Characterising the Small Heat Shock Proteins: Structure and Regulation

Heidi Ysabella Gastall

University of Oxford

Systems Biology Doctoral Training Centre

St. Anne's College

This thesis is submitted for the degree of Doctor of Philosophy in Systems Biology

Hilary 2016

Abstract

The small heat shock protein HSP27 is expressed throughout the body and involved in a variety of cellular processes. Its multiple roles may all be mediated by its ability to chaperone partially folded protein substrates, and its malfunction is implicated in a range of diseases such as motor neuropathy and neurodegeneration. Here I describe structural, dynamical and functional investigations into this protein, particularly regarding its regulation by oxidation and phosphorylation. The foundations of this work were a protein crystal structure obtained on the oxidised core domain of the wild-type protein, which forms a disulphide-locked dimer, and additional structures that contain neuropathy-associated mutations. Observations of a structurally variable region, namely the β2-strand at the beginning of the core domain, led us to probe the nature of this region in solution and examine whether phosphorylation at the nearby phosphoserine site led the β2 to bind more strongly to the core domain, as phosphorylation elicits HSP27 oligomer disruption and activity modulation. I am able to postulate on the structural effect of mutation using the mutant crystal structures, and report the first structure showing the interface formed by the HSP27 core domain dimer in the 'antiparallel III' register. This finding, in tandem with the unusual nature of the intact intra-subunit disulphide bond reported here, motivated studies into the role of this disulphide in the redox response of the HSP27 chaperone. These include dynamical studies on the subunit exchange of the protein and functional investigations showing that the disulphide is formed in the cell and that chaperone behaviour depends on the oxidation state. Moreover, I show that the redox potential of the disulphide is affected by mutation, as is the change in chaperone response on oxidation. Taken together, I suggest that HSP27 is regulated by oxidation through the intra-dimer disulphide, which allows it to respond to oxidative stress within the cell.

Declaration

This thesis is a summary of research conducted at the Department of Chemistry, University of Oxford between September 2012 and April 2016. The work described here is wholly my own or specifically indicated in the text if the outcome of collaboration (please see below). It has not, neither in whole nor in part, been submitted for a degree, diploma or any other academic qualification at any other academic institution.

Heidi Ysabella Gastall
Oxford, April 2016

A substantial amount of the experimental work presented in this thesis was undertaken in collaboration with other researchers. Where I have included work undertaken by others in the results chapters, I have acknowledged their contribution in the figure legends. In all cases, I was involved throughout the data acquisition and analysis stages, often providing the samples, and explaining how the results tied into the larger project. For clarity, I list them here.

Art Laganowsky helped design the chaperone assay experiment mentioned in Chapter 1, the constructs for crystallisation trials and led the refinement of the wild-type crystal structure along with some of the R136W structure. Tim Allison refined the T151I and T164A crystal structures, and conducted the final refinement of R136W. I also undertook some refinement as well as being trained during this process for each of the structures, providing the perspective on the biological implications of particular models (for example, possible ligand identities for the additional electron density in the R136W structure), and conducting all of the crystallisation trials.

Matteo Degiacomi and Patty Lee conducted the molecular dynamics simulations and square wave voltammetry respectively, with my overview and description of the experiment, based on my observations in the wild-type crystal structure and in the glutathione assay in solution. Similarly, Reid Alderson conducted the NMR experiments in light of the crystal structure and my observations in solution and the gas phase. Anthea Rote and Heath Ecroyd performed the full-length chaperone assays, FRET subunit exchange experiments and some of the Western blots, and prepared the plasmid DNA for their experiments on the FL. Otherwise, I conducted all of the preparation of plasmid DNA for this research. I conducted all other experimental work.

Chapter 2 was prepared as a book chapter review in co-authorship with Dale A. Shepherd. While he and I wrote separate sections of the first draft, all authors edited and produced the final piece as presented here.

The way I have presented the crystallography introduction in Chapter 1 relies heavily on the structure used by Gale Rhodes in *Crystallography Made Crystal Clear: A Guide for Users of Molecular Models*. I am grateful for the numerous online guides explaining general and specific aspects of protein crystallography, which along with the books, reviews and software release articles have shaped my theoretical understanding of this subject.

Acknowledgements

I have many people to thank for their training, support and practical help during the course of my DPhil, and especially during the time I have been writing my thesis. One of the most appealing aspects of scientific research has been its collaborative nature, allowing me to work with a myriad of people who each bring a different perspective and contribution to the research. I feel like the DPhil and thesis themselves have been the product of a collective effort, as so many people have given their time and energy to help me to achieve it.

Firstly I am very grateful to Justin Benesch who has always made himself available, offered insightful advice and feedback, given me freedom to explore, encouraged and challenged me. Moreover, he ensured that I had many opportunities to learn, and encouraged me from the beginning that a DPhil would be highly rewarding. Together with the groups of Carol Robinson and Andy Baldwin, his research group has been full of people who are dedicated, highly skilled and intelligent. They always desire to understand more, and are incredibly generous with their time and expertise. This makes for a really effective training and research environment, and I really appreciate the time I have had here.

Art Laganowsky provided a lot of the practical training in the first half of my project, and moreover was always incredibly generous with his time, energy and ideas. The HSP27 project described herein evolved from one of the side projects he suggested I undertake to learn about protein crystallography, though we worked on several other projects together which were each exciting and challenging. Through his energetic determination, he demonstrated that many techniques and projects can be mastered to answer a research question.

Tim Allison has patiently fielded many questions about crystallography, after he perhaps unwittingly first offered to help me harvest crystals. Since then, he conducted and/or polished refinement of several of the mutant crystal structures, and has given me hours of his time explaining and training me. He also advised me on how to use the Pulsar software. Additionally, along with Jani Bolla, he gave me very helpful feedback on the crystallography sections in this thesis. I am very grateful to them both.

Reid Alderson joined the lab not long after Art had left, and in him I found a colleague who was equally as passionate about discovery. His enthusiasm for the project, along with his expertise in NMR, have been so valuable and I have really enjoyed working with him. He too offered useful feedback on this thesis.

I had the pleasure of working with Matteo Degiacomi on the molecular dynamics simulations (as well as a few other side projects). His enthusiasm, skill and energy, peppered with some trademark gesticulations, produce excellent work and he has always so willingly explained to me the details of his experiments, along with gently correcting my French grammar!

Heath Ecroyd, Anthea Rote and BJ Jovcevski have been long-standing collaborators. I am really grateful for their dedication and insight on the work I have included here, and am delighted that our collaboration has been so fruitful.

Miranda Collier and Daisy Nicholls have both joined the HSP27 wing of the Benesch group with dedication and good judgement, with Daisy also taking on the baton for the chimera project described in Chapter 1. I am really looking forward to seeing the outcomes of this work in the very capable hands of Reid, Daisy and Miranda.

Patty Lee and Richard Compton undertook the electrochemical characterisation described in Chapter 7. Patty patiently applied her expertise in this largely unchartered application, and kindly returned comments to me on the section in this thesis, as well as explaining her approach and helping me to rationalise the results.

Dale Shepherd agreed to write the book chapter with me, and three years later probably thought it was a good idea. Moreover, his MS and IM-MS expertise was invaluable, and along with Jon Hopper, Fran Kondrat and Gill Hilton, shaped much of my practical understanding of how mass spectrometers work. Mike Marty freely and kindly offered much advice on the use of UniDec and how to get the most from it. Jason Yen provided help with the Western blots, generously sharing his time and experience.

Iva Pritisanac has always willingly shared her knowledge of NMR spectroscopy and furthermore, been a constant encouragement both academically and personally. She is always willing to help practically too, which has led her to spend several very late nights in the office on my behalf both before a job interview and for the DPhil (this being no exception), and I am incredibly grateful to her.

Fran Kondrat has been a friend and advisor from her first day in the lab, always encouraging, willing and patient to discuss the details of my project and wet lab work. I learnt a lot from Georg Hochberg, who was always happy to discuss, give me ideas and remind me that sometimes research is tough, but nonetheless was a stalwart worker and great inspiration. Eamonn Reading enthused those around him with his passion for science as well as giving me practical help. Idi and Dom offered much fun and friendship, as well as casual conversation about MS. Erika Wu constantly cheered me on and always offered me a place to stay after I had moved from Oxford, often at the slightest notice.

I would like to express my gratitude to the Systems Biology DTC for funding and to all of the members therein for their contribution to my scientific education. I am particularly grateful to Elspeth Garman for her guidance and organisation.

It will not be surprising to anyone that my family has been a constant source of encouragement to me throughout my DPhil. My parents, sister and brother-in-law, and wider but very close family have all always known that I could get to the point of writing this preface. They have prayed for me every day for whatever particular hurdle I was facing –

whether it be the infamous 'Chapter 4' of the Christmas holidays, or the peptide binding experiment when I was fighting to acquire suitably and consistently resolved spectra. Their constant words of encouragement and faith have made this journey both joyful and fruitful. And today I am delighted to write that our family has grown by one – my new niece (also known as 'Mimo' hitherto) timed her arrival perfectly as the sweetest possible reward. I cannot wait for the years that we will spend together discovering and learning about God and the world around us.

Similarly, I have had many friends who have supported me in prayer or practical help. This includes Anna and Lucie (long-term companions in life and academic pursuit); my community group and friends at Emmanuel church who were always a source of prayer and fun, including my dear friend and library coworker, Debs, who has been a great encouragement; all of my wonderful housemates during my time in Oxford, not least Abi, a faithful, witty and justice-loving friend who always waited up to check I'd got home safely from the lab; Emily and Jen for their continued support and long-term loan of a laptop; Rachel and Alistair, Joy and Dan, who so kindly opened their homes and ovens to me during the final months of my time in Oxford; my new housemates Sian and Leanne who have also welcomed me with food and care every evening and then sent me off to write; my friends and colleagues on the DTC, especially Angie and Jia, who taught me more about God's delight in our work and our lives; Bex, Stuart and Stuart, who encouraged and prayed for me from afar. There are so many people whom I can't mention by name now (it's time to print…), but have been such loyal friends throughout the past four years.

Finally, I would like to thank the Father of all things, who loves to give his children good gifts. My time in Oxford has brought me such joy and stimulation, where I've been able to learn and explore, not only about the sHSPs, mass spectrometry and structural biology, but about nature, rowing (!) and people. I am so grateful to live in such a beautiful city which is full of stories – I have frequently had such delight when cycling through just to see these elaborate buildings. I really feel that God has provided for me in so many small and large ways, from the provision of computers, wonderful people who have welcomed me into their homes, the generosity of many who have given their time and energy to train and teach me or give comments on my thesis, the number of perfectly timed encouraging words, as well as so many logistical factors that have all made this process easier, and reminded me that He is such a generous and loving Father.

'*how much more will your heavenly Father give good gifts to those who ask him.*'

Matthew 7:11

Contents

Chapter 1: Introduction

1 Biological Motivation 1
1.1 sHSP structure 2
1.2 sHSP function 6
2 Technical Strategy 8
2.1 MS 9
2.1.1 Nano-electrospray ionisation 10
2.1.2 Mass analysers 14
2.1.3 The detector 18
2.1.4 Collision-induced dissociation 22
2.1.5 Ion mobility MS 23
2.1.6 Estimating CCS values from structural models 26
2.2 Protein Crystallography 28
2.2.1 Protein crystallisation 29
2.2.2 X-ray diffraction 31
2.2.3 The unit cell and its space group 32
2.2.4 The reciprocal lattice 33
2.2.5 The requirement for formation of diffraction spots 34
2.2.6 Indexing, integration, merging and scaling 35
2.2.7 The relationship between the structure factor and electron density 36
2.2.8 Molecular replacement 39
2.2.9 Phase improvement and phase extension 42
2.2.10 Model building and refinement 44
2.2.11 Validation 50
2.2.12 Biological relevance 51
3 Conclusion 53
4 References 54

Chapter 2: Native mass spectrometry of the proteostasis machinery

1 Introduction 71
2 Holdases: Small Heat-Shock Proteins 74
2.1 Quantifying the variable oligomerisation of sHSPs 75
2.2 Monitoring the quaternary dynamics of sHSPs 80
2.3 Disentangling chaperone-substrate interactions 82
2.4 From stoichiometries to structures using ion mobility MS 84
3 Foldases: ATP-dependent molecular chaperones 87

3.1 Comparing the myriad interactions of HSP70 and HSP90 87
3.2 Insights into HSP60 beyond its role as a native MS standard 91
4 Degradases: the Proteasome 94
5 Conclusions and Perspectives 97
6 References 100

Chapter 3: Structural Insights into HSP27

1 Introduction 114
1.1 The sHSP HSP27 114
1.2 Existing structural insight on HSP27 and the mammalian sHSPs 116
2 Results and Discussion 121
2.1 In search of a full-length model of oxidised HSP27 121
2.2 A new structural model for the core domain of HSP27 124
2.2.1 Optimising crystallisation of the core domain 125
2.2.2 The oxidised core domain of human HSP27 displays multiple interfaces 127
2.2.3 Comparison of c27-2 with other structural models 137
2.2.4 Are the interactions of β2 an artefact of crystallisation? 146
2.2.5 Significance of the domain swapped β2 148
3 Conclusion 149
4 Methods 151
4.1 Materials and reagents 151
4.2 Constructs 151
4.3 Protein expression and purification 152
4.4 Crystallisation trials 153
4.5 Structure determination 154
4.6 Mass spectrometry 156
5 References 157

Chapter 4: Investigating Regulation of HSP27 by Phosphorylation

1 Introduction 167
2 Results and Discussion 169
2.1 MD simulations validate that the β2 easily dissociates 169
2.2 NMR detects a disordered β2 strand in solution 175
2.3 Oligomeric distribution of c27-2 176
2.3.1 An aside on an unexpected adduct 183
2.4 Investigating competitive binding between the C-terminus and β2 strand with native MS 184
2.5 NMR studies show binding of the β2 strand is enhanced by phosphorylation 192
2.6 Continuing investigation on the mechanism of phosphorylation 196

3 Conclusion 199
4 Methods 201
4.1 Materials and reagents 201
4.2 Constructs 201
4.3 Protein expression and purification 202
4.4 MD simulations 202
4.5 Mass spectrometry 203
4.5.1 Oligomeric distribution 203
4.5.2 Peptide competition experiments 204
4.6 NMR spectroscopy 206
5 References 208

Chapter 5: The Effect of Mutation on the Core Domain of HSP27

1 Introduction 214
1.1 Mutations to the HSP27 core domain have been identified in motor neuropathy 214
2 Results and Discussion 216
2.1 Structural models of the core domain with R136W, T151I and T164A mutations 216
2.1.1 T164A interrupts a hydrogen bonding network across the minor β-sheet 220
2.1.2 Mutation of T151 destabilises a 3_{10} helical turn 224
2.1.3 c27-2(R136W) forms a register-shifted dimer 230
2.2 Additional electron density in the R136W structure 234
2.2.1 A population of β-mercaptoethanol cysteines at the dimeric interface 234
2.2.2 Unmodelled electron density near to the β2 strand 236
2.3 Registration shifts of the intra-dimer interface 240
2.4 The intersubunit disulphide is rare and sterically strained 243
3 Conclusion 245
4 Methods 248
4.1 Materials and reagents 248
4.2 Constructs 248
4.3 Protein expression and purification 248
4.4 Crystallisation trials 249
4.5 Structure determination 250
4.6 F_o-F_c map 254
5 References 255

Chapter 6: Redox Regulation of HSP27

1 Introduction 260
2 Results and Discussion 264
2.1 The disulphide forms within the cell 264
2.2 Using IM-MS to probe the conformation of the core domain: solution-mediated effects 266
2.3 The disulphide stabilises the intra-dimer interface and significantly slows subunit exchange 273
2.4 Solution NMR spectroscopy reveals multiple populations of the interface in solution 286
2.5 Disulphide formation affects chaperone behaviour 288
2.6 Native MS of the core domain shows S135F oxidises less readily 291
2.7 Using IM-MS to probe the conformation of the core domain: effects of mutation 294
2.8 Motor neuropathy mutations affect the redox potential of the HSP27 disulphide 298
2.9 Mutations affect the change in chaperone activity of the full length on reduction 305
3 Conclusion 310
4 Methods 314
4.1 Materials and reagents 314
4.2 Construct preparation 314
4.3 Protein expression and purification 314
4.4 Western blots 315
4.5 Ion mobility mass spectrometry 316
4.5.1 Estimating the CCS values of API, APII and APIII c27-2 317
4.6 Monitoring SE of c27-2 318
4.7 Monitoring SE of FL HSP27 320
4.8 NMR spectroscopy 320
4.9 Creatine phosphokinase aggregation assays 321
4.10 Native MS 321
4.11 Measuring relative redox potential with glutathione-mediated reduction potential and SDS-PAGE 322
4.12 Measuring redox potential with square wave voltammetry 323
5 References 324

Summary

Abbreviations

1D	One dimensional
2D	Two dimensional
ACD	α-crystallin domain
API	Antiparallel I
APII	Antiparallel II
APIII	Antiparallel III
ATP	Adenosine triphosphate
BME	β-mercaptoethanol
c27-1	Core domain of HSP27 – construct 1 as outlined in Chapter 3, Figure 2.3 and PDB ID 4MJH.
c27-2	Core domain of HSP27 – construct 1 as outlined in Chapter 3, Figure 2.3
cABC	A core domain of αB-crystallin
CCS	Collision cross section
CID	Collision-induced dissociation
CPK	Creatine phosphokinase
CSP	Chemical shift perturbation
CTP	C-terminal peptide
DC	Direct current
DT	Drift time
DT-IMS	Drift-time ion mobility spectrometry
DTT	Dithiothreitol
EDTA	Ethylene diamine tetraacetic acid
EM	Electron microscopy
ESI	Electrospray ionisation
FAD	Flavin adenine dinucleotide
FL	Full length; a protein construct comprising the full amino acid sequence of the protein
FPLC	Fast protein liquid chromatography
FRET	Fluorescence resonance energy transfer
HFF	Human foreskin fibroblasts
HSF	Human skin fibroblasts
HSP	Heat shock protein
HSP27 3D	Triple phosphomimic of HSP27
HSQC	Heteronuclear single quantum coherence
IAM	Iodoacetamide
IDPs	Intrinsically disordered proteins
IgG	Immunoglobulin G
IM	Ion mobility

IMAC	Immobilized metal ion affinity chromatography
IM-MS	Ion mobility mass spectrometry
IPTG	Isopropyl thiogalactose
MCP	Micro-channel plate
MD	Molecular dynamics
MES	2-ethanesulfonic acid
MS	Mass spectrometry
NCS	Non-crystallographic symmetry
nESI	Nano electrospray ionisation
NMR	Nuclear magnetic resonance
NOE	Nuclear Overhauser enhancement
PBS	Phosphate buffered saline
PDB	The RCSB Protein Data Bank, found at http://www.rcsb.org/pdb/home/home.do
PDB file	A file in the format required for deposition in the PDB, referring to the coordinates of a high resolution protein structure and associated molecules within the experiment.
PEG	Polyethylene glycol
PMSF	Phenylmethylsulfonyl fluoride
PN	Proteostasis network
PTM	Post-translational modification
PX	Protein crystallography
QToF	Quadrupole Time-of-Flight
RF	Radio frequency
RMSD	Root-mean-square deviation
RMSF	Root mean squared fluctuation
SAXS	Small-angle X-ray scattering
SDS-PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
SE	Subunit exchange
SEC	Size-exclusion chromatography
SEC-MALS	Size-exclusion chromatography multi-angle light scattering
sHSP	Small heat shock protein
SPG	Succinic Acid, Sodium Dihydrogen Phosphate and Glycine
SRIG	Stacked ring ion guide
TCA	Trichloroacetic acid
TEV	Tobacco Etch Virus
TLS	Translation, Libration, Screw tensors
ToF	Time-of-flight mass analyser
WT	Wild type

Chapter 1: Introduction

1 Biological Motivation

The sHSPs are ubiquitously expressed across all domains of life and critical for cellular proteostasis[1]. They are characterised by small monomer masses (approximately 12-42 kDa) that comprise a conserved α-crystallin domain (ACD) flanked by variable N- and C-termini. These monomers typically assemble into large oligomers which may be monodisperse[2,3], access a small number of stoichiometries, or be incredibly polydisperse accessing multiple stoichiometries over a wide range of masses[4–6]. The oligomers are dynamic, capable of readily exchanging subunits with one another[7,8] and of varying their stoichiometry and size in response to the environment[9–12].

The sHSPs are molecular chaperones, interacting with (partially) unfolded protein substrates to help maintain their native structure and prevent their improper association[13–17]. Notably, the sHSPs are stress proteins, displaying structural transitions, activity and upregulated expression under cellular stress such as temperature[6,10,12,18], ischemic injury[9], and oxidative stress[19,20], rather than their primary role being in *de novo* folding. Additionally, post-translational modification can interrupt their oligomeric profile, protective activity and cellular localization[5,13,21,22].The presence of other sHSPs and partially unfolded substrate both

lead to a disruption of sHSP homo-oligomers, through formation of sHSP hetero-oligomers[7,8,23], or sHSP-substrate complexes[10,24]. Taken together, these multiple modes of modulation produce flexibility in the sHSP network to respond to an assortment of cellular stresses.

The sHSPs' molecular chaperone function may also account for their involvement in numerous cellular processes (*e.g.* cytoskeletal maintenance, apoptosis inhibition, cell differentiation to name a few[25]), such that their multiple roles within the cell are all mediated by interactions with partially unfolded substrate[26]. It may also explain the implication of the human sHSPs in a diverse range of disease states such as cancer, cataract, motor neuropathies, neurodegenerative disorders, and cardiac myopathy[25,27,28]. It is the gravity of these conditions, alongside many others associated with protein misfolding, that has motivated the study of the sHSPs and their role in the cell. However, despite decades of study, our understanding of their mechanisms of action remains limited. For example, there is yet much to discover about the mechanisms that mediate the sHSP response to stress, their interaction with substrates, and their cellular regulation.

1.1 sHSP structure

The sHSPs have been difficult to characterise structurally to provide a rationale for their function. This is due to their heterogeneous and dynamic nature, both along the polypeptide chain and between oligomers. Non-metazoan proteins typically access a small number of stoichiometries under non-stressed conditions[29], and so have been studied most successfully by protein crystallography. The three oligomeric structures solved to atomic resolution all show the sHSP adopts polyhedral or ring-like cages with dimers forming the protomeric building

block along the edges[2,3,30] (Figure 1 – 1a). The dimer is formed via association of the ACDs on each subunit. Each ACD shows an immunoglobulin-like β-sandwich comprising up to nine β-strands, and in non-metazoans the β6-strand of one subunit forms backbone hydrogen bonds with the β2 of its partner to form an inter-subunit, composite β-sheet (Figure 1 – 1a). This has been recapitulated in truncated constructs comprising non-metazoan ACDs alone, which also form dimers, such that there are several high-resolution structures available from yeast[30], wheat[3], archeal[2,31] and bacterial sHSPs[32].

In addition to the intra-dimer interface, the two C-termini of a dimer bind to neighbouring dimers and thus mediate higher-order oligomerisation. This inter-dimer binding involves an IXI/V motif (where X is any amino acid) anchoring within the groove created by the β4 and β8 strands of the ACD (this groove contains numerous hydrophobic side chains) and hydrogen bonding between neighbouring residues at each interface, as well as ionic interactions being possible[2]. There is also substantial evidence for this IXI binding in animal sHSPs both by crystallography[33,34] and NMR[35,36] (with binding across the groove possible in both directions[37]), and mutation of the IKI motif to FKF in a bacterial sHSP decreased the species size from 24mer to dimer[38].

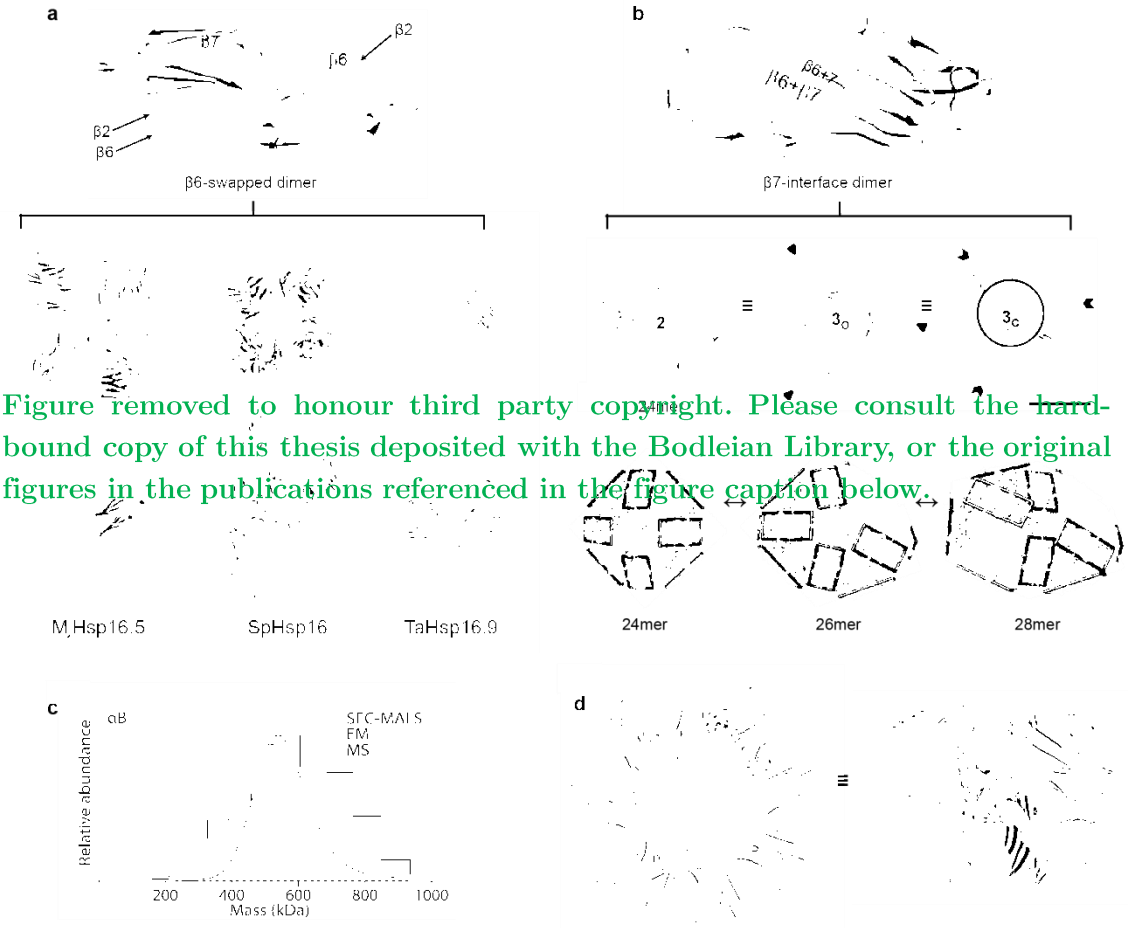


Figure removed to honour third party copyright. Please consult the hard-bound copy of this thesis deposited with the Bodleian Library, or the original figures in the publications referenced in the figure caption below.

Figure 1 – 1 The sHSPs form polyhedral and ring-like cages. a) In non-metazoan structures published to date, the dimer interface is formed between the β6 strand on the ACD of each subunit forming a composite β-sheet through hydrogen bonding to the β2 of its partner. These dimers then assemble into high order oligomers through the association of the C-termini, and with additional interaction of some of the N-termini (the N-termini are not always resolved). b) The metazoan dimer demonstrates the dimeric interface between two extended β6+7 strands, to form an extended β-sheet. Structural models of the dimeric building block have then been used to construct pseudo-atomic models by modelling with spatial restraints provided by cryo-EM[39] (density envelope in beige) or IM-MS (blue). c) A wide range of oligomeric species (and masses) of αB-crystallin is observed by native MS, EM or SEC-MALS[37], demonstrating its polydispersity. d) The crystal structure of the 32mer of Sip1[12]. Figure modified from Haslbeck and Vierling[13], Braun et al.[39], Baldwin et al.[40], Hochberg and Benesch[37], Fleckenstein et al.[12]

In the oligomeric structures (Figure 1 – 1a) the N-termini were wholly or partially unresolved, indicating that they were disordered within the crystal. Those that were observable were buried in the centre of the sphere. Along with other studies, this has indicated that they are buried[41], dynamic[42], occupy disordered states and non-equivalent conformations[30,43], and have a

propensity for helical secondary structure[30,44] (in both non-metazoan and metazoan structures).

In contrast to the non-metazoan ACDs described above, the metazoan ACDs demonstrate a different dimeric binding mode – the interface is formed between a β-strand on each subunit which comprises both the β6 and β7 residues (the 'β6+β7', Figure 1 – 1b). (The nomenclature for the β-strands employed across the field follows the original labelling in the *Mj*HSP16.5 and *Ta*HSP16.9 structures[2,3].) The main chain hydrogen bonding between these two 'β6+β7' strands leads to an extended β-sheet across the top of the dimer. This was first observed in protein constructs where the dynamic N- and C-termini had been removed to facilitate crystallisation or NMR interrogation, as many of the metazoan sHSPs are polydisperse (Figure 1 – 1c), as well as having flexible N and C-termini[45]. While other dimeric interfaces have been suggested[46,47], the β6-swapped dimer and β7-interface dimer are by far the most observed and so far have only been demonstrated in non-metazoan and metazoan sHSPs respectively (metazoan examples are human αB-crystallin[34,48–51], bovine and zebrafish αA-crystallin[34,52], human HSP27[20,51,53], human and rat HSP20[48,54]). Pseudo-atomic models of some of the oligomers formed by αB-crystallin have been developed using the dimeric building blocks along with hybrid structural approaches[35,39,55] (Figure 1 – 1b). Recently, the cryoEM and crystal structures of one of the oligomers of Sip1 (a developmental sHSP of *Caenorhabditis elegans*) demonstrated a 32mer formed from two symmetric rings each comprising eight dimers (Figure 1 – 1d), with the N-termini buried within the cage cavity, the IXI motif of the C-termini binding to neighbouring β4-β8 grooves (both intra-ring and inter-ring, depending on the position of the subunit), and the N and C-termini of subunits at the

poles being poorly resolved, indicative of increased flexibility[12]. This metazoan oligomer structure provided further weight to the hybrid models developed for αB-crystallin.

1.2 sHSP function

The N-terminal domain[43,56], ACD[33] and C-terminal domain[57] all appear to confer chaperone activity in different sHSPs and different substrates (*et cetera*), consistent with several binding sites being available to chaperone misfolded proteins dependent on their nature[43,58]. The N and C terminal sequences are non-conserved with variable length and sequence, whereas the ACD is highly conserved and the motif by which the sHSPs are defined[13]. However, even the ACD shows differences in behaviour from the same organism for the same substrate. For example, of the ten human sHSPs, αB-crystallin and HSP27 (HSPB5 and HSPB1 respectively[59]) have attracted the most attention due to their high profile in disease; polydisperse and dynamic nature; and constitutive and inducible expression in multiple tissues throughout the body[25]. αB-crystallin has been employed as the paradigm for polydisperse sHSPs in structural studies. Constructs comprising the core domains of these proteins ('cABC' and 'c27-1'), have 71% identical or highly similar residues (> 0.5 in the Gonnet PAM 250 matrix[60,61]) and possessed contrasting chaperone affinity *in vitro* for α-lactalbumin (amorphous aggregation), and fibril-forming κ-casein and $A\beta_{1\text{-}42}$[51]. When highly dissimilar residues in the poorly chaperoning c27-1 are substituted for their counterparts from cABC, the ability of c27-1 to chaperone aggregating α-lactalbumin is markedly improved (Figure 1 – 2, green traces). Indeed, even when simply three dissimilar regions that lie on the same edge of the β-sandwich (that including the β7 - β8 loop[51]) are substituted, there is an improvement in activity (Figure 1 – 2,

red trace). Thus the ACD encodes substrate specificity even though highly conserved. This allows the sHSPs to protect a broad range of distinct substrates when combined with their regulation by the numerous factors described above. Notably, HSP27 contains a cysteine residue on its β6+7 strand, which is able to form a disulphide with its equivalent on the dimeric partner[53]. This covalent modification provides it with an extra level of differentiation from other sHSPs.

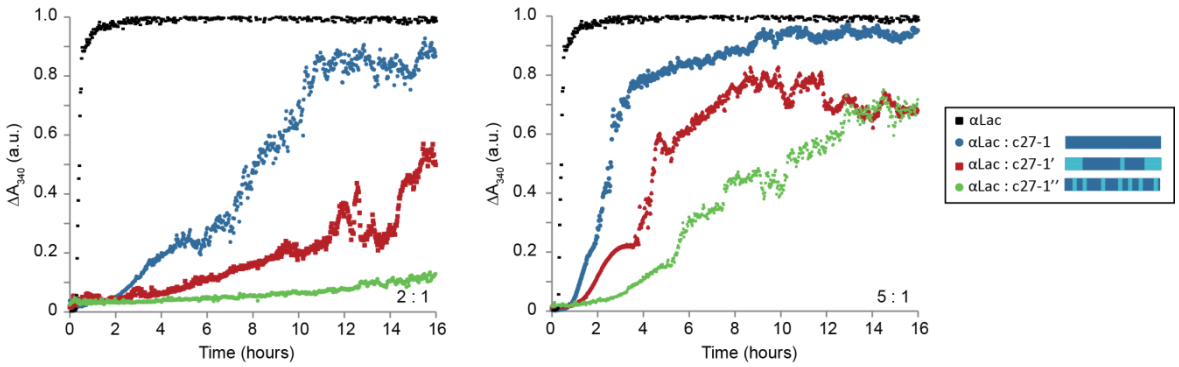


Figure 1 – 2 The highly conserved ACD encodes differences in chaperone activity *in vitro*. **Aggregation of 200 μM α-lactalbumin with 20 mM DTT at 37°C in the presence of 100 μM (left panel) or 40 μM (right panel) c27-1. HSP27 to αB-crystallin amino acid substitutions increases the ability of the core domain to prevent aggregation of the substrate, as monitored by the change in light scattering over time. c27-1 – the WT HSP27 core domain; c27-1' – the HSP27 core domain with the termini and two residues in the β7-β8 loop substituted by their equivalents in cABC (these contain several dissimilar residues that all cluster on the same side of the β-sandwich in the monomer[51]); c27-1'' - the HSP27 core domain with the 12 most strongly dissimilar residues between cABC and c27-1 substituted by their equivalents in cABC (schematic, c27-1 and cABC dark blue and light blue respectively). The c27-1 WT and cABC plasmid was generously provided by A. Laganowsky[51] in pET28a vectors with an N-terminal His-tag. Chimeras were produced by replacing c27-1 with synthetic genes (geneBlocks, IDT) in the pET28a vector. Proteins were expressed in E. Coli BL21(DE3) cells and purified by IMAC without tag cleavage with the buffers described in Section 4.3 of Chapter 3, and exchanged into 50 mM sodium phosphate, 100 mM NaCl, 2 mM EDTA, pH7.0 using a Superdex 75 26/60. Chaperone assays were performed as described elsewhere[51] using a Fluostar Optima plate reader (BMG Lab Technologies). The aggregation of α-lactalbumin was induced with DTT (both Sigma) at 37°C. Graphs show the average of triplicate samples at each time point (every 1.5 minutes for 16 hours), normalised to the mean light scattering of α-lactalbumin at the final three time points.**

αB-crystallin investigations have produced a wealth of insight into the subtleties of this complex protein, and by extrapolation, the other mammalian sHSPs. However, the high

clinical relevance of HSP27 and its additional complexity due to the intra-subunit disulphide demands the structural and dynamical characterisation of HSP27 in its own right. This thesis reports investigations of this protein, stemmed from observations in a new crystal structure of the core domain, and extended to the behaviour of the β2 and β3 on phosphorylation; the effect of neuropathy-associated mutation; and the role of the disulphide bond in the regulation of this protein, particularly in light of its response to oxidative stress. I hope to highlight how regulation of HSP27, either by post-translational modification or oxidation, can be rationalised from a combination of structural and functional insights.

2 Technical Strategy

In order to conduct structural and functional investigations into HSP27, we employed a range of techniques and experimental approaches. The structural exploration relies upon X-ray crystallography data and important dynamical understanding is gained from 'native' mass spectrometry (MS) of the intact protein complex. The work is based on a simplified system that was constructed by manipulating the protein sequence at the genetic level. These methods are presented within a wider network of technologies, which were each employed because of the additional insight they could bring. Here I will provide a brief introduction to native MS and protein crystallography, with a focus on the approaches employed in this project.

2.1 MS

Protein complexes can be maintained intact within a mass spectrometer, and so can be interrogated in their functional states using a wide range of MS experiments (this is often termed 'native MS' as an analogy with native SDS-PAGE where non-covalent interactions are also maintained). Chapter 2 provides an overview of the types of MS approaches that can be employed, and discusses the strengths of native MS which render it so powerful for the investigation of heterogeneous and dynamic protein systems, which are manifold within the proteostasis network.

Quadrupole Time-of-Flight (QToF) mass spectrometers were used in this project (Figure 1 – 3). They comprise two mass analysers (the quadrupole and ToF) which select and separate the ions based on their mass-to-charge ratio (m/z). Analytes are ionised and introduced to the gas phase by a nano electrospray source, transmitted through the various chambers of the mass spectrometer, before final detection. I will first describe the ionisation, analysis and detection of the macromolecular ions, before describing two additional technologies (collision-induced dissociation and ion mobility) that have revolutionised the MS field and are also employed in the Waters Synapt QToF used in this project (Figure 1 – 3)}.

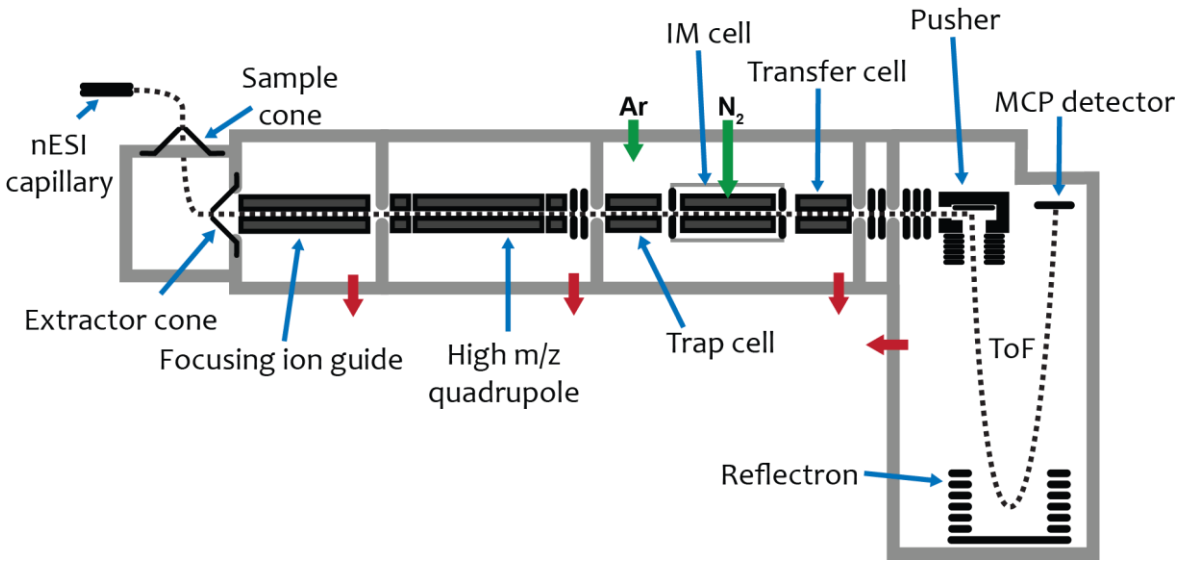


Figure 1 – 3 Schematic of a Synapt G1 QToF mass spectrometer capable of ion mobility separation. The spectrometer is modified for transmission of high m/z species with higher pressure in the initial chambers, a low frequency (high mass) quadrupole, and a pusher that introduces ion packets to the ToF with low sampling frequencies to allow higher m/z ions to reach the detector within each pulse. The chambers are under different degrees of vacuum (red arrows indicate pumping), Gases can be introduced in a controlled manner (green arrows).

2.1.1 Nano-electrospray ionisation

In order to ionise protein complexes, introduce them to the gas phase and deliver them intact to the mass spectrometer, nano-electrospray ionisation (nESI) was employed. This is a miniaturised variant of the widely-used electrospray ionisation (ESI). Since it was initially coupled to MS[62], ESI has revolutionised the study of biological macromolecules by MS, most notably within the field of proteomics where MS is used to identify peptide fragments of proteins, and its impact was recognised by the partial award of the 2002 Nobel Prize in Chemistry (J. B. Fenn).

nESI (and ESI) is conducted at atmospheric pressure. The protein sample in solution is placed within a gold-plated capillary and a high electrical potential (typically 1-1.6 kV) is applied between the capillary and the sample cone at the entrance of the mass spectrometer (Figure 1

– 4). Under the influence of the electric field, positive electrolytes in solution will move towards the tip of the capillary and form a cone (a 'Taylor cone'[63]). When the repulsion of the like charges and the force of the electric field become sufficiently large to overcome the surface tension of the cone, a filament will protrude from the cone and small droplets will break from the end. These droplets (containing solvent, electrolytes and in a small fraction, the proteinaceous analyte[64]) will then accelerate towards the entrance of the mass spectrometer under the influence of the electric and pressure gradients. Along the way, as solvent evaporates from the droplets, their volume shrinks and their charge increases. When their charge-to-surface-area ratio is sufficiently high, they will undergo further asymmetric fission events where small daughter droplets are emitted from the parent droplet (carrying 1% of the mass but 15-25% of the charge[65]). Each of these daughter droplets will in turn undergo multiple fission events with concomitant size decrease. There has been much discussion over the way in which the final 'naked' analyte ion is produced, though it is widely accepted that for folded biological macromolecules, the droplet evaporation and fission continues until the charge left on the analyte is inherited from the droplet that contained it an instant before[64,66]. The analyte ions are then guided into the mass spectrometer by the electrical and pressure gradients. In Waters QToF instruments, the capillary is typically positioned such that the sample cone is located on the edge of the plume of droplets, and the subsequent extraction cone is positioned perpendicular[67]. This 'Z source' layout reduces the number of neutral molecules entering the spectrometer (Figure 1 – 3).

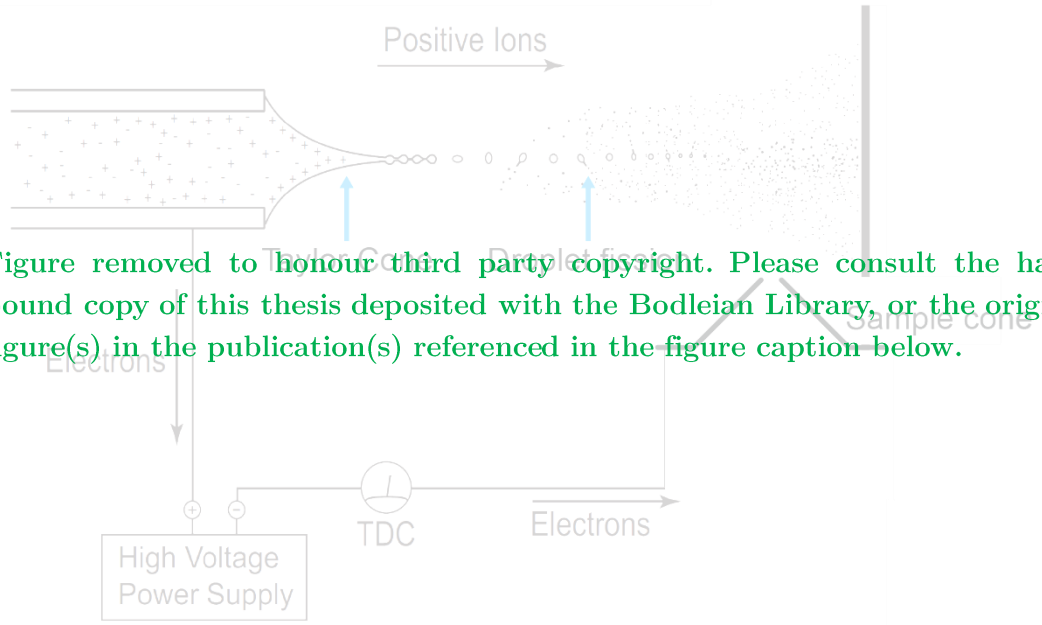


Figure 1 – 4 The electrospray ionisation process to produce positively charged analyte ions. Figure modified from Kebarle and Verkerk 2009[65].

ESI results in multiply charged species from the addition of protons (and sometimes cations) to the analyte, $[\text{Protein} + n\text{H}]^{n+}$, with a narrow range of n for folded proteins which varies due to several parameters including the surface area of the protein[65,68]. This results in a 'charge state distribution' (*e.g.* Figure 1 – 7b, orange species) which is typically Gaussian[68]. This multiple charging brings the m/z value of very large complexes into the easily-tractable range of the mass spectrometer. Additionally, the numerous peaks represent multiple observations of the macromolecular ion so improve confidence in the assignment. ESI results in little fragmentation (*i.e.* it is a 'soft ionisation' technique) which along with the multiple charging highlighted its utility for native MS.

nESI is now favoured over ESI for native MS as the droplet sizes are substantially smaller (µm-range versus <200 nm[64]), resulting in more efficient desolvation. This means that heating and organic solvents, which are often employed with ESI to facilitate desolvation but would

disrupt non-covalent complexes, can be avoided. Additionally, nESI requires less sample, leads to fewer salt adducts[64], is less likely to produce non-specific oligomers as a result of the co-habitation of a droplet[66], and results in higher ionisation efficiency, and seems to preserve non-covalent complexes more successfully due to fewer collisions in desolvation[64], all making it the technique choice for the interrogation of non-covalent complexes.

nESI is undertaken in a volatile buffer such as ammonium acetate, which is most widely used, typically at neutral pH. Contrary to the typical solution conditions in protein preparation and structural biology, volatile buffers desolvate easily during the ESI process, and reduce the number of cationic adducts (most notably sodium and potassium ions) detected in the mass spectrum[69]. However, trace amounts of cations often persist after buffer exchange[70], exemplified by equally spaced peaks following the protonated protein peak at each charge state. These adduct peaks can be used to confirm the assigned charge[‡] but as the signal for the protein is distributed over multiple peaks for each charge state, they will also reduce the signal-to-noise ratio and lead to less accurate mass determination if not well resolved[71]. As such it is necessary to optimise sample preparation; the capillary shape and position such that efficient desolvation occurs; the acceleration voltages used in the source region of the mass spectrometer for further adduct removal without dissociation of the specific complexes under investigation; and to analyse the data appropriately to record accurate mass measurement and species abundance. Additionally, neutral solvent vapour can be introduced into the source region, which reduces the amount of adduction[71]. This has the additional advantage of charge

[‡] The spacing between the peaks for a protonated and cationised species will be $\frac{Mass\ of\ cation - Mass\ of\ proton}{Charge}$.

reduction which gives better peak separation and promotes the maintenance of intact protein:ligand complexes due to a lower kinetic energy of the ions[70].

On entry to the rough vacuum of the mass spectrometer, the ion plume expands significantly due to charge-charge repulsion and the pressure gradient. They are focused by an ion guide prior to entry into the quadrupole. Transmission of large macromolecular species is aided by an increased pressure in this region, as collision with neutral gas molecules decreases the radial momentum of the ions so that they are more efficiently focused by the ion guide[64,72].

2.1.2 Mass analysers

Quadrupole analysers select ions based on their m/z ratio by inducing stable and unstable ion trajectories. They comprise four parallel rods (Figure 1 – 5) to which a direct current potential and radio frequency (RF) potential are applied. Each rod has the same potential as the rod opposite, with a potential of +Φ on one pair and –Φ on the other where:

$$\Phi = U - V\cos(2\pi\nu t) \quad (1)$$

where U is the amplitude of the DC voltage, t is time, and V and ν are the amplitude and frequency of the RF voltage respectively[73]. The value of ν is kept constant[64] while U and V are varied ($U < V$)[73].

Figure 1 – 5 A quadrupole mass analyser. Each rod has the same potential applied as its partner opposite (+Φ or –Φ, Eqn. (1)). Ions will be deflected from their path through the middle of the rods by the applied voltages, according to their m/z values (blue). Any that maintain a stable trajectory will pass through the quadrupole and onto further stages of the mass spectrometer (red). Analyser position is shown within a Synapt QToF instrument (see Figure 1 – 3). Figure kindly provided by Ellie Moore and modified.

Let us consider the +Φ rods first. As the focused ion beam enters the quadrupole, the trajectories of the positive ions will be destabilised by the RF potential, which is periodically changing polarity and so will cyclically repel and attract the ions. Ions with a low mass or large charge will be most greatly accelerated away from or towards the rods. Only ions with stable trajectories will be transmitted through the exit. If an ion is sufficiently deflected from its trajectory through the centre of the rods, it will either exit through the side of the quadrupole or hit a transiently negative rod and be discharged. If the magnitude of V is increased, ions with an increasing value of m/z will be destabilised. The small DC potential on both rods will act to repel any ions between them and so further stabilise any ions with large m/z.

However, the negative DC potential on the -Φ rods will act to attract the positive ions passing through the quadrupole, and so deflect their path. The RF potential may prevent this destabilisation by modifying the ion path in the opposite direction. It will have a greater effect on ions with a low m/z.

In this way, the +Φ rods will select for ions above a given m/z (determined by the value of U and V), and the -Φ rods will select for ions below a given m/z, such that only ions permitted by both +Φ and -Φ voltages will be transmitted. The quadrupole can therefore be used to select for a narrow range of m/z values. Additionally, if U=0 ('RF only mode'), the quadrupole will act as a broad mass filter, allowing a range of m/z ions to be transmitted efficiently through the mass spectrometer (~0.8 – 5 $(m/z)_{set}$, where is $(m/z)_{set}$ is determined by the value of V[72]). This permits the user to set the quadrupole to scan between specific values where the m/z area of interest lies, and is the primary mode employed for the work herein.

The maximum m/z value that can selected by a quadrupole is:

$$(m/z)_{max} = \frac{7 \times 10^6 \, V_{max}}{\nu^2 r_0^2} \quad (2)$$

Where r_0 is the inner radius between the rods, V_{max} is the maximum RF amplitude applied, and ν is again the frequency. Therefore, quadrupoles with a low frequency are employed to transmit non-covalent complexes[72,74] (typically 300 kHz, which can mass select up to m/z = 32000 Th and transmit to over 85000 Th[72]).

The quadrupole can also be employed as a mass analyser in its own right – a detector records the intensity of ions transmitted through the quadrupole as the values of U and V are linearly increased, and thus a well-resolved mass spectrum over a range of m/z values is produced. However in MS of protein complexes, they are more commonly employed as a mass filter before further interrogation of the ions in a collision cell ('Trap' and 'Transfer', Figure 1 – 3) or IM cell, finally followed by mass analysis. This overcomes a poorer resolution of the quadrupole selection which is introduced on reducing its frequency of operation[72].

The mass analysis in a QToF instrument is performed by the Time of Flight (ToF) analyser. In its simplest form, a ToF analyser comprises a constant electric field followed by a field-free drift region and detector. Ions are accelerated by the electric field such that their electrical potential energy is converted to kinetic energy.

$$zeV = \frac{1}{2}mv^2 \tag{3}$$

Where z is the charge on the ion, e is the elementary charge, V is the electrical potential (typically 9.1 kV), m is the mass of the ion and v is its velocity. So ions of the same charge will gain the same amount of kinetic energy (but their velocities will differ if their masses differ).

Therefore an ion traverses a drift region of length d in time t (its time-of-flight) where:

$$t = \frac{d}{v}$$
$$t = \sqrt{\frac{m}{z} \cdot \frac{d^2}{2eV}} \tag{4}$$

Conversely,

$$\frac{m}{z} = \frac{2eV}{d^2} \cdot t^2 \tag{5}$$

Thus the m/z value of each ion can be calculated from the time it arrives at the detector. Notably, as mass analysis depends only on t, ToF analysers have a theoretically unlimited mass range which is why they have proven so successful in the field of native MS[64,75].

As ToF analysis depends on pulses of ions being introduced to the drift region by the pusher, the ToF tube is oriented orthogonally to the continuous ion beam from the earlier chambers

(Figure 1 – 3). This reduces any initial distribution of kinetic energy along with electrostatic lenses that focus the ions as they enter the ToF. Despite this, ions of the same charge will not necessarily possess the same kinetic energy within the drift region, due to collisions within the imperfect vacuum of the ToF, possible decomposition, and initial spread of the ion packet. This range of kinetic energies will alter the time-of-flight for ions of the same m/z, resulting in broader peaks which are less accurate and more difficult to resolve.

To overcome this problem, an ion mirror ('reflectron') is employed at the base of the ToF tube (Figure 1 – 3). The electric field in the reflectron will decelerate the ions so that they return in the opposite direction. If one ion has equal m/z to another but a higher kinetic energy, it will traverse the field free regions of the ToF tube in less time. However, it will also penetrate the reflectron more deeply and so spend longer within the ion mirror, such that the two ions arrive at the detector at the same time[76]. Additionally, the reflectron ameliorates the resolution of the instrument as it effectively doubles the length of the ToF tube.

2.1.3 The detector

Once ions have been separated within the ToF tube, they are detected with a detector containing a micro-channel plate (MCP, Figure 1 – 6). This comprises a thin plate (*ca.* 4 cm diameter) containing many parallel microscopic channels that each act as an electron multiplier. Thus if an incident positive ion (red, Figure 1 – 6) arriving at the MCP has sufficient energy, it will cause the emission of a small number of electrons that begin an electron emission cascade. At the end of the channels, the large number of electrons hits a

scintillator, causing the emission of photons which are then amplified and converted to an electrical signal in a photomultiplier tube which is recorded digitally.

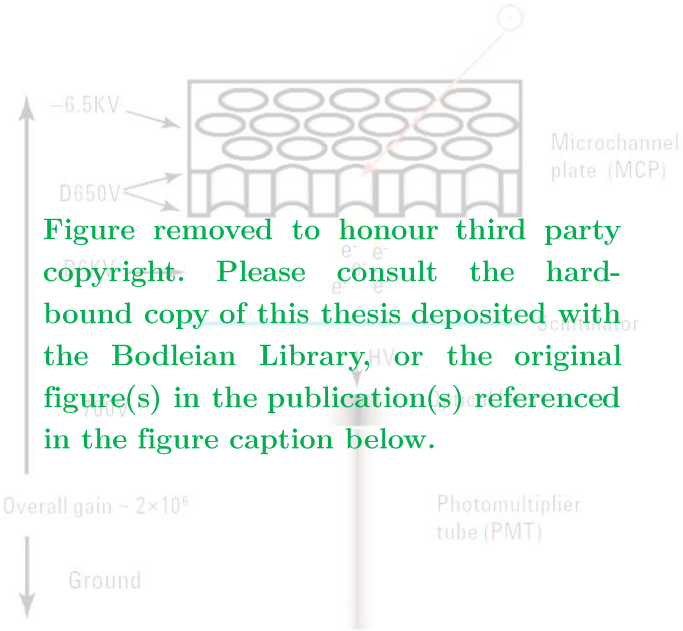


Figure removed to honour third party copyright. Please consult the hard-bound copy of this thesis deposited with the Bodleian Library, or the original figure(s) in the publication(s) referenced in the figure caption below.

Figure 1 – 6 An example of an MCP detector employed at the end of a ToF mass analyser. Figure from Agilent's Time-of-Flight Mass Spectrometry Technical Overview (www.agilent.com/chem, 5990-9207EN).

As an example of the utility of native MS in studying protein complexes, Figure 1 – 7 shows some spectra collected on variants of the sHSP Acr2 from *Mycobacterium smegmatis*, using the MS set-up described above. These variants were designed to investigate the role of the C-terminus in the association of the protein. SEC profiles showed that the FL protein eluted from the SEC column much earlier than constructs missing the final eight residues (Acr2(ΔC8)) or where the C-terminal IAI motif had been mutated to FAF (Acr2(I140F,I142F)). However, SEC was unable to resolve the apparent masses of the smaller species occupied by these variants. MS spectra demonstrated that the Acr2(ΔC8) sample contained a dimer and monomer (including a truncated form), as well as an octamer,

producing a rationale for the two peaks seen on the SEC column. Acr2(I140F,I142F) showed clear peaks for the dimer but not the monomer, and numerous overlapping peaks at higher m/z corresponding to higher order oligomers, despite the apparently similar SEC profiles for the two samples. This demonstrates that the C-terminus of Acr2 is critical in the oligomerisation of Acr2 to form the 12, 18 and 24mers of the FL (including impeding oligomerisation with steric bulk on the IAI motif), consistent with the role of the C-terminal IXI motif in the oligomerisation of other sHSPs. It also demonstrated the value of the mass resolution afforded by MS.

