:MDM2 complex is therefore a highly desirable therapeutic strategy; releasing, reactivating and stabilizing p53 levels, thus providing an attractive cancer therapy drug target. To date, numerous p53:MDM2 protein-protein interaction (PPI) antagonists have been identified, including cis-imidazolines [13] [14], “stapled” peptides [15, 16], terphenyls [17], oligobenzamides [18], spiro-oxindoles [19], chromenotriazolopyrimidines[20, 21] and Benzodiazepinedione [22]. The cis-imidazoline Nutlin-3 is composed of enantiomers a and b, of which enantiomer a is 150 times more potent, and binds MDM2 in the p53 peptide groove, mimicking the three p53 residues responsible for the bulk of binding interactions [13]. Nutlin-3 is effective in numerous cell lines, and is able to arrest or induce apoptosis in proliferating cancer cells with micromolar concentrations [23]. The drug candidate RITA (reactivation of p53 and induction of tumour cell apoptosis, NSC 652287) has been shown to restore wild-type p53 function in tumour cells by preventing the p53:MDM2 interaction [24]. In contrast to the Nutlins, which bind MDM2 in its N-terminal hydrophobic pocket [13], RITA binds to p53 N-terminal domain with estimated K_D = 1.5 nM [24]. It is hypothesised that RITA binds outside of the
The use of Ion Mobility Mass Spectrometry to probe Modulation of the Structure of p53 and of MDM2 by Small Molecule Inhibitors

p53/MDM2 binding cleft, allosterically exerting its effect via a conformational change in the highly disordered N-terminus of p53 (Np53) [24].

Since its advent in the 1970’s [25, 26], the hybrid gas phase technique Ion Mobility-Mass Spectrometry (IM-MS) has gained credibility as a tool to study the conformations adopted by proteins and peptides in the gas phase. IM-MS is especially effective in its use for studying IDPs [27-29] due to its ability to observe conformations adopted by analytes on a millisecond time scale [30, 31]. IM-MS provides information regarding charge, mass and shape of an analyte. The simplest setup of IM-MS is that of drift time IM-MS (DT IM-MS) [32]. Ions are separated by their mobility (K) as they traverse a drift cell of known length filled with buffer gas to a known pressure and temperature. Ions travel down a weak electric field (5-50 V cm⁻¹) colliding with buffer gas molecules which counter their progress until an equilibrium drift velocity, proportional to the electric field, is reached. The mobility (K) of an ion is the ratio between the drift velocity (v_d) and the applied electric field (E). The mobility of an ion can be used to calculate the rotationally averaged collision cross section (CCS, Ω, Å²) using Equation 1.7 [33]:

\[
\text{Equation 1.7} \quad K_0 = \frac{3ze}{16\pi} \left( \frac{2\pi}{\mu k_B T}\right)^{0.5} \frac{1}{\Omega}
\]

Where \( K_0 \) is the reduced mobility; \( z \) 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 Boltzmann constant and \( T \) is the gas temperature.

Here we employ native mass spectrometry, DT IM-MS, circular dichroism (CD) and hydrogen-deuterium exchange coupled to mass spectrometry (HDX-MS) to observe the conformations of N-terminal p53 domain (Np53) and the N-terminal domain of MDM2 (N-MDM2) both in the gas phase and in solution. We also probe the binding and conformational changes conferred by small molecule inhibitors; Nutlin-3 for N-MDM2, and RITA for Np53. Further information about DT IM-MS, CD and HDX-MS methodology can be found in Appendix B.
The use of Ion Mobility Mass Spectrometry to probe Modulation of the Structure of p53 and of MDM2 by Small Molecule Inhibitors

3.4 Materials and Methods

Expression and purification of both Np53 (residues 1-100) [34] [35] and N-MDM2 (residues 1-126) [36] have been previously described. Before the analysis reported here, the protein samples were thawed and dialysed in 50mM ammonium acetate using Bio-RAD micro bio-spin chromatography columns (Bio-Rad Laboratories, Inc). Concentrations of purified proteins were measured by the Thermo Scientific NanoDrop Spectrophotometer ND 1000 (Thermo Scientific, USA). Small molecule RITA (2,5-Bis(5-hydroxymethyl-2-thienyl) furan, NSC 652287) was reconstituted in 100% IPA and stored at -20 °C. Before analysis, RITA was thawed and diluted to 100 µM and an IPA concentration of 5% using 50 mM ammonium acetate. Nutlin-3 was reconstituted in 100% DMSO and stored at -80 °C. Before analysis, Nutlin-3 was thawed and diluted to 500 µM and a DMSO concentration of 1% using 50 mM ammonium acetate.

MS and IM-MS experiments were performed on Np53 and N-MDM2 from solutions buffered with ammonium acetate (pH 6.8). Np53 samples were incubated with 5 % IPA for 30 minutes at 37 °C to account for the solvent present in the RITA sample. N-MDM2 samples were incubated with 0.5 % DMSO for 30 minutes at room temperature to account for the solvent present in the Nutlin-3 sample. Binding experiments were performed on Np53 with RITA in a 1:2 protein:ligand ratio, samples were incubated for 30 minutes at 37°C. Binding experiments were performed on N-MDM2 and Nutlin-3 in a 1:10 protein:ligand ratio, samples were incubated for 30 minutes at room temperature. All MS and DT IM-MS data were acquired on an in-house modified quadropole time-of-flight mass spectrometer (Waters, Manchester, UK) [30] containing a copper drift cell of length 5.1 cm. Ions were produced by positive nano-electrospray ionisation (nESI) with a spray voltage of 1.3-1.62 kV. Helium was used as the buffer gas, its pressure measured using a baratron (MKS Instruments, UK). Buffer gas temperature and pressure readings (294.31 K – 303.69 K and 3.518 – 3.898 Torr, respectively) were taken at each drift voltage and used in the analysis of drift time measurements. The drift voltage across the cell was varied by decreasing the cell body potential from 60 V to 15 V, with arrival time measurements taken at a minimum of five distinct voltages. Instrument parameters were kept as constant as possible and are as follows: cone voltage: 114-119 V, source temperature: 80 °C.
The use of Ion Mobility Mass Spectrometry to probe Modulation of the Structure of p53 and of MDM2 by Small Molecule Inhibitors

nESI tips were prepared in-house using a micropipette puller (Fleming/Brown model P-97, Sutter Instruments Co., USA) using 4” 1.2 mm thin wall glass capillaries (World Precision Instruments, Inc., USA) and filled with 10-20 µL of sample.

Data was analysed using MassLynx v4.1 software (Waters, Manchester, UK), Origin v9.0 (OriginLab Corporation, USA) and Microsoft Excel. Experiments were carried out in triplicate. Ion arrival time distributions were recorded by synchronisation of the release of ions into the drift cell with mass spectral acquisition. The collision cross section distributions (CCSD) are derived from arrival time data using Equation 1.9: [33]

\[
\Omega_{\text{avg}} = \frac{(18\pi)^{1/2}}{16} \left( \frac{1}{m} + \frac{1}{m_b} \right)^{1/2} \frac{z e}{(k_B T)^{3/2} \rho L^2} t_d \frac{V}{L}
\]

Where Ω is the collision cross section (Å²); 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; ρ 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. For these experiments where the CCS has been evaluated experimentally with helium as the buffer gas and using a drift tube with a linear field we use the convention \( ^{12}\text{COS}_{\text{He}} \) to report our collision cross section values in the context of the mobility technique employed as well as the buffer gas used.

HDX-MS experiments were carried out using a fully automated LEAP autosampler system (HTS PAL, Leap Technologies, Carrboro, NC, USA) previously described [37, 38] and an online Acquity UPLC M-class HDX System (Waters Inc, Manchester, UK). Np53 and RITA were mixed at a 1:2 protein:ligand ratio and incubated for 30 minutes at 37 °C before analysis. Stock protein solutions (50 µM Np53 ± 100 µM RITA, with 5% IPA) were diluted to 10 µM with equilibration buffer. 3.8 µl protein solution was incubated with D_2O (54.2 µl labelling buffer) and incubated at 18 °C for 15, 30, 60 or 120 seconds. Following deuterium on-exchange, 50 µl of the labelled protein solution was quenched by adding 50 µl of quench buffer at 1 °C, and samples were passed across an immobilized pepsin column (Enzymate BEH pepsin column, Waters Inc, Manchester, UK) at 100 µL min\(^{-1}\) (H2O + 0.1% Formic acid, 20 °C). The resulting peptides were trapped on a UPLC BEH C_{18} Van-Guard Pre-column (Waters Inc, Manchester, UK) and then gradient eluted (1 minute loading time, 8% - 85% ACN + 0.1% formic acid gradient, 40µl min\(^{-1}\), 1°C) across a UPLC BEH C_{18} column.
The use of Ion Mobility Mass Spectrometry to probe Modulation of the Structure of p53 and of MDM2 by Small Molecule Inhibitors

(Waters Inc, Manchester, UK) before undergoing electrospray ionization and analysis using a Synapt G2Si mass spectrometer (Waters Inc, Manchester, UK). Data was analysed using ProteinLynx Global Server (PLGS) (Waters, Manchester, UK), Dynamx v1.0 (Waters, Manchester, UK) and Origin v9.0 (OriginLab Corporation, USA).

100 % sequence coverage was obtained for Np53 ± RITA. Selected peptides were restricted to be present in all 3 repeats of 0 second incubation time experiments.

3.5 Results

3.5.1 Modulation of N-terminal p53 by RITA

In the absence of RITA and under near neutral solution pH conditions, the mass spectra of Np53 (Figure 3.1a) presents a broad monomeric charge state distribution (CSD) range 5 ≤ z ≤ 12, with three major signals corresponding to the ions [M+6H]⁶⁺, [M+7H]⁷⁺ and [M+8H]⁸⁺, of which the [M+6H]⁶⁺ species is most intense. Upon incubation of Np53 with RITA we observe a shift in the CSD towards lower charge states. Specifically, Np53 in the presence of RITA (Figure 3.1b) exhibits a significant decrease in intensity of the [M+7H]⁷⁺ and [M+8H]⁸⁺ species, along with an increase in the intensity of the [M+5H]⁵⁺ species, an appearance of the [M+4H]⁴⁺ species and a loss of the high charge states z > 10. Although source conditions were carefully controlled to give gentle ionisation of the sample, the Np53:RITA complex was not strong enough to be retained during desolvation at any protein:ligand ratio (data not shown).
The use of Ion Mobility Mass Spectrometry to probe Modulation of the Structure of p53 and of MDM2 by Small Molecule Inhibitors

Figure 3.1 n-ESI mass spectra recorded for a) wild-type Np53 and b) wild-type Np53 in the presence of RITA. Np53 was incubated for 30 minutes at 37°C with addition of 5% IPA. Binding studies were carried out at a 1:2 Np53:RITA ratio, incubated for 30 minutes at 37 °C with 5% IPA content.

DT IM-MS was performed on Np53 both in the absence and presence of RITA. The collision cross section distribution ($^{\text{DT}}\text{CCSD}_{\text{loc}}$) (Figure 3.2a top panel) shows the Np53 [M+6H]$^{6+}$...
The use of Ion Mobility Mass Spectrometry to probe Modulation of the Structure of p53 and of MDM2 by Small Molecule Inhibitors

charge state presents as two conformational families; a more populated compact form (denoted C\(_1\), blue Gaussian distribution) centered at \(\sim 1250\ \text{Å}^2\) and a low intensity extended form (denoted X, green Gaussian distribution) centered at \(\sim 1500\ \text{Å}^2\). Two conformations are also observed for [M+7H]\(^{7+}\) (Figure 3.2b), which are assigned to X and a more intense larger distribution, centered at \(\sim 1750\ \text{Å}^2\), which is assigned to an unfolded form of the protein (U, purple Gaussian distribution). [M+8H]\(^{8+}\) (Figure 3.2c), is also made up of U, along with low intensity signal from a still more extended form (U\(_2\), gold Gaussian distribution), although this latter distribution is poorly resolved. [M+9H]\(^{9+}\) (Appendix B Figure S 6) presents in three conformational families; X, U and U\(_2\), of which the most extended U\(_2\) is most populated.

![Collision cross section distributions](image)

Figure 3.2 Collision cross section distributions \(^{\text{DTCCSD}_{\text{He}}}\) arising from arrival time distributions (ATDs) at drift voltage 35V for Np53 in the absence (top panels) and presence (bottom panels) of RITA. Distributions shown for the a) [M+6H]\(^{6+}\) b) [M+7H]\(^{7+}\) and c) [M+8H]\(^{8+}\) species. In the absence of RITA, Np53 was incubated with 5% IPA for 30 minutes at 37°C. Binding studies were carried out at a 1:2 Np53:RITA ratio incubated at 37°C for 30 minutes with 5% IPA. \(^{\text{DTCCSD}_{\text{He}}}\) are normalized to the intensity of the ion peak in the corresponding mass spectrum. Conformational families are denoted by hatched Gaussian curves showing novel compact conformational family C\(_0\) in red, compact conformational family C\(_1\) in blue, extended conformational family X in green, and unfolded conformational family U in purple.
Upon incubation with RITA the $^{DT}$CCSD$_{tie}$ for $[\text{M}+6\text{H}]^{6+}$ is significantly altered (Figure 3.2a bottom panel); we no longer observe the extended conformer X, observe a reduction in the population of compact conformer $C_1$, and the induction of a highly populated novel conformational family centered at $\approx 950 \text{ Å}^2$, $C_0$ (red Gaussian distribution). The $[\text{M}+7\text{H}]^{7+}$ CCSD (Figure 3.2b) is also altered by the presence of RITA, with loss of conformer U, and induction of both conformers $C_1$ and $C_0$. This change is accompanied by a decrease in intensity of this charge state. $[\text{M}+8\text{H}]^{8+}$ (Figure 3.2c) behaves similarly to $[\text{M}+7\text{H}]^{7+}$, with loss of conformer $U_2$, and induction of conformers X and $C_1$. We observe an increase in the intensity of the $[\text{M}+5\text{H}]^{5+}$ species (Appendix B Figure S 5) along with the appearance of a highly compact form of the protein $C_0$. This compaction is evident in all charge states, for example $[\text{M}+9\text{H}]^{9+}$ (Appendix B Figure S 6) has lost the population of the unfolded conformer $U_2$ upon incubation with RITA, alongside a reduction in intensity of conformers X and U and induction of highly populated $C_1$ conformational family. This alteration of the conformational spread as shown by the $^{DT}$CCSD$_{tie}$ is supported in solution by CD. Appendix B Figure S 7 shows the secondary structure content of Np53 increases upon incubation with RITA, supporting the hypothesis that RITA induces a novel conformer of Np53. Structural analysis using DiChroWeb [39] using CONTILL algorithm [40, 41] predicted that Np53 is 32% disordered, and upon incubation with RITA the level of disorder reduced to 28%.

Hydrogen-deuterium exchange coupled to mass spectrometry (HDX-MS) was used to ascertain if the conformational changes induced by RITA could be mapped in the solution phase. Np53 was incubated for varying time points in deuterated buffer, and the mass shift of peptides was determined. Np53 shows a significant uptake of deuterium at the shortest experimental time point of 15 seconds for a large proportion of peptides detected (Figure 3.3a-d). From the mass spectrometry data of each peptide, we observe no significant difference between deuterium uptake in the absence or presence of RITA indicating that we cannot sample the interconverting solution conformations for this highly dynamic protein over the longer timescale of the HDX-MS experiment. The butterfly plot in Figure 3.3g depicts the overall deuterium uptake differences between Np53 in the absence and presence of RITA. Each set of points along the x-axis represent a peptide, with time points denoted by different coloured points and lines (15 (yellow), 30 (red), 60 (blue) and 120 (black) seconds incubation time). Grey bands indicate the error in the uptake level and vertical lines indicate the sum of uptake differences for each time point. Several peptides
show deuterium uptake differences slightly above the error, but all at less than 1 Da, indicating that this protein is highly dynamic with or without RITA, for example, the greatest deuterium uptake difference in Np53 in the absence and presence of RITA being 0.271 Da, for a peptide spanning residues 23-39 with a [M+2H]^{2+} charge state.

Figure 3.3 a) c) deuterium uptake mass spectra for Np53 peptide residues 23-30 at time points 0, 15, 30, 60 and 120 seconds in the absence (a) and presence (c) of RITA. b) d) deuterium uptake charts for Np53peptide residues 53-63 at time points 0, 15, 30, 60 and 120 seconds in the absence (b) and presence (d) of RITA. e) f) relative deuterium uptake plots for Np53 in the absence (black points) and presence (red points) of RITA for peptide residues 23-30 and 53-63, respectively. g) butterfly plot of uptake difference comparison for Np53 in the absence and presence of RITA. X-axis denotes each peptide detected, in ascending residue number. Y-axis denotes uptake difference (Daltons). Coloured lines show uptake difference at 15 (yellow), 30 (red), 60 (blue) and 120 (black) seconds incubation time. Grey band indicates the error associated with each peptide. Vertical lines are the summed difference of all time points for each peptide.
Mass Spectra for MDM2 (Figure 3.4a) sprayed from native conditions with 50 mM ammonium acetate and 0.5% DMSO show a broad bimodal CSD spanning charge states $5 \leq z \leq 14$. The most intense species is $[M+10H]^{10+}$ with significant intensity also in $[M+7H]^{7+}$ and $[M+6H]^{6+}$. We observe low intensity $[D+11H]^{11+}$, $[D+13H]^{13+}$ and $[D+15H]^{15+}$ dimers, which means that the species attributed to $[M+5H]^{5+}$ will also contain some $[D+10H]^{10+}$ (and the $[M+6H]^{6+}$ some $[D+12H]^{12+}$ etc.) but since the flanking unique m/z dimers are of an intensity of $<5\%$ we ignore this contribution. Upon incubation with Nutlin-3 (Figure 3.4b) we see a CSD shift towards the lower charge states, with the $[M+6H]^{6+}$ species most intense, although the CSD range is retained. Binding of one Nutlin-3 molecule to MDM2 is observed at the $[M+5H]^{5+}$, $[M+6H]^{6+}$ and $[M+7H]^{7+}$ charge states. The shift in the N-MDM2 CSD upon incubation with Nutlin-3 is substantially greater than that caused by DMSO alone (Appendix B Figure S 8) suggesting that Nutlin-3 confers a structural or conformational change in N-MDM2.
The use of Ion Mobility Mass Spectrometry to probe Modulation of the Structure of p53 and of MDM2 by Small Molecule Inhibitors

Figure 3.4 nESI mass spectra of a) 50 µM N-terminal MDM2 (residues 1-126) sprayed from 50 mM ammonium acetate + 0.5% DMSO. b) 50 µM N-MDM2 + Nutlin-3 in a 1:10 protein:ligand ratio, incubated for 30 minutes at room temperature. Final DMSO concentration 0.5%. Monomeric species are denoted by single grey spheres, dimeric species by two grey spheres and N-MDM2 bound to Nutlin-3 by a single grey sphere and small blue sphere.

DT IM-MS analysis reveals that MDM2 in the absence of Nutlin-3 presents as two conformational families at all charge states (Appendix B Figure S 9). The [M+5H]^{5+} charge state (Figure 3.5a) presents as two conformers centered at \(~1000\) and \(~1250\) Å\(^2\), referred to as \(C_1\) (black Gaussian curve) and \(C_2\) (red Gaussian curve), respectively. The [M+6H]^{6+} charge state (Figure 3.5b) presents as three conformers, the compact \(C_1\) and \(C_2\) families and a more extended family, \(X\) (blue Gaussian curve) centered at \(~1400\) Å\(^2\). The [M+7H]^{7+} charge state , when bound to Nutlin-3, undergoes a compaction event to produce a single conformational family centered at \(~1250\) Å\(^2\), corresponding to conformer \(C_2\) (Figure 3.5d). This effect is again seen for the [M+7H]^{7+} charge state (Figure 3.5f), which presents as a single conformer corresponding to \(X\) when bound to Nutlin-3. These altered conformations remain, even when Nutlin-3 is not bound to N-MDM2. Figure 3.5 bottom panels show the DTCCSD\(_{\text{he}}\) of N-MDM2 in the presence of Nutlin-3, but not bound in a complex. The [M+5H]^{5+} species undergoes a minor change in the DTCCSD\(_{\text{he}}\) with an increase in conformational family \(C_2\) compared with the bound complex. Charge states [M+6H]^{6+} and [M+7H]^{7+} remain in the single conformational families \(C_2\) and \(X\), respectively, even after the ligand is no longer bound.
3. The use of Ion Mobility Mass Spectrometry to probe Modulation of the Structure of p53 and of MDM2 by Small Molecule Inhibitors

Figure 3.5 Collision cross section distributions \( ^{\text{DT}} \text{CCSD}_{\text{He}} \) derived from arrival time distributions (ATDs) for N-MDM2 sprayed from 50mM ammonium acetate (+ 0.5% DMSO) for the a) \([\text{M+5H}]^{5+}\), b) \([\text{M+6H}]^{6+}\) and c) \([\text{M+7H}]^{7+}\) charge states. Top panel represents N-MDM2, middle panel represents N-MDM2 bound to Nutlin-3 and bottom panel represents N-MDM2 incubated with Nutlin-3 but in apo-form with no small molecule bound. Proteins were incubated with Nutlin-3 in a 1:10 protein:ligand ratio, for 30 minutes at room temperature. CCSDs were taken at a drift voltage of 35 V and are normalized to the intensity of the ion species in the corresponding mass spectrum however to allow greater visibility, the \([\text{M+5H}]^{5+}\), \([\text{M+5H}]^{5+}\) bound to Nutlin-3 and \([\text{M+7H}]^{7+}\) bound to Nutlin-3 \( ^{\text{DT}} \text{CCSD}_{\text{He}} \) have been magnified x 10. Conformational families are depicted by coloured Gaussian curves; C\(_1\) (black), C\(_2\) (red), X (blue) and U (green).

3.6 Discussion

The MS spectra for Np53 in the absence of RITA (Figure 3.1a) corroborates previous reports of disorder for the N-terminus of p53,[2] [3] a broad charge state range 5 \( \leq z \leq 12\), indicative of a disordered system [42, 43] [43] with numerous residues available for protonation in solution. We are unable to preserve the binding of small molecule RITA to Np53, suggesting the binding is lower affinity than that reported previously (\( K_D = 1.5 \) nM [24]) or that it proceeds principally by hydrophobic interactions that are significantly
The use of Ion Mobility Mass Spectrometry to probe Modulation of the Structure of p53 and of MDM2 by Small Molecule Inhibitors

diminished in the absence of solvent, resulting in loss of ligand during desolvation. We observe a narrowing of the CSD for Np53 on ligand incubation, we also observe the isolated ligand (data not shown), which also supports our assertion of ligand dissociation during desolvation. This CSD shift toward lower charge states suggests conformational tightening induced by RITA, a hypothesis that is supported by DT IM-MS data. The $^{DT}\text{CCSD}_{He}$ for Np53 is significantly altered in the presence of RITA at all charge states present, with loss of larger conformational families and induction of more compact conformers. We observe a compact conformer $C_0$ for [M+5H]$, [M+6H]$, and [M+7H] charge states, which is not present in the absence of RITA. Whilst the [M+8H]$^8$ species does not contain any of the $C_0$, it no longer contains conformer $U$, rather is populated by the more compact conformers $C_1$ and $X$, although conformer $X$ is poorly resolved. [M+9H]$^9$ undergoes loss of conformer $U_2$ with induction of compact conformer $C_1$ at a much lower $^{DT}\text{CCSD}_{He}$. The use of IM-MS to discern conformational tightening due to ligand binding has been previously reported [29], and along with these findings provides an exciting prospect as a method for screening inhibitors to conformationally dynamic systems. As RITA is predicted to bind outside of the p53:MDM2 hydrophobic binding pocket [24], it has been asserted that the observed inhibition proceeds via a conformational change, which in turn will allosterically prevent the binding of MDM2. Our IM-MS data is evidence for the conformational modulation of Np53 by RITA. The induction of a smaller conformation is corroborated by CD results, which show an increase in secondary structure, and a decrease in the disordered content when analysed using the CONTILL algorithm. We do note, however, that the calculated differences in structural content predictions for Np53 are minimal, with only a 4% decrease in disordered content, this is less informative than the clear conformational change provided by IM-MS.

The use of HDX-MS reinforces the view that Np53 is conformationally dynamic in solution; high levels of deuterium uptake are observed after 15 seconds incubation, with very little further uptake at longer incubation times. This suggests that backbone amides are solvent exposed and free to exchange with deuterium. When uptake was compared after RITA incubation we observe no significant changes in deuterium uptake for Np53 (Figure 3.3g). While there are several peptides which exhibit deuterium uptake differences outside of the error, the greatest difference is 0.271 Da for a [M+2H]$^2$ peptide. As we see no difference greater than 1 Da, the mass difference between a hydrogen and deuterium atom, we can infer that there is no significant structural difference between Np53 in the absence and presence of RITA on the timescale of these experiments. Our shortest time step (15 s) is
insufficient to observe the conformational changes occurring as the protein has enough
time to rearrange back to its original conformations. In contrast, the isolated gas phase
conformers exiting the electrosprayed droplets appear trapped in distinct conformers at
least over the time scale of our IM-MS experiments. We estimate this time to be ~15 ms
including the transmission of ions to our drift cell [30], which is short enough to retain the
conformational changes induced by RITA such that they can be observed. Both of the
solution approaches indicate conformational flexibility and some slight change in structural
content in the presence of the ligand, IM-MS provides a more definitive, readout of the
modulation of conformation to Np53 in the presence of RITA.

We can contrast the results observed for the RITA interaction with Np53 with that for the
well-studied drug candidate Nutlin-3 with MDM2. N-MDM2 presents with a wide CSD (5≤z
≤14), again suggesting a disordered protein. DT IM-MS shows that N-MDM2 presents in the
gas-phase in two conformational families, potentially assignable to the previously reported
“open” and “closed” position of the lid mini-domain [6, 36, 44]. When incubated with
Nutlin-3, we observe a substantial CSD shift towards the lower charge states which cannot
be attributed to the effect of DMSO alone (Appendix B Figure S8), again suggesting some
conformational effect conferred by Nutlin-3 binding. We observe binding of Nutlin-3 to N-
MDM2 over three charge states, [M+5H]^{5+}, [M+6H]^{6+} and [M+7H]^{7+}. As there is no binding
to the more extended high charge states, Nutlin-3 may only be able to bind N-MDM2 in a
compact conformation, which is transferred to the gas phase as low change state complex.
DT IM-MS analysis showed that Nutlin-3 configures N-MDM2 into a more compact and
inflexible conformer. The [M+5H]^{5+} charge state retains both conformational families upon
Nutlin-3 binding, however the larger conformer at DTCCS_{he} ~1250 Å^2 was greatly reduced.