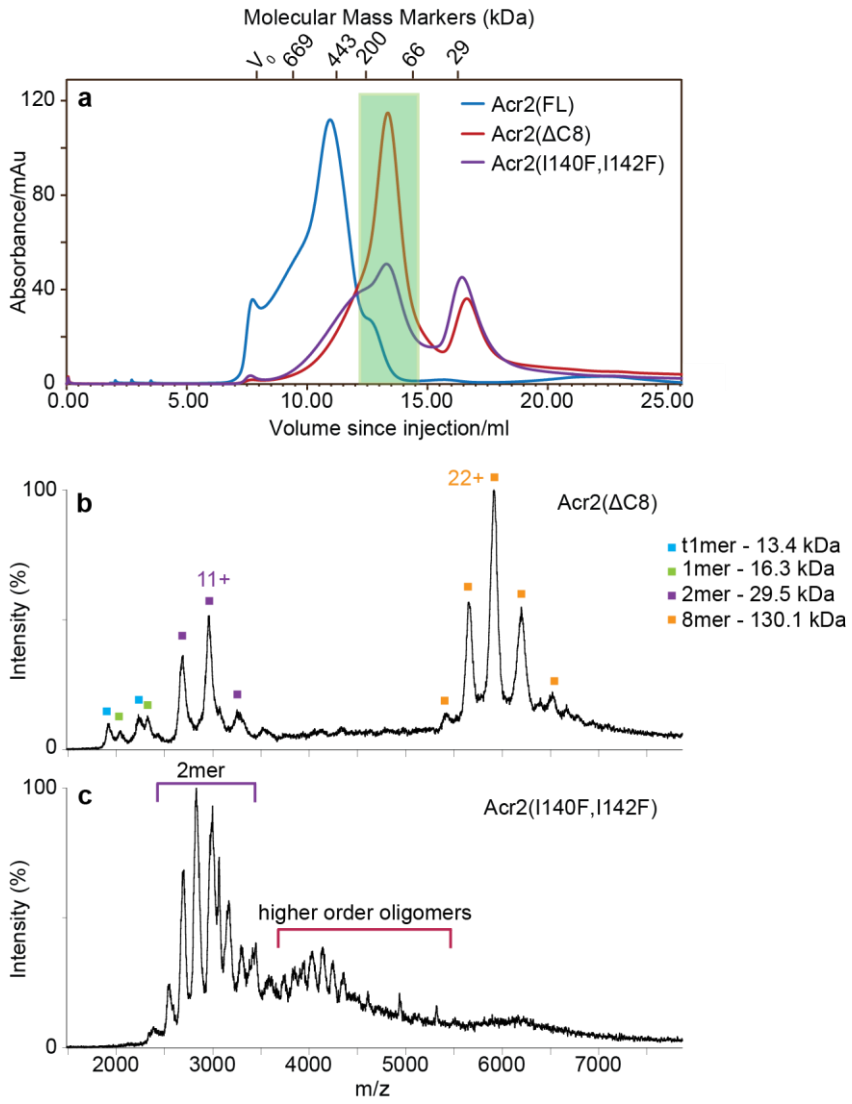


Figure 1 – 7 SEC profiles of sHSP Acr2 variants (a) and their mass spectra collected on a QToF instrument (b and c).

In addition to the minimal components required to build a mass spectrometer (namely the ionisation source, mass analyser and detector, as described above), numerous instrument modifications have been developed to elucidate further insight on ions during their transmission through the mass spectrometer. For protein complex ions, collision-induced dissociation CID and ion mobility greatly enlarged the MS investigations possible, and both are employed regularly in the investigation of the sHSPs.

2.1.4 Collision-induced dissociation

Acceleration of the analyte ions in the presence of neutral gas molecules causes them to undergo energetic collisions. This can be employed to enhance desolvation of solvent molecules and buffer ions (including the cations discussed in Section 2.1.1)[64], or for intentional disruption of the protein complex itself. The latter is typically undertaken in a 'collision cell', which is filled with an inert collision gas such as argon (the trap and transfer cells on a Synapt instrument, Figure 1 – 3). The protein complex ions are accelerated into the cell by a potential difference, typically 50-200V when trying to induce dissociation. This drop in potential energy leads to an increase in kinetic energy of the ions proportional to their charge and the potential difference[64]. In turn, the kinetic energy is converted to internal energy on collisions with the bath gas. As the accelerating voltage is increased, there is increasingly removal of adducts, possible restructuring of the protein ion, and then unfolding and dissociation of a highly charged monomer which leaves the 'stripped' complex ion with concomitant decrease in charge[77]. This collision induced dissociation ('CID') mechanism can cause multiple sequential stripping events and fragmentation of the stripped monomers at the highest acceleration voltages[77]. The stripped oligomers have better peak separation due to their lower charge, allowing the assignment of complex spectra[6,78]. This is particularly useful when ions at a specific m/z are selected in the quadrupole and then exposed to CID to produce dissociation products and identify the subcomplexes of the species[79], though CID is highly useful in many MS experiments ranging from topology-elucidation[80] to the stability afforded by different ligands to membrane proteins[81].

2.1.5 Ion mobility MS

MS capabilities have been further greatly expanded by coupling to ion mobility (IM) spectroscopy. IM discriminates between ions based on their transport through electric fields. In IM-MS, the ions are then identified by their m/z value. The simplest implementation is a linear drift tube (DT-IMS) where ions are accelerated through a gas-filled cell by a constant field E. Ions with a larger collision cross section (CCS) will undergo more collisions with the neutral bath gas and so be more retarded in their transit through the cell. If the electric field is weak the velocity of an ion, v, is directly proportional to its mobility constant, K[82]:

$$v = KE \tag{6}$$

K in turn is inversely proportional to the CCS (Ω) under appropriate experimental conditions[83].

IM extended the application of native MS in several ways. Firstly, it introduced an additional dimension of separation in the IM-MS experiment, such that peaks that would overlap in the m/z dimension could be separated according to their drift time through the cell[84,85]. Next, it provided information on the shape of proteins from the CCS parameter, allowing the user to observe conformational variation[86,87] and discriminate between topological models for proteins, including those of highly dynamic and hetereogeneous protein complexes[11,55], or complexes containing a variety of subunits[80,88]. Recently, IM-MS has been used to follow the gas phase unfolding of proteins when collisionally activated[89–92], allowing determination of the stability afforded by ligands that was not detected by solution phase measurements[91].

The use of IM-capable QToF instruments, such as the Waters Synapt series, has been highly fruitful on targets ranging from the small molecule through the peptide and up to the protein complex level[87]. Synapt spectrometers incorporate IM after the quadrupole and a collision cell (trap cell, Figure 1 – 3), and before a second collision cell (transfer cell, Figure 1 – 3). Rather than a linear drift tube, their ion mobility separation is performed with a travelling wave ion mobility cell[93,94] comprising a stacked ring ion guide (SRIG) with both an RF and 'travelling wave' DC voltage applied (Figure 1 – 8). The RF potential is applied to adjacent electrodes with opposite phases, which radially confines ions as they pass through the aperture in the centre of the electrodes. To produce the travelling wave, a transient DC voltage is applied between two adjacent rings, then the next two rings then the next and so on (this is replicated at multiple points along the SRIG, so that there are multiple peaks and troughs of the wave travelling forwards)[93]. If the DC voltage is sufficiently high, ions are driven away from the potential hills and so propelled forward, which reduces the time it takes for them to transit the guide (this produces an effective ion guide[93], such as those found in Synapt instruments in the trap and transfer cells and prior to the quadrupole, Figure 1 – 3). However, if the height of the travelling wave (*i.e.* the magnitude of the DC potential) is below a critical value, ions that are being retarded by collisions with the neutral bath gas may not always be pushed forwards by the DC wave; instead collisions with the gas may prevent their propulsion forward by the wave, and ions are said to have 'rolled over' the wave (*i.e.* the wave of high potential has overtaken them). The transit time of these ions will thus be longer; the ion will have a lower mobility on account of its larger CCS causing more collisions with the bath gas. In the extreme, if the ions are not propelled forward by the wave at all, they may become trapped in

the cell. However, ideal separation is achieved when the ions 'surf' the waves[83], being propelled forwards and occasionally rolling over, with ions with increasingly large CCS rolling over the waves more frequently than smaller ions.

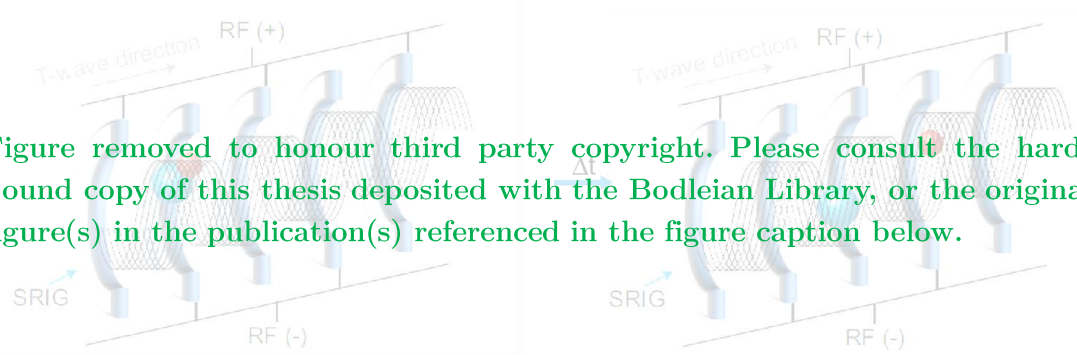


Figure removed to honour third party copyright. Please consult the hard-bound copy of this thesis deposited with the Bodleian Library, or the original figure(s) in the publication(s) referenced in the figure caption below.

Figure 1 – 8 Schematic of a travelling wave IM cell. Opposite phases of an RF voltage are applied to neighbouring ring electrodes to produce radial confinement in the SRIG. A DC potential is then applied to two neighbouring rings, and also repeated at an electrode pair six pairs further down all along the device (61 electrode pairs, 185 mm[94]). Over time, the DC voltage is transferred to the next electrode pair (and its equivalents six pairs further down etc.) and then to the next, and so on. This produces a wave with multiple peaks of potential that propel the ions forward (the potential is superimposed in black here, where the y-axis shows the amplitude and the x-axis shows the wave period along the cell; in space the potential is applied radially through the ring electrode). Colisions with the bath gas retard the ions under the influence of the travelling wave and may cause them to roll over some of the peaks of DC potential. Ions with a greater CCS are retarded more and so undergo more roll-over events (blue versus red sphere). Figure modified from Lanucara *et al.* 2014[87].

Roll-over events can be increased by raising the gas pressure, lowering the travelling wave height or increasing the travelling wave velocity (as an ion will undergo a greater number of collision per µs and have an increased chance of successfully rolling over). This in turn increases the separation between ions of different mobilities, which leads to better characterisation of their CCS values. These parameters are varied to optimize separation while still ensuring that all the ions have traversed the collision cell before the next ion packet is released into the entrance of the IM cell[95].

In each IM collection, ions are gated by the last electrode on the trap cell so that they can be introduced as an ion packet into the IM cell. They then transit the IM cell in different

amounts of time according to their mobility (and CCS). For each packet of ions released into the IM cell, ions are detected in two hundred sequential bins after their IM separation, transmission through the transfer cell and ion optics, and analysis in the ToF. Thus each bin is linked to a mass spectrum of the ions leaving the IM cell at a particular time point. Ions with a low CCS will leave the IM cell quickly, and so be detected in some of the earliest spectra recorded.

While CCS can be calculated directly from DT-IMS drift times and potentials, travelling wave IM relies on calibration (this is because the theory underpinning DT-IMS is better characterised than that of travelling wave IM[83]). By comparing the drift time of protein ions under particular conditions with the drift time of calibrants of known CCS, alongside the mass of the gas molecules and the mass and charge of the ions identified from the m/z dimension, it is possible to calculate their CCS[96]. An accurate calibration requires that the original CCS measurements of the calibrants was undertaken in the same bath gas[97] (nitrogen here), and that calibrants over a range of mobilities are used[83]. Additionally, if using the CCS values to model the topology of the complex in solution, it is critical that the accelerating voltages in the stages before the IM cell are as soft as possible to prevent gas-phase unfolding.

2.1.6 Estimating CCS values from structural models

In order to provide a structural basis for the CCS calculated from IM-MS, the CCS of each candidate model is calculated which allows selection of those candidates that best agree with the experimental IM data[55]. Multiple algorithms have been proposed to estimate the rotationally averaged CCS of a structure (averaged because the molecules will be randomly

oriented as they enter the IM cell so there is a spread in drift times recorded for the population). Each algorithm incorporates different assumptions, approximations and limitations depending on the particular molecular targets and applications for which it was developed. There is a trade-off between the number of approximations employed and the computational expense of the calculation, which is considerable for macromolecular species. The trajectory method[98] (TJM) is the most computationally expensive and most detailed approach, modelling the long range interactions of a gas probe incident on the target structure, and considered one of the most accurate sources of CCS approximation[99,100]. The projection approximation (PA) is the simplest method, using the averaged projected area of the target[101] and the size of the gas probe in IM as an estimate of the CCS[102]. This leads to a systematic underestimation of the CCS, which can be exploited to calibrate PA CCS estimates (CCS_{PA}) with TJM estimates. This has been shown to produce agreement between the two approximations within 2 % on test sets encompassing structures across a wide range of sizes[99,100]. However, even though the TJM is the 'gold standard' for CCS estimation, it slightly overestimates the CCS values calculated from experiment (CCS_{exp}). Therefore, an empirical relationship can be used to predict the experimental CCS for a given structure[103], namely,

$$CCS_{exp} = CCS_{PA} \times 1.14 \tag{7}$$

The PA has recently been implemented extremely efficiently[99], rendering it immediately accessible for rapid, accurate CCS estimation, and thus the selection of candidate topological models based on experimental CCS values[11,55].

2.2 Protein Crystallography

A great deal of our structural understanding of proteins on the molecular level is based on observations from protein crystallography. In this technique, the diffraction of X-rays by prepared protein crystals is analysed to produce a map of the electron density, into which a model of the protein structure can be built. This is incredibly powerful, and in the very low resolution regime (>4 Å) allows observation on the assembly, assembly interfaces, subunit fold and presence of α-helices[104]. For crystals with a slightly better resolution, β-sheets can be identified, and the electron density surrounding the main chain and side chains is more defined, such that some side chain conformations can be modelled (> 2.7 Å). In the 'medium' resolution range (*ca.* 2.7 – 2 Å)[105], main chain carbonyls and ordered waters are observed[104], and side chain identity and their rotamers can be assigned with increasing confidence into the high resolution range (2 – 1.2 Å), such that individual atoms can be resolved (1.5 Å). 'Atomic' resolution structures (< 1.2 Å) may even show defined electron density for electron poor hydrogen atoms[105].

With the detail accessible by protein crystallography, we can develop a structural rationale for the function of a protein. For example we can characterise protein:ligand and protein:protein interaction sites, identify unexpected binding partners[106], and observe the architecture of active sites and the influence of substitution. This information can then be applied in concert with that from a cohort of complementary techniques to understand the behaviour of the protein in its native *in vivo* environment.

For example the first crystal structure of a sHSP, HSP16.5 from *Methanococcus jannaschii*[2], showed that the protein exhibited a β-sandwich monomer fold, that monomers associated into dimers with formation of a β2-β6 interface, and that these dimers formed the edges of a 24-membered octahedral assembly. This detailed information provided the sHSP community with the first residue-level information on the structure of the protein, and guided a multitude of studies to further investigate the role of the interfaces, model assembly of the oligomers, and form hypotheses on the substrate binding and chaperone activity of this protein and the sHSP family.

Here I will briefly describe some of the underpinning theory of X-ray diffraction within the framework of the practical steps required to go from a purified protein to producing a reliable structural model that can be biologically interpreted.

2.2.1 Protein crystallisation

In order to achieve a measurable intensity of signal, conventional X-ray diffraction requires an ordered array of the protein macromolecules. This is achieved by preparing the protein in such a way as to encourage protein crystallisation, where the protein precipitates out of solution to form a three-dimensional crystal lattice composed of many millions to billions of protein molecules. The X-rays are then diffracted by the electron density in each of these identically oriented molecules, and X-rays diffracted at the same angle will constructively interfere if the angle satisfies a geometrical relationship (Bragg's Law, see below). This produces a strong signal that can be measured on a detector.

Preparation of protein crystals generally involves overexpression and purification of the protein target to produce a high concentration of pure protein. This can then be used in crystallisation trials in the presence of a crystallisation buffer in various different strategies. In this project I use a vapour diffusion crystallisation strategy whereby the aqueous protein solution (typically *ca.* 15 mg/ml) is mixed in near equal volume ratios with an aqueous crystallisation solution and a droplet of this mixture is suspended within a sealed well above a reservoir of the same crystallisation solution. As the concentration of the solutes in the reservoir solution is higher than that in the protein droplet there is net water transfer within the closed system as water evaporates from the droplet and condenses on the reservoir solution surface (*i.e.* vapour diffusion). This causes both the protein and precipitant concentration in the protein droplet to slowly rise as the system comes to equilibrium, hopefully enough so that nucleation occurs and the protein precipitates out of solution ('precipitant' refers to the solutes in the crystallisation buffer which are used to encourage precipitation). This precipitation causes the protein concentration in the solution to drop slightly, which encourages growth of the formed crystals[107]. This is a hanging drop arrangement, and alternatively sitting drop vapour diffusion was used in which the protein droplet sits on a pedestal surrounded by a moat of the reservoir solution.

Crystallisation can be the most time-consuming and laborious step in the process, with well-diffracting crystals of difficult targets sometimes taking years to optimise or remaining enigmatic. The experimentalist may employ a range of approaches to improve chances of crystallisation. For example, screening matrices of solution conditions that have varying buffer and precipitant identities, concentrations and pH (commercial screens are available that help

with this, based on conditions that have previously produced successful 'hits'); varying protein concentration and the temperature at which the multi-welled crystallisation plates are kept; co-crystallisation with a ligand (which may help crystallisation as well as providing biological insight); and conducting limited proteolysis to identify the most rigid section of the protein that might be more likely to form ordered crystals if the full length protein has highly flexible regions. Previous knowledge on successful crystallisation of similar protein targets can guide this search (as has been the case for the mammalian sHSPs), and any promising crystallisation conditions are further optimized in the hope of producing well-diffracting crystals.

Notably, crystallisation is not required for all forms of crystallography and the construction of X-ray free-electron lasers over the past seven years with ultra-bright high energy X-rays has allowed serial femtosecond crystallography, which can be conducted on a solution of micro-crystals and may advance towards single-particle structure determination in the future[108].

Once sufficiently large crystals have been produced, such that they can be harvested with a small nylon loop, they are briefly soaked in a cryoprotectant (here generally prepared from the reservoir solution with 20 % glycerol) before transfer to a storage puck in liquid nitrogen. This protects the crystal during data acquisition and on its transit to the X-ray source, which was in this project a synchrotron light source.

2.2.2 X-ray diffraction

At a synchrotron, a series of particle accelerators accelerate electrons to near the speed of light. As the path of the negatively charged electrons is bent by powerful magnetic fields into

near-circular motion in the 'storage ring', the electrons emit high energy electromagnetic radiation in the X-ray region. Additional magnetic insertion devices cause the electrons to wiggle and thus increase the intensity of the X-ray radiation further. The emitted X-rays are channelled tangentially to the ring into the 'beamlines' where they are focused and processed to produce monochromatic X-rays all travelling in the same direction, at tunable wavelengths[107]. For example, the I04 beamline at the Diamond Light Source has a tunable wavelength of 0.708 - 2.066 Å. The protein crystal is mounted within the beamline experimental hutch, and X-rays directed at the crystal are diffracted by the electron density in the many repeating units of the crystal. This diffraction is detected by a detector (here Pilatus 6M detectors, in which X-rays incident on the silicon surface produce an electrical signal counted by an integrated circuit behind each pixel) and a diffraction pattern of numerous spots or 'reflections' is produced at each angle of orientation of the crystal, characteristic of the crystal contents.

2.2.3 The unit cell and its space group

The unit cell is the smallest repeating unit of the crystal that when translated in three dimensions reproduces the crystal exactly. It has edges **a**, **b**, **c** of length a, b, c and angles between the edges of α, β, γ. The relationship between the edges and between the angles defines the 'lattice system' of the unit cell and the lattice type is defined by the position of lattice points (points within the crystal that have identical environments).

The asymmetric unit is the largest unit in the crystal which does not possess internal symmetry itself but on application of certain symmetry operators, will reconstruct the unit cell. These symmetry operators are described by the space group of the crystal (and unit cell).

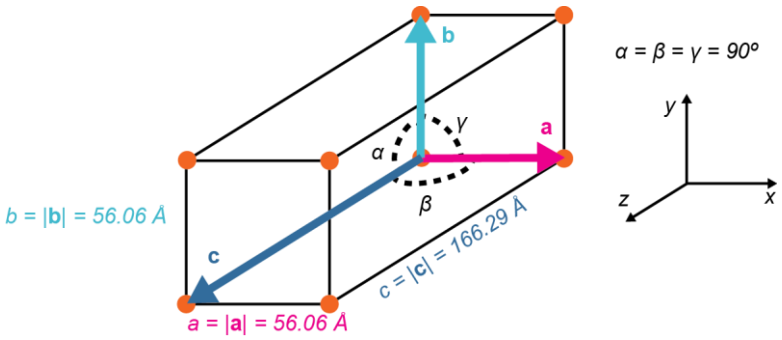


Figure 1 – 9 Example of a tetragonal unit cell.

For example, in some of the crystals presented in this work, we observed the P 4_3 2_1 2 space group, which describes a tetragonal lattice system of the unit cell with $\mathbf{a = b \neq c}$ and $\alpha = \beta = \gamma = 90°$C. The 'P' represents the lattice type, which is primitive here, meaning that the unit cell has lattice points only at its vertices (rather than also at its centre or on some or all of its faces). Together the lattice system and lattice type determine the 'Bravais' lattice of the crystal. The '4_3 2_1 2' of the space group describes the symmetry operators required to construct the unit cell from the asymmetric unit, and the details of the architecture of a P 4_3 2_1 2 space group or any other can be found in crystallography space group tables. It is possible for multiple types of unit cell to describe the crystal, and the one with highest symmetry is usually chosen by convention.

2.2.4 The reciprocal lattice

The crystal lattice (in real space) contains numerous planes which transect lattice points on unit cells across the crystal in a periodic fashion. Each of these planes is part of a set of

equivalent parallel planes throughout the lattice with the same orientation, and each set is defined by the indices *hkl* (the 'lattice' or 'Miller' indices) which describe into how many segments the set of planes cuts the **a**, **b**, and **c** edges of a unit cell. The reciprocal lattice of the real lattice can be constructed by drawing the point at which the perpendicular line to the lattice plane *hkl* is $1/d_{hkl}$ from the origin, where d_{hkl} is the spacing between the lattice planes and the origin is the same point for the real space lattice and the reciprocal lattice and is an arbitrary real lattice point. The reciprocal lattice point drawn in this way is also given the indices *hkl*. By repeating the process over all of the lattice planes, the lattice points over all of the reciprocal lattice can be constructed. The distances between the lattice points of the real lattice are inversely proportional to the distances between the reciprocal lattice points, and for real unit cells where $\alpha = \beta = \gamma = 90°$ (as is the case in this thesis), the reciprocal unit cell has dimensions parallel to the **a, b, c** axes of the unit cell with length 1/a, 1/b, 1/c.

2.2.5 The requirement for formation of diffraction spots

X-rays incident on each of set of lattice planes at an angle θ will be reflected at angle θ. The spots or 'reflections' of these X rays (as detected by the detector) are only produced if there is constructive interference of the diffracted X-rays. The spacing of the spots within the observed diffraction pattern is characteristic of the unit cell because constructive interference only occurs when a point *hkl* in the reciprocal lattice intersects with the surface of a theoretical sphere of reflection (the 'Ewald' sphere) because at that point Bragg's law of constructive interference holds true ($2d\sin\theta = n\lambda$, meaning the separation between the parallel reflecting lattice planes, d_{hkl}, multiplied by twice the sine of the angle of incidence θ of the X-rays on the lattice planes is equal to an integer number of the X-ray wavelength λ). As the distances

between the points of the reciprocal lattice are inversely proportional to the distances between the lattice points of the real lattice, (which are the dimensions of the unit cell for a primitive lattice), the spacing of the reflections thus describe the unit cell of the crystal. The reflected X-rays will pass through the *hkl* point in reciprocal space lying on the Ewald sphere, so the *hkl* point indicates at what angle the ray will be reflected, and so where the reflection spot can be expected on the detector.

The unit cell dimensions together with systematic absences of reflections in the diffraction pattern can be used to determine the space group of the real lattice unit cell. This in turn relates to the symmetry of the reciprocal lattice (specifically, the 'Laue' group of the reciprocal lattice) for which there is a specific fraction of the reflections that are unique. Thus, an appropriate data collection strategy can be designed from the initial diffraction images and characterisation of the unit cell, such that the number of measurable reflections can be optimised and there are multiple observations of each unique reflection to improve accuracy (this is expressed in the 'redundancy' of the data).

2.2.6 Indexing, integration, merging and scaling

Once all the diffraction images have been recorded, the individual reflections must be indexed (with the Miller indices *hkl*). This is usually performed by autoprocessing software, which first of all determines the cell parameters accurately and the angle of orientation of the crystal in a particular frame (and from this the angle of orientation of all frames, because the oscillation range is known). It will then calculate a prediction of the diffraction image for the crystal in a particular orientation, and compare it to the experimental image[109]. The fit between the two is

optimised (with curation from the user) and then the program will assign the *hkl* indices from the prediction to the experimental reflections. The program will then integrate the intensity for each reflection (iMosflm[110] and XDS[111] were used for indexing and integration here). The images are then scaled across the dataset and multiple observations of the same reflection are merged together to give an average intensity (for example with XSCALE[111] or Aimless[112]).

2.2.7 The relationship between the structure factor and electron density

The reflection *hkl* is produced by a constructively interfering set of X-rays that have been diffracted from the same set of lattice planes within the crystal. The reflection can be described by a 'structure factor' $\mathbf{F}_{hkl}$ which is the sum of all the contributions of diffraction from each individual atom within the unit cell to the reflection *hkl*.

For example, we can define the atomic structure factor $\mathbf{f}_{j,hkl}$ of atom j in the unit cell, which describes the contribution of diffraction from atom j to the reflection *hkl*:

$$\boldsymbol{f}_{j,hkl} = f_j \, e^{2\pi i(hx_j + ky_j + lz_j)} \tag{8}$$

Where h, k and l are the indices of the reflection, x_j, y_j and z_j are the coordinates of atom j in the real unit cell, and f_j is the scattering factor of j due to the number of electrons it possesses (*i.e.* it depends on the elemental identity of atom j).

$\mathbf{F}_{hkl}$ is then the sum of the atomic structure factors for all n atoms in the unit cell,

$$\mathbf{F}_{hkl} = \sum_{j=1}^{n} f_j \, e^{2\pi i(hx_j + ky_j + lz_j)} \tag{9}$$

So $\mathbf{F}_{hkl}$ is a Fourier sum of n terms which describe the reflection *hkl*.

This equation implies that every atom in the unit cell contributes to $\mathbf{F}_{hkl}$ and the reflection *hkl*, which is true for all reflections (*i.e.* all values of *h*, *k* and *l*).

The structure factor can also be defined in terms of each infinitesimal volume within the unit cell at point (*x,y,z*), with average electron density $\rho(x,\ y,\ z)$:

$$\mathbf{F}_{hkl} = \iiint_{x\,y\,z} \rho(x,y,z)\, e^{2\pi i(hx+ky+lz)}\, dx\, dy\, dz \tag{10}$$

So $\mathbf{F}_{hkl}$ is also the sum of contributions to the scattering from the electron density in each infinitesimal volume *dxdydz* (*dV*). In this form it can be seen that $\mathbf{F}_{hkl}$ is the Fourier transform of $\rho(x,\ y,\ z)$, which means that $\rho(x,\ y,\ z)$ is also the Fourier transform of $\mathbf{F}_{hkl}$:

$$\rho(x,y,z) = \frac{1}{V}\sum_{h}\sum_{k}\sum_{l}\mathbf{F}_{hkl}e^{-2\pi i(hx+ky+lz)} \tag{11}$$

where *V* is the volume of the unit cell, and a triple sum is used because there are discrete values of *h*, *k* and *l*. Thus the unknown electron density of the crystal $\rho(x,\ y,\ z)$ can be determined from the structure factors of all the reflections.

ρ(x, y, z) is periodic due to the periodic fluctuation in electron density of repeating unit cells throughout the crystal on the real space x, y, z axes. $\mathbf{F}_{hkl}$ is therefore also periodic (Eqn. (10)) and so has amplitude, frequency and phase.

As $\mathbf{F}_{hkl}$ is a complex number (Eqn.(9)), it can be depicted as a complex vector composed of its real parts described by vector **A** and its imaginary parts described by vector **B** (Figure 1 – 10). The amplitude of $\mathbf{F}_{hkl}$ is given by its length $|\mathbf{F}_{hkl}|$ where

$$|\mathbf{F}_{hkl}|^2 \propto \mathrm{I}_{hkl} \tag{12}$$

And I_{hkl} is the intensity of spot *hkl*, which was extracted in the integration, merging and scaling steps. The frequency of $\mathbf{F}_{hkl}$ is given by the values of h, k and l, which represent the frequency of the three dimensional periodic wave function in the x, y and z directions, and are the h, k and l indices of the individual reflection that $\mathbf{F}_{hkl}$ describes, so are also known experimentally. However the phase of $\mathbf{F}_{hkl}$, α, is unknown and must be determined in order to calculate $\rho(x, y, z)$.

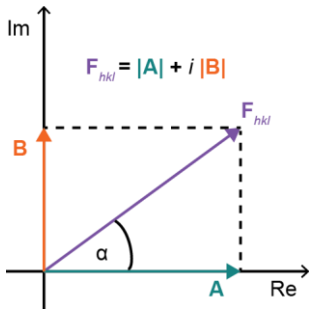


Figure 1 – 10 An Argand diagram depicting the vector $\mathbf{F}_{hkl}$ as a sum of its imaginary and real parts.

The amplitude of $\mathbf{F}_{hkl}$ can be expressed in terms of the amplitudes of **A** and **B** according to:

$$|\mathbf{A}| = |\mathbf{F}_{hkl}| \cos \alpha_{hkl}$$

$$|\mathbf{B}| = |\mathbf{F}_{hkl}| \sin \alpha_{hkl} \tag{13}$$

So,

$$\mathbf{F}_{hkl} = |\mathbf{A}| + i|\mathbf{B}| = |\mathbf{F}_{hkl}| (\cos \alpha_{hkl} + i \sin \alpha_{hkl}) = |\mathbf{F}_{hkl}| e^{i\alpha_{hkl}} \tag{14}$$

Where the last term results from the general equality $\cos\theta + \mathrm{i} \sin\theta = e^{i\theta}$ (Euler's formula).

Eqn. (14) can be used to express the phase of $\mathbf{F}_{hkl}$ explicitly in Eqn. (11):

$$\rho(x, y, z) = \frac{1}{V} \sum_h \sum_k \sum_l |\mathbf{F}_{hkl}| e^{-2\pi i(hx+ky+lz-\alpha'_{hkl})} \tag{15}$$

Where α'_{hkl} is the phase expressed in cycles rather than radians, *i.e.* $\alpha'_{hkl} = \alpha_{hkl}/2\pi$.

As α'_{hkl} (and α_{hkl}) are unknown, they must be estimated from additional data (see below). Estimates of α'_{hkl} can then be used with the experimental I_{hkl} ($|\mathbf{F}_{hkl}|$) values at each known *hkl* along with the known volume of the unit cell V to produce an approximation of $\rho(x, y, z)$. A model of identifiable features in $\rho(x, y, z)$ can then be used to estimate $\mathbf{F}_{hkl}$ (Eqn.(10)) which will contain improved phase estimates. These can in turn be used to calculate an improved approximation of $\rho(x, y, z)$ which facilitates construction of an improved model and so on. This iterative process will be described in greater detail below, but first I will discuss how the initial phase estimates can be prepared.

2.2.8 Molecular replacement

Estimates of the phases of the structure factors can be calculated in several ways, and the method of choice depends on several factors including the availability of similar structures, successful production of heavy atom 'derivative' crystals that are isomorphic (*i.e.* that have the same space group and crystal packing), and tunable wavelength X-ray sources. In this project, we phased all of the structures using molecular replacement (with the Phaser software[113]), which I will discuss in more detail here.

When the structure of the protein under investigation (the 'target' structure) is expected to be similar to another protein for which the structure is known (the 'phasing model' structure), it is possible to use the structure factor phases of the phasing model (calculated from Eqn.(9)) as an estimate for the phases of the target protein. As the core domain structure of sHSPs is well-conserved this is often the case when solving new sHSP structures, and this approach has been employed in the past[34,48,51]. The phase estimates, α'_{hkl}, and the known amplitudes of the

target structure factors, $|\mathbf{F}_{obs}|$, can be used in Eqn.(15) to calculate the first electron density approximation.

If the phasing model and target are not isomorphous a unit cell of the model structure must first be constructed with the same symmetry as the target protein. However, its position and orientation within the unit cell will be different to those of the target, and the structure factors of each reflection depend on the position of each atom in the unit cell (Eqn.(9)). Therefore in order to produce sufficient phase estimates, the model must be moved to replicate the orientation and position of molecules within the unit cell of the target. The reorientation is achieved by comparing the 'Patterson maps' of the model and target. This requires the calculation of the Patterson function:

$$\mathrm{P}(u,v,w) = \frac{1}{V}\sum_{h}\sum_{k}\sum_{l}|\mathbf{F}_{hkl}|^2\, e^{-2\pi i(hu+kv+lw)} \tag{16}$$

As calculation of the Patterson function only requires knowledge of the *hkl* values of each reflection and their intensity, it can be calculated for the target as well as the phasing model. This allows the construction of the respective Patterson maps, which is a plot of P(*u,v,w*) over coordinates *u, v, w*. A Patterson map represents vectors between atoms in the real lattice unit cell. Thus a peak in the Patterson map at coordinates (*u,v,w*) represents a vector between atoms at (x_1,y_1,z_1) and (x_2,y_2,z_2) where $x_2 - x_1 = u$, $y_2 - y_1 = v$, $z_2 - z_1 = w$.

The (0,0,0) points of the model and target Patterson maps are then overlaid, and the model molecule systematically rotated, which also rotates its Patterson map. The overlap between the two maps is optimised by scoring each orientation according to a 'rotation' function (for

example, the product of $P(u,v,w)^{target}$ with $P(u,v,w)^{model}$ at each (u,v,w) point in the map will be maximal when the most peaks overlap).

Once the model has been rotated to the orientation of the target, it can be translated to superimpose with the target. This is achieved by comparing the structure factor amplitudes of the target (from the spot intensities, $|\mathbf{F}_{obs}|$) and the model in a particular trial position ($|\mathbf{F}_{calc}|$, calculated from Eqn.(9)). The comparison can be scored by an R-factor:

$$R = \frac{\sum ||\mathbf{F}_{obs}| - |\mathbf{F}_{calc}||}{\sum |\mathbf{F}_{obs}|} \tag{17}$$

Where the sum is for the $|\mathbf{F}_{obs}|$ and $|\mathbf{F}_{calc}|$ values over all (h, k, l).

The optimal position of the model will give the lowest R-factor. Once the phasing model has been reoriented and repositioned to superimpose on the target within the unit cell, its phases can be calculated and used as a first estimate of the phases of the target.

In the absence of a suitable search molecule for molecular replacement, the phase estimates must be determined by another approach such as isomorphous replacement or anomalous scattering, both of which require a 'derivative' of the crystal containing heavy atoms. The heavy atoms are introduced by co-crystallisation; soaking protein crystals in a solution of heavy atoms; expression of the protein in a micro-organism auxotrophic for methionine which is grown in the presence of selenomethionine such that all the methionines contain heavy selenium rather than sulphur; or in some cases the heavy atoms are present in the native protein (*e.g.* metalloproteins). In isomorphous replacement, the diffractive contribution of the heavy atoms to each reflection can be determined through comparison of the reflection

intensities for the native and derivative crystals, which allows the position of the heavy atoms in the unit cell to be calculated. The structure factors of the heavy atoms can thus be produced (Eqn.(9)) which in turn allows phase estimates of the protein structure factors to be calculated. In anomalous scattering, the structure factors due to the heavy atoms are similarly determined. Use of an additional X-ray wavelength that is absorbed by the heavy atoms and re-emitted with a different phase leads to non-equal intensities in symmetry-related reflections (*i.e.* anomalous scattering) and allows the structure factors of the derivative crystal to be calculated. Together these sets of structure factors can be used to calculate the phase of the structure factors for the native crystal. There are several isomorphous replacement and anomalous scattering approaches, with slight variations to the simplest schemes described here. Nonetheless, the final result of either approach or molecular replacement is a set of initial phase estimates for the target protein.

2.2.9 Phase improvement and phase extension

It is now possible to calculate an initial electron density map from Eqn.(15) using $|\mathbf{F}_{hkl}|$ (*i.e.* $|\mathbf{F}_{\text{obs}}|$ which is equal to $\sqrt{\text{I}_{hkl}}$ from the experimental reflection data) and α'_{hkl} (*i.e.* α_{calc}, the phase estimates). Phase estimates may not be available for every reflection – that is, every value of (*h, k, l*) - due to the resolution of the phasing data, so only those terms for which there is an α'_{hkl} estimate are included in the Fourier sum of Eqn.(15). Additionally, the terms are usually weighted according to the reliability of the particular phase estimate (a 'figure of merit').

This initial electron density map will be crude, but should allow discrimination of the protein molecules from bulk water, and hopefully also determination of the boundary of each molecule

(the 'molecular envelope') resembling protein structure. If this is not observable, further data collection is required[107].

From this point on, the primary aim is to improve the phase estimates by identifying features in the map. This will improve the electron density map and will ultimately result in the production of a complete molecular model with well-determined phases and few errors. The whole process is guided by appropriate prior knowledge of protein structure and the particular crystal, and it is the combination of this additional information with the reflection data that allows the determination of improved phase estimates.

In order to improve the phases before any molecular model is introduced, density modification can be used along with symmetry averaging if non-crystallographic symmetry is present. Density modification applies knowledge about protein crystals to reduce noise and sharpen density corresponding to protein. For example, knowledge that the density in bulk solvent regions is nearly constant or 'flat', and the histograms of electron density in protein regions are similar and distinct from that of noise. Non-crystallographic symmetry (NCS) may exist between molecules within the asymmetric unit such that the molecules are identical but will never have the same orientation to the X-ray beam during data collection[107]. Patterson methods can again be used to identify the presence and identity of such symmetry operators, and the signal from the symmetry related molecules can be averaged to improve the signal-to-noise ratio in symmetry averaging.

These methods lead to an improved electron density. Gradually phase estimates for reflections at higher resolution are computed from Eqn.(10) and are included with the corresponding

experimental $|\mathbf{F}_{hkl}|$ as terms in the Fourier sum of Eqn.(15). Though estimates at all the (*h, k, l*) values can be computed from the nascent electron density map, this initial model would give poor α'_{hkl} estimates and so would not guide the process appropriately. Instead by adding the weaker, higher resolution terms to the Fourier sum gradually, the experimental data can be used to steer the modelling of the electron density. At some point, the map will become detailed enough that a molecular structure can begin to be constructed within it, and the structure factors (and phase estimates) will be determined from Eqn.(9). Again if the partial model is correct, the phase estimates will improve and the electron density map will become more detailed, allowing more of the protein to be modelled into the density.

2.2.10 Model building and refinement

Initially model building (or 'map fitting') may involve skeletonisation (drawing lines through regions of maximum density in the hope of detecting some general structural features such as α-helices) and chain tracing (fitting fragments or the whole of the polypeptide chain into the density by deciding feasible positions for the main chain atoms). As the map becomes more detailed, it hopefully becomes possible to identify small stretches of the sequence by identifying density characteristic of particular side chains (*e.g.* bulky tryptophans or electron rich methionines). Short stretches of the sequence may be imported from fragment databases based on previous structures and are used as initial models of the local conformations. In this way, model building is improved with prior knowledge of the protein sequence and the abundance of previous experimental observations. In turn, this allows more terms to be added to the Fourier sum from which the electron density is computed, again improving the detail of the map.

Model building in real space is interspersed with rounds of automated refinement in reciprocal space, in which the structural model is varied while the structure factor amplitudes of the model $|\mathbf{F}_{calc}|$ are compared to the experimental $|\mathbf{F}_{obs}|$ values through a target function. The nature of the target function depends on the particular software used, but will generally minimise the difference between $|\mathbf{F}_{calc}|$ and $|\mathbf{F}_{obs}|$, along with additional 'restraints' such as bond length. The software used in this project (Refmac 5[114] and phenix.refine[115]) both employ likelihood target functions, in which the likelihood of the model given the reflection data is maximised (or the minus log likelihood is minimised).

The restraints used in reciprocal space refinement are based on prior knowledge and are included as additional terms in the target function to be optimised. They frequently include optimising the similarity of molecules related by NCS and comparing the bond lengths, bond angles and Van der Waals contact distances of the model to preferred values known from small molecule crystallography and high resolution protein structures. The restraints are gradually lifted as refinement progresses to ensure that refinement is driven by the reflection data, rather than the model being optimised according to geometry.

Real space model building and reciprocal space refinement will be sequentially looped. Subsequent stages of model building will involve building more of the model and modifying it to better fit within the electron density using knowledge of secondary structure characteristics (*e.g.* the carbonyl bonds point in alternating directions within a β-sheet, with the sidechains being perpendicular to the sheet and alternating in direction) and preferred conformations ('rotamers') of the side chains. As incorrect atomic coordinates will degrade the electron

density map through poorer phase estimates, the process is driven by the density. On the other hand, the electron density becomes more accurate as the model across the unit cell improves, so previously unassignable features can be included or corrected due to the influence of their neighbours.

The iterative process of improving the model is guided by numerous validation tools, the most noteworthy being the R-factor (or 'R_{work}') which is calculated as in Eqn. (17) but with $|\mathbf{F}_{calc}|$ structure factor amplitudes calculated from the most recent refined model. R_{work} should decrease as the model better agrees with the data. However, if a sufficiently large number of parameters are used R_{work} can be made arbitrarily small[116], so an additional parameter R_{free} is also used. This is calculated with Eqn. (17) but on a test set of randomly chosen reflections that are removed before refinement. As the model has never been refined against these reflections, R_{free} gives a fairer measure of how closely the model predicts the experimental data. Both R_{work} and R_{free} should fall during refinement, and the gap between them should also decrease as the model better describes the data. The final cycles of refinement should show that neither values are still declining (*i.e.* both values have converged).

The manual map fitting described above can be performed automatically in some software packages, such that the iterative model building and refinement steps are all executed without user input. For example, in the Phenix software suite the phenix.autobuild procedure undertakes model-building, density modification and refinement[115]. The output of this automated model building can be very successful, such that in the best cases the user is only required to curate a few remaining errors.

The electron density map is increasingly constructed using the phases calculated from the current model (*i.e.* there are more terms with phases calculated from the model in the Fourier sum of Eqn.(15)). It is therefore important to mitigate the effect of model bias in the refinement process, in order that the experimental structure factors guide the model building process (via the reflection intensities). This can be achieved by calculating electron density maps based on alternative Fourier syntheses in which $|\mathbf{F}_{calc}|$ is subtracted from a multiple of $|\mathbf{F}_{obs}|$ in the amplitude of each Fourier term. For example, in this project the $m\mathrm{F}_{\mathrm{obs}}$ - $D\mathrm{F}_{\mathrm{calc}}$ map is used which is constructed from

$$\rho(x,y,z) = \frac{1}{V}\sum_{h}\sum_{k}\sum_{l}(m|\mathbf{F}_{obs}| - D|\mathbf{F}_{calc}|)e^{-2\pi i(hx+ky+lz-\alpha'_{calc})} \qquad (18)$$

where m is the figure of merit for each model phase and D is an estimate of atomic coordinate errors in the current model[107]. This difference map removes the amplitude of the model structure factor from each experimental structure factor, for each value of (h,k,l). This leads to positive 'density' in regions where the unit cell contains greater electron density than the model suggests (as $|\mathbf{F}_{obs}| > |\mathbf{F}_{calc}|$), and negative density where the model predicts greater electron density than is actually supported by the data (as $|\mathbf{F}_{obs}| < |\mathbf{F}_{calc}|$). Thus part of a model surrounded by a region of negative density should be moved, possibly to a region of adjacent positive density.

Additionally, we employed a $2m\mathrm{F}_{\mathrm{obs}}$ - $D\mathrm{F}_{\mathrm{calc}}$ difference map, which is easier to interpret than an $m\mathrm{F}_{\mathrm{obs}}$ - $D\mathrm{F}_{\mathrm{calc}}$ map yet still accounts for model bias. It is calculated with

$$\rho(x,y,z) = \frac{1}{V}\sum_{h}\sum_{k}\sum_{l}(2m|\mathbf{F}_{obs}| - D|\mathbf{F}_{calc}|)e^{-2\pi i(hx+ky+lz-\alpha'_{calc})} \qquad (19)$$

Apart from drastic errors in the model, this map should be positive everywhere and resembles the electron density map calculated from Eqn.(15) similar to a molecular surface, facilitating model building into the density. The contour maps constructed from Eqns.(18) and (19) are respectively referred to as an 'F_o-F_c' and '$2F_o$-F_c' map herein, but in fact each of the maps is the figure-of-merit weighted version presented above (closely related to 'σA-weighted' maps).

Atomic coordinates are not the only parameters refined during the process. The occupancy of each atom j, which is the fraction of molecules within the crystal in which atom j occupies the position described in the model[107], can also be refined. The occupancy is sometimes constrained to a value of 1 in early rounds of refinement, and then refined in the final stages once the model and concomitant phases have been improved, such that they are better able to indicate unassigned density. Occupancy refinement may lead to a more accurate model and importantly reports on multiple conformations of atoms within the crystal (for example, a side chain may occupy multiple rotamers). Another parameter that is frequently refined is the atomic displacement parameter (ADP, 'B factor' or 'temperature factor'). The ADPs report on the spread of the atomic positions of atom j around the position described in the model, *i.e.* how much an atom oscillates about its position and varies within each unit cell of the crystal. An atom with a high B-factor has a greater freedom of movement within the molecule. Atomic displacements can be described by an isotropic B-factor though they are likely anisotropic in reality, which can be described by six parameters for each atom. Anisotropic B-factor refinement requires a concurrent increase in the number of observations to meaningfully determine the parameters to describe all of

the atoms. This is only possible for high resolution structures[105]. However for medium and low resolution data, TLS refinement can be used to model anisotropic ADPs using only a few parameters[114]. In this, the protein chain is treated as a number of rigid bodies composed of atoms that all move in the same way. The anisotropic displacement parameters are defined for each rigid body (rather than for each individual atom separately) by describing its motion with translation (T), libration (L) and screw (S) tensors. Each rigid body is described by 20 parameters in total and from these anisotropic displacement parameters can be defined for each constituent atom[114]. In this way, anisotropic B-factors can be defined for lower resolution structures with fewer reflection observations. The rigid body segments can be assigned with default parameters when integrated in refinement software (*e.g.* Refmac assigns one group per chain[114]), or more precisely by grouping based on estimates of the B-factor (as performed by the TLSMD web server[117]).

Care must be taken to appropriately use the numerous utilities offered by refinement software as well as avoid 'overfitting' the data which produces a misleading level of agreement between the model and experimental data because of the low ratio between the number of observations (*i.e.* reflections) and the number of parameters describing the molecular model. This is demonstrated by a large difference between R_{work} and R_{free} values[114]. The low observation: parameter ratio for protein crystallography is ameliorated using the prior knowledge about the protein as described above.

2.2.11 Validation

Once the refinement procedure is complete, the validity of the molecular model must be assessed, *i.e.* how well the model agrees with the original data. In fact, the model building itself is guided by many of the same tools indicating the appropriateness of the model. For example, two important indicators of the model's validity are the R_{work} and R_{free} described above. The final values allow the users of a crystallographic model to judge the quality of the model themselves. The R-factors expected for accurate models will depend on the resolution of the data, though some guidelines are suggested[105]. Validation tools, for example those within Coot[118] and Molprobity[119] used herein, may also highlight peptide bonds that deviate significantly from planar, report the average root mean squared deviations from ideal bond lengths and bond angles, and compare the side chain dihedral angles to preferred angles within a structural database. Additionally, parameters that have not themselves been refined are useful unbiased indicators of the quality of the model. For example, the torsion angles of the polypeptide backbone φ and ψ are usually evaluated within the context of a Ramachandran plot. Any angles that lie outside the allowed regions of the plot would need to have some justification (*e.g.* local structural constraints that outweigh the energy cost of steric repulsion within the backbone).

Therefore structural abnormalities must either be justified or they must be corrected and further refinement undertaken, and concurrent measures employed to remove some of the bias introduced by this incorrect part of the model in the prior refinement. One such approach is simulated annealing, a molecular dynamics approach in which the model is 'heated' such that it moves and can escape local energy minima within the energy landscape. The model is then

cooled gradually to the temperature of data collection in the hope of finding its preferred conformation[107]. This strategy is regularly used in the refinement process, and not just invoked to remove bias from particular errors.

When modelling a structural feature of particular note or non-protein atoms in the model such as those in ligands, water and solutes, it is good practice to only add them to the model prior to the last round of refinement. This prevents putatively modelled ligands biasing the electron density through the calculated phases and so falsely appearing to be present in the crystal (*i.e.* model bias[120]). However, as a more complete model can aid the refinement process they are sometimes added in earlier refinement rounds, then removed and the model subjected to simulated annealing to attempt to remove the bias. The resulting 'omit' F_o-F_c map should then have positive density in the region if the feature is truly supported by the experimental structure factors.

The model must be critically evaluated in light of the way it has been determined, and the output of the validation tools. When it has been ascertained that the model is a sufficiently faithful representation of the true molecular structure, it is important to investigate whether the conformation is a result of the packing of the molecule within the crystal.

2.2.12 Biological relevance

Once validation is complete, the functional implications of the structure can finally be explored. In this thesis, I refer to the root mean square deviation (RMSD) in atomic positions between atoms in two different structures. This is calculated on aligned structures with

$$RMSD = \sqrt{\frac{\sum_{i=1}^{N} d_{ii}^2}{N}} \quad (20)$$

where d_{ii} is the distance between atom i in Structure A and atom i in Structure B, and N is the number of pairs of identical atoms (one in each structure). Herein the RMSD is usually calculated on Cα atoms within the backbone on residues resolved in both structures, and the two structures are aligned in sequence and structure but without any rejection of outliers in the alignment (with the PyMol 'align' tool).

Protein crystallography has long been a foundation of structural biology. Though there can be numerous obstacles to overcome before a valid structural model is obtained, the level of detail afforded by this technique warrants the time and financial cost of the experiment. It provides an atomic description of a molecule which can be used in concert with other structural and dynamical techniques to provide a rationale for the functional properties of a protein. Recent breakthroughs in single particle cryoEM[121] will allow it to contest with crystallography for some of the most coveted targets in structural biology into the future. However, this does not negate the utility of protein crystallography for a wide range of protein targets (for example soluble and membrane proteins, protein:nucleotide complexes), over a very wide mass range (anything from small molecules to large molecular machines such as the ribsosome[122] and RNA polymerase[123]). Additionally, familiarity in the wider scientific audience, the substantial expertise of crystallographers across the globe, and the robustness of crystallographic models following decades of discussion and software development in the community all stand in the favour of this established technique in the decades to come. Ultimately, protein

crystallography is invaluable because the insight it permits will provide fundamental contributions to our understanding of how proteins work.

3 Conclusion

In this chapter, I have provided an overview of the biological motivation behind this project, and two of the central techniques employed in the study of HSP27 herein. Native MS and protein crystallography have each provided a wealth of insight on the structure and behaviour of the sHSPs, which have been maximised by applying a combination of experimental techniques. This work similarly benefitted from the use of a range of approaches, with a structural and functional focus. Before describing these investigations, I will first provide a general review of the contribution of native MS to the study of the dynamic and heterogeneous systems of the proteostasis network, of which the sHSPs are notable examples. I hope that this will add further detail on the nature of these proteins, as well as demonstrating the utility of this particular technique.

4 References

1. Haslbeck, M., Franzmann, T., Weinfurtner, D. & Buchner, J. Some like it hot: the structure and function of small heat-shock proteins. *Nat. Struct. Mol. Biol.* **12,** 842–846 (2005).

2. Kim, K. K., Kim, R. & Kim, S. H. Crystal structure of a small heat-shock protein. *Nature* **394,** 595–9 (1998).

3. van Montfort, R. L. M., Basha, E., Friedrich, K. L., Slingsby, C. & Vierling, E. Crystal structure and assembly of a eukaryotic small heat shock protein. *Nat. Struct. Biol.* **8,** 1025–30 (2001).

4. Aquilina, J. A., Benesch, J. L. P., Bateman, O. a, Slingsby, C. & Robinson, C. V. Polydispersity of a mammalian chaperone: mass spectrometry reveals the population of oligomers in alphaB-crystallin. *Proc. Natl. Acad. Sci. U. S. A.* **100,** 10611–6 (2003).

5. Jovcevski, B. *et al.* Phosphomimics Destabilize Hsp27 Oligomeric Assemblies and Enhance Chaperone Activity. *Chem. Biol.* **22,** 186–195 (2015).

6. Benesch, J. L. P. *et al.* The Quaternary Organization and Dynamics of the Molecular Chaperone HSP26 Are Thermally Regulated. *Chem. Biol.* **17,** 1008–1017 (2010).

7. Bova, M. P., Mchaourab, H. S., Han, Y. & Fung, B. K.-K. Subunit Exchange of Small Heat Shock Proteins: Analysis of oligomer formation of αA-crystallin and HSP27 by fluorescence resonance energy transfer and site-directed truncations. *J. Biol. Chem.*

275, 1035–1042 (2000).

8. Sobott, F., Benesch, J. L. P., Vierling, E. & Robinson, C. V. Subunit exchange of multimeric protein complexes: real-time monitoring of subunit exchange between small heat shock proteins by using electrospray mass spectrometry. *J. Biol. Chem.* **277,** 38921–38929 (2002).

9. Baldwin, A. J., Lioe, H., Robinson, C. V., Kay, L. E. & Benesch, J. L. P. αB-Crystallin Polydispersity Is a Consequence of Unbiased Quaternary Dynamics. *J. Mol. Biol.* **413,** 297–309 (2011).

10. Stengel, F. *et al.* Quaternary dynamics and plasticity underlie small heat shock protein chaperone function. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 2007–12 (2010).

11. Shepherd, D. A., Marty, M. T., Giles, K., Baldwin, A. J. & Benesch, J. L. P. Combining tandem mass spectrometry with ion mobility separation to determine the architecture of polydisperse proteins. *Int. J. Mass Spectrom.* **377,** 663–671 (2015).

12. Fleckenstein, T. *et al.* The Chaperone Activity of the Developmental Small Heat Shock Protein Sip1 Is Regulated by pH-Dependent Conformational Changes. *Mol. Cell* 1–12 (2015). doi:10.1016/j.molcel.2015.04.019

13. Haslbeck, M. & Vierling, E. A First Line of Stress Defense: Small Heat Shock Proteins and Their Function in Protein Homeostasis. *J. Mol. Biol.* **427,** 1537–1548 (2015).

14. Jakob, U., Gaestel, M., Engel, K. & Buchner, J. Small heat shock proteins are molecular chaperones. *J. Biol. Chem.* **268,** 1517–20 (1993).

15. Mchaourab, H. S., Dodson, E. K. & Koteiche, H. A. Mechanism of Chaperone Function in Small Heat Shock Proteins: TWO-MODE BINDING OF THE EXCITED STATES OF T4 LYSOZYME MUTANTS BY A-CRYSTALLIN. *J. Biol. Chem.* **277,** 40557–40566 (2002).

16. Mchaourab, H. S., Godar, J. A. & Stewart, P. L. Structure and Mechanism of Protein Stability Sensors: Chaperone Activity of Small Heat Shock Proteins. *Biochemistry* **48,** 3828–3837 (2009).

17. Cheng, G., Basha, E., Wysocki, V. H. & Vierling, E. Insights into Small Heat Shock Protein and Substrate Structure during Chaperone Action Derived from Hydrogen/Deuterium Exchange and Mass Spectrometry. *J. Biol. Chem.* **283,** 26634–26642 (2008).

18. Kourtis, N., Nikoletopoulou, V. & Tavernarakis, N. Small heat-shock proteins protect from heat-stroke-associated neurodegeneration. *Nature* **490,** 213–218 (2012).

19. Rogalla, T. *et al.* Regulation of Hsp27 Oligomerization, Chaperone Function, and Protective Activity against Oxidative Stress/Tumor Necrosis Factor by Phosphorylation. *J. Biol. Chem.* **274,** 18947–18956 (1999).

20. Rajagopal, P., Liu, Y., Shi, L., Clouser, A. F. & Klevit, R. E. Structure of the α-crystallin domain from the redox-sensitive chaperone, HSPB1. *J. Biomol. NMR* 0–5 (2015). doi:10.1007/s10858-015-9973-0

21. Hayes, D., Napoli, V., Mazurkie, A., Stafford, W. F. & Graceffa, P. Phosphorylation

Dependence of Hsp27 Multimeric Size and Molecular Chaperone Function. *J. Biol. Chem.* **284,** 18801–18807 (2009).

22. Bryantsev, A. L., Chechenova, M. B. & Shelden, E. A. Recruitment of phosphorylated small heat shock protein Hsp27 to nuclear speckles without stress. *Exp. Cell Res.* **313,** 195–209 (2007).

23. Aquilina, J. A., Shrestha, S., Morris, A. M. & Ecroyd, H. Structural and functional aspects of hetero-oligomers formed by the small heat shock proteins αB-crystallin and HSP27. *J. Biol. Chem.* **288,** 13602–9 (2013).

24. Stengel, F. *et al.* Dissecting heterogeneous molecular chaperone complexes using a mass spectrum deconvolution approach. *Chem. Biol.* **19,** 599–607 (2012).

25. Arrigo, A.-P. & Gibert, B. Protein interactomes of three stress inducible small heat shock proteins: HspB1, HspB5 and HspB8. *Int. J. Hyperth.* **29,** 409–22 (2013).

26. Arrigo, A.-P. in *Mol. Chaperones Methods Protoc.* (eds. Calderwood, S. K. & Prince, T. L.) **787,** 105–119 (Humana Press, 2011).

27. Cox, D., Carver, J. A. & Ecroyd, H. Preventing α-synuclein aggregation: The role of the small heat-shock molecular chaperone proteins. *Biochim. Biophys. Acta - Mol. Basis Dis.* **1842,** 1830–1843 (2014).

28. Evgrafov, O. V *et al.* Mutant small heat-shock protein 27 causes axonal Charcot-Marie-Tooth disease and distal hereditary motor neuropathy. *Nat. Genet.* **36,** 602–606 (2004).

29. Tanguay, R. M. & Editors, L. E. H. *The Big Book on Small Heat Shock Proteins*.

30. Hanazono, Y. *et al.* Nonequivalence Observed for the 16-Meric Structure of a Small Heat Shock Protein, SpHsp16.0, from Schizosaccharomyces pombe. *Structure* **21,** 220–228 (2013).

31. Hanazono, Y., Takeda, K., Yohda, M. & Miki, K. Structural studies on the oligomeric transition of a small heat shock protein, StHsp14.0. *J. Mol. Biol.* **422,** 100–8 (2012).

32. Hilario, E., Martin, F. J. M., Bertolini, M. C. & Fan, L. Crystal Structures of Xanthomonas Small Heat Shock Protein Provide a Structural Basis for an Active Molecular Chaperone Oligomer. *J. Mol. Biol.* **408,** 74–86 (2011).

33. Hochberg, G. K. a *et al.* The structured core domain of αB-crystallin can prevent amyloid fibrillation and associated toxicity. *Proc. Natl. Acad. Sci. U. S. A.* **111,** E1562–70 (2014).

34. Laganowsky, A. *et al.* Crystal structures of truncated alphaA and alphaB crystallins reveal structural mechanisms of polydispersity important for eye lens function. *Protein Sci.* **19,** 1031–43 (2010).

35. Jehle, S. *et al.* Solid-state NMR and SAXS studies provide a structural basis for the activation of αB-crystallin oligomers. *Nat. Struct. Mol. Biol.* **17,** 1037–1042 (2010).

36. Delbecq, S. P., Jehle, S. & Klevit, R. Binding determinants of the small heat shock protein, αB-crystallin: recognition of the 'IxI' motif. *EMBO J.* **31,** 4587–4594 (2012).

37. Hochberg, G. K. A. & Benesch, J. L. P. Dynamical structure of αB-crystallin. *Prog. Biophys. Mol. Biol.* **115,** 11–20 (2014).

38. Takeda, K. *et al.* Dimer structure and conformational variability in the N-terminal region of an archaeal small heat shock protein, StHsp14.0. *J. Struct. Biol.* **174,** 92–99 (2011).

39. Braun, N. *et al.* Multiple molecular architectures of the eye lens chaperone αB-crystallin elucidated by a triple hybrid approach. *Proc. Natl. Acad. Sci.* **108,** 20491–20496 (2011).

40. Baldwin, A. J. *et al.* The polydispersity of alphaB-crystallin is rationalized by an interconverting polyhedral architecture. *Structure* **19,** 1855–1863 (2011).

41. Shi, J., Koteiche, H. a, McHaourab, H. S. & Stewart, P. L. Cryoelectron microscopy and EPR analysis of engineered symmetric and polydisperse Hsp16.5 assemblies reveals determinants of polydispersity and substrate binding. *J. Biol. Chem.* **281,** 40420–8 (2006).

42. Wintrode, P. L., Friedrich, K. L., Vierling, E., Smith, J. B. & Smith, D. L. Solution Structure and Dynamics of a Heat Shock Protein Assembly Probed by Hydrogen Exchange and Mass Spectrometry †. *Biochemistry* **42,** 10667–10673 (2003).

43. Mainz, A. *et al.* The chaperone αB-crystallin uses different interfaces to capture an amorphous and an amyloid client. *Nat. Struct. Mol. Biol.* (2015). doi:10.1038/nsmb.3108

44. Hilton, G. R., Lioe, H., Stengel, F., Baldwin, A. J. & Benesch, J. L. P. Small Heat-Shock Proteins : Paramedics of the Cell. *Top. Curr. Chem.* **328,** 69–98 (2013).

45. Mainz, A. *et al.* The chaperone [alpha]B-crystallin uses different interfaces to capture an amorphous and an amyloid client. *Nat Struct Mol Biol* **advance on,** (2015).

46. Liu, L., Chen, J., Yang, B. & Wang, Y. Crystal structure and function of an unusual dimeric Hsp20.1 provide insight into the thermal protection mechanism of small heat shock proteins. *Biochem. Biophys. Res. Commun.* **458,** 429–434 (2015).

47. Stamler, R., Kappé, G., Boelens, W. & Slingsby, C. Wrapping the α-Crystallin Domain Fold in a Chaperone Assembly. *J. Mol. Biol.* **353,** 68–79 (2005).

48. Bagnéris, C. *et al.* Crystal Structures of α-Crystallin Domain Dimers of αB-Crystallin and Hsp20. *J. Mol. Biol.* **392,** 1242–1252 (2009).

49. Clark, A. R., Naylor, C. E., Bagnéris, C., Keep, N. H. & Slingsby, C. Crystal structure of R120G disease mutant of human αb-crystallin domain dimer shows closure of a groove. *J. Mol. Biol.* **408,** 118–134 (2011).

50. Jehle, S. *et al.* αB-Crystallin: A Hybrid Solid-State/Solution-State NMR Investigation Reveals Structural Aspects of the Heterogeneous Oligomer. *J. Mol. Biol.* **385,** 1481–1497 (2009).

51. Hochberg, G. K. A. *et al.* The structured core domain of αB-crystallin can prevent amyloid fibrillation and associated toxicity. *Proc. Natl. Acad. Sci. U. S. A.* **111,** E1562–70 (2014).

52. Laganowsky, A. & Eisenberg, D. Non-3D domain swapped crystal structure of truncated zebrafish alphaA crystallin. *Protein Sci.* **19,** 1978–1984 (2010).

53. Baranova, E. V *et al.* Three-dimensional structure of α-crystallin domain dimers of human small heat shock proteins HSPB1 and HSPB6. *J. Mol. Biol.* **411,** 110–22 (2011).

54. Weeks, S. D. *et al.* Molecular structure and dynamics of the dimeric human small heat shock protein HSPB6. *J. Struct. Biol.* **185,** 342–354 (2014).

55. Baldwin, A. J. *et al.* The Polydispersity of αB-Crystallin Is Rationalized by an Interconverting Polyhedral Architecture. *Structure* **19,** 1855–1863 (2011).

56. Giese, K. C., Basha, E., Catague, B. Y. & Vierling, E. Evidence for an essential function of the N terminus of a small heat shock protein in vivo, independent of in vitro chaperone activity. *Proc. Natl. Acad. Sci.* **102,** 18896–18901 (2005).

57. Ylikallio, E. *et al.* Truncated HSPB1 causes axonal neuropathy and impairs tolerance to unfolded protein stress. *BBA Clin.* **3,** 233–242 (2015).

58. Jaya, N., Garcia, V. & Vierling, E. Substrate binding site flexibility of the small heat shock protein molecular chaperones. *Proc. Natl. Acad. Sci.* **106,** 15604–15609 (2009).

59. Kampinga, H. H. *et al.* Guidelines for the nomenclature of the human heat shock proteins. *Cell Stress Chaperones* **14,** 105–11 (2009).

60. Sievers, F. *et al.* Fast, scalable generation of high-quality protein multiple sequence

alignments using Clustal Omega. *Mol. Syst. Biol.* **7,** 1–6 (2011).

61. Goujon, M. *et al.* A new bioinformatics analysis tools framework at EMBL-EBI. *Nucleic Acids Res.* **38,** W695–9 (2010).

62. Yamashita, M. & Fenn, J. B. Electrospray ion source. Another variation on the free-jet theme. *J. Phys. Chem.* **88,** 4451–4459 (1984).

63. Fernández de la Mora, J. The Fluid Dynamics of Taylor Cones. *Annu. Rev. Fluid Mech.* **39,** 217–243 (2007).

64. Benesch, J. L. P., Ruotolo, B. T., Simmons, D. A. & Robinson, C. V. Protein complexes in the gas phase: technology for structural genomics and proteomics. *Chem. Rev.* **107,** 3544–3567 (2007).

65. Kebarle, P. & Verkerk, U. H. Electrospray: from ions in solution to ions in the gas phase, what we know now. *Mass Spectrom. Rev.* **28,** 898–917 (2009).

66. Mehmood, S., Allison, T. M. & Robinson, C. V. Mass Spectrometry of Protein Complexes: From Origins to Applications. *Annu. Rev. Phys. Chem.* **66,** 453–474 (2015).

67. Kondrat, F. D. L., Struwe, W. B. & Benesch, J. L. P. in *Struct. Proteomics High-Throughput Methods* (ed. Owens, R. J.) **1261,** 349–371 (Springer New York, 2015).

68. Konijnenberg, A., Butterer, A. & Sobott, F. Native ion mobility-mass spectrometry and related methods in structural biology. *Biochim. Biophys. Acta - Proteins*

Proteomics **1834,** 1239–1256 (2013).

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

70. Hopper, J. T. S. & Oldham, N. J. Alkali Metal Cation-Induced Destabilization of Gas-Phase Protein–Ligand Complexes: Consequences and Prevention. *Anal. Chem.* **83,** 7472–7479 (2011).

71. Hopper, J. T. S., Sokratous, K. & Oldham, N. J. Charge state and adduct reduction in electrospray ionization-mass spectrometry using solvent vapor exposure. *Anal. Biochem.* **421,** 788–790 (2012).

72. Sobott, F., Hernández, H., McCammon, M. G., Tito, M. A. & Robinson, C. V. A Tandem Mass Spectrometer for Improved Transmission and Analysis of Large Macromolecular Assemblies. *Anal. Chem.* **74,** 1402–1407 (2002).

73. Hoffmann, E. De & Stroobant, V. *Mass Spectrometry Principles and Applications. Mass Spectrom. Princ. Appl.* (John Wiley & Sons Ltd., 2007). doi:10.1002/9783527654703.ch11

74. Sharon, M. & Robinson, C. V. The role of mass spectrometry in structure elucidation of dynamic protein complexes. *Annu. Rev. Biochem.* **76,** 167–193 (2007).

75. Lössl, P., Snijder, J. & Heck, A. J. R. Boundaries of Mass Resolution in Native Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **25,** 906–917 (2014).

76. Sharon, M. & Robinson, C. V. The Role of Mass Spectrometry in Structure Elucidation of Dynamic Protein Complexes. *Annu. Rev. Biochem.* **76,** 167–193 (2007).

77. Benesch, J. L. P. Collisional activation of protein complexes: Picking up the pieces. *J. Am. Soc. Mass Spectrom.* **20,** 341–348 (2009).

78. Aquilina, J. A., Benesch, J. L. P., Bateman, O. A., Slingsby, C. & Robinson, C. V. Polydispersity of a mammalian chaperone: Mass spectrometry reveals the population of oligomers in αB-crystallin. *Proc. Natl. Acad. Sci.* **100,** 10611–10616 (2003).

79. Benesch, J. L. P., Aquilina, J. A., Ruotolo, B. T., Sobott, F. & Robinson, C. V. Tandem Mass Spectrometry Reveals the Quaternary Organization of Macromolecular Assemblies. *Chem. Biol.* **13,** 597–605 (2006).

80. Zhong, Y., Hyung, S.-J. & Ruotolo, B. T. Ion mobility-mass spectrometry for structural proteomics. *Expert Rev. Proteomics* **9,** 47–58 (2012).

81. Laganowsky, A. *et al.* Membrane proteins bind lipids selectively to modulate their structure and function. *Nature* **510,** 172–175 (2014).

82. Bohrer, B. C., Merenbloom, S. I., Koeniger, S. L., Hilderbrand, A. E. & Clemmer, D. E. Biomolecule Analysis by Ion Mobility Spectrometry. *Annu. Rev. Anal. Chem.* **1,** 293–327 (2008).

83. Shvartsburg, A. & Smith, R. Fundamentals of traveling wave ion mobility spectrometry. *Anal. Chem.* **80,** 9689–9699 (2008).

84. Hilton, G. R. *et al.* C-terminal interactions mediate the quaternary dynamics of αB-crystallin. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **368,** 20110405 (2013).

85. Woods, L. A., Radford, S. E. & Ashcroft, A. E. Advances in ion mobility spectrometry–mass spectrometry reveal key insights into amyloid assembly. *Biochim. Biophys. Acta - Proteins Proteomics* **1834,** 1257–1268 (2013).

86. Zhou, M. *et al.* Mass Spectrometry of Intact V-Type ATPases Reveals Bound Lipids and the Effects of Nucleotide Binding. *Science (80-.).* **334,** 380–385 (2011).

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

88. Politis, A. *et al.* A mass spectrometry-based hybrid method for structural modeling of protein complexes. *Nat. Methods* **11,** 403–6 (2014).

89. Zhong, Y., Han, L. & Ruotolo, B. T. Collisional and Coulombic Unfolding of Gas-Phase Proteins: High Correlation to Their Domain Structures in Solution. *Angew. Chemie Int. Ed.* **53,** 9209–9212 (2014).

90. Niu, S., Rabuck, J. N. & Ruotolo, B. T. Ion mobility-mass spectrometry of intact protein–ligand complexes for pharmaceutical drug discovery and development. *Curr. Opin. Chem. Biol.* **17,** 809–817 (2013).

91. Allison, T. M. *et al.* Quantifying the stabilizing effects of protein–ligand interactions in the gas phase. *Nat. Commun.* **6,** 8551 (2015).

92. Hyung, S.-J., Robinson, C. V. & Ruotolo, B. T. Gas-Phase Unfolding and Disassembly Reveals Stability Differences in Ligand-Bound Multiprotein Complexes. *Chem. Biol.* **16,** 382–390 (2009).

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

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

95. Michaelevski, I., Kirshenbaum, N. & Sharon, M. T-wave Ion Mobility-mass Spectrometry: Basic Experimental Procedures for Protein Complex Analysis. *J. Vis. Exp.* 1–7 (2010). doi:10.3791/1985

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

97. Bush, M. F. *et al.* Collision cross sections of proteins and their complexes: a calibration framework and database for gas-phase structural biology. *Anal. Chem.* **82,** 9557–65 (2010).

98. Mesleh, M. F., Hunter, J. M., Shvartsburg, A. A., Schatz, G. C. & Jarrold, M. F. Structural Information from Ion Mobility Measurements: Effects of the Long-Range Potential. *J. Phys. Chem.* **100,** 16082–16086 (1996).

99. Marklund, E. G., Degiacomi, M. T., Robinson, C. V., Baldwin, A. J. & Benesch, J. L. P. Collision Cross Sections for Structural Proteomics. *Structure* **23,** 791–799 (2015).

100. Bleiholder, C., Wyttenbach, T. & Bowers, M. T. A novel projection approximation algorithm for the fast and accurate computation of molecular collision cross sections (I). Method. *Int. J. Mass Spectrom.* **308,** 1–10 (2011).

101. Mack, E. Average cross-sectional areas of molecules by gaseous diffusion methods. *J. Am. Chem. Soc.* **47,** 2468–2482 (1925).

102. von Helden, G., Hsu, M. T., Gotts, N. & Bowers, M. T. Carbon cluster cations with up to 84 atoms: structures, formation mechanism, and reactivity. *J. Phys. Chem.* **97,** 8182–8192 (1993).

103. Benesch, J. L. P. & Ruotolo, B. T. Mass spectrometry: come of age for structural and dynamical biology. *Curr. Opin. Struct. Biol.* **21,** 641–9 (2011).

104. Blow, D. *Outline of Crystallography for Biologists*. (OUP Oxford, 2002). at <https://books.google.com/books?id=14dgAwAAQBAJ&pgis=1>

105. Wlodawer, A., Minor, W., Dauter, Z. & Jaskolski, M. Protein crystallography for non-crystallographers, or how to get the best (but not more) from published macromolecular structures. *FEBS J.* **275,** 1–21 (2008).

106. Muñoz, I. G. *et al.* Crystal structure of the open conformation of the mammalian chaperonin CCT in complex with tubulin. *Nat. Struct. Mol. Biol.* **18,** 14–9 (2011).

107. Rhodes, G. *Complementary Science : Crystallography Made Crystal Clear (3rd Edition).* (Academic Press, 2014). at <http://site.ebrary.com/lib/bodleian/docDetail.action?docID=10138286>

108. Doerr, A. Single-molecule structure determination. *Nat. Methods* **8,** 45–45 (2011).

109. Smyth, M. S. & Martin, J. H. X Ray Crystallography. *J. Clin. Pathol. Mol. Pathol.* **53,** 8–14 (2000).

110. Leslie, A. G. W. & Powell, H. R. in *Evol. Methods Macromol. Crystallogr. Struct. Path to Underst. Mech. Action CBRN Agents* (eds. Read, R. J. & Sussman, J. L.) 41–51 (Springer Netherlands, 2007). doi:10.1007/978-1-4020-6316-9_4

111. Kabsch, W. XDS. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66,** 125–132 (2010).

112. Evans, P. R. & Murshudov, G. N. How good are my data and what is the resolution? *Acta Crystallogr. Sect. D Biol. Crystallogr.* **69,** 1204–1214 (2013).

113. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40,** 658–674 (2007).

114. Murshudov, G. N. *et al. REFMAC* 5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **67,** 355–367 (2011).

115. Adams, P. D. *et al.* PHENIX : a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66,** 213–221 (2010).

116. Karplus, P. A. & Diederichs, K. Linking Crystallographic Model and Data Quality. *Science (80-.).* **336,** 1030–1033 (2012).

117. Painter, J. & Merritt, E. A. TLSMD web server for the generation of multi-group TLS models. *J. Appl. Crystallogr.* **39,** 109–111 (2006).

118. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66,** 486–501 (2010).

119. Chen, V. B. *et al.* MolProbity : all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66,** 12–21 (2010).

120. Terwilliger, T. C. *et al.* Iterative-build OMIT maps: map improvement by iterative model building and refinement without model bias. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **64,** 515–524 (2008).

121. Nogales, E. The development of cryo-EM into a mainstream structural biology technique. *Nat. Methods* **13,** 24–27 (2016).

122. Ben-Shem, A. *et al.* The Structure of the Eukaryotic Ribosome at 3.0 Å Resolution. *Science (80-.).* **334,** 1524–1529 (2011).

123. Engel, C., Sainsbury, S., Cheung, A. C., Kostrewa, D. & Cramer, P. RNA polymerase I structure and transcription regulation. *Nature* **502,** 650–655 (2013).

Chapter 2: Native mass spectrometry of the proteostasis machinery

This chapter describes the success achieved in applying native MS to the study of the intricate systems of the proteostasis network. It was prepared for publication in the book 'Protein Misfolding, Aggregation and Disease: Insights from Mass Spectrometry', with equal contribution from Dr. Dale A. Shepherd and myself. For the initial draft, I wrote the Foldases and Conclusions and Perspectives sections. As part of the editing process, I extensively amended all sections of the manuscript after ensuring I was familiar with all the research articles described. I have included it here as an overview of the unique contribution afforded by native MS to knowledge of the sHSPs and the proteostasis network, particularly because other techniques employed in this work are highly characterised and their insight on the sHSPs is well documented[1–4]*. Moreover, I feel it highlights the fruitful application of MS among multiple complementary technologies to investigate these highly complicated and dynamic systems. This is an approach I employed and orchestrated in the study of HSP27, as presented in Chapter 3 – 6 of this thesis.*

1 Introduction

The proteostasis network consists of an array of macromolecular machines tasked with maintaining the cell in a functional steady-state. Two major clusters in the network, the molecular chaperones and the protein degradation machinery, act to subdue the deleterious effects associated with improper polypeptide folding. The proteostasis machinery is highly diverse and complex, including proteins exhibiting dynamic structures, populating a range of oligomeric states, and interacting with multiple binding partners. These factors conspire to make these proteins challenging for structural biologists, and have necessitated the development of novel approaches to elucidate their molecular details. One such approach is native MS, a method that has enabled important insights on the structure and function of a wide range of proteostasis network components. In turn, these complex systems have stimulated the development of the technique itself. This chapter gives an overview of recent native MS studies conducted on molecular chaperones and the degradation machinery, illustrating the utility of the approach to study oligomerisation, interactions, quaternary dynamics and the three-dimensional structure of proteins and protein complexes. The diversity of information that can be gained from native MS experiments demonstrates their key role in understanding molecular mechanism in general, and the proteostasis machinery in particular.

Proteins require their correct three-dimensional structure in order to carry out their intended functions. However environmental factors, for example temperature, pH, salts and the presence or absence of other biomolecules[5,6], can dramatically affect protein structure and force them to

explore previously inaccessible regions of their conformational energy landscape. When allowed to do this, many proteins will populate structures that expose buried hydrophobic residues to the bulk solvent, which can lead to loss of function, or gain of toxic function[7,8]. Oftentimes this partial unfolding increases the propensity for aggregation into forms that are detrimental to the homeostasis of the cellular environment. As a result, many important diseases are associated with protein misfolding and the formation of insoluble aggregates, including Parkinson's, Alzheimer's, and type-II diabetes[9].

Mechanisms are therefore required to minimise the perturbative effects of environmental fluctuations and stresses, and so maintain protein homeostasis. This "proteostasis" is ensured by a wide and diverse range of proteins and macromolecular assemblies, two major groups of which are the molecular chaperones and the protein degradation machinery (Figure 2 - 1)[10]. In general terms, the molecular chaperones prevent aggregation by direct interaction with a partially folded substrate, either trapping it in a recoverable state ready for further processing, or catalysing refolding directly[11]. The former is termed "holdase" activity, with the most prominent exponents being the sHSPs[12]. The latter, "foldases", include the HSP70 and HSP90 systems and the cylindrical chaperonins (HSP60), which stimulate folding in a manner dependent on ATP binding and hydrolysis[11]. The protein degradation machinery works very closely with the molecular chaperones and includes the ubiquitin-proteasome system and the Clp proteases, the former being dominant in eukaryotes and archaea. These "degradases" act to recognise and break down either aberrant proteins[13], or other proteins requiring tight control of their cellular

concentrations[14]. The myriad of proteins, their interconnected functions, and tight regulation thereby combine to make the proteostasis network extremely intricate and complex.

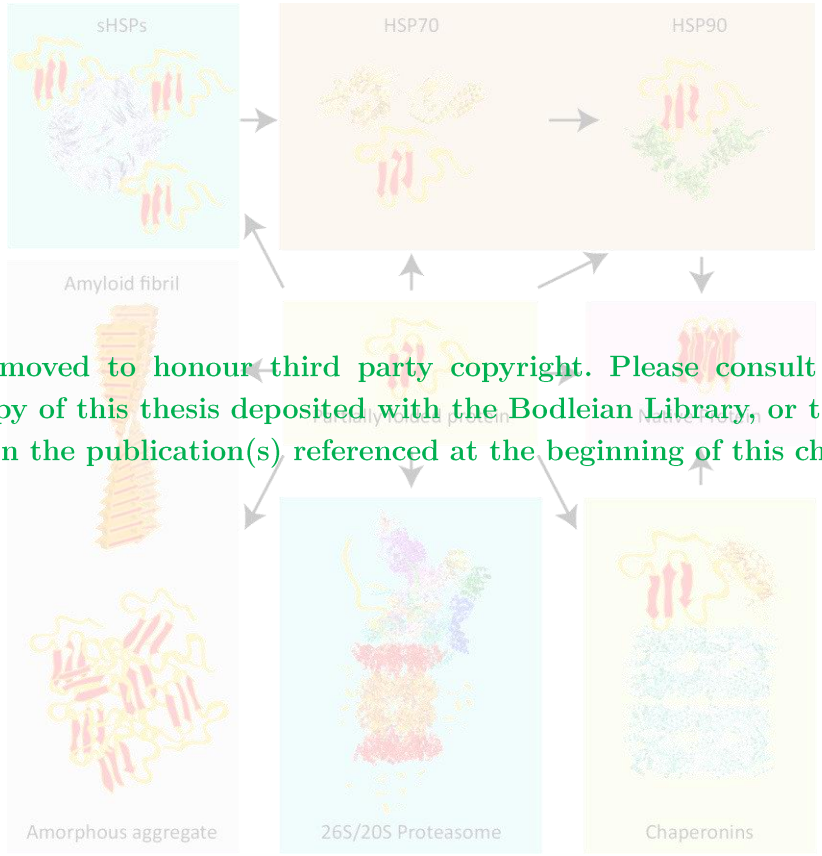


Figure 2 - 1 Molecular chaperones in the proteostasis network. A partially folded protein (yellow) may experience a number of fates in the cell. Spontaneously, it may (re)fold into its native conformation (pink), or aggregate either into stable amyloid fibrils or amorphous aggregates (violet). To prevent such aggregation, the cell can process the polypeptide by a number mechanisms. The unfolded species can be bound by the holdase sHSPs (turquoise) and stabilised. The substrate in these complexes can then be recruited by the HSP70 and HSP90 systems (orange) for ATP-dependent refolding, which can also act directly on the partially folded protein. Alternatively, the (re)folding can be catalysed by the large chaperonin complexes (green). Rather than chaperone-assisted refolding, the protein may instead be targeted for degradation, for instance by the proteasome (blue). (Note: directional arrows are shown to illustrate the fates of the partially folded protein, but reversibility of at least some of the processes is known).

MS has been instrumental to the study of proteostasis, as demonstrated by important proteomics studies revealing the response of the proteome to stress and ageing[15,16]. MS-based proteomics was made possible by the development, more than two decades ago, of soft ionisation techniques to

enable the transfer of intact polypeptides into vacuum. Alongside the miniaturisation of ESI sources to allow the use of aqueous buffers and low sample volumes, and the modification of mass spectrometer ion transfer stages, these technological advances culminated to allow the analysis of biomolecular complexes held together by non-covalent interactions[17,18]. Such native MS studies benefit from high speed and separative power, such that small amounts of heterogeneous mixtures may be analysed, and provide distinct benefits to complement traditional approaches such as X-ray crystallography and NMR spectroscopy. These capabilities have rendered native MS very useful for structural biologists and biophysicists, providing structural and dynamical insights into a wide range of protein assemblies, including those associated with nucleic acids and membranes[19–24]. Native MS has brought significant insight to the structure and function of the protein machinery involved in proteostasis. In turn, the demands of studying these intricate and inter-connected complexes have driven the development of the technique itself: indeed, the molecular chaperone SecB was one of the earliest examples of a protein complex to be studied by means of native MS[25]. Here we describe the development of native MS approaches and the insights they have provided into the stoichiometry, dynamics, structure and interactions of the machinery responsible for the maintenance of proteostasis.

2 Holdases: Small Heat-Shock Proteins

sHSPs are low molecular weight proteins that assemble into oligomers with a varying number of subunits, and are in many cases strongly up-regulated under cellular stress conditions[26,27]. The canonical function of sHSPs, which does not hold for all members of the family[28], is that they act

as ATP-independent holdases, trapping misfolded proteins in a refolding-competent state before recruitment of ATP-dependent systems to facilitate refolding or degradation[3,29–31]. They exhibit a common domain architecture consisting of a variable N-terminal region, a β-sheet rich core domain, and a relatively short and unstructured C-terminal tail harbouring a well-conserved 'IXI' motif through which they can co-assemble with other sHSP monomers to form large molecular weight oligomers. This modular structure also enables great flexibility in assembly, and as a result many sHSPs populate an ensemble of oligomers and are thereby difficult targets for structural biology investigations[4,29].

2.1 Quantifying the variable oligomerisation of sHSPs

Many sHSPs are polydisperse, that is they exist as a range of oligomers comprising different numbers of subunits. Native MS data of such samples can therefore be complex due to the overlap of multiple peaks (*i.e.* the charge-state envelopes) for each of the oligomers present in the ensemble[32]. This can frustrate peak assignment, and therefore make mass-measurement difficult or impossible by standard methods. The most established method for deconvolution of these spectra is a CID approach[33]. In this procedure, ions are accelerated into a low pressure of inert gas, resulting in collisions that convert their kinetic energy into internal modes, and lead to dissociation[34]. The general CID mechanism involves the expulsion of a highly charged monomer from a complex of n subunits, giving rise to signal at high m/z corresponding to stripped complexes with *n-1* subunits and disproportionately few charges. The separation between peaks in the stripped complex region is thereby greatly improved, and may allow their unambiguous

assignment and mass-measurement[33]. As a result, the intensities of the signal corresponding to the different charge state-envelopes can be extracted and used to quantify the oligomeric distribution of the ensemble. This approach was developed and demonstrated for bovine αB-crystallin, revealing for the first time the true extent of the polydispersity that had plagued structural studies of this protein[35]. The CID approach has been used to characterise oligomeric distributions for a variety of polydisperse mammalian sHSPs, including human αB-crystallin[36–38], αA-crystallin[39] and HSP27[40]. Notably, the mass distributions obtained in this way fit very well with data from solution-phase techniques, but offer vastly improved resolution of separation[41] (Chapter 1 Figure 1 - 1)..

An interesting showcase of this approach is the case of HSP26[42], the stress-inducible sHSP from *Saccharomyces cerevisiae*. Mass spectra obtained for *Sc*HSP26 reveal that it undergoes dissociation as a function of temperature, such that the abundance of oligomers decreases and both monomers and dimers became more prevalent (Figure 2 - 2A). Interestingly, the signal corresponding to oligomers narrows and shifts to higher m/z with increasing temperature. To interrogate this change, the oligomers were submitted to CID, such that *n-1*, *n-2*, and *n-3* stripped oligomers were formed by successive emission of monomers (Figure 2 - 2B). The effective resolution of the resulting data is such that the oligomers can be unambiguously assigned (Figure 2 - 2C), and their relative abundances quantified (Figure 2 - 2D). A comparison of the distributions obtained at 25°C and 43°C showed that while all the populated stoichiometries comprise an even number of subunits, and hence reveal the presence of a dimeric building-block, they shifted from 24-42 subunits at 25°C to predominantly 40 subunits at 43°C.

Similar temperature-dependent changes in oligomerisation have also been demonstrated for the homologous plant sHSPs HSP16.9 from *Triticum aestivum*[43], and HSP18.1 from *Pisum sativum*[44], which morph from monodisperse dodecamers into both lower and higher-order oligomers at elevated temperatures. Such changes in quaternary organisation at heat-shock temperatures are likely to be important for the activity of some sHSPs.

The precision of native MS for analysing such changes in oligomeric distribution has allowed the study of the finer details of sHSP regulation by mutation or post-translational modification. For instance, it was shown that mimicking phosphorylation of residues in the N-terminal region of αB-crystallin, which is known to modulate chaperone activity, also affects its oligomeric organization by destabilizing the dimer interface and therefore augmenting the tendency of the protein to populate odd stoichiometries[36,38,45]. Similarly, variations in pH were found to affect the dimer interface, such that it was weakened under the acidic conditions associated with ischemia[38]. The oligomeric distributions obtained in all these cases could then be fitted to a simple biophysical self-assembly model[41], allowing the quantification of the free energies of the inter- and intra-dimer interfaces, and how they varied upon perturbation[38].

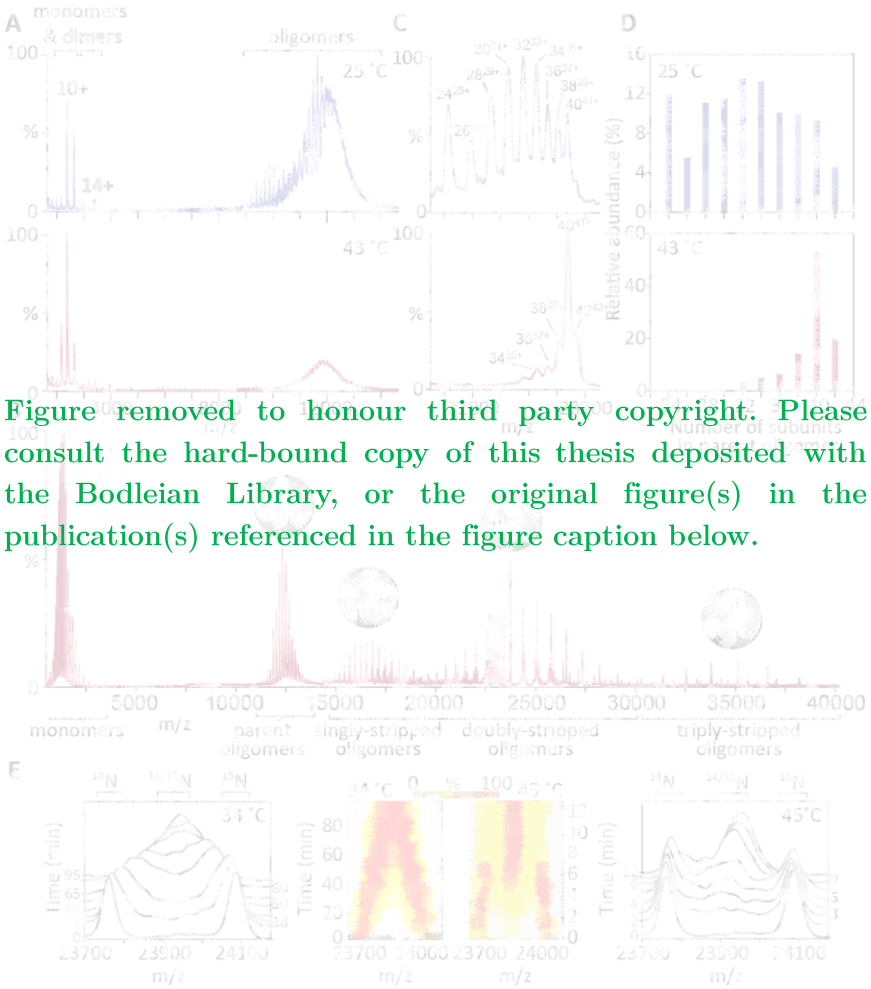


Figure removed to honour third party copyright. Please consult the hard-bound copy of this thesis deposited with the Bodleian Library, or the original figure(s) in the publication(s) referenced in the figure caption below.

Figure 2 - 2 Self-assembly of ScHSP26[42]. A) Native mass spectra of ScHSP26 at 25 °C (upper, blue) and 43 °C (lower, red) show a shift to both higher-order oligomers, and dissociation into monomers and dimers at heat-shock temperatures. B) CID was used to disentangle the overlapping charge states of ScHSP26 at 43 °C. This yields highly charged monomers in the low m/z region (extended polypeptide, grey) and baseline-resolved 'stripped' oligomers in the high m/z region of the spectrum (spheres missing one, two or three segments). C) An expansion of part of the CID mass spectrum (B, grey box) for 25 °C (upper, blue) and 43 °C (lower, red) allows the assignment of the doubly stripped stoichiometries. D) Based on the intensities in the mass spectra, and the predictable nature of CID, the oligomeric distribution of the polydisperse ensemble in solution can be back-calculated for 25 °C (upper, blue) and 43 °C (lower, red). E) Subunit exchange was monitored at 34 °C (left) and 45 °C (right) by incubating ScHSP26 with its ^{15}N-labelled equivalent. Mass spectra recorded in real time show that peaks corresponding to unlabelled and labelled (^{15}N) ScHSP26(n-2)$^{(n-2)+}$ (where n-2 is the number of subunits in the doubly stripped oligomers) coalesce over time, indicative of the formation of hetero-oligomers. The subunit exchange is not only much faster at 45 °C, but also appears co-operative, with both un-exchanged and fully exchanged oligomers detected simultaneously. Fitting of the experimental data to a kinetic model (middle) reveals that a single rate constant can account for the 34 °C data, whereas a minimum of two rate constants are required to reproduce that obtained at 45°C.

An interesting result was obtained from such analysis of the oligomeric distribution of αB-crystallin upon alanine-scanning the C-terminal region[46]. Remarkably, it was found that the removal of interacting residues in the C-terminus, thereby weakening the inter-dimer interactions, resulted in an increased propensity to populate even stoichiometries, *i.e.* a strengthening of the intra-dimer interface. In other words, the two interfaces are allosterically coupled, such that energy can be transferred between them. This was corroborated by experiments monitoring the binding of C-terminal peptides of αB-crystallin to a dimeric core domain construct: the dissociation constants obtained revealed the communication between the inter- and intra-dimer interfaces. This negative compensation acts to ensure that the mean oligomer size remains essentially constant, which may be important in maintaining the solubility of αB-crystallin at its high concentration in the eye lens[46].

The combination of CID with native MS has therefore provided unparalleled resolution of the oligomeric distributions of sHSPs, as well as enabled thermodynamic insights into their self-assembly. Nevertheless, alternative approaches to CID for the interpretation of complex mass spectra have been developed, and their utility for the sHSPs has already been demonstrated. These include performing positive-charge reduction through the addition of electrons[47], or through the development of data deconvolution strategies[48,49]. The combination of both computational and experimental techniques is likely to be particularly profitable, such that sets of MS data are analysed in concert. This has been demonstrated with UniDec software, where all the data corresponding to differentially stripped oligomers produced in a CID spectrum are combined to give an improved description of the solution distribution[48]. Initiatives such as these

that combine cutting edge MS experimental strategies with objective data analysis will result in ever more accurate and precise quantification of the stoichiometries populated by sHSPs.

2.2 Monitoring the quaternary dynamics of sHSPs

The assembly and disassembly dynamics of sHSPs are considered to be important to their function, with the equilibrium between oligomers and smaller subunits likely a mechanism of regulating their function[50]. A popular hypothesis is that the rapid subunit-exchange processes the proteins undergo act to transiently expose substrate-binding surfaces. The speed of MS analysis allows the monitoring in real time of reactions on the second timescale, and has been exploited for the study of sHSP subunit exchange. This experiment is performed by following the rate of change in oligomer mass arising from the incorporation of subunits of a different mass[51]. This is achieved either through employing isotopic labels, or by using an isoform or homologous protein. The major advantage is this approach does not require the addition of artificial fluorescent tags that could affect the protein structure and exchange rates.

It is thought that many sHSPs have the ability to co-assemble into hetero-oligomers. Two dodecameric sHSPs from *Arabidopsis thaliana*, HSP17.6 and HSP18.1, were shown to undergo subunit exchange with each other to form a distribution of hetero-dodecamers at equilibrium[52]. Notably, only hetero-oligomers with an even number of each protein were populated, suggesting that exchange was mediated by the transfer of dimeric subunits. This contrasted with a previous study of the subunit exchange between HSP18.1 from *Pisum sativum* and HSP16.9 from *Triticum aestivum*, in which the transfer of both dimers and monomers was observed[53]. The data

showed that the dimers exchanged faster than monomers, indicating a lower barrier to the breakage of inter-dimer interactions than intra-dimer interactions. By monitoring the reaction at different temperatures, the extraction of activation energies is possible, facilitating a quantitative kinetic analysis of the process[52].

An interesting temperature-dependent behaviour was observed when monitoring the quaternary dynamics of polydisperse *Saccharomoyces cerevisiae* HSP26[42]. At temperatures near optimal for growth, the subunit exchange between unlabelled and ^{15}N-labelled HSP26 occurred via gradual mixing of subunits into hetero-oligomers (34 °C, Figure 2 - 2E). However, at heat-shock temperatures, the exchange proceeds by a different mechanism, with both un-exchanged and fully exchanged oligomers being present simultaneously (45 °C, Figure 2 - 2E). Fitting the data reveals that at least two different exchange processes, with markedly different temperature dependencies, are required to account for the observed behaviour[42]. Similar subunit exchange experiments were used to characterise the effects of alanine-scanning the C-terminal tail of αB-crystallin, using the same mutants that displayed the allosteric effects evidenced by the oligomeric distributions[46]. MS showed the exchange rates to be much greater for the mutants relative to the wild-type, behaviour that could only be rationalized by a simultaneous increase in both the subunit association and dissociation rates.

Together these studies highlight the power of MS in interrogating quaternary dynamics. Not only can overall exchange rates be monitored in real time, but MS allows both the identification of the exchanging species and extraction of their individual rates and Arrhenius parameters. It thus

provides detailed biophysical insights which can be correlated with functional studies on chaperone activity and regulation in order to obtain mechanistic insight.

2.3 Disentangling chaperone-substrate interactions

A key to understanding the activity of sHSPs will be elucidating the nature of the interactions between the chaperone and the misfolded substrate. While much remains to be discovered in this area, native MS has enabled the field to make significant steps forward. For instance, incubation of αB-crystallin with the amyloidogenic protein β-2-microglobulin (β2m) did not yield the typical β2m pre-fibrillar oligomers usually observed in steady-state with monomeric protein after several days, rather the distribution was still biased toward the monomer[54]. From this data it is clear that αB-crystallin can inhibit amyloid formation in sub-stoichiometric amounts, likely by binding the lowly populated misfolded state of β2m[55], arresting further aggregation and promoting spontaneous refolding. Given the emerging use of MS as a means to study amyloid formation and for the detection of pre-fibrillar oligomers[22], this may provide a useful assay for the analysis of chaperone-substrate interactions.

In contrast to fibrillar aggregation, it has been shown that mammalian sHSPs can form stable complexes with misfolded substrates *en route* to forming amorphous aggregates[56]. Similar complexes have been observed with plant sHSPs and a range of model substrates[57], but have been difficult to study in detail due to their heterogeneity. The most detailed information as to the stoichiometries of interaction have come from MS, employing methods developed for analyzing polydisperse systems. Complexes between *Pisum sativum* HSP18.1 and luciferase were

analysed by means of tandem-MS[44], and were revealed to consist of a variable number of both sHSP and luciferase monomers (Figure 2 - 3A). Extrapolating the MS data according to gel-filtration traces revealed that more than 300 individual stoichiometries are populated during the protection process. This heterogeneity is bewildering, particularly when considering the capacity of sHSPs to coassemble, and reveals these chaperones to be unique in the diversity and plasticity of interactions they make.

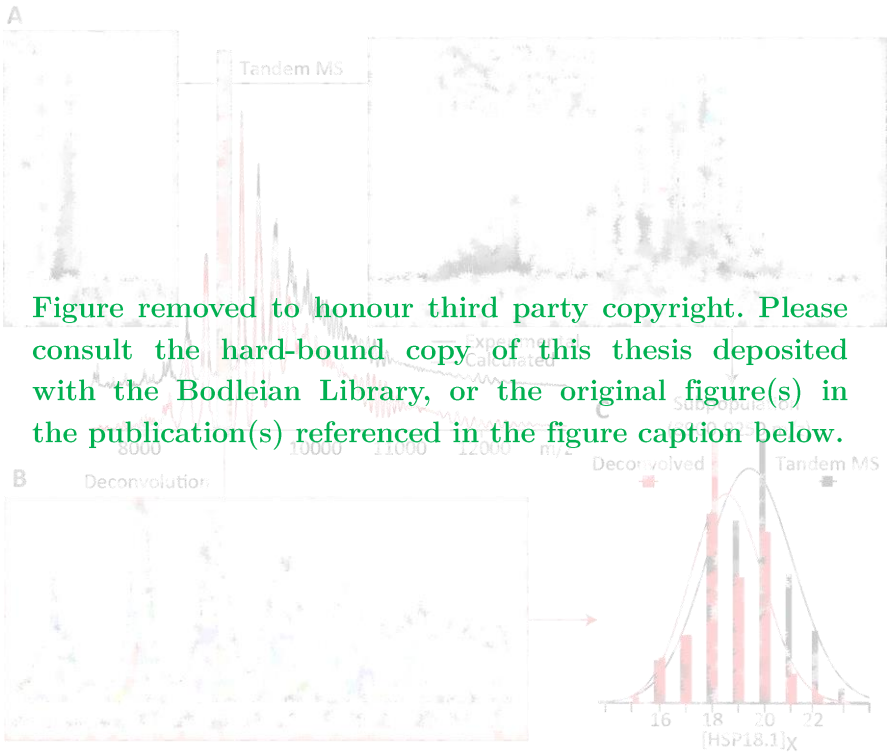


Figure removed to honour third party copyright. Please consult the hard-bound copy of this thesis deposited with the Bodleian Library, or the original figure(s) in the publication(s) referenced in the figure caption below.

Figure 2 - 3 Quantifying sHSP-substrate stoichiometries[44,49]. A) Tandem MS can be used to extract the stoichiometries formed between PsHSP18.1 and the model substrate luciferase. The stoichiometries are assigned from the stripped complexes in CID spectra (grey box, right). B) Alternatively, a spectrum deconvolution approach can be employed, with theoretical spectra calculated for a range of sHSP-substrate stoichiometries and their abundances optimised to best fit the data. The lines show the experimental (black) and theoretical spectra (coloured). The calculated spectrum (A, red) is the sum of multiple simulated mass spectra for individual stoichiometries. C) The resulting distribution agrees well with data obtained from tandem MS, revealing the compatibility of the two methods.

Building on this work, a deconvolution algorithm was developed for interpreting mass spectra of sHSP-substrate complexes[49], foregoing the need for CID-MS analysis. This approach yielded

stoichiometric distributions that are in good agreement with results obtained by CID-MS (Figure 2 - 3B,C). Interestingly, it was shown that the native quaternary structure of the substrate was reflected in the distribution, suggesting that protection occurs early during denaturation[49]. In general, the combined data showed that the number of sHSP subunits in the ensemble scale with the number of protected substrates, and revealed the preferential incorporation of an even number of sHSP subunits into the complexes, highlighting the likely importance of HSP18.1 dimers in chaperone action. The presence of two-dimensions of heterogeneity is remarkable considering that HSP18.1 is essentially monodisperse in the absence of substrate, and emphasizes the power of MS in deconvolving heterogeneity. Therefore, while there are still many challenges to be faced in the study of sHSP-substrate interactions, it is clear that MS will play a significant role in overcoming them.

2.4 From stoichiometries to structures using ion mobility MS

Ion mobility spectrometry (IM) allows the size measurement of ions in vacuum, providing semi-orthogonal information to MS alone. By measuring the time taken for an ion to traverse a gas-filled device of known length and electric field strength, the collision cross-section (CCS) can be determined[58]. Structural information can then be obtained by comparison of the experimental values to theoretical CCS values calculated from candidate structures. Given the great utility of MS for the analysis of polydisperse proteins, and the absence of structural data for them, IM-MS represents an excellent opportunity for obtaining architectural insights.

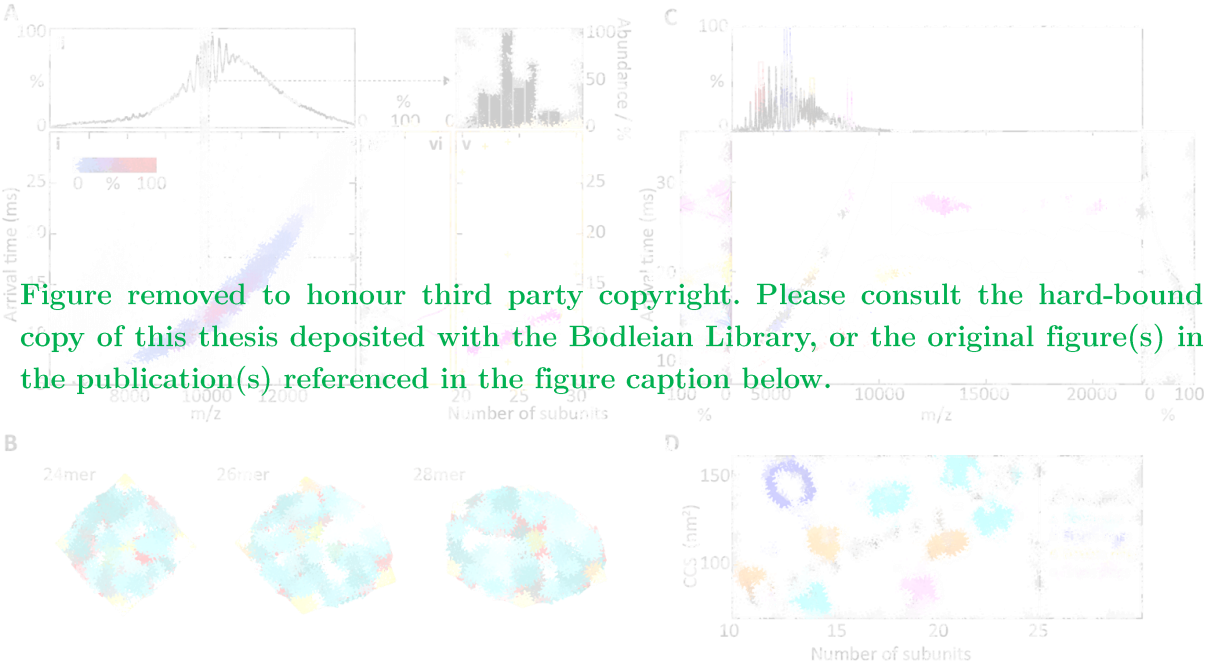


Figure removed to honour third party copyright. Please consult the hard-bound copy of this thesis deposited with the Bodleian Library, or the original figure(s) in the publication(s) referenced in the figure caption below.

Figure 2 - 4 Obtaining structural models for polydisperse proteins using IM-MS[59,60]. A) An IM-MS spectrum for polydisperse αB-crystallin is complex and convoluted in both m/z and arrival time dimensions (i,ii). When a subpopulation of oligomers are mass-selected (i,ii, grey slice) and subjected to CID their underlying oligomeric distribution can be extracted (iii). Similarly, the arrival time can be recorded for that same mass-selected sub-population preserved intact within the mass spectrometer (iv). Theoretical CCS values were calculated for a database of structural models, built based on the crystal structure of the core domain and knowledge of subunit connectivity (v, orange crosses). A linear trend in CCS versus number of subunits (v, purple), weighted according to oligomer abundances, returns an excellent fit to the data (vi). Notably, some of the candidate models agree extremely well with the linear fit (v, ringed) and therefore are considered as likely structures for the oligomeric states interrogated. B) The structures obtained by IM-MS for αB-crystallin suggest an interconversion mechanism that proceeds by insertion of additional protomeric dimers at vertices of the polyhedral framework. This causes minimal rearrangement of the overall structure, providing a rationale for the facile inter-conversion of oligomers. C) An alternative approach to this off-line combination of tandem- and IM-MS is to perform CID after IM separation, within the same experiment, such that dissociation products are time-aligned with the parent. TaHSP16.9 is polydisperse at low pH, with a resulting highly complex IM-MS spectrum. By performing tandem-IM-CID-MS, selecting a range of different m/z windows (red, blue, orange, magenta), it was possible to obtain the arrival times for oligomers and simultaneously determine their stoichiometries from their stripped dissociation products. In this way, oligomers containing between 4 and 24 subunits were characterised. D) Comparison of the experimental CCSs obtained for the larger oligomers with the CCSs of models built from a combination of structural restraints. Architectures based on polyhedra and double-rings fit the data best, consistent with other members of the sHSP family.

However, while conducting IM experiments on monodisperse protein assemblies is now relatively straightforward, extracting accurate CCSs relies on knowledge of protein complex mass and charge state[58]. This is non-trivial when dealing with polydisperse ensembles such as those

populated by many sHSPs, where multiple charge states of a range of stochiometries overlap in *m/z*. A strategy was developed to tackle the problem which employs tandem-MS to determine the mass and charge distribution of a selected region of the native mass spectrum, and then exploits this information to fit the associated IM dimension (Figure 2 - 4A). Thus, likely architectures of the most populated stoichiometries in the αB-crystallin spectrum were determined, with excellent agreement between the measured CCSs of the 24mer, 26mer and 28mer and the theoretical CCS of polyhedral models constructed *in silico* from the X-ray structure coordinates of the αB-crystallin core domain (Figure 2 - 4B)[59]. Moreover, the architectures of the models suggested a plausible mechanism of interconversion between them[59].

To refine this methodology, a novel CID and IM strategy was developed to allow the measurement of the CCSs of the individual oligomers within the ensemble, without the need for data fitting[60]. In this approach, IM is performed prior to CID such that the drift time of the ions are measured before they are dissociated to find their oligomer identity. As such, stoichiometry and CCS determination are performed in the same experiment. The effectiveness of this strategy was demonstrated by application to *Triticum aestivum* HSP16.9, a dodecamer that becomes polydisperse on lowering the pH. Numerous regions of the HSP16.9 mass spectrum were isolated, and the arrival time distributions and CID spectra were collected for all of the underlying oligomeric species in each *m/z* window (Figure 2 - 4C). By comparing the calculated CCS values for each oligomer to several candidate structures, double-ring and polyhedral architectures were proposed to underlie the polydispersity of the protein. The next step in such IM-MS based investigations of sHSPs is to reduce the error in the CCS measurement and its implementation in

modeling, allowing much larger sets of models to be filtered according to experimental data, which will ultimately result in higher resolution structures.

3 Foldases: ATP-dependent molecular chaperones

Molecular chaperones that facilitate folding of nascent chains, or refolding of non-native states, using energy supplied by ATP binding and hydrolysis are known as foldases[12]. The most well-known of these are the HSP60, HSP70, and HSP90 families, all of which have proved more amenable to traditional structural biology techniques than the sHSPs, as they populate fewer stoichiometries than their polydisperse colleagues. Nonetheless, native MS studies have contributed significantly to the characterisation of their structure, function and interaction networks.

3.1 Comparing the myriad interactions of HSP70 and HSP90

HSP70 (in concert with its co-chaperone HSP40) is essential for *de novo* folding as well as refolding of non-native states (Figure 2 - 1). It acts through a kinetic partitioning mechanism[11], trapping an aggregation-prone protein and then releasing it upon interaction with nucleotide exchange factors for subsequent folding or re-binding. Each HSP70 monomer consists of a nucleotide-binding domain and a C-terminal substrate-binding domain, separated by a short hydrophobic linker. Native MS has been used together with other biophysical techniques to study the oligomerisation behaviour of HSP70[61]. By designing modified constructs of HSP70 to isolate the regions of the protein that contribute to oligomerisation, the authors were able to identify

the C-terminal region of the substrate-binding domain as interacting with the inter-domain linker of another monomer. Moreover, they demonstrated that a variant constituting the substrate-binding domain alone could accommodate two substrates per dimer, supporting a model in which the substrate-binding site is distinct from the oligomerisation region on the C-terminus.

Once HSP70 has interacted with the substrate protein, it may pass it to HSP90 (Figure 2 - 1). ATP-hydrolysis induces a conformational change in HSP90 allowing it to close on its substrate[62] and thus either maintain substrate structure or induce its transition to a functional state[63]. This occurs in cooperation with many co-chaperones and regulators, which affect different parts of the cycle and may depend on the nature of the substrate[63]. Understanding these highly dynamic and heterogeneous systems demands techniques that are able to identify multiple states rapidly and simultaneously.