For the [M+6H]^{6+} and [M+7H]^{7+} ions, Nutlin-3 binding configures the protein into a single
conformer with a narrow DTCCSD_{he}, indicating less dynamics. This single conformational
family was centered at a DTCCS_{he} ~1250 Å^2 for [M+6H]^{6+}, corresponding to conformational
family C2, and ~1400 Å^2 for [M+7H]^{7+} corresponding to family X. We see loss of both the C2
and X families for [M+6H]^{6+} and loss of C2 and U for [M+7H]^{7+}, suggesting much lower
flexibility of the protein when bound to Nutlin-3.

Interestingly, it appears as for Np53, that the ligand free N-MDM2 in the IM-MS
experiments also retains a ‘memory’ of its in solution Nutlin-3 bound state. DTCCSD_{he} of N-
MDM2, incubated with Nutlin-3 but in its apo-form, show similar conformers than those
which retain binding of Nutlin-3 (Figure 3.5, bottom panels). This suggests that Nutlin-3

3. The use of Ion Mobility Mass Spectrometry to probe Modulation of the Structure of p53
and of MDM2 by Small Molecule Inhibitors
The use of Ion Mobility Mass Spectrometry to probe Modulation of the Structure of p53 and of MDM2 by Small Molecule Inhibitors

binds a higher proportion of analyte molecules than we observe, but is not retained fully during desolvation. The apo \([M+5H]^{5+}\) species is not only compact, suggesting that it rearranges back to the free N-PM22 conformer, or that some of it arises from a conformer in solution that is incapable of binding Nutlin. The apo \([M+6H]^{6+}\) and \([M+7H]^{7+}\) remain in tight, single conformational families, and a much lower proportion of the Nutlin-bound N-PM22 presents in the \([M+5H]^{5+}\) charge state (Figure 3.4) supporting our hypothesis that Nutlin-3 is unable to bind as well to the very compact conformer C_1. For the larger conformational families, N-PM22 seems unable to rearrange back to its original conformations within the timescale of desolvation and analysis.

3.7 Conclusions

Multiple techniques have been used to probe the binding of small molecule inhibitors RITA and Nutlin-3 to N-terminal p53 (Np53) and N-terminal MDM2 (N-PM22), respectively. Native mass spectrometry of Np53 shows a shift in the CSD towards the lower charge states and loss of the more extended charge states upon incubation with RITA. IM-MS of Np53 reveals two conformational families in the absence of RITA. Upon incubation with RITA, Np53 is configured into a novel, more compact conformer C_0 with loss of the more extended conformational family. We are able to retain this conformational tightening in the gas-phase on the time scale of our DT IM-MS experiments, even though we are unable to preserve the RITA:Np53 complex in the gas phase. HDX-MS data highlights the disordered nature of Np53, with no discernible conformations visible on a longer timescale. Very little differences are noted between the deuterium on-exchange of Np53 in the absence and presence of RITA, and we are unable to locate RITA induced conformational changes.

The nESI mass spectrum of N-terminal MDM2 shows a wide range of charge states (5≤ z ≤14) indicative of a disordered protein [42, 43]. The bimodal distribution suggests the protein may possess a more compact and more extended conformer. Indeed, DT IM-MS results show the protein presents as two conformational families at all charge states. Upon incubation with Nutlin-3, we observe ligand binding to the forms of the protein that present to the gas phase with low charge states \([M+5H]^{5+}\), \([M+6H]^{6+}\) and \([M+7H]^{7+}\), suggesting selective binding to a compact conformer of MDM2, or possibly that more
3. The use of Ion Mobility Mass Spectrometry to probe Modulation of the Structure of p53 and of MDM2 by Small Molecule Inhibitors

extended forms lose Nutlin-3 on transfer to the gas phase. The bound species of MDM2 are compact at all three charge states, with [M+6H]^{6+} and [M+7H]^{7+} forming a single conformational family centered at a CCS_{he} of the middle conformational family exhibited by apo-N-DM2. These conformational changes are likely retained by ions which lose the bound Nutlin-3 molecule during desolvation, indicating that the protein is unable to rearrange during the experiment. IM-MS is presented as a promising technique able to track conformational changes in unstructured proteins on a millisecond timescale.

3.8 Footnote

Since the publication of this manuscript, Wanzel, et al. [45] have published data showing that the sensitivity of p53wt cells to RITA does not proceed via a p53 dependent pathway, but rather via a DNA damage pathway. Using novel technique CRISP-Cas9, Wanzel, et al. showed that RITA activity directly correlates with the ability to induce DNA damage, and cells with RITA resistance also display remarkable cross-resistance to DNA crosslinking compounds such as cisplatin and additionally exhibit increased DNA cross-link repair. They conclude that whilst RITA binds to N-terminal p53, it acts primarily as a cross-linking drug whose activity is limited by DNA repair.

3.9 References

3. The use of Ion Mobility Mass Spectrometry to probe Modulation of the Structure of p53 and of MDM2 by Small Molecule Inhibitors

The use of Ion Mobility Mass Spectrometry to probe Modulation of the Structure of p53 and of MDM2 by Small Molecule Inhibitors


The use of Ion Mobility Mass Spectrometry to probe Modulation of the Structure of p53 and of MDM2 by Small Molecule Inhibitors

4
The Contribution of Proline to the Conformational Landscape of the p53 Transcriptional Activation Domain
4.1 Declaration

This chapter consists of one draft publication awaiting submission:


As the first author, I carried out all MS and IM-MS experiments and compiled all figures. I also drafted and edited the manuscript. GC and WB expressed and purified the p53TAD samples and provided the ensemble structures used for Figure 4.4. LGM carried out the computational calculations shown in Figure 4.4, Figure S 18 and Figure S 19. VRB carried out the SAXS measurements used for Figure 4.4.
4. The Contribution of Proline to the Conformational Landscape of the p53 Transcriptional Activation Domain

4.2 Abstract

Intrinsically disordered proteins (IDPs) lack a stable three-dimensional structure but are able to carry out numerous specific functions. Many IDPs undergo disorder-to-order transitions coupled with folding-upon binding, often utilising transient helical secondary structure (PreSMoS). It has been shown that proline residues may be responsible for altering the helical content of IDPs. Here, we study the N-terminal transactivation domain of the tumour suppressor protein p53 (p53TAD) (residues 1-73, mass 8.5 kDa) using ion mobility mass spectrometry (IM-MS) and computational methods. Wild-type p53TAD (WT p53) yields a distinctive bi-nodal conformational signature by IM-MS which is modulated by the selective mutation of prolines for alanine. For all constructs, compact forms (collision cross section ($^{DT}_{CCS_{He}}$) ~900 Å²) are present, but significant differences are seen in the population of more extended states. Mutating the prolines of p53 that flank the Mdm2-binding region results in increased contribution from conformers with $^{DT}_{CCS_{He}}$ ~1430-1600 Å² and reduces the contribution from more extended states ($^{DT}_{CCS_{He}}$ > 1800 Å²). When all prolines are substituted with alanine (p53$_{PallA}$), NMR and circular dichroism (CD) show enhanced helicity and IM-MS yields a conformational family (H-state) at ~1600 Å² which is notably more dominant than in other constructs. We compare our data to that obtained by IM-MS from the similarly sized structured protein Ubiquitin (8.6 kDa) and note an excellent agreement of the p53TAD H-state with the $^{DT}_{CCS_{He}}$ assigned to the helical rich A-state of Ubiquitin (CCS ~1650 Å²). We also compare our IM-MS data with CCS values derived from computational ensembles of structures of WT p53 weighted to both NMR and SAXS data. These results do not access the compact forms of p53TAD seen using IM-MS however map well to the more extended node although IM-MS is able to delineate more discrete conformational sub-states adopted by p53TAD.

4.3 Introduction

Intrinsically disordered proteins (IDPs) are characterised by their lack of stable tertiary structure and are probably best described as a dynamic, heterogeneous ensemble of structures. These dynamic conformational ensembles are necessary for meditating protein-
The Contribution of Proline to the Conformational Landscape of the p53 Transcriptional Activation Domain

Proline residues, the structure of which restricts possible backbone torsion angles [8], are enriched in IDPs [9, 10] and in PreSMoS containing IDPs, are most commonly found flanking these transient α-helical regions [5]. It has been demonstrated that the positioning of proline residues in a polypeptide chain is linked with the helix forming propensity and stability of a given stretch of sequence [11], controlling the level of helicity which in turn may govern the binding mechanism and/or the strength of the interactions between IDPs and their targets by modulating the lifetime of the bound state [5].

Figure 4.1 Composite figure showing a) mass spectrum of WT p53 sprayed in negative nano-electrospray (nESI) mode, b) mass spectrum of p53PallA sprayed in negative nESI mode, c) sequence alignments depicting p53 polypeptide chain and proline replacements made. Prolines are highlighted in yellow. Green residues represent alanine replacements of proline. Residues
The Contribution of Proline to the Conformational Landscape of the p53 Transcriptional Activation Domain

highlighted red represent those involved in the Mdm2 binding interaction. Secondary structure representations for each p53 construct showing residual helicity locations determined by NMR shown for each construct.

The tumour suppressor p53, predicted to be ca. 50% disordered in its monomeric state [12], has been reported to be involved in over 300 protein-protein interactions within the cell [13], many of which involve the N-terminal transactivation domain (p53TAD). Free cellular p53TAD exists in a rapid equilibrium between fully disordered and partially helical forms [2, 14, 15]. The domain contains three PreSMos; one helical (residues 18-26) and two short turns (residues 40-44 and 48-53)[2]. The N-terminal helical PreSMos corresponds to the residues which form a stable α-helix in complex with Mdm2 (residues 19-25). p53TAD is enriched with proline residues, which flank all three PreSMos (Figure 4.1). Mutation of one or more of these prolines causes an increase in the residual helicity levels of the domain[16]. NMR spectroscopy showed that mutation of a single proline P27 to alanine in the 12 amino acid C-terminal flanking region of the Mdm2 binding PreSMos gives rise to an increase in maximum helical content from 28% to 64% [16]. The increase in helicity was observed to occur in the Mdm2-binding region of p53TAD, as illustrated in Figure 4.1c. In contrast, mutation of N-terminal flanking prolines P12 and P13 has little effect on residual helicity and when both N- and C-terminal prolines were replaced there was no further increase in helicity above that observed for P27 alone. In this study, we have added a mutant form of p53TAD where all prolines are replaced with alanine, here circular dichroism (CD) and NMR show a further increase in helicity with particular helix formation in the Mdm2-binding region and at residues 58-71 (Figure 4.1, Figure 4.3c, Appendix C Figure S 10). An increase in the residual helicity following proline substitution at position 27 resulted in an increase in binding affinity for Mdm2, altered p53 dynamics, impaired target gene expression and failure to induce cell cycle arrest upon DNA damage[16].

In this study we use mass spectrometry along with ion mobility separation coupled to mass spectrometry (IM-MS) to study the effect of proline mutation on p53TAD. The hydrophobic environment of the mass spectrometer allows intramolecular interactions to be studied in and environment without contributions from protein-solvent interactions, providing an attractive approach to probe how the proline content of a polypeptide chain and the residual helicity impact upon the conformational landscape of an IDP. Stabilisation of helices in the gas-phase[17, 18] coupled with the favourable correlation between
monomeric IDP charge states and surface area estimates [19] allows conformational changes as a result of altered helicity to be studied. IM-MS is a gas-phase electrophoretic technique. In addition to providing information regarding protein mass, charge and shape, it is uniquely capable of identifying the relative populations of conformational substates adopted by a protein. A packet of ions is injected into a drift cell filled with inert neutral gas. The ions are allowed to drift along a weak electric gradient, and are impeded by buffer gas molecules. Ions with a more compact geometry will undergo fewer collisions with the buffer gas compared with more extended ions which will take longer to reach the end of the drift cell. Drift times are converted into rotationally averaged collision cross section ($^{DT}\text{CCS}_{\text{acc}}; \Omega; \text{Å}^2$), which can be compared with values computed from X-ray crystallography, NMR, and molecular modelling. It has been shown that dynamic systems such as IDPs[18, 20-22], weak complexes[23] and fibrillar systems[24-27] can be probed using IM-MS. Comparisons can be made with data from other techniques can inform on the differences between solution and gas-phase behaviour. Parameters such as radius of gyration ($R_g$) measurements[25, 28], Stoke radii ($R_s$) measurements[29] derived from experimental or computational data sets have been used.

4.4 Methods

4.4.1 Mass Spectrometry and Ion Mobility Mass Spectrometry

WT and mutant versions of human p53 (residues 1-73) were expressed and purified as previously described[14], and stored at -80°C. Prior to analysis, samples were thawed at room temperature, diluted to 50 µM with ammonium acetate and buffer exchanged into a solution of 50 mM ammonium acetate (pH 7.2) using Slide-A-Lyzer dialysis cassettes (Fisher Scientific, UK). Concentrations were confirmed spectroscopically using a Nanodrop 2000 Spectrophotometer.

Circular dichroism (CD) spectra were acquired using a Chirascan CD spectrometer (Applied Photophysics, Ltd.) in the far UV region 190-260 nm with a spectral bandwidth of 1 nm, 0.5 nm step size, 0.1 mm path length and recorded at 5 seconds per point. p53TAD samples were prepared to 1mg/ml in 50 mM ammonium acetate. Data was viewed using Pro Data
The Contribution of Proline to the Conformational Landscape of the p53 Transcriptional Activation Domain

Mass spectrometry and ion mobility mass spectrometry experiments were performed on a quadrupole time-of-flight mass spectrometer (Waters, Manchester, UK) in-house modified to include a 5.1 cm drift cell [30]. Ions were produced by nano-electrospray ionisation (nESI) using a Z-spray source and a spray voltage of -1.31 to -1.62 kV. The drift cell was filled with helium buffer gas with average pressures of 4 Torr. The voltage across the drift cell was varied from 12 V cm\(^{-1}\) to 2 V cm\(^{-1}\), with arrival time measurements taken at a minimum of 5 distinct voltages. Pressure and temperature readings were taken at each voltage and used in the analysis of drift time measurements. The mobility of an ion of interest was obtained from a linear plot of average arrival time vs. pressure/temperature, and the rotationally averaged collision cross section (\(^{0\text{T}}\text{CCS}_{\text{inv}}, \Omega, \text{Å}^2\)) calculated using Equation 1.7:

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

Where \(K_0\) is the reduced mobility, \(z\) is the ion charge, \(e\) is the elementary charge, \(N\) is the gas number density, \(\mu\) is the reduced mass of the ion-neutral pair, \(k_B\) is the Boltzmann constant, and \(T\) is the gas temperature.

The \(^{0\text{T}}\text{CCS}_{\text{inv}}\) distribution plots (CCSDs) are derived from raw arrival time data using Equation 1.9 [31].

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

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 \(t_\text{D}\) 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 \((t_\text{d})\). The value for \(t_\text{d}\) is calculated by
The Contribution of Proline to the Conformational Landscape of the p53 Transcriptional Activation Domain

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 ($t_0$):

Equation 1.8  

\[ t_0 = t_a - t_0 \]

MS and IM-MS data was analysed using MassLynx 4.1 (Waters, Manchester, UK) and Origin 9.1 software. Conformational families were fitted to CCSDs using the Peak Analyser tool in Origin 9.1. The minimum number of conformational families required to satisfy total CCSD spread for all charge states of all p53 constructs were selected; the centroid value for each conformational family was held constant for all charge states. Cumulative peak fits for every charge state were summed to produce a total CCSD for each p53 construct.

Information on the conformational families chosen can be found in Appendix C Table S1. A small amount of CCSD area not encompassed by selected conformers (Appendix C Figure S12 - Figure S16) suggests that there are further unresolved conformational families we are unable to distinguish at our experimental resolution. CCSD Y-axis intensity is proportional to the ATD area for each charge state normalised to the mass spectrum peak intensity for each p53 construct.

4.4.2 Molecular Modelling

4.4.2.1 Ensemble generation

An initial pool of 1 million WT p53 structures was generated using TraDES software [32]; the entire pool was then reweighted using the Broad Ensemble Generation with Reweighting (BEGR) algorithm [33, 34] to create smaller ensembles representing the experimental results in terms of NMR chemical shifts (alpha carbon, CA; beta carbon, CB; and carbonyl carbon, CO) and SAXS data. Five independent ensembles were generated for each experimental setup (NMR or SAXS) containing between 5,000-20,000 structures. A subset of structures was subsequently selected for CCS calculations. For the calculations of random structures, 25,000 structures were randomly selected from the 1M pool of structures.
4.4.2.2 Theoretical collision cross sections

Theoretical collision cross sections were calculated for p53TAD models from SAXS and NMR ensembles using the exact hard sphere model as implemented in EHSSrot software [35, 36]. Structures were selected from a large pool of models based on their BEGR weights, \(R_g\), and SASA. A total of ~85,000 structures were evaluated during CCS calculations.

4.4.2.3 Derivation of CCSs from \(R_g\)

The geometric metric of radius of gyration, which determines the compactness of the molecular system has been shown previously [37] to give excellent linear correlation with CCS. A plot of \(R_g\) versus CCS results in a straight line fit with correlation coefficient between 0.75-0.85. For each of the ensembles (SAXS, NMR and random), a plot of CCS vs \(R_g\) was used to determine the linear regression equations that can be used to scale up \(R_g\) of all structures to their theoretical CCSs.

Further details regarding the use of molecular dynamics, simulated annealing and principle component analysis (PCA) used for Figure S 19 can be found in Appendix C.

4.5 p53 constructs exhibit a distinct conformational phenotype

Negative nESI produces spectra of p53 constructs with a charge state range of \(-4 \leq z \leq -10\) (Figure 4.1a,b, Appendix C Figure S 11). The peak corresponding to the \([M-5H]^{4-}\) charge state is dominant for all constructs, however we see a further distribution of ions in charge states \(z \geq 6\). \(p53^{\text{PallA}}\) (Figure S 11e) exhibits a different distribution of charge states compared with all other constructs, with an intense peak corresponding to ions of charge \([M-4H]^{5+}\) and less ions presenting in charge states \(z \geq 6\). With a net charge of -15 at pH 7 for p53 residues 1-73, we expect to observe high charge states in negative nESI due to preference for proton loss from solvent accessible charged residues during the electrospray
process. Interestingly however, the median charge state is low, and we observe no peaks corresponding to $z \geq -10$. 
Figure 4.2 Cumulative collision cross section distributions (CCSDs) derived from arrival time
distributions for all charge states observed of a) WT p53 b) p53\textsuperscript{P12,13A}, c) p53\textsuperscript{P27A}, d) p53\textsuperscript{P3xA} and e) p53\textsuperscript{PallA} sprayed under negative nESI conditions. X-axis denotes collision cross section (DTCCS\textsubscript{He}, Å\textsuperscript{2}).
The Contribution of Proline to the Conformational Landscape of the p53 Transcriptional Activation Domain

the relative proportion of ions in the annotated conformers are shown and are labelled A-G with their corresponding colour. Conformer centered at \( \sim 880 \text{ Å}^2 \) in red, \( \sim 990 \text{ Å}^2 \) in yellow, \( \sim 1100 \text{ Å}^2 \) in blue, \( \sim 1250 \text{ Å}^2 \) in green, \( \sim 1430 \text{ Å}^2 \) in magenta, \( \sim 1600 \text{ Å}^2 \) in solid orange and \( \sim 1800 \text{ Å}^2 \) in purple. The cumulative curve for all fitted conformers is shown as a red line. Raw data is shown in grey.

Cumulative CCSDs for all charge states of WT p53, p53\(^{P12,13A}\), p53\(^{P27A}\), p53\(^{P3xA}\) and p53\(^{PallA}\) are shown in Figure 4.2a-e and, whilst each construct presents in the same distribution of conformations A-G, significant differences between the populations of these conformers are evident (charge state specific CCSDs are shown in Appendix C Figure S 12 - Figure S 16). An intense, narrow feature, centered at \( \sim 880-990 \text{ Å}^2 \) and comprising conformers A and B, indicating compact states is detected in all the constructs (Figure 4.2). This feature is observed in several of the lower charge states (Appendix C Figure S 12- Figure S 16). With removal of proline, we observe an decrease in the population of conformers E and G accompanied by an increase in the population of conformer F, which becomes dominant when all prolines are removed (Figure 4.2).

WT p53 exhibits a well-defined transition from the compact states A and B into a more extended conformer E, centered at \( \sim 1430 \text{ Å}^2 \), and a further minor population of conformer G centered at \( \sim 1800 \text{ Å}^2 \), observed as a shoulder. The near baseline resolution between compact and extended conformers arises from a sharp transition with the removal of one proton from the [M-5H]\(^5\) charge state to the [M-6H]\(^6\) (Appendix C Figure S 12) and the shoulder is as a result of a further extension of the protein between the [M-8H]\(^8\) and [M-9H]\(^9\) charge states.

p53\(^{P12,13A}\) ions present with a greater intensity in the intermediate (C and D) and extended (E and F) conformational families compared with the WT protein (Figure 4.2b cf. Figure 4.2a). This is as a result of a smooth increase in \( \text{DTCCS}_{\text{He}} \) with charge state (Appendix C Figure S 13). Interestingly, a significant increase in intensity of conformer F centered at \( \sim 1600 \text{ Å}^2 \) is observed (coloured in orange Figure 4.2), in contrast with its minimal contribution to the cumulative CCSD for WT p53.

Figure 4.2c shows the cumulative CCSD for p53\(^{P27A}\), shown to possess regions of increased residual helicity (63% compared with WT p53 (28%) [16]). The transition from compact to extended conformers is similar to p53\(^{P12,13A}\), with a smooth charge state dependent extension (Appendix C Figure S 14). Intermediate conformers D and E (centered at \( \sim 1250 \text{ Å}^2 \)
and ~1430 Å²) are observed as distinct shoulders at a lower intensity compared with WT p53 and a significant proportion of ions present in conformer F at ~1600 Å². This conformer is the dominant species over three charge states (Appendix C Figure S14) in contrast with WT p53, where it is observed at a low intensity in the [M-8H]⁶⁺ charge state only. Unlike WT p53, p53P27A presents with very little intensity in the extended conformer G.

p53P3xA, with proline replacement both N-terminal (P12, P13) and C-terminal (P27) to the Mdm2 binding site has been shown to exhibit an increased helicity of 66%[16] in the Mdm2-binding region; significantly higher that WT p53 (28%), but similar to that of single substitution mutant p53P27A (Figure 4.1c) [16]. To this end, we could assume that the conformational spread of p53P3xA may be similar to that of the sum of p53P12,13A and p53P27A. Indeed, we observe conformers C, D and E at significant intensity, with little distinction between. Interestingly, the protein retains compact conformers A and B at high charge states (Appendix C Figure S15), in contrast to all other constructs. We observe a similar proportion of ions in conformer F compared with p53P12,13A and p53P27A, and a greater proportion compared with WT p53. There is very little intensity in the most extended conformer G, suggesting this construct is more restricted by stabilising non covalent interactions than WT p53.

p53Pala, with all Proline residues mutated to Alanine, presents with the same charge state range as all p53 constructs but an altered charge state distribution, with an increase for the [M-4H]⁴⁺ ions (Figure 4.1b). The cumulative CCSD for p53Pala (Figure 4.2e) exhibits significant alteration in comparison to WT p53. A dominant narrow distribution corresponding to the conformational family F (centered at ~1600 Å²) is observed at a significantly increased intensity compared with WT p53 (Figure 4.2a vs Figure 4.2e). It appears the proportion of ions in conformer F is correlated with the mutation of prolines, and is present over four charge states of p53Pala, suggesting it is a stable structural state lying within a local energy minima.

4.6 Can IM-MS delineate the increased helical content?

We observe a significant increase in the proportion of conformer F (centered at 1600 Å²) coupled with a decrease in the populations of conformers E and G with increased helicity of
The Contribution of Proline to the Conformational Landscape of the p53 Transcriptional Activation Domain

the p53 constructs (Figure 4.2, Figure 4.3c). NMR and CD experiments report increased helicity in p53\textsuperscript{PallA} compared with other constructs; the CD spectrum shows a typical helical signature, with features at 215-230 nm (Figure 4.3a). The NMR data also shows significant residual helicity along the backbone of p53\textsuperscript{PallA} (Appendix C Figure S 10). Together, these results suggest that conformer F is mainly helical (denoted p53 H-state). It is instructive to compare p53\textsuperscript{PallA} to Ubiquitin, which has a very similar mass (bovine Ubiquitin 8565 Da, p53\textsuperscript{PallA} 8194 Da). Ubiquitin is an extensively studied model system, with several assigned, well defined conformational states; including a folded, native state (N-state), a partially folded elongated state with a high propensity for helical structure (A-state), and an unfolded state (U-state) [38-42]. The A-state of Ubiquitin is most prevalent in acidic solution conditions containing methanol.

Extensive work using mass spectrometry, IM-MS and molecular modelling has been carried out to link the gas-phase conformers displayed by Ubiquitin to solution phase structures. Bowers et al. [43] and Clemmer et al. [44, 45] have shown the A-state of Ubiquitin presents as several conformational families with distinctive sharp features at $^{DT\text{CCS}}_{\text{He}} \sim 1650-1700$ Å\textsuperscript{2} for the [M+8H]\textsuperscript{8+} (Figure 4.3d) [44] and [M+9H]\textsuperscript{9+} [45] charge states and correlated these conformers with calculated models of the A-state (1640 Å\textsuperscript{2}) [44] (Figure 4.3d, inset).

Recently, a second A-state of Ubiquitin, the A’-state, was noted using IM-MS at higher charge states. This is also mainly helical in structure and presents over $^{DT\text{CCS}}_{\text{He}}$ values 1525 – 1740 Å\textsuperscript{2} [45]. Additional work using IM-MS has shown the hydrophobic environment of the mass spectrometer to have a stabilising effect on $\alpha$-helical structures [17, 18, 43, 46-48]. As such, we compare the experimental $^{DT\text{CCS}}_{\text{He}}$ values for [M+8H]\textsuperscript{8+} Ubiquitin sprayed from 70:30 water:methanol conditions with the CCSD for p53\textsuperscript{PallA}. The dominant extended conformer F of p53\textsuperscript{PallA} (Figure 4.3b) correlates well with all conformers that have been assigned to the A-state of Ubiquitin [44] (Figure 4.3d), shown in the green lines and highlighted by the green box. This experimental correlation further supports the NMR and CD results which suggest that the p53TAD distribution F centered at $\sim 1600$ Å\textsuperscript{2} is a mostly helical conformer.
4. The Contribution of Proline to the Conformational Landscape of the p53 Transcriptional Activation Domain

Figure 4.3 a) circular dichroism spectra showing each p53TAD construct: WT p53 in black, p53P12,13A in red, p53P27A in blue, p53P3xA in magenta, p53PallA in green. X-axis denotes wavelength (nm), b) cumulative CCSD for p53PallA depicting conformational family at ~1600 Å² highlighted in orange. Figure modified from Figure 2e, c) plot showing the relative area of each fitted conformational family from Figure 4.2, showing alterations between WT and mutant p53TAD constructs. Each conformational family is shown in its corresponding colour, and its contribution to the total CCSD shown for (from left to right) WT, p53P12,13A, p53P27A, p53P3xA and p53PallA, respectively. d) CCSD for [M+8H]⁺ charge state of Ubiquitin sprayed from 70:30 water:methanol, with conformers attributed to α-state of Ubiquitin plotted as green lines and highlighted with green box. Figure adapted from [44]. Structure of the proposed Ubiquitin A-state shown as inset, figure adapted from [43].