Native MS was used to characterise the stoichiometry of HSP90 complexed to the co-chaperone Cdc37 and a cyclin-dependent kinase substrate, Cdk4[64]. The mass spectra revealed formation of $HSP90_2$-$Cdc37_1$-$Cdk4_1$ and $Cdc37_2$-$Cdk4_1$ complexes, the former of which was contrary to their previous prediction[65], demonstrating the utility of obtaining high-mass resolution data when characterising such intricate chaperone systems. Building on this capability, an extensive study employed native MS to probe part of the interaction network of HSP90 and HSP70, and the association of two co-chaperones: Hop and FKBP52[66]. The stoichiometry and assembly order of complexes formed by these proteins had been a matter of debate, but the authors were able to explore all possible complexes that these four proteins form by varying the concentrations of the different components, and the order in which they were added. This was facilitated by fitting

simulated spectra to the complicated experimental data in order to elucidate the relative contributions of each species[67]. Moreover, measuring the intensity of the peaks in the mass spectrum for each of the different sub-complexes as a function of concentration allowed the authors to construct models of complex formation and elucidate the dissociation constants (Figure 2 - 5A). Interestingly, the similarity in these values between the different complexes implies that the population of these complexes within the network can be finely tuned in response to cellular stress.

This work was extended to describe the interaction of the HSP70/40 and HSP90 cycles[68,69]. The study showed that the small dimeric population of HSP70 present at low micromolar concentrations[61] increased upon phosphorylation. The authors then employed a series of native MS experiments, alongside chemical cross-linking MS, to determine the intermediates involved in the loading of substrate onto a monomer of HSP70, and its subsequent dimerisation with another HSP70 molecule within a $HSP90_2$-HSP70-Hop complex. This resulted in a stable hexameric intermediate (Figure 2 - 5B) that, on addition of p23 co-chaperone, readily released the $HSP90_2$-$p23_2$-substrate complex, revealing that the substrate had been transferred from HSP70 to HSP90 (Figure 2 - 5C). As a consequence of the resolution of multiple species present simultaneously in the mass spectra, the authors were able to present the individual complexes involved in the partnering between these two cycles. These studies on the foldases highlight how native MS is a powerful means to interrogate and quantify the interactions within multi-component networks featuring many co-populated states.

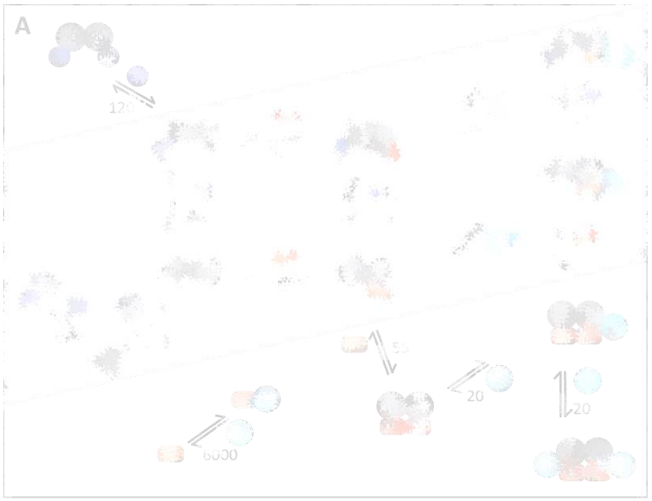


Figure removed to honour third party copyright. Please consult the hard-bound copy of this thesis deposited with the Bodleian Library, or the original figure(s) in the publication(s) referenced in the figure caption below.

Figure 2 - 5 Characterising the foldase network of HSP70, HSP40 and HSP90 by using native MS[66,68]. A) Dissociation constants (values in nM) were extracted for the various complexes formed when incubating HSP70, HSP90, Hop and FKBP52, thus permitting quantification of the interactions in this complex network. The grey shading denotes complexes that are likely to form, based on the measured cellular concentrations of the four proteins. B) A mass spectrum obtained at equimolar ratios of $HSP90_2$ (grey), Hop (red), HSP70 (blue) and substrate (orange) with a catalytic concentration of HSP40 shows the assembly of a stable "client loading complex" (grey circle) from intermediate species, and the requirement for Hop. The higher abundance of the complex containing two copies of HSP70 (grey versus green circles) suggests that HSP70 dimerization is an important step in this process. C) Addition of an equimolar amount of the p23 co-chaperone (purple) produces a substantial population of the $HSP90_2$-$p23_2$-substrate complex (purple circle), showing transfer of substrate from the 'client loading complex' (grey circle) has occurred.

3.2 Insights into HSP60 beyond its role as a native MS standard

The HSP60 (or chaperonin) protein class is subdivided into group I (found in bacteria, chloroplasts and mitochondria) and group II, which are found in archaea and the eukaryotic cytosol. Both classes encourage the formation of the native state by encapsulation in an "Anfinsen cage", and are involved in *de novo* protein folding downstream of HSP70 as well as refolding[70]. GroEL, from *E. coli*, is the paradigm for the type I chaperonins, and is the most studied protein in this family. It has been ever-present in native MS development since the 1990s, when GroEL was used to demonstrate the wide mass range attainable, and that native MS could identify the subcomplex connectivity of a macromolecular species[71]. GroEL has since become a *de facto* standard for MS proof-of-principle experiments. For example, the application of surface-induced dissociation to GroEL produced a range of product ions, most notably a heptameric dissociation product, indicative of the sub-structure underpinning the native 14-mer[72].

However GroEL has been a subject of significant interest in its own right. An influential study followed the refolding of gp23, the major capsid protein of bacteriophage T4, by complexes of GroEL associated with the bacteriophage cochaperonin gp31[73]. When the co-chaperone was absent, the authors found that up to two gp23 substrates can bind the GroEL 14-mer. When gp31 and ADP were added, a single species was detected corresponding to $GroEL_1$-$gp31_1$-$gp23_1$, indicating that gp31 association in the presence of ADP reduced the affinity of a second gp23 molecule. When ATP was used instead of ADP, the GroEL-gp31-gp23 complex was replaced by GroEL alone, GroEL-gp31 and a hexamer of gp23. This implied that the gp23 had been refolded into an oligomerisation-competent state. Conversely, if the *E. coli* co-chaperone GroES was

added instead of gp31, no refolded gp23 was produced, in agreement with the *in vivo* requirement of gp31[74]. This work was followed by further studies from the same laboratory on the nature of the GroEL chaperonin[75,76], including the use of IM-MS to measure the CCSs of fourteen different complexes between GroEL, the cochaperonins GroES and gp31, nucleotides and various substrates[77]. The CCS of GroEL was not found to change significantly between the *apo*- and *holo*-forms, consistent with substrate binding within one of the GroEL cavities[77].

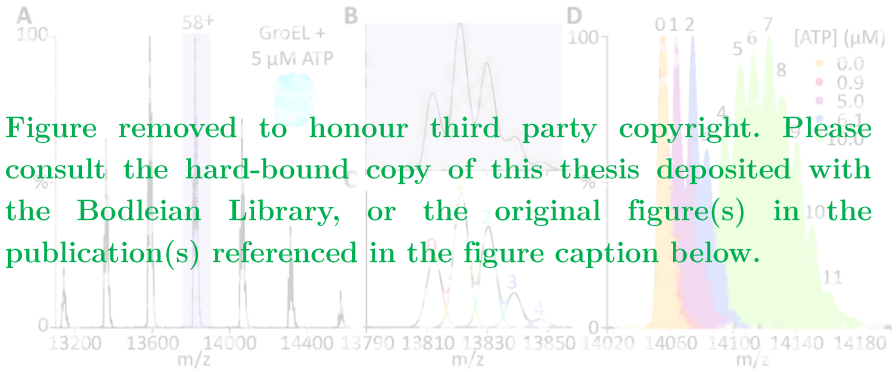


Figure 2 - 6 Elucidating the details of ATP-binding to GroEL[78]. A) An excellently resolved mass spectrum of 1 µM GroEL in the presence of 5 µM ATP showing "fine structure" in the charge states. B) Expansion of the 58+ peak demonstrates a series of ATP-bound forms of the GroEL. C) Deconvolution of these states allows the quantification of the relative abundances of the GroEL tetradecamer carrying between 0 and 4 molecules of ATP. The summed deconvolution in is in close agreement with the experimental distribution (B). D) A larger number of ATP molecules are bound with increasing ATP concentration. Superimposed spectra are for the 57+ charge state with 0-10 µM ATP, and peaks are labelled with the number of bound ATP molecules to which they correspond. The extracted abundances allowed the calculation of Hill coefficients of binding, and provided strong evidence for the nested model of allosteric co-operativity.

Native MS has also been employed to measure the binding affinity of ATP to GroEL[78]. Very well resolved mass spectra were obtained, revealing sequential binding of up to 14 ATP molecules (Figure 2 - 6). ATP-binding number distributions were extracted from the area under the peak for each GroEL species as a function of ATP concentration (Figure 2 - 6D), allowing the calculation of the Hill-coefficient of ATP-binding. The authors observed an increase in positive cooperativity with an increasing number of bound ATP, and then a decrease in cooperativity

when nearly all the sites were occupied. Furthermore, the 14 binding constants for sequential binding of ATP were calculated from the ATP concentration and provided evidence for the nested model of allostery in GroEL, for which experimental evidence was previously unavailable[79].

Other Group I chaperonins have been studied by means of native MS. For instance, an investigation into the oligomeric state of chaperonin 60.2 from *Mycobacterium tuberculosis* demonstrated that it formed only a monomeric species without ATP, and assembled into oligomers with ATP (14-mer, and all stoichiometries from heptamers down to monomers), thus answering ongoing discussion about the architecture of the species[80]. Native MS has also been employed to probe the assembly and heterogeneous character of chloroplast co-chaperonins, which were able to encase unfolded Rubisco when complexed with *E. Coli* GroEL in the presence of ATP, revealing their unusual hetero-oligomerisation capabilities[81].

CCT (TriC) is the eukaryotic Group II chaperonin. It comprises two octameric rings and is found in the cytosol. Along with HSP70, it prevents accumulation of toxic oligomers in Huntington's Disease and other neurodegenerative diseases[11], which made the elucidation of its structure and function very attractive. When a structure of bovine CCT was obtained, it contained tubulin bound to its inner cavities, which was a natural substrate remaining after extraction. This was complemented by MS which verified the association of tubulin to the CCT complex and demonstrated tubulin dissociation following incubation with ATP[82].

Archaea contain up to five genes encoding group II chaperonin subunits. A study combining genetic work and native MS showed that the three *cct* genes in *Haloferax volcanii* are each non-

essential when the others remain within the genome[83]. Mass spectra of the wild type CCT complexes allowed the identification of their stoichiometry, and revealed that they contained a combination of the different subunits encoded by the *cct* genes rather than a unique stoichiometry[83]. While the HSP60 family, and GroEL in particular, has proven extremely useful to mass spectrometrists for technique development, these studies demonstrate how in return MS has provided key insights into the assembly and function of these proteins.

4 Degradases: the Proteasome

Many proteins require their cellular concentrations to be tightly controlled in order to carry out their function, and to prevent toxic effects. In addition, proteins that have been trapped and bound by holdases or foldases but are not in a state that allows refolding into the functional form, must be disposed of in a controlled way. The fate of many of these proteins is that they are marked by ubiquitin ligases for recognition and degradation by the proteasome[84]. The full proteasome is referred to as the 26S particle, which is in turn made up of the 20S core and the 19S regulatory unit. A key study concerned the architecture of the 19S proteasome lid from *S. cerevisiae*, which is responsible for deubiquitination and contains nine subunits, and whose arrangement eluded structural biologists for some time. Tandem-MS was used to gain information on sub-complexes, and to generate information on the location of proteins within the full complex, be it in the core or on the periphery. This allowed the generation of a subunit interaction map and structural model of the 19S lid complex[85], which has since been validated by cryo-EM studies[86].

A more recent study focusing on the interacting partners of the proteasome from *S. cerevisiae* showed the structural organisation of the 19S base complex[87]. Additionally, interactions were observed between this complex and a variety of assembly chaperones, and their stoichiometries were deduced[87]. Further interaction maps and pseudo-atomic structures for the proteasome lid, as well as an assembly intermediate of the base, have since been proposed based on constraints from cross-linking, native MS, IM and tandem-MS[88,89]. These are in good agreement with previous models and demonstrate the utility of an integrative MS approach in studying difficult structural targets.

A study on the 20S particle from *Rhodococcus erythropolis* monitored the assembly from individual subunits in real-time[90]. The authors observed the initial formation of α-β heterodimers (a complex that had not been detected by previous experiments), followed by the assembly of unprocessed half-proteasomes, before the formation of intact 28-mers. This study demonstrates the use of MS in monitoring self-assembly in real time, particularly the assembly of large macromolecular complexes made up of subunits a great deal lower in mass than the final species. Recently, a study of the 20S proteasome of *Methanosarcina thermophila* again demonstrated the utility of the alternative dissociation products produced by surface induced dissociation, identifying an α_7-β_7 subcomplex that reflects the heptameric ring structure of the 20S precursor[91].

In the past decade, it has become apparent that many regulatory proteins lack defined three-dimensional structure[92]. These intrinsically disordered proteins have been shown to recognise binding partners either through short linear motifs, or folding-by-binding events[92]. It has been shown recently that such proteins are substrates for the 20S proteasome[93] rather than the 26S, as

there is no need for the deubiquitinating and unfolding activities of the 19S regulatory particle. An elegant MS-based study demonstrated a negative feedback regulatory mechanism for the 20S proteasome with respect to the degradation of disordered proteins[94] (Figure 2 - 7). The protein NQO1 is a cytosolic enzyme that reduces cellular quinones in the presence of a cofactor FAD[95]. The researchers discovered that FAD-bound NQO1 binds directly to the proteasome[94]. This binding was shown to halt proteolytic activity and therefore protect intrinsically disordered proteins from degradation. As apo-NQO1 was itself shown to be disordered, it was hypothesised that it would be degraded by the proteasome.

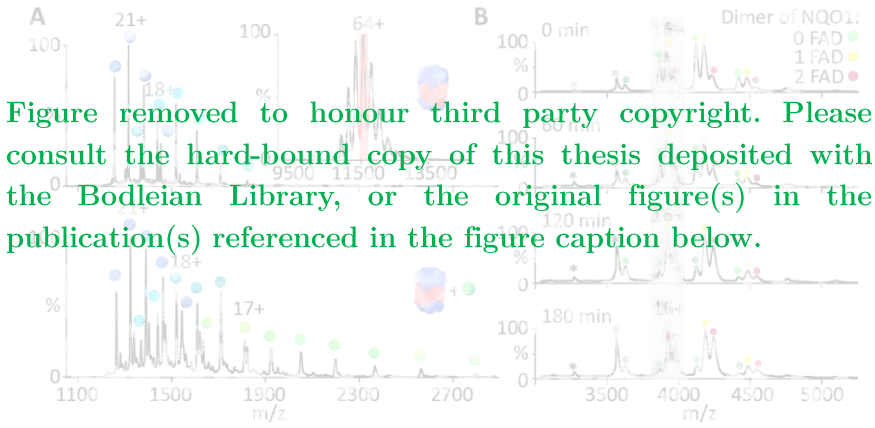


Figure 2 - 7 Using native MS to probe the regulation of the proteasome[94]. A) Mass selection and CID of the 64+ peak of the 20S proteasome (inset, red slice) yields signal for dissociated monomers corresponding to the two types of α-subunits (upper, blue and cyan spheres). After incubation with the regulator NQO1, an equivalent experiment yields signal for both the α-subunits and NQO1 (green), indicating that NQO1 was bound to the proteasome. B) When following the proteolysis of NQO1 by the proteasome over time, it is notable that the apo form (green) is degraded much faster than the FAD-bound forms (yellow, red). This is consistent with the observation that apo-NQO1 is intrinsically disordered, and therefore behaves as a 20S proteasome substrate.

By using MS to monitor a solution containing 20S proteasome and NQO1, the authors were able to demonstrate that apo-NQO1 was degraded faster than singly-bound NQO1 which was in turn degraded faster than holo-NQO1 (Figure 2 - 7B). These data showed that upon loss of FAD the fold of NQO1 is destabilised, effectively making it a 20S proteasome substrate and losing the

ability to protect intrinsically disordered proteins from degradation. This study highlights the dependence of proteasome activity on the levels of FAD, and hence the metabolic state of the cell, a crucial factor in proteostasis. More recently, the same group identified another 20S proteasome regulator, DJ-1[96]. They again employed MS in their characterisation of the behaviour of the protein, including showing binding of the DJ-1 protein to the proteasome, but not to the partially folded proteins whose proteasomal degradation it inhibits. Along with degradation and cellular assays, they thus built upon their previous study to suggest a model for the 20S proteasome regulation by DJ-1 and NQO1. These studies on the proteasome, together with others on the Clp protease pathway[97,98], demonstrate the wide ranging applications of native MS to protein degradation. They include the assessment of the stability of a protein fold, the study of binding of a small protein to a large macromolecular complex, and the monitoring of degradation of a protein in real time.

5 Conclusions and Perspectives

Here we have described how MS is an invaluable tool for the investigation of the proteostasis machinery. Its rapid development has facilitated further inquiry into the nature of these molecular machines, and indeed molecular chaperones themselves have been the subject of the studies progressing the application of native MS. While this chapter has focused on the application of native MS to the proteostasis network, there remains a wealth of research on the molecular chaperones using other MS-based approaches: proteomics, and hydrogen/deuterium exchange, foot-printing, or chemical cross-linking. Native MS is particularly powerful for proteins

which, due to their heterogeneity and highly dynamic nature, are refractory to traditional structural biology techniques such as protein crystallography and NMR spectroscopy. Its initial use in identifying the stoichiometries of monodisperse macromolecular protein complexes and sub-complexes has been expanded to the determination of the stoichiometries within heterogeneous mixtures, their relative abundance, dynamics, and even likely architectures.

Improvement and additions to the instrumentation, methodologies and analyses associated with MS will fuel further and more sophisticated investigations. For instance, recent technical improvements in desolvation and resolution have allowed the interrogation of highly heterogeneous samples, including the analysis of protein complex microheterogeneity[99]. This will be complemented by alternatives to CID that offer different dissociation routes or do not depend on the initial charge of the ion. Coupling improvements in the MS dimension with refinements in IM separation promises to provide an unprecedented wealth of structural information on a wide range of complexes and systems, as exemplified by the proteostasis network. With the development of complementary computational approaches continuing apace[48,100–102], the prospect of *bona fide* structure-determination from MS remains an attainable, if long-term, goal for the field.

The ultimate aim of researchers in the field of proteostasis it to understand the interactions made between components of the network in health and in disease. Protein misfolding diseases are strongly related to age, which is possibly due to the coupling of the pathways that regulate proteostasis with those that control longevity[103]. As a result, the investigation of interactions between aggregation-prone proteins and the molecular chaperone and degradation machinery will

likely drive much of the future native MS work in proteostasis. These interactions are inherently dynamic and difficult to capture, and relatively few investigations have applied native MS to such complexes thus far. However, the studies that have been conducted have established the framework and approaches required to effectively probe chaperone-substrate complexes which can now be applied to the myriad of species within the proteostasis network.

Amyloid, which consists of proteins that have not been successfully refolded by the proteostasis machinery before deposition, have been the subject of numerous MS investigations, including some that have attempted to identify and characterize inhibitors (see other chapters in this volume). An alternative to manipulating amyloidogenesis with therapeutics that interrupt aggregation is to develop molecules that target underperforming molecular chaperones or degradation assemblies, thereby stimulating the cellular machinery in its housekeeping tasks. It is the primary objective of native MS studies in this field to gain sufficiently deep insight into the proteostasis network in order that activators can be developed to combat diseases associated with proteostasis deficiency. Due to the common mechanisms underlying many of these conditions, the concept of panacea drugs for diseases of proteostasis remains a tantalizing possibility.

6 References

1. Kulig, M. & Ecroyd, H. The small heat-shock protein αB-crystallin uses different mechanisms of chaperone action to prevent the amorphous versus fibrillar aggregation of αlactalbumin. *Biochem. J.* **448,** 343–352 (2012).

2. Haslbeck, M. & Buchner, J. in *Stress Responses Methods Protoc.* (ed. Christine M. Oslowski) **1292,** 39–51 (Springer New York, 2015).

3. Haslbeck, M. & Vierling, E. A First Line of Stress Defense: Small Heat Shock Proteins and Their Function in Protein Homeostasis. *J. Mol. Biol.* **427,** 1537–1548 (2015).

4. Delbecq, S. P. & Klevit, R. E. One size does not fit all: The oligomeric states of αB crystallin. *FEBS Lett.* **587,** 1073–1080 (2013).

5. Wang, W., Nema, S. & Teagarden, D. Protein aggregation--pathways and influencing factors. *Int. J. Pharm.* **390,** 89–99 (2010).

6. Stigter, D., Alonso, D. V & Dill, K. E. N. A. Protein stability : Electrostatics and compact denatured states. *Proc. Natl. Acad. Sci. U. S. A.* **88,** 4176–4180 (1991).

7. Eichner, T. & Radford, S. E. A diversity of assembly mechanisms of a generic amyloid fold. *Mol. Cell* **43,** 8–18 (2011).

8. Chiti, F. & Dobson, C. M. Protein Misfolding , Functional Amyloid , and Human Disease. *Annu. Rev. Biochem.* (2006). doi:10.1146/annurev.biochem.75.101304.123901

9. Powers, E. T., Morimoto, R. I., Dillin, A., Kelly, J. W. & Balch, W. E. Biological and chemical approaches to diseases of proteostasis deficiency. *Annu. Rev. Biochem.* **78,** 959–91 (2009).

10. Balch, W. E. Adapting Proteostasis for Disease Intervention. *Science (80-.).* **916,** (2011).

11. Kim, Y. E., Hipp, M. S., Bracher, A., Hayer-hartl, M. & Hartl, F. U. Molecular Chaperone Functions in Protein Folding and Proteostasis. *Annu. Rev. Biochem.* **82,** 323–355 (2013).

12. Richter, K., Haslbeck, M. & Buchner, J. The heat shock response: life on the verge of death. *Mol. Cell* **40,** 253–66 (2010).

13. Vilchez, D., Saez, I. & Dillin, A. The role of protein clearance mechanisms in organismal ageing and age-related diseases. *Nat. Commun.* **5,** 5659 (2014).

14. Geng, F., Wenzel, S. & Tansey, W. P. Ubiquitin and proteasomes in transcription. *Annu. Rev. Biochem.* **81,** 177–201 (2012).

15. Walther, D. M. *et al.* Widespread Proteome Remodeling and Aggregation in Aging C. elegans. *Cell* **161,** 919–932 (2015).

16. Wallace, E. W. J. *et al.* Reversible, Specific, Active Aggregates of Endogenous Proteins Assemble upon Heat Stress. *Cell* **162,** 1286–1298 (2015).

17. Loo, J. A. Studying noncovalent protein complexes by electrospray ionization mass spectrometry. *Mass Spectrom. Rev.* **16,** 1–23 (1997).

18. Hilton, G. R. & Benesch, J. L. P. Two decades of studying non-covalent biomolecular assemblies by means of electrospray ionization mass spectrometry. *J. R. Soc. Interface* **9,** 801–816 (2012).

19. Benesch, J. L. P. & Ruotolo, B. T. Mass spectrometry: come of age for structural and dynamical biology. *Curr. Opin. Struct. Biol.* **21,** 641–649 (2011).

20. Sharon, M. Structural MS pulls its weight. *Science (80-.).* **340,** 1059–60 (2013).

21. Konijnenberg, A., Butterer, A. & Sobott, F. Native ion mobility-mass spectrometry and related methods in structural biology. *Biochim. Biophys. Acta - Proteins Proteomics* **1834,** 1239–1256 (2013).

22. Williams, D. M. & Pukala, T. L. Novel insights into protein misfolding diseases revealed by ion mobility-mass spectrometry. *Mass Spectrom. Rev.* **32,** 169–187 (2013).

23. Uetrecht, C., Rose, R. J., van Duijn, E., Lorenzen, K. & Heck, A. J. R. Ion mobility mass spectrometry of proteins and protein assemblies. *Chem. Soc. Rev.* **39,** 1633–55 (2010).

24. Thalassinos, K., Pandurangan, A. P., Xu, M., Alber, F. & Topf, M. Conformational States of Macromolecular Assemblies Explored by Integrative Structure Calculation. *Structure* **21,** 1500–1508 (2013).

25. Smith, V. F., Schwartz, B. L., Randall, L. L. & Smith, R. D. Electrospray mass spectrometric investigation of the chaperone SecB. *Protein Sci.* **5,** 488–94 (1996).

26. Beck, M. *et al.* Visual proteomics of the human pathogen Leptospira interrogans. *Nat.*

Methods **6,** 817–823 (2009).

27. Malmström, J., Beck, M., Schmidt, A., Lange, V. & Deutsch, E. W. Proteome-wide cellular protein concentrations of the human pathogen Leptospira interrogans. *Nature* **460,** 762–766 (2009).

28. Carra, S. *et al.* Different anti-aggregation and pro- degradative functions of the members of the mammalian sHSP family in neurological disorders. *Philos. Trans. R. Soc. B Biol. Sci.* **368,** 20110409 (2013).

29. Hilton, G. R., Lioe, H., Stengel, F., Baldwin, A. J. & Benesch, J. L. P. Small Heat-Shock Proteins : Paramedics of the Cell. (2012). doi:10.1007/128

30. Treweek, T. M., Meehan, S., Ecroyd, H. & Carver, J. A. Small heat-shock proteins : important players in regulating cellular proteostasis. *Cell. Mol. Life Sci.* **72,** 429–451 (2015).

31. Mchaourab, H. S., Godar, J. A. & Stewart, P. L. Structure and Mechanism of Protein Stability Sensors: Chaperone Activity of Small Heat Shock Proteins. *Biochemistry* **48,** 3828–3837 (2009).

32. Benesch, J. L. P. & Robinson, C. V. Mass spectrometry of macromolecular assemblies: preservation and dissociation. *Curr. Opin. Struct. Biol.* **16,** 245–51 (2006).

33. Benesch, J. L. P., Aquilina, J. A., Ruotolo, B. T., Sobott, F. & Robinson, C. V. Tandem mass spectrometry reveals the quaternary organization of macromolecular assemblies.

Chem. Biol. **13,** 597–605 (2006).

34. Benesch, J. L. P. Collisional activation of protein complexes: Picking up the pieces. *J. Am. Soc. Mass Spectrom.* **20,** 341–348 (2009).

35. Aquilina, J. A., Benesch, J. L. P., Bateman, O. A., Slingsby, C. & Robinson, C. V. Polydispersity of a mammalian chaperone: mass spectrometry reveals the population of oligomers in alphaB-crystallin. *Proc. Natl. Acad. Sci. U. S. A.* **100,** 10611–6 (2003).

36. Aquilina, J. A. *et al.* Phosphorylation of αB-Crystallin Alters Chaperone Function through Loss of Dimeric Substructure. *J. Biol. Chem.* **279,** 28675–28680 (2004).

37. Benesch, J. L. P., Ayoub, M., Robinson, C. V. & Aquilina, J. A. Small Heat Shock Protein Activity Is Regulated by Variable Oligomeric Substructure. *J. Biol. Chem.* **283,** 28513–28517 (2008).

38. Baldwin, A. J., Lioe, H., Robinson, C. V., Kay, L. E. & Benesch, J. L. P. αB-Crystallin Polydispersity Is a Consequence of Unbiased Quaternary Dynamics. *J. Mol. Biol.* **413,** 297–309 (2011).

39. Aquilina, J. A. *et al.* Subunit exchange of polydisperse proteins: mass spectrometry reveals consequences of αA-crystallin truncation. *J. Biol. Chem.* **280,** 14485–14491 (2005).

40. Aquilina, J. A., Shrestha, S., Morris, A. M. & Ecroyd, H. Structural and functional aspects of hetero-oligomers formed by the small heat shock proteins αB-crystallin and HSP27. *J. Biol. Chem.* **288,** 13602–9 (2013).

41. Hochberg, G. K. A. & Benesch, J. L. P. Dynamical structure of αB-crystallin. *Prog. Biophys. Mol. Biol.* **115,** 11–20 (2014).

42. Benesch, J. L. P. *et al.* The quaternary organization and dynamics of the molecular chaperone HSP26 are thermally regulated. *Chem. Biol.* **17,** 1008–17 (2010).

43. Benesch, J. L. P., Sobott, F. & Robinson, C. V. Thermal dissociation of multimeric protein complexes by using nanoelectrospray mass spectrometry. *Anal. Chem.* **75,** 2208–14 (2003).

44. Stengel, F. *et al.* Quaternary dynamics and plasticity underlie small heat shock protein chaperone function. *Proc. Natl. Acad. Sci.* **107,** 2007–2012 (2010).

45. Ecroyd, H. *et al.* Mimicking phosphorylation of αB-crystallin affects its chaperone activity. *Biochem. J.* **401,** 129–141 (2007).

46. Hilton, G. R. *et al.* C-terminal interactions mediate the quaternary dynamics of αB-crystallin. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **368,** 20110405 (2013).

47. Lermyte, F., Williams, J. P., Brown, J. M., Martin, E. M. & Sobott, F. Extensive Charge Reduction and Dissociation of Intact Protein Complexes Following Electron Transfer on a Quadrupole-Ion Mobility-Time-of-Flight MS. *J. Am. Soc. Mass Spectrom.* **26,** 1068–1076 (2015).

48. Marty, M. T. *et al.* Bayesian Deconvolution of Mass and Ion Mobility Spectra: From Binary Interactions to Polydisperse Ensembles. *Anal. Chem.* **87,** 4370–4376 (2015).

49. Stengel, F. *et al.* Dissecting Heterogeneous Molecular Chaperone Complexes Using a Mass Spectrum Deconvolution Approach. *Chem. Biol.* **19,** 599–607 (2012).

50. Hochberg, G. K. A. & Benesch, J. L. P. in *Big B. Small Heat Shock Proteins* (eds. Tanguay, R. M. & Hightower, L. E.) 87–100 (Springer International Publishing Switzerland, 2015). doi:10.1007/978-3-319-16077-1_3

51. Kondrat, F. D. L., Struwe, W. B. & Benesch, J. L. P. in *Struct. Proteomics High-Throughput Methods* (ed. Owens, R. J.) **1261,** 349–371 (Springer New York, 2015).

52. Painter, A. J. *et al.* Real-time monitoring of protein complexes reveals their quaternary organization and dynamics. *Chem. Biol.* **15,** 246–53 (2008).

53. Sobott, F., Benesch, J. L. P., Vierling, E. & Robinson, C. V. Subunit exchange of multimeric protein complexes. Real-time monitoring of subunit exchange between small heat shock proteins by using electrospray mass spectrometry. *J. Biol. Chem.* **277,** 38921–9 (2002).

54. Esposito, G. *et al.* Monitoring the interaction between β2-microglobulin and the molecular chaperone αB-crystallin by NMR and mass spectrometry: αB-crystallin dissociates β2-microglobulin oligomers. *J. Biol. Chem.* **288,** 17844–58 (2013).

55. Eichner, T., Kalverda, A. P., Thompson, G. S., Homans, S. W. & Radford, S. E. Conformational conversion during amyloid formation at atomic resolution. *Mol. Cell* **41,** 161–72 (2011).

56. Kulig, M. & Ecroyd, H. The small heat-shock protein αB-crystallin uses different mechanisms of chaperone action to prevent the amorphous versus fibrillar aggregation of α-lactalbumin. *Biochem. J.* **448,** 343–352 (2012).

57. Basha, E., O'Neill, H. & Vierling, E. Small heat shock proteins and α-crystallins: dynamic proteins with flexible functions. *Trends Biochem. Sci.* **37,** 106–17 (2011).

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

59. Baldwin, A. J. *et al.* The Polydispersity of αB-Crystallin Is Rationalized by an Interconverting Polyhedral Architecture. *Structure* **19,** 1855–1863 (2011).

60. Shepherd, D. A., Marty, M. T., Giles, K., Baldwin, A. J. & Benesch, J. L. P. Combining tandem mass spectrometry with ion mobility separation to determine the architecture of polydisperse proteins. *Int. J. Mass Spectrom.* **377,** 663–671 (2015).

61. Aprile, F. A. *et al.* Hsp70 oligomerization is mediated by an interaction between the interdomain linker and the substrate-binding domain. *PLoS One* **8,** e67961 (2013).

62. Wegele, H., Müller, L. & Buchner, J. Hsp70 and Hsp90--a relay team for protein folding. *Rev. Physiol. Biochem. Pharmacol.* **151,** 1–44 (2004).

63. Wandinger, S. K., Richter, K. & Buchner, J. The Hsp90 chaperone machinery. *J. Biol. Chem.* **283,** 18473–7 (2008).

64. Vaughan, C. K. *et al.* Structure of an Hsp90-Cdc37-Cdk4 complex. *Mol. Cell* **23,** 697–707 (2006).

65. Roe, S. M. *et al.* The Mechanism of Hsp90 Regulation by the Protein Kinase-Specific Cochaperone p50cdc37. *Cell* **116,** 87–98 (2004).

66. Ebong, I. *et al.* Heterogeneity and dynamics in the assembly of the Heat Shock Protein 90 chaperone complexes. *Proc. Natl. Acad. Sci. U. S. A.* **108,** 17939–17944 (2011).

67. Morgner, N. & Robinson, C. V. Massign: an assignment strategy for maximizing information from the mass spectra of heterogeneous protein assemblies. *Anal. Chem.* **84,** 2939–48 (2012).

68. Morgner, N. *et al.* Hsp70 Forms Antiparallel Dimers Stabilized by Post-translational Modifications to Position Clients for Transfer to Hsp90. *Cell Rep.* **11,** 759–769 (2015).

69. Schmidt, C., Beilsten-Edmands, V. & Robinson, C. V. The joining of the Hsp90 and Hsp70 chaperone cycles yields transient interactions and stable intermediates – insights from mass spectrometry. *Oncotarget* **6,** 18276–18281 (2015).

70. Melkani, G. C., Zardeneta, G. & Mendoza, J. A. On the chaperonin activity of GroEL at heat-shock temperature. *Int. J. Biochem. Cell Biol.* **37,** 1375–85 (2005).

71. Rostom, A. A. & Robinson, C. V. Detection of the Intact GroEL Chaperonin Assembly by Mass Spectrometry. *J. Am. Chem. Soc.* **121,** 4718–4719 (1999).

72. Zhou, M., Jones, C. M. & Wysocki, V. H. Dissecting the large noncovalent protein

complex GroEL with surface-induced dissociation and ion mobility-mass spectrometry. *Anal. Chem.* **85,** 8262–7 (2013).

73. van Duijn, E. *et al.* Monitoring macromolecular complexes involved in the chaperonin-assisted protein folding cycle by mass spectrometry. *Nat. Methods* **2,** 371–376 (2005).

74. Laemmli, U. K., Beguin, F. & Gujer-Kellenberger, G. A factor preventing the major head protein of bacteriophage T4 from random aggregation. *J. Mol. Biol.* **47,** 69–85 (1970).

75. van Duijn, E., Heck, A. J. R. & van der Vies, S. M. Inter-ring communication allows the GroEL chaperonin complex to distinguish between different substrates. *Protein Sci.* **16,** 956–65 (2007).

76. van Duijn, E. *et al.* Tandem mass spectrometry of intact GroEL-substrate complexes reveals substrate-specific conformational changes in the trans ring. *J. Am. Chem. Soc.* **128,** 4694–702 (2006).

77. van Duijn, E., Barendregt, A., Synowsky, S., Versluis, C. & Heck, A. J. R. Chaperonin complexes monitored by ion mobility mass spectrometry. *J. Am. Chem. Soc.* **131,** 1452–9 (2009).

78. Dyachenko, A., Gruber, R., Shimon, L., Horovitz, A. & Sharon, M. Allosteric mechanisms can be distinguished using structural mass spectrometry. *Proc. Natl. Acad. Sci. U. S. A.* **110,** 7235–9 (2013).

79. Sharon, M. & Horovitz, A. Probing allosteric mechanisms using native mass spectrometry.

Curr. Opin. Struct. Biol. **34,** 7–16 (2015).

80. Fan, M. *et al.* The unusual mycobacterial chaperonins: evidence for in vivo oligomerization and specialization of function. *Mol. Microbiol.* **85,** 934–44 (2012).

81. Tsai, Y.-C. C., Mueller-Cajar, O., Saschenbrecker, S., Hartl, F. U. & Hayer-Hartl, M. Chaperonin cofactors, Cpn10 and Cpn20, of green algae and plants function as hetero-oligomeric ring complexes. *J. Biol. Chem.* **287,** 20471–81 (2012).

82. Muñoz, I. G. *et al.* Crystal structure of the open conformation of the mammalian chaperonin CCT in complex with tubulin. *Nat. Struct. Mol. Biol.* **18,** 14–9 (2011).

83. Kapatai, G. *et al.* All three chaperonin genes in the archaeon Haloferax volcanii are individually dispensable. *Mol. Microbiol.* **61,** 1583–97 (2006).

84. Finley, D. Recognition and Processing of Ubiquitin-Protein Conjugates by the Proteasome. *Annu. Rev. Biochem.* **78,** 477–513 (2009).

85. Sharon, M., Taverner, T., Ambroggio, X. I., Deshaies, R. J. & Robinson, C. V. Structural Organization of the 19S Proteasome Lid: Insights from MS of Intact Complexes. *PLoS Biol.* **4,** e267 (2006).

86. Lasker, K. *et al.* Molecular architecture of the 26S proteasome holocomplex determined by an integrative approach. *Proc. Natl. Acad. Sci.* **109,** 1380–1387 (2012).

87. Sakata, E. *et al.* The Catalytic Activity of Ubp6 Enhances Maturation of the Proteasomal Regulatory Particle. *Mol. Cell* **42,** 637–649 (2011).

88. Politis, A. *et al.* A mass spectrometry–based hybrid method for structural modeling of protein complexes. *Nat. Methods* **11,** 403–406 (2014).

89. Politis, A. *et al.* Topological Models of Heteromeric Protein Assemblies from Mass Spectrometry: Application to the Yeast eIF3:eIF5 Complex. *Chem. Biol.* **22,** 117–128 (2015).

90. Sharon, M., Witt, S., Glasmacher, E., Baumeister, W. & Robinson, C. V. Mass Spectrometry Reveals the Missing Links in the Assembly Pathway of the Bacterial 20 S Proteasome. *J. Biol. Chem.* **282,** 18448–18457 (2007).

91. Ma, X., Loo, J. A. & Wysocki, V. H. Surface induced dissociation yields substructure of Methanosarcina thermophila 20S proteasome complexes. *Int. J. Mass Spectrom.* **377,** 201–204 (2015).

92. Dyson, H. J. & Wright, P. E. Intrinsically unstructured proteins and their functions. *Nat. Rev. Mol. Cell Biol.* **6,** 197–208 (2005).

93. Tsvetkov, P. *et al.* Thermo-resistant intrinsically disordered proteins are efficient 20S proteasome substrates. *Mol. Biosyst.* **8,** 368–73 (2012).

94. Moscovitz, O. *et al.* A mutually inhibitory feedback loop between the 20S proteasome and its regulator, NQO1. *Mol. Cell* **47,** 76–86 (2012).

95. Dinkova-Kostova, A. T. & Talalay, P. NAD(P)H:quinone acceptor oxidoreductase 1 (NQO1), a multifunctional antioxidant enzyme and exceptionally versatile cytoprotector.

Arch. Biochem. Biophys. **501,** 116–23 (2010).

96. Moscovitz, O. *et al.* The Parkinson's-associated protein DJ-1 regulates the 20S proteasome. *Nat. Commun.* **6,** 6609 (2015).

97. Andersson, F. I. *et al.* Structure and Function of a Novel Type of ATP-dependent Clp Protease. *J. Biol. Chem.* **284,** 13519–13532 (2009).

98. Mikhailov, V. A., Ståhlberg, F., Clarke, A. K. & Robinson, C. V. Dual stoichiometry and subunit organization in the ClpP1/P2 protease from the cyanobacterium Synechococcus elongatus. *J. Struct. Biol.* **192,** 519–527 (2015).

99. Rosati, S., Yang, Y., Barendregt, A. & Heck, A. J. R. Detailed mass analysis of structural heterogeneity in monoclonal antibodies using native mass spectrometry. *Nat. Protoc.* **9,** 967–76 (2014).

100. Allison, T. M. *et al.* Quantifying the stabilizing effects of protein–ligand interactions in the gas phase. *Nat. Commun.* **6,** 8551 (2015).

101. Sivalingam, G. N., Yan, J., Sahota, H. & Thalassinos, K. Amphitrite: A program for processing travelling wave ion mobility mass spectrometry data. *Int. J. Mass Spectrom.* **345-347,** 54–62 (2013).

102. Marklund, E. G., Degiacomi, M. T., Robinson, C. V., Baldwin, A. J. & Benesch, J. L. P. Collision Cross Sections for Structural Proteomics. *Structure* **23,** 791–799 (2015).

103. Lopez-Otin, C., Blasco, M. A., Partridge, L., Serrano, M. & Kroemer, G. The Hallmarks

of Aging. *Cell* **153,** 1194–217 (2013).

Chapter 3: Structural Insights into HSP27

1 Introduction

1.1 The sHSP HSP27

HSP27 (HSPB1) is expressed throughout the body[1] and performs a range of roles including modulation of the cytoskeleton[2,3], as an anti-apoptotic and anti-necrotic[4] agent, and as an effective molecular chaperone[1,5–7]. It is implicated in a range of disorders including neurodegenerative disease[6], motor neuropathies[8–10], and has received particular attention in light of its anti-apoptotic role which can mediate tumour growth and cancer therapy resistance[4]. While affecting widely varied substrates and cell type, each of these disease mechanisms may be directly or indirectly related to the chaperone function of this protein[11], interacting with a range of client proteins[7] to moderate substrate stability and degradation throughout the body.

Full-length HSP27 has the canonical sHSP domain structure of a structurally heterogeneous N-terminus followed by the ACD and finally a flexible C-terminus. The N-terminus carries three main sites that can be phosphorylated in HSP27: S15, S78 and S82. The N-terminus of full-length, monodisperse sHSPs is often unresolved in their crystal structures[12–14], leading to the suggestion that it is partly intrinsically disordered or highly dynamic and accessing a

number of states[15,16]. The ACD comprises seven β-strands in two β-sheets (as observed in the crystal structure of the core domain c27-1[17]), and also a cysteine at position 137 that can result in disulphide-locked dimers[18] by formation of a disulphide bond across the β6+7 intra-dimer interface. While the C-terminus is again highly flexible, the IPV motif is able to bind into a groove formed by the β4 – β8 groove in the core domain[17], and by analogy with other sHSPs it is purported to mediate binding to neighbouring subunits in the higher order oligomers[13,19]. The two sections of the C-terminus on either side of the IPV motif are termed the 'tail' and 'extension'[16,20].

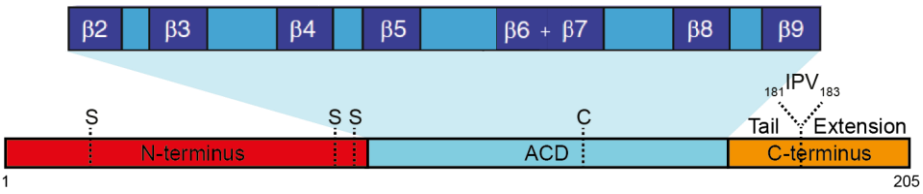


Figure 3 – 1 The sequence characteristics of FL HSP27. The N-terminus carries three phosphosites at S15, S78 and S82. The ACD comprises seven β-strands, with a cysteine residue at position 137 on the extended β6+7 strand. The C-terminus then carries the conserved 'IXI' motif of IPV, which separates the tail from the extension.

Phosphorylation of HSP27 at sites 15, 78 and 82 is mediated by MAPKAP-kinase 2 and 3[21,22] (mitogen-activated-protein-kinase-activated kinase) which are likely to be the responsible kinases *in vivo*[23]. Phosphorylation is induced under stress at these sites[5] and causes an interruption of oligomerisation such that there is a prevalence of low molecular weight species[24,25], postulated to be tetramers[21] or dimers[24], or a mixture of the two[15]. While an imperfect proxy[24] for the phosphorylated form, serine-to-aspartate mutations at residues 15, 78 and 82 to mimic one of the negative charges of the phosphoryl group have proven a useful tool as a mimic of the phosphorylated form. They provide a fully pseudo-phosphorylated form of the protein, and allow systematic interrogation of the contribution of each individual site[21,26].

This is not possible with *in vitro* phosphorylation with the MAPKAP kinases, which phosphorylate all three sites simultaneously and with variable efficiencies[21]. The focus of previous mass spectrometry studies on the effect of sHSP phosphorylation have been phosphomimics of FL αB-crystallin[27,28], which accessed a greater number of odd stoichiometries than the WT, such that the triple phosphomimic had very little of the preference for even stoichiometries exhibited by WT αB-crystallin[28,29].

1.2 Existing structural insight on HSP27 and the mammalian sHSPs

Despite decades of study, our understanding of the mechanisms of action of sHSPs remains limited, for example the mechanisms that mediate the sHSP response to stress, their interaction with their substrates, and their regulation in the cell.

The plasticity and particularly the polydispersity of the mammalian sHSPs has rendered them refractory to protein crystallisation, such that there are not any full length models of the mammalian proteins. Additionally, the often large molecular weight of the oligomers has hindered determination of atomic-resolution models by NMR. The full-length atomic-resolution models that do exist are restricted to the non-metazoan sHSPs which populate a small number of distinct stoichiometries or are monodisperse[12,13,30-32]. Mammalian αB-crystallin and HSP27 demonstrate a particularly large oligomeric distribution with up to 10 - 40 subunits for αB-crystallin[33,34] and 6 – 30 for HSP27 with substantial dimer and tetramer populations[26,35].

This is demonstrated in Figure 3 – 2, which shows full-length HSP27 at 20 µM. The spectrum shows a series of peaks between 5000 – 17000 Th. While at first it appears as if the peaks are poorly resolved, this spectrum actually occurs from the overlap of the charge series of many different species[34], corresponding to a range of stoichiometries with sequential addition of a dimer subunit[26,35].

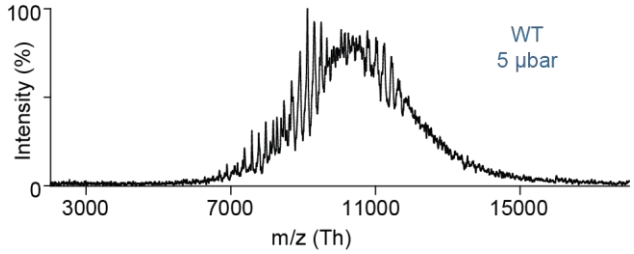


Figure 3 – 2 Mass spectra of full-length WT HSP27 at 20 µM. At a 5 µbar pressure in the initial ion focusing guides and quadrupole, and an acceleration voltage of 80 V into the collision cell, a range of larger oligomer stoichiometries are detected.

One approach to overcome the problem of polydispersity has had significant success. It involves expression of the core domain of the small heat shock protein alone, and its subsequent analysis. Core domain constructs typically comprise the α-crystallin domain with a varying but small number of the flanking residues of the N- and C-termini. The crystallisation of a core domain was first reported in 2009 for cABC and rat HSP20[36], which demonstrated that the ACD of αB-crystallin comprised seven β-strands in two β-sheets, and that a dimer was formed between two ACDs through the formation of an extended β-sheet between the β6+7 interface, as suggested by NMR[37]. The β2 strand of cABC was only resolved in one monomer of the dimer, where it formed part of the four stranded β-sheet (β2, β3, β8 and β9, called the 'minor' sheet herein), implying that β2 was disordered in the second dimer[36]. This was in keeping with the solution-state NMR data on another cABC construct which predicted

the β2 to be disordered in solution, and solid-state NMR of the FL αB-crystallin which observed several different environments for this region[37].

Preparing core domain constructs, either by design[37] or by identification among proteolytically stable fragments[38,39], has since become an established approach for crystallisation[18,20,40,41] and NMR interrogation[42,43] of metazoan sHSPs. cABC has received particular attention, in part due to αB-crystallin's established position as a paradigm for mammalian sHSPs[42] but also because of the implication of αB-crystallin itself in a multitude of disease states[6,44]. Crystallisation of cABC with a greater section of the C-terminus incorporated, such that it contained the IPI motif of αB-crystallin, demonstrated binding of the C-terminus via this motif in the β4-8 groove[20], and also captured a different relative register of the two monomers at the intra-dimer interface (termed the 'AP' antiparallel registration)[20]. A third register was later captured[17] indicating the possible plasticity at this interface, as well as variability in the binding of the C-terminus to the ACD, exemplified in this structure through the co-crystallisation with a C-terminal peptide ('CTP') mimicking the area surrounding the IPI motif[17]. Binding of the CTP was in the opposite conformation to the binding of the cABC C-terminus observed previously[20]. This co-crystallisation strategy not only facilitates crystallisation but has thus offered further illumination on the interaction of the core domain with the C-terminus, which is important for oligomerisation.

The core domain of human HSP27 has already permitted two structures to be proposed from X-ray crystallography. The first, published by Baranova, Weeks *et al.* in 2011, revealed that the core domain of HSP27 exhibited a very similar monomeric structure to cABC, with two β-

sheets folding in an IgG-like β-sandwich[18]. However, instead of a dimer formed by association between the two β6+7 strands, the authors observed the protein in a hexameric β-barrel-like ring with non-canonical interaction between the β4 and β7 strands. Using SEC and SAXS, they showed that the protein was largely dimeric in solution and so classified the β4-β7 interaction as an artefact of crystallisation. In 2014 Hochberg, Ecroyd *et al.* successfully crystallised an HSP27 core domain in a dimeric asymmetric unit, with the intra-dimer interface between the β6+7 strands on each monomer[17]. This was with the c27-1 construct used in the chaperone assay shown in Chapter 1. The crystallisation had been undertaken with 1 mM DTT and produced a similar monomeric and dimeric structure to cABC, with the IgG-like β-sandwich, β6+7 intra-dimer interface and the binding of a synthetic peptide mimicking the C-terminus in the same β4 – β8 binding groove as observed for cABC[17]. The core domain was crystallised in the APII register at the intra-dimer interface, similar to the first cABC structure[36]. The most significant differences were the different crystal packing; less curvature in the extended β-sheet for c27-1; a larger distance between the two β-sheets in the c27-1 monomer due to residues that were bulkier or interrupted a salt bridge; and the change of a charge network across the upper β-sheet (β4, 5, 6+7) to a hydrogen-bonding network with the c27-1 disulphides in a position capable of disulphide bond formation across the intra-dimer interface.

In order to interrogate the different registers of cABC and the role of monomers versus dimers in its chaperone activity, the authors introduced a disulphide to cABC (E117C) which was consistent with the intra-dimer disulphide found in c27-1. In so doing, they indirectly probed the possible role of this disulphide in c27-1, albeit it in the artificial context of cABC.

A third structure of the core domain of HSP27 has recently been proposed from solution NMR restraints[43]. The ten-member ensemble again exhibited the canonical monomer and dimer form, but differed from the c27-1 structure above with a disordered N-terminal end (rather than forming a β2 in the minor sheet as observed for c27-1) modelled on resonances from three residues within the thirteen-residue stretch as peaks for all other residues in this region could not be resolved. Additionally, the ten models of the ensemble showed a greater curvature of the extended β-sheet than the c27-1 crystal structure, reminiscent of a variable degree of curvature observed for the same sheet in cABC[17,41,45]. Finally, the authors collected the model restraints on oxidised c27, and so were able to model the disulphide between the C137 residues (Figure 3 – 1) on either β6+7 strand of the intra-dimer interface, which can only be formed in the APII register.

The influence of this disulphide bond has been discussed in the literature[5,43,46], and its characteristics are unusual – it resides on hydrogen-bonded residues within an anti-parallel β-sheet that links two different subunits ('cross-strand')[47], suggestive of a role as a 'redox switch'. This is particularly notable because of the protective action of HSP27 under oxidative stress[21,48–50]. The progress in the structural studies of c27 detailed above have highlighted the need to investigate the natural disulphide in human HSP27. As our first step, we thus undertook crystal trials in air oxidizing conditions with the aim of solving a structure of the disulphide-bound dimer by X-ray diffraction. This would complement the progress achieved in the previous crystal structures by capturing the structural implications of disulphide formation, and the solution NMR models by producing atomic coordinates of the oxidised protein with greater precision. We used five different human HSP27 fragments, including a

full-length triple phosphomimic and two core domain constructs, in the hope of obtaining the structure of a longer stretch of HSP27 or with higher resolution. This chapter details these studies, conducted in the hope of shedding further light on the structure-function relationship of this major human sHSP.

2 Results and Discussion

2.1 In search of a full-length model of oxidised HSP27

The possibility of gaining structural insight on full-length mammalian sHSPs, which is so elusive because of their polydisperse nature[44], always incites an immediate response. When an increasing number of S82D, S78D and finally S15D phosphomimicking mutations were introduced to HSP27, there was a progressive reduction in the average molecular weight of species formed, such that the oxidised triple phosphomimic ('3D') appeared to form dimers alone in the native mass spectrum[26]. This shift in assembly equilibrium towards low molecular weight species was consistent with previous studies[21,24] and the low molecular weight species could be identified as the dimer with confidence due to the high mass resolution of mass spectrometry. This report thus invited the possibility of crystallising the full-length protein in the oxidised state. If there was a sufficient population of dimers to co-precipitate, this would not only act as a nucleation scaffold, but as the crystals grow the equilibrium between dimers and a small population of higher order oligomers would be constantly re-established, thus reducing the overall population of higher order oligomers. As HSP27 3D was observed to form monomers and non-covalent dimers under reducing conditions[26], crystallisation conditions

under air oxidation would not only provide the possibility of capturing the intra-dimer disulphide as discussed in the Introduction, it would also likely facilitate crystallisation by reducing plasticity at the dimer interface.

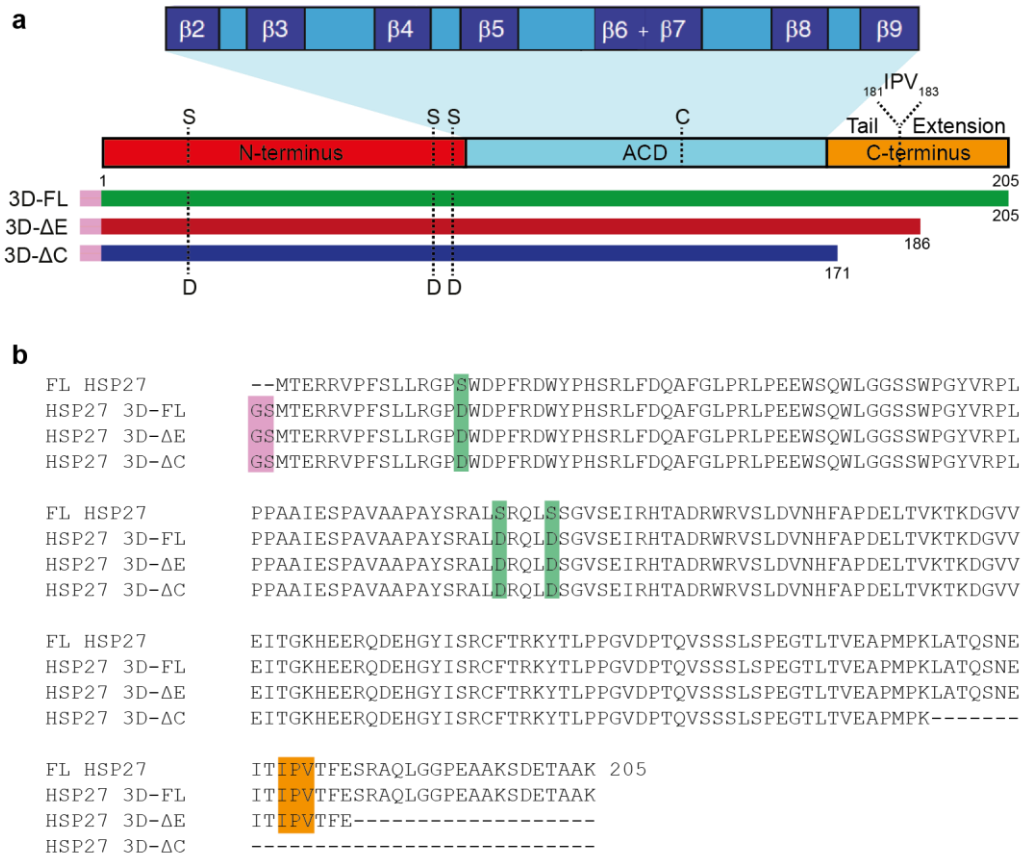


Figure 3 – 3 The HSP27 constructs used in this section. a) Schematic highlighting sequence features. b) Sequence of the final protein after tag cleavage. The serine to aspartate residues are highlighted in green; the IPV motif in yellow. 3D-FL comprises the whole of the full-length (FL) sequence; 3D-ΔE is lacking the flexible C-terminal extension after E186, but still incorporates the residues involved in hydrogen bonding of the C-terminus to the ACD[17]; 3D-ΔC lacks the majority of the C-terminus (upto K171 is included due to the resolution of this in a core domain structure[17]). Each construct was recombinantly expressed in a modified pET28a 6H-TEV-//HSP27// and a 6H-MBP-TEV-//HSP27// vector for improved yield and so retains a GS motif on the N-terminus after purification, highlighted in pink. TEV – tobacco etch virus protease cleavage site; MBP – maltose binding protein.