4.7 Comparisons to computational data

A large pool of 1M structures of WT p53 was generated and assessed in order to represent the NMR chemical shifts (α-carbon (CA), β-carbon (CB) and carbonyl carbon (CO)) and SAXS measurements. Each ensemble was used for CCS calculations resulting in three separate
The Contribution of Proline to the Conformational Landscape of the p53 Transcriptional Activation Domain

data sets (NMR weighted, SAXS weighted and random). Figure 4.4 shows the computational CCS ($^{\text{COMP}}$CCS) results. All CCSDs, from both NMR and SAXS-based structures show $^{\text{COMP}}$CCS values which overlay well with the intermediate and extended conformers measured experimentally, however we observe no overlap with the compact states of the WT p53. The randomly selected structures (Random, Figure 4.4a), showed the widest CCS distribution ranging from 1135 Å² to 2223 Å², with an average of 1670 Å². The models used to represent the SAXS data (Figure 4.4b) were larger, with CCSD ranging from 1430 to 2224 Å², with an average of 1830 Å². The ensemble used to represent the NMR results (Figure 4.4c) was found to have CCSD ranging from 1218 to 2014 Å² with an average centered at 1580 Å², lower than both SAXS weighted and random ensembles. This CCSD fits most accurately with the second node of the experimental CCSD, with highest population at 1450 Å² and a lower population of structures in the more extend conformers up to 2000 Å². The NMR ensemble is more likely to give a more accurate representation of the secondary structure elements than the random or SAXS ensemble, since the selected structures were found to have appropriate CA, CB and CO chemical shifts, whereas the SAXS ensemble will have more accurate representation of the global structural properties (i.e. $R_g$ distribution) but poorer local secondary structure elements.
The Contribution of Proline to the Conformational Landscape of the p53 Transcriptional Activation Domain

Figure 4.4 Experimental and computational CCSDs for WT p53: a) CCSD from ensemble containing randomly selected structures, b) CCSD from SAXS weighted ensemble, c) CCSD from ensemble weighted to NMR chemical shifts (alpha carbons). CCSDs for computational models were obtained.
The Contribution of Proline to the Conformational Landscape of the p53 Transcriptional Activation Domain

by converting $R_g$ values to CCSs using equations of best fit as described in the experimental methods. Correlation graphs can be found in Appendix C Figure S 18. Secondary structure (SS) composition was calculated using the DSSP algorithm for each structure. Structures were binned according to their $R_g$ and the average secondary structure content for each bin was determined. The SS was then normalised to the size of the bin and CCS distribution. Elements representing components with secondary structure (β-strands, α-helix, 3_10-helix, π-helix, turn and bend) were summed together and reported as ‘structured’ (grey) whereas unstructured elements (none) as ‘disordered’ (cyan). d) Experimental cumulative CCSD derived from arrival time distributions for all charge states observed of WT p53 and example structures taken from a range of CCS values. Orange distribution denotes conformational family centered at 1600 Å². Figure reproduced from Figure 2a. X-axis denotes collision cross section, either derived experimentally ($^{DT}\text{CCS}_{He}$, Å²) or computationally ($^{EINS}\text{CCS}_{He}$, Å²)

We observe that the more extended node of our experimental IM-MS results correlates well with computationally calculated structures of WT p53 and can infer that these intermediate and extended conformations of WT p53 are found in solution, with the models of higher secondary structural content from the NMR weighted ensemble mapping well with the postulated H-state. It must be noted that the approach used to convert $R_g$ measurements from SAXS data into CCS values is not ideal since the $R^2$ of CCS vs $R_g$ plots (Appendix C Figure S 18) is significantly less than 1, however it enables a quick assessment of large pool of structures for their prospective CCSs and generation of CCSD plots for the entire SAXS, NMR or random ensemble to be used as a comparison with IM-MS experimental data. Neither random, SAXS weighted nor NMR weighted structures overlay with the compact conformers A and B observed during our IM-MS experiment (Figure 4.4). Interestingly, we observe these compact conformations at high intensity (Figure 4.2), particularly at low charge states (Appendix C Figure S 12). It may be that these conformers of WT p53, and perhaps mutant p53, are kinetically trapped during our experiments or that the protein has formed intramolecular interactions in the absence of solvent in order to ‘self-solvate’, forming very compact states as has been previously observed for p53[49], and suggested to occur for all IDPs[28]. To identify the extent of compaction, we calculated the smallest theoretical CCS ($^{CALC}\text{CCS}$) for each p53 sequence based on the average density of native-like, globular proteins in the protein databank (http://www.rcsb.org/) [50-52]. We also calculated the most extended CCS possible, based on a fully extended polypeptide chain ignoring a-typical backbone contributions from proline residues. The $^{CALC}\text{CCS}$ range
from the constant-density estimate and cylindrical-chain estimate spans almost all of the WT p53 IM-MS results, ranging from ~550 Å² to ~2500 Å² (Appendix C Figure S 19), lacking only the most compact and most extended theoretical CCS values. We suggest that the inclusion of experimental compact conformers A and B within the theoretical possible range, along with studies showing IDPs adopt compact states[9], supports the biological relevance for compact states of the protein, although they could exist in a molten globule-like, self-solvated structure. We have attempted to produce computational structures which reside in these compact states using a series of methods including molecular dynamics (MD), simulated annealing and MD with enhanced sampling techniques utilising the principle component analysis. Structures produced using a number of these computational methods have EHSS-CCS well-matched with the compact conformer centered at ~990 Å² (Appendix C Figure S 19) and an example structure can be seen in Figure 4.4d showing retention of secondary structure. These calculations demonstrate that IM-MS allows for excellent delineation of a wide range of conformations sampled by WT p53, including those not observed using NMR and SAXS measurements, but with biological and structural relevance.

In future work, computational results from ensembles of structures calculated for p53PallA will be analysed to determine if computational structures are able to delineate the alterations to the populations of conformations observed compared with WT p53, as can be observed using IM-MS.

4.8 Conclusions

A range of biophysical and computational techniques have been employed in this study to track the conformational changes associated with changes in proline-dependent helicity of the transactivation domain of p53. Although nESI mass spectra for each p53 construct were observed with a similar charge state range and distribution, we observe significant differences in the conformational spread of the protein with proline removal using IM-MS. We note a reduction in the extension of the protein, a smoother increase in DTCCSHe with charge and an increase in the intensity of intermediate conformers, in particular a conformational family of DTCCSHe ~1600 Å², as we replace prolines flanking the Mdm2 binding site. p53PallA, with removal of all proline residues, produces a highly helical
The Contribution of Proline to the Conformational Landscape of the p53 Transcriptional Activation Domain

construct, as shown by both NMR and CD. We observe this using IM-MS as a well-defined, narrow Gaussian distribution of ions corresponding to the ~1600 Å² conformational family. We compare this H-state of p53 with the A-state of Ubiquitin, a well-studied protein of near identical size, and found an excellent agreement. All p53TAD constructs present with highly populated compact conformers in our gas-phase experiments. Computational CCS calculations were performed on structures of WT p53 from ensembles of structures that were either randomly selected, weighted to SAXS Rg measurements, and weighted to NMR chemical shifts and were found to correlate excellently with the intermediate and extended conformations determined experimentally, however less so with the compact states of the protein. We conclude that these conformers can be compared to solution-phase data. Whilst it has been shown that IDPs undergo a degree of collapse in the gas-phase[49, 53], we note that the compact conformers observed with IM-MS lie within the calculated theoretical range of a protein of average density, suggesting that some structural elements are likely to remain. This suggests that these conformations may be kinetically trapped in the electrospray process, although structures of similar CCS can be calculated using numerous molecular dynamics methods. IM-MS provides a powerful and informative view into the conformational populations presented by a highly disordered protein domain and can delineate conformational changes that arise as a result of a single residue mutation.

4.9 References

4. The Contribution of Proline to the Conformational Landscape of the p53 Transcriptional Activation Domain


The Contribution of Proline to the Conformational Landscape of the p53 Transcriptional Activation Domain


43. Wytenbach, T. and M.T. Bowers, Structural Stability from Solution to the Gas Phase: Native Solution Structure of Ubiquitin Survives Analysis in a Solvent-Free Ion
The Contribution of Proline to the Conformational Landscape of the p53 Transcriptional Activation Domain


Proline residues mediate promiscuous binding of p53 to Mdm2
5.1 Declaration


As a first author on this publication I carried out all experiments and compiled the data for all figures. I also drafted and edited the manuscript. GD and WB expressed and purified the p53TAD and Mdm2 samples.
5.2 Abstract

Knowledge of protein-protein interactions, the structure and dynamics of the complexes they form and the active binding conformation of the individual components is of paramount importance in understanding intercellular function. Many protein systems are dynamic with respect to both conformation and interaction, so in the course of a protein’s ‘lifespan’ it will sample many configurations and bind to multiple targets, which is particularly true for intrinsically disordered proteins (IDPs). The transactivation domain of the IDP p53 (p53TAD) undergoes a folding-upon-binding event, forming a stable helical structure in the presence of its most important interaction partner, the E3 Ubiquitin ligase murine double minute-2 (Mdm2). Regulation of this disorder-to-order transition is thought to be carried out on a primary structure level; proline has been shown to mediate the residual helicity of the transactivation domain of p53, and its mutation implicated in alterations to both Mdm2-binding affinity and downstream gene signalling patterns. Here we use mass spectrometry (MS) in tandem with ion mobility mass spectrometry (IM-MS) to interrogate the interaction between these two highly disordered proteins. To investigate this, we use a series of mutants in which proline residues flanking the Mdm2-binding interface are replaced with alanine. We show that the p53TAD mutants display altered conformational landscapes. We investigate the impact of these p53TAD mutations on the binding stoichiometry and affinity for N-terminal Mdm2 (N-Mdm2). We then probe structure of the dynamic complex formed by p53TAD and N-Mdm2 and track the conformations involved in binding, by delineating conformational differences between WT and mutant p53 when in the presence of, but not bound to, N-Mdm2.

5.3 Introduction

The interaction between p53 and Mdm2 represents perhaps the best-studied relationship between a tumour suppressor gene which acts as a transcription factor and an oncogene, which functions primarily as an E3 protein ligase [1]. Mdm2 is the primary cell regulator of p53, downregulating its activity in normal cells via a negative feedback loop [2]. In addition to directly blocking the transcriptional activity of p53, Mdm2 also exports p53 from the nucleus and tags it via ubiquitination for degradation by the 20S proteasome [3-5]. p53 and
Proline residues mediate promiscuous binding of p53 to Mdm2

Mdm2 interact via their respective N-terminal domains, forming a stable complex in which p53 inserts an α-helix into a hydrophobic binding pocket on the surface of the N-terminal domain of Mdm2 (N-Mdm2) [6]. This intimate relationship is tightly regulated by an array of post-translational modifications, which in turn dictate the activity and stability of both p53 and Mdm2.

The N-terminus of p53 consists of an acidic transcription activation domain (TAD) and a proline-rich domain. The TAD is often further subdivided into ill-defined subdomains TAD1 (residues 1-42) and TAD2 (residues 43-63) [7, 8]. The p53-TAD is a promiscuous binding site for a multitude of interaction partners including negative regulators Mdm2/Mdm4, which trigger ubiquitination and ultimately degradation of p53 [6, 9, 10], modules of the transcriptional machinery [11-13], and the transcriptional coactivators p300/CBP [14, 15]. In solution, p53-TAD exists in equilibrium between intrinsically disordered and partially helical conformations [16-20], although residual secondary structure is observed in three conserved regions of residual helicity [16]. These regions of residual helicity in an otherwise disordered domain are called pre-structured motifs (PreSMoS), the second of which (residues 18-26) forms the interaction between p53 and Mdm2. To date, the only high resolution structures of p53-TAD that have been solved are those of p53 in complex with its binding partners, where the p53 fragments used form an α-helix [6, 13, 21-25]. The interaction between p53 and Mdm2 is best understood though a crystal structure of N-Mdm2 (residues 25-109) bound to 15-residue peptide of p53-TAD (residues 15-29) by Kussie et al. [6] (Figure 5.1). This structure shows most of the Mdm2-binding PreSMoS of p53 (residues 18-26) folded into a stable α-helix and inserted into a hydrophobic binding pocket on the surface of Mdm2, with residues Phe19, Trp23, and Leu26 providing the bulk of the interaction energy. As yet, there is little information regarding the transition of p53 from disordered with some residual helicity to a rigid stable α-helical structure and the mechanisms by which this structural change is controlled in the cell.

It is thought that intrinsic disorder in proteins provides a mechanism by which multiple, highly specific interactions between an IDP and its diverse target proteins can be made using overlapping binding sites; by forming multiple structural isomers each able to bind to a specific partner. The evolutionary advantages of these structural isomers, and the disorder-to-order transitions which form them, lie in their ability to maintain the binding of a single protein to many partners via a one-to-many binding mechanism allowing many IDPs, including p53, to operate as “hub” proteins.
Proline residues mediate promiscuous binding of p53 to Mdm2

Figure 5.1 Composite figure showing side and top views of X-ray crystal structure of human p53TAD (residues 15-29) in complex with N-Mdm2 (PDB; 1YCR [6]). p53TAD shown as a magenta helix with grey turns, N-Mdm2 shown as a surface with charged residues denoted by blue and red. Residues Phe19, Trp23 and Leu25 of p53 which insert into the binding cleft are shown and labelled. Protein sequences used in experiments are shown for p53 (top, residues 1-73) and MDM2 (bottom, residues 17-125). Underlined sequences represent the residues present in x-ray crystal structure depicted. p53 residues involved in the MDM2 interaction are highlighted in red. Residues incorporating the MDM2 disordered lid region are shown in grey.

In the last fifteen years mass spectrometry (MS) in accord with ion mobility mass spectrometry (IM-MS) has established itself as a method capable of providing unique insights into the structure and dynamics of proteins and their complexes [26-30]. IM-MS is
Proline residues mediate promiscuous binding of p53 to Mdm2

a gas-phase technique in which ions are separated on the basis of their charge, mass and shape. IM-MS gives a coarse-grained view of the conformational populations presented by a protein in the form of a rotationally averaged collision cross section (CCS, Å²). The solvent free environment of a mass spectrometer is ideally suited to the study of intrinsic interactions and how they contribute to structure. As such, IM-MS is an attractive tool for the elucidation of IDP structure; capturing detail on the conformational range of a macromolecule at a single conformer level, rather than the averaged ensemble of conformations yielded by alternative biophysical techniques. This is highly informative when assigning conformers to dynamic proteins, and can be used as a diagnostic tool whereby the conformations of a protein involved in a binding event can be tracked through observation of conformer-specific depletion from the unbound population. Alternatively, conformational changes occurring in the presence of a protein-protein interaction partner can also be observed. Although IM-MS does not provide the resolution of some traditional biophysical techniques, the CCS of molecular ions delivered by IM-MS allows comparisons to be made with NMR, X-ray crystallography and molecular dynamics coordinates.

Upon binding to partner proteins, regions of p53TAD nascent helical structure that are present in the native state fold entirely and rigidify [18]. Studies have suggested that the positioning of proline residues within a polypeptide chain may be partially responsible for regulating disorder-to-order transitions and the functional disorder of IDPs as a whole; allowing IDPs to form highly specific but low affinity interactions by mediating the level and location of residual structure [31, 32]. p53TAD contains a number of highly conserved prolines, with several prolines flanking its three PreSMoS; in particular the Mdm2-binding PreSMoS has two N-terminal (P12, P13) and one C-terminal (P27) flanking prolines. It has been shown using NMR that replacement of these prolines with alanine has a direct impact on the level of p53TAD residual helicity [20, 33]; replacement of P27 (p53\(^{P27A}\)) results in a significant increase in residual helicity of the Mdm2-binding PreSMoS from a maximum of 28% to 64% [20]. In addition to this, p53\(^{P27A}\) exhibits a greater binding affinity for Mdm2, and induces downstream genes with an altered amplitude and frequency [20]. Interestingly, the replacement of P12 and P13 (p53\(^{P12,13A}\)) has only a mild effect on the residual helicity of p53TAD, and the addition of these two mutations to the single P27 mutation (p53\(^{P27A}\)), does not increase the helicity beyond that for P27 alone (66%) [20].

More recently, we have studied these WT and mutant p53TAD constructs using MS and IM-MS and showed that all p53TAD constructs present in multiple conformations in the gas phase with distinct signatures. Upon proline replacement the extended forms of p53TAD
Proline residues mediate promiscuous binding of p53 to Mdm2

become less populated, with intermediate conformations becoming more dominant. p53\textsuperscript{P27A}, where all prolines are replaced with alanine was shown by circular dichroism and NMR to exhibit a large increase in residual helicity compared with the WT protein and other mutants. This result was accompanied by observation of a dominant intermediate conformer we postulated to be mainly helical.

In this study, we report mass spectrometry (MS) and ion mobility mass spectrometry (IM-MS) findings on the binding of WT and mutant p53TAD constructs to N-Mdm2. We determine if the noted increase in residual helicity of mutant p53TAD constructs results in a comparable increase in binding affinity for Mdm2, as reported for p53\textsuperscript{P27A} [20]. The conformational diversity of the p53-Mdm2 complexes is interrogated to determine if proline is implicated in the final complex structure. We also track the changes to the conformational spread of monomeric WT and mutant p53TAD when incubated with Mdm2 at increasing concentration. Here we use IM-MS as a diagnostic tool, tracking the conformations of p53TAD when incubated with, but not bound to, increasing concentrations of N-Mdm2, to determine an active conformation of p53TAD required for binding to N-Mdm2 or if conformational changes in p53TAD occurring during the binding event.

5.4 Materials and Methods

5.4.1 Protein preparation

Both the WT and mutant p53TAD constructs and the N-Mdm2 protein was kindly supplied by Dr. Gary Daughdrill (University of South Florida, USA) and their purification has been previously described [18]. The amino acid sequences of proteins are shown below:

p53TAD (residues 1-73)

\[
\text{MEEPQSDPSV EPPLSQETFS DLWKLLPENN VLSPLPSQAM DDLMLSPDDI EQWFTEDPGP DEAPRMPEAA PPV}
\]

Mdm2-N (residues 17-125)
Proline residues mediate promiscuous binding of p53 to Mdm2

\[
\text{SQIP ASEQETLVRP KPLLKLLKS VGAQKDTYT}
\]

\[
\text{KEVLFYLGQY IMTKRLYDEK QQHVYCSND LLGDLFGVPS FSVKEHRKIY}
\]

\[
\text{TMIYRLNVV NOQESSDSGT SVSE}
\]

p53TAD and N-Mdm2 samples were thawed and dialysed into 50mM ammonium acetate overnight at 4°C using Slide-A-Lyzer dialysis cassettes (Fischer Scientific, UK). Protein concentrations were confirmed post dialysis spectrophotometrically (NanoDrop Spectrophotometer ND 1000 Thermo Scientific, USA).

5.4.2 Mass Spectrometry and Ion Mobility Mass Spectrometry

Drift tube (DT) IM-MS experiments were performed on an in-house modified quadrupole time-of-flight mass spectrometer, the MoQToF [34]. Ions were produced by positive nano-electrospray (nESI) using Z-spray source, within a spray voltage range 1.32 to 1.55 kV and a source temperature of 80°C. nESI glass capillaries were prepared in-house from thin-walled borosilicate capillaries (Precision Instruments, Stevenage, UK) using a Flaming/Brown micropipette puller Model P-97 (Sutter Instrument Company, Novato, CA, USA). The drift cell was filled with helium buffer gas at the average pressure of 3.8 - 4.2 Torr (5.1 – 5.6 mBar) and a temperature of 295 - 302 K. The electric potential across the cell was varied from 60 to 15 V (12 to 2 Vcm⁻¹) and arrival time distributions (ATDs) recorded at eight drift voltages with three analytical repeats taken.

The CCS distribution plots are derived from raw arrival time data at drift voltage 35 V using Equation 1.9 below [35]:

\[
\Omega_{\text{avg}} = \frac{1}{16} \left( \frac{1}{m_b} + \frac{1}{m} \right)^{1/2} \frac{ze}{(K_bT)^{1/2}} \rho \frac{1}{L^2} \frac{1}{t_dV}
\]

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
5.
Proline residues mediate promiscuous binding of p53 to Mdm2

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 ($t_0$):

Equation 1.8

\[ t_0 = t_a - t_0 \]

MS and IM-MS data was analysed using MassLynx 4.1 (Waters, Manchester, UK) and Origin
9.1 software. Conformational families were fitted to CCSDs using the Peak Analyser tool in
Origin 9.1. The minimum number of conformational families required to satisfy total CCSD
spread for all charge states of all p53 constructs were selected; the centroid value for each
conformational family was held constant for all charge states.

For the summed CCSD in Appendix D Figure S 24, cumulative peak fits of each charge state
of the complex were summed to produce a total CCSD for each p53 construct. Information
on the conformational families chosen can be found in Appendix D Table S 2. A small
amount of CCSD area not encompassed by selected conformers (Figure 5.2a (inset),
Appendix D Figure S 21 and Figure S 22) suggests that there are further unresolved
conformational families we are unable to distinguish at our experimental resolution. CCSD
Y-axis intensity is proportional to the ATD area for each charge state normalised to the
mass spectrum peak intensity for each p53 construct.

The degree of binding for WT and mutant p53 with N-Mdm2 were calculated (Figure 5.3c).
Total complex intensity was calculated by summing all peaks corresponding to the complex.
The same was carried out for free p53 protein. The fraction of total protein in complex was
calculated using Equation 5.1:

Equation 5.1

\[ \text{Fraction of Complex ion intensity} = \frac{\Sigma \text{PL}}{\Sigma \text{P} + \Sigma \text{PL}} \]

Where $\Sigma \text{P}$ is the total intensity of free p53TAD, and $\Sigma \text{PL}$ is total intensity of the complex.

This was then plotted against the concentration of ligand (in this case N-Mdm2 added). As
the peaks corresponding to the p53:Mdm2 complex and low charge states of p53$^{\text{P12,13A}}$ are
wide due to retention of buffer salt ions, reducing the intensity of the complex (Appendix D
Figure S 23a), we define $\Sigma \text{P}$ and $\Sigma \text{PL}$ as the summed intensity of all protein and protein+salt...
Proline residues mediate promiscuous binding of p53 to Mdm2

peaks present. In calculating these values, we assume that the response factor between p53TAD constructs is the same.

5.5 Increase in residual helicity of p53 results in a change in conformation

nESI mass spectra of WT and mutant p53TAD constructs are shown in Figure 5.2a and Appendix D Figure S 20. For WT p53 (residues 1-73, calculated $M_r = 8558.45$ Da, observed $M_r = 8559$ Da), a monomeric charge state distribution is observed with $4 \leq z \leq 9$ (Figure 5.2a) and the most abundant species corresponding to the monomeric [M+5H]$^{5+}$. This is indicative of a compact protein with a limited number of solvent-accessible residues available for protonation. This charge state range and general charge state distribution is observed for all mutant p53TAD constructs (Appendix D Figure S 20), with some slight intensity differences. p53PallA shows most difference in comparison to WT p53; with an intense peak corresponding to the [M+4H]$^{4+}$ charge state, and a reduction of the populations of ions presenting in the [M+6H]$^{6+}$ charge state compared with the other p53TAD constructs.

Using IM-MS, we have previously shown the collision cross section distributions (CCSD, Å$^2$) of each charge state of WT p53TAD sprayed in negative electrospray mode (Chapter 4). These CCSDs are derived directly from arrival time distributions (ATDs) and allow the conformational spread of the protein to be observed for each individual charge state. In this study, we focus on the interaction of p53TAD with N-Mdm2, a basic protein. As such, we have used positive electrospray ionization mode for all of our experiments. We observe some alterations to the CCSDs for WT and mutant p53TAD constructs between ionization modes, which we attribute to the presence of multiple acidic residues in the p53TAD sequence. WT p53 [M+5H]$^{5+}$ (Figure 5.2a, inset) presents in a dominant compact conformational family centered at ~990 Å$^2$ (shown in yellow), with less intense contributions from families centered at ~880 Å$^2$ (red) and ~1100 Å$^2$ (blue). With the addition of a proton to form the [M+6H]$^{6+}$ charge state, we observe an altered CCSD incorporating a greater proportion of intermediate conformational families centered at ~1100 Å$^2$ (blue) and ~1250 Å$^2$ (green). This extension of the protein continues smoothly with increase in charge, with the ions presenting with charge [M+9H]$^{9+}$ exhibiting
5.
Proline residues mediate promiscuous binding of p53 to Mdm2

conformers ranging from $\sim 1250 \text{ Å}^2$ to $\sim 1800 \text{ Å}^2$.\(^1\) (Appendix D Figure S 21) The CCSDs for mutant p53TAD constructs can be found in Appendix D Figure S 22.

![Figure 5.2 a) nESI mass spectrum of p53TAD (residues 1-73) sprayed from 50 mM ammonium acetate. Inset; CCSDs for intense species $[M+5H]^{5+}$ and $[M+6H]^{6+}$ with conformers fitted. Conformer at $\sim 880 \text{ Å}^2$ in red, $\sim 990 \text{ Å}^2$ in yellow, $\sim 1100 \text{ Å}^2$ in blue and $\sim 1250 \text{ Å}^2$ in green. b) nESI mass spectrum of N-Mdm2 sprayed from 50 mM ammonium acetate pH 6.8. c) Waterfall plot showing the collision cross section distributions (CCSDs) derived from arrival time distributions (ATDs) for each charge state of N-Mdm2 recorded at drift voltage 35 V. x-, y- and z-axes denote collision cross section (CCS, \(\text{Å}^2\)), relative intensity, and charge state respectively. Several charge states are labelled.]