We therefore designed three constructs, including the full-length (FL) protein, which incorporated the S15D, S78D and S82D mutations (Figure 3 – 3). We used an N-terminal His-tag and TEV cleavage site for recombinant expression of all the proteins, and purified them

according to Section 4.3. 3D-ΔE readily aggregated on purification, and so we performed crystallisation trials on 3D-FL and 3D-ΔC.

We also performed MS on the species to confirm their oligomeric distribution. The mass spectrum for FL HSP27-3D shows multiple overlapping peaks in the region 5000 – 17000 Th (Figure 3 – 4) which resembles the spectrum of HSP27-WT at the same concentration (Figure 3 – 2) and is indicative of a range of high order stoichiometries with overlapping charge states[34]. This somewhat surprising result is detected when instrument parameters that favour the transmission of large protein complexes are employed (namely a higher pressure in the ion guide (a radio frequency multipole) and the initial chambers of the mass spectrometer, which helps to collisionally focus the large oligomeric species such that their radial momentum is diminished[51], and a higher acceleration voltage through the collision cell). This revealed that while the triple phosphomimic has a greater relative population of dimers than the WT, there is still a significant population of the higher-order oligomers present within the sample at 20 µM monomeric concentration[24,26] (Figure 3 – 4). Similar behaviour was also observed for the 3D-ΔC construct. The population of the higher order oligomers was thus higher than originally reported by native MS, and likely contributed to the failure of crystallisation trials, as the highly polydisperse population would impede formation of ordered crystals of the dimer. Indeed, even if the protein had accessed the dimeric form alone, the flexible nature of the N-terminus in particular may have hindered crystallisation[39].

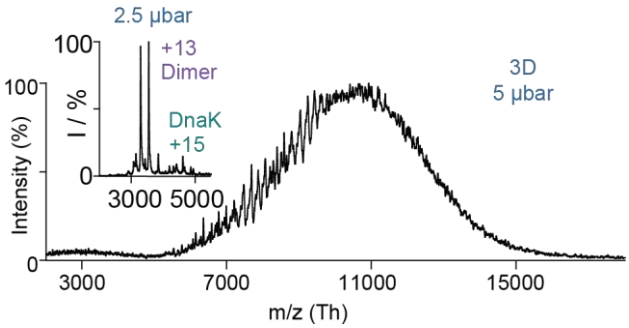


Figure 3 – 4 Spectrum of FL HSP27-3D at 20 µM. While it can appear that only low mass species are present in the sample when a low source pressure is used (2.5 µbar, 20V collision voltage, inset), on raising the pressure to 5 µbar to favour the transmission of larger oligomers (and 80 V collision voltage), the presence of a range of high order oligomers becomes apparent. Low mass species *ca.* 46129 Da, consistent with the dimer; 69161 Da, likely E. Coli DnaK, which is often co-purified in small amounts due to the interaction and chaperone activity of these proteins.

Nonetheless, this study proved valuable as it had led us to detect the higher order species in the phosphomimicking construct, supporting observations of a concentration-dependent bias towards the small multimers[24], and refuting occupation of the small stoichiometries exclusively[52]. HSP27-3D thus populates the dimer to a higher degree than the WT. This is supported by the observation of only dimers (and DnaK) at low source pressures for HSP27-3D (Figure 3 – 4 inset), whereas at the same pressure for HSP27-WT the population of large oligomers can still be detected alongside dimers, albeit with a diminished abundance. This implies that the relative abundance of dimers is much higher in the HSP27-3D sample than for the WT, showing the triple mutation to mimic phosphorylation shifts the oligomer equilibrium towards the dimeric species[26].

2.2 A new structural model for the core domain of HSP27

Our next strategy in order to crystallise the disulphide-bound dimer was to design two new constructs comprising the core domain of HSP27 (c27-2 and c27-3, Figure 3 – 5a). As

discussed in the Introduction, this approach has had significant success in elucidating atomic-resolution data for multiple human sHSPs[18,20,36,39] (Figure 3 – 5b). We also employed co-crystallisation with a synthetic peptide mimicking the C-terminus (EITIPVTFE$_{178\text{-}186}$, 'CTP' Figure 3 – 5), as this approach has previously provided additional structural insight and facilitated crystallisation[17]. The constructs were shorter than the c27-1 construct described in Chapter 1 by either five or six residues on the C-terminus (Figure 3 – 5). The omitted residues had not been resolved in either chain A or chain B of the previous structure of the c27-1 dimer[17] (Figure 3 – 5b) so their removal may have facilitated better order in the crystal, and so greater resolution.

2.2.1 Optimising crystallisation of the core domain

The c27-2 and c27-3 constructs were introduced to a modified pET28a vector with an introduced N-terminal histidine tag and TEV protease cleavage site. They were expressed and purified as described in the Methods section, with immobilised metal affinity chromatography and TEV cleavage under reducing conditions (5 mM β-mercaptoethanol (BME)), followed by size exclusion chromatography under air oxidation. Initial co-crystallisation trials yielded numerous small plate, rod and micro-crystals after 2.5 weeks for both constructs, with single crystals developing after 3.5 weeks. On raising the concentration of protein, c27-2 produced single tetragonal crystals and c27-2 produced single hexagonal crystals after twelve days. Crystallisation was thus optimised by hand, as detailed in Section 4.4 according to the conditions in Table 3 - 5 therein. c27-2 produced numerous crystals large enough for harvest which were used in the X-ray diffraction experiment using synchrotron radiation.

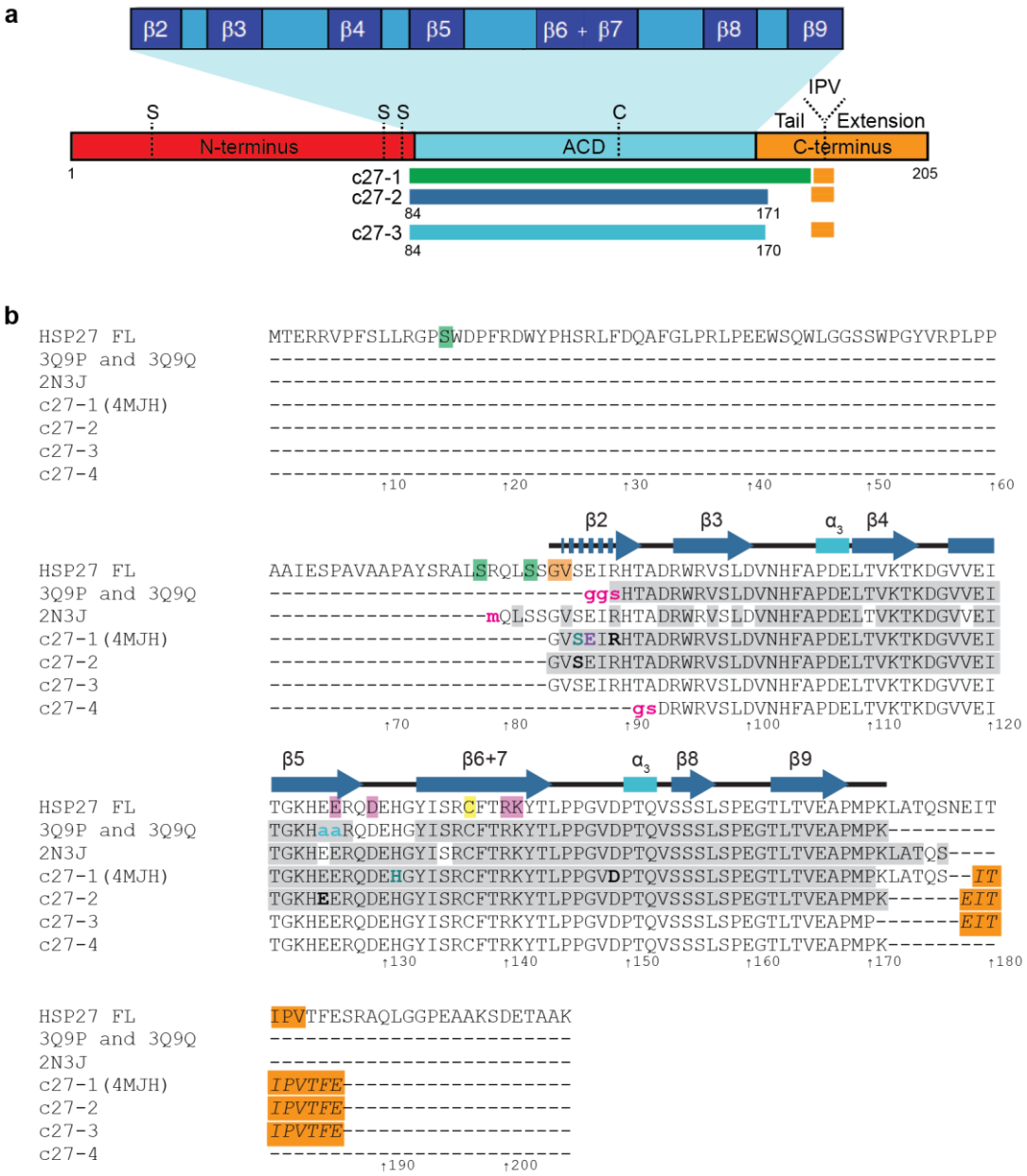


Figure 3 – 5 Core domain constructs of HSP27. a) Schematic showing the core domain constructs described in Chapter 1 (c27-1) and in the crystallisation trials presented here (c27-2, c27-3). b) Sequence alignment of HSP27 FL (UniProt ID: P04792) followed by the ACD constructs of the previously published HSP27 structures (PDB IDs: 4MJH[17], and 3Q9P and 3Q9Q[18] from protein crystallography; 2N3J from NMR[43]), and the constructs in (a) along with a fourth form used for ongoing solution studies (c27-1 is the same construct as that in 4MJH). Residue numbers for HSP27 FL are shown. Secondary structure (arrow – β-strand; box – 3_{10} α-helix) was assigned using DSSP[68,69] on the chain A coordinates of the c27-2 model presented herein, except for the β2 strand which was disordered but had previously been observed bound to the β3 strand (in c27-1(4MJH)[17], dashed arrow). Non-native residues remaining after tag cleavage (lower case, pink) or substituted to aid crystallisation (lower case, light blue) are shown. (*Legend continues on next page.*)

2.2.2 The oxidised core domain of human HSP27 displays multiple interfaces

The best crystal of c27-2 diffracted to 2.25 Å (Table 3 - 6 in Section 4.5) and data processing indicated a $P4_32_12$ tetragonal space group. This is an improvement on 4MJH with resolution 2.6 Å, and the 3Q9P and 3Q9Q which diffracted to 2.0 and 2.2 Å respectively but did not capture the biological dimer interface or the loop between β5 and β6+7. The c27-1 structure (PDB 4MJH) was used for molecular replacement and the final structural model after refinement is presented in Figure 3 – 6.

Figure 3 – 6a shows the packing of the asymmetric unit within the unit cell of the crystal (with dimensions a = 56.06 Å, b = 56.06 Å, c = 166.29 Å, $\alpha = \beta = \gamma = 90°$). The crystal packing is more complex than that of c27-1 (which showed a dimer of dimers), showing runaway behaviour through association of the β2 strand of one chain on one dimer with two other dimers (discussed in more detail below). The asymmetric unit comprises two protein chains, chain A has all 88 residues resolved (84-171, Figure 3 – 5) and chain B has only residues 93-170, with one C-terminal peptide bound to chain A.

(Continuing from previous page.) **Shading highlights the following features: grey – all residues that could be resolved in at least one of the chains for each pdb entry; orange – IPV motif; orange and italics – the synthetic CTPs used for co-crystallisation; light orange – the GV residues observed in c27-2 which bound in the groove normally occupied by the IPV motif of the CTP; green – phosphorylatable serines; yellow – the cysteine at 137 which can form a disulphide across the intra-dimer interface; pink – residues with side chain interactions across the intra-dimer interface (see text). The following residues discussed in Chapters 3 and 4 are in bold text: green – S86 and H131 hydrogen bonding pair between chain A and B in 4MJH which may stabilise the bound conformer of the β2-strand; purple – E87 which forms the last main chain hydrogen bond to be broken between β2 and β3 in MD simulations of the ACD; black – R89 and D149' form the only hydrogen bond due to the crystal lattice that may stabilise the β2-strand in the 4MJH structure. Alignment was performed using the Clustal Omega webserver[70,71].**

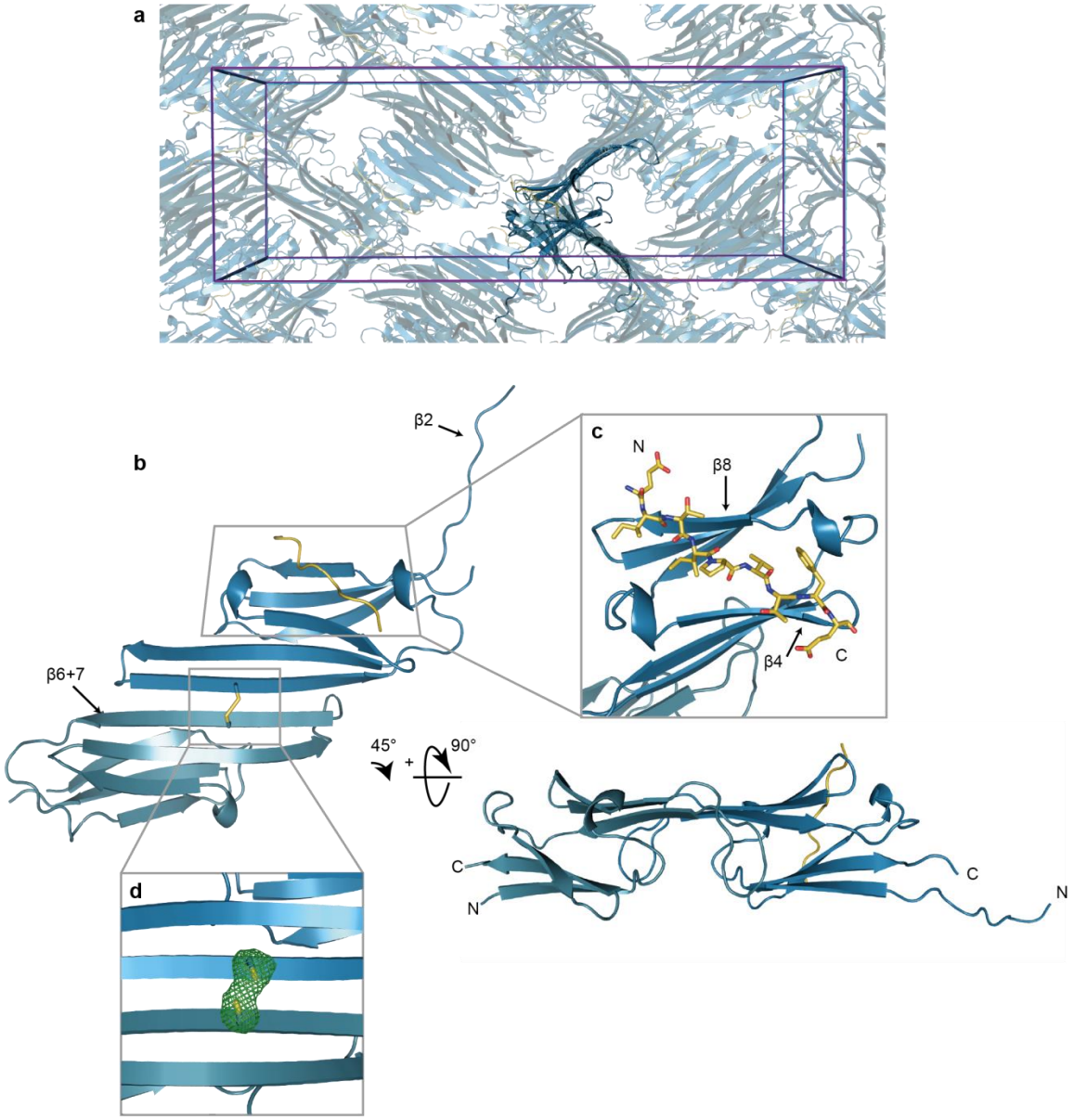


Figure 3 – 6 Crystal structure of c27-2. (a) The crystal packing of c27-2 within the unit cell. The asymmetric unit (dark blue) comprises chain A (foreground), chain B (behind chain A and oriented at approximately 90° to chain A), and the CTP (yellow). (b) The c27-2 dimer is formed through formation of an extended β-sheet with the interface between a β6+7 strand on each monomer. Rotation of the dimer shows the β-sandwich fold in each monomer, which come together to form the extended β-sheet and a shared groove between the two minor β-sheets. (c) The peptide ligand is found binding to the β4 – β8 groove on chain A, in an anti-parallel direction to β8. (d) An F_o-F_c omit map produced when the cysteine side chains are removed from the model (green), overlaid on the final model (blue). The unaccounted positive electron density is consistent with the cysteine side chains pointing towards one another, such that a disulphide can form between the sulphur atoms (yellow). The map is displayed at a contour level of 0.1838 e.Å^{-3}, which corresponds to 2.97 σ in Coot, within 2.2 Å of the sulphur and C_β of the cysteine sidechains in the final model. N and C denote the N- and C-termini of the protein or peptide chain.

The biological unit can be constructed from the asymmetric unit using symmetry operator (–x

+ 1/2, y + 1/2, -z + 3/4) on chain B (Figure 3 – 6b). Similarly to c27-1, c27-2 was found to form the canonical IgG-like β-sandwich fold within each monomer, with an intersubunit β-sheet formed between the β6+7 strands of the two monomers and the C137 opposite one another on this interface showing that it was in the antiparallel II registration[20] (APII). Notably, the thiol groups of the cysteine side chains were rightly oriented to form a disulphide bond in the model that best fit the electron density with significant electron density between the sulphur atoms. While there may be a very small population of the cysteines in an alternate conformation (for example, with one of the thiols binding a BME molecule as hypothesised for the additional peaks in the mass spectra (Chapter 4)), there was little evidence for this in the electron density. However, the distance between the sulphur atoms of the two cysteines is 2.6 Å, which would have been dismissed in a PDB-wide study of disulphides[47]. The disulphide-bound dimer is clearly formed in solution (Chapter 6) and routinely observed on SDS-PAGE gels on air oxidation[18]. In order to verify whether the disulphide was indeed maintained in the crystal, we constructed an F_o-F_c map using a model where the cysteines were stripped to glycines followed by simulated annealing in the Phenix suite[53] (to remove some of the bias in the electron density associated with the prior refinement cycles using a model in which the thiols were facing one another), followed by a final round of refinement in Refmac[54]. This map can be seen in Figure 3 – 6d at 3σ and demonstrates the disparity in the electron density if the structural model does not include the thiols pointing towards one another. The region of electron density that is unassigned between the two β-sheets in the F_o-F_c map, along with the density between the two sulphurs in the $2F_o$-F_c map

during refinement, led us to conclude that the disulphide was maintained in the crystal (Figure 3 – 6b).

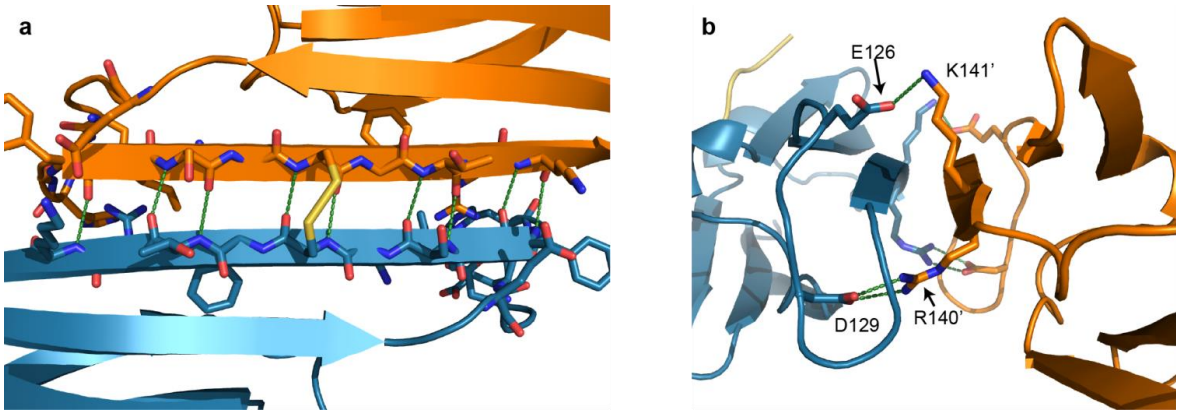


Figure 3 – 7 The intra-dimer interface of c27-2. (a) Main chain hydrogen bonds (green) formed between the residues of chain A (blue) and chain B (orange). (b) A side view of the side chain interactions across the interface, looking down the β6+7 strand of chain A. The concurrent interactions between the R140 and K141 of chain A with the D129' and E126' of chain B are visible in the background.

The APII registration of c27-2 mediates numerous main chain hydrogen bonds and several side chain electrostatic interactions across the dimer interface (Figure 3 – 7 and Chapter 5 Table 5 - 2), as well as the covalent bond between the two cysteines. The side chain interactions involve both sets of K141-E126 and D129-R140 present at the interface (Figure 3 – 7 and Figure 3 – 5b). These are significant because both the K141Q mutation[55] and the sporadic or autosomal dominant R140G[8] have been identified in motor neuropathy. The conserved R140 residue also aligns with R120 in αB-crystallin (R120G is linked to desmin-related myopathy) and R116 in αA-crystallin (associated with dominant cataract)[8,10]. Thus the structure of the WT core domain presented here gives some rationale to the effect of these mutations: mutation of positively charged lysine to neutral glutamine will interrupt the salt bridge formed with glutamate 126; removal of the positively charged arginine side chain on mutation to glycine will similarly nullify its electrostatic interaction with aspartate. The

dimeric interface of HSP27 will thus be weakened and so the dynamics of the full-length oligomers will be modified[29]. Additionally, mutations to S135 and R136, which are in close proximity to the C137 of the disulphide, have also been identified in motor neuropathy[10]. These are discussed in further detail in Chapter 5 and Chapter 6.

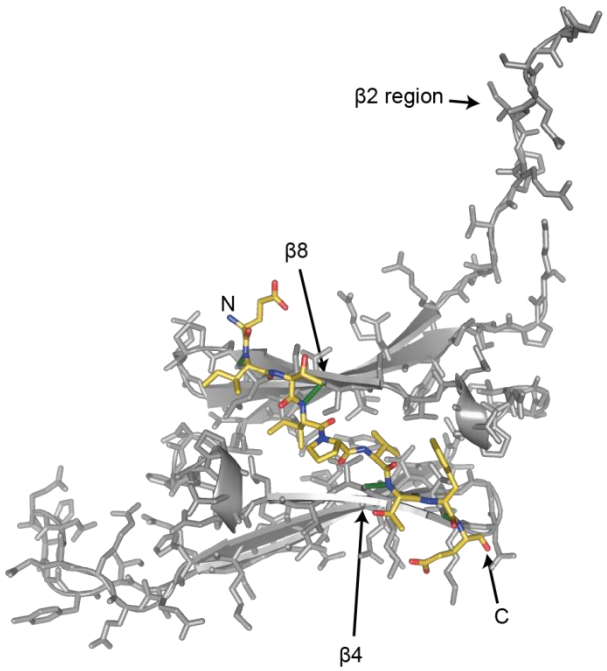


Figure 3 – 8 Binding of the CTP (yellow) to chain A of c27-2 (grey) is mediated by hydrogen bonds with the main chain of β4 and β8 (green lines link hydrogen-bonded pairs) and the burial of the hydrophobic isoleucine and valine side chains within the groove between β4 and β8. Chain A is shown in grey here for clarity, but it is the same chain as that presented in Figure 3 – 6c.

The C-terminal peptide can be found in its canonical binding groove between the β4 and β8 strands of chain A with the IPV motif sitting within the hydrophobic groove and the flanking residues forming hydrogen bonds with the β4 and β8 (Figure 3 – 6c). It forms three main chain hydrogen bonds with β4, and three main chain hydrogen bonds with β8 (via residues I179, I181, T184, E186 on the peptide, shown in greater detail in Figure 3 – 8). It is modelled here in the antiparallel orientation with respect to the β8 strand, but considering the resolution of the data and the palindromic nature of the peptide (EITIPVTFE), it is possible

that both orientations are present in the crystal, as modelled in Figure 3 – 11c (Figure 3 – 11b shows the CTP as presented here). Interestingly, the β4-β8 groove of the B chain is not host to a CTP (perhaps due to the ratio of 2.4 CTP to 1 c27-1 dimer in the crystallisation, which was limited by the solubility of the peptide) but it is not vacant. Rather it is occupied by an unexpected partner.

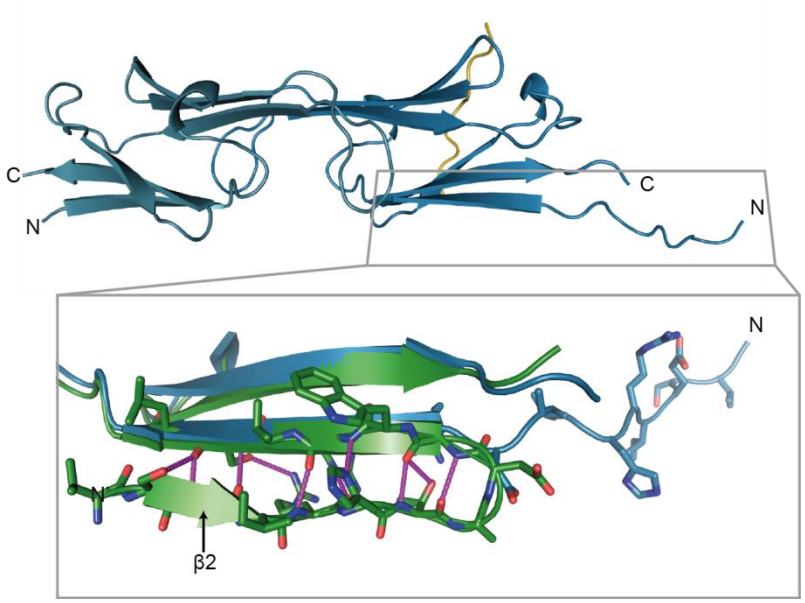


Figure 3 – 9 The β2 of c27-2 chain A has been crystallised in an extended conformation (blue), whereas it forms part of the minor β-sheet in each chain of the structure of c27-1 (4MJH, green) through ten possible hydrogen bonds (dashed magenta batons). N and C denote the termini. The β2 and β3 of c27-1 have been overlaid in stick representation, only the side chains of c27-2 β2 have been shown.

The N-terminal end of c27-2 is highly extended. This is contrary to the structures of cABC[17,20,36] and c27-1 which contain the β2 strand within the β3,9,8-sheet (or unresolved in one chain of the dimer[36]). This 'unravelled' β2 strand of c27-2 is highlighted in Figure 3 – 9 (blue) and contrasted with the c27-1 β2 that loops back to form part of the minor β-sheet (green). Though a random coil, the unravelled β2 strand is resolved. This may be because of its hydrogen bonding with the unravelled β2 strand of another dimer before its further

interaction with the β4-β8 groove of the B chain of a third dimer (Figure 3 – 10). Figure 3 – 10a again shows the crystal packing of c27-2, with the three interaction partners of one chain A (purple) highlighted. These are the chain B of the biological dimer (purple), another chain A with which there is reciprocal hydrogen bonding between the unravelled β2-strands (green), and a third partner in which the N-terminal end of the β2 lies within the β4 – β8 hydrophobic groove of chain B (blue). The corresponding interactions occur for every chain A and chain B in the crystal lattice and mediate the unusual crystal packing noted above.

The positioning of the end of the β2 in the β4 – β8 of chain B is stabilised through the burial of V85 in the hydrophobic groove (where the G84 occupies the typical position of P182 of the C-terminal IPV motif), and three main chain hydrogen bonds (two to β8 residues and one to a residue on β4, Figure 3 – 11a). This can be compared to the binding of the CTP (Figure 3 – 11b), and that of the near-palindromic CTP if modelled in the opposite direction (i.e. parallel to β8, Figure 3 – 11c). We chose the conformer modelled in Figure 3 – 11b as it was more convincing within the $2F_o$-F_c map, but both conformers were possible, in keeping with bi-directional binding of the C-terminus observed for αB-crystallin[17,20]. The CTP binding is stabilised by more hydrogen bonds with the β4 and β8 strands (three hydrogen bonds to each) than the binding of the unravelled β2, and also occludes a greater hydrophobic area. If the c27-2 extended further on the N-terminus, it is possible that the unravelled β2 would form a greater number of hydrogen bond interactions. However, the GV is flanked by a serine on the N-terminal side (*i.e.* it is not an 'IXI' motif, Figure 3 – 5b), so the C-terminus with the IPV motif could be expected to exclude a greater amount of solvent on binding to this hydrophobic groove.

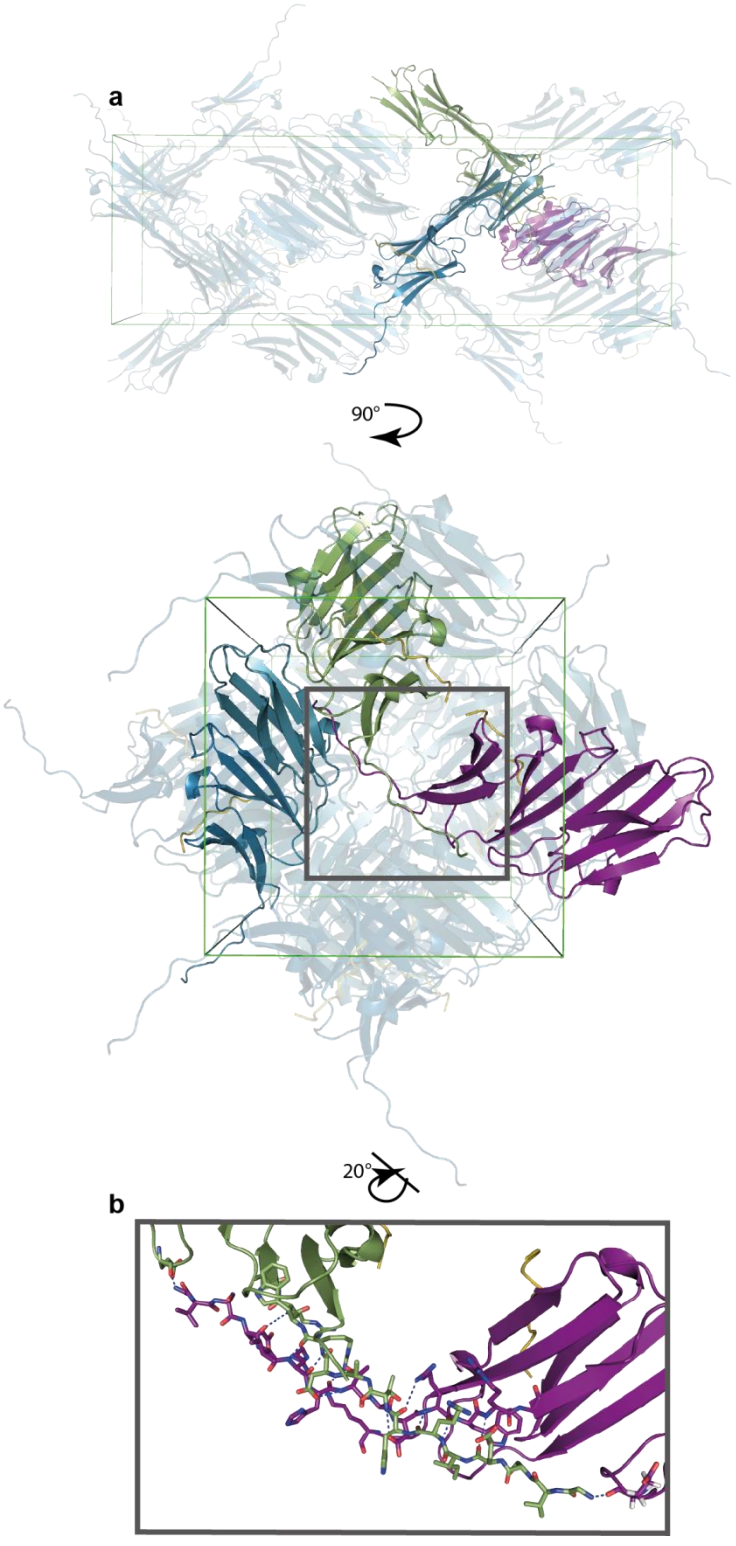


Figure 3 – 10 The unravelled β2 of c27-2 chain A is stabilised by hydrogen-bonding to neighbours in the crystal. (a) Chain A of the dimer (purple) forms reciprocal hydrogen bonds with the chain of another dimer (green) before interacting with chain B of a third dimer (blue). (b) An expanded view showing the hydrogen bonding formed between the two extended β2 strands.

That said, the core domain of human HSPB6 was observed as a tetramer in a crystal

structure[39], in which the binding between dimers was mediated by a similar interaction of the β2-containing region with the β4-β8 groove. There is not competitive binding with the C-terminus as this protein lacks an IXI motif in the C-terminus (Figure 6 – 1), which may allow a more weakly binding β2-region to occupy this site. There, a sole valine or a valine and threonine were seen buried in the hydrophobic groove (in a VPT motif), and mutation of the valine caused HSPB6 to form dimers rather than tetramers in solution[39]. This suggests that the same may be possible with the SGV sequence in the β2 of HSP27. If the β2 - β2 hydrogen bonding and the β2 binding to the β4-β8 groove are present in solution, albeit with a likely lower affinity than the C-terminus, they present a novel mode of binding for polydisperse human small heat shock proteins. Moreover, competitive binding between the C-terminus and β2 region may moderate the oligomeric state and form part of HSP27's response to its environment.

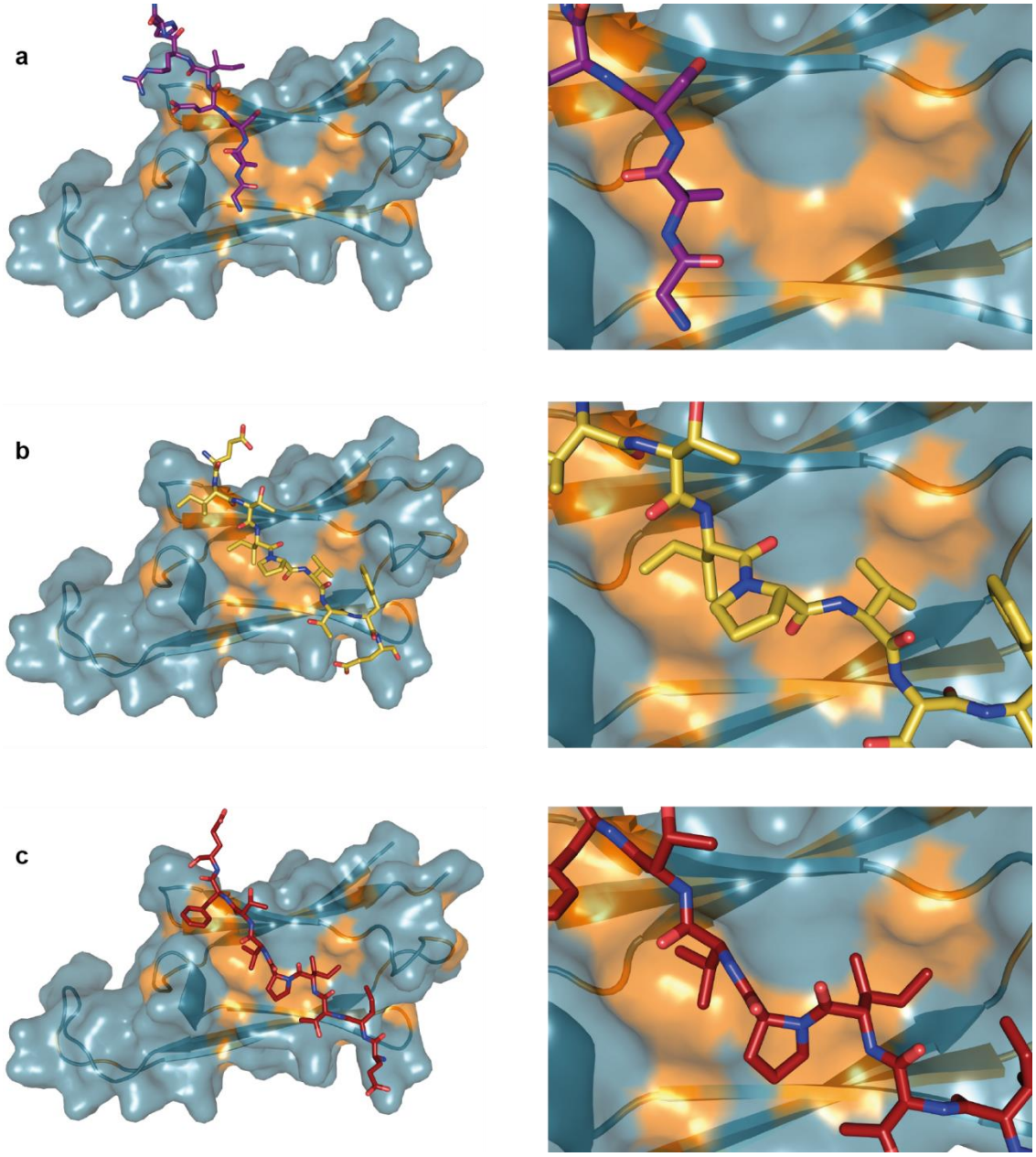


Figure 3 – 11 Multiple modes of binding to the β4 – β8 groove. (a) The hydrophobic groove (orange) between the β4 and β8 strands on chain B is patched by V85 of the unravelled β2 (purple) in c27-2. (b and c) Superposition of chain A with chain B shows the binding of the CTP to chain A also leads to patching of the hydrophobic groove, this time by I181 and V183. (b) The CTP as modelled in the c27-2 structure (yellow). (c) The CTP if modelled in the opposite direction (red) has a valine in the same pocket as in (a). All six panels shown are overlaid on chain B, but as chain A superposes well onto chain B, this is almost identical to the binding seen in the crystal structure (Figure 3 – 8).

2.2.3 Comparison of c27-2 with other structural models

	3Q9P[18]	3Q9Q[18]	4MJH[17]	2N3J[43]	c27-2(WT)
Resolution (Å)	2.0	2.2	2.6	-	2.25
Method	PX	PX	PX	NMR	PX
Crystallisation or solution conditions	2.2M ammonium sulphate, 0.2M di-ammonium tartrate	2.2M ammonium sulphate, 0.2M di-ammonium tartrate	0.1 M Tris, 0.2 M magnesium chloride, 30% PEG 4000	50 mM sodium phosphate, 100 mM sodium chloride, 1 mM PMSF, 0.1 mM EDTA, 90% H_2O / 10% D_2O	0.2M ammonium sulphate, 0.1M sodium cacodylate trihydrate, 22-24 % w/v PEG 8000
pH	5.5	5.5	8.5	7.5	6.5
Temperature (K)	277	277	298	295, 310	293
Protein concentration (mM)				1.0	
V_M (Å^3/Da)	2.95	2.59	1.93		3.16
Solvent content (%)	58	53	36		61

Table 3 - 1 Comparison of HSP27 core domain structures, solved by protein crystallography (PX) or NMR. V_M is the Matthew's coefficient (the unit cell volume divided by the mass of protein (and peptide) in the unit cell[56]) and was calculated using the Matthews_coef tool in CCP4[57], along with the solvent content of the crystal (to 2 s.f.)[56].

In order to compare the c27-2 structure with the three previously published HSP27 core domain structures solved by X-ray diffraction[17,18] or solution NMR[43] (Figure 3 – 5, Table 3 - 1), the RMSD was calculated between the Cαs on the A chains of each of the structures and c27-2. They are listed in Table 3 - 2.

	3Q9P	3Q9Q	4MJH(c27-1)	4MJH(c27-1)
Residues for alignment	90-125, 133-171	90-127, 133-170	85-170	93-170
No. of aligned atoms (No. of possible aligned atoms)	74 (75)†	74 (76)†	86 (86)	78 (78)
All-atom RMSD / Å	2.0	1.9	8.3	0.7

Table 3 - 2 The monomer root mean square deviation to 1 d.p. in the Cαs between the c27-2 structure presented here and the two previously published structures of the HSP27 core domain. Alignment was performed on chain A of c27-2, 3Q9Q and 4MJH. †Alignment of all paired residues is not possible with 3Q9P and 3Q9Q due to multiple conformations of the C137 in 3Q9P and R136 and C137 in 3Q9Q.

The all-atom RMSD between Cαs was calculated using the 'align' command in PyMol which performed a sequence alignment followed by a structural superposition, without any rejection of outliers. The RMSD was then calculated from the distance between the Cαs on the aligned pairs of residues. This was the most appropriate structural alignment tool given that each construct had almost identical sequence identity (excluding a few residues after tag cleavage and the E125A and E126A mutations in 3Q9P and 3Q9Q, Figure 3 – 5). Alignment was performed solely on the Cαs of the residues resolved in chain A of both structures, which varied between the constructs (Figure 3 – 5). The residues used for the alignment are noted in the first row of Table 3 - 2 and Table 3 - 3. The co-crystallised peptides of c27-2 and c27-1 were excluded (but visually overlay well after alignment).

The RMSD is very similar for both crystal forms reported by Baranova, Weeks *et al.* (3Q9P and 3Q9Q). The superposition produced by this pairwise alignment is presented in Figure 3 – 12. A significant difference is that there are 5 - 7 unresolved residues in the loop between the β5 and β6+7 strands which lead to shorter β-strands in this region of 3Q9P and 3Q9Q. This disordered region contained the alanine mutations introduced by the authors to aid crystallisation[38], but in the crystal form presented here the loop is stabilised by interactions

across the dimer interface and crystal contacts. 3Q9P and 3Q9Q also contain the final residues of the β2 in an extended conformation (overlaying well with the β2 of c27-2).The construct was truncated more on the N-terminus than c27-2, starting at H90 (with an N-terminal GSS remaining from recombinant expression, Figure 3 – 5), so this unravelling may have resulted from the incomplete β2 sequence. However, the authors selected this fragment after limited proteolysis[38] of full length HSP27, which indicates that this region is exposed in the full length protein as it was susceptible to proteolysis by trypsin.

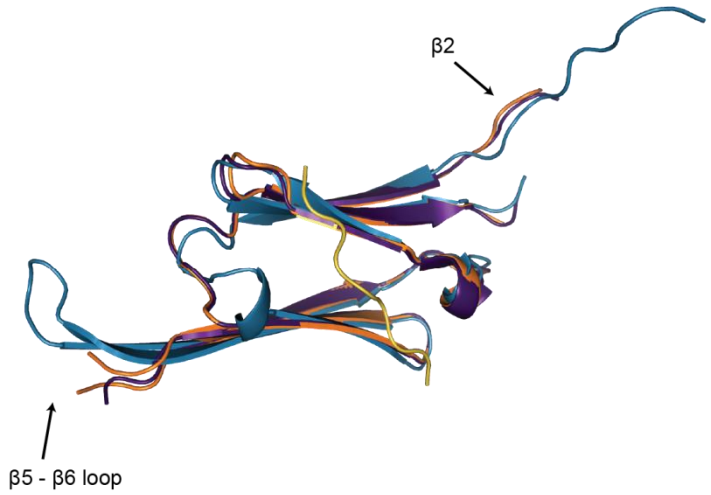


Figure 3 – 12 Superposition of the monomers of c27-2 and the two crystal forms crystallised by Baranova, Weeks *et al.* PDB files 3Q9P (purple) and 3Q9Q (chain A, orange) overlaid with c27-2 chain A (blue).

Alignment of the A chains of c27-2 and c27-1 (4MJH) result in a large all-atom RMSD of 8.3 Å. By far the greatest contribution to this is the β2, which partakes in the minor β-sheet in c27-1 but is in the extended conformation in c27-2. Indeed, when residues 85-92 are excluded from the alignment, the all-atom RMSD decreases to 0.7 Å, indicating a closer agreement of c27-2 with this structure than 3Q9P and 3Q9Q (Table 3 - 2). The main remaining differences are the divergent positions of the loop coming into the β3 (perturbed by the excluded β2) and

the final two C-terminal residues, with some slight variations in the position of the loops and β5. The superposition formed by this pairwise alignment is presented in Figure 3 – 13a. The unravelled β2 of c27-2 is indicated, but the rest of the monomers are very similar, including the degree of bend in the hairpin loop between β5 and β6+7, and the position of the co-crystallised CTPs. The CTPs are in the same orientation (antiparallel to β8) and both have the isoleucine and valines of the IPV buried in the hydrophobic groove.

Overlay of the B chains of c27-2 and c27-1 produces a similar RMSD value (Table 3 - 3) to chain A without the β2, indicating similarity within the β-sandwich of both monomers. As the β2 is not resolved at all in chain B of c27-2 and these calculations were performed only on residues resolved in both species, the RMSD is not affected by multiple conformations of the β2. However the fact that c27-1 has three residues of the β2 resolved in the minor β-sheet (in the same position as in chain A but with more structural heterogeneity as there are fewer residues resolved), but the β2 is not resolved at all in c27-2 chain B, indicates that the β2 is again accessing different conformations in the crystals.

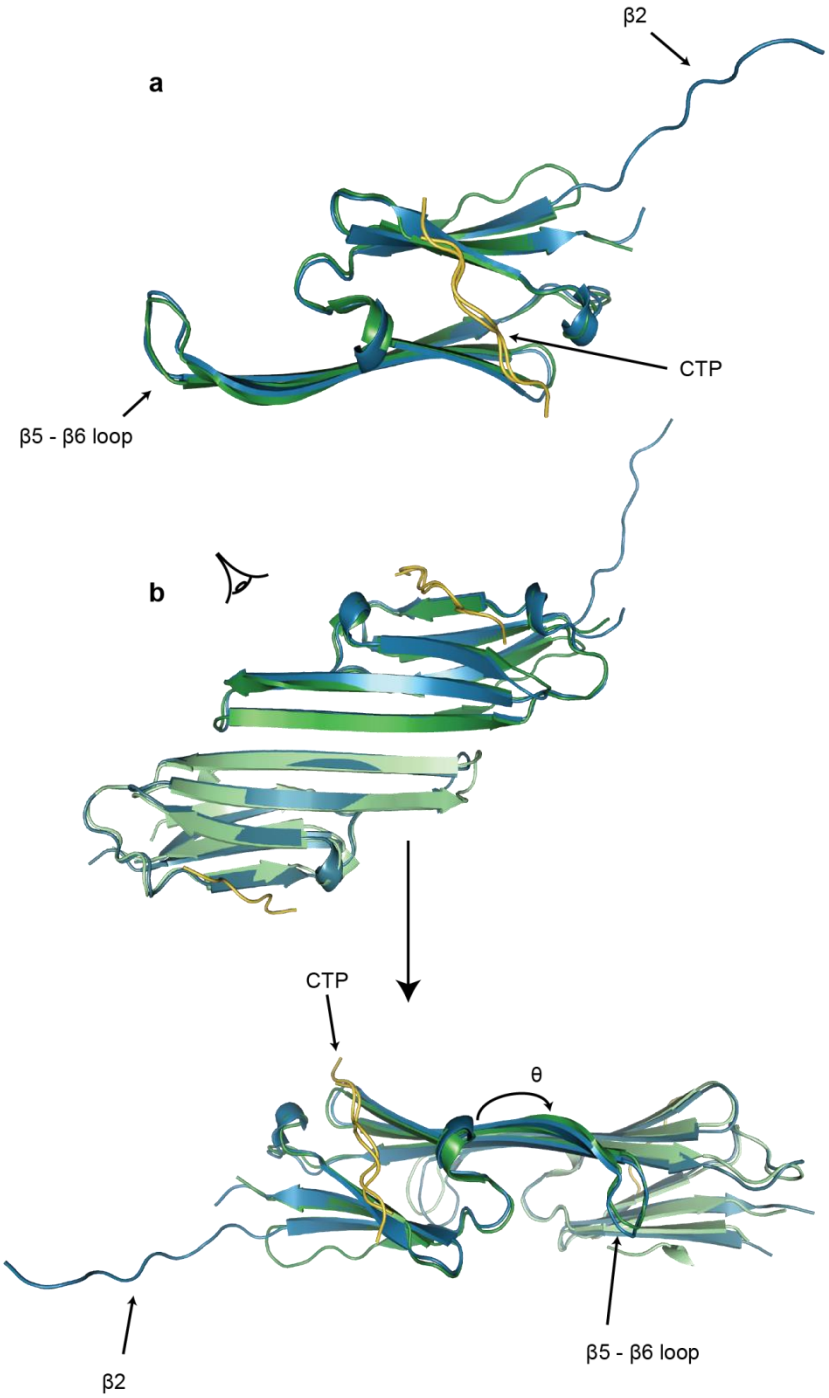


Figure 3 – 13 Superposition of the crystal structures of c27-2 (blue) and c27-1 (PDB 4MJH, green). (a) Superposition of the monomers of chain A shows a very similar structure of the core domain, except for the conformation of the β2 strand. b) Strong overlap of the whole dimer indicates the relative position of the monomers in the dimer is also maintained. The view in the lower panel is from the vantage point indicated in the upper panel. The C-terminal peptides (CTP) of both structures are shown in yellow.

Alignment of both chains within c27-1 and c27-2 (excluding the β2 of chain A), again produces a low value of the non-refined all-atom RMSD (Table 3 - 3), indicating that the

relative orientation of the monomers within the dimer is also maintained within the two crystal forms. The corresponding superposition is shown in Figure 3 – 13b. The relative orientation is particularly clear in the lower panel, where the extended β-sheet formed by the two monomers has the same degree of bend between the two structures and is in keeping with the curvature of a cABC structure solved at pH 9[36]. c27-1 and c27-2 were crystallised at pH 8.5 and pH 6.5 respectively (Table 3 - 1), and their similar curvature along this sheet is in contrast to cABC, for which a varying θ with pH has been detected in the APII form[41].

	Chain A	Chain B	Chains A+B
Residues for alignment	93-170	94-169	93-170, 94-169
No. of aligned atoms (No. of possible aligned atoms)	78 (78)	76 (76)	154 (154)
All-atom RMSD / Å	0.7	0.8	0.9

Table 3 - 3 The root mean square deviation to 1 d.p. between the Cαs in the c27-1 crystal structure (PDB 4MJH) and the c27-2 structure presented here. The values for the pairwise comparison of the Cαs in the A chains are those displayed in the final column of Table 3 - 2, i.e. excluding the β2 strand. Only the residues resolved in both structures were used for the alignment and comparison.

However, in the NMR solution structure, this sheet is more curved[43]. The authors have deposited an ensemble of ten structures (PDB 2N3J), the most notable feature is the flexibility of the N and C-termini in this longer construct (residues 80-176, Figure 3 – 14c). The position of the N-terminus is determined by three residues distributed along the region of residues 80-94 region (Figure 3 – 5b), as resonances for the other residues are not detected, which the authors noted implied lack of structure and exchange with water in this region[43], corroborating our findings.

There is a greater difference between 2N3J and c27-2 than the other structures, probably due to the increased uncertainty inherent in predicting the solution structure from NMR restraints, but also partially affected by the solution structure capturing states of the protein

currently exhibiting dynamic behaviour, rather than confined within a crystal. The all-atom RMSD between the A chain monomer of c27-2 and 2N3J State 5 (one of the most different to c27-2 on visual inspection) is 2.4 Å (Table 3 - 4), with the main contributions arising from the different positions of the loop regions, and slight variation in the relative position of the β-strands (Figure 3 – 14a). β5 is shortened due to a longer β6+7 loop, but similarly to c27-2 this loop bends down in all the structures in the ensemble (Panel (c)). The general features of the core domain are still decidedly similar, including the β-sandwich fold and β6+7 dimer interface. The N-terminus of the construct (residues 80 – 92, including the β2 region of c27-2 at 84 – 92) was excluded from this alignment to allow more representative alignment of the β-sandwich core domain, but is extended in all of the states of the ensemble (green, Figure 3 – 14c).

	State 5 Chain A	State 5 Chains A + B	State 10 Chains A+B
Residues for alignment	93-171	93-171, 93-170	93-171, 93-170
No. of aligned atoms (No. of possible aligned atoms)	79 (79)	157 (157)	157 (157)
All-atom RMSD / Å	2.4	2.6	2.7

Table 3 - 4 The root mean square deviation to 1 d.p. between the Cαs in the HSP27 core domain solution structure (PDB 2N3J) and the c27-2 structure presented here. The values for the pairwise comparison of the Cαs in the A chains excluding the β2 strand. Only the residues resolved in both structures were used for the alignment and comparison (2N3J was longer than the c27-2 structure in all cases).

Alignment of both chains of the dimer results in similar RMSD values for State 5 and State 10 (Table 3 - 4), and reflects a greater curvature in the extended sheet of 2N3J. This curvature is exhibited both in the angle between the two β4 – 7 β-sheets, and also a bend along the β6+7 strands, which in turn cause some deviation in the other parts of each monomer. Nonetheless, the two dimer structures are again in close agreement with each other as to the behaviour of the core domain and flanking regions of the terminal domains.

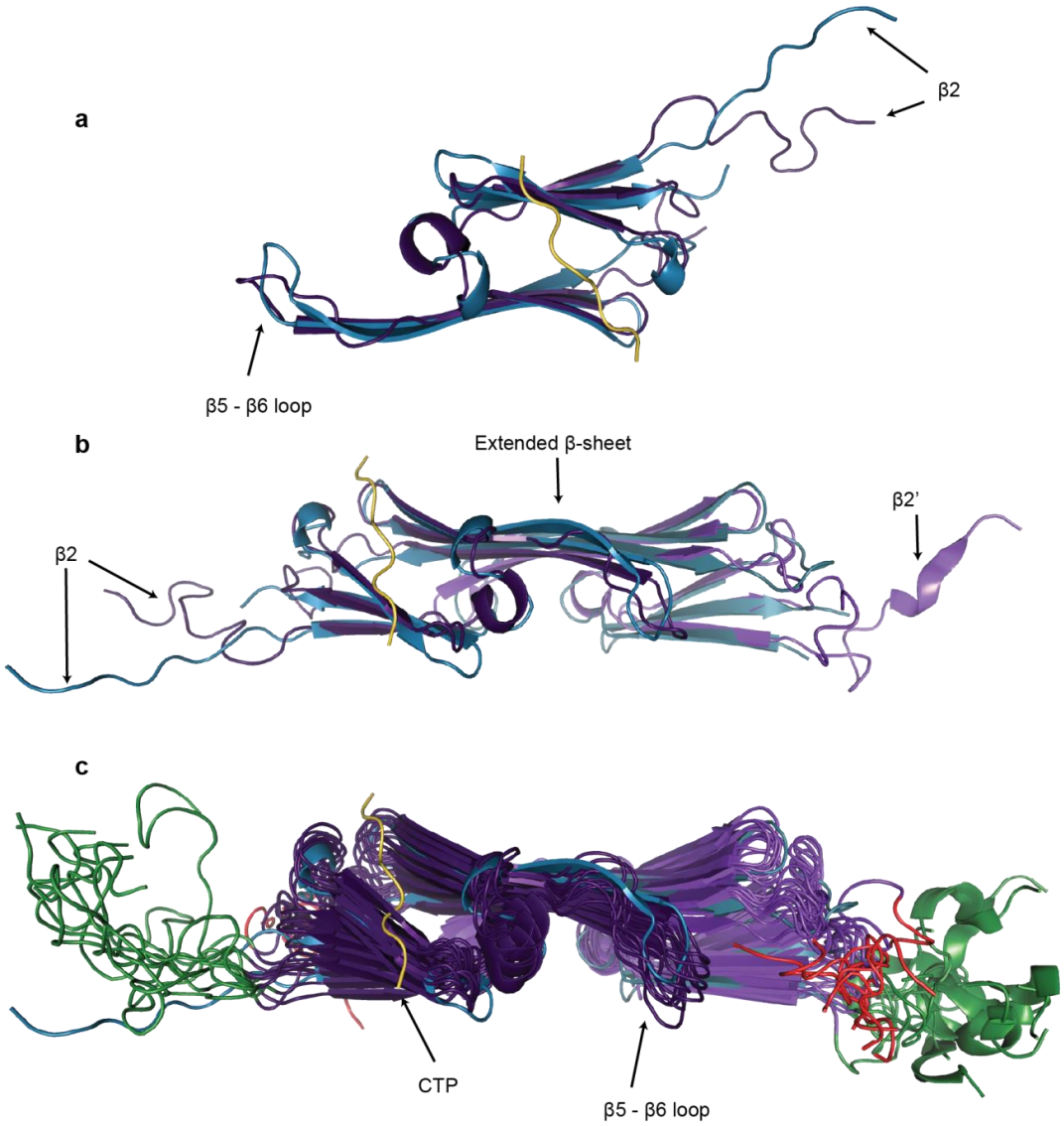


Figure 3 – 14 Superposition of the solution structure 2N3J (purple) with the crystal structure of c27-2 (blue, CTP in yellow). a) Superposition of the monomers of chain A again shows a similar structure of the core domain, including an extended β2 strand. 2N3J is represented by State 5 of the deposited ensemble. b) 2N3J exhibits a greater curvature of the extended β-sheet, as demonstrated between State 5 and c27-2. c) The NMR solution structure ensemble overlaid on the structure of c27-2, after alignment of both chains of c27-2 and State 10. The unused regions for the alignment correspond to the truncated N-terminal region (80 - 92 including β2, green), and the truncated C-terminus (residues 172 – 176, red). The N-terminal region exhibits considerable disorder in solution, as does the shorter region of the C-terminus.

2.2.4 Are the interactions of β2 an artefact of crystallisation?

Within the c27-1(4MJH) structure and the c27-2 structure presented above, there are three possible conformations of the β2: hydrogen bonding to β3 within the minor β-sheet (c27-1(4MJH) chains A and B); hydrogen bonding with two symmetry mates in an extended conformation (c27-2 chain A); and unresolved, presumably due to high disorder (c27-2 chain B). The β2 may populate these different states dependent on its environment in the crystal alone, but not in solution. We first looked at the crystal structures to see if there was any evidence for a particular preference of one of the conformations, due to the crystal setting.

Regarding location of the β2 in the c27-1(4MJH) structure (*i.e.* within the minor β-sheet), there is only one additional hydrogen bond that arises between the β2 of chain A and a symmetry mate due to the crystal packing, which is between R89 and D149' on its neighbour (Figure 3 – 5). There is substantial hydrogen bonding between β2 and β3 in chain A of 4MJH (Figure 3 – 9), and additionally a hydrogen bond could form between the S86 side chain on β2 and H131 on chain B (Figure 3 – 5), which lies on the loop between β5 and β6 that bends down around the shared groove. The position of the loop is very similar in c27-1 and c27-2 (Figure 3 – 13) so this interaction would also be maintained in the c27-2 crystal (there is a deviation in the conformer of the H131 side chain in c27-2, presumably due to the absence of this interaction), and all of these intra-dimer interactions will likely be accessible in solution (this has been supported by recent NMR data suggestive of a change in the chemical environment of the H131 between constructs with and without the β2 region). Similarly, H131 of chain A is in close proximity to the β2 on chain B in 4MJH and there is also a hydrogen bond between S86 of chain B and E125' on a symmetry mate, though residues 90-93 are not

resolved. Therefore, there is not a substantial difference in the interactions formed with neighbouring molecules for the β2 on chain B in either crystal structure. For β2 on chain A, it is likely that the large number of hydrogen bonds formed in either conformation (unravelled and bound (Figure 3 – 10), or in the minor β-sheet in c27-1(4MJH) (Figure 3 – 9)) act to stabilise the β2 in these conformations. The terminal glycine of the unravelled β2 in c27-2 may also form one hydrogen bond to the β5-6 loop of a third symmetry mate, though this crystal contact will be negligible in comparison to the other interactions it forms (Figure 3 – 10b).

c27-1(4MJH) has a smaller Matthew's coefficient than the other structures (V_M, Table 3 - 1). This is the ratio of the unit cell volume to the mass of protein within the unit cell, and so the c27-1(4MJH) crystal had a higher protein density than c27-2. Concomitantly, the estimated solvent content of the c27-1(4MJH) is smaller[56]. Thus it may be that the higher density of the crystal has reduced the conformational freedom of the β2, such that the minor-sheet bound form is favoured. Indeed, the cleft surrounding the non-resolved β2 in c27-2 is larger than that near the β2 in 4MJH, which would allow the β2 to access more conformations. However, there is not an obvious rationale as to why β2 of c27-2 has been induced to form the novel domain swapping form presented here, suggesting that all three conformers may be present in solution, albeit possibly with different populations. Thus it will be important to verify the behaviour of the β2 in solution.

2.2.5 Significance of the domain swapped β2

In addition to human HSPB6 mentioned above (Section 2.2.2), domain swapping has also previously been observed in other mammalian sHSPs[20]. In rat HSP20 (HSPB6), like human HSPB6 and the c27-2 structure presented here, the β2 binds to the β4 - β8 hydrophobic groove on a neighbouring dimer in the crystal of the core domain. Notably, neither of these proteins possess a C-terminal IXI motif[36] (Figure 6 – 1). In 2009, Bagnéris and colleagues reported the first crystal structure of the core domain of αB-crystallin alongside this core domain of rat HSP20[36]. The β2 of rat HSP20 was also observed in an unravelled conformation which also binds the β4 - β8 groove. Similarly to the HSP27 structure presented here, one sole valine (V67) resides in the hydrophobic groove, and hydrogen bonding occurs between flanking residues on the β4 and β8 strands (these interactions seem to be partly mediated by the rigidity of the neighbouring P68). V67 corresponds to M68 of αB-crystallin on sequence alignment[36] (Figure 6 – 1), which shows multiple resonances in solid-state NMR of full-length αB-crystallin, indicating this region exists in a range of environments[37]. Human HSPB6 is homologous to rat HSP20 and displayed widespread reciprocity of the β2-binding, with the same chain donating its N-terminus and receiving an extended N-terminus from the same partner[39] to form the tetramer. The residues in the groove vary, with one or two sidechains residing in the groove from one of three possible motifs (Figure 6 – 1). Similarly to the CTP of c27-2, their interactions again seem to be facilitated by the rigidity of a neighbouring proline residue.

Observation of this interaction in HSP27 is particularly intriguing because of its polydispersity and critically, because of the proximity of the S78 and S82 residues to the β2 region (c27-1

and c27-2 begin at G84). Phosphorylation at these sites is known to interrupt the oligomerisation of HSP27, with a much higher population of smaller oligomers[21,26] and moderates the chaperone activity of the protein *in vitro* and *in vivo*[4,21,24,26,58]. Phosphorylation at the equivalent sites weakens the dimer interface of αB-crystallin[27,28] with a concomitant strengthening of the edge interaction[29]. The effect of phosphorylation could be mediated by the behavior of this β2, both in its dissociation from the β3 strand and the possible resulting interactions with other dimers.

There are also multiple reports implicating the region affected by this variability in the function of the protein, both through the neuropathy-associated mutation G84R[9] and the maintained chaperone activity of fragments containing the region[59,60].

Taken together, these studies and the structure of c27-2 lead us to postulate that the activity and oligomerisation of the protein could be regulated in the cell by the effect of phosphorylation on the association of the β2 strand.

3 Conclusion

In this chapter, we have introduced a new structural model of the HSP27 core domain, in which we observed three sites of interaction and possible modulation of the subunit within the oligomer, all of which may contribute to the cellular stress response of this protein.

We found that co-crystallisation of c27-2 with the CTP reproduced the established C-terminal interaction with the β4-8 groove well[17]. This interface is already well-studied[16,44], and in FL

human αB-crystallin it has been found to mediate oligomerisation into large molecular weight species[61], be modulated by pH and temperature[33], and be allosterically linked to the β6+7 intra-dimer interface[29].

We have presented the first crystal structure with the intact intra-dimer disulphide, which acts to strengthen the intra-dimer interface (see Chapter 5 and Chapter 6 for a detailed discussion of the role of the disulphide, and its moderation by neuropathy-associated mutations). In the absence of the disulphide, the subunits of HSP27 3D are found to populate both monomers and dimers[26], indicating that the intra-dimer interface exists in an equilibrium between the bound and unbound form. We observe that this interface is formed through the main chain hydrogen bonding between the β6+7 strands, and salt bridges formed between side chains across the interface. Formation of the covalent disulphide will strengthen this interface and likely accounts for the solely even stoichiometries populated by oxidised HSP27[26].

Finally, we have also observed the β2 region of the core domain in an extended conformation, which then interacts with two neighbouring dimers within the crystal lattice. This implies that the β2 is able to unravel, domain swap with other subunits, and compete with the C-terminus, which implicates it in the modulation of oligomerisation. While the C-terminus of full-length αB-crystallin, and by similarity HSP27, is suggested to be sufficient for oligomer formation[19,33], the N-terminus is thought to determine the particular stoichiometries accessed[20] and appears to partly mediate oligomerisation in non-metazoan structures[13,32]. This behaviour is even more noteworthy because the β2 region is immediately downstream of two phosphorylation sites of

HSP27 which are known to interrupt the oligomerisation and regulate the activity of HSP27[21,24,26].

The structure of the ACD of HSP27 presented here has thus begun to shed light on a possible structural mechanism mediating the effect of phosphorylation, as well as provide a model for the intra-dimer disulphide of HSP27, which is thought to mediate its response to oxidative stress[5,43].

4 Methods

4.1 Materials and reagents

All reagents were purchased from Sigma unless otherwise specified.

4.2 Constructs

HSP27 3D-FL, 3D-ΔE, 3D-ΔC, were amplified from a non-tagged HSP27 3D-FL plasmid (a generous gift from Heath Ecroyd, University of Wollongong) and c27-2 and c27-3 were amplified from the c27-1 plasmid[17] (kindly provided by Art Laganowsky, University of Oxford). Amplification was undertaken with polymerase chain reaction (PCR) (primers purchased from Integrated DNA Technologies; Phusion polymerase from Agilent). After gel separation and purification, the gene fragments were introduced to a modified pET28a vector (6H-TEV-//gene//) digested with BamHI and XhoI (New England Biolabs, U.S.A.), using the Infusion HD Cloning kit (Clontech). Transformation was performed in Stellar competent cells (Clontech) and cultures grown in Luria Broth (Fisher) with kanamycin (50 µg/ml) ready

for harvest (MiniPrep kit, ThermoFisher). Successful mutation was determined by plasmid sequencing (SourceBioScience). MBP was later amplified and introduced into the 3D-FL, 3D-ΔE, and 3D-ΔC in a similar manner.

4.3 Protein expression and purification

BL21 (DE3) Gold cells (Agilent) were transformed with the pET28 vector 6H-TEV-[3D-FL]; 6H-TEV-[3D-ΔE]; 6H-TEV-[3D-ΔC]; 6H-MBP-TEV-[3D-FL]; 6H-MBP-TEV-[3D-ΔE]; 6H-TEV-MBP-[3D-ΔC]; 6H-TEV-[c27-2] or 6H-TEV-[c27-3]. 1l cultures were grown in Luria Broth (Fisher) with kanamycin (50 µg/ml) at 37°C until OD_{660} = 0.6-0.8, then induced with IPTG at 500 µM for three hours. Cells were harvested by centrifugation, washed in PBS and stored at -80°C. Cell pellets were resuspended and disrupted by sonication or microfluidisation in Buffer A (50 mM Tris, 300 mM NaCl, 20 mM imidazole, 5 mM BME, pH 8) supplemented with 1 EDTA-free protease inhibitor cocktail tablet (one per 1L pellet, Roche), followed by clarification by centrifugation. The first purification step was immobilised metal affinity chromatography (IMAC) conducted on a HisTrap HP column and Akta FLPC system (both GE Life Sciences) in Buffer A with elution in Buffer B (50 mM Tris, 300 mM NaCl, 500 mM imidazole, 5 mM BME, pH 8), followed by a desalting column (HiPrep, GE Life Sciences) or overnight dialysis in Buffer DS (20 mM Tris, 150 mM NaCl, 20 mM imidazole, 5 mM BME, pH 8). Overnight cleavage with TEV protease (S219V mutant purified in-house, pRK793, Addgene plasmid #8827) was conducted at room temperature and followed by a reverse IMAC step in Buffer A and then size exclusion chromatography (SEC) in SEC Buffer (20 mM Tris, 150 mM NaCl, pH 8) on a Superdex 75 column 26/60 (GE Life Sciences).

4.4 Crystallisation trials

The proteins were initially prepared for crystallisation at 10.0 mg/ml (c27-2, c27-3) and 10.1 mg/ml (HSP27-3D) in SEC buffer. CTP in water was added to the c27-2 and c27-3 solutions to a final concentration of 2.5 mM. Crystal trials were conducted using sitting-drop vapour-diffusion in 96-well plates and the commercial screens PactPremier, Midas and SaltRX, and JCSG+ for HSP27-3D. The second crystal trials were conducted with c27-2 at 14.8 mg/ml with 1.78 mM CTP; c27-3 at 13.8 mg/ml with 1.66 mM CTP; HSP27-3D-ΔC at 29.1 mg/ml to 1.8 mM CTP (*i.e.* approximately 1.2 CTP: monomer in each case); and HSP27-3D at 10.1 mg/ml and 69.1 mg/ml. These trials were undertaken with the Pact Premier, JCSG+, Index, 'Crystal Screen' and PegRx screens (HSP27-3D was used at either 10.1 mg/ml or 69.1 mg/ml). The ratios of the well solution to protein solution were 0.1 µl: 0.2 µl, 0.1 µl: 0.1 µl and 0.2 µl: 0.1 µl with 80-90 µl of well solution in the reservoir. The best initial hits of c27-2 and c27-3 were found with the 'Crystal Screen' commercial screen and were then optimised with hand-set hanging drop vapour-diffusion in 1:1, 1:2 and 2:1 ratio of the protein-peptide solution to well solution. The drops were 2-4.5 µl with a 1 ml volume of the reservoir solution (the reservoir solutions which gave the most successful crystallisation are listed in Table 3 - 5). Crystal trials were undertaken at 20°C and 4°C, but all the crystals used for data collection were grown at 20°C in all cases. The crystals were then harvested with brief transfer into a 20 % glycerol cryoprotectant made with the reservoir solution. Typically the hand-set crystals began to form within one week, and were fully formed within three weeks.

Construct	Reservoir Solution	Method	Cryoprotectant
c27-2	0.2M Ammonium Sulphate, 0.1M Sodium cacodylate trihydrate pH6.5, 22-24 % w/v PEG 8000	Sitting drop vapour diffusion	20% glycerol
c27-3	- 0.1M HEPES sodium pH7.5, 1.0 – 1.4 M Sodium citrate tribasic dehydrate - 0.2M Ammonium sulphate, 0.1M MES monohydrate pH6.5, 26-36% PEG monomethyl ether 5000	Sitting drop vapour diffusion	

Table 3 - 5 – Crystallisation conditions for the c27-2 and c27-3 constructs. The conditions for c27-3 refer to those in which crystals were optimised, but diffraction data was not collected as the crystals were smaller than those produced from c27-2 and so difficult to harvest. For c27-2, the conditions listed correspond to the crystal which gave the best diffraction data and is presented in this chapter.

4.5 Structure determination

Data for c27-2 were collected on the I04-1 beamline at Diamond Light Source, Didcot, UK using a radiation wavelength of 0.920 Å at 100 K. They were then processed using XDS and XSCALE[62]. The protein crystallised in the tetragonal space group $P\,4_3\,2_1\,2$ and diffracted to 2.25 Å with the unit cell dimensions a = 56.06, b = 56.06, c = 166.29 Å. The key data collection and refinement statistics are listed in Table 3 - 6. This was followed by molecular replacement conducted using Phaser[63] and the c27-1 structure (PDB 4MJH) as a search molecule.

Refinement was undertaken within Phenix[53] using TLS motion determination[64]. Coot[65] was used to analyse the electron density maps and correct conformational abnormalities that were detected using the Coot validation tools. Ramachandran Statistics within Coot classified 94.67% in preferred regions, 4.73% in allowed regions and 0.59% were outliers.

RMSD values were calculated with the 'align' tool in PyMol (The PyMOL Molecular Graphics System, Version 1.2r3pre, Schrödinger, LLC) and the images prepared with the same package.

	c27-2
Data collection	
Crystal system, space group	Tetragonal, $P\,4_3\,2_1\,2$ (#96)
Unit cell parameters a, b, c (Å) α, β, γ (°)	56.06, 56.06, 166.29 (90, 90, 90)
Resolution range (Å)	28.69 – 2.25 (2.31-2.25)
No. of measurements	332431 (24472)
No. of unique reflections	13304 (932)
Redundancy	25.0 (26.3)
Completeness (%)	99.9% (100%)
$I/\sigma(I)$	32.30 (2.24)
Wilson B value (Å^2)	59.66
R_{meas}	0.077 (1.821)
$CC_{1/2}$	1.000 (0.804)
Refinement	
Resolution (Å)	28.69 – 2.25 (2.308-2.250)
R_{work}	0.2215
No. reflections	12638
R_{work} / R_{free}	0.2215 / 0.2645
No. atoms	
Protein	1297
Ligand/ion	73
Water	34
R_{free}	0.2645
Amino acids‡	88 + 9, 78
No. of water molecules	34
No. of others	9
Mean B^* (Å^2)	
Protein	64.2
Water	52.42
Other	78.9
RMSD from target values	
Bond lengths (Å)	0.017
Bond angles (deg)	1.855
Dihedral angles (deg)	7.361

Ramachandran	
Most favoured (%)	94.67
Allowed (%)	4.73
Disallowed (%)	0.59

Table 3 - 6 – Data collection and refinement statistics for c27-2. *The mean residue temperature factor B was calculated with the BAverage program within the CCP4[57] suite. ‡(Chain A length + peptide, Chain B length). R_{meas}[66] is the multiplicity-independent measure of data quality where $R_{meas} = \frac{\sum_h \sqrt{\frac{n_h}{n_h-1}} \sum_i^{n_h} |\bar{I}_h - I_{h,i}|}{\sum_h \sum_i^{n_h} I_{h,i}}$ for observation i of reflection h with multiplicity n_h, intensity $I_{h,i}$ and average $\bar{I}_h$. R_{work} and R_{free} are defined in Chapter 1.

4.6 Mass spectrometry

The HSP27-WT and HSP27-3D samples used for the spectra in Figure 3 – 2 and Figure 3 – 4 were generously provided by Heath Ecroyd and prepared as detailed elsewhere[26] before buffer exchange into 200 mM ammonium acetate. pH 6.9. They were then equilibrated at 37 °C for one hour before MS analysis on a Q-ToF2 instrument (Waters, UK) in positive ESI mode with the following parameters: 1.6 kV capillary voltage; 60 V sample cone; 10 V extraction cone; 80 V collision cell 'energy'; 4.7 - 5.0 µbar backing pressure (*i.e.* in the source ion guide and quadrupole). For the transmission of smaller species in the inset to Figure 3 – 4 , 2.4 µbar backing pressure and 20 V collision cell voltage were used.

Gold-plated capillaries were prepared in-house[67] and spectra were analysed and calibrated in Waters MassLynx software using a CsI calibrant spectrum.

5 References

1. Arrigo, A.-P. & Gibert, B. Protein interactomes of three stress inducible small heat shock proteins: HspB1, HspB5 and HspB8. *Int. J. Hyperth.* **29,** 409–22 (2013).

2. Kayser, J. *et al.* The Small Heat Shock Protein Hsp27 Affects Assembly Dynamics and Structure of Keratin Intermediate Filament Networks. *Biophys. J.* **105,** 1778–1785 (2013).

3. Doshi, B. M., Hightower, L. E. & Lee, J. The role of Hsp27 and actin in the regulation of movement in human cancer cells responding to heat shock. *Cell Stress Chaperones* **14,** 445–457 (2009).

4. Garrido, C. Size matters: of the small HSP27 and its large oligomers. *Cell Death Differ.* **9,** 483–5 (2002).

5. Almeida-Souza, L. *et al.* Increased monomerization of mutant HSPB1 leads to protein hyperactivity in Charcot-Marie-Tooth neuropathy. *J. Biol. Chem.* **285,** 12778–86 (2010).

6. Cox, D., Carver, J. A. & Ecroyd, H. Preventing α-synuclein aggregation: The role of the small heat-shock molecular chaperone proteins. *Biochim. Biophys. Acta - Mol. Basis Dis.* **1842,** 1830–1843 (2014).

7. Arrigo, A. P. Human small heat shock proteins: Protein interactomes of homo- and

hetero-oligomeric complexes: An update. *FEBS Lett.* **587,** 1959–1969 (2013).

8. Houlden, H. *et al.* Mutations in the HSP27 (HSPB1) gene cause dominant, recessive, and sporadic distal HMN/CMT type 2. *Neurology* **71,** 1660–1668 (2008).

9. James, P. A., Rankin, J. & Talbot, K. Asymmetrical late onset motor neuropathy associated with a novel mutation in the small heat shock protein HSPB1 (HSP27). *J. Neurol. Neurosurg. Psychiatry* **79,** 461–463 (2008).

10. Evgrafov, O. V *et al.* Mutant small heat-shock protein 27 causes axonal Charcot-Marie-Tooth disease and distal hereditary motor neuropathy. *Nat. Genet.* **36,** 602–606 (2004).

11. Arrigo, A.-P. in *Mol. Chaperones Methods Protoc.* (eds. Calderwood, S. K. & Prince, T. L.) **787,** 105–119 (Humana Press, 2011).

12. Kim, K. K., Kim, R. & Kim, S. H. Crystal structure of a small heat-shock protein. *Nature* **394,** 595–9 (1998).

13. van Montfort, R. L. M., Basha, E., Friedrich, K. L., Slingsby, C. & Vierling, E. Crystal structure and assembly of a eukaryotic small heat shock protein. *Nat. Struct. Biol.* **8,** 1025–30 (2001).

14. Takeda, K. *et al.* Dimer structure and conformational variability in the N-terminal region of an archaeal small heat shock protein, StHsp14.0. *J. Struct. Biol.* **174,** 92–99 (2011).

15. Haslbeck, M. & Vierling, E. A First Line of Stress Defense: Small Heat Shock Proteins and Their Function in Protein Homeostasis. *J. Mol. Biol.* **427,** 1537–1548 (2015).

16. Hilton, G. R., Lioe, H., Stengel, F., Baldwin, A. J. & Benesch, J. L. P. Small Heat-Shock Proteins : Paramedics of the Cell. *Top. Curr. Chem.* **328,** 69–98 (2013).

17. Hochberg, G. K. A. *et al.* The structured core domain of αB-crystallin can prevent amyloid fibrillation and associated toxicity. *Proc. Natl. Acad. Sci. U. S. A.* **111,** E1562–70 (2014).

18. Baranova, E. V *et al.* Three-dimensional structure of α-crystallin domain dimers of human small heat shock proteins HSPB1 and HSPB6. *J. Mol. Biol.* **411,** 110–22 (2011).

19. Baldwin, A. J. *et al.* The Polydispersity of αB-Crystallin Is Rationalized by an Interconverting Polyhedral Architecture. *Structure* **19,** 1855–1863 (2011).

20. Laganowsky, A. *et al.* Crystal structures of truncated alphaA and alphaB crystallins reveal structural mechanisms of polydispersity important for eye lens function. *Protein Sci.* **19,** 1031–43 (2010).

21. Rogalla, T. *et al.* Regulation of Hsp27 Oligomerization, Chaperone Function, and Protective Activity against Oxidative Stress/Tumor Necrosis Factor by Phosphorylation. *J. Biol. Chem.* **274,** 18947–18956 (1999).

22. Clifton, A. D., Young, P. R. & Cohen, P. A comparison of the substrate specificity of

MAPKAP kinase-2 and MAPKAP kinase-3 and their activation by cytokines and cellular stress. *FEBS Lett.* **392,** 209–214 (1996).

23. Stokoe, D., Engel, K., Campbell, D. G., Cohen, P. & Gaestel, M. Identification of MAPKAP kinase 2 as a major enzyme responsible for the phosphorylation of the small mammalian heat shock proteins. *FEBS Lett.* **313,** 307–313 (1992).

24. Hayes, D., Napoli, V., Mazurkie, A., Stafford, W. F. & Graceffa, P. Phosphorylation Dependence of Hsp27 Multimeric Size and Molecular Chaperone Function. *J. Biol. Chem.* **284,** 18801–18807 (2009).

25. Kato, K., Hasegawa, K., Goto, S. & Inaguma, Y. Dissociation as a result of phosphorylation of an aggregated form of the small stress protein, hsp27. *J. Biol. Chem.* **269,** 11274–8 (1994).

26. Jovcevski, B. *et al.* Phosphomimics Destabilize Hsp27 Oligomeric Assemblies and Enhance Chaperone Activity. *Chem. Biol.* **22,** 186–195 (2015).

27. Ecroyd, H. *et al.* Mimicking phosphorylation of αB-crystallin affects its chaperone activity. *Biochem. J.* **401,** 129–141 (2007).

28. Aquilina, J. A. *et al.* Phosphorylation of αB-Crystallin Alters Chaperone Function through Loss of Dimeric Substructure. *J. Biol. Chem.* **279,** 28675–28680 (2004).

29. Hilton, G. R. *et al.* C-terminal interactions mediate the quaternary dynamics of αB-crystallin. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **368,** 20110405 (2013).

30. Hilario, E., Martin, F. J. M., Bertolini, M. C. & Fan, L. Crystal Structures of Xanthomonas Small Heat Shock Protein Provide a Structural Basis for an Active Molecular Chaperone Oligomer. *J. Mol. Biol.* **408,** 74–86 (2011).

31. Hanazono, Y., Takeda, K., Yohda, M. & Miki, K. Structural studies on the oligomeric transition of a small heat shock protein, StHsp14.0. *J. Mol. Biol.* **422,** 100–8 (2012).

32. Hanazono, Y. *et al.* Nonequivalence Observed for the 16-Meric Structure of a Small Heat Shock Protein, SpHsp16.0, from Schizosaccharomyces pombe. *Structure* **21,** 220–228 (2013).

33. Baldwin, A. J., Lioe, H., Robinson, C. V., Kay, L. E. & Benesch, J. L. P. αB-Crystallin Polydispersity Is a Consequence of Unbiased Quaternary Dynamics. *J. Mol. Biol.* **413,** 297–309 (2011).

34. Aquilina, J. A., Benesch, J. L. P., Bateman, O. A., Slingsby, C. & Robinson, C. V. Polydispersity of a mammalian chaperone: Mass spectrometry reveals the population of oligomers in αB-crystallin. *Proc. Natl. Acad. Sci.* **100,** 10611–10616 (2003).

35. Aquilina, J. A., Shrestha, S., Morris, A. M. & Ecroyd, H. Structural and functional aspects of hetero-oligomers formed by the small heat shock proteins αB-crystallin and HSP27. *J. Biol. Chem.* **288,** 13602–9 (2013).

36. Bagnéris, C. *et al.* Crystal Structures of α-Crystallin Domain Dimers of αB-Crystallin and Hsp20. *J. Mol. Biol.* **392,** 1242–1252 (2009).

37. Jehle, S. *et al.* αB-Crystallin: A Hybrid Solid-State/Solution-State NMR Investigation Reveals Structural Aspects of the Heterogeneous Oligomer. *J. Mol. Biol.* **385,** 1481–1497 (2009).

38. Baranova, E. V., Beelen, S., Gusev, N. B. & Strelkov, S. V. The taming of small heat-shock proteins: crystallization of the α-crystallin domain from human Hsp27. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* **65,** 1277–1281 (2009).

39. Weeks, S. D. *et al.* Molecular structure and dynamics of the dimeric human small heat shock protein HSPB6. *J. Struct. Biol.* **185,** 342–354 (2014).

40. Laganowsky, A. & Eisenberg, D. Non-3D domain swapped crystal structure of truncated zebrafish alphaA crystallin. *Protein Sci.* **19,** 1978–1984 (2010).

41. Clark, A. R., Naylor, C. E., Bagnéris, C., Keep, N. H. & Slingsby, C. Crystal structure of R120G disease mutant of human αb-crystallin domain dimer shows closure of a groove. *J. Mol. Biol.* **408,** 118–134 (2011).

42. Delbecq, S. P., Jehle, S. & Klevit, R. Binding determinants of the small heat shock protein, αB-crystallin: recognition of the 'IxI' motif. *EMBO J.* **31,** 4587–4594 (2012).

43. Rajagopal, P., Liu, Y., Shi, L., Clouser, A. F. & Klevit, R. E. Structure of the α-crystallin domain from the redox-sensitive chaperone, HSPB1. *J. Biomol. NMR* 0–5 (2015). doi:10.1007/s10858-015-9973-0

44. Hochberg, G. K. A. & Benesch, J. L. P. Dynamical structure of αB-crystallin. *Prog.*

Biophys. Mol. Biol. **115,** 11–20 (2014).

45. Jehle, S. *et al.* Solid-state NMR and SAXS studies provide a structural basis for the activation of αB-crystallin oligomers. *Nat. Struct. Mol. Biol.* **17,** 1037–1042 (2010).

46. Chalova, A. S., Sudnitsyna, M. V, Semenyuk, P. I., Orlov, V. N. & Gusev, N. B. Effect of disulfide crosslinking on thermal transitions and chaperone-like activity of human small heat shock protein HspB1. *Cell Stress Chaperones* **19,** 963–72 (2014).

47. Indu, S., Kochat, V., Thakurela, S., Ramakrishnan, C. & Varadarajan, R. Conformational analysis and design of cross-strand disulfides in antiparallel β-sheets. *Proteins* **79,** 244–60 (2011).

48. Préville, X. *et al.* Mammalian Small Stress Proteins Protect against Oxidative Stress through Their Ability to Increase Glucose-6-phosphate Dehydrogenase Activity and by Maintaining Optimal Cellular Detoxifying Machinery. *Exp. Cell Res.* **247,** 61–78 (1999).

49. Zavialov, A. *et al.* The effect of the intersubunit disulfide bond on the structural and functional properties of the small heat shock protein Hsp25. *Int. J. Biol. Macromol.* **22,** 163–73 (1998).

50. Mehlen, P., Kretz-Remy, C., Préville, X. & Arrigo, A. P. Human hsp27, Drosophila hsp27 and human αB-crystallin expression-mediated increase in glutathione is essential for the protective activity of these proteins against TNFα-induced cell death. *EMBO*

J. **15,** 2695–706 (1996).

51. Benesch, J. L. P., Ruotolo, B. T., Simmons, D. A. & Robinson, C. V. Protein complexes in the gas phase: technology for structural genomics and proteomics. *Chem. Rev.* **107,** 3544–3567 (2007).

52. Chávez Zobel, A. T., Lambert, H., Thériault, J. R. & Landry, J. Structural instability caused by a mutation at a conserved arginine in the alpha-crystallin domain of Chinese hamster heat shock protein 27. *Cell Stress Chaperones* **10,** 157–66 (2005).

53. Adams, P. D. *et al.* PHENIX : a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66,** 213–221 (2010).

54. Murshudov, G. N. *et al. REFMAC* 5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **67,** 355–367 (2011).

55. Ikeda, Y. *et al.* A clinical phenotype of distal hereditary motor neuronopathy type II with a novel HSPB1 mutation. *J. Neurol. Sci.* **277,** 9–12 (2009).

56. Matthews, B. W. Solvent content of protein crystals. *J. Mol. Biol.* **33,** 491–497 (1968).

57. Winn, M. D. *et al.* Overview of the CCP4 suite and current developments. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **67,** 235–242 (2011).

58. Bryantsev, A. L., Chechenova, M. B. & Shelden, E. A. Recruitment of phosphorylated small heat shock protein Hsp27 to nuclear speckles without stress. *Exp. Cell Res.* **313,**

195–209 (2007).

59. Thériault, J. R. *et al.* Essential role of the NH2-terminal WD/EPF motif in the phosphorylation-activated protective function of mammalian Hsp27. *J. Biol. Chem.* **279,** 23463–71 (2004).

60. Nahomi, R. B., DiMauro, M. A., Wang, B. & Nagaraj, R. H. Identification of peptides in human Hsp20 and Hsp27 that possess molecular chaperone and anti-apoptotic activities. *Biochem. J.* **465,** 115–125 (2015).

61. Baldwin, A. J. *et al.* Quaternary dynamics of αB-crystallin as a direct consequence of localised tertiary fluctuations in the C-terminus. *J. Mol. Biol.* **413,** 310–20 (2011).

62. Kabsch, W. XDS. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66,** 125–132 (2010).

63. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40,** 658–674 (2007).

64. Painter, J. & Merritt, E. A. Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **62,** 439–450 (2006).

65. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66,** 486–501 (2010).

66. Diederichs, K. & Karplus, P. A. Improved R-factors for diffraction data analysis in

macromolecular crystallography. *Nat. Struct. Biol.* **4,** 269–275 (1997).

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

68. Kabsch, W. & Sander, C. Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22,** 2577–2637 (1983).

69. Joosten, R. P. *et al.* A series of PDB related databases for everyday needs. *Nucleic Acids Res.* **39,** D411–D419 (2011).

70. Sievers, F. *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7,** 1–6 (2011).

71. Goujon, M. *et al.* A new bioinformatics analysis tools framework at EMBL-EBI. *Nucleic Acids Res.* **38,** W695–9 (2010).

Chapter 4: Investigating Regulation of HSP27 by Phosphorylation

1 Introduction

Protein phosphorylation is exceedingly widespread, modulating aspects of every fundamental cellular process, from growth to memory[1]. This common PTM appears on all members of the human sHSP family[2], is implicated in the role of this protein under stress[3] and non-stress conditions[4], and though the nature of its reported effect on oligomerisation and chaperone activity varies, it is widely accepted to modulate the structure and function of the protein[2,5–7]. For example, it has been reported to have a variable effect on substrate aggregation *in vitro*, with both deterioration[6] and amelioration[5,7] of HSP27 chaperone activity being observed, possibly reflecting a substrate-dependent activity or variable chaperoning mechanisms. Additionally, HSP27-3D conferred poorer protection against oxidative stress on mouse L929 cells than the WT[6], but was found to better prevent protein precipitation in neuro-2a cell lysates[7], illustrating that sHSP behavioural modulation by phosphorylation is not straightforward.

Similarly, the mechanisms of regulation of the ATP-independent sHSPs, found across all domains of life[2,8] and critical for healthy cell survival, remain ill-defined. This is due to the

high level of complexity exhibited by the sHSP network. Expression of the sHSPs is upregulated on stress at the transcriptional[9] and even translational level[2]. Additionally, the constitutive or stress-induced sHSP population is regulated by the stresses themselves such as pH[10], temperature[10–13] and oxidation[14]. They also exhibit a direct response to the cellular environment including the presence of substrate[2,11], and the availability of other sHSPs for hetero-oligomerisation[8,13]. Finally, they are also regulated by the cellular machinery through PTM[2]. The numerous layers of sHSP regulation, paired with their associated dynamic and heterogeneous association as well as the independent characteristics of each sHSP member within an organism, paints a complex and presumably highly-tuned network of these chaperones which enables them to fulfil their multiple roles within the cell.

The behaviour observed for the β2 region of HSP27 in Chapter 3 may shed some light on one of these multiple modes of regulation, and therefore it was essential that we next probe whether our observations carry any significance in solution. This chapter thus details the experiments to test the validity of the new binding schemes we observed. Namely, whether the flexible behaviour is recapitulated in solution; whether there is any evidence for domain swapping and the competition between the β2 and the C-terminus for binding to the ACD; and whether the behaviour of the β2 is moderated by phosphorylation.

2 Results and Discussion

2.1 MD simulations validate that the β2 easily dissociates

In order to verify whether our observation of the extended β2 region was relevant in solution, we first turned to molecular dynamics (MD) simulations to see whether the β2 would exhibit a higher propensity to unravel than the rest of the protein. The dimer of the c27-1 model deposited in the Protein Data Bank (PDB identifier: 4MJH) was used to build a starting model for the simulations as this has the β2 in the bound conformation. c27-1(4MJH) is missing residues 84-85 and 90-93 on chain B, so these were constructed from the equivalent positions in chain A. Additionally, the disulphide was introduced between the two C137 side chains in case this was necessary for β2 dissociation, as c27-1(4MJH) had been crystallised under reducing conditions.

The simulation was performed in explicit water solvent, and run for 630 ns after minimisation and an initial period of equilibration (see Methods). The first frame of this simulation after equilibration is shown in the top left panel of Figure 4 – 1. The structure is coloured according to the root mean squared fluctuation (RMSF) for the C_αs of each residue across all frames, which is a measure of the relative degree of flexibility for the protein chain over the whole simulation. Blue indicates little fluctuation and red marks regions that are highly flexible. The β2 region for chain A and chain B are marked as β2 and β2' respectively, and both regions exhibit a higher RMSF than the rest of the two chains, as indicated by the pink and red colouring. Indeed during the simulation the β2 chains fluctuate significantly about their

position next to the β3 strand, often breaking and reforming the hydrogen bonds with the β3. Additionally, the β2 on chain A completely unravels without any hydrogen bonds involving residues 84-91 (i.e. the whole of the β2).

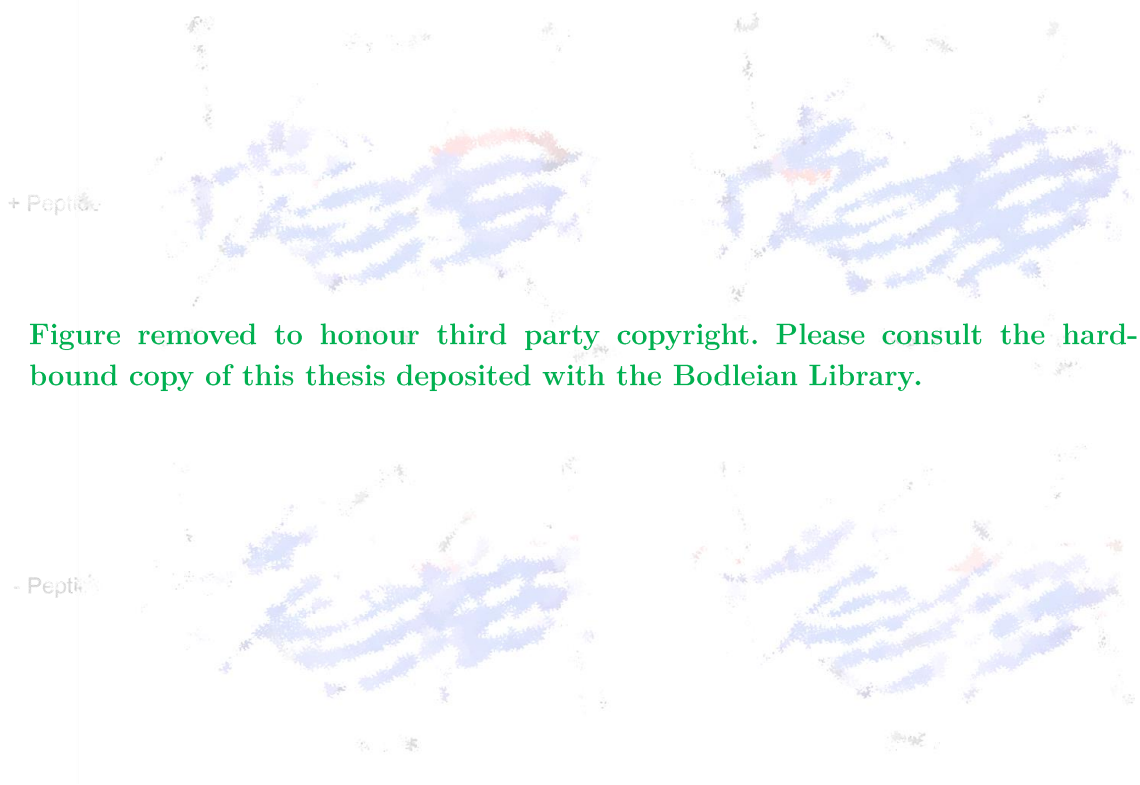


Figure removed to honour third party copyright. Please consult the hard-bound copy of this thesis deposited with the Bodleian Library.

Figure 4 – 1 MD simulations of the core domain, with and without peptide and with and without the cysteines oxidised to the disulphide, all show increased flexibility in the β2 strand, with several residues completely dissociating in the course of the simulation. The RMSF in Å is indicated on the same colour scale of blue through white to red, where blue depicts a low RMSF and red indicates a higher degree of flexibility. MD simulations were performed with Matteo T. Degiacomi (University of Oxford), figure panels courtesy of M.T.D..

The RMSF is less than 2.5 Å for most of the protein and less than 1.5 Å for the β-sheet core, but 3 – 10 Å for the β2 (the residue at the very beginning of the core domain has an RMSF of 10 Å as this moves the most on unravelling). The RMSF for atom i is directly related to the crystallographic temperature factor B for atom i[16] via:

$$RMSF_i^2 = \frac{3}{8\pi^2} B_i \quad (1)$$

So the RMSF calculated in the simulations can act as a predictor of the temperature factor for that atom in the crystal structure. So a C_α RMSF of 3 – 10 Å corresponds to B factors of 240 – 2600 Å^2. Not unsurprisingly, this is much higher than the B values recorded for β2 in c27-1(4MJH), as the positional variation of atoms (and residues) with such B factors would be so great as to prevent their detection with X-ray diffraction. Thus, this simulation is in keeping with the unresolved β2 of chain B in c27-2, indicating an equilibrium between the bound and unbound state. As the model system could not include additional molecules nearby without substantial computational expense, the unravelled β2 does not have the opportunity to associate with another dimer. However, the simulation does support our hypothesis that the β2 can access an extended conformation in solution.

The two crystal structures of c27-1(4MJH) and c27-2 were solved from crystals grown at pH 8 and pH 6.5 respectively, and the β2 of c27-1(4MJH) through S86 on chain A can form a hydrogen bond with H131 in the β5-6 loop of chain B. The pK_a of the nitrogen in the histidine imidazole ring is 6.04 (for the nitrogen with a lone electron pair in the plane of the ring (Figure 4 – 2, left panel)). Thus there will be a higher population of doubly protonated histidines at pH 6.5 than pH 8, which may interrupt the hydrogen bonding to the S86 side chain observed at pH 8. In order to ascertain whether the β2 behaviour was a result of pH, we ran another simulation under the same conditions, with this histidine doubly protonated (Figure 4 – 2, right panel). This simulation produced similar behaviour of the β2, so the flexibility of the β2 is not solely dependent on this hydrogen bond with H131.

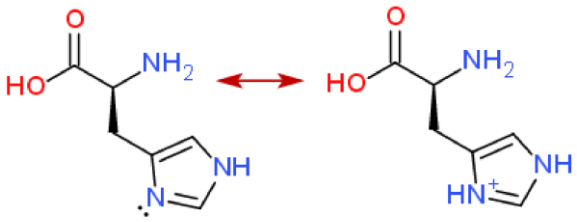


Figure 4 – 2 The two protonation states of histidine used in these simulations. Figure made with the ChemDoodle webtool.

We then altered the c27-1(4MJH) MD model again to account for two further possible allosteric contributions to the β2 flexibility, with H131 in the singly protonated state. c27-1(4MJH) was collected under reducing conditions, so we also ran simulations with and without the disulphide. Additionally, the absence of the C-terminal peptide on chain B in the crystal structure of c27-2, for which the β2 is not resolved (indicating high disorder), and our hypothesis that the unravelled β2 may compete with the C-terminal peptide at the β4 - β8 groove, led us to question whether the presence of the C-terminal peptide may reciprocally influence the behaviour of the β2. The initial frames after equilibration in the different conditions are shown in the remaining panels of Figure 4 – 1, coloured according to the RMSF for each simulation on the same scale. In all cases, the β2 is the most flexible part of the protein, indicated by the white and red banding on residues therein. There was not a noticeable difference in the β2 behaviour between the simulations. While some of the β2 regions unravelled more than others due to the stochastic nature of this experiment; the difference between the different conditions was no more pronounced than between the A and B chains in each simulation.

The simulations demonstrate that the interaction of the β2 within the minor sheet is labile, with each simulation showing loss of the β-strand structure before partial or complete unravelling of this region. It took at least 200 ns before full unravelling was first observed in one simulation. Again, as the process is stochastic, the time that it takes for the β2 to unravel varies between chains. This indicates that the β2 is able to exist in the bound conformation, rather than dissociating immediately, but that it also readily accesses the unbound state. Notably, residues 87-89 retain β-strand secondary structure more than the rest of the β2 strand, and the hydrogen bond involving the carbonyl oxygen of residue E87 seems to stabilise the β2 in the bound conformation. Indeed, breaking of this bond is reproducibly the last event prior to full loss of secondary structure.

The four simulations together demonstrate the behaviour of the rest of the core domain. The lighter blue shading on the loops connecting the major and minor β-sheets, along with the β5-6 loop, indicates their increased flexibility relative to the residues within the β-sheets. The β9 strands are indicated and again have a low RMSF, perhaps not surprising for a strand within the middle of a β-sheet, but demonstrating that nearby truncation has not caused this strand to be unduly flexible, and suggesting that the unravelling of the β2 is similarly not an artefact of truncating the full length protein to the core domain alone. Finally, the IXI motif of the C-terminal peptide in full-length αB-crystallin is bound only 2 % of the time, though localised near the β4 – β8 groove when not bound[17]. In these simulations, the CTP has a low RMSF (blue banding) despite its interaction with the binding groove being mediated only by six-hydrogen bonds and the burial of the isoleucine and valine residues in the hydrophobic groove.

So it is much less flexible than the β2, which can form nine hydrogen bonds to the β3 when in the β-sheet.

As there was not a significant difference between the simulations run in the different conditions in Figure 4 – 1, each chain from all four simulations was taken as a repeat of the experiment, and the mean C_α RMSF for each residue calculated. The first eight residues show a higher RMSF than any other part of the protein (Figure 4 – 3). Taken together, these results indicate that the β2 is somewhat disordered in solution and is not dependent on the oxidation state or the binding of the peptide.

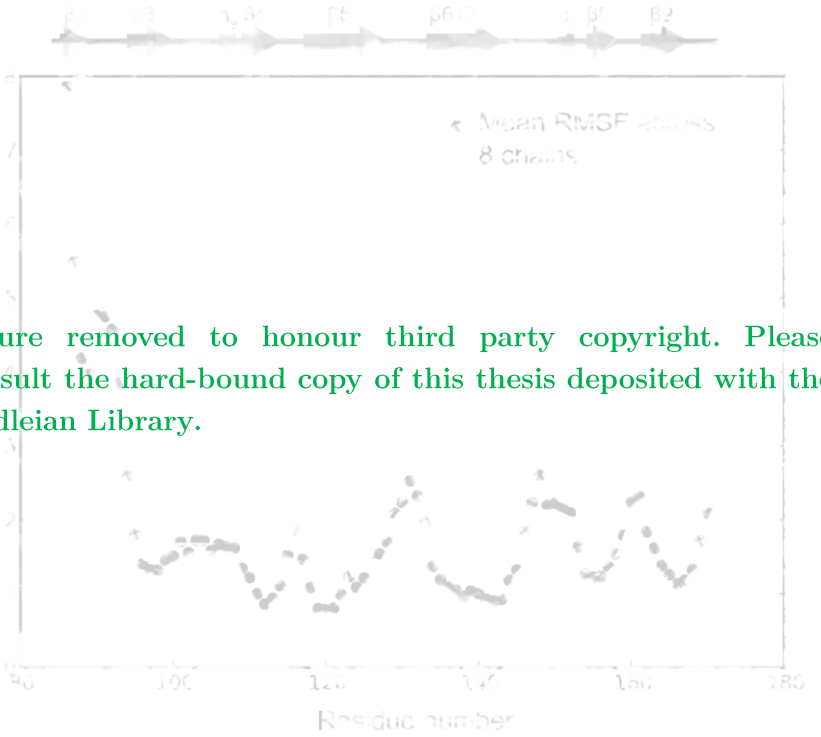


Figure removed to honour third party copyright. Please consult the hard-bound copy of this thesis deposited with the Bodleian Library.

Figure 4 – 3 The mean RMSF for the C_α of each residue taken from all eight chains in the simulations in Figure 4 – 1. The first eight residues, which comprise the β2 strand, show much more fluctuation than the rest of the protein. Graph courtesy of M.T.D..

2.2 NMR detects a disordered β2 strand in solution

Having observed the propensity of the β2 to dissociate from the β-sheet within the theoretical framework of MD simulations, we next turned to solution NMR to establish whether there is experimental evidence for the flexibility of the β2.

This was achieved by calculating generalized order parameters (S^2) for the N-H bond vectors of each residue[18] in c27-2. 2D ^{1}H-detected ^{15}N R_1, R_2, and {^{1}H}-^{15}N heteronuclear NOE experiments were conducted at two magnetic field strengths (14.1 T and 11.7 T[19]), on 1 mM oxidized c27-2 at 298K (Figure 4 – 4). (Only ^{15}N R_1 and R_2 data were collected on the 11.7 T spectrometer). S^2 values close to 1.0 for c27-2 would describe rigid regions of the protein, whereas values below 0.2 would indicate completely disordered and flexible regions[18], as is observed for IDPs. Residues within flexible loops and termini of folded proteins give rise to S^2 values near and below 0.6.

The calculated S^2 values for the β-strands and α-helices of c27-2 lie between 0.80 and 0.95 (Figure 4 – 4b), indicating their retained order in solution[18]. The S^2 values for the loop regions and the β2 were generally lower, between ~0.6 and 0.8. This implies that these regions are mobile on the picosecond timescale, but not completely disordered (S^2 values >0.6 rather than <0.2). The most disordered regions are the β2 strand and the C-terminal K171 residue, which is consistent with the variable position of this residue in the crystal structures (Chapter 3). Therefore the S^2 values for the β2 region are much lower than would be expected if it solely occupied the state bound to the β3. These data thus demonstrate that while the β2 is not completely disordered in solution, perhaps due to its equilibrium with the bound β-sheet state

or its interaction with other binding sites as hypothesized in Chapter 3, it populates a dynamic conformation in solution and is inconsistent with an ordered β-strand.

These results were supported by secondary ^{13}C chemical shifts ($\Delta\delta^{13}C$), which indicated local disordered secondary structure for β2[20], and $^{1}H^{N}$ chemical shift temperature coefficients, which demonstrated that the β2 region is solvent-exposed[21].

Figure 4 – 4 Analysis of backbone mobility indicates the β2 is disordered in solution. a) Generalised order parameters S^2 calculated from ^{15}N relaxation data collected at 11.7 and 14.1 T at 298 K on 1 mM oxidised c27-2. b) The order parameters from (a) plotted onto the c27-2 structure according to the colour bar shown. T. Reid Alderson performed the NMR experiments. Figure courtesy of T.R.A..

2.3 Oligomeric distribution of c27-2

Having established that the β2 exists in an equilibrium with an unbound and disordered state, we then turned to native MS to determine the oligomeric distribution of the core domain. The c27-2 within the crystal seems to oligomerise with three partners (the dimeric partner, and the two binding partners of the β2), each of which in turn interacts with multiple partners to form a runaway complex. However, based on previous observations with c27-1(4MJH) and cABC,

the SEC elution profile during purification, and the solubility of c27-2 which did not suggest the formation of large aggregates, we had postulated that the biological unit contained within the crystal was dimeric. Native MS can provide a direct measure of the oligomeric population of proteins[22] and observe the individual species present within an ensemble[23].

In order to do this, the conditions for the transmission of the macromolecular ions must be optimised to maintain the complex intact throughout its journey through the mass spectrometer. Specifically, the voltages used to accelerate the ions through each section of the spectrometer must be low enough to prevent collision-induced dissociation of the protein while maintaining the path of the ions to the detector and even interrupting non-specific adduction of salt and solvent[24]. This is also highly dependent on the pressures used in the initial vacuum stages of the mass spectrometer, which must be low enough to prevent scatter of the ions but high enough to retain some collisions to collisionally focus the ions on their path through the spectrometer[25,26]. Concomitantly, a low protein concentration and small capillary orifice size must be used to prevent non-specific oligomers, which are an artefact of the ESI process[26].

The mass spectrum of c27-2 is presented in Figure 4 – 5. The peaks between 1950 – 3300 Th represent a charge state distribution of one species, with sequential peaks corresponding to the protein ion with one fewer charge (from one fewer proton or cation). From the distance between the peaks on the m/z scale and the knowledge that each peak is related to its neighbour by a difference in charge, $\Delta z = \pm 1$, one can calculate the mass of the protein ion.

The peaks labelled with blue squares correspond to a species with mass 19662.75 ± 3.85 Da[†], which compares well to the theoretical mass of the c27-2 dimer (19663.8 Da). This charge state series dominates the spectrum, indicating that the dimer is the most abundant species in oxidised c27-2, and confirming our hypothesis that the dimer is the biological unit in the crystal structure. Moreover, reduction of the protein leads to a significant population of monomers (Chapter 6), consistent with the disulphide 'locking' the dimer interface between the two β6+7 strands.

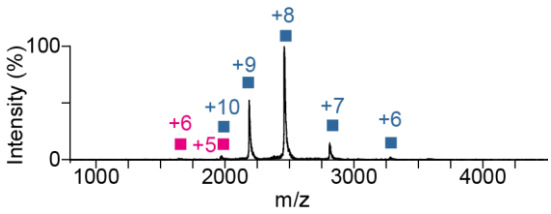


Figure 4 – 5 The native mass spectrum of c27-2 at a 20 µM monomeric concentration and a sample cone voltage of 150 V.

There is also a second, very lowly abundant species in the mass spectrum, indicated with the pink squares. This corresponds well with a monomer species at 9831.51 ± 2.09 Da (*cf* 9831.9 Da for the theoretical mass), and peaks for charges +6 and +7. This is distinguishable from the dimer, despite the possible overlap of the monomer peaks with the +12 and +10 peaks of the dimer, because of the absence of the +11 dimer peak and the Gaussian profile of a charge state distribution[27]. This means that the peak at 1638 Th can be attributed to the monomer

[†] The error in the masses discussed here is the standard deviation in the mass extracted from the sequential charge states of a particular species; it thus reflects the variation in the m/z position of centroids modelling the peaks within a charge state distribution within a specific spectrum. While it is a helpful guide to the relative certainty of the mass extracted from the charge state envelope, it can sometimes misleadingly imply that the values calculated from the mass spectrum are inaccurate. Thus, it should be noted that manual assignment of the individual peaks normally returns a mass within a few Daltons of the expected mass, at the most.

+6 peak rather than the dimer +12 peak. It is clear from the total areas under the two charge states that the monomer abundance is minimal, though it does indicate that there is a very small monomer population in solution.

This is further evidenced when IM is used to separate the ions based on their CCS, as shown by the arrival time distribution (Figure 4 – 6b). This is the arrival time of the ions at the detector, which results from the different drift times of the ions in the IM cell. Together, the separation in the m/z dimension and in the drift time dimension can produce a powerful 2D separation of species in the 2D spectrum (Figure 4 – 6c). This spectrum was collected with a lower potential on the sample cone (10V rather than 150V in Figure 4 – 5), producing less collision-induced cleaning of the macromolecular ions and thus wider peaks[24] (the lower accelerating voltage was used to prevent collision-induced unfolding of the ions, which would affect the drift times). Nonetheless, the peaks corresponding to the non-adducted dimer give a mass of 19662.54 ± 3.22 Da for this charge state distribution, again very close to the theoretical value of 19663.8 Da. The dimer accounts for 97.7 % of the signal within the spectrum (as determined from the areas of the peaks using the UniDec software[28] and the charge state envelopes identified in Figure 4 – 6c). This shows that the peak at 1974 Th is mainly due to the dimer (z = +10) rather than the monomer (z = +5), as assumed for the 1D spectrum above.