Similarly to p53, the N-terminus of Mdm2 (N-Mdm2) also belongs to a family of IDPs. The NMR experiments performed by Uhrinova et al.\(^{[36]}\) revealed multiple conformations and an unstable nature of ligand-free N-Mdm2 (residues 2-118). The mass spectrum obtained for N-terminal Mdm2 sprayed from near native conditions (50 mM ammonium acetate) is illustrated in Figure 5.2b. It displays a wide monomeric charge state distribution ($7 \leq z \leq$
Proline residues mediate promiscuous binding of p53 to Mdm2

14); characteristic of IDPs. The central spectral peak can be assigned to charge state [M+7H]+, evidence of a predominant compact structure of the protein, but another distribution centered at [M+9H]+ suggests more extended states of N-Mdm2 are also preserved. CCSDs derived from ATDs for N-Mdm2 are shown in Figure 5.2c. A single conformer can be observed at ~1500 Å^2 for low charge states [M+7H]^+ and [M+8H]^+. A widening of the CCSD at charge states 9 ≤ z ≤ 11, with a tailing effect at a larger CCS of ~2100 Å^2, indicates a second conformer. Indeed, at high charge states 12 ≤ z ≤ 14 we observe at least two conformers at CCS ~1700 Å^2 and ~2200 Å^2.

5.6 p53 residual helicity correlates with Mdm2 binding affinity

In the study of p53-Mdm2 complex formation, WT and all mutant p53TAD constructs were found to bind N-Mdm2 as 1:1 complex only (Figure 5.3a,b, Appendix D Figure S 23). Proteins were incubated up to one hour in varying ratios (p53TAD:N-Mdm2 10:1, 5:1, 2:1, 1:1); an increase in the incubation time did not alter the binding strength (data not shown) indicating an instant binding mechanism. 1:1 p53TAD:N-Mdm2 was found to produce the strongest complex signal for all p53 constructs. The complex was observed as peaks corresponding to the [p53:Mdm2+8H]^+ and [p53:Mdm2+9H]^+, although for p53^PallA, [p53:Mdm2+7H]^+ was also observed (Appendix D Figure S 23c). Ion envelopes of unbound monomeric ensembles for p53TAD and N-Mdm2 were also detected, which may be as a result of the transient and dynamic character of the interaction between the two proteins, ensuring a low-affinity but high specificity binding mode. We observe a reduction in the intensity of the p53TAD [M+5H]^+ and in particular the [M+6H]^+ charge states with increasing ratios of N-Mdm2.
Proline residues mediate promiscuous binding of p53 to Mdm2

Proline residues mediate promiscuous binding of p53 to Mdm2

Mutant p53TAD constructs display markedly different binding affinities for N-Mdm2 compared with the WT protein. p53

P27A (Figure 5.3b, Figure 5.3c, red line) displays a significant increase in binding affinity for N-Mdm2. Figure 5.3c shows proportion of p53

P27A :N-Mdm2 complex as a function of N-Mdm2 concentration, and shows that the gas-phase \( K_d \) for p53

P27A (red) is lower than for WT p53 (black). We observe a significant proportion of complex at low N-Mdm2 concentrations, and at 1:1 ratio almost the entire sample is in complex. This result correlates with that observed by Borcherds, et al. [20] who observed by ITC a significant increase in p53 binding affinity for N-Mdm2 with replacement of Pro27.

p53

P12,13A (Figure 5.3c, blue line) also shows an increased affinity for MDM2 in comparison with WT p53, although proportion of the bound complex remains less than that observed for p53

P27A . p53

P3xA exhibits greater affinity in comparison to WT p53 (Figure 5.3c, magenta line), although a lower affinity in comparison with p53

P27A ; no complex is observed at 10:1 ratio and we observe very little increase in the proportion of complex at ratios above 5:1. This suggests that the replacement of P12 and P13 in addition to P27, whilst having little effect on residual helicity [20], negatively impacts upon N-Mdm2 binding affinity. This result is in contrast to ITC data which showed a slight decrease in the \( K_d \) [20]. p53

PallA , shown in Chapter 4 to be mainly helical by NMR and circular dichroism (CD), exhibits an increased binding affinity for N-Mdm2, similar to that of p53

P27A (Figure 5.3c, green line). Interestingly, the complex produced by p53

PallA :N-Mdm2 presents in an altered charge state range, with the appearance of a peak corresponding to [p53:Mdm2+7H]+ and a reduction in the relative intensity of the [p53:Mdm2+9H]+ compared with all other constructs (Appendix D Figure S 23c).

The removal of Prolines flanking the Mdm2 binding site has been shown to increase the residual helicity of p53TAD [20], however MS data shows that the level of p53TAD residual helicity does not match the binding affinity for N-Mdm2. The binding affinity of p53

P12,13A correlates with its increased residual helicity (32 % compared with WT 28 %) [20], displaying a moderate increase in the populations of the bound state compared with WT p53, when the widening of the complex peaks is accounted for. Our MS data for p53

P27A (Figure 5.3c, red line) shows excellent correlation with the ITC data described by Borcherds et al.[20], showing a significant increase in the binding affinity for N-Mdm2 which corresponds to its increased residual helicity (64 %). Interestingly however, replacing
prolines P12 and P13 in addition to P27 seems to have a negative impact on the affinity of p53\(^{P27A}\) for N-Mdm2 compared with replacing P27 alone (Figure 5.3c red line vs magenta line), in contrast with NMR data which showed a slight increase in residual helicity to 66%. The binding affinity of p53\(^{P27A}\) (Figure 5.3, green line) is similar to that of p53\(^{P27A}\). This may be expected, as an increased level of helicity outside of the Mdm2-binding region should have no impact on the complex association or dissociation.

5.7 Residual helicity does not affect complex conformation

The p53:Mdm2 complex presents in low charge states [p53:Mdm2+8H]\(^{8+}\) and [p53:Mdm2+9H]\(^{9+}\), with the additional [p53:Mdm2+7H]\(^{7+}\) observed for p53\(^{PallA}\). This indicates a compact structure, where the highly charged disorder signature has been tamed by the interaction. The CCSDs for each charge state of the p53TAD:N-Mdm2 complexes for each p53TAD construct are shown in Appendix D Figure S24, each spanning approximately the same CCS range, from \(~1000\ \text{Å}^2\) to \(~2500\ \text{Å}^2\), with two ill-resolved or interconverting conformers centered at \(~1600\ \text{Å}^2\) and \(~1900\ \text{Å}^2\). Comparison of \(\text{DTCCS}_{\text{He}}\) values for the complex with that of the respective monomers reveals that the complex is compact, with only slightly larger \(\text{DTCCS}_{\text{He}}\) than that found for the smallest N-Mdm2 conformer sized 1500 \text{Å}^2 (Figure 5.2c). This may be due to the p53TAD forming a close interaction with N-Mdm2, and/or due to conformational rearrangement of the N-Mdm2 structure. Uhrinova et al. [37] argued that N-Mdm2 undergoes expansion upon binding to p53; a conformational change attained through rearrangement of its two sub-domains resulting in a more ordered and open conformation of the binding cleft. Schon et al. [9] also reported extensive changes in chemical shifts when N-Mdm2 (residues 2-125) is complexed with p53TAD peptides (residues 15-29 and 17-26). These perturbations were not only restricted to the binding cleft but could be mapped to the other parts of the protein, suggesting extensive conformational rearrangement. Interestingly, we observe a compaction in free N-Mdm2 when in the presence of p53TAD but unbound (Appendix D Figure S25), contrasting the observations by Uhrinova et al. [37]. We used MOBCAL [38-40] to calculate the theoretical CCS value \(\text{CALC}_{\text{TM}}\) of the p53:Mdm2 complex from the coordinates obtained from X-ray crystallography measurements[6] (PDB; 1YCR). The \(\text{CALC}_{\text{TM}}\) of the complex is 1270 \text{Å}^2, smaller than the conformational families observed experimentally (\(~1600\ \text{Å}^2\) and \(~1900\ \text{Å}^2\)), although within the CCSD spread exhibited by the complex (Appendix D Figure S
Proline residues mediate promiscuous binding of p53 to Mdm2

As the crystal structure does not utilise the full N-terminus of p53 (only residues 15-29 included in the structure[6]) used in our IM-MS measurements (residues 1-73), it can be inferred that the complex matches relatively well with the theoretical value and corresponds well with a restructuring of N-Mdm2 when binding to p53TAD to a more homogenic complex.

The similarities in the CCSDs of the complex formed by WT and mutant p53TAD constructs suggest that the altered residual helicity of mutant p53TAD does not impact on the final complex conformation. From this we can infer that the replacement of proline does not affect binding mode, only altering the binding affinity. Wide CCSDs and the observation of at least two conformational families indicate the complex possesses flexibility or exists in multiple interconverting conformations on our experimental timescale (~15 ms including time spend in the drift cell). To simplify, structured, folded proteins with little conformational flexibility in solution, would give rise to narrower CCSDs, typically presenting one stable conformation. In turn, less structured or unfolded proteins would give rise to wider ATDs often with multiple conformations, as seen here. Analysis of the width of ATDs can act as an indicator of intrinsic disorder in proteins and their complexes, semi quantifying the conformational flexibility and spread of structures. Combined, we observe a complex which is compact; the binding of p53TAD to N-Mdm2 results in compaction of N-Mdm2 (Appendix D Figure S25) and yields a complex ~100-300 Å² larger than the smallest conformer of free N-Mdm2. This structure matches well with the CCS_CCSTM, suggesting the complex is tightly associated but has not undergone collapse during the electrospray process or under the vacuum of the mass spectrometer, and as such we postulate that N-Mdm2 undergoes a folding-upon-binding event when interacting with p53TAD.

5.8 Mutant p53TAD undergoes distinct conformational changes during Mdm2 binding

Due to the transient nature of the p53TAD and the high nanomolar/low micromolar reported K_d of the p53TAD-N-Mdm2 interaction [20, 33, 41, 42], the conformational distribution of free p53TAD in the presence of N-Mdm2 but not bound may give insight into any conformational preferences of, or conformational changes occurring during, the
Proline residues mediate promiscuous binding of p53 to Mdm2

p53:Mdm2 binding event. We have previously reported the gas-phase templating of flexible proteins where we are able to observe the bound form of a protein even when no longer associated with its ligand/binding partner [43].

CCSDs of p53TAD [M+5H]^{5+} in the presence of N-Mdm2, but not bound, are shown in Figure 5.4. For WT p53, we observe that free p53TAD undergoes a compaction when in the presence of N-Mdm2 (Figure 5.4a); we observe total loss of the conformer at ~1100 Å^2 (shown in blue) and significant depletion of the conformer at ~990 (shown in yellow), with the majority of the ions presenting in the compact conformer at ~880 Å^2 (shown in red).

Figure 5.4 CCSDs derived from ATDs for the [M+5H]^{5+} charge states of a) WT p53, b) p53^{P12,13A}, c) p53^{P27A}, d) p53^{P3xA}, e) p53^{PallA} in the absence (top row) and presence of N-Mdm2 at varying ratios. Conformers are fitted and coloured according to Appendix D Table S 2.

p53^{P12,13A} and p53^{P27A} [M+5H]^{5+} (Figure 5.4b and Figure 5.4c, top panels) present with wider CCSDs and an increased proportion of ions in the ~880 Å^2 conformer (shown in red) and the ~1100 Å^2 conformer (shown in blue) compared with WT p53. p53^{P12,13A} exhibits depletion of the conformer at ~1100 Å^2 upon incubation with N-Mdm2 at higher ratios (Figure 5.4b), although the CCSD does not alter significantly. p53^{P27A} also undergoes a slight reduction in the intensity of both the intermediate conformer at ~1100 Å^2, with almost complete loss at 1:1 ratio, and the compact conformer ~880 Å^2 (shown in red), with the conformer 990 Å^2 (shown in yellow) retained.
p53\textsuperscript{P3xA}, shown by NMR to have an increased residual helicity compared with p53\textsuperscript{P27A} \cite{20}, was shown by MS to have a reduced binding affinity (Figure 5.3c). We observe a significantly different CCSD for p53\textsuperscript{P3xA} compared with all other constructs; a large population of ions presents in the compact conformer centered at \(\sim 750 \text{ Å}^2\) (shown in grey) with conformers \(\sim 880 \text{ Å}^2\) (red) and \(\sim 990 \text{ Å}^2\) (yellow) also intense. Upon incubation with N-Mdm2 (Figure 5.4d), p53\textsuperscript{P3xA} undergoes a significant conformational change, resulting in reduction of the compact \(\sim 750 \text{ Å}^2\) and \(\sim 880 \text{ Å}^2\) conformers. We observe a large proportion of the ions in the \(\sim 1100 \text{ Å}^2\) conformer (blue) at the 10:1 ratio. This conformer at \(\sim 1100 \text{ Å}^2\) is then depleted with increasing ratios of N-Mdm2.

p53\textsuperscript{PallA} presents in a wide, ill-defined CCSD encompassing conformers at \(\sim 880 \text{ Å}^2\), \(\sim 990 \text{ Å}^2\), with a small proportion of ions in the \(\sim 1100 \text{ Å}^2\) conformer. Upon incubation with N-Mdm2, we observe a decrease in the intensity of the \(\sim 1100 \text{ Å}^2\) conformer, producing a narrower CCSD consisting of predominantly the conformer at \(\sim 990 \text{ Å}^2\).

We observe significant depletion of [M+6H]\textsuperscript{6+} for all p53TAD constructs upon incubation with N-Mdm2 (Figure 5.3, Appendix D Figure S 23), indicating that p53TAD ions presenting in this charge state may be involved in the binding event. The corresponding CCSDs for the WT and mutant p53TAD [M+6H]\textsuperscript{6+} are shown in Figure 5.5. WT p53 (Figure 5.5a) in the absence of N-Mdm2 presents in conformers at \(\sim 880 \text{ Å}^2\), \(\sim 990 \text{ Å}^2\), \(\sim 1100 \text{ Å}^2\) and \(\sim 1250 \text{ Å}^2\) (red, yellow, blue and green, respectively), with most intensity in the intermediate states \(\sim 1100 \text{ Å}^2\) and \(\sim 1250 \text{ Å}^2\). Upon incubation with N-Mdm2 we observe a significant decrease in the intensity of the intermediate charge states \(\sim 1100 \text{ Å}^2\) (blue) and \(\sim 1250 \text{ Å}^2\) (green). For WT p53, we observe a conformational shift of free p53 to produce ions in a more compact conformation for both the [M+5H]\textsuperscript{5+} (Figure 5.4a), and [M+6H]\textsuperscript{6+} (Figure 5.5a) charge states. With a high reported \(K_d\), corroborated by our MS experiments (Figure 5.3c), we anticipate that WT p53 will undergo many binding events with N-Mdm2 in solution and thus the conformational changes associated with these binding events will be captured on the timescale of our IM-MS experiments. We therefore suggest that this is a conformational change, rather than the loss of the intermediate conformers due to conformer-specific binding.
Proline residues mediate promiscuous binding of p53 to Mdm2

Figure 5.5 CCSDs derived from ATDs for the \([\text{M}+6\text{H}]^{6+}\) charge state of a) WT p53, b) p53\(^{12,13A}\), p53\(^{27A}\), p53\(^{33A}\) and e) p53\(^{49A}\). Conformational families are fitted and denoted by their colour in Appendix D Table S2. X-axis denotes collision cross section (\(\text{CCS}_{\text{incl}}, \text{Å}^2\)), y-axis denotes relative intensity of the area of the CCSD normalised to the intensity of the corresponding peak in the mass spectrum.

p53\(^{12,13A}\) (Figure 5.5b) presents in a similar conformational spread to the WT protein, encompassing the same conformers: \(\sim 880 \text{ Å}^2\), \(\sim 990 \text{ Å}^2\), \(\sim 1100 \text{ Å}^2\) and \(\sim 1250 \text{ Å}^2\), but with an increased proportion of ions in the more compact conformers. Upon incubation with N-Mdm2, we observe a reduction in the intensity of the compact conformer at \(\sim 880 \text{ Å}^2\), with almost complete loss at 2:1 and 1:1 ratios of p53:N-Mdm2.

p53\(^{27A}\), which shows a significant increase in the binding affinity for N-Mdm2 (Figure 5.3c, red line) [20], also displays the same conformational spread as observed for WT p53, but with an increased intensity of the compact conformer \(\sim 880 \text{ Å}^2\) (Figure 5.5c). Upon incubation with N-Mdm2, we observe an almost complete depletion of compact conformers \(\sim 880 \text{ Å}^2\) and \(\sim 990 \text{ Å}^2\) at all ratios. This is a similar trend to that observed for p53\(^{12,13A}\), however the significant CCSD alterations occur at a lower ratio. Indeed, at ratios 5:1 and higher, the free population of p53\(^{27A}\) is low intensity and presents only of the intermediate conformer at \(\sim 1250 \text{ Å}^2\). Both p53\(^{12,13A}\) and p53\(^{27A}\) display similar trends when incubated with N-Mdm2; we observe minimal changes to the \([\text{M}+5\text{H}]^{5+}\) species but significant depletion of the compact states of the \([\text{M}+6\text{H}]^{6+}\). As they exhibit an increased binding affinity compared with WT p53 (Figure 5.3c black vs red/blue) we suggest that the compact conformers of p53\(^{12,13A}\) and p53\(^{27A}\) selectively bind to N-Mdm2, a mechanism which occurs to a greater extent for the more helical p53\(^{27A}\).
p53\textsuperscript{P3xA} [M+6H]\textsuperscript{6+} possesses an additional compact conformer centered at ~750 Å\textsuperscript{2} in the absence of N-Mdm2 (Figure 5.5d), as observed for the [M+5H]\textsuperscript{5+} charge state. Upon incubation with N-Mdm2, the intensity of this compact conformer is reduced, along with the reduction in the intensity of the intermediate conformers at ~1100 Å\textsuperscript{2} and ~1430 Å\textsuperscript{2}. p53\textsuperscript{P3xA} exhibits the compact conformer at ~750 Å\textsuperscript{2} in both the [M+5H]\textsuperscript{5+} and [M+6H]\textsuperscript{6+} ions, which is depleted alongside more intermediate conformers upon incubation with N-Mdm2. The novel compact conformer centered at ~750 Å\textsuperscript{2} adopted by p53\textsuperscript{P3xA} may be responsible for the unexpectedly weak N-Mdm2 binding affinity compared with p53\textsuperscript{P27A} (Figure 5.3c). We suggest that the protein may be required to conformationally rearrange from this compact state in order to bind to N-Mdm2, thereby reducing the complex formation as observed in our experiments.

Free p53\textsuperscript{PallA} [M+6H]\textsuperscript{6+}, in the absence of N-Mdm2, is extended; the CCSD is significantly altered compared with the other p53TAD constructs for both [M+5H]\textsuperscript{5+} (Figure 5.4e) and the [M+6H]\textsuperscript{6+} (Figure 5.5e). For the [M+6H]\textsuperscript{6+}, we observe a wide, ill-defined distribution encompassing intermediate conformers ~1250 Å\textsuperscript{2} and ~1400 Å\textsuperscript{2} (green and magenta, respectively) and more extended states at ~1600 Å\textsuperscript{2} and ~1800 Å\textsuperscript{2} (orange and purple). Upon the incubation of this construct with N-Mdm2 at a 10:1 ratio we observe a significant conformational change; complete loss of the extended states of the protein and the appearance of the compact conformers centered at ~880 Å\textsuperscript{2} and ~990 Å\textsuperscript{2}. This distribution is then depleted with increasing ratios of N-Mdm2. This significant CCSD change could arise via several binding mechanisms; for instance p53\textsuperscript{PallA} may be required to undergo a significant rearrangement into a more compact state to facilitate the correct binding of p53\textsuperscript{PallA} to N-Mdm2. We postulate however, that the more extended states of p53\textsuperscript{PallA} are highly helical. IM-MS has previously observed helical states of proteins with a \textit{DTCCS}_{hel} larger than that observed for the disordered state [44-46], and we have previously shown that the extended states of p53\textsuperscript{PallA} compare well to a helical state of Ubiquitin, a protein of similar mass (Chapter 4). The binding affinity of p53\textsuperscript{PallA} for N-Mdm2 is similar to that observed for p53\textsuperscript{P27A} (Figure 5.3c, red line vs green line). The complex of p53\textsuperscript{PallA}:N-Mdm2 presents in an altered charge state range compared with other p53TAD constructs, however exhibits the same conformational preferences, presenting in the same conformers (Appendix D Figure S24). Together, we infer that the extended helical states of p53\textsuperscript{PallA} contain the pre-stabilised α-helix of the Mdm2-binding PreSMos and are therefore able to bind directly to N-Mdm2 with little conformational alteration, resulting in their depletion from the free protein.
5. Proline residues mediate promiscuous binding of p53 to Mdm2

5.9 Conclusions

The p53-Mdm2 interaction remains one of the best studied complexes, and structural information regarding the complex structure and dynamics, the interaction mechanisms of the free monomers and regulation of these processes is required in order to enable design of small, drug-like ligands for inhibition [47]. Here, we use MS and IM-MS to probe a series of p53TAD mutants in which prolines are replaced with alanine, forming p53TAD constructs with increased residual helicity. We show that the proline-dependent increase in residual helicity results in an increase in binding affinity for N-Mdm2, although this result is not linear with residual helicity content.

Mittal et al. [41] have recently shown that free p53TAD is present in structural isomers which closely match the Mdm2-bound form, although overall the conformational landscape of the disordered domain was relatively flat. This correlates well with our IM-MS data, which shows that both WT and mutant p53TAD adopt a range of conformations both in the absence and presence of N-Mdm2. The conformational spread of the p53:Mdm2 complex spans a range of over ~1500 Å², and encompasses two broad conformational families, indicating the flexibility of these IDPs is retained even whilst in complex.

We use IM-MS as a diagnostic tool to track alterations to the populations of p53TAD conformers upon N-Mdm2 binding. We show proline-dependent differences in the conformational preference of the binding event. WT p53 undergoes a conformational change upon incubation with N-Mdm2, forming a more compact conformation which is retained in the unbound population. p53⁠₃₁₂,₁₃ₐ and p53⁠₉₂₇ₐ exhibit N–Mdm2 dependent depletion of their compact conformers, suggesting conformational selection of these structures during binding. p53⁠₃₃ₐ displays a highly compact CCSD in the absence of N-Mdm2, which then undergoes rearrangement upon binding to Mdm2, indicating this very compact conformer is unable to bind. We suggest that this conformer may be kinetically trapped in our gas-phase experiments, forming a very compact and possibly molten globule-like state. p53⁠₉₃ₐ, shown to be highly helical, presents in extended conformations which become depleted upon Mdm2-binding. We suggest that these extended conformers are helical in structure and are able to form a complex with N-Mdm2 without extensive structural rearrangement.
Proline residues mediate promiscuous binding of p53 to Mdm2

The systems presented here undoubtedly exhibit behaviour common to all IDPs: promiscuous interactions, persistent flexibility in complex, conformational diversity, folding coupled to binding and partial folding in isolation. This gas-phase study has provided insights into the implications of disorder regulation using proline; and has shown that disorder-to-order transitions can be modulated at the primary structure level, although the conformational and structural impacts of proline replacement are not linear. IM-MS can delineate conformational differences in proteins comprising a heterogeneous ensemble. Although no differences between the structures of the final p53:Mdm2 complexes are observed, IM-MS is able to track conformer-specific binding between WT and mutant p53TAD and N-Mdm2.

5.10 References

5. Proline residues mediate promiscuous binding of p53 to Mdm2


5. Proline residues mediate promiscuous binding of p53 to Mdm2


6
Protein-protein Interactions of MAGE-A4 with gankyrin and p53
6.1 Declaration

This chapter consists of one draft paper awaiting submission:


As a second author on this publication I carried out all MS and IM-MS experiments and compiled the data for Figure 6.2, Figure 6.4, Figure 6.7 and Figure 6.8. I also drafted and edited the MS and IM-MS sections of manuscript. YH expressed and purified the MAGE-A4, gankyrin and p53 samples, carried out NMR experiments and drafted all other sections of the manuscript.
6.2 Abstract

Melanoma antigen A4 (MAGE-A4) belongs to MAGE protein family and specifically to type I MAGEs. Type I MAGEs are overexpressed in cancer cells and are associated with cancer malignancy and poor patient prognosis. The functions of these proteins remain poorly understood. Here we report protein-protein interactions of MAGE-A4 with the oncoprotein gankyrin and p53 tumour suppressor using nuclear magnetic resonance techniques and native mass-spectrometry. In addition, both electrophoretic mobility shift assay and native mass-spectrometry demonstrate that MAGE-A4 disrupts the interaction of p53 with its response elements. Our findings shed light on the biological functions of MAGE-A4, which can be exploited in anti-cancer therapy.

6.3 Introduction

Melanoma antigen 4 (MAGE-A4) belongs to MAGE family, which in turn forms a broader family of cancer-testis antigen (CTA) along with GAGE, XAGE, PAGE and SAGE etc [1-6]. MAGE family can be divided into two subfamilies based on their expression pattern: type I and type II [7]. While the expression of type I MAGEs is restricted to germline and cancer cell, type II subfamily is expressed in somatic tissues. The expression of type I MAGEs are repressed by hypermethylation in CpG rich promoters blocking the access of transcriptional factors [8, 9]. However in tumour cells, hypomethylation can induce the expression of type I MAGEs. The mechanism of transcriptional activation of type I MAGEs has not been revealed yet.

Type I MAGEs can be subdivided into MAGE-A, MAGE-B and MAGE-C groups. MAGE-A group consists of 12 genes (MAGE-A1 to MAGE-A12) with MAGE-A7 being pseudogene [10]. All of MAGE family proteins share a highly conserved domain termed as MAGE homology domain (MHD) spanning over 200 amino acid residues from the central region to near the C-terminus [11]. The remaining N-terminal region is less conserved among the family and is expected to be disordered as it is rich in disorder promoting residues such as Ser, Pro and Glu etc. MHD consists of two winged-helix termed as WH-A and WH-B [12].
The expression of type I MAGEs has been correlated with cancer malignancy and poor patient prognosis [13-17]. Although the molecular mechanisms of MAGEs in cancer development remain poorly understood, emerging data suggest that they promote cell growth and evasion of cell cycle arrest and/or apoptosis through down-regulation of p53 tumour suppressor. For example, MAGE-A2 was reported to interact with the DNA binding domain (DBD) of p53 and suppresses its function either by recruiting histone deacetylase [18] or by steric hindrance [19]. Furthermore, various MAGEs have been identified to bind with E3 RING ubiquitin ligases enhancing their ligase activity [12]. In particular, MAGE-C2 was found to bind with TRIM28 ubiquitin ligase, promoting the degradation of p53 in proteasome dependent manner [12]. MAGE-A4 was reported to suppress the expression of p53 downstream pro-apoptotic and cell cycle arrest genes such as Bax and p21 [20]. However it is not known how MAGE-A4 suppresses the expression of those genes.