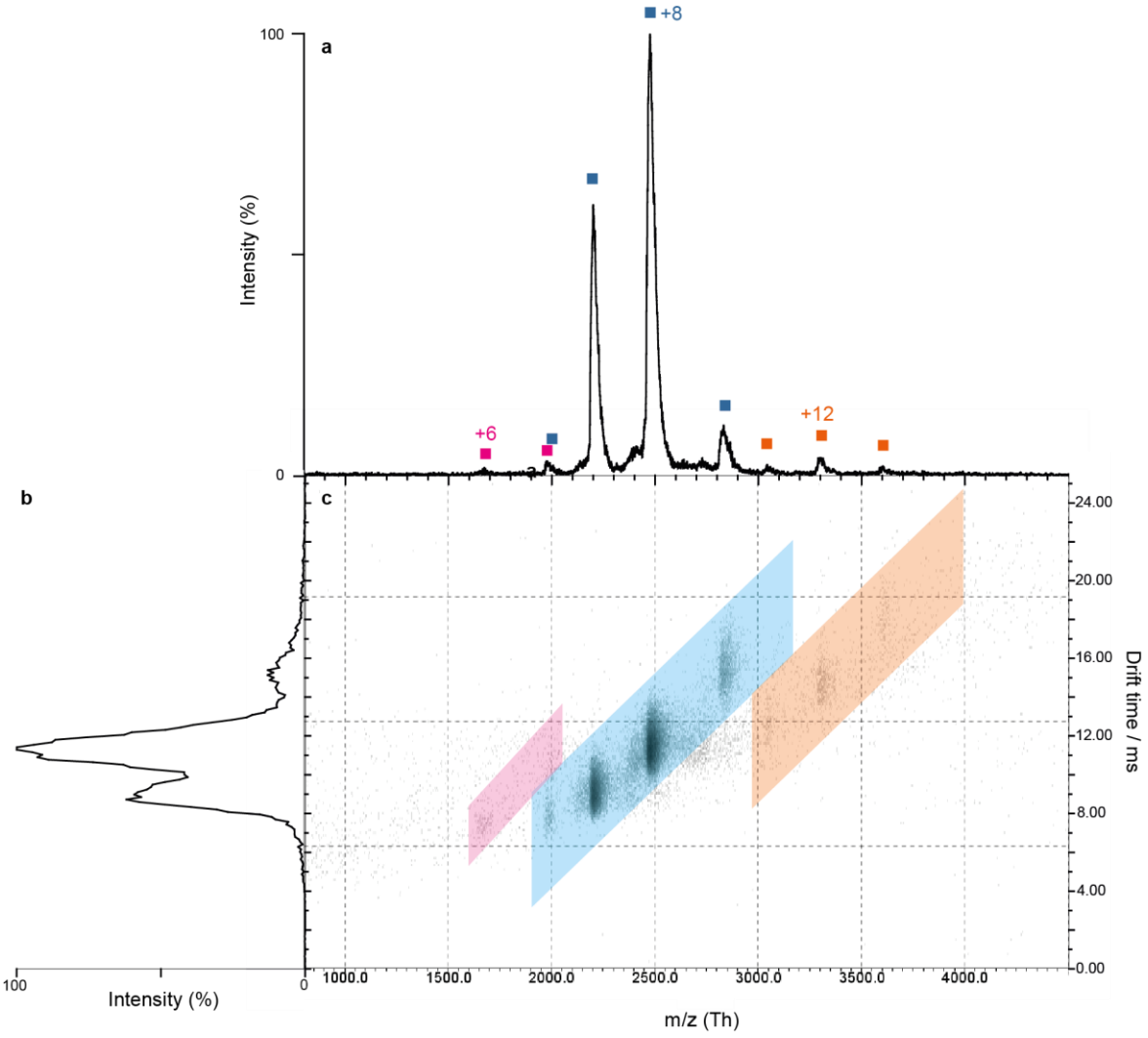


Figure 4 – 6 Ion mobility MS of c27-2 under 'soft' instrument acceleration voltages. A) The 1D mass spectrum, b) the arrival time distribution ('mobilogram'), c) the combination of the two dimensions of separation allows three individual species to be identified (shaded regions). The intensity of the 2D plot is presented on a base-10 logarithmic scale; the 1D plot intensities are linear. Monomeric concentration is 20 µM; this sample is the same as that used for the spectrum in Figure 4 – 5. Sample cone voltage = 10 V.

The most notable part of this spectrum is the appearance of a third species (orange shading, Figure 4 – 6c). The intensity of the 2D plot is plotted on a base-10 logarithmic scale, so the species shaded with pink and orange are very lowly abundant relative to the dimer in blue. Assignment of the orange species reports a mass of 39474.13 ± 31.08 Da, which is in keeping with a tetramer of c27-2 (theoretical mass 39327.6 Da). Accurate mass assignment is impeded

by the low abundance of this species, which means that the non-adducted peak is difficult to identify. Nonetheless, the measurement is within 0.4 % of the theoretical mass, and moreover the species shaded in orange displays drift time behaviour which is in keeping with a larger, more highly charged species than the dimer. This tetrameric species represents approximately 2 % of the signal in the spectrum, and likely represents a non-native dimer of dimers of the protein. This is in part a non-specific artefact of the electrospray process, resulting from the statistical possibility of having two dimers within one fission-produced droplet that then undergoes final solvent evaporation[26]. This would result in an artificial dimer of dimers being transmitted through the mass spectrometer. Additionally, it is possible that the core domain can assemble transiently in solution to form a non-specific dimer of dimers, which the sensitive MS experiment is able to detect. This is weakly supported by the possibility of crystallising these core domains in a dimer of dimer form[29] and indeed the interactions suggested for c27-2 herein.

The species highlighted in pink is even less abundant (less than 0.5 %), and corresponds to a monomer (mass approximately 9860.01± 19.86 Da), again mass assignment is hindered by the low signal to noise of this species. From the 2D spectrum, it is evident that the highest charge state in this series is the +6 (see assignment in Panel (a)). Post-desolvation dissociation has been minimised in this spectrum with low acceleration voltages throughout the mass spectrometer, and so it is likely that this minor population reflects a small amount of core domain in which the disulphide has been reduced (a similarly small amount of reduced protein was detected in solution subunit exchange, see Chapter 6). This reduced population would be

able to access both the noncovalent dimer and monomer states (please see Chapter 6 for the evidence of the monomer-dimer distribution accessed on removal of the disulphide bond).

Crystal packing of c27-2 shows that the unravelled β2 strand not only lies within the β4-8 groove of a neighbouring dimer, but also implies possible interaction with an independent dimer via hydrogen bonding of their respective unraveled β2 strands (Chapter 3). However, the mass spectrum of c27-2 does not report runaway oligomers, as would be expected from the crystal structure and has been reported for the core domain of αB-crystallin which retained its C-terminus[30]. As discussed above, the mass spectrum of c27-2 at 20 μM monomeric concentration shows a small population of tetramers, likely non-native, but is otherwise devoid of large oligomeric species and rather is dominated by the dimer. Additionally, NMR diffusion experiments undertaken on the protein over the range 0.2 – 1 mM solely detected a dimeric species, and size exclusion chromatography coupled to multi-angle light scattering (SEC-MALS) also detected an overwhelming dimer population at 50 μM to 1 mM. In addition, when mass spectra were collected on a peptide comprising residues 86-96 of c27-2, dimers were only observed at high concentration (300 μM) and low abundance, and did not associate more than a random peptide of the same length. Of course, the largely main chain hydrogen bonding in this region in the crystal structure (Chapter 3) may not be sequence specific. However, based on the evidence discussed in this section, namely that the dimer is the strongly predominant species between 20 μM to 1 mM, we are confident that the runaway domain swapping in the crystal structure is not replicated in solution.

Nevertheless, experiments employing peptides to mimic the β2 did yield some interesting results with regard to its binding to the core domain. These are discussed in the following sections, and arose from our observations of the crystal lattice discussed in Chapter 3.

2.3.1 An aside on an unexpected adduct

It is sometimes possible to see a fourth species in the mass spectrum of c27-2, which corresponds to a mass of 9906 Da. This species shows a similar drift time profile to the monomer and is likely one c27-2 monomer with an adduct of 75 Da. We hypothesised this could correspond to the binding of β-mercaptoethanol (BME, 78.13 Da).

BME could remain after purification (required for activity of TEV cleavage but also included in the IMAC buffers) despite extensive buffer exchange on the SEC column and sometimes by further dialysis or SEC. Unfortunately, when we replaced BME with DTT in the purification protocol, we saw a similar adduction of DTT. This is significant because inclusion of a BME model accounted well for additional electron density around the cysteine in one model of the core domain (Chapter 5), and so it was necessary to mitigate the effect of this phenomenon in our other experiments, both by extensive buffer exchange and by identification of the remaining small population of this species. For example, the sample in Figure 4 – 5 had been prepared as in Section 4.3, but was further incubated with 400 fold excess of DTT for one hour before buffer exchange on a desalting column and equilibration at 37 °C for two hours (though complete air oxidation seems to occur as soon as the protein is buffer exchanged). The incubation with DTT seems to have successfully competed with the bound BME species, with the absence of any peaks corresponding to the 9906 Da component.

Despite this experimental hurdle, it is clear from the mass spectra of the buffer exchanged sample that oxidised c27-2 exists chiefly as a dimer, and it demonstrates that mass spectrometry is well-placed to analyse the constituents of a protein sample.

2.4 Investigating competitive binding between the C-terminus and β2 strand with native MS

In order to test whether the unravelled β2 competes with the C-terminus to bind to the β4 – β8 groove, we designed a competition experiment whereby the binding of the C-terminal peptide (CTP) was monitored in the presence of one of three peptides mimicking the area surrounding the β2-strand ('β2-peps'): a non-phosphorylated WT sequence ('β2-WT'), a scrambled version of this peptide disrupting the GVS motif ('β2-scram') or a β2-pep with phosphorylation of the pS82 site ('β2-pS82'). The peptides extend from residue 80 to 88, to encompass slightly more of the full-length sequence than the c27-2 construct (Figure 4 – 7 and Chapter 3).

If the β2-WT does indeed compete with the CTP, we would expect that it will reduce the number of c27-2+CTP species with a concomitant increase in the number of c27-2+β2-WT species, relative to the same experiment with the β2-scram. Additionally, if phosphorylation affects the strength of the binding of the β2-pep to the β4-8 groove, we would again expect a change in the relative populations of these species. If our hypothesis is true that phosphorylation interrupts oligomerisation by increased competition with the C-terminus, we would expect an even greater reduction in c27-2+CTP species. These predictions assume that

the effective concentration of the integral β2 region of each c27-2 chain is so high that it outcompetes binding of the free β2-pep with β3.

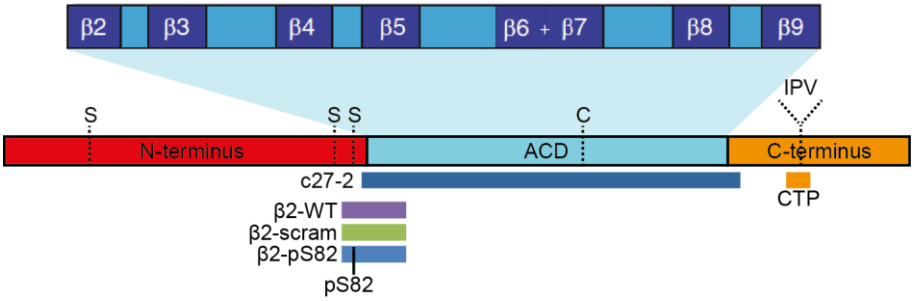


Figure 4 – 7 The synthetic peptides used in this experiment mimic the IPV motif and surrounding sequence in the C-terminus (CTP, Ac-EITIPVTFE$_{178-186}$); the area surrounding the GV residues in the β2 region, including the S82 residue and the first five residues of the c27-2 construct (β2-WT, Ac-QLSSGVSEI$_{80-88}$); a scrambled version of this peptide (β2-scram, Ac-VEGSQSISL); and phosphorylated version of this peptide (β2-pS82, Ac-QLpSSGVSEI$_{80-88}$).

Some representative spectra are shown in Figure 4 – 8, to demonstrate the position of the peaks before describing each full spectrum. The upper panel shows a mass spectrum of a sample of 5 μM c27-2 with 7.5 μM CTP and 40 μM β2-WT (1:3:16 dimer:CTP:β2-pep ratio). The 2400 – 4000 Th region of the mass spectrum is shown in which peaks corresponding to the dimers appear when a reservoir of acetonitrile is placed in the source housing, underneath the nano-electrospray capillary. This causes a reduction in charge state and can also promote the integrity of protein:ligand complexes[31], as discussed briefly in Chapter 1. Each dimer peak (denoted by the charge state) is followed by a series of peaks corresponding to one or two of the β2-WT and CTP binding to the dimer.

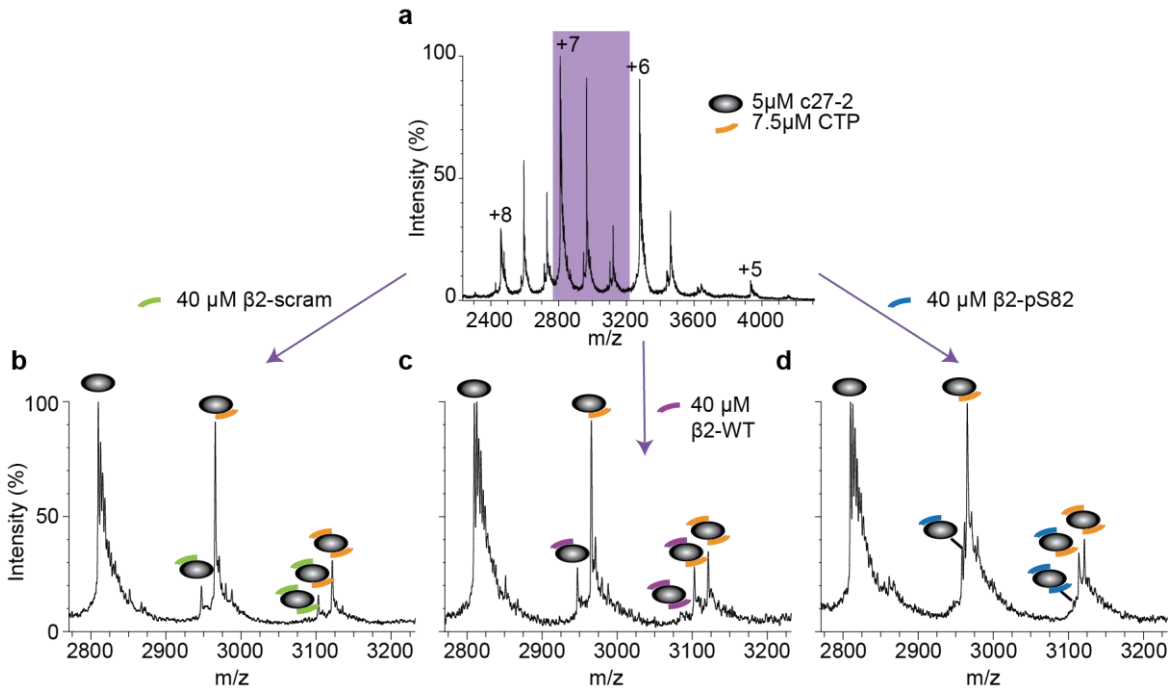


Figure 4 – 8 MS experiment to determine the degree of competitive binding between the C-terminus and β2-region of HSP27. Each spectrum is collected for a sample of 5 μM c27-2: 7.5 μM CTP: 40 μM of (a and c) β2-WT, (b) β2-scram, or (d) β2-pS82. The lower panels show the expanded region of the +7 charge state (as demonstrated for the sample with β2-WT by purple shading, upper panel).

The concentration of the β2-peps used was much higher than the CTP, as the CTP binds much more strongly than the NTP, perhaps not unsurprising as this is the primary interaction with the β4-8 groove typically observed[29,32].

For clarity, the +7 charge state is expanded and displayed for two additional samples containing 40 μM of the β2-scram and β2-pS82 in place of the β2-WT (lower panels). In each spectrum, the dimer peak at 2809 Th is followed by two peaks between 2950 and 2980 Th corresponding to the dimer with one β2-pep or CTP bound, and three peaks between 3080 and 3150 Th for the dimer with two β2-peps, one β2-pep and one CTP, or two CTPs bound. The spectra with β2-WT and β2-scram appear very similar at first glance (this was also the case at lower concentrations of peptide). However, the peaks for the dimer+β2-WT and

dimer+β2-WT+CTP are more intense than their β2-scram counterparts, suggesting the dimer has a greater affinity for β2-WT. This is in keeping with spectra collected in the absence of the CTP which also indicated a greater propensity for the β2-WT to bind than the β2-scram at a six fold excess to the c27-2 dimer. The peaks for dimer+2*β2-WT and dimer+2*β2-scram in this sample are very small, making their quantification more difficult. The most noticeable difference in the three spectra is that the peaks in the β2-pS82 spectrum for the dimer+β2-pS82 and dimer+2*β2-pS82 are positioned close to the following peak corresponding to dimer+CTP and dimer+β2-pS82+CTP respectively. This results from the higher mass of the phosphorylated β2-pS82, requiring careful extraction of the underlying areas.

The synthetic peptides used in this experiment were commercially-sourced and the samples unfortunately retained some of the salt from their preparation. Despite being low in concentration, due to the evaporative reduction of droplet size in the nESI process, these salt and solvent adducts can reach a high concentration in the electrospray droplets as the protein is ionized, and lead to a dispersion of the protein signal over a higher number of masses which each have an additional metal cation attached (sodium is the most prevalent here, with some potassium and water adducts also detectable in the peptide only peaks at low m/z). It is thus necessary to consider the abundance of the additional protein species reported in these extra adduct peaks. Also, the peptides did not easily dissolve in aqueous solution. To improve our quantification of the spectra, we thus varied the concentrations of the constituent molecules and the buffer conditions, and then used a mass spectrum deconvolution algorithm with restraints from manual assignment of the spectra to extract the areas below each peak[28].

Figure 4 – 9 shows spectra collected with 20 µM c27-2: 2µM CTP: 107.5 µM β2-pep in 50mM ammonium acetate pH 6.9. This dimer:CTP:β2-pep ratio of 1:0.2:10.75 gives more equal amounts of CTP-bound and β2-pep-bound dimers, making area extraction more accurate, and produced better resolved spectra after further instrument optimisation. Confident peak assignment was achieved by employing ion mobility MS to separate species that have similar m/z values but different mobilities (as demonstrated for the c27-2 protein alone in Section 2.3).

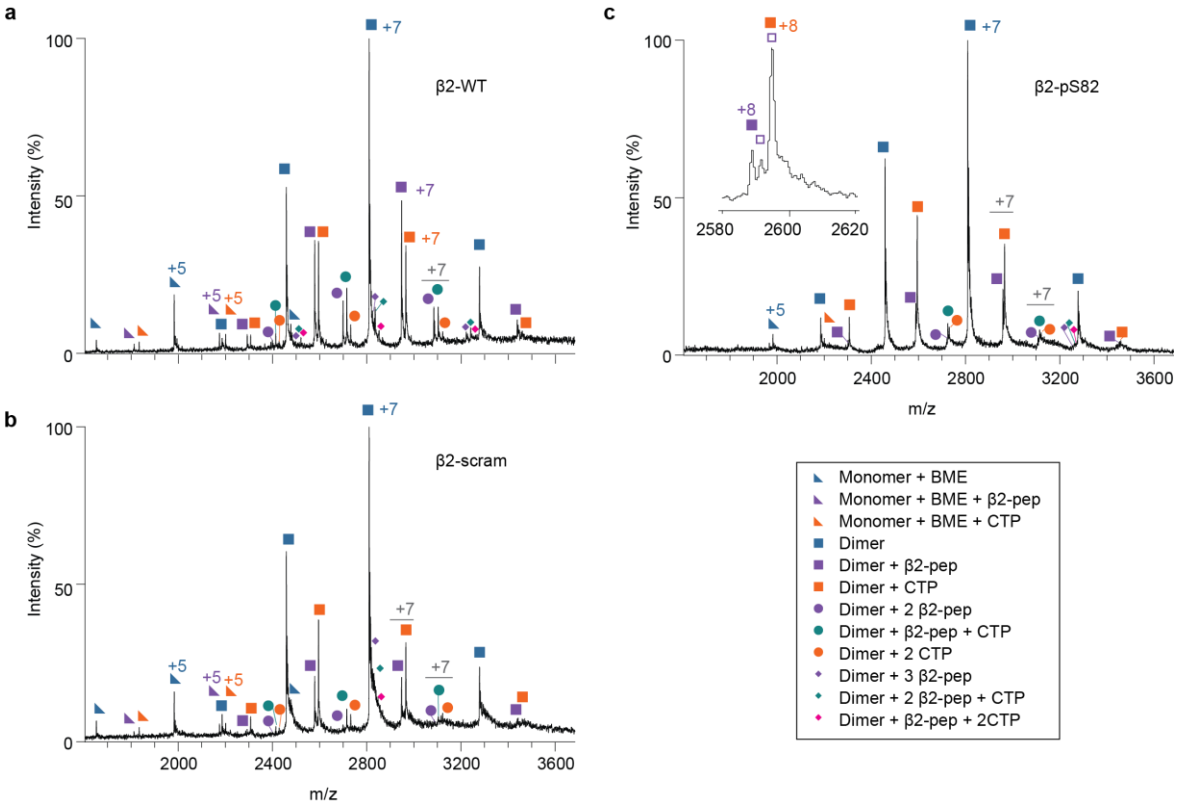


Figure 4 – 9 Competitive binding between each β2-pep and CTP to c27-2. Representative spectra collected with 20 µM c27-2: 2 µM CTP: 107.5 µM of β2-WT (a), β2-scram (b), or β2-pS82 (c). The closer peptide-bound peaks in the β2-pS82 spectrum ((c) inset, +8 charge state) result in some overlap of the salt-adducted species (hollow purple squares) with neighbouring peaks.

The spectrum in the presence of β2-WT (Figure 4 – 9a) again displays a larger peak area (and height) for the dimer+β2-WT (20621.38±9.48 Da), dimer+2*β2-WT (21578.58±7.72

Da) and dimer+β2-WT+CTP (21709.42±9.00 Da) species than the equivalent species with β2-scram (Figure 4 – 9b). This is quantified across the whole charge state distribution in Figure 4 – 10, where each bar shows the mean population of each species across three repeats for each type of β2-pep. The β2-pep bound species are consistently lower with the β2-scram than the β2-WT, except in the case where three peptides are bound. This may be because three peptides bound to the dimer represents non-specific binding if the CTP and β2-pep both bind in the β4-8 groove, or could be a result of the small intensity of these peaks and their proximity to highly intense peaks such as the +7 dimer peak (for the dimer with mass 19659.67±7.56 Da).

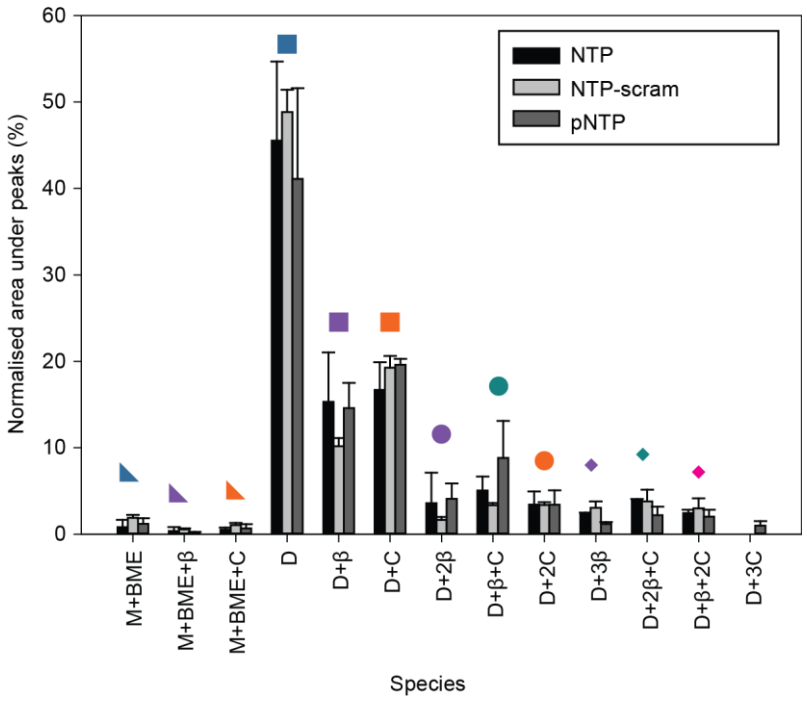


Figure 4 – 10 Extraction of the areas under each charge state series in Figure 4 – 9 reveals the relative abundance of each species. Monomer (M), β2-pep (β), CTP (C) and dimer (D) are denoted in the labels on the abscissa. n = 3 for each type of β2-pep, bar height indicates the mean relative area for the whole charge state distribution for each species in the mass spectrum, error bars indicate the range.

The spectra also contained small peaks corresponding to the c27-2 monomer with an adduct of 73 Da (blue triangles, 9906.23±0.90 Da), attributed to the monomer with the BME adduct (Section 2.3.1). This species seems less able to form the dimer than the unmodified protein based on the absence of a concurrent peak for the dimer+BME or dimer+2*BME species. There is also a small population of this monomer with either the β2-pep or CTP bound.

While the peaks in the β2-pS82 are closer together, it is possible to resolve them and their sodium adducted counterparts (see inset, Figure 4 – 9c). Due to the closer position of the peaks, the peaks with the sodium adducts for one species sometimes overlapped with a neighbouring peak for another species. For example, the peaks in the inset correspond to a dimer+β2-pS82 at 2588.6 Th (purple square, 20704.19±2.81 Da), the same species with a sodium adducted at 2591.4 Th (hollow purple square, 20727.36±5.20 Da), and we would expect the species with two adducted sodiums at 2594.2 Th (hollow purple square). This would overlap with the dimer+CTP at 2594.9 Th (orange square, 20749.96±2.15 Da). As this did not occur with the β2-WT and β2-scram, it was thus necessary to quantify the area in the overlapping peak which belonged to each species. This was achieved by fitting an exponential decay to the quantified dimer+sodium species, and using this to estimate the area of the next adduct peak for each species based on the preceding, non-overlapping peaks (please see Methods for more details).

The extracted areas in Figure 4 – 10 show that there is not a significant difference between the abundance of dimer+β2-pep species with the β2-WT or β2-pS82, nor is there a detectable difference in the abundance of the dimer+CTP species in the presence of any of the β2-peps.

Therefore, this experiment does not support our hypothesis that the interaction between the β2 region and β4-8 groove observed in the crystal structure is present in solution (as the gas phase distribution of species will represent the relative abundance of species present in the solution-phase, even if there is a small amount of gas-phase dissociation). It is possible that there is competition between the strongly binding CTP and weakly binding β2-WT or β2-pS82, but that this is outside the fold-concentrations investigated here. The lower affinity of the β2-pep is consistent with its lack of an IXI motif (in the c27-2 crystal structure, the GV occluding the hydrophobic groove is flanked by hydrophilic residues, Chapter 3) versus this canonical motif within the CTP, even though the neighbouring sequences are able to hydrogen bond to the β4 and β8 strands in both peptides (Chapter 3). With the high β2-pep concentration and concomitant low CTP concentration, the conditions used here are reaching the extremes of those tractable by native MS. At higher peptide concentrations, the salt impurities will become prohibitive. However, we can conclude that in the crystallisation solution, where the ratio of β4-8 binding groove: β2: CTP was 1.00:1.00:1.18 (*cf* 1.00:5.38:0.10 in the MS experiment in Figure 4 – 9), the absence of the CTP from half of the monomers has not occurred due to the competitive binding of the unravelled β2. Rather the behaviour observed in the crystal structure is more strongly governed by the binding affinity of the CTP for the β4-8 groove and the β2 binding takes advantage of the vacant β4-8 grooves due to the CTP concentration used. This β2 conformation may also be artificially enhanced in crystallisation, for example by the positive selection in the screening process on conformations that will pack into ordered crystals. However, the crystal structure did lead us to identify the flexibility of the β2 region, and there is a difference in the binding of the β2-WT and β2-

scram, which though subtle was also corroborated by spectra in the absence of the C-terminus. This suggests that there is a specific interaction of the β2-WT species with the dimer. In order to investigate whether the β2-WT is able to bind the β4-8 at higher fold concentrations, and determine if its binding is enhanced by phosphorylation, we next turned to solution NMR spectroscopy. NMR spectroscopy is able to directly observe the behaviour of individual atoms and their corresponding residues, and is also well-suited to monitoring interactions with K_D values in the mM regime[33].

2.5 NMR studies show binding of the β2 strand is enhanced by phosphorylation

The results in Section 2.4 indicate that an extraneous β2 strand is able to bind the c27-2 dimer, despite the presence of integrated β2 strands on the dimer. In order to investigate this binding further at a higher dimer: β2-pep ratio, and determine if there was any difference on phosphorylation of the peptide, we turned to solution NMR. As ^{1}H-^{15}N heteronuclear single quantum correlation (HSQC) experiments generally produce a spectrum with a peak for each backbone amide (excluding prolines)[33], it is possible to follow changes in the environment of individual residues on addition of a ligand or an increase in protein concentration. It can therefore shed light on protein:protein interactions and protein:ligand interactions[33].

The ^{1}H-^{15}N HSCQ spectrum of 200 μM ^{15}N-labelled reduced c27-2 is shown in blue in Figure 4 – 11a. The spectrum contains correlation cross-peaks for most of the N-H groups within the polypeptide backbone, apart from prolines which lack an amide proton, as well as resonances

for some additional side chain NH groups. Overlaid in green and red are spectra collected under the same conditions in the presence of 2 mM β2-WT, or 2 mM β2-pS82 respectively. Some of the positions of the cross-peaks have shifted (*i.e.* there are chemical shift perturbations) in the presence of the β2-WT in this 20 fold concentration to the dimer. The CSPs are small for ligand binding (the largest is 0.2 ppm *cf* 0.7 ppm on reduction of the disulphide) but indicate that some of the c27-2 residues are experiencing different chemical environments, which implies binding of the extraneous β2-WT. Furthermore, the spectrum in the presence of the phosphorylated β2-pS82 shows different peak profiles for some cross-peaks, for example those at (~8.2 ppm, 115 ppm) and (~9.4ppm, 121 ppm), implying that phosphorylation affects the binding of the peptide.

More importantly, there is also a widespread attenuation of peak intensity on peptide binding. This is represented in Figure 4 – 11b, which shows the intensity of each peak in the presence of the peptide, relative to its intensity in the apo-form. The peaks have been identified through comparison to an assigned ^{1}H-^{15}N HSQC spectrum of oxidised c27-2, where each peak had been assigned to backbone resonances (with a few side chain NH groups) using 3D NMR experiments[34]. This allows the confident identification of the peaks for the reduced sample, which is directly possible because of the small CSPs observed. The relative peak intensity for each residue is shown in the presence of β2-WT, β2-pS82 and β2-scram. In the presence of the 20-fold concentration of β2-WT (10-fold to each monomer), there is a drop in peak intensity for many residues, particularly within the regions 101 – 107 and 122 - 140. A change in peak intensities reports on binding because it indicates exchange between multiple states in the intermediate exchange regime, where the exchange rate coefficient $k_{ex} \approx |\Delta\nu|$, the difference

in chemical shifts[35]. These mutliple states can correspond to the bound and unbound state. Residues 134 – 138 could not be identified due to the low signal to noise of these peaks due to peak broadening, which occurs on reduction of c27-2 and is concomitant with exchange on the micro-second to milli-second timescale and probably monomer-dimer exchange (see Chapter 6 for further discussion).

The binding of β2-WT and β2-pS82 is likely specific as the intensity attentuation is not present with the scrambled peptide β2-scram (Figure 4 – 11b, black). Moreover, it is even more pronounced in the presence of the phosphorylated β2-pep, indicating that phosphorylation strengthens the binding of the β2-pep.

The intensities of residues 84 – 93 have been excluded from this plot for clarity, as their intensity is increased by 1 – 2.5. The increase of intensity (rather than the decrease seen for the rest of the protein) is suggestive of an increase in one particular state of this β2-containing region (rather than exchange between multiple states in the apo c27-2), and is likely due to the increased population of the disordered state of β2 on competition with the β2-pep for the binding site of the bound state.

Together, these results indicate that binding of β2 to c27-2, while weak, is specific. Moreover, phosphorylation increases the strength of this binding.

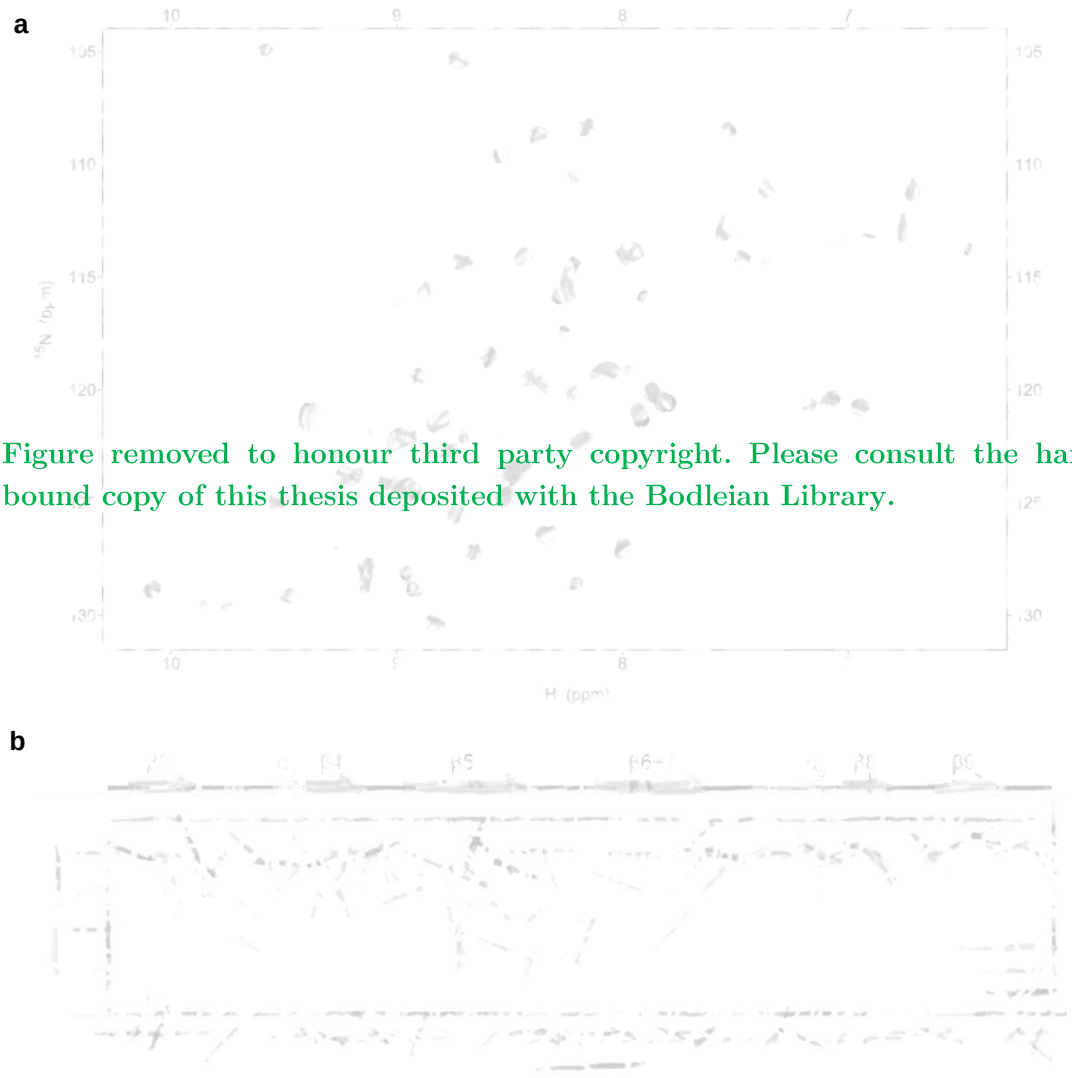


Figure 4 – 11 Phosphorylation modifies the association of the β2 strand as observed in ^{1}H-^{15}N HSCQ spectra. a) Spectra of 200 µM ^{15}N-labelled reduced c27-2 in the apo form (blue) or in the presence of 2 mM β2-WT (green) or 2 mM β2-pS82 (red). Peptide binding results in small CSPs. For example, two peaks in the region (~8.2 ppm, 115 ppm) corresponding to D93 and S98, and the two peaks (~9.4ppm, 121 ppm) corresponding to A169 and K141 (which is involved in a salt bridge across the intra-dimer interface). Spectra were recorded at pH 7.0, 298 K and plotted at the same contour level. Data with the β2-scram excluded for clarity. b) Ratio of peak intensity in the presence of peptide to the intensity in the apo form. Values are calculated from the spectra in (a), with additional data for 200 µM reduced c27-2 with 2mM β2-scram. Peaks were assigned by comparison with an assigned spectrum for the oxidised c27-2. NMR experiments were performed by T. Reid Alderson. Figure courtesy of T.R.A..

2.6 Continuing investigation on the mechanism of phosphorylation

The c27-2 structure presented in the last chapter has led us to question the role of the β2 region in the effect of phosphorylation (specifically, the mechanism by which phosphorylation disrupts oligomerisation and modifies chaperone activity). The results discussed in Sections 2.1 - 2.5 have demonstrated that the β2 accesses a significant disordered state in solution, and that binding of the β2 to the core domain is enhanced by phosphorylation. There remain some important questions regarding this interaction, specifically: Where does the β2 bind on the core domain? How does binding of the β2 affect the oligomerisation of the full-length protein?

We have already made some progress towards answering these questions. Firstly, when we removed the β2 region from the c27-2 construct (residues 93-171, 'c27-4'), we found that there were significant CSPs along the length of the β3 region in the ^{1}H-^{15}N HSQC spectrum, suggesting we may have removed a population of β3-bound β2 as well as the unravelled population. This is in keeping with the higher order parameter for β2 in c27-2 than expected for a completely disordered state (ca. >0.6 vs 0.2, Section 2.2) and the findings in Section 2.5 suggesting the β2 had been displaced from a bound state into a more disordered form.

However, it is not yet conclusive as to where the β2-pep binds. There are widespread ^{1}H-^{15}N HSQC CSPs and intensity drops across the protein when the β2-pep is added in excess, with a notable absence in the β8 strand, refuting our hypothesis of competitive binding between the β2 and the C-terminus to the β4 - β8 groove. The widespread CSPs would be consistent with binding of β2 alongside β3 to form part of the β-sheet. This would be contrary to our previous assumption that the β2 of c27-2 would prevent any association in this region because of the

high effective concentration (Section 2.4), and is possibly because of the high K_D (and low affinity) for the binding of the integral β2.

When the β2-WT or β2-pS82 peptide is added to c27-4 in excess, two peaks (G116 and W95) display complete shifts to their position when in the monomeric form (the c27-4 is observed at low concentration). This suggests that the binding of the peptide is interrupting the monomer ↔ dimer equilibrium to bias the monomeric state, and provides a rationale for the widespread CSPs on peptide binding as being due to disruption of the dimer.

These findings do not preclude domain swapping mediated by the β2 strand, as a solid-state NMR study on full-length αB-crystallin noted an interaction of β3 (D78 and V76) with the β2 of a different dimer (W60, F61 and S59, which is a phosphorylation site of αB-crystallin)[36]. This followed work by the same group where they had shown that the M68 (in the β2 region) had multiple resonances in the solid state NMR spectra of full-length αB-crystallin, indicating that it exists in heterogeneous environments in the oligomer[37].

We have also designed co-crystallisation experiments to try to capture the interaction of the phosphorylated β2-pep with the core domain, including peptides that incorporate either or both of the phosphorylation sites. Additionally, we have designed some longer protein constructs to incorporate phosphomimics for the S78 and S82 sites, to study the effect of phosphorylation on the oligomerisation of the protein. Initial MS experiments have indicated that there is not an increase in higher order oligomers (as could be postulated with domain swapping of the longer β2) but that the reduced double phosphomimic has a higher monomer:dimer ratio than the reduced WT equivalent when prepared under the same

conditions, supporting the finding of an interrupted intra-dimer interface on binding of the phosphorylated β2-pS82.

The results discussed here are relevant more widely than HSP27. For example, phosphorylation of αB-crystallin, which has been studied in much greater detail than HSP27, results in an increased population of odd-numbered oligomers, such that they become comparable to even-numbered stoichiometries[38,39]. This indicates that the intra-dimer interface has been weakened, consistent with our observation of β2-WT and β2-pS82 binding shifting the monomer ↔ dimer equilibrium of reduced c27-4 to increase the population of monomers. Of course, this effect will compete with the stability afforded to this interface by the covalent disulphide bond in oxidised HSP27, whereas the effect on the αB-crystallin intra-dimer interface will be more direct. The regulation of disulphide formation and thus the intra-dimer interface will be discussed in detail in Chapter 5 and Chapter 6. However, it is clear here that modulation of the oxidation state will moderate the effect of phosphorylation on this interface, the strength of this intra-dimer interface is in turn related to the strength of binding of the C-terminus in the β4-8 groove[40]. Indeed, phosphorylation at the three phosphosites on the N-terminus of αB-crystallin leads to an increased rate of subunit exchange (F. D. L. Kondrat, manuscript in preparation), which could be mediated by monomers due to interruption of the intra-dimer interface. However, rather than binding to the α-crystallin domain, phosphorylative oligomer disruption could also involve restructuring of the N-terminus, which has traditionally been suggested as the mechanism for oligomer disruption[7,39,41]. For example, oxidised HSP27 shifts to lower stoichiometries on phosphorylation at S78 and S82 (the effect is more pronounced with the double phosphomimic than the single phosphomimics) though it

retains the disulphide-locked dimeric substructure[7]. Therefore, phosphorylation can interrupt oligomerisation even if the intra-dimer interface is maintained. Taken together, these findings suggest a complex fine-tuning of the interfaces of this protein through regulation and in response to cellular stress.

3 Conclusion

The aim of this chapter has been to test the hypothesis presented in Chapter 3 following the solution of the crystal structure of c27-2, namely the following: that the β2 is able to unravel in solution; it is then able to participate in domain swapping with domains on other dimers as mediated by β2 – β2 interactions and β2 – β4-8 interactions; the β2 is able to compete with the C-terminus for binding to the β4-8 groove; and that phosphorylation moderates the binding of the β2.

MD simulations indicated that the β2-containing region can fully dissociate from the minor β-sheet as well as reside within it, NMR S^2 order parameters showed that the β2-containing region is somewhat disordered in solution. Mass spectra of the intact complexes showed that the dimer constitutes the major population in the oxidised form, and notably did not evidence the runaway oligomers anticipated from the crystal structures. An MS competition assay demonstrated that a peptide containing the β2 region bound specifically to the dimer (versus the scrambled peptide), including when it was phosphorylated, though it did not show that binding of the CTP had been disrupted or distinguish a stoichiometric ratio difference between the phosphorylated and non-phosphorylated forms. However, when the dimer-peptide

system was interrogated at a higher excess of peptide by ^{1}H-^{15}N HSQC NMR, phosphorylation led to a greater (though subtle) perturbation in the peak chemical shifts and intensity attenuation than the non-phosphorylated form, indicating stronger binding.

Therefore, while we do not find any evidence for domain swapping or competition between the C-terminus and β2-pep, our data do support a model of phosphorylative regulation mediated by binding of the region surrounding β2. Binding of the β2-pS82 appears to interrupt the strength of the intra-dimer interface, indicating allostery between these two sites as well as the allostery between the intra-dimer interface and the binding of the C-terminus to the core domain[40], suggesting a fine-tuning between these multiple interfaces by the cellular conditions of pH, temperature, oxidation and phosphorylation.

Of course, ultimately the effect of phosphorylation must be studied within the context of the full length (FL) protein. However, as FL HSP27 is a difficult target for protein NMR, introduces substantial complexity to the mass spectra, and is possibly intractable by protein crystallography, the core domain systems presented here have proven effective models in the study of HSP27 structure and regulation. The work undertaken on this system will hopefully guide and be transferable to studies of the full-length protein.

sHSP phosphorylation has been an important area of study, as it demonstrates that the cellular machinery is able to exercise control over sHSP structure and activity, rather than experience a solely passive modulation of sHSP behaviour. The observations I have described in this chapter are starting to suggest a structural rationale for the oligomeric and functional consequences of this modification in the dynamic N-terminus. This has led to multiple

hypotheses that can be tested in the future, and represents an exciting opportunity to probe the structure-function relationship of these highly responsive proteins.

4 Methods

4.1 Materials and reagents

All reagents were purchased from Sigma unless otherwise specified.

4.2 Constructs

HSP27 c27-2, c27-3, c27-4 were amplified from the c27-1 plasmid, the longer constructs briefly discussed in Section 0 were amplified from an HSP27 WT plasmid (generous gift of Heath Ecroyd, University of Wollongong). Amplification was undertaken with polymerase chain reaction (PCR) (primers purchased from Integrated DNA Technologies; Phusion polymerase from Agilent), with introduced serine-aspartate mutations at S78 and S82 for the longer constructs in Section 0. After gel separation and purification, the gene fragments were introduced to a modified pET28a vector (6H-TEV-//gene//) digested with BamHI and XhoI (New England Biolabs, U.S.A.), using the Infusion HD Cloning kit (Clontech). Transformation was performed in Stellar competent cells (Clontech) and cultures grown in Luria Broth (Fisher) with kanamycin (50 µg/ml) ready for harvest (MiniPrep kit, ThermoFisher). Successful mutation was determined by plasmid sequencing (SourceBioScience).

4.3 Protein expression and purification

BL21 (DE3) Gold cells (Agilent) were transformed with the pET28 vector 6H-TEV-[c27-2]; 6H-TEV-[c27-3] or 6H-TEV-[c27-4] etc. 1l cultures were grown in Luria Broth (Fisher) with kanamycin (50 µg/ml) at 37°C until OD_{660} = 0.6-0.8, then induced with IPTG at 500 µM for three hours. Cells were harvested by centrifugation, washed in PBS and stored at -80°C. Cell pellets were resuspended and disrupted by sonication or microfluidisation in Buffer A (50 mM Tris, 300 mM NaCl, 20 mM imidazole, 5 mM BME, pH 8) supplemented with 1 EDTA-free protease inhibitor cocktail tablet (one per 1L pellet, Roche), followed by clarification by centrifugation. The first purification step was immobilised metal affinity chromatography (IMAC) conducted on a HisTrap HP column and Akta FPLC system (both GE Life Sciences) in Buffer A with elution in Buffer B (50 mM Tris, 300 mM NaCl, 500 mM imidazole, 5 mM BME, pH 8), followed by a desalting column (HiPrep, GE Life Sciences) or overnight dialysis in Buffer DS (20 mM Tris, 150 mM NaCl, 20 mM imidazole, 5 mM BME, pH 8). Overnight cleavage with TEV protease (S219V mutant purified in-house, pRK793, Addgene plasmid #8827) was conducted at room temperature and followed by a reverse IMAC step in Buffer A and then size exclusion chromatography (SEC) in SEC Buffer (20 mM Tris, 150 mM NaCl, pH 8) on a Superdex 75 column 26/60 (GE Life Sciences).

4.4 MD simulations

Four simulations were prepared using the c27-1 crystal structure (PDB 4MJH), with missing residues on chain B reconstructed from the coordinates of the corresponding atoms on chain

A. The four simulations comprised dimeric c27-1 with and without the disulphide, in either the oxidised or reduced state (Figure 4 – 1). Each system was immersed in a TIP3P water box, and neutralized with 0.15 M of Na^+ and Cl^- ions.

All systems were simulated using the Amber99SB force field[42] on the NAMD2.9[43] MD engine, with the SHAKE algorithm constraining heavy atom distances, and PME treating the electrostatic interactions in periodic boundary conditions.

All the systems were first minimized with 2000 conjugate gradient steps, and were subsequently simulated using a 2 fs time-step. Systems were first equilibrated in the nPT ensemble for 0.5 ns, with a 10 $kcal.mol^{-1}$ constraint on the C_α atoms. Conditions of 300K and 1 atm were imposed with Langevin dynamics, using a damping constant of 1 /ps, a piston period of 200 fs, and a piston decay of 50 fs. Constraints were subsequently removed, and the systems were simulated in the nVT ensemble at 300K for 1 ns. Finally, production runs for all four systems were run in the nPT ensemble (300K and 1 atm) for at least 0.5 µs.

4.5 Mass spectrometry

4.5.1 Oligomeric distribution

The spectra in Section 2.3 were collected on c27-2 as prepared in Section 4.3, but with further incubation with 400 fold excess of DTT for one hour before buffer exchange on a desalting column into 200 mM ammonium acetate pH 6.9 and equilibration at 37 °C for two hours. The spectra were collected on 20 µM of sample using the gold-plated capillaries prepared in-house[22] on a Synapt G1 mass spectrometer in positive mode and ion-mobility mode (Waters, UK).

The spectrum in Figure 4 – 5 was collected with the following parameters: 1.4 kV capillary voltage; 150 V sample cone; 4 V extraction cone; 5 V trap; 4 V transfer; 2 $mlmin^{-1}$ flow and 0.03 mbar pressure of the trap and transfer gas (argon); 250 ms^{-1} IMS wave velocity; 6 V IMS wave height; 22 $mlmin^{-1}$ flow and 0.5 mbar pressure of the IMS drift gas (nitrogen); 0.05 mbar in the quadrupole and ion guides. The spectrum in Figure 4 – 6 was collected under the same conditions except 10V sample cone, and 5.5 IMS wave height. Spectra were calibrated in Waters MassLynx software using a CsI calibrant spectrum.

4.5.2 Peptide competition experiments

The synthetic peptides were purchased from BioMatik at >98% purity, all with acetylation on the N-terminus as the N-terminal glutamine of β2-WT was likely to be unstable (advice gratefully received from David Vraj, BioMatik). In the spectra in Figure 4 – 8 the protein and peptides were prepared in 200 mM ammonium acetate pH6.9. c27-2 was prepared as described in Section 4.3, then buffer exchanged into 200 mM ammonium acetate pH 6.9 with a BioRad Biospin P-30 column and equilibrated for 5 hours at 37°C. The peptides were dissolved in 200 mM ammonium acetate pH 6.9 and incubated overnight at 37°C with agitation. The CTP and β2-pep were less soluble than the β2-scram and the β2-pS82, despite the exact amino acid identity of β2-pep and β2-scram, and so were heated further at 40°C for several hours. The concentration of c27-2 was measured by absorbance at 280 nm with the theoretical extinction coefficient 8480 Mcm^{-1} and the peptide concentration determined from the initial mass of the lyophilized powder.

The mass spectra were collected on a Synapt G1 QToF mass spectrometer (Waters, UK) in positive, ion mobility mode with the following parameters: 1.2 kV capillary voltage; 150 V sample cone; 4 V extraction cone; 20 V trap; 5 V transfer; 6 $mlmin^{-1}$ flow and 0.06 mbar pressure of the trap and transfer gas (argon); 250 ms^{-1} IMS wave velocity; 7 V IMS wave height; 22 $mlmin^{-1}$ flow and 0.6 mbar pressure of the IMS drift gas (nitrogen); 0.1 mbar pressure in the quadrupole and RF ion guide, 4.9 mbar backing pressure. Gold-plated capillaries were prepared in-house[22] and a reservoir of acetonitrile was placed in the source region[31]. Spectra were calibrated in Waters MassLynx software using a CsI calibrant spectrum.

For the spectra in Figure 4 – 9, peptides were dissolved in 50mM ammonium acetate pH 6.9, and quantified with a BCA Assay (Thermo Pierce), except in the case of the CTP whose low concentration was outside of the dynamic range. The concentration of this was thus estimated from the mass of the initial lyophilized powder, and the same solution was used across all the samples for fair comparison. c27-2 was prepared as described in Section 4.3, then buffer exchanged into 50 mM ammonium acetate pH 6.9 with a BioRad Biospin P-30 column.The mass spectra were collected as above with modification to the following parameters: 1.4 kV capillary voltage; 40 V sample cone; 40 V trap; 4 V transfer; 1.5 $mlmin^{-1}$ flow and 0.03 mbar pressure of the trap and transfer gas (argon); 4.5 mbar backing pressure, 0.03 µbar pressure in the quadrupole and RF ion guide.

The spectra were assigned in MassLynx using the 2D data (m/z and drift time dimensions). This assignment was then used to restrain the deconvolution algorithm in the UniDec software to search for the assigned species including additional sodiums (using the Oligomer Tool) in

the 1D data (m/z dimension) and the peak areas were extracted using the Data Collector function. For the spectra with the β2-pS82, the areas under the overlapping peaks were apportioned by modelling an exponential decay to each dimer+n*peptide species and its sequential sodium adducted species. This exponential was derived from the dimer species, which did not overlap in m/z space with any other species, and had up to six adducted sodiums ($y = 0.1705e^{-0.433x}$ where y is the peak height and x-1 is the number of adducts).

4.6 NMR spectroscopy

c27-2 samples isotopically labelled uniformly with ^{15}N alone were prepared by growing BL21 (DE3) Gold E.Coli cells transformed with the pET28 6H-TEV-c27-2 plasmid in M9 minimal medium. ^{15}N ammonium chloride was used as the sole nitrogen source with 4 g/L of ^{12}C-glucose. The cells were induced with IPTG, harvested and purified as described in Section 4.3 before buffer exchange into 30 mM NaH_2PO_4, 2 mM EDTA, 2 mM NaN_3, pH 7.0 (NMR buffer) on a Superdex S75 26/60 or 10/300 (GE Life Sciences).

Uniformly ^{13}C and ^{15}N labelled c27-2 for the 3D resonance assignments and ^{2}H fractionally, ^{15}N uniformly labelled c27-2 for the CPMG experiments were produced similarly. Insight from those experiments that informed the work herein are mentioned where relevant in the text.

All experiments were recorded on an 11.7 T or a 14.1 T Varian-Inova spectrometer equipped with a 5 mm *z*-axis gradient triple resonance, room temperature probe, and the spectra were processed with NMRPipe[44] and visualized in Sparky (T. D. Goddard and D. G. Kneller, SPARKY 3, University of California, San Francisco).

The 2D ^{1}H-^{15}N heteronuclear single-quantum coherence (HSQC) spectra for c27-2 and c27-4 were recorded with eight scans and 128 increments. The spectra shown in Section 2.5 are with 200 µM ^{15}N-labelled reduced c27-2 (5 mM BME) in the apo form or in the presence of 2 mM β2-pep, at 298 K. The spectra discussed in Section 0 were on ^{15}N-labelled reduced c27-4 (5 mM BME) with addition of β2-pS82 in a 200 µM: 2 mM ratio or 50 µM: 5.6 mM ratio.

The ^{1}H chemical shift temperature coefficients mentioned in Section 2.2 were determined from ^{1}H-^{15}N HSQC spectra at 290, 293, 295, and 298 K.

The ^{15}N heteronuclear Overhauser enhancements (hetNOEs), longitudinal (T_1, the inverse of R_1), and transverse (T_2, the inverse of R_2) relaxation rates were determined using standard pulse sequences at 14.1 T, with R_1 and R_2 data also acquired at 11.7 T. The peak fitting and R_1 and R_2 determination were performed using the FuDA program.

Generalized order parameters (S^2) for c27-2 N-H bond vectors were calculated with model-free analysis using ModelFree4.1. An axially symmetric diffusion model of c27-2 was used to determine a global correlation time (τ_c) of 13.65 ns and $D_{\parallel}/D_{\perp} = 1.52$. These values were used alongside the c27-2 crystal structure in the model-free analysis.

5 References

1. Ubersax, J. A. & Ferrell Jr, J. E. Mechanisms of specificity in protein phosphorylation. *Nat. Rev. Mol. Cell Biol.* **8,** 530–541 (2007).

2. Haslbeck, M. & Vierling, E. A First Line of Stress Defense: Small Heat Shock Proteins and Their Function in Protein Homeostasis. *J. Mol. Biol.* **427,** 1537–1548 (2015).

3. Garrido, C. Size matters: of the small HSP27 and its large oligomers. *Cell Death Differ.* **9,** 483–5 (2002).

4. Bryantsev, A. L., Chechenova, M. B. & Shelden, E. A. Recruitment of phosphorylated small heat shock protein Hsp27 to nuclear speckles without stress. *Exp. Cell Res.* **313,** 195–209 (2007).

5. Hayes, D., Napoli, V., Mazurkie, A., Stafford, W. F. & Graceffa, P. Phosphorylation Dependence of Hsp27 Multimeric Size and Molecular Chaperone Function. *J. Biol. Chem.* **284,** 18801–18807 (2009).

6. Rogalla, T. *et al.* Regulation of Hsp27 Oligomerization, Chaperone Function, and Protective Activity against Oxidative Stress/Tumor Necrosis Factor by Phosphorylation. *J. Biol. Chem.* **274,** 18947–18956 (1999).

7. Jovcevski, B. *et al.* Phosphomimics Destabilize Hsp27 Oligomeric Assemblies and Enhance Chaperone Activity. *Chem. Biol.* **22,** 186–195 (2015).

8. Haslbeck, M., Franzmann, T., Weinfurtner, D. & Buchner, J. Some like it hot: the

structure and function of small heat-shock proteins. *Nat. Struct. Mol. Biol.* **12,** 842–846 (2005).

9. Rajasekaran, N. S. *et al.* Human αB-Crystallin Mutation Causes Oxido-Reductive Stress and Protein Aggregation Cardiomyopathy in Mice. *Cell* **130,** 427–439 (2007).

10. Baldwin, A. J., Lioe, H., Robinson, C. V., Kay, L. E. & Benesch, J. L. P. αB-Crystallin Polydispersity Is a Consequence of Unbiased Quaternary Dynamics. *J. Mol. Biol.* **413,** 297–309 (2011).

11. Stengel, F. *et al.* Quaternary dynamics and plasticity underlie small heat shock protein chaperone function. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 2007–12 (2010).