Interestingly, tumour suppressing activity was also reported for MAGE-A4. For example, MAGE-A4 was found to be expressed non-small cell lung cancers and promote apoptosis [21]. Furthermore, MAGE-A4 was shown to interact with the oncoprotein gankyrin suppressing its tumorigenic activity [22].

In this study, we have examined the interaction of MAGE-A4 with gankyrin and p53 DNA-binding domain (DBD) using heteronuclear quantum coherence (HSQC) nuclear magnetic resonance (NMR), native mass-spectrometry (MS) and ion mobility mass spectrometry (IM-MS). The results indicate that MAGE-A4 can interact with gankyrin and p53 DBD. Furthermore electrophoretic mobility shift assay (EMSA) studies and native nanoelectrospray ionisation mass spectrometry (nESI-MS) experiments reveal that excess amount of MAGE-A4 is capable of blocking the DNA binding ability of p53 DBD, displacing promoter DNAs of p53 response elements (such as Bax, PUMA and p21) from p53 DBD. These findings shed light on the biological functions of MAGE-A4, which remain poorly understood, and can have implications in anti-cancer therapy.

### 6.4 Materials and Methods

#### 6.4.1 Cloning
MAGE-A4 was cloned into pGEX-6p-1 (GE Healthcare) between EcoRI and XhoI to express GST-tagged MAGE-A4 using forward primer:

5'-AAAAAAGAATTCATGTCTTCTGAGCAGAAGAGTCAG-3',

reverse primer: 5'-AAAAAAACTCGAGTCAGACTCCCTCTTCTCTAA-3' and vector pGEX-4T-1 containing MAGE-A4 DNA sequence [23] as a template. The GST-tag is removable using PreScission® protease (GE Healthcare). Gankyrin DNA was cloned into plasmid pET-22b(+) using forward primer:

5'AAAAAAACATATGGGCAGCAGCCATCATCATCATCATCACACAGCATGGAGGGGTGTGTGTC TAACCTA 3'

and reverse primer: 5' AAAAAAAACTCGAGTAAACCTCCACCATCTTCTCTT 3', and gankyrin cDNA clone IRATp970C0117D (Source Bioscience) as a template DNA to express 6 × His-tagged gankyrin. p53 DBD (amino acid residues: 94-312) were previously cloned in the lab into pRSET (A) (Invitrogen) expression vector.

### 6.4.2 Protein Expression and Purification

The expression plasmid of MAGE-A4 was transformed into E. coli C41 (DE3) cells. Single colonies were inoculated into 5 ml LB media containing 100 μg/ml ampicillin, followed by incubation at 37 °C overnight. Each of these overnight mini-culture was added into 500 ml LB supplemented with 100 μg/ml ampicillin. The cultures were allowed to grow at 37 °C till OD600 reached 0.6. The cultures were then allowed to cool down to 21 °C and the protein expression was induced with 1 mM isopropyl β-D-thiogalactoside (IPTG) overnight. The cells were harvested by centrifugation and lysed using 5 ml/g cell of BigBuster® protein extraction reagent (Novagen) containing 1 tablet of cOmplete® EDTA- free protease inhibitor (Roche), 125 U of Benzonase nuclease (Novagen) and 5 mM dithiothreitol (DTT). The soluble fraction of the cell lysate was collected by centrifugation and the GST-tagged MAGE-A4 was purified using GST GraviTrap columns (GE Healthcare). The GST-tagged protein was incubated with 5 U/mg protein of PreScission® protease (GE Healthcare) at 4 °C overnight. The protein sample was then passed again through GST GraviTrap column to remove GST, followed by further purification using 26/60 Superdex 200 prep grade gel-filtration column connected to ÄKTAprime® system equipped with PrimeView software (all
Protein-protein Interactions of MAGE-A4 with gankyrin and p53 from GE Healthcare). Gankyrin expression vector and p53 DBD expression vector were also transformed into the above-mentioned bacterial cells, and the protein expressions were induced as described above. Gankyrin protein was purified using HisTrap HP column (GE Healthcare), followed by further purification using 26/60 Superdex 200 prep grade gel-filtration column. p53 DBD was purified using HiTrap® SP HP cation exchange column (GE Healthcare) followed by further purification using HiTrap® Heparin HP column.

6.4.3 Protein Labelling with $^{15}$N for $^{1}$H-$^{15}$N HSQC NMR

MAGE-A4 protein was labelled with $^{15}$N for HSQC NMR by supplementing minimal media with $^{15}$NH$_4$Cl as nitrogen source. C41 (DE3) *E. coli* cells were transformed with the expression vector of GST-tagged MAGE-A4 as described above. A single colony was inoculated into LB media (5 ml), followed by incubation at 37 °C till OD$_{600}$ reached 0.6-0.8. Minimal media (50 ml) in 250 ml flasks were each inoculated with 5 ml of mini-culture, followed by incubation at 37 °C overnight. The cells were harvested by centrifuging at 1,600 x g for 10 minutes. The cell pellet was re-suspended in 10 ml labelled minimal media, which was inoculated into 500 ml labelled media in 2.5 L flask. The cultures were grown at 37 °C till the OD$_{600}$ reached 0.8, and the protein expression was induced with 1 mM IPTG at 21 °C. The cultures were incubated at 21 °C overnight. The protein was purified, followed by tag removal and further purification using gel-filtration column as described above.

6.4.4 $^{1}$H-$^{15}$N HSQC NMR

MAGE-A4 protein labelled with $^{15}$N was concentrated to 150 μM in 50 mM phosphate buffer (pH 7.2) supplemented with 50 μM NaCl and 5 mM DTT. D$_2$O was added to the labelled protein sample to 10%. The protein sample was transferred to an NMR tube and the signals were obtained at 25 °C using Bruker Avance III 600 MHz equipped with TCI cryoprobe. Unlabelled gankyrin was titrated to the labelled MAGE-A4 and the signals were collected again. Same method was used to examine the protein-protein interaction between MAGE-A4 and p53 DBD.
6.4.5 MS and IM-MS for Protein-Protein Interaction Studies

Proteins MAGE-A4, p53 DBD and gankyrin were snap frozen in liquid Nitrogen and stored at -80 °C. Before analysis, samples were thawed and buffer exchanged into 50 mM ammonium acetate (pH6.8) using BIO-RAD micro bio-spin chromatography columns (BIO-RAD Laboratories, Inc). Protein concentration was measured using a Thermo Nanodrop.

nESI mass spectrometry experiments were performed on a Waters Synapt G2Si mass spectrometer (Waters, Manchester, UK). Tuning parameters were kept as constant as possible and are as follows: capillary voltage, 1.33-1.45 kV; cone voltage, 30-50 V; Trap collision energy, 2-6 V; source temperature, 80 °C.

IM-MS experiments were carried out on an in-house modified QToF2 (Waters, Manchester, UK) mass spectrometer incorporating a 5.1 cm drift cell [24]. Proteins were prepared at a 1:1 ratio, with each protein present at a concentration of 20 µM. The drift cell was filled with Helium at an average pressure 3.8 – 3.95 Torr measured by a baratron (MKS Instruments) and a temperature between 294 – 297 K. The electric potential across the cell was lowered from 60 to 20 V and arrival time distributions (ATDs) recorded at 7 drift voltages. The ATD was converted into a collision cross section distribution (CCSD, Å²) using Equation 1.9 shown below:

\[
\Omega_{\text{avg}} = \frac{(18\pi)^{1/2}}{16} \left[ \frac{1}{m_b} + \frac{1}{m} \right]^{1/2} \frac{ze}{(k_BT)^{1/2} \rho} \frac{1}{L^2} \frac{t_dV}{L^2}
\]

Where \(\Omega\) is the rotationally averaged collision cross section \((\text{CCS}_{\text{He}}, \text{Å}^2)\), \(m\) and \(m_b\) are the masses of the ion and the buffer gas, \(z\) is the ion charge, \(e\) 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 cell length, \(V\) is the voltage applied across the drift cell (here 35 V) and \(t_d\) is the drift time. The arrival time of the ions \((t_a)\) includes the time spent within the mass spectrometer by outside of the drift cell, also 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 \((t_d)\):
All processing was carried out using Mass Lynx V4.1 (Waters Corporation) and Origin 9.0 (OriginLab Corporation, USA) software. Intensities of the CCSD peaks were taken directly from ATD area in Mass Lynx V4.1 software (Waters Corporation). Gaussian curves were fitted with fixed centroid and width values to show abundance of conformational families present at each charge state.

6.4.6 MS for Protein-DNA Interaction Studies

Promoter DNA’s p21, Bax and PUMA were prepared in 50 mM ammonium acetate (pH 6.8). Proteins p53 DBD and MAGE-A4 were buffer exchanged into 50 mM ammonium acetate (pH 6.8) using BIO-RAD micro bio-spin chromatography columns (BIO-RAD Laboratories, Inc). Protein:DNA experiments were carried out at 20 µM:2 µM, 20 µM:4 µM, 20 µM:10 µM and 20 µM:20 µM ratios all with a final volume of 15 µL. Protein:DNA mixtures were incubated for 15 minutes at room temperature before analysis.

To observe the effect of MAGE-A4 on the p53 DBD:DNA interaction, p53 DBD and promoter DNA was incubated in a 1:1 ratio with both samples at a concentration of 30 µM for 15 minutes at room temperature. MAGE-A4 was then added at a range of concentrations (15 µM, 30 µM, 60 µM and 90 µM) to a final volume of 15 µL.

nESI mass spectra were obtained using a Micromass Q-ToF Ultima Global mass spectrometer (Waters, UK). Tuning parameters were kept as constant as possible and are as follows: capillary, 1.36 – 1.6 kV; cone, 35V; source temperature, 60 °C.

6.4.7 EMSA

EMSA was carried out to examine whether MAGE-A4 is capable of binding with p53 downstream gene promoter DNAs as well as to investigate whether MAGE-A4 has effects on the interaction between p53 DBD and such DNAs. The protein samples were concentrated in 50 mM Tris buffer (pH 7.2) supplemented with 150 mM NaCl, 5% glycerol
and 5 mM DTT. The DNA promoters were 5'-fluorescein labelled to enable sensitive
detection under UV light than ethidium bromide. For testing whether MAGE-A4 can bind
with the DNA promoters, the DNA concentration was kept constant (2 μM) while the
concentration of the protein was increased by twice. Each mixture of DNA and MAGE-A4
was prepared in 20 μl and incubated on ice for one hour. For examining the effects of
MAGE-A4 on p53 DBD-DNA complex, p53 DBD and DNA were mixed at constant
concentrations (1 and 4 μM, respectively), while MAGE-A4 was titrated so that its
concentration increases by twice in a volume of 20 μl. The samples were incubated on ice
for one hour. The samples were run on a 0.7% agarose gel with Tris/Borate/EDTA buffer at
a constant current of 80 V at 4 °C for 1 hour. The DNA bands were visualised under UV
light.

6.5 Results

6.5.1 Protein-Protein Interaction between MAGE-A4 and Gankyrin Probed by
\(^1\)H-\(^{15}\)N HSQC NMR

A spectrum of the \(^{15}\)N labelled MAGE-A4 alone was recorded (Figure 6.1a). Then the
labelled MAGE-A4 was titrated with unlabelled seven molar equivalent of gankyrin (Figure
6.1b). The spectrum of MAGE-A4 alone and that of titrated mixture are overlaid in Figure
6.1. The HSQC NMR spectrum of MAGE-A4 alone shows large signals around 8.5 ppm of \(^1\)H
chemical shift as well as smaller signals dispersed in a wider range of chemical shift. This
indicates that the protein contains disordered region(s) as well as a folded domain. After
adding gankyrin, some of the dispersed peaks disappear or weaken. This may be due to
complex formation resulting in slower tumbling of the protein complex and thus line-
broadening. Furthermore, the weakening or disappearance of these peaks may also be
caused by intermediate exchange rate on the NMR timescale. This result indicates complex
formation between MAGE-A4 and gankyrin. As the changes in the signals are observed in
those corresponding to the folded domain but not those in unstructured regions (around
8.5 ppm), it is expected that this interaction is mediated by the MHD of MAGE-A4 and not
the N-terminal region, which is predicted to be disordered.
6.
Protein-protein Interactions of MAGE-A4 with gankyrin and p53

![Diagram a](hsqc_600.nv)

![Diagram b](hsqc_600+7eq_gan.nv)
Protein-protein Interactions of MAGE-A4 with gankyrin and p53

6.5.2 Protein-Protein Interactions between MAGE-A4 Proteins and Gankyrin Probed by Nanoelectrospray Ionisation Mass Spectrometry (nESI-MS) and Ion Mobility Mass Spectrometry (IM-MS)

nESI-MS was used to probe MAGE-A4 interaction with gankyrin. Individual spectra for MAGE-A4 and gankyrin are shown in Appendix E Figure S 26 and Figure S 27, respectively and the spectrum for MAGE-A4 has been previously described [25]. MAGE-A4 presents as a very wide multimodal charge state distribution $7 \leq z \leq 37$, indicative of a disordered protein.
in solution [26], with intense peaks at [M+11H]$^{11+}$, [M+15H]$^{15+}$, and [M+29H]$^{29+}$. Gankyrin present in a charge state range 6 ≤ z ≤ 25, with 90 % of the intensity in the [M+9H]$^{9+}$, [M+10H]$^{10+}$ and [M+11H]$^{11+}$. We observe a loss of the highest charge states 24 ≤ z ≤ 37 of MAGE-A4, and observe loss of low intensity high charge states 11 ≤ z ≤ 24 of gankyrin when the two proteins are incubated together. The mass spectra for an equimolar mixture of MAGE-A4 and gankyrin (10 µM:10 µM) is show in Figure 6.2a. The spectrum shows that both MAGE-A4 and gankyrin predominantly present as monomers, with lower intensity peaks attributable to MAGE-A4 in complex with gankyrin. The concentration of each protein was altered from 10 to 80 µM and at different ratios (data not shown), no change was seen in the relative intensity of the complex indicating a kD that is > 50 µM. The relative amount of complex also did not alter with incubation time, indicating that the kinetics of binding is fast (data not shown). Heterodimers due to MAGE-A4:gankyrin complexes [MA4:Gnk+zH]$^{z+}$ are seen as peaks with m/z 4382, 4090, 3834 and 3609, attributable to charge states (z) from 14-17 and a complex mass of 61339 Da which compares well with that predicted from the sequence. Lower abundance species corresponding to MAGE-A4 dimers in complex with a gankyrin monomer (peaks corresponding to [2MA4:Gnk+20H]$^{20+}$ and [2MA4:Gnk+21H]$^{21+}$), and MAGE-A4 monomer in complex with gankyrin dimer (peak corresponding to [MA4:2Gnk+19H]$^{19+}$) were also observed.

Drift time ion mobility mass spectrometry (DT-IM-MS) of the MAGE-A4:gankyrin heterodimer (Figure 6.2b) reveals that the complex presents in a single conformational family centered at $^{DT}$CCS$_{He}$ ~3200 Å$^2$ for lower charge states [MA4:Gnk+14H]$^{14+}$ and [MA4:Gnk+15H]$^{15+}$. The higher charge states exhibit an additional extended conformational family centered at $^{DT}$CCS$_{He}$ ~4000 Å$^2$. For charge state [MA4:Gnk+16H]$^{16+}$, the more compact conformational family is prevalent, for the most extended [MA4:Gnk+17H]$^{17+}$ charge state the larger conformational family is most abundant.
6. Protein-protein Interactions of MAGE-A4 with gankyrin and p53

Figure 6.2 a) nESI mass spectrum of MAGE-A4 and gankyrin. Inset spectra indicated a zoom of the higher m/z species present. Several charge states are denoted; single grey spheres represent MAGE-A4 and single red sphere represents gankyrin. Double or triple spheres represent dimers and trimers respectively. Asterisks denote contaminant peaks in spectra. b) DT-IM-MS CCSDs converted from ATDs for the heterodimeric MAGE-A4:gankyrin complex at charge states [MA4:Gnk+14H]^{14+} – [MA4:Gnk+17H]^{17+}. X-axis denotes drift time collision cross section measured in Helium (OT_CCS exp, Å²). Conformational families are denoted by coloured Gaussian curves in purple and orange.

6.5.3 HSQC NMR for the protein-protein interaction between MAGE-A4 and p53 DBD

It was reported that MAGE-A2 interacts with p53 DNA binding domain (DBD) and represses its transactivational function by recruiting histone deacetylase 3 (HDAC 3) to p53 [18] or by sterically preventing the interaction between p53 and DNAs [19]. In addition, overexpression of MAGE-A4 was reported to induce growth in normal keratinocytes by inhibiting growth arrest and apoptosis through repression of p53 downstream pro-apoptotic and cell cycle arrest genes such as BAX and p21 [20]. It is not known whether MAGE-A4 is capable of binding with p53. As MAGE-A4 is highly homologous to MAGE-A2 sharing 68% of amino acid sequence identity, we examined whether MAGE-A4 interacts with p53 DBD using ^1H-^{15}N HSQC NMR.
MAGE-A4 protein was labelled with $^{15}$N and the HSQC spectrum was recorded (Figure 6.3a). Three molar equivalent of unlabelled p53 DBD was titrated into the labelled MAGE-A4 and the spectrum is shown in Figure 6.3b. The spectrum of MAGE-A4 alone and that obtained after adding p53 DBD are overlaid in Figure 6.3c. After adding p53 DBD, some of the dispersed peaks disappear or weaken (such as signals around 10.5-10.6 ppm which are likely to correspond to Trp side-chains Ne-He ). This indicates a complex formation between p53 DBD and MAGE-A4, causing line-broadening due to increase in the size of the molecule and a slower tumbling rate, and thus reduction in signals. The reduction in the signal intensity may also be due to intermediate exchange regimen in the NMR timescale. Again the changes in the signals are observed in those that correspond to residues in a folded region but not those in unstructured region. Thus it is expected that this interaction is mediated by the MHD of MAGE-A4 and not the N-terminal, which is predicted to be disordered.
Protein-protein Interactions of MAGE-A4 with gankyrin and p53

\[\text{\textvisiblespace}\]

\[\text{\textvisiblespace}\]
6. Protein-protein Interactions of MAGE-A4 with gankyrin and p53

Figure 6.3 $^1$H-$^{15}$N HSQC NMR probing protein-protein interaction between MAGE-A4 and p53 DBD. The spectrum of the $^{15}$N labelled MAGE-A4 protein alone is shown in black (a) while the one obtained after titrating p53 DBD is shown in red (b). The two spectra are overlaid to see the difference (c). Both labelled MAGE-A4 and p53 DBD proteins were prepared in 50 mM NaPi buffer (pH 7.2) containing 50 mM NaCl and 5 mM DTT. The concentration of AT MAGE-A4 and p53 DBD were 156 μM and 235 μM, respectively. The spectra were recorded at 25 °C.

6.5.4 Protein-protein interaction between MAGE-A4 and p53 probed by nESI-MS and IM-MS

Native nESI-MS was employed to probe the MAGE-A4:p53 interaction and characterise its stoichiometry (Figure 6.4). Individual spectra of MAGE-A4 and p53 DBD are shown in Appendix E Figure S 26 and Figure S 28, respectively. p53 DBD presents in a bimodal distribution with peaks attributable to charge states ($z$) $7 \leq z \leq 19$. Low intensity dimeric species [$D+15H$]$^{15+}$ is also observed. MAGE-A4 and p53 DBD were mixed at 1:1 molar ratio with a concentration of 10 μM each. The mass spectrum in Figure 6.4a shows both proteins predominantly present as monomers, with low abundance MAGE-A4:p53DBD complex. The MAGE-A4:p53DBD complex presents in a 1:1 stoichiometry with signals corresponding to [MA4:p53+14H]$^{14+}$, [MA4:p53+15H]$^{15+}$, [MA4:p53+16H]$^{16+}$ and [MA4:p53+17H]$^{17+}$. We also observe some low intensity trimeric species, comprising dimeric MAGE-A4 in complex with p53 monomer (charge states [2MA4:p53+19H]$^{19+}$, [2MA4:p53+20H]$^{20+}$ and [2MA4:p53+21H]$^{21+}$). The concentration of each protein was altered from 10 to 100 uM and at different ratios, no change was seen in the relative intensity of the complex indicating again a $K_d$ that is > 50 μM. The relative amount of complex also did not alter with incubation time, indicating that the kinetics of binding is fast (data not shown).

DT-IM-MS was performed and CCSDs derived from arrival time distributions recorded. The CCSDs for MAGE-A4:p53 DBD complex at charge states [MA4:p53+14H]$^{14+}$ - [MA4:p53+17H]$^{17+}$ are shown in Figure 6.4b. The complex presents as a single broad conformational family centered at $^{DT}CCS_{ne} \sim 3300$ Å$^2$ at all charge states. We observe a slight widening of the CCSD at the higher charge states but see no evidence for a resolved second conformational family.
6.5.5 Interaction studies of MAGE-A4 with p53 downstream pro-apoptotic and cell cycle arrest DNA promoters using EMSA

As we observe an interaction between MAGE-A4 and p53, we questioned whether MAGE-A4 displaces p53 downstream gene promoter DNAs from p53, which may result in reduced expression of such genes. Before investigating this question, we carried out EMSA studies to first test whether MAGE-A4 is capable of binding with such DNAs, which may also lead to reduction in the expression of p53 downstream genes. We used 30-mer dsDNA promoters of BAX and PUMA as models of pro-apoptotic response element and p21 as a cell cycle arrest model (Figure 6.5). The DNA concentration was maintained constant (2 μM), while the protein was titrated so that its concentration increases by twice. The EMSA data show that p53 is capable of binding with all of these promoter dsDNAs, as can be observed from
the slower migration of these DNAs upon titrating p53. However titration of MAGE-A4 did not affect the migration of the DNAs in the gel, ruling out the possibility of interaction between MAGE-A4 and such promoter DNAs.

**A**

<table>
<thead>
<tr>
<th>Bax (µM)</th>
<th>2</th>
<th>2</th>
<th>2</th>
<th>2</th>
<th>2</th>
<th>2</th>
<th>2</th>
<th>2</th>
<th>2</th>
<th>2</th>
<th>2</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>protein (µM)</td>
<td>0</td>
<td>2</td>
<td>4</td>
<td>8</td>
<td>16</td>
<td>32</td>
<td>64</td>
<td>0</td>
<td>2</td>
<td>4</td>
<td>8</td>
<td>16</td>
</tr>
</tbody>
</table>

**B**

<table>
<thead>
<tr>
<th>PUMA (µM)</th>
<th>2</th>
<th>2</th>
<th>2</th>
<th>2</th>
<th>2</th>
<th>2</th>
<th>2</th>
<th>2</th>
<th>2</th>
<th>2</th>
<th>2</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>protein (µM)</td>
<td>0</td>
<td>2</td>
<td>4</td>
<td>8</td>
<td>16</td>
<td>32</td>
<td>64</td>
<td>0</td>
<td>2</td>
<td>4</td>
<td>8</td>
<td>16</td>
</tr>
</tbody>
</table>
Figure 6.5 EMSA studies to test whether MAGE-A4 is capable of binding to p53 pro-apoptotic and pro-cycle arrest elements: (A) BAX, (B) PUMA and (C) p21. The protein and DNA concentrations are indicated. The triangle symbol indicates an increase in concentration of the respective protein.

6.5.6 The Effects of MAGE-A4 on the DNA-binding ability of p53 DBD, analysed using EMSA

As we could not observe the interaction between MAGE-A4 and p53 response element promoters (BAX, p21 and PUMA), we next investigated whether interaction between MAGE-A4 and p53 DBD blocks the DNA-binding of p53 DBD and thus displaces DNA from p53 DBD. p53 DBD and p53 response element promoters (BAX, PUMA and p21) were mixed at constant concentrations (4 μM and 1 μM) to form p53 DBD-DNA complexes, and increasing concentration of MAGE-A4 was titrated to each complex. The results are shown in Figure 6.6. It was observed that MAGE-A4 displaces the DNA from p53 DBD upon titration. The displacement is obvious when MAGE-A4 is over 16-fold in molar excess relative to p53 DBD.
Protein-protein Interactions of MAGE-A4 with gankyrin and p53

Figure 6.6 EMSA studies to examine the effects of MAGE-A4 on p53 DBD:DNA complexes. MAGE-A4 was titrated to A) p53 DBD-Bax complex, B) p53 DBD:PUMA complex and C) p53 DBD-p21
complex. The concentrations of the proteins and DNAs are indicated. The triangle symbol indicates increase in the concentration of MAGE-A4.

6.5.7 Protein-DNA interaction studies using nESI-MS

p53 DBD binding affinity and stoichiometry to DNA promoter sequences p21, Bax and PUMA was probed. Protein and DNA were incubated in increasing ratios to reach an equimolar concentration (20 µM:20 µM) (Figure 6.7a, Appendix D Figure S 29a, Figure S 30a). We observed monomeric p53, free DNA fragments and binding of p53 DBD to each DNA promoter sequence as a 4:1 p53:DNA ratio, in accordance with previous studies [27-29]. The complexes present in charge states $18 \leq z \leq 22$ for p21 and Bax DNA promoter sequences and $17 \leq z \leq 21$ for PUMA DNA promoter sequence. Whilst free p53 DBD and dsDNA was observed at all ratios, increasing ratios of DNA resulted in an increased intensity of the complex peaks. MAGE-A4 was incubated with DNA promoter sequences in increasing ratios to equimolar ratio (20 µM:20 µM). The nESI mass spectra produced (Figure 6.7b, Appendix E Figure S 29b and Figure S 30b) show no interaction between MAGE-A4 and DNA promoter sequences p21, Bax, and PUMA at any ratio. We observed a raised baseline, likely due to undesolvated DNA fragments.
Protein-protein Interactions of MAGE-A4 with gankyrin and p53

Figure 6.7 nESI mass spectra showing a) p53 DBD incubated in increasing ratios up to 1:1 and concentration 20 µM:20 µM with p21 DNA promoter sequence. Highlighted blue m/z region depicts p53 DBD monomeric species and free dsDNA, which are further denoted using blue spheres and DNA helix cartoon, respectively. Red highlighted region indicates peaks corresponding to a 4:1 p53 DBD:p21 DNA complex. b) MAGE-A4 incubated in increasing ratios up to equimolar concentrations (20 µM:20 µM) with p21 DNA promoter sequence. MAGE-A4 monomeric species are denoted by single Tilda symbols, dimeric MAGE-A4 by double Tilda symbols. Several charge states have been labelled.

Native MS was employed to probe the binding between MAGE-A4 and p53 DBD in competition with the p53 DBD:DNA interaction. Figure 6.8 shows nESI mass spectra of p53 DBD and p21 DNA promoter sequence held at equimolar concentration (30 µM:30 µM) with increasing concentrations of MAGE-A4. We observe a decrease in the intensity of the p53 DBD:DNA complex upon increase in ratio of MAGE-A4 up to a 1:1:3 p53 DBD:DNA:MAGE-A4. We do not directly observe a MAGE-A4:p53 DBD complex. The trend is also observed for PUMA DNA promoter sequence, although the effect is less pronounced.
Protein-protein Interactions of MAGE-A4 with gankyrin and p53 (Appendix E Figure S 30). Interestingly, MAGE-A4 addition appears to completely ablate the p53:Bax DNA interaction even at a low concentration (Appendix E Figure S 29).

Figure 6.8 nESI mass spectra showing MAGE-A4 inhibition of p53 DBD:DNA interaction. Concentrations of p53 DBD and p21 DNA promoter sequence were kept constant (30 µM), with increasing concentration of MAGE-A4 added. Black Tilda corresponds to MAGE-A4 signals, blue
Protein-protein Interactions of MAGE-A4 with gankyrin and p53

sphere corresponds to p53 DBD signals, and free dsDNA depicted by cartoon DNA helix. Red highlighted m/z region shows peaks corresponding to p53 DBD:p21 DNA complex.

6.6 Discussion

The first discovered MAGE protein was MAGE-A1 which was identified as an antigen recognised by cytotoxic T lymphocytes in melanoma patients [30]. The gene of MAGE-A1 was later cloned successfully [31, 32]. Since then, more than 50 genes of MAGE family have been identified. MAGE protein family consists of two subfamilies: type I and type II. While type II subfamily is expressed in somatic cells, the expression of the type I subfamily is restricted to cancer cells and germlines. Due to its expression pattern, type I MAGEs have been considered as a potential target for cancer immunotherapy. For example GlaxoSmithKline attempted to develop a cancer immunotherapy using peptides derived from MAGE-A3 [33].

The expression of type I MAGEs is correlated with cancer malignancy and poor patient prognosis. Although the functional mechanism of MAGE proteins in cancer cells is poorly understood, MAGE-A2 has been revealed to interact with p53 (specifically with its DNA binding domain), suppressing its transactivational functions. Contrary to other type-I MAGE proteins, tumour suppressor activity was also reported for MAGE-A4. For example MAGE-A4 was observed to promote apoptosis in cancer cells [21]. In addition, MAGE-A4 was identified as an interacting partner of the oncoprotein gankyrin, suppressing its tumorigenic activity [22]. In addition, as part of its oncogenic activity, MAGE-A4 has been reported to suppress p53 downstream pro-apoptotic and pro-cell cycle arrest genes [20]. In order to elucidate these two contradictory activities for MAGE-A4, we employed NMR techniques alongside nESI-MS and the complementary technique DT-IM-MS to probe the interaction of MAGE-A4 with gankyrin and p53 DBD. Our results show that MAGE-A4 can interact with both gankyrin and p53 DBD. Although it is unknown which region or domain in MAGE-A4 is involved in these interactions, the NMR signals which exhibited changes upon titrating gankyrin or p53 DBD correspond to residues located in a folded domain or region, whilst the signals arising from unfolded region did not show any shift or change in the intensity. We therefore hypothesise that the folded MHD is responsible for these
interactions as the rest of the protein (mainly the N-terminal ≈100 amino acid residues) is predicted to be unfolded.

Native MS shows that whilst the majority of protein remains as unbound monomeric species, MAGE-A4 is able to form a low intensity complex with gankyrin and p53 DBD. Proteins with regions of disorder often exhibit a ‘one-to-many’ mechanism of binding, undergoing high specificity but low affinity binding events to multiple partners [34, 35]. Native MS has previously been used to observe low affinity interactions of disordered proteins [36, 37]. MAGE-A4 predominantly interacts with p53 DBD and gankyrin as a 1:1 complex. We observe low intensity trimeric species incorporating both dimeric MAGE-A4 and dimeric gankyrin, showing the dimeric forms of each protein, observed in the individual spectra, can bind. In contrast, we only observe trimeric species between MAGE-A4 and p53 DBD with dimeric MAGE-A4 but monomeric p53 DBD. This suggests that although we observe low intensity [M+17H]^{17+} dimer in the p53 DBD spectra (Appendix E Figure S 28), this does not interact with MAGE-A4.

IMMS analysis reveals that the MAGE-A4:gankyrin complex presents in at least two conformational families, suggesting some level of flexibility retained in the complex. We observe the retention of conformational families across several charge states, indicative of stable structures that do not increase in Δ^1 CCS_Hi with charge. IM-MS analysis shows that the MAGE-A4:p53 DBD complex presents in a single conformational family across all four charge states observed (Figure 6.4b). Low charge states of monomeric MAGE-A4 ([M+10H]^{10+}) and p53 DBD ([M+7H]^{7+}), have been shown to present at Δ^1 CCS_Hi ~2300 Å^2 and ~1500 Å^2 respectively [25]. The complex presents at Δ^1 CCS_Hi ~3200 Å^2 (Figure 6.4b) suggesting a relatively compact complex is formed by the two monomers. The lack of CCS increase across heterodimer charge states observed also suggests a stable structure; although we do note that the CCSDs are broad, indicating some dynamics within the complex. From these results, we can infer that the mostly folded proteins gankyrin and p53 DBD have a structuring effect upon the much more disordered MAGE-A4 protein when binding.

To further investigate the impact of MAGE-A4 on the DNA-binding ability of p53 DBD, we carried out in vitro EMSA studies using BAX and PUMA promoter DNAs as pro-apoptotic p53 response elements as well as p21 promoter DNA as a cell cycle arrest p53 response element. We titrated increasing amounts of MAGE-A4 into preformed p53DBD:DNA complexes. Our data showed that increasing amount of MAGE-A4 displaces the pro-
Protein-protein Interactions of MAGE-A4 with gankyrin and p53

apoptotic promoters (BAX and PUMA) as well as pro-cell cycle arrest promoter (p21). These results were further supported by nESI-MS experiments, which first showed a 4:1 stoichiometry for p53 DBD:DNA but no association between MAGE-A4 and any DNA promoter sequence (Figure 6.7, Appendix E Figure S 29, Figure S 30). We observe strong binding for both p21 and PUMA DNA promoter sequences, but weaker interaction for Bax promoter sequence. nESI-MS was then used to track alterations in p53:DNA complex in response to an increase in MAGE-A4 concentration. For both p21 and PUMA promoters we observe a smooth decrease in p53:DNA complex intensity with increasing MAGE-A4 (Figure 6.8, Appendix E Figure S 30c), however for Bax DNA sequence we observe total loss of the p53:DNA complex upon incubation with minimal MAGE-A4 (Appendix E Figure S 29). We hypothesise that MAGE-A4 may be able to disrupt this weaker interaction more readily. Interestingly, we observe no MAGE-A4:p53 DBD complex signals upon loss of the p53:DNA complex, indicating that the corresponding peaks were either not resolved due to the raised baseline, or the complex is not retained under the experimental conditions. These findings demonstrate that excess amount of MAGE-A4 can abrogate the DNA-binding capacity of p53 DBD. These results are consistent with the previous report that overexpression of MAGE-A4 promotes tumour growth while suppressing p53 downstream genes such as BAX and p21 [20].

Based on our data, we propose the following model to explain the biological functions of MAGE-A4 (Figure 6.9). By interacting with gankyrin, MAGE-A4 acts as a tumour suppressor by suppressing its tumorigenic activity. On the other hand, MAGE-A4 can function as an oncoprotein by binding with p53, disrupting its interaction with its response elements (such as BAX, PUMA and p21), and thus its transactivational functions. The results presented here results give the first insights into the biological functions of MAGE-A4, which may be exploited through further study in anti-cancer therapy.
Protein-protein Interactions of MAGE-A4 with gankyrin and p53

Figure 6.9 Proposed model for the biological functions of MAGE-A4.

6.7 References


208
Protein-protein Interactions of MAGE-A4 with gankyrin and p53


Protein-protein Interactions of MAGE-A4 with gankyrin and p53


7

Conclusions and Outlook
Conclusions and Outlook

Whilst the important relationship between protein structure and protein function has always been a central tenet in the study of biological systems, understanding the cellular roles of protein regions which lack any stable structure has become a focus of research over the last 10 years. The study of intrinsically disordered protein (IDP) structure (or lack thereof), dynamics, interaction partners and mechanisms of binding has expanded dramatically over recent years, predominantly due to their implication in a wide range of diseases and pathologies. The work of this thesis was carried out with the aim of using mass spectrometry (MS) and the hybrid technique ion mobility mass spectrometry (IM-MS) to probe the structure and conformational dynamics of several IDPs implicated in tumourgenesis and the propagation of cancer, along with several of their known or postulated cellular interactors and small molecule inhibitors. IM-MS has been demonstrated to be a useful tool for the elucidation of the conformers adopted by IDPs; in contrast to other biophysical techniques it is able to report on multiple co-existing conformers, lacks bias towards the folded state of a protein, and can be used to probe an isolated protein in a heterogeneous sample with no interference. The highly hydrophobic environment of the mass spectrometer allows the intramolecular interactions of a protein that influence conformation to be probed in the absence of solvent. IM-MS measurements provide information on the size and shape of proteins in the form of a rotationally averaged collision cross section (CCS). CCS values derived from IM-MS measurements are complementary to other biophysical techniques, allowing comparisons to be made to coordinates from NMR and x-ray crystallography experiments and structures produced by molecular modelling. In this way, we are able to link our gas-phase IM-MS conformations with those which may be present in solution.

This thesis begins with a study in Chapter 2 of the gas-phase unfolding of several proteins of increasing intrinsic disorder; cytochrome c, p53 DNA binding domain, and the N-terminal domain of MDM2. Cytochrome c is a model protein of stable structure coordinated by a heme group. The transcription factor p53 binds to specific DNA promoter sequences via its DNA-binding domain (p53DBD), a mostly folded domain which contains multiple flexible loops giving it substantial flexibility. The N-terminal domain of the E3 ubiquitin ligase Mdm2 contains a disordered “lid” region which exists in multiple conformations. Variable temperature IM-MS was used to probe the conformations of these three proteins at buffer gas temperatures ranging from 200 K to 571 K. Multiple conformations were observed at room temperature for each protein. Cytochrome c underwent a charge state dependent unfolding pathway; the lower charge states retained compact structures, whereas the
Conclusions and Outlook

Higher charge states firstly underwent compaction and then extension before a sharp unfolding event above 450 K. p53DBD followed a similar pathway, however unfolded in a much smoother manner and was observed to aggregate above its $T_{\text{melt}}$. Mdm2 also unfolded in a charge state dependent manner; low charge states first collapsed slightly above room temperature, before extending significantly at higher temperatures. Higher charge states unfolded smoothly with temperature increase before a significant compaction was observed at the highest temperatures, presumably into a molten globule state. The intermediate conformational states occupied by these proteins during their temperature-dependent unfolding were captured by IM-MS, whereas they may not be observable in solution.

Chapter 3 contains the analysis of the N-terminal domains of both p53 (Np53) and Mdm2 (N-MDM2) using several biophysical techniques. Np53 was shown by MS, IM-MS, hydrogen deuterium exchange coupled to MS (HDX-MS) and circular dichroism (CD) to be highly disordered. Np53 was then incubated with RITA (reactivation of p53 and induction of tumour cell apoptosis), a small molecule ligand shown previously to interact with p53 and inhibit the p53:MDM2 interaction. The complex between Np53 and RITA could not be preserved, indicating a high $K_d$, and HDX-MS analysis showed no alterations to the uptake of deuterium by Np53 in the presence of RITA. Interestingly however, CD showed an increase in secondary structure and IM-MS revealed a significant conformational compaction of Np53 at all charge states upon incubation with RITA, despite no retention of the bound form. This templating effect allowed the conformational effect of the ligand on the protein to be observed, even when the binding to RITA was lost upon desolvation. The conformational spread of N-MDM2 was then probed by MS and IM-MS and again showed the dynamic nature of this domain. When incubated with known p53:MDM2 inhibitor Nutlin-3, peaks corresponding to a 1:1 complex and a shift in the charge state distribution of N-MDM2 were observed. Again, a compaction of N-MDM2 when binding to Nutlin-3 could be observed by IM-MS and the templating effect was noted for free N-MDM2 in the presence of Nutlin-3. This study provided an excellent example of how IM-MS could be used to screen for drug compounds from a library; it is able to report not only on binding affinity and stoichiometry, but on the structural impact of ligand binding, which may be used to guide subsequent small molecule design.

In Chapter 4 the conformational and dynamic properties of wild-type and mutant constructs of the p53 transactivation domain (p53TAD) are compared using MS and IM-MS.
In this study, the conformational effects of proline replacement with alanine are investigated. Proline has been postulated to be partially responsible for controlling the degree of disorder adopted by IDPs, and mutation of prolines which flank the p53TAD residues which interact with Mdm2 has been previously shown to increase the helicity of the domain. These mutants were studied by MS and IM-MS. MS revealed minimal differences to the charge state range and distribution of the p53TAD mutants; however IM-MS provided evidence of interesting conformational alterations. For all constructs a compact node containing two conformers and a more extended node encompassing several intermediate and extended conformers were observed. Upon proline replacement we observed a decrease in the population of extended conformers and the increase in the proportion of ions presenting in a specific intermediate conformer. This was most distinct for the mutant with all prolines replaced (p53\textsuperscript{PallA}), confirmed by CD and NMR to have a high helical content, which led to the postulation that this conformer was mainly helical. In order to compare our IM-MS data with solution-phase data, computational calculations were performed on several ensembles of model structures trained on NMR and SAXS experimental data. The CCSs of these structures mapped well to the extended node of our IM-MS data, suggesting these conformers are present in solution. The compact conformers may be kinetically trapped during the gas-phase experiments, forming molten-globule like states. Future work will detail molecular modelling results for the p53\textsuperscript{PallA} to attempt to recreate the significant conformational changes computationally. It will be interesting to observe if the computational ensembles are able to delineate between conformers to the extent that we observe using IM-MS.

Chapter 5 is a continuation of the study discussed in Chapter 4, containing analysis of the same mutant p53TAD constructs. Here, the binding of both WT and mutant p53TAD to Mdm2 is investigated using MS and IM-MS. Here, we observed that the increased helicity of the mutant p53TAD constructs correlates with an increased binding affinity for Mdm2 compared with WT p53. The mutants were shown to form complexes of the same CCS as the WT protein, although a difference between the conformational preferences for binding was observed between mutants. Here, IM-MS is used as a diagnostic tool to track the conformers depleted from the free p53TAD population upon incubation with increasing Mdm2 concentrations. This type of experiment could be useful in investigations of binding mode for small molecule ligands or protein interaction partners; alterations to the design of the ligand/therapeutic may result in a change to the active conformation targeted for binding. This can be monitored quickly by IM-MS to determine if a specific structural
change, for example the alteration of protein conformation such that it can no longer bind to an interaction partner, has been reached.

The conformational analysis of melanoma-associated antigen 4 (MAGE-A4) is presented in Chapter 6. Structural information regarding this protein is limited; the structure of the central homology domain (MHD) has been solved, however the full length protein is thought to be disordered. A recent study, included in the Appendices, used NMR, CD, MS, IM-MS and molecular modelling to demonstrate the significant flexibility of MAGE-A4 protein both in solution and in the gas-phase. In Chapter 6, the protein:protein interactions of MAGE-A4 are probed. MS is used to show the weak interaction of MAGE-A4 with p53DBD and the oncoprotein gankyrin, as well as the conformations of the complexes formed which are shown by IM-MS to be stable but with some dynamics. Here, a significant folding-upon-binding event of the large disordered MAGE-A4 is shown, forming a stable complex. MS was then employed to show the competitive inhibition of the p53DBD-DNA interaction by MAGE-A4 for several promoter sequences.

Whilst MS is becoming more widely used for the high-throughput study of small molecules, expansion of IM-MS to become a more high throughput method for protein study would enable the structural impacts of ligand binding to macromolecules to be investigated rapidly, with minimal sample volume and at low concentrations. This could be used as an initial characterisation of the ligand binding mechanism, whilst simultaneously reporting on the kinetics and stoichiometry of the interaction. There are several platforms for high throughput analysis which can be coupled to mass spectrometers, including the Advion NanoMate, which allows automated ionisation of a large number of samples.

For larger, more complicated protein systems, it is feasible for IM-MS to become a routine technique in the initial characterisation of these proteins and their complexes. As shown in Chapters 3 and 6, a single IM-MS experiment can provide information regarding individual protein dynamics in a heterogeneous sample, protein:protein interaction affinity, conformational preferences in binding, determine complex size and monitor structural changes that may occur in the individual monomers upon binding. Whilst techniques such as NMR and X-ray crystallography are likely to remain the gold standard for protein characterisation due to their atomic resolution, disordered proteins are often not amenable to study by x-ray crystallography due to their inherent flexibility, and are not available in great enough quantities for study by NMR. This positions IM-MS as an excellent
initial or alternative technique able to provide low resolution structural information to guide the use of further techniques.

The hybrid technique IM-MS has been shown to provide informative data regarding the structure and dynamics of flexible proteins in the absence of solvent. The conformational diversity of an individual protein, and alterations to this conformational landscape as a result of a number of intrinsic or extrinsic factors, can be readily revealed by IM-MS, providing much needed insights into complex and unstable systems of biological or pharmaceutical relevance.
Appendix A

Supplementary information for Chapter 2.

Cytochrome c

As the final component of the electron transport chain, cytochrome c functions to accept electrons from cytochrome bc$_1$ via its heme group before transferring to the cytochrome oxidase complex in the mitochondrial respiratory chain. A ~12 KDa globular protein, cytochrome c contains four α-helices surrounding a Heme molecule coordinating an Iron atom. Cytochrome c can be considered one of the most extensively studied proteins by mass spectrometry [1] due to its ease of availability and reliable ionisation. Cytochrome c is often used as a protein standard, providing a model system against which gas phase protein structure and folding data can be compared, and as such its folding pathways in solution are relatively well understood [2]. Gas phase studies thus far on cytochrome c have revealed a diverse range of conformations present at intermediate charge states ($z \geq 7$) [1, 3-6], highlighting the conformational heterogeneity present even in proteins with well-defined tertiary structures. Interestingly, cytochrome c purified from differing organisms shows dissimilar folding behaviour with very little change in structure [7].

p53

The transcription factor p53, dubbed the “Guardian of the Genome” [8], is a tumour suppressor protein implicated in a vast range of cellular processes including cell-
cycle arrest, DNA repair, metabolism and transrepression [9-11]. It plays a central role in tumour suppression pathways, blocking tumour development by triggering cellular senescence or apoptosis in the event of cellular stress signals resulting from DNA damage, oncogene activation or telomere erosion [11]. Cellular levels of p53 are closely regulated by MDM2 and its partner MDMX in an auto-regulatory feedback loop which targets p53 for degradation via the E3 ligase pathway. Further control of p53 is carried out via post-translational modifications and interactions with a number of signalling proteins [12].

The p53 protein is divided into multiple functional domains, the N-terminal transcriptional transactivation domain (residues 1-61), proline rich domain (residues 62-94), the central DNA-binding domain (DBD) (residues 94-292), the tetramerization domain (residues 325-355) and the C-terminal regulatory domain (residues 363-393). We study the thermal stability of the ‘core’ of p53 (residues 94-312) containing the DBD shown in Figure 2.1. Inactivation of p53 by mutation or suppression is an almost universal marker of human cancers, with over 95% of tumour-derived mutations mapped to the ‘core’ of the protein [13]. In solution this domain is structured as a highly unstable β-sandwich composed of two antiparallel β-sheets with a small β-hairpin, with a low solution melting temperature = 315-317 K (~42-44°C) [14].

Cloning of wild-type human p53 DBD (a.a. residues 94-312) was performed using pRSET(A) vector [15, 16]. The protein was expressed at 37°C in Escherichia coli C41 (DE3) cells (Avidis) and grown up to an O D600 ~0.8. The temperature was reduced to 22°C before cell cultures were induced with 1 mM isopropyl β-D-thiogalactoside (IPTG) overnight. Cells were harvested at 4°C and lysed using Bug Buster Protein Extraction Reagent (Novagen) supplemented with EDTA-free
protease inhibitor tablets (Roche) and benzonase nuclease (Novagen) [13, 17]. The soluble fraction was loaded onto an SP Sepharose column (GE Healthcare) and eluted with a NaCl gradient followed by gel filtration chromatography using HP26/60 Superdex 200 column (GE Healthcare). Protein samples were snap-frozen in liquid nitrogen and stored in 25 μL aliquots at -80 ºC. Before analysis, the protein samples were thawed and dialysed in 50mM ammonium acetate for two hours at 4°C using Slide-A-Lyzer dialysis cassettes (Fischer Scientific, UK). Concentration of purified proteins were measured by the Thermo Scientific NanoDrop Spectrophotometer ND 1000 (Thermo Scientific, USA) and calculated using the known mass extinction coefficients (for p53 DBD ε280 = 17130 cm−1·M−1).

MDM2

MDM2 is a multidomain, multifunctional 55 KDa IDP with roles as an E3-ubiquitin ligase, molecular chaperone and in translational control. The domains of MDM2 consist of an intrinsically disordered “lid” mini-domain (residues 1-24), the N-terminal domain, the disordered central acidic domain containing a nuclear localisation signal, the zinc finger domain and the C-terminal RING domain, with functions including co-ordinating the activity of MDM2 as an adaptor protein during E2-mediated ubiquitin transfer and dimerization with MDMX RING domain to reduce auto-ubiquitylation and stabilize MDM2 [18].

In this study we utilize the N-terminal 126 residues of MDM2 (NT-MDM2) comprising the structured N-terminal domain and the intrinsically disordered “lid” mini-domain. The N-terminal is a globular domain containing a hydrophobic pocket, the binding site for the transcriptional transactivation domain of p53, targeting the tumour suppressor for degradation by the proteasome [19, 20]. The disordered “lid”
mini-domain has been previously shown to exist in multiple interconverting conformations [21] and functions to regulate the proteins E3 ligase function by switching between a ‘closed’ and ‘open’ conformation [22, 23]. This conformational change reveals the N-terminal hydrophobic pocket for binding to p53. The disordered nature of the “lid” mini-domain has prevented inclusion of this region in the crystal structure of the MDM2 N-terminal domain. This partially disordered domain has been shown to have a $T_{\text{melt}} = 348$ K [22].

Human NT-MDM2 (residues 1-126), provided by Dr. E. Worrall, was kept at -80°C. Expression and purification has been previously described [22]. Before analysis, the sample was thawed and dialysed using BIO-RAD micro bio-spin chromatography columns (Bio-Rad Laboratories, Inc) containing Tris buffer at 4 °C. Tris buffer was exchanged to 50 mM Ammonium Acetate (AmAc) pH 6.8, prepared with LC-MS grade H$_2$O at 4 °C, before 20-75 µM of sample was passed through the buffer-exchanged chromatography column. The concentration of protein was measured by a NanoDrop Spectrophotometer ND 1000 (Thermo Scientific, USA) and calculated using the known mass extinction coefficients (for MDM2 $\epsilon_{280} = 51380 \text{ cm}^{-1} \cdot \text{M}^{-1}$).

**IM-MS Further Experimental Details**

The simplest IM-MS set-up is that of a linear drift time ion mobility coupled to mass spectrometry (DT IM-MS). Ions are separated by their mobility (k) by traversing a drift cell of known length filled with buffer gas at a known temperature and pressure. The electrostatic force generated by a weak electric field (5-50 V cm$^{-1}$) provides the forward motion of the ions, whilst collisions with buffer gas molecules retard the progress of ions until they reach an equilibrium drift velocity that is proportional to the electric field. The mobility of an ion can be given as the ratio between the drift
velocity (vd) and applied electric field (E). This can be converted to the rotationally averaged CCS (Ω, Å2) using a rearrangement of the following equation [19];

Equation 1.7

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

Where z, ion charge state; e, elementary charge; N, gas number density; \( \mu \), reduced mass of the ion (kg); \( k_B \), Boltzmann constant (1.381 e10-23 J K-1); T, temperature (K); \( K_0 \), reduced mobility (measured mobility corrected to standard pressure and temperature; 760 Torr and 273.15 K, respectively).

**Supplementary Figures**
Figure S 1 a) CCSD for cytochrome c [M+5H]⁵⁺ at increasing buffer gas temperature. As no [D+9H]⁹⁺ or [D+11H]¹¹⁺ is present in the spectra at 571 K, we assume the CCSD for [M+5H]⁵⁺ at 571 K comprises monomer only. As such, this enduring monomeric species is mapped as highlighted by hatched Gaussian distributions at all temperatures. In doing this we have made the assumption that the dominant species in the ATD at 300 K is monomer at ~1200 Å², and that a substantial portion of this conformational family remains so at all temperatures. This is the finding of Mao et al. (Figure S 1b) under conditions where they have no evidence in the mass spectrum for any dimeric species. The coincident dimeric [D+10H]¹⁰⁺ and other monomeric conformational families are represented by the grey intensity in the figure. Based on our data for the [D+11H]¹¹⁺ charge state dimer (main text) we would suggest that the dimer is not well separated from the monomer at room temperature. Lower x-axis denotes CCS in Å² for [M+5H]⁵⁺ monomer and upper x-axis denotes CCS in Å² for [D+10H]¹⁰⁺. b) CCSD for cytochrome c [M+5H]⁵⁺ at increasing buffer gas temperature as taken by Mao, et al. [4]. It is of note that the injection energies used by (1500eV) are significantly higher than used in this study, however Mao, Y. et al. report no change in CCS distributions at lower injection energies. Part b is reprinted with permission from ref 4. Copyright 1999 American Chemical Society.
Figure S 2 nESI mass spectra of 50 µM DBD p53 sprayed from 50 mM Ammonium Acetate at increasing buffer gas temperatures. Monomeric species are denoted by a single grey sphere.
Figure S 3 nESI mass spectra of 50µM N-terminal MDM2 sprayed from 50 mM Ammonium Acetate at varying drift cell temperatures. Monomeric species are denoted by a single grey sphere, dimeric species by two grey spheres.
Figure S4 CCS distributions at increasing buffer gas temperature for MDM2 [M+13H]$^{13+}$ and [M+15H]$^{15+}$ species, respectively.

References for Appendix B

Appendix A

Appendix B

Supplementary Information for Chapter 3

**IM-MS Theory**

Ion mobility coupled to mass spectrometry (DT IM-MS) was utilised to distinguish the conformational families present in Np53. IM-MS separates analytes on the basis of their mass (m), charge (z) and shape to give a rotationally averaged collision cross section (CCS, \( \Omega, \text{Å}^2 \)). Ions are separated by their mobility (k) as they pass through a drift cell filled with a buffer gas of known temperature and pressure. A weak electric field (E) (5-50 V cm\(^{-1}\)) provides the forward motion of the ions along the drift cell, whilst collisions with buffer gas molecules slow the progress until ions reach a constant drift velocity (\( \nu_d \)), allowing calculation of the mobility using Equation S 1:

\[
\text{Equation S 1 } \quad \nu_d = KE
\]

The mobility of an ion is often expressed as reduced mobility (\( K_0 \)) which is the measured mobility K standardised for temperature (T), 273 K and pressure (P), 760 Torr. Using this parameter, we can calculate the buffer gas dependent CCS using:

\[
\text{Equation 1.7 } \quad K_0 = \frac{3ze}{16N} \left( \frac{2\pi}{\mu k_B T} \right)^{0.5} \frac{1}{\Omega}
\]

Where \( K_0 \) is the reduced mobility, \( z \) 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 Boltzmann constant, \( T \) is the gas temperature and \( \Omega \) is collision cross section.
The raw arrival time \( t_a \) of an ion includes the time the ion spends outside the drift cell but within the mass spectrometer, also known as the dead time \( t_0 \). This can be calculated by taking an average of the intercept of a linear plot of average arrival time versus pressure/temperature and the drift time can be calculated using Equation 1.8:

\[
\text{Equation 1.8} \quad t_D = t_a - t_0
\]

The CCS of a protein conformer can be used in comparison to coordinates obtained from biophysical techniques such as X-ray crystallography or NMR and from co-ordinates generated from computational studies [1-3].

**HDX-MS Theory**

Hydrogen deuterium exchange coupled to mass spectrometry (HDX-MS) was used to monitor solution phase conformational changes in Np53 in the absence and presence of RITA. In a HDX-MS experiment, the target protein is incubated with excess D\(_2\)O labelling buffer and on-exchange of solvent deuterons occurs for a set length of time. The continuous on-exchange labelling is then followed by quenching with a low pH buffer. Subsequently, the target protein is digested using an acid stable protease such as pepsin. Desalting and separation of the peptides by reversed-phase LC is followed by analysis using electrospray ionisation (ESI) mass spectrometry. The quench and LC steps of the experiment are carried out at low temperature (~0 °C) to minimize the back-exchange of Deuterium to Hydrogen. The digestion is carried out at the optimum pepsin digestion temperature (~20 °C) but kept to a minimum of time. Analysis of the resulting peptides monitors the deuteration level as a function of time by calculation of the mass shift in the isotopic distributions of individual peptides, enabling production of uptake graphs. Exposure of a protein to a D\(_2\)O containing environment will allow both side-chain and backbone Hydrogens to exchange freely to Deuterium, increasing the mass of the protein by 1 Da. Regions of the protein which are protected from the D\(_2\)O solution, by ligand binding or the fold of the protein, exhibit lower levels or a slower rate of on-exchange compared with solvent accessible backbone amides, which exchange freely and quickly. This can be used to relate to conformational dynamics of the protein on the timescale of
the experiment. Uptake rates can be mapped onto protein structures created from x-ray crystallography or NMR data points.

Materials and Methods

Protein expression and purification

The expression and purification of MDM2 1-126 has been described previously [4]. The expression and purification of wild-type N-terminal p53 (Np53) purification has also been described previously [5, 6]. In brief both proteins were expressed as glutathione-S-transferase-tagged proteins and purified from soluble Escherichia coli lysates. Cells were lysed and the proteins cleaved off column. The GST tag was also cleaved from the protein and all samples were snap-frozen in liquid nitrogen and stored in 25 µL aliquots at -80 ºC prior to analysis described in the main text.

HDX-MS buffer and pH

The equilibrium buffer (pH 7) was made up as 5 mM di-potassium hydrogen phosphate and 5 mM potassium di-hydrogen phosphate prepared in H₂O. The Labelling buffer (pH 6.66) used 5 mM di-potassium hydrogen phosphate and 5 mM potassium di-hydrogen phosphate prepared in D₂O. For the low pH quench buffer (pH 2.66) 50 mM di-potassium hydrogen phosphate and 50 mM potassium di-hydrogen phosphate were prepared in H₂O. The pH of all buffers was measured with a Jenway 3505 pH meter.

Circular dichroism

Proteins were prepared to 50 µM using 50 mM ammonium acetate with 5% IPA. Protein samples were incubated with RITA which was diluted with ammonium acetate to 100 µM and an IPA concentration of 5%. Np53 proteins in the presence and absence of RITA were incubated for 30 minutes at 37 ºC prior to analysis.
Circular dichroism (CD) spectra were acquired using a Chirascan CD spectrometer (Applied Photophysics, Ltd.) in the far UV region 180-260 nm with a spectral bandwidth of 1 nm, 0.5 nm step size and 0.1 mm path length. Three repeats of all experiments were taken and an average used. Data was analysed using Chriascan v 4.2.17 (Applied photophysics Ltd), Pro Data viewer (Applied photophysics Ltd) and CDNN v2.1 (Gerald Böhm).

Supplementary Figures
Appendix B

Figure S5 Collision cross section distribution (CCSD) derived from arrival time distributions (ATDs) for the [M+5H]$^{9+}$ charge state of wild-type Np53. Protein was analysed in the absence (top panel) of RITA and in a 1:2 protein:ligand ratio with RITA (bottom panel). Both samples were incubated for 30 minutes at 37°C with 5% IPA. CCSDs were taken at a drift voltage of 35 V. Hatched Gaussian curves indicate conformational families $C_0$ (red) and $C_1$ (blue). CCSD intensity is normalized to the intensity of the ion peak in the mass spectrum.
Figure S 6 Collision cross section distribution (CCSD) derived from arrival time distributions (ATDs) for the [M+9H]9+ charge state of wild-type Np53. Protein was analysed in the absence (top panel) of RITA and in a 1:2 protein:ligand ratio with RITA (bottom panel). Both samples were incubated for 30 minutes at 37°C with 5% IPA. CCSDs were taken at a drift voltage of 35 V. Hatched Gaussian curves indicate conformational families present; C1, X,U and U2 in blue, green, purple and gold respectively. CCSD intensity is normalized to the intensity of the ion peak in the mass spectrum, however [M+9H]9+ CCSD intensity is x10 to allow visibility of conformers present.

Figure S 7 Circular dichroism spectra of wild-type Np53 (incubated at 37 °C for 30 minutes with 5% IPA) and wild-type Np53 incubated with RITA (1:2 protein:ligand incubated at 37 °C for 30 minutes) in red and blue, respectively.
Figure S 8 nESI mass spectra of a) 50 µM N-MDM2 in 50 mM ammonium acetate b) 50 µM N-MDM2 + 0.5% DMSO and c) 50 µM N-MDM2 : 500 µM Nutlin-3 (0.5% DMSO). Single grey spheres denote monomeric species, double grey spheres denote dimeric species, and blue small spheres denote Nutlin-3 molecules.
Figure S9 Waterfall plot representing in vacuo conformations of N-terminal MDM2 + 0.5% DMSO sprayed from 50 mM ammonium acetate. The x, y, z axis show the collision cross section (DTCCSHe, Å^2), charge state (range: 5≤z≤14) and the relative intensity, respectively. DT IM-MS data taken at a drift voltage of 35V is shown.
References for Appendix C


Appendix C

Supplementary information for Chapter 4.

Molecular dynamics and Simulated Annealing

All molecular dynamics simulations were carried out using sander module in Amber14 with the amber99SBildn force field. Random structures were taken from the NMR and SAXS ensembles of WT p53TAD, representing a range of compact and extended protein conformations. Four different charge states of the protein were generated by neutralization of negatively charged amino acid residues; histidine residues were considered neutral. For each charge state, multiple (5-15) protomers were considered. After adding hydrogen atoms, the protein was minimised in vacuo using an infinite radial cut-off of 999 Å. The system was gradually heated to 300 K using a weak coupling coefficient of 2 ps$^{-1}$ with a time step of 2 fs. The SHAKE algorithm was used on all bonds involving hydrogen atoms. Following heating and equilibration, 20 ns of molecular dynamics was carried out in the absence of any cut-offs. Each of the protomers has then undergone simulated annealing process involving 300 rounds of repeated heating to 1200 K for 100 ps and cooling to 50 K for 50 ps, where the time step was reduced to 1 fs to ensure energy conservation. CCSs were calculated for structures extracted along the MD simulation and for each cooling cycle of the SA process. Models that exhibited small CCSs following MD and SA procedures were energy minimised to optimise the position of the side chains, reducing their CCSs by 1-3 %.

Structural features of the IDPs such as the secondary structure (DSSP), radius of gyration and solvent accessible surface area (SASA) were determined using cpptraj module from the Amber suite.

Principle Component Analysis compact conformer generation
PCA is a procedure often used to determine correlations between data sets and variables; here it was used to enhance conformational sampling of the compact states of the system. The Complementary Coordinates (“CoCo”) package was used to analyse the MD trajectories to determine the conformational variability of the ensemble and generate 100 new seeds for further short MD simulations. Each of the seeds was initially energy minimised and subjected to thermal equilibration at 300 K and to 2 ns of MD simulations. The PCA approach yielded significant improvements in conformational sampling of the compact states of the p53 protein, often inaccessible by MD and SA procedures.

**Supplementary Figures**

<table>
<thead>
<tr>
<th>Conformer CCS (Å²)</th>
<th>Denoted in text</th>
<th>Colour</th>
</tr>
</thead>
<tbody>
<tr>
<td>750</td>
<td>-</td>
<td>Grey</td>
</tr>
<tr>
<td>880</td>
<td>A</td>
<td>Red</td>
</tr>
<tr>
<td>990</td>
<td>B</td>
<td>Yellow</td>
</tr>
<tr>
<td>1100</td>
<td>C</td>
<td>Blue</td>
</tr>
<tr>
<td>1250</td>
<td>D</td>
<td>Green</td>
</tr>
<tr>
<td>1430</td>
<td>E</td>
<td>Magenta</td>
</tr>
<tr>
<td>1600</td>
<td>F</td>
<td>Orange</td>
</tr>
<tr>
<td>1800</td>
<td>G</td>
<td>Purple</td>
</tr>
</tbody>
</table>

Table S1 Centroid CCS values for all conformations assigned to experimental data.
Figure S 10 Positive Cα random-coil chemical shift deviations (ΔδCα) denoting levels of helicity within p53<sup>PtoA</sup>. X-axis shows residue number of the protein, Y-axis shows ΔδCα (left) and % helicity (right).

Figure S 11 nESI spectra in negative ionisation mode sprayed from 50 mM ammonium acetate of a) WT p53 b) p53<sup>P1213A</sup>, c) p53<sup>P27A</sup>, d) p53<sup>P3xA</sup> and e) p53<sup>PallA</sup>
Figure S 12 Collision cross section distributions derived from arrival time distributions for all charge states of WT p53 sprayed from 50 mM ammonium acetate in negative ionisation mode. Gaussian functions employed to represent conformational families are plotted as solid green lines, and the sum of Gaussian functions as a red line.
Figure S 13 Collision cross section distributions derived from arrival time distributions for all charge states of p53\textsuperscript{112,138} sprayed from 50 mM ammonium acetate in negative ionisation mode. Conformational families with centroid denoted in Table S1 are shown in green. Cumulative curve fits are shown in red.
Appendix C

Figure S 14 Collision cross section distributions derived from arrival time distributions for all charge states of p53\textsuperscript{P27A} sprayed from 50 mM ammonium acetate in negative ionisation mode. Conformational families with centroid denoted in Table S1 are shown in green. Cumulative curve fits are shown in red.
Figure S 15 Collision cross section distributions derived from arrival time distributions for all charge states of p53<sup>P34A</sup> sprayed from 50 mM ammonium acetate in negative ionisation mode. Conformational families with centroid denoted in Table S1 are shown in green. Cumulative curve fits are shown in red.
Figure S 16 Collision cross section distributions derived from arrival time distributions for all charge states of p53\(^{\text{PALLA}}\) sprayed from 50 mM ammonium acetate in negative ionisation mode. Conformational families with centroid denoted in Table S1 are shown in green. Cumulative curve fits are shown in red.
Appendix C

Figure S 17 nESI spectra in positive ionisation mode sprayed from 50 mM ammonium acetate of a) WT p53 b) p53$^{P1213A}$, c) p53$^{P27A}$, d) p53$^{P3xA}$ and e) p53$^{PallA}$

Figure S 18 CCS$_{EHSS}$ vs $R_g$ for (top left) random, (top right) SAXS weighted and (bottom left) NMR weighted ensembles. There is a strong correlation between radius of gyration and CCSs for a
limited range of values, permitting adequate conversion of $R_g$ to CCS and vice-versa. The slope and intercept of the plots is dependent on the ensemble, hence small differences are observed for the data sets.

Figure S 19 CCSD derived from arrival time distributions for all charge states of WT p53. Orange overlaid line represents the range of theoretical CCS values calculated using the constant density calculation and the cylindrical-chain estimates. Blue line represents CCS range calculated from NMR weighted ensemble of model structures. Red line represents CCS range calculated from SAXS weighted ensemble of model structures. Purple line represents CCS range of structures calculated using simulated annealing methods. Pink line represents CCS range for structures generated by neutralisation of negatively charged amino acids and subject to MD. Navy line represents CCS range for structures generated with multiple protonation sites. Green line represents CCS range for structures generated using the Complementary Coordinates (CoCo) package during MD simulations.
Supplementary information for Chapter 5

Supplementary Figures

<table>
<thead>
<tr>
<th>Conformer CCS (Å²)</th>
<th>Colour</th>
</tr>
</thead>
<tbody>
<tr>
<td>750</td>
<td>Grey</td>
</tr>
<tr>
<td>880</td>
<td>Red</td>
</tr>
<tr>
<td>990</td>
<td>Yellow</td>
</tr>
<tr>
<td>1100</td>
<td>Blue</td>
</tr>
<tr>
<td>1250</td>
<td>Green</td>
</tr>
<tr>
<td>1430</td>
<td>Magenta</td>
</tr>
<tr>
<td>1600</td>
<td>Orange</td>
</tr>
<tr>
<td>1800</td>
<td>Purple</td>
</tr>
</tbody>
</table>

Table S 2 conformational families assigned to CCSDs for WT and mutant p53TAD and the corresponding coloured shading applied.
Figure S 20 nESI mass spectra sprayed from 50 mM ammonium acetate in positive electrospray mode for a) p53\textsuperscript{P12,13A}, b) p53\textsuperscript{P27A}, c) p53\textsuperscript{P3xA} and d) p53\textsuperscript{PaliA}. Charge states are denoted above main each peak.
WT p53

Figure S 21 CCSDs derived from ATDs for each charge state of WT p53.
The CCSDs for each of the p53TAD mutants are shown in Figure S 22. Although they exhibit a similar charge state distribution as WT p53, the mutant p53TAD constructs have clear differences in their conformational landscape. p53^{P12,13A} presents in a similar CCSD for the [M+5H]^{5+} charge state, and exhibits the same conformers for the [M+6H]^{6+} charge state compared with WT p53, however retains more compact conformational families, with the conformer centered at ~880 Å² (red) most dominant at charge states [M+6H]^{6+} and [M+7H]^{7+}. p53^{P27A} presents in a similar distribution for the intense [M+5H]^{5+} charge state, but retains the compact ~880 Å² conformer at a greater intensity at [M+6H]^{6+} charge state compared with WT p53. p53^{P3xA} exhibits an additional compact conformational centered at ~750 Å² at significant intensity both [M+5H]^{5+} and [M+6H]^{6+} charge states which is not observed for WT p53. p53^{PallA}, shown by NMR and CD to possess significant residual helicity, exhibits a wide CCSD incorporating compact conformers at ~880 Å² and ~990 Å² at [M+5H]^{5+}. The [M+6H]^{6+} charge state is highly extended in comparison with WT p53, presenting intermediate conformers at ~1250 Å² and ~1430 Å² with no population of the compact conformers observed for other constructs. Although the MS spectra alone are indicative of a compact protein, the large spread of conformational families, with differences of over ten percent, are evidence of the intrinsic conformational diversity of WT and mutant p53TAD. The prevalence of intrinsic structural flexibility is signature behaviour.
of transactivation domains of transcription factors [1], suggesting that WT p53 is a compact disordered protein.

Figure S 24 Overlaid cumulative CCSDs for the [M+8H]$^9_8$ and [M+9H]$^9_9$ charge states of the p53:Mdm2 complex using each WT and mutant p53TAD construct. X-axis denotes collision cross section.
Figure S 25 Overlaid CCSDs derived from arrival time distributions for N-Mdm2 in the absence and presence of p53TAD. A) [M+6H]** N-Mdm2 alone (purple) and in the presence of WT p53TAD (blue). B) [M+7H]** charge state of N-Mdm2 alone (purple) and in the presence of WT p53TAD (blue). X-axis denotes collision cross section (Å²).

References for Appendix E

Appendix E

Supplementary information for Chapter 6

Supplementary Figures

Figure S 26 nESI mass spectrum of 10 µM MAGE-A4 in 50 mM ammonium acetate. Monomeric species are denoted with single Tilda, dimeric species with double Tilda. Several charge states are labelled. MAGe-A4 presents with a very broad, multimodal charge state range ($\Delta z$) $7 \leq z \leq 37$, with intense peaks at [M+11H]$^{11+}$ and [M+15H]$^{15+}$. We observe low intensity dimers across a broad range of charges $13 \leq z \leq 45$. 
Monomeric species denoted by red spheres, dimeric species by double red spheres. Gankyrin presents in a multimodal distribution with most intensity in the $[\text{M+9H}]^9^+$, $[\text{M+10H}]^{10^+}$ and $[\text{M+11H}]^{11^+}$ charge states. Low intensity $[\text{D+13H}]^{13^+}$ and $[\text{D+15H}]^{15^+}$ species are also observed.
Figure S 28 nESI mass spectrum of 10 µM p53 DNA binding domain in 50 mM ammonium acetate. Monomeric species are denoted with single blue spheres, dimeric species by double blue spheres. Several charge states are also denoted.
Figure S 29 nESI spectra of a) p53 DBD incubated incubated in increasing ratios up to 1:1 and concentration 20 µM:20 µM with Bax DNA promoter sequence. Highlighted blue m/z region depicts p53 DBD monomeric species and free dsDNA, which are further denoted using blue spheres and DNA helix cartoon, respectively. Red highlighted region indicates peaks corresponding to a 4:1 p53 DBD:p21 DNA complex. b) MAGE-A4 incubated in increasing ratios up to equimolar concentrations (20 µM:20 µM) with Bax DNA promoter sequence. MAGE-A4 monomeric species are denoted by single Tilda symbols, dimeric MAGE-A4 by double Tilda symbols. Several charge states have been labelled. c) MAGE-A4 inhibition of p53 DBD:DNA interaction. Concentrations of p53 DBD and Bax DNA promoter sequence were kept constant (30 µM), with increasing concentration of MAGE-A4 added. Black Tilda corresponds to MAGE-A4 signals, blue sphere corresponds to p53 DBD signals, and free dsDNA depicted by cartoon DNA helix. Red highlighted m/z region shows peaks corresponding to p53 DBD:p21 DNA complex.
Appendix E

Figure S 30 nESI spectra of a) p53 DBD incubated in increasing ratios up to 1:1 and concentration 20 µM:20 µM with PUMA DNA promoter sequence. Highlighted blue m/z region depicts p53 DBD monomeric species and free dsDNA, which are further denoted using blue spheres and DNA helix cartoon, respectively. Red highlighted region indicates peaks corresponding to a 4:1 p53 DBD:p21 DNA complex. b) MAGE-A4 incubated in increasing ratios up to equimolar concentrations (20 µM:20 µM) with PUMA DNA promoter sequence. MAGE-A4 monomeric species are denoted by single Tilda symbols, dimeric MAGE-A4 by double Tilda symbols. Several charge states have been labelled. c) MAGE-A4 inhibition of p53 DBD:DNA interaction. Concentrations of p53 DBD and PUMA DNA promoter sequence were kept constant (30 µM), with increasing concentration of MAGE-A4 added. Black Tilda corresponds to MAGE-A4 signals, blue sphere corresponds to p53 DBD signals, and free dsDNA depicted by cartoon DNA helix. Red highlighted m/z region shows peaks corresponding to p53 DBD:p21 DNA complex.
Appendix F

This appendix consists of one published perspective article:


As an author on this publication carried out all MS and IM-MS experiments, produced Figure 7 and 8, along with Supplementary Figure 2. I also drafted and edited the MS and IM-MS sections. Y.H expressed and purified WT and mutant MAGE-A4, carried out the NMR and CD experiments, drafted and edited the manuscript. A.A performed the structure-based molecular modelling.
Consequences of point mutations in melanoma-associated antigen 4 (MAGE-A4) protein: Insights from structural and biophysical studies

Yoshio Hagiwara1, Lina Sieverling1, Farina Hanif1, Jeney Anton1, Eleanor R. Dickinson1, Tam T. T. Bu2, Antonina Andreeva1, Perdita E. Barran1, Ernesto Cota3 & Penka V. Nikolova1

The Melanoma-Associated Antigen A4 (MAGE-A4) protein is a target for cancer therapy. The function of this protein is not well understood. We report the first comprehensive study on key cancer-associated MAGE-A4 mutations and provide analysis on the consequences of these mutations on the structure, stability and biophysical studies. Based on Nuclear Magnetic Resonance and Circular Dichroism, these mutations had no significant effects on the structure and folding of the protein. Some mutations affected the thermal stability of the protein remarkably. Native mass spectrometry of wild-type MAGE-A4 showed a broad charge state distribution suggestive of a structurally dynamic protein. Significant intensity was found in relatively low charge states, indicative of a predominantly globular form and some population in more extended states. The latter is supported by Ionen Mobility measurements. The MAGE-A4 mutants exhibited similar features. These novel molecular insights shed further light on better understanding of these proteins, which are implicated in a wide range of human cancers.

Cancer/testis (CT) antigens are a large family of proteins typically expressed in germ cells. CT antigens can also be overexpressed in an aberrant manner in various types of tumors such as melanoma, sarcoma, lung cancer, prostate cancer, breast cancer, ovarian cancer and a range of other cancers1,2. The first discovered CT antigen family member was melanoma-associated antigen A1 (MAGE-A1), which was identified as an antigen recognized by cytotoxic T lymphocytes in melanoma patients3, and was successfully cloned later. Since then, extensive chromosomal sequencing identified more than 50 genes of the MAGE family4,5. Based on their expression pattern, the MAGE family can be divided into two superfamilies: type-I and type-II. While the expression of type-I subfamily is restricted to germ line and cancer cells, type-II subfamily is expressed in normal somatic tissues. Type-I MAGEs can be subdivided into MAGE-A, -B and -C groups. MAGE-A contains 12 genes (MAGE-A1 to A12) where MAGE-A7 is a pseudogene. MAGE proteins consist of nearly 100 amino acid residues long N-terminal region, followed by two tandem helical domains with each terminal as WH-A and WH-B, respectively6. The N-terminal is rich in disorder promoting residues such as Ser, Pro, Glu and arg, and thus this region is predicted to be disordered. The C-terminal region that spans over the two helical domains is highly conserved among type-I and type-II MAGEs and is known as MAGE Homology Domain (MHD). The crystal structure of MHD has been determined in a free state (MAGE-A4, PDB ID: 2AVb, MAGE-A3, PDB ID: 4AV0) and in complex with NSE1 (MAGE-G1, PDB ID: 3NWO)7,8.

The biological functions of MAGE proteins remain poorly understood. However, many reports correlate over-expression of type-I MAGEs with cancer malignancy, tumor growth and poor patient prognosis. For example, MAGE-A2 was reported to promote tumor growth in normal oral keratinocytes and inhibits cell cycle arrest.  

1King's College London, Faculty of Life Sciences & Medicine, Institute of Pharmaceutical Science, Franklin-Wilkins Building, 150 Stamford St, London, SE1 9NH, UK; 2Michael Barber Centre for Collaborative Mass Spectrometry, Manchester Institute of Biotechnology, The University of Manchester, 131 Princess Street, Manchester, M1 7DN, UK; 3Biomolecular Spectroscopy Centre, King's College London, The Wolfson Wing, Hodgkin Building, London SE1 1UL; 4NRC LMB, Francis Crick Avenue, Cambridge, CB2 0QH; Imperial College London, Faculty of Natural Sciences, London, SW7 2AZ. Correspondence and requests for materials should be addressed to P.V.N. (email: perk@nikolova@kcl.ac.uk)
Appendix F

through down-regulation of p53 targets41. MAGE-A4 was also shown to abrogate p53-dependent growth arrest and apoptosis in normal oral keratinocytes25. The mechanism(s) of action of MAGE-A proteins that can lead to tumor growth have not yet been fully elucidated at a molecular level. However, MAGE-A2 was reported to interact with the DNA binding domain of p53 and to suppress p53 transactivation by recruiting histone deacetylase (HDAC)39 or by inhibiting the interaction between p53 and DNA through steric hindrance39. Most recently, MAGE-A antigens were identified as being causal contributors in the development of tamoxifen-resistant breast cancer25. Specifically, MAGE-A2 protein was reported to localize to the nucleus and to form complexes with p53 and ER, alpha, resulting in repression of the p53 pathway while increasing the ER dependent signalling. This report showed that in tamoxifen-treated breast cancer patients, there was a significant link between MAGE expression and reduced overall survival. In addition, various MAGE proteins such as MAGE-A2, A3 and A6, have been reported to interact with E3 ubiquitin ligase such as TRIM58, enhancing their ubiquitin ligase activity and reducing p53 protein level through a proteasome-dependent pathway25.

In contrast, tumor suppressor activity was also reported for MAGE-A4. MAGE-A4 was found to interact with the liver oncogene Gankyrin, suppressing its tumorigenic activity26. A truncated form of MAGE-A4 (the C-terminal 197 amino acids) was reported to induce apoptosis by interacting with POZ domain/zinc finger transcription factor Nizi-127.

To date, there are 104 MAGE-A4 DNA mutations documented in the Catalogue of Somatic Mutations in Cancer (COSMIC) database (http://cancer.sanger.ac.uk/cosmic/). Among these mutations, 76 are missense mutations, 52 of which are located within the MHD. In addition, Caballero et al.16 identified ten missense mutations in melanoma patients and two missense mutations in tumor samples (breast cancer cell line and glioma)16.

In this study, we have selected nine MAGE-A4 mutations to investigate the effects of these mutations on the structure, folding and stability of MAGE-A4, all of which are reported by Caballero et al.16. Eight of these mutations are located within the MHD (E138K, P149S, G153D, E221K, E222K, E242K, P267S and R269C) and one near the C terminus (G316R). The overall criteria for the selection of these mutations was that they mapped within the MHD of MAGE-A4, which is the only domain that had known X-ray structure. Our results reveal that although the mutations have marginal effects on the structure significantly the thermal stability of the protein. In addition, both WT and mutant MAGE-A4 contain a large disordered or unfolded region as well as a structured hydrophobic core. Native mass spectrometry (MS) shows that all of the MAGE-A4 proteins are present primarily in a monomeric form with a low abundant fraction of dimeric species. Ion mobility mass spectrometry (IM-MS) shows the WT MAGE-A4 to present in multiple conformers, whilst mutant MAGE-A4 proteins show significant compaction in the gas-phase. This study provides new information for one of the MAGE-A proteins, namely, MAGE-A4 with regard to its biophysical properties. We have performed structure-based homology modelling, mass spectrometry (MS), circular dichroism (CD) and nuclear magnetic resonance (NMR) to provide comprehensive analyses of the MAGE-A4 protein and its key cancer associated mutants. The findings could be exploited for cancer-based drug design and/or to inform further cancer therapies.

Results

Assessing the folding of the MAGE-A4 proteins using Far-UVCD at different temperatures. In order to assess the folding state and secondary structure of MAGE-A4 and examine the effects of mutations on them, the proteins were subjected to CD in the Far-UV region (260-195 nm). The locations of these selected mutations in the amino acid sequence and the crystal structure of MAGE-A4 MHD are shown in Figs 1 and 2, respectively. The spectra were recorded at 360°, 370° and 90°C, followed by cooling down to 37°C and 20°C respectively. The results are presented in Fig. 1. The CD spectra of each protein at 6, 20 and 37°C were essentially identical, indicating that the WT and mutant MAGE-A4 proteins maintain their native folded state at physiological temperature. In addition, no significant difference between the WT and the mutants was observed in the spectra at each temperature, suggesting little effects of these mutations on the secondary structure. When the temperature was raised to 90°C, all of the proteins were thermally denatured. When the temperature was cooled down to 37° and 20°C, the spectra still clearly differed from those obtained at 6, 20 and 37°C, demonstrating that the proteins did not recover their native folded state. This indicates that the thermal unfolding of MAGE-A4 proteins is irreversible. We employed the K2D3 programme (http://k2d.jysk.ca/13) to estimate the secondary structure elements of WT MAGE-A4 at physiological temperature (i.e. 37°C), and this was estimated to be 29.9% α-helix and 17.1 % 3-sheets.

The effects of mutations on the thermal stability of MAGE-A4. We employed differential scanning fluorimetry (DSF) to determine the apparent melting temperature (Tm) of MAGE-A4 proteins and assess the effects of the mutations on the thermal stability of the protein. The melting curves of MAGE-A4 proteins are shown in Fig. 4. All of the curves have the classic sigmoidal shape showing two different phenomena: unfolding and aggregation. The unfolding process of each protein occurs in the range of 20°C. The apparent Tm of WT MAGE-A4 is 48.3°C. The least thermally stable protein was E242K with the Tm = 44.6°C, followed by P267S (Tm = 44.6°C), E138K (Tm = 44.7°C), R269C (Tm = 46.2°C) and P149S (Tm = 47.3°C). The difference between the Tm of E242K and that of WT (ΔTm) was ~5.7°C and thus significant. Nearly 10% fraction of E242K mutant unfolds at physiological temperature. The most stable mutant was E221K, showing Tm = 54.6°C and ΔTm of 5.3°C. The difference in the Tm between the most stable mutant (E221K) and that of the least stable mutant (E242K) was remarkable (11°C). Other mutants (G153D, E222K) and 3-sheets are shown similar thermal stability to the WT with relatively similar Tm within ~1°C difference compared to the WT. These results demonstrate that some mutations can dramatically enhance (such as E221K) or reduce (such as E242K) the thermal stability of the protein. Interestingly, the mutations that showed significant difference in the thermal stability of the proteins (E242K and E221K) affect solvent-exposed residues.
Figure 1. Multiple sequence alignment of MAGE-A proteins. Sequences are colored according to the Clustal coloring scheme. The MHD region is underlined in red. Residues that are mutated in this study are indicated with a star.

Figure 2. Crystal structure of MAGE-A4 MHD (PDB ID: 2WA0). The structure depicts the secondary structure helices in red and strands in yellow. The mutated residues that have been investigated in this study are shown in blue. Note that P267 and G316 are not visible in the structure. The figure was produced using Pymol.

Molecular modeling. Using Modeller we modelled the point mutations in the MAGE-A4 structure in order to understand the structural basis for the observed differences in the thermal stability. The residues Glu242 and Glu221 have the most pronounced effect on the stability. The Glu242 residue is located at the N-cap of helix 3 of the second winged helix domain. It H-bonds to the exposed backbone NH of the residues Arg244 and Lys245 in the first helical turn (Fig. 5a). This residue is fairly conserved amongst MAGE-A4 homologs suggesting the importance of this position. The mutations to Lys could cause a loss of the H-bonding network and this can account for the drastic loss of stability we have observed. The Glu221 residue is located on the second helix of the second winged helix domain. Mutation to Lys could allow the formation of a salt bridge between Lys221 and Glu213 and this can potentially increase the stability of the mutant E221K MAGE-A4 protein (Fig. 5b). The effect of the remaining mutations on the structure is less obvious. The Pro267 and Arg269 residues are located at...
Figure 3. Far-UV CD spectra of WT and mutant MAGE-A4 proteins. Spectra in the region of 260–195 nm were obtained at 6°C, 20°C, 37°C, 90°C, 37°C after heating to 90°C, and at 20°C after heating to 90°C. (a) WT, (b) E138K, (c) P149S, (d) G153D, (e) E221K, (f) E224K, (g) E243K, (h) P267S, (i) R269C and (j) G316R.

A loop region that is disordered in the crystal structure. Mutation of both residues may have a destabilizing conformational effect that can vary in magnitude depending on the residue location within the loop. The Pro149 and Gly153 residues are located on the third helix of the first winged helix domain. Both residues are probably important for the stability of the alpha-helix, particularly Pro149, as it is located at the N-terminal turn of the helix and is entropy-favoured in this position. A substitution of Gly153 with a residue with side chain and charge such as Glu would be sterically unfavoured as it may clash with the side chain of the residue Ile300. The Glu138 residue
Figure 4. Thermal unfolding curves of MAGE-A4 and its cancer-associated mutants recorded by DSF. The data were recorded using MX3005P machine (Stratagene) with the excitation and emission of SYPRO Orange dye set to be 470 nm and 570 nm, respectively. The proteins were diluted to 2 μM in 50 mM Tris buffer containing 150 mM NaCl, 5 mM DTT and 5% glycerol, pH 7.2.

Figure 5. (a) Side by side comparison of the structure of wild type MAGE-A4 (PDB ID: 2WA0) and the theoretical model of the E242K mutant. (b) Side by side comparison of the structure of wild type MAGE-A4 (PDB ID: 2WA0) and the theoretical model of the E221K mutant.

is located at helix 2 of the first winged helix domain. Mutation of this residue to lys may be unfavoured due to charge repulsion as this residue is in close vicinity to Lys242, Lys245 and Arg239.

1D proton NMR of MAGE-A4 proteins. 1D proton NMR was employed to examine the folding and conformation of MAGE-A4 protein and to assess the effects of the mutations in such properties. The spectra of the
Figure 6. 1D Proton NMR Spectra of WT MAGE-A4 and its cancer-associated mutants. (a) Whole spectra. (b) Zoomed-in spectra within 0.7 and −2 ppm. Each protein was concentrated to 200 μM in 50 mM phosphate buffer (pH 7.2) containing 50 mM NaCl and 5 mM DTT. The spectra were recorded at 600 MHz, 25°C.
Figure 7. nESI mass spectrum for 20 μM WT MAGE-A4. WT MAGE-A4 sprayed from aqueous solution containing 50 mM Ammonium Acetate at pH 6.8. The protein exhibits a wide, multimodal, monomeric charge state range 10 ≤ z ≤ 44. Several charge states are denoted with single spheres along with their respective charge.

Figure 8. Collision cross section distributions (CCSDs) derived from arrival time distributions (ATDs) for WT MAGE-A4 at drift voltage 35 V. The x-, y- and z-axes represent collision cross section (CCS, Å²), intensity directly acquired from IM-MS experiment, and charge state, respectively. Significant charge states have been labelled.

proteins are shown in Fig. 6a. The differences in the spectra between the WT and mutant proteins are minimal and located within the chemical shift of methyl protons (below 0.8 ppm). The NMR spectra for all of the tested MAGE-A4 proteins showed excellent dispersion of peaks between −1 and 0.8 ppm that are typical of ring-current shifted proton located above or below the plane of aromatic side-chain. This indicates the presence of a structured and disordered regions. These results suggest that the WT and mutant MAGE-A4 proteins contain a folded structure as well as disordered region(s). Figure 6b zooms in the chemical shift region below 0.8 ppm where the difference in the peaks is observed. While the spectra of P267S and R269C overlap completely with that
of WT, the spectra of other mutants differ from that of WT in the chemical shift region from 0.7 to -0.2 ppm. These small changes in the positions and patterns of the peaks reflect minor changes in the hydrophobic core caused by such mutations even though these mutations are located on the solvent-exposed surface of the protein. These mutations may somehow change the position or orientation of the planes of aromatic and/or methyl groups, changing the distances and/or angles of the interactions between methyl groups and aromatic planes.

**M5 and IM-M5 of MAGE-A4.** We employed MS and drift time IM-MS to probe the structure of WT and mutant MAGE-A4. The mass spectrum for WT MAGE-A4 is shown in Fig. 7. The protein presents in a wide charge state distribution (CCSDs) 10 ≤ z ≤ 44. The most intense charge state is [M + 11H]^{11+}, although the distribution is multimodal and other intense species include [M + 15H]^{15+}, [M + 20H]^{20+} and [M + 28H]^{28+}. The CCSDs was easily measureable, changing significantly with ionic strength and cone potential. At the salt concentration increased the intensity of the higher charge states decreased somewhat, but they were still present over the same range (data not shown). The protein predominately presents as a monomer of mass 35,679 (cf. predicted mass of 35,684 Da) with very low abundance peaks assigned to a dimeric form (Supplementary Fig. S1). The wide CCSDs is typical of an intrinsically disordered protein, with many solvent accessible non-neutral sites^{59}. The intense [M + 11H]^{11+} species indicates WT MAGE-A4 may possess some stable secondary structure, or at the very least a globular form which is built around the solved MHD structure.

IM-MS reveals that WT MAGE-A4 presents in multiple conformational families (Fig. 8). The low charge states [M - 10H]^{10+}, [M + 11H]^{11+} and [M + 12H]^{12+} present in a single conformational family at collision cross section (CCS) ~2400 Å^2, likely corresponding to a stable compact form of the protein. At intermediate charge states 11 ≤ z ≤ 29 MAGE-A4 presents in at least two poorly resolved conformational families centered at ~3200 and 4500 Å^2. At high charge states 21 ≤ z ≤ 26 the protein exhibits very wide collision cross section distributions (CCSDs) with multiple conformational families present ranging from ~3500 to 7500 Å^2, indicative of a disordered protein with large extended states. The highest charge states 27 ≤ z ≤ 29 exhibit a single broad CCSD centered on ~4000 Å^2, possibly due to gas-phase collapse. Charge states higher than [M + 29H]^{29+} were not of sufficient intensity to obtain IM MS data.

Mutant MAGE-A4 spectra are shown in Supplementary Fig. S2. All mutants present a wide CCSDs similar to that of WT MAGE-A4. For each case the protein presents as mainly monomer, exhibiting a multimodal distribution with the [M + 11H]^{11+} species being the most intense. E138K MAGE-A4 presents in a monomeric charge state range 8 ≤ z ≤ 33 with some low intensity low charge dimers. F149S MAGE-A4 presents in charge state range 8 ≤ z ≤ 33, with a higher abundance of dimeric species compared with the WT protein. G153D MAGE-A4 presents as a charge state range 7 ≤ z ≤ 19 with significantly fewer high charge states compared with the WT protein. As for WT MAGE-A4, G153D MAGE-A4 presents with very low abundance dimeric species. E221K MAGE-A4 presents in charge state range 7 ≤ z ≤ 21 with some low intensity low charge dimers. E221K MAGE-A4 presents a charge state range 7 ≤ z ≤ 20, again with the [D + 15H]^{15+} and [D + 17H]^{17+} dimeric species. E242K presents in charge state range 7 ≤ z ≤ 18 with low intensity [D + 15H]^{15+} dimer. This charge state range is narrower, and dimer abundance is lower than the WT MAGE-A4 protein. P267S MAGE-A4 presents in a monomeric charge state range 7 ≤ z ≤ 22. The dimeric species are low intensity with charge state range 15 ≤ z ≤ 39, the highest abundance of dimeric species of WT and mutant MAGE-A4. R269C presents as a monomeric charge state range 7 ≤ z ≤ 31 and with [D + 15H]^{15+} and [D + 17H]^{17+} dimers. G316R presents with a monomeric charge state range 7 ≤ z ≤ 31 and a dimeric charge state range 15 ≤ z ≤ 25.

All mutants present with a narrower charge state range compared with the WT MAGE-A4, with G153B, E242K and R269C displaying a significantly narrower distribution, centered on lower charges suggestive of a more compact form being dominant in solution. MAGE-A4 mutants P149S, P267S and G316R display an increased abundance of dimeric species compared with the WT protein.

We performed IM MS on the MAGE-A4 mutants which exhibited significant thermal stability differences, the most stable, E221K, and the least stable, E242K MAGE-A4 (Supplementary Fig. S3). E221K MAGE-A4 presents in similar conformational families to the WT protein for low charge states 9 ≤ z ≤ 12, with a single conformational family centered at ~2400 Å^2. At intermediate charge states 13 ≤ z ≤ 16 E221K MAGE-A4 exhibits two conformational families centered at ~2600 and ~4000 Å^2; similar to that of WT MAGE-A4 but with a narrower distribution. The high charge states 17 ≤ z ≤ 20 present in much more narrow CCSDs for that observed in the WT protein. We see no highly extended conformational states, instead observing a broad unresolved distribution at ~3500-5000 Å^2. The E242K MAGE-A4 protein again exhibits conformational differences in comparison with WT MAGE-A4. The trend of a single conformational family at low charge states, with multiple conformational families presenting at charge states 12 ≤ z ≤ 23 remains consistent. We again however do not observe the more unfolded states present in the high charge states of WT MAGE-A4, and observe consistently much narrower CCSDs than both WT and E221K MAGE-A4. Several charge states are directly compared for the WT, E221K and E242K MAGE-A4 proteins (Supplementary Fig. S4), highlighting the similar CCSDs for the compact, low charge states but significant differences in the CCSD width at both intermediate and especially high charge states for the two mutant MAGE-A4 proteins. The mutations which impact on thermal stability, both positively and negatively, appear to also have an effect on the conformation landscape of MAGE-A4. The narrowing of the CCSDs suggest that E221K and E242K MAGE-A4 proteins possess less flexibility than WT MAGE-A4, preferably presenting in more compact conformational states.

**Discussion**

Despite being discovered more than 20 years ago, biological functions of the MAGE family of proteins still remain poorly understood. Most studies were focused on developing anti-cancer immune therapy using the exclusiveness of type-I MAGE expression in cancer cells. For example, MAGE-A3 has been studied as a candidate for developing immune therapy^{53-54}. Emerging data highlight their active roles in promoting cancer growth and
malignance35. MAGE-A2 has been reported to repress the activity of tumor suppressor protein p53 through recruitment of HDAC or by inhibiting the interaction between p53 and DNA.36,37

In this study, we have selected MAGE-A4 and nine of its mutants identified in human cancer cells to obtain a first insight into their structural and biophysical properties. We observed that the selected point mutations have little effects on the structural integrity and folding of the protein probably since most of these mutated residues are located on solvent-exposed surface. However, some mutations, such as E221K and E242K, had significant effects on the thermal stability. The apparent $T_m$ of E242K MAGE-A4 was 5.7°C lower than that of WT protein. A small fraction (10%) of MAGE-A4 mutant unfolds at physiological temperature. The decreased thermal stability coupled with unfolding at physiological temperature caused by E242K mutation may reduce the functional activities of MAGE-A4 protein in vivo.

Both NMR and native MS showed for the first time that these MAGE-A4 proteins contain a compact folded structure as well as disordered regions. Most of such disordered regions are expected to be located within the N-terminal 109 residues due to presence of disorder promoting residues. Interestingly, we observed a very small fraction of dimeric species in the native MS. It has been reported that MAGE-A4 cannot form a native dimer.18 In addition, while this article was under review, a new report emerged suggesting that MAGE-A3 protein remains monomeric in solution17. It is not known whether such dimeric species possess different functional properties compared to the monomers.

Of note, the percentage of MAGE-A4 positive patient tumors observed in select cancer subtypes has been reported as follows: 67% in ovarian cancers, 19-35% in lung cancers, 22% in colon cancers and 13% in breast cancers.19 The role of MAGE-A1 in cancers is not fully understood especially at the molecular level. In fact, there are conflicting reports regarding its role in cancer. It has been reported to have oncogenic properties by inhibiting p53 downstream genes such as BAX and p21, and thus promoting cell growth in normal oral keratinocytes19. However, MAGE-A4 was also reported to have tumor suppressive functions as it was shown to promote apoptosis in non-small cell lung cancers.20 It is not understood how MAGE-A4 exerts such opposing effects. We hypothesize that such observations may arise from different interacting partners of MAGE-A4. As part of its tumor suppressing mechanism, MAGE-A4 has been indicated to bind with gas6, suppressing its oncogenic activity.20 However, to date, none of these interactions have been identified for MAGE-A4, which can lead to cancer growth.

As the point mutations selected for this study are located on the solvent-exposed surface of MAGE-A4, even though they do not affect structural integrity of MHD as observed in this study, they may have a significant impact on the protein-protein interaction(s), which may enhance or repress tumor growth. In addition, conformational changes have been shown for E221K and E242K mutants in NMR experiments. These may also have effects on the protein-protein interaction(s), modifying the conformational flexibility of MAGE-A4. Furthermore, the impact of the mutations on the thermal stability and folding of the protein may have significant effects on the function of this protein in vivo. Further investigation needs to be conducted to identify other proteins that interact with MAGE-A4 and assess the impact of these mutations on such interactions to understand whether these mutations have different functional roles in cancer.

Given the significance of MAGE-A4 in biology and specifically in cancer, it is important to fully understand the functions of this protein at the molecular level in order to exploit it in translational research. Specifically, the mutations may affect the interactions with key binding cell cycle proteins. The latter are subject to further investigations aiming to identify and characterize such binding proteins in the hope to better understand the role of these mutations in cancer.

Methods

MAGE-A4 cloning and site-directed mutagenesis. The DNA of WT MAGE-A4 was cloned into pGEX 6p-1 (GE Healthcare) to express GST-tagged MAGE-A4 using primer 1: 5'-AAAAAAGGAATCAGTCTCTGAGAGCAGA AGAAGTCG-3' as a forward primer, primer 2: AAAAACTGTTAAGACTCCCTCTCTCTCTAA-3' as a reverse primer, and vector pGEX 6-1 containing MAGE-A4 DNA sequence as a template. MAGE-A4 mutants were generated using QuickChange Lightning Site-Directed Mutagenesis Kit (Agilent Technologies).

Protein expression and purification. Proteins were expressed in E. coli C41 (DE3) cells. Single colonies were inoculated into 5 mL LB media (Fermentas) supplemented with 100 µg/mL ampicillin, followed by incubation at 37°C overnight. 5 mL of the overnight culture was added into 500 mL LB with 100 µg/mL ampicillin, followed by incubation at 37°C until OD$_{600}$ reached 0.6. The culture was then allowed to cool down to 21°C and the protein expression was induced with 1 mM isopropyl β-D-thiogalactoside overnight. The cultures were centrifuged at 11300 × g, 4°C for 20 min. The cells were collected and lysed using 5 μg/mL cell lyser (Roche) 125 U of Benzonase (Novagen) and 3 mM dithiothreitol (DTT). The cell lysates were centrifuged at 20000 × g, 4°C for 60 min. The soluble fraction was collected and the protein was purified using GST GraviTrap column (GE Healthcare). The GST-tagged protein was incubated with 5 μg/mL protein of PreScission protease (GE Healthcare) at 4°C overnight. The protein was passed again through GST GraviTrap column to remove the GST, followed by further purification using 36/60 Superdex 200 column connected to ÄKTAprime+ system equipped with Prime View software (GE Healthcare). The purified protein was concentrated using Amicon Ultra-15 centrifugal filter unit, 10,000 MWCO (Merck Millipore) and stored at −80°C.

CD. Similarities in CD absorbance and CD spectra of MAGE4 proteins were acquired on the Chirascan Plus® spectrometer, equipped with a Quantum Northwest TCI250 Peltier unit (Applied Photophysics, Leatherhead, UK). The instrument was flushed continuously with pure evaporated nitrogen gas throughout the measurements. Far-UV CD spectra were recorded with a 2.0 nm spectral bandwidth, a 1.0 nm step size and a 1.5 s instrument integration time.
time-point. A 0.5 mm Suprasil rectangular cell (Hellma UK Ltd) was employed in the region 260–195 nm. Far-UV CD spectra were recorded at room temperature (20°C), cooled to 6°C, then heated to high temperatures (70°C and 90°C) and re-cooled to 27°C and 20°C. The temperature was measured directly with a thermocouple probe in the sample solution. Protein samples were concentrated to 0.2 mg/ml and were buffer exchanged to 20 mM Tris buffer containing 50 mM NaCl and 1 mM diethyldithiocarbamic acid (DTT), pH 7.2. All CD spectra were acquired at 26 C with the baseline subtracted and then corrected for concentration and path length and expressed in terms of ε M⁻¹ cm⁻¹ per amino acid residue (MW = 113). Protein secondary structure prediction was performed using K2D3 (http://k2d3.igc.gda.i.c). 

**DF5**. Protein samples were diluted to 2 mg/ml in 50 mM Tris buffer containing 150 mM NaCl, 5 mM DTT, and 5% glycerol. SYPRO® Orange stain (Sigma-Aldrich) was added to the sample to be diluted 500 times, 20 µl of each sample was transferred to 96-well plate, and the plate was sealed with a transparent tape. The samples were run in Mx3005P machine (Stratagene) connected to MxPro software and the temperature was increased from 25°C to 95°C. The scan was excited at 480 nm and the emission at 570 nm was measured. The unfolding curve was generated by normalizing the fluorescence values using equation:

\[ \frac{F_T - F_{	ext{min}}}{F_{\text{max}} - F_{\text{min}}} \]

where \( F_T \) is fluorescence at temperature \( T \), \( F_{\text{max}} \) is maximum fluorescence, and \( F_{\text{min}} \) is minimum fluorescence.

**Molecular modeling.** Models were built with Modeler® using the structure of human MAGEA4 as template (PDB ID: 2WAA). 

**1D proton NMR.** Protein samples were concentrated to 200 µM and the buffer was exchanged to 50 mM phosphate sodium buffer containing 50 mM NaCl and 5 mM DTT. 500 µl of each protein sample was mixed with 50 µl of deuterium oxide and transferred to an NMR tube. The proton NMR spectra were recorded using Bruker Avance III 600 at 25°C, 600 MHz.

**MS.** Prior to analysis, samples were thawed and dialyzed using Bio-RAD micro spin chromatography columns (Bio-Rad Laboratories, Inc.) into 50 mM Ammonium Acetate and diluted to 20 µg/ml. Protein concentrations were confirmed by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE). Mass spectrometric experiments were performed on a Micro Mass Quadrupole-Time of Flight Ultima Global (Waters, Manchester, UK) and ion mobility mass spectrometry measurements were made on an in-house modified quadrupole time-of-flight mass spectrometer including a 5.1 cm drift tube. Ions were produced by positive nano-electrospray ionisation (nESI), with a spray voltage range of 1.55–1.69 kV and a source temperature of 60°C. nESI glass capillaries were prepared in-house from thin-walled borosilicate capillaries (Precision Instruments, Stevenage, UK) using a flaming/Brown micro pipette puller Model P-97 (Sutter Instrument company, Novato, CA, USA). The IM-MS drift cell was filled with helium gas, to pressure of 3.56–3.75 Torr, measured using a baratron (MKS Instruments). The electric potential of the drift cell was varied from 12 to 2 Vac. Ion arrival time distributions (ATDs) were recorded by synchronisation of the release of ions into the drift cell with mass spectral acquisition. ATDs were converted into collision cross section distributions using equation:

\[ \Omega_{\text{avg}} = \frac{1}{2} \left( \frac{m_e}{m_i} + 1 \right) \left( \frac{1}{k_B T} \right)^{1/2} \frac{\sigma e}{\rho V} \]

where \( \Omega \) is the rotationally averaged collision cross section (Å²), \( m_e \) and \( m_i \) are the masses of the ion and the buffer gas, \( e \) is the ion charge, \( e \) is the electron charge, \( k_B \) is the Boltzmann constant, \( T \) is the gas temperature, \( l \) is the drift cell length, \( V \) is the voltage applied across the drift cell (35 V) and \( t_0 \) is the drift time. The arrival time of the ions (\( t_0 \)) includes the time spent within the mass spectrometer by outside of the drift cell, also known as the dead time (\( t_d \)). The value for \( t_d \) is calculated by taking an average value of the intercept from a linear plot of average arrival time vs. pressure/temperature and was subtracted from the arrival time to calculate drift time (\( t_0 \)).

\[ t_0 = t_l - t_d \]

All processing was carried out using Mass Lynx V4.1 (Waters Corporation) and Origin 9.0 (Origin Lab Corporation, USA). Intensities of the CCS peaks were directly taken from peak intensity during IM-MS experiments for three analytical repeats, in Mass Lynx V4.1 software (Waters Corporation).

**References**
Appendix F

Appendix F

www.nature.com/scientificreports


Acknowledgements

We thank Dr. Claudius Schneider for the MAGE-A4 clone. We wish to acknowledge the support of professor Brian Sutton during the course of this study.

Author Contributions

Y.H., F.H., and A.J. purified proteins. Y.H. and T.B. performed CD experiments. Y.H. and E.C. performed NMR and DSD and analyzed data, A.A. performed structure-based molecular modeling, E.R.D. and P.E.B. performed MS experiments and analyzed data. P.V.N. and Y.H. participated in project design, Y.H., T.T.E.B., L.S., E.R.D., I.C., A.A., F.E.B., and P.V.N. analyzed data and participated in the preparation of the manuscript, P.V.N. conceived and supervised the project. All authors read and provided feedback on the manuscript.