12. Benesch, J. L. P. *et al.* The Quaternary Organization and Dynamics of the Molecular Chaperone HSP26 Are Thermally Regulated. *Chem. Biol.* **17,** 1008–1017 (2010).

13. Aquilina, J. A., Shrestha, S., Morris, A. M. & Ecroyd, H. Structural and functional aspects of hetero-oligomers formed by the small heat shock proteins αB-crystallin and HSP27. *J. Biol. Chem.* **288,** 13602–9 (2013).

14. Rajagopal, P., Liu, Y., Shi, L., Clouser, A. F. & Klevit, R. E. Structure of the α-crystallin domain from the redox-sensitive chaperone, HSPB1. *J. Biomol. NMR* 0–5 (2015). doi:10.1007/s10858-015-9973-0

15. Almeida-Souza, L. *et al.* Increased monomerization of mutant HSPB1 leads to protein hyperactivity in Charcot-Marie-Tooth neuropathy. *J. Biol. Chem.* **285,** 12778–86 (2010).

16. Kuzmanic, A. & Zagrovic, B. Determination of Ensemble-Average Pairwise Root Mean-Square Deviation from Experimental B-Factors. *Biophys. J.* **98,** 861–871 (2010).

17. Hochberg, G. K. A. & Benesch, J. L. P. Dynamical structure of αB-crystallin. *Prog. Biophys. Mol. Biol.* **115,** 11–20 (2014).

18. Lipari, G. & Szabo, A. Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and range of validity. *J. Am. Chem. Soc.* **104,** 4546–4559 (1982).

19. Kay, L. E., Torchia, D. A. & Bax, A. Backbone dynamics of proteins as studied by 15N inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease. *Biochemistry* **28,** 8972–8979 (1989).

20. Wishart, D. S. Interpreting protein chemical shift data. *Prog. Nucl. Magn. Reson. Spectrosc.* **58,** 62–87 (2011).

21. Cierpicki, T. & Otlewski, J. Amide proton temperature coefficients as hydrogen bond indicators in proteins. *J. Biomol. NMR* **21,** 249–261 (2001).

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

23. Benesch, J. L. P. & Ruotolo, B. T. Mass spectrometry: come of age for structural and dynamical biology. *Curr. Opin. Struct. Biol.* **21,** 641–9 (2011).

24. Benesch, J. L. P. Collisional activation of protein complexes: Picking up the pieces. *J.*

Am. Soc. Mass Spectrom. **20,** 341–348 (2009).

25. Kondrat, F. D. L., Struwe, W. B. & Benesch, J. L. P. in *Struct. Proteomics High-Throughput Methods* (ed. Owens, R. J.) **1261,** 349–371 (Springer New York, 2015).

26. Benesch, J. L. P., Ruotolo, B. T., Simmons, D. A. & Robinson, C. V. Protein complexes in the gas phase: technology for structural genomics and proteomics. *Chem. Rev.* **107,** 3544–3567 (2007).

27. Konijnenberg, A., Butterer, A. & Sobott, F. Native ion mobility-mass spectrometry and related methods in structural biology. *Biochim. Biophys. Acta - Proteins Proteomics* **1834,** 1239–1256 (2013).

28. Marty, M. T. *et al.* Bayesian Deconvolution of Mass and Ion Mobility Spectra: From Binary Interactions to Polydisperse Ensembles. *Anal. Chem.* **87,** 4370–4376 (2015).

29. Hochberg, G. K. A. *et al.* The structured core domain of αB-crystallin can prevent amyloid fibrillation and associated toxicity. *Proc. Natl. Acad. Sci. U. S. A.* **111,** E1562–70 (2014).

30. Laganowsky, A. *et al.* Crystal structures of truncated alphaA and alphaB crystallins reveal structural mechanisms of polydispersity important for eye lens function. *Protein Sci.* **19,** 1031–43 (2010).

31. Hopper, J. T. S., Sokratous, K. & Oldham, N. J. Charge state and adduct reduction in electrospray ionization-mass spectrometry using solvent vapor exposure. *Anal. Biochem.* **421,** 788–790 (2012).

32. Delbecq, S. P., Jehle, S. & Klevit, R. Binding determinants of the small heat shock protein, αB-crystallin: recognition of the 'IxI' motif. *EMBO J.* **31,** 4587–4594 (2012).

33. Goldflam, M., Tarragó, T., Gairí, M. & Giralt, E. in *Protein NMR Tech.* (eds. Shekhtman, A. & Burz, D. S.) **831,** 233–259 (2012).

34. Sattler, M., Schleucher, J. & Griesinger, C. Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. *Prog. Nucl. Magn. Reson. Spectrosc.* **34,** 93–158 (1999).

35. Kleckner, I. R. & Foster, M. P. An introduction to NMR-based approaches for measuring protein dynamics. *Biochim. Biophys. Acta - Proteins Proteomics* **1814,** 942–968 (2011).

36. Jehle, S. *et al.* Solid-state NMR and SAXS studies provide a structural basis for the activation of αB-crystallin oligomers. *Nat. Struct. Mol. Biol.* **17,** 1037–1042 (2010).

37. Jehle, S. *et al.* αB-Crystallin: A Hybrid Solid-State/Solution-State NMR Investigation Reveals Structural Aspects of the Heterogeneous Oligomer. *J. Mol. Biol.* **385,** 1481–1497 (2009).

38. Ecroyd, H. *et al.* Mimicking phosphorylation of αB-crystallin affects its chaperone activity. *Biochem. J.* **401,** 129–141 (2007).

39. Aquilina, J. A. *et al.* Phosphorylation of αB-Crystallin Alters Chaperone Function through Loss of Dimeric Substructure. *J. Biol. Chem.* **279,** 28675–28680 (2004).

40. Hilton, G. R. *et al.* C-terminal interactions mediate the quaternary dynamics of αB-crystallin. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **368,** 20110405 (2013).

41. Thériault, J. R. *et al.* Essential role of the NH2-terminal WD/EPF motif in the phosphorylation-activated protective function of mammalian Hsp27. *J. Biol. Chem.* **279,** 23463–71 (2004).

42. Hornak, V. *et al.* Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins Struct. Funct. Bioinforma.* **65,** 712–725 (2006).

43. Phillips, J. C. *et al.* Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **26,** 1781–802 (2005).

44. Delaglio, F. *et al.* NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **6,** 277–293 (1995).

Chapter 5: The Effect of Mutation on the Core Domain of HSP27

1 Introduction

1.1 Mutations to the HSP27 core domain have been identified in motor neuropathy

In 2004, Evgrafov *et al.* identified five independent missense mutations of HSP27 (HSPB1) in patients with Charcot-Marie-Tooth disease Type II (CMT2), and distal hereditary motor neuropathies (distal HMN). Both of these diseases have heterogeneous genetic and clinical profiles, but their phenotypes do exhibit similar characteristics: progressive degeneration of the peripheral nerves; distal muscle weakness and atrophy; and reduced tendon reflexes. Additionally, CMT2 patients exhibit sensory abnormalities in the distal limbs. Evgrafov and colleagues used a haplotype analysis to identify the position of the disease-linked locus[1] in a family with CMT2, followed by sequencing of the genes at that position in CMT2 and distal HMN patients[2]. This identified the R127W, S135F, R136W and T151I mutations, which are all within the α-crystallin domain of HSP27, and the P182L mutation of the IPV motif involved in the reciprocal binding of the C-terminus. Only one mutation in HSPB1 was identified in each affected family; they did not appear together. The disease-linked residues were well conserved in HSP27 orthologs in a range of higher eukaryotes[2].

Additionally, in 2011 Lin *et al.*[3] reported a fifth mutation associated with motor neuropathy within the α-crystallin domain, T164A. This was identified in a Taiwanese Han Chinese male, who was heterozygous for the A>G nucleotide mutation (amino acid T164A missense mutation). There they also demonstrated the conservation of T164 among HSPB1 orthologs in mammals and higher order eukaryotes, and described the symptoms of Charcot–Marie-Tooth disease Type 2 in the patient and his family. These included slowly progressive weakness, mild to severe muscle atrophy, and mild to severe sensory loss in the distal limbs leading to multiple injuries and ulcers.

The c27-2 core domain contains the sites of four of the five mutations identified by Evgrafov *et al.*[2] and the T164A mutation identified by Lin *et al.*[3] It is thus an ideal system with which the effect of the mutations on the core domain can be determined.

This chapter discusses the result of crystallisation trials of the c27-2 ACD-peptide system with the individual mutations introduced. Structures for three of the mutants were determined, and their characteristics are discussed here. The most notable difference between the structures was a change in the register of the inter-dimer interface in the R136W c27-2 structure. Though it is well documented for αB-crystallin, this is the first time that this extended registration has been reported for HSP27 and thus the first observation where the C137 residues are not able to form the inter-subunit disulphide. This caused us to question the role of the disulphide bond within the inter-subunit interface, especially as its position is very rare within proteins in the Protein Data Bank[4].

2 Results and Discussion

2.1 Structural models of the core domain with R136W, T151I and T164A mutations

Mutation	Disease
(WT)	-
R127W	HMN2B
S135F	CMT2F and HMN2B
R136W	CMT2F
T151I	HMN2B
T164A	CMT2F

Table 5 – 1 Mutations of HSP27 in the core domain associated with motor neuropathy[2,3]. CMT2F – Charcot-Marie-Tooth disease Type II, with the mutation locus on chromosome 7q11-q21. HMN2B - distal hereditary motor neuropathy.

In order to probe the structural effect of mutation on the core domain of HSP27, five of the mutations identified in motor neuropathy were introduced to the c27-2 system (Table 5 – 1). The mutations were introduced into modified pET28a carrying the c27-2 gene with an N-terminal histidine tag and a TEV protease cleavage site under a T7 promoter. Proteins were expressed in *E. Coli*, purified with immobilised metal affinity chromatography with tag cleavage, and size exclusion chromatography. The mutated c27-2 genes were codon optimised for expression in *E. Coli*, so these experiments are not sensitive to changes in co-translational folding resulting from the nucleotide sequence level.

Purified protein was subjected to crystallisation trials with 14.8 mg/ml of protein and 1.78 mM of the C-terminal peptide with five commercial crystallisation screens. All five of the

mutants produced hits within the screen, which were optimised to produce crystals for data collection.

Unfortunately the R127W and S135F c27-2 crystals did not produce suitable diffraction data, and so these targets were not pursued further. However, optimisation of the crystallisation of R136W, T151I and T164A produced diffracting crystals and the data is discussed below. All three structures were refined after molecular replacement using the c27-2 WT as a phasing model.

The fifth mutation identified in the 2004 study discussed in the Introduction was P182L[2], which lies within the IPV motif of the C-terminus (Chapter 4). P182L and P182S[5] mutations have both been associated with distal HMN. The P182L/S mutation could also be studied within the ACD-peptide system employed here, as it lies within the C-terminal peptide. Unfortunately however, low solubility of the synthetic peptide containing the P182L mutation prevented its investigation. Nonetheless, we are able to draw some insight from the crystal structure of the WT (Chapter 4). The hydrophobic and polar nature of the P182L and P182S substitutions respectively imply that it is not the gain of interactions but rather the loss of the proline that mediates the detrimental effect. Alongside consideration of the crystal structure, which captures the proline side chain pointing out from the molecule (**Figure 5 – 1**) such that its steric restriction seemingly facilitates binding of isoleucine and valine in the hydrophobic groove and hydrogen bonding of the flanking C-terminal residues to the β4 and β8 strands, this suggests that the unfavourable change on replacement of the proline is the increased rotational freedom afforded to the C-terminus. Indeed, in a study where the residues of the C-

terminus of full-length αB-crystallin were systematically mutated to alanine[6], the equivalent proline to P182 in the IPI motif was found to have the greatest effect on the strength of the interaction between the C-terminus and β4-8 groove, indicating its critical role in C-terminal binding.

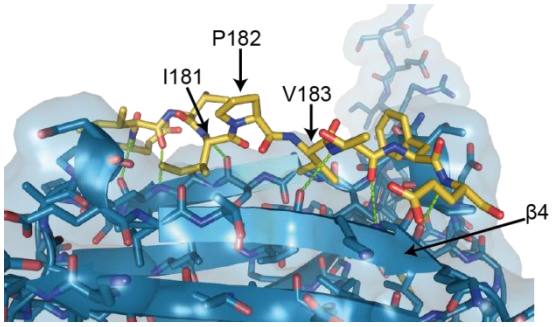


Figure 5 – 1 Proline 182 acts as a scaffold for binding of the C-terminus on chain A of c27-2 (blue, cartoon and stick representation with surface). The CTP (yellow) binds across the β4-β8 groove with the sterically restricted proline pointing outwards from the interface. The I181 and V183 side chains are buried within the hydrophobic groove between the β4 and β8 strands, and the remaining C-terminal residues form six main chain hydrogen bonds with residues on the β4 and β8 strands (light green).

Similarly, the WT structure can shed light on the effect of the R127W and S135F mutations. In the c27-2(WT) crystal, the R127 forms part of the β5 strand (forming main chain hydrogen bonds with I134). Its side chain does not partake in a direct interaction with any other residues, rather it hydrogen bonds to a sulphate molecule through which it is part of a widespread hydrogen bonding network involving side chains from several parts of the protein (the sulphate could be substituted by another hydrogen-bond acceptor, *e.g.* water, in the cellular environment). For example, the sulphate is also bound by H103 on the β3-β4 loop connecting the minor and major sheets of the β-sandwich; and hydrogen bonds to a water that interacts with R136 at the interface and through D129 to R140' on the partner dimer (Figure

5 – 2). Additionally, the R127 secondary amine may be able to hydrogen bond to a water that in turn hydrogen bonds E108 between the 3_{10} helix and β4. Introducing the steric bulk of tryptophan with the R127W mutation may interrupt the hydrogen bonding network and may therefore exert influence on the dimer interface as well as the β3-4 loop. Alternatively, it may lead to strain on the β5 and β6+7 strands as the W127 main chain twists under the influence of the hydrophobic side chain.

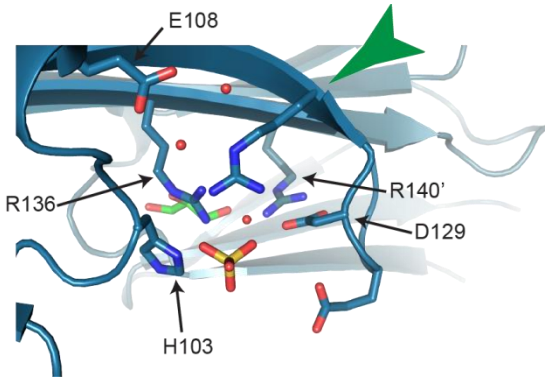


Figure 5 – 2 The R127 sidechain forms part of a widespread hydrogen-bonding network which involves residues from the β3-β4 loop, at the intra-dimer interface and in the partner monomer. The R127W mutation (indicated by green arrow) will interrupt this network. Sulphate and glycerol observed in the crystal are shown as sticks in yellow and green respectively, water is shown as red spheres.

The S135 in c27-2(WT) forms main chain hydrogen bonds with T139' in the monomer partner, as part of the intra-dimer interface (Figure 5 – 3). The S135 sidechain can hydrogen bond to a nearby water but unlike R127 it is not involved in an extensive hydrogen-bond network. It is most likely that a bulky and hydrophobic phenylalanine substitution (S135F) will cause steric repulsion and exclusion of water at this interface, possibly affecting the strength of the hydrogen bonding across the intra-dimer interface.

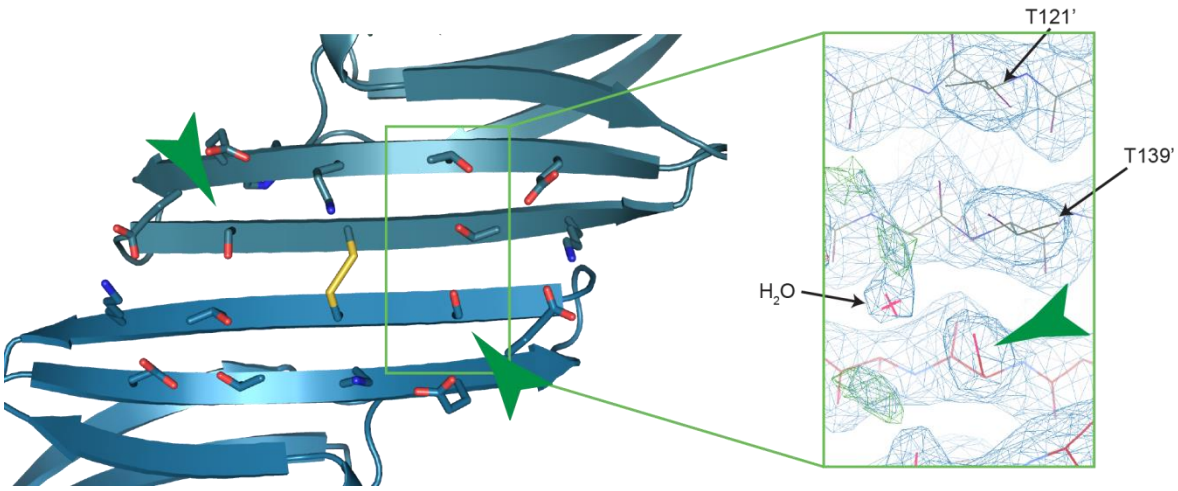


Figure 5 – 3 S135 forms part of the intra-dimer interface through hydrogen-bonding of its main chain to T139'. Its side chain is able to hydrogen bond a water (inset). The S135F mutation would introduce a bulky hydrophobic group on this face of the β-sheet, which is otherwise dominated by polar interactions.

2.1.1 T164A interrupts a hydrogen bonding network across the minor β-sheet

The T164A structural model is very similar to the WT (Figure 5 – 4). It crystallised with the same packing which is consistent with a tetragonal P 4_3 2_1 2 space group, at a resolution of 2.9 Å. It is presented in Figure 5 – 4 where the bound C-terminal peptide and extended β2 on chain A are evident. This region binds into the β4-β8 groove of a symmetry mate through the GV motif, as observed for the WT. This occurs in the absence of a second C-terminal peptide in the groove of chain B.

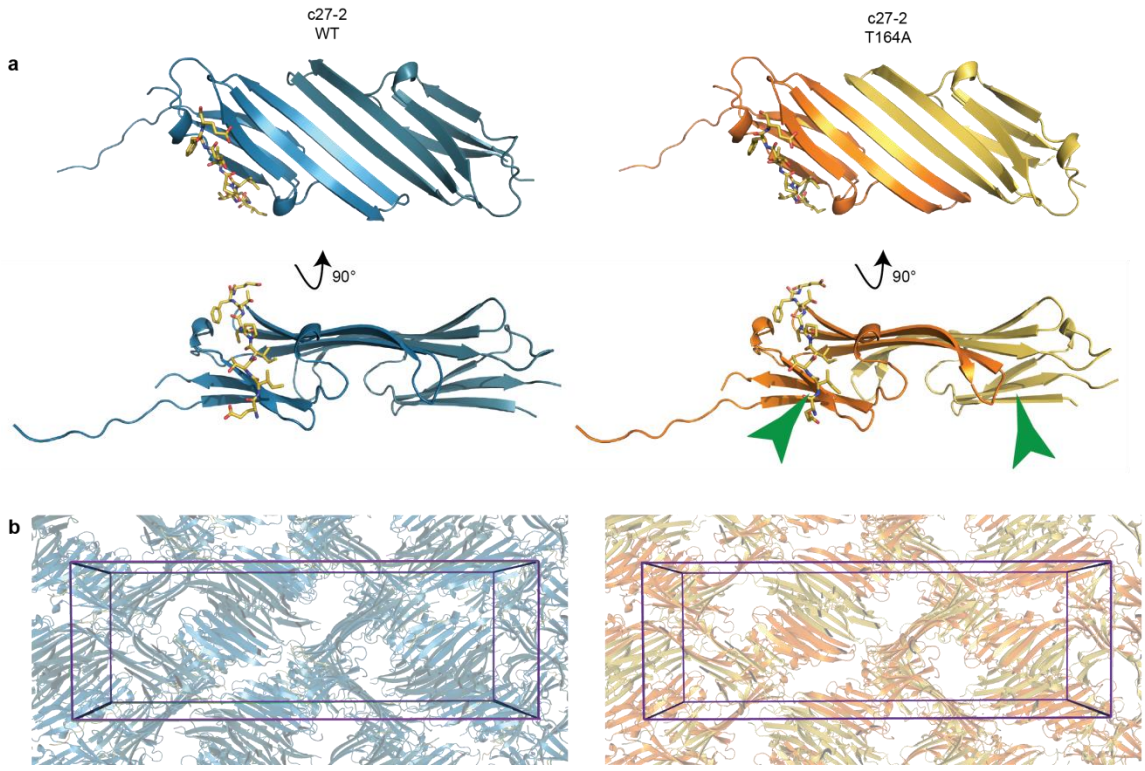


Figure 5 – 4 The structure of T164A c27-2 (orange) in comparison with WT c27-2 (blue). (a) The c27-2(T164A) structure exhibits the unravelled β2-strand which can bind into the binding groove of a neighbouring dimer; formation of the characteristic dimer through an extended β-sheet; and the binding of one CTP (yellow) per dimer, and so is very similar to the WT. Green arrows indicate the T164A mutation sites. (b) The crystal packing of the proteins is also similar, with almost identical unit cell dimensions (Table 5 – 6).

The effect of the T164A mutation is subtle, without an obvious large scale structural effect on the macromolecule. It is most obvious in chain B, where the threonine side chain in the WT is able to hydrogen bond to the S98 side chain and a nearby water molecule (Figure 5 – 5). On mutation to an alanine, this hydrogen bonding is interrupted, which would interrupt a small hydrogen bonding network across the outward face of the β3-β9-β8 sheet. This hypothesis was corroborated by the high resolution structure of R136W c27-2 at 1.45 Å (Section 2.1.3), where it was possible to distinguish two populated conformers of T164. Indeed, the main conformer of T164 (78% occupancy) can hydrogen bond to a neighbouring water molecule and the S98 side chain, for which it is possible to model the rotamer with confidence. The other conformer

(22%) can hydrogen bond S156 and one of two neighbouring waters. This is shown in Figure 5 – 6. In a study of FL HSP27, the T164A mutant exhibited a similar chaperone activity to HSP27(WT) in *in vitro* assays, but was observed to have a poorer thermal stability and a larger range of oligomeric mass[7], so it is possible that by interrupting this network, the mutation has an allosteric effect on the interfaces formed by the monomer.

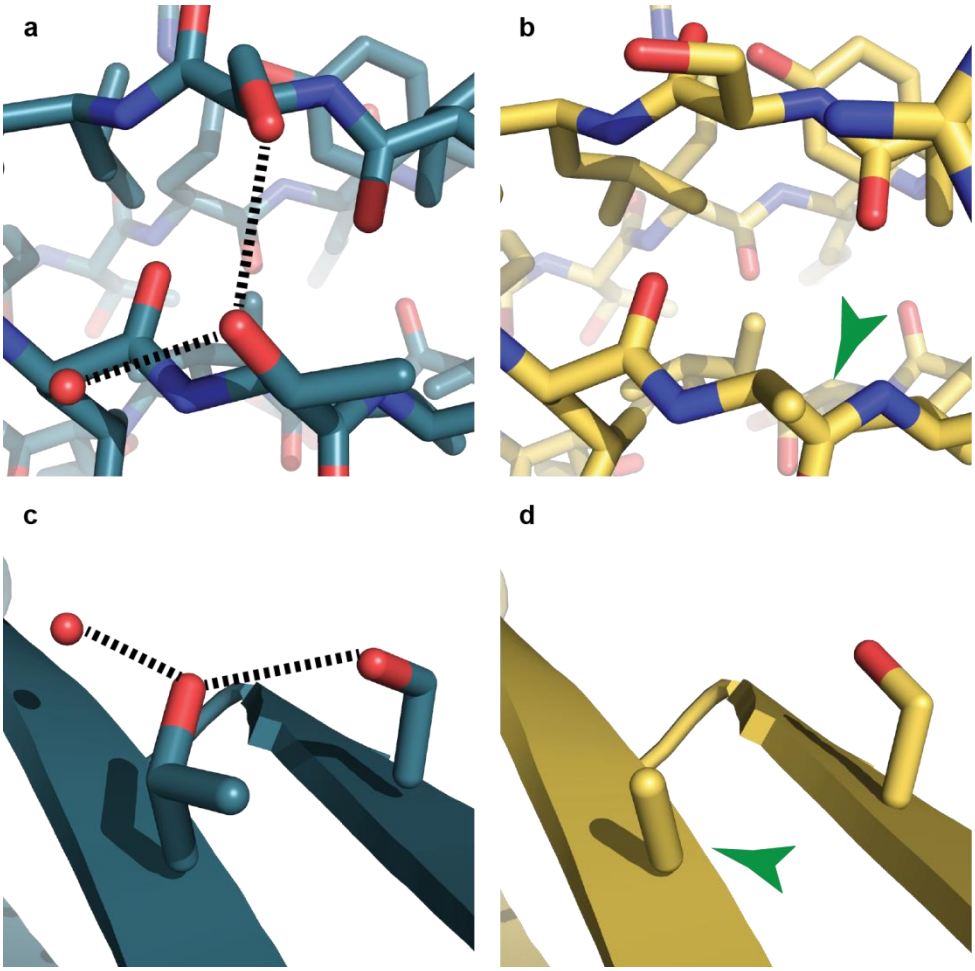


Figure 5 – 5 Comparison of the T164A mutation site in the c27-2(WT) (a and c) and c27-2(T164A) (b and d) structures. The T164A mutation in the β9-strand interrupts the hydrogen bonds formed by T164 to S98 and water (red sphere) as part of a small hydrogen bond network. (c and d) The side chains of the same residues, from the perspective at the C-terminal end of the β9 strand. Dashed lines indicate hydrogen-bonds formed within the WT (blue) that are absent in the T164A mutant (yellow).

The cysteines at the inter-dimer interface are modelled towards one another with a S-S distance slightly longer than in the WT (3.0 Å versus 2.6 Å). The large S-S distance could be in part due to the lower resolution of the structure and higher temperature factors in this region in the T164A model, and so the certainty with which we can model the exact position of the atoms is lower (the temperature factor, B, of the sulphurs are 96.26 and 104.00 in T164 versus 60.37 and 61.74 in the WT). It is likely that disulphides would also be present in these structures as the cysteines are aligned in the same way as the WT, for which the disulphide is demonstrated with an F_o-F_c map in Chapter 4, and the crystals were also formed under oxidising conditions with T164A, which is easier to oxidise than the WT (see Chapter 6).

The DSSP (Dictionary of Protein Secondary Structure)[8,9] algorithm assigns one more residue in both the β5 and β6+7 strands of T164A c27-2 at the β5-6 loop end. There is very little difference between the two models in this region, and so it must lie on a borderline for the DSSP criteria. In order to be consistent amongst the structures presented here, I have used this DSSP assignment for the β5 and β6+7 strands in the figures in this chapter. The use of DSSP was required as the automatic secondary structure assignment in PyMol was not assigning the β-strands of the R136W structure effectively.

The statistical measures of the model are presented in Table 5 – 6 in the Methods section.

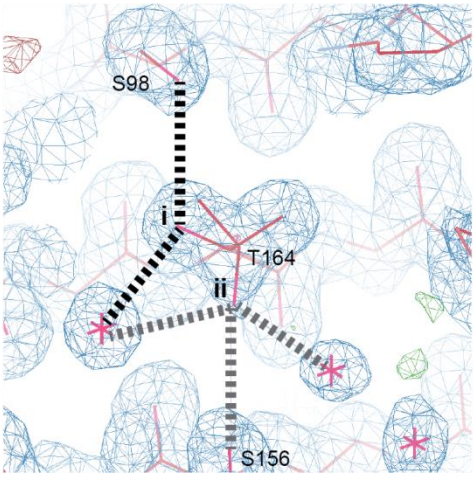


Figure 5 – 6 The T164 residue in the R136W structure. The R136W c27-2 $2F_o$-F_c map (blue mesh, 1σ) clearly indicates the presence of two conformers of T164A ('i' and 'ii'), as present in the molecular model (pink sticks). Grey dashed lines indicate the possible hydrogen bonds of the less populated 'ii' conformation, black dashed lines indicate the hydrogen bonds possible in conformer 'i'.

2.1.2 Mutation of T151 destabilises a 3_{10} helical turn

T151I c27-2 crystallised with tetragonal symmetry (also assigned the P 4_3 2_1 2 space group), but with a slightly elongated unit cell compared to the WT and T164A. Thus its symmetry partners are in slightly different positions than those in the WT crystal. Nonetheless, T151I is also very similar to the WT (Figure 5 – 7). T151I exhibits the unravelled β2 strand on chain A, which binds to the β4-8 groove of a symmetry partner of chain B with the hydrophobic valine modelled pointing into the groove. There is one C-terminal peptide bound in the β4-8 groove of chain A, but none interacting with chain B. The T151I crystals diffracted to 3.1 Å, so we have not modelled waters in the structure, however the resolution is certainly sufficient to conclude that the T151I mutation only leads to a subtle change in the structure of the core domain, as captured in these crystals.

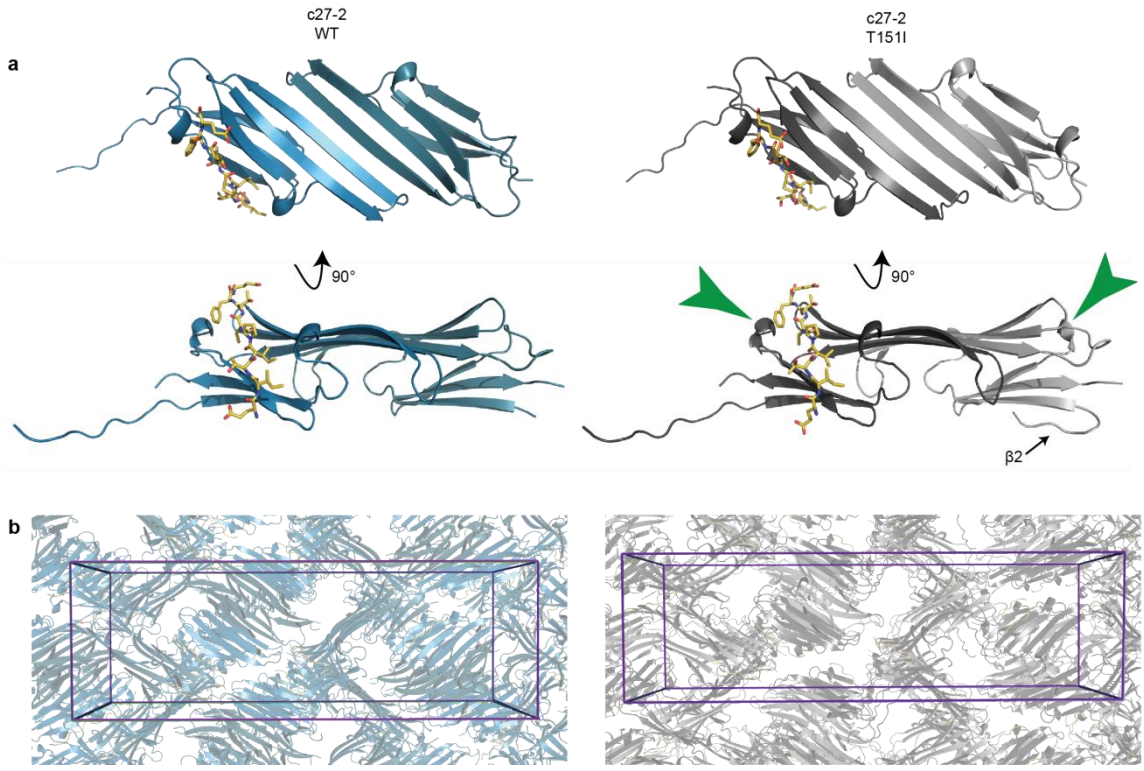


Figure 5 – 7 The structure of T151I c27-2 (grey) in comparison with WT c27-2 (blue). (a) The c27-2(T151I) structure also captures the dimeric biological unit formed through the interface between two β6+7 strands and exhibits the binding of one CTP (yellow) per dimer. The β2-strand is unravelled on chain A, which goes onto form crystal contacts with two further dimers within the crystal lattice. Notably the β2 region of chain B is more resolved than in the WT or T164A structures, and is in a conformation able to hydrogen bond to the β3. Green arrows indicate the T151I mutation sites within the α-helical turn in the β7-β8 loop. (b) The crystal packing of c27-2(T151I) is similar to the c27-2(WT) but with a slightly elongated unit cell (Table 5 – 6).

The most notable difference is that more of the β2 of chain B can be resolved in T151I c27-2, and it is seen to be folded back along the side of the β3 strand (Figure 5 – 7a). This may in part be due to a stabilising interaction with a symmetry mate in the lattice, which encourages the β2-strand to adopt the ordered conformation. It may also be due to an interaction with H131 in the β5-6 loop that was observed in c27-1 and may confer stability to the β2-strand and be a possible node for allosteric communication between the dimeric interface and β2.

The main chain is resolved from I88 onwards, as are the side chains of H90 – R94, neither of which could be modelled in the WT chain B. While DSSP does not assign this is as a β-

strand, some of the residues are able to form hydrogen bonds with the β3. Thus, this region is more ordered than in chain B of WT or T164A c27-2, and captured in a similar mode to cABC[10,11] or the c27-1 core domain[12].

The T151 residue of the WT resides within a short five-membered 3_{10} helix which lies on the loop between the β7 and β8 strands (Figure 5 – 8a). This is formed through a hydrogen bond between the carbonyl oxygen of D149 to the amide nitrogen of Q152 (*i.e.* the donor), and a hydrogen bond between the P150 CO and V153 NH (inset, Figure 5 – 8a), thus defining the 3_{10} helix (hydrogen bonding between the CO of residue *i* with the NH of residue *i+3*). Chain B of c27-2(WT) contains a similar helix to chain A (as assigned by the DSSP algorithm[8,9]), albeit looser. There is also a short 3_{10} helix formed by residues $_{106}$PDE$_{108}$ in chain A and chain B.

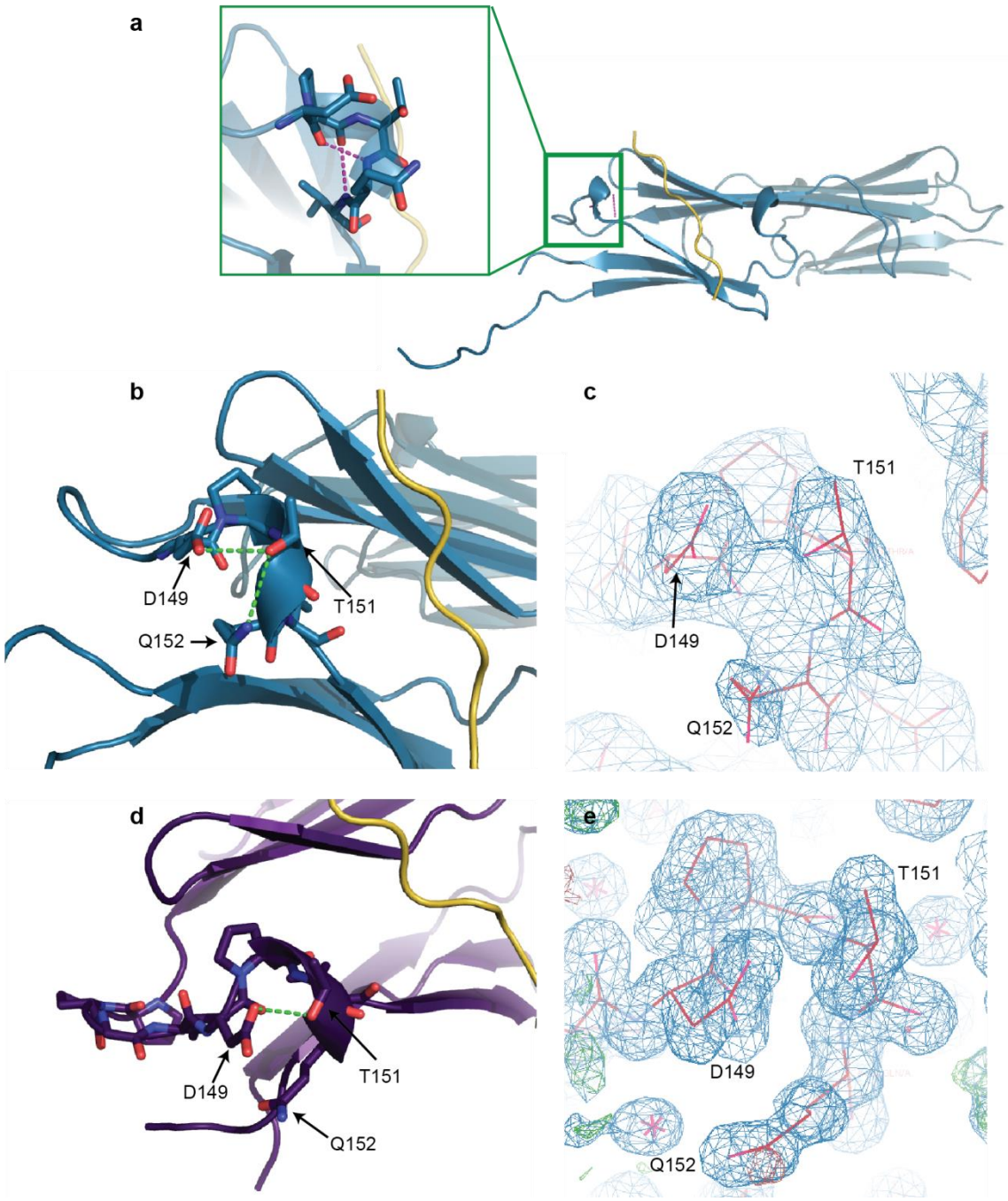


Figure 5 – 8 The T151 residue is positioned in the middle of a short 3_{10} helix in the WT and R136W structures. (a) Position of the helix on the loop linking the major and minor β-sheets between β7 and β8. The helix is defined by two main chain hydrogen bonds between D149 and Q152 and P150 with V153 (inset, hydrogen bonds in magenta). (b and c) T151 can form two hydrogen bonds (green) to D149 and Q152 sidechains in c27-2(WT). (d and e) Only the D149 and T151 are close enough to form a hydrogen bond (green) in the c27-2(R136W) model, the data are consistent with Q152 being further away (e). c27-2(WT) is displayed in blue, c27-2(R136W) in purple, CTP in yellow, $2F_o$-F_c map at 1 σ in blue mesh, F_o-F_c map at 3 σ in red and green mesh.

T151 in the WT offers stability to this helix through sidechain hydrogen bonding to D149. The WT model also suggests that the T151 OH group can accept a hydrogen bond from the Q152 sidechain (Figure 5 – 8b). However, formation of this second hydrogen bond between T151 and Q152 is not observed in the R136W structural model (Figure 5 – 8d), where there is a 5.1 Å distance between the oxygen acceptor and the nitrogen donor on Q152 due to the extension of the Q152 side chain. Both of these conformers are supported by the electron density in the $2F_o$-F_c Fourier synthesis on chain A (blue mesh, Figure 5 – 8, panels c and e), with a small negative region in the F_o-F_c map near to the Q152 nitrogen in the c27-2(R136W) structure (red mesh. Figure 5 – 8e) but this region was still best modelled by the conformer shown. As the T151 – Q152 sidechain hydrogen bond is possible, as demonstrated in the c27-2(WT) structure, it is likely that it is formed at least some of the time in solution or that the three sidechains are involved in a hydrogen-bond network with water in this solvent-exposed region of the dimer. The T151I mutation will interrupt this hydrogen bond network as the aliphatic isoleucine cannot form hydrogen bonds (Figure 5 – 9).

The T151I mutation will therefore introduce greater conformational flexibility to the loop region, perhaps interrupting a substrate binding site, or producing an allosteric effect within the β6+7 strand or the β8 strand, both of which are important for oligomerisation. FL HSP27 carrying the T151I mutation was observed to have a similar efficacy as the WT in cell survival and *in vivo* chaperone assays[13] under heat shock, suggesting that the deleterious effect of the mutation may present itself in other ways. It is also noteworthy that the equivalent position to T151 is occupied by a leucine in cABC, which was more active than c27-1 in *in vitro* chaperone assays[12].

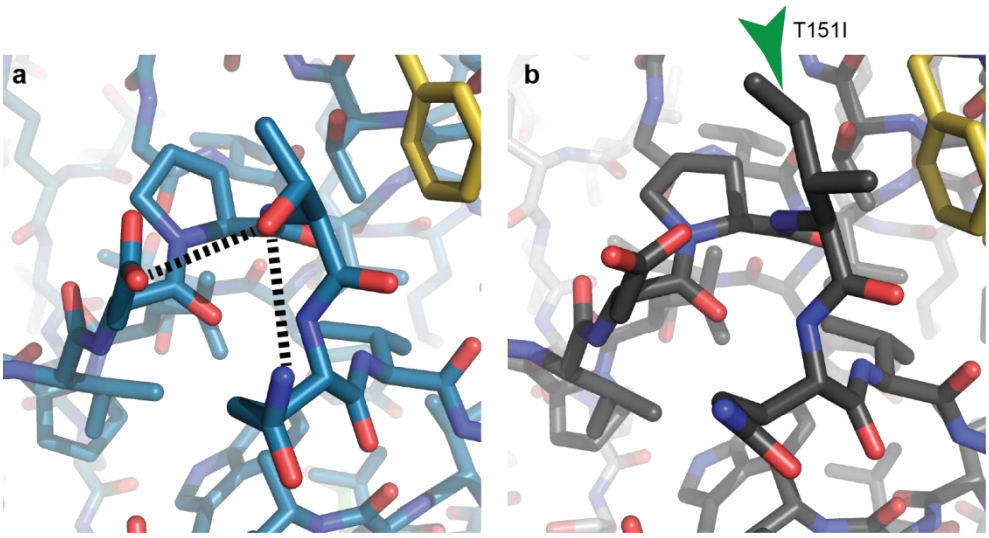


Figure 5 – 9 Comparison of the T151I mutation site in the c27-2(WT) and c27-2(T151I) structures. Dashed lines indicate hydrogen bonds formed within the WT (blue) that are absent in the T151I mutant (grey) when the threonine is mutated to isoleucine.

As is the case in the WT, the loop between β5 and β6+7 curves down (Figure 5 – 7a). This loop could be stabilised in chain B by a hydrogen bond with a symmetry mate in the crystal lattice. However, chain A is not in the proximity of symmetry mates at this loop, so the various observed conformers of this loop are not solely artefacts of crystallisation. cABC also exhibits conformational flexibility in this region[14].

The cysteines are modelled towards each other with a 3.4 Å S-S distance. Like c27-2(T164A), c27-2(T151I) was also easier to oxidise with oxidised glutathione (see Chapter 6), so again it is likely that the larger S-S distance is a result of the poorer resolution of the structure than the WT, so a greater uncertainty in the modelling of the atomic coordinates.

2.1.3 c27-2(R136W) forms a register-shifted dimer

The structures of both T151I and T164A showed only subtle differences to the structure of the WT c27-2. Surprisingly, the crystals of c27-2 R136W revealed a noticeably different arrangement, with a difference in unit cell; peptide stoichiometry; disorder of the β2 strand; and most notably in the registration of the two monomers relative to one another (discussed in Section 0).

The unit cell of the R136W c27-2 crystal is 35% the size of the WT unit cell, and is consistent with an orthorhombic C 2 2 2_1 symmetry (Figure 5 – 10). This reflects the smaller asymmetric unit which comprises only one c27-2 chain, as well as one C-terminal peptide chain, 37 waters and one adducted β-mercaptoethanol (discussed in Section 2.2.1). The highest resolution shell is 1.45 Å, the highest of all the structures here, which allows the confident designation of side chain orientation within the model (Figure 5 – 10b).

On applying the symmetry operators of the C 2 2 2_1 space group, the canonical dimer is formed through β-sheet hydrogen-bonding between the β6+7 strands. However, as can be observed in Figure 5 – 10c, the angle between these two strands, θ, is smaller in R136W than it was in the WT. This arises because the two monomers in the R136W structure are 'shifted' by two residues at the dimer interface, such that each cysteine sits opposite the S135 of its partner. This arrangement is described as an 'antiparallel III' registration[11] whereas the WT, T151I and T164A structures reported above are all in the antiparallel II registration, where the cysteine thiol side chains sit opposite one another. The three registrations of the inter-dimer interface have been well reported for the core domain of αB-crystallin[10,11,12], but this is

the first time an alternate registration has been observed for an HSP27 variant as the published crystallographic[12] and NMR[15] structures in which this inter-dimeric interface is maintained are in the APII state. Notably, the structure of R136W c27-2 has a smaller θ than the APIII structure of cABC[12].

In this model two peptides are bound per dimer, which agrees with previous observations in c27-1 and the core domain of αB-crystallin[12]. The β2 strand is not resolved in this structure which is due to added disorder in this region of the polypeptide chain. The unravelled β2 of the WT, T151I and T164A gains order due to its binding to the β4-β8 groove of a symmetry mate in the crystals. This implies there is either competition between the C-terminal peptide and the β2 to bind to the β4-β8 binding, or the absence of a second bound C-terminal peptide in the WT, T151I and T164A allowed the anchoring of the unravelled β2 strand in a non-specific interaction. The density of the c27-2(R136W) structure is higher than the other structures (lower V_M and higher solvent content, Table 5 – 6), despite the disordered β2 and larger volume of the dimer. Thus the higher density may simply reflect the particular lattice formed rather than any specific variation in structure, and is comparable to the density of c27-1(4MJH) (Chapter 3).

In the WT, R136 coordinates water which in turn coordinates to D129 in an intra-chain network. Mutation may interrupt this single interaction and so make the APIII state more favourable (Figure 5 – 10). It is possible that the main chain of H103 interacts with the aromatic ring of the tryptophan of R136W in a T-form pi-interaction in its place. However, a R136L mutation has been identified in both distal HMN and CMT2 patients[16], suggesting

that in this R136W mutant it is removal of the arginine side chain rather than the specific identity of the mutant that is the crucial factor.

Due to the different antiparallel registration of this protein, the identity of the partners in hydrogen-bonding pairs across the interface has also changed (Section 0). One particular noteworthy change is in the side chain of E126. This faces upwards out of the extended β-sheet in the WT and forms a salt bridge across the interface with K141 on the other chain. In c27-2(R136W), it faces downwards and hydrogen-bonds with the intra-chain H124 sidechain. Histidine 124 is homologous to H104 in αB-crystallin, which when mutated makes a more active chaperone[17].

The original PyMol representation of R136W had shorter β-strands at several positions. In the WT and R136W models, there was not an obvious difference in the distances and angles between putative hydrogen bonding partners on the β6+7 strands. In order to produce a fairer comparison of the R136W structure with the WT structure, we analysed both with DSSP[9,8] and subsequently modelled the β6+7 strand as a continuous strand in R136W. DSSP also recognised the β4, β5 and β9 as being longer β-strands than PyMol assigned, so they are also presented according to the DSSP assignment in the figures herein. (Both DSSP and PyMol are assigning hydrogen bonds and secondary structure on the coordinates for each atom provided in the PDB file. Both of the software packages use similar criteria for the detection of hydrogen bonds. However, PyMol's secondary structure assignment is optimised for aesthetics resulting in slightly different results from DSSP for those regions where the secondary structure is less well-defined.)

The β5 strand of R136W is longer than that of the WT (Figure 5 – 10c), such that the β5-6 loop is much shorter and does not curve downwards into the shared groove as observed for the WT. This is mediated in part by the interface H-bonding of earlier residues in R136W (H131 hydrogen bonds to K141 in R136W rather than Y133 hydrogen bonding K141 in the WT, which are both main chain interactions). It is also associated with the interruption of the D129-H_2O-R136 network discussed above (Figure 5 – 10).

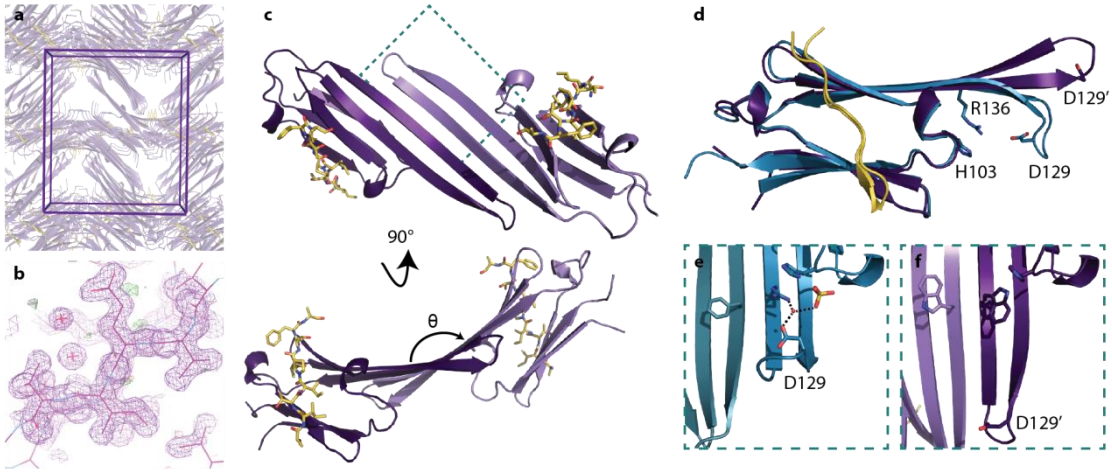


Figure 5 – 10 The crystal structure of R136W c27-2. (a) The unit cell is much smaller than for the WT, because there is only one chain and peptide in the asymmetric unit. (b) The 1.45 Å resolution allows confident modelling of side chain orientations. The $2F_o$-F_c map is shown at 1 σ (0.3590 eÅ^{-3}) in purple. (c) The R136W c27-2 does not have the β2-region resolved, and there are two peptides bound per dimer. Notably, main chain hydrogen bonding across the intra-dimer interface now occurs between different partners, to place the dimer in the antiparallel III registration. This moderated interface results in a large curve in the extended β-sheet, denoted by θ. (d) The overlaid monomers of WT (blue) and R136W (purple) c27-2 demonstrate a significant variability in the orientation of the β5-6 loop containing D129. In the WT APII state, the loop curves in so that D129 can coordinate water (e). This same water is coordinated by a sulphate and R136, and this network is interrupted in the presence of the R136W mutation and the APIII state (f).

Both T164 and T110 occupy multiple conformations. There were three possible conformations that could fit the electron density (see Figure 5 – 6 for a snapshot of the T164 residue on the R136W model and its $2F_o$-F_c map in blue). We thus chose the two most likely rotamers based

on their prevalence in the PDB, using the Rotamer Analysis tool in Coot, and modelled them with 78% and 22% occupancy after refinement (or 56% and 44% for T110) to account for the data.

2.2 Additional electron density in the R136W structure

2.2.1 A population of β-mercaptoethanol cysteines at the dimeric interface

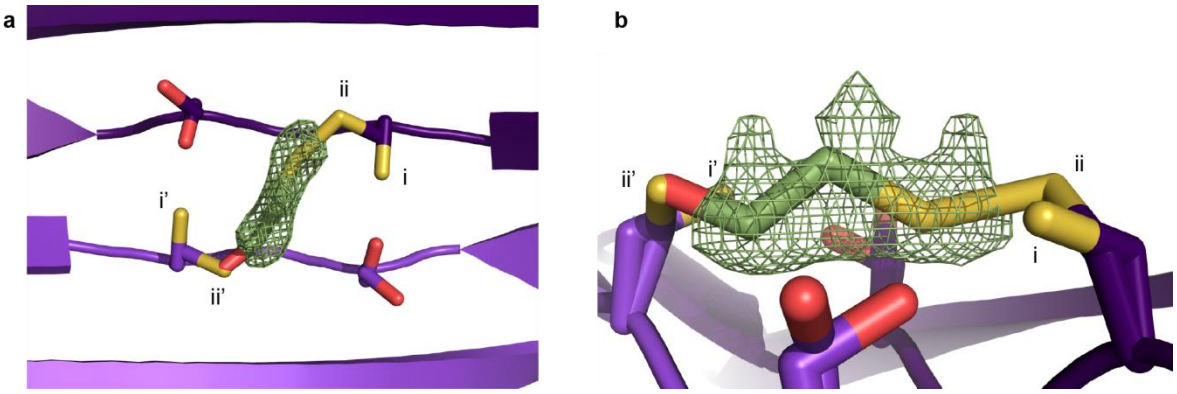


Figure 5 – 11 Alternate conformations of C137 and S135 in the c27-2 R136W structure, as viewed from above the extended β-sheet (a) or along the β6+7 strand (b). The F_o-F_c omit map (green) was calculated by removing the BME from the model, followed by simulated annealing and a final round of refinement. It is shown here within 4 Å of the BME molecule on the final model at a level of 0.3769 e.Å-3 which corresponds to a cut-off of RMSD (σ) = 3 in Coot.

One of the most notable differences of R136W c27-2 to the WT and other mutants is its shift into a different registration state at the dimer interface. This is accompanied by the residence of the cysteine C137 side chain in two occupancies: across the dimer interface where it is able to hydrogen bond to S135 (conformer (i) in Figure 5 – 11) or with the sulphur atom above the main chain (conformer (ii) in Figure 5 – 11). In the WT, S135 forms two main chain hydrogen bonds with T139, and its side chain forms a hydrogen bond with water.

Conformer (ii) rationalises electron density found between the two β-strands, which we have modelled as BME bound to the cysteine side chain in conformer (ii). BME was used in the

early steps of the purification of all of the core domains to help with cell lysis and during cleavage of the histidine affinity tag where it was used for the required reduction of TEV protease. Despite extensive buffer exchange into buffers without reducing agent, a small population of monomers with an adduct of *ca.* 76 Da was often observed in the mass spectra of the core domains, which we hypothesised was BME (Chapter 5 Section 2.3). Figure 5 – 11 shows two views of an F_o-F_c map for BME in the structure of R136W (observed structure factors – calculated structure factors in the absence of BME), shown at 0.3769 eÅ^{-3} (which is equal to a contour level of 3σ). This demonstrates that there is a substantial amount of electron density at this location, which is in keeping with the size and conformation of a BME molecule. Binding of BME at this site would result in C137 on one subunit existing as conformer (ii), and simultaneously C137' on its partner subunit must be in conformer (i') to avoid steric clash with the BME molecule. This would then make it available for hydrogen bonding to S135, which itself populates two occupancies that is presumably dependent on the availability of C137. As the asymmetric unit consists of only one monomer and peptide, we modelled the BME and each conformer of S135 and C137 with 50% occupancy.

2.2.2 Unmodelled electron density near to the β3 strand

The R136W model also shows additional extraneous electron density around the β3 strand. This is depicted in Figure 5 – 12a which shows the F_o-F_c map in green at 11 Å around the $RSVL_{96\text{-}99}$ sequence in the β3 strand, again at a cutoff of 3σ (0.3900 e.Å^{-3}). The black oval represents a two-fold rotational axis (with symmetry operator $(x,-y,-z)$).

During the model refinement, we left the extended electron density unmodelled after considering several options. Firstly, the density resembles a peptide structure and so we modelled some of the unresolved residues from the β2 region, with the rest of the strand being disordered. The density lies next to the β3 and so would be consistent with the position of the β2 in previous models. In particular, in one of the molecular dynamics simulations we had observed the three terminal residues forming a small β-strand in the β2 position. A short β-strand would be consistent with the density as it lies on an axis of symmetry so only half of it can be accounted in each unit cell. However, it was difficult to satisfy the geometrical restraints and also to justify why the interceding residues would lie in regions of very low electron density despite restricted flexibility. We also tried a model with succinic acid and glycine molecules in this region (which were present in the mother liquor during crystallisation), but we did not feel that they explained the density well. Finally, we considered whether the N-terminus of another chain was binding in this pocket, akin to the cross-dimer interactions observed in the WT c27-2 crystal lattice. While this is conceivable, we decided that there was not evidence in the data to support this in our model, particularly regarding the interceding residues.

This final consideration of cross-dimer binding was of particular interest as binding interactions have been observed in this position before. In their 2011 paper[14], Clark *et al.* discussed three observations of binding at this position between the smaller β-sheets of each monomer, which they term the 'shared groove'. Two of these occurrences involved small molecules from the crystallisation conditions of human αB-crystallin (PDB 3L1G) and bovine αA-crystallin (PDB 3L1F)[11], and in the c27-2 WT structure presented in Chapter 4, we also have modelled four glycerol molecules within this groove. The third related to the authors' earlier structure of rat HSP20[10], discussed in Chapter 4 here for the binding of its β2-strand to the β4-β8 groove of a neighbouring dimer. The C-terminus of a neighbouring dimer was also captured in this shared groove in rat HSP20. This is recreated in Figure 5 – 12b and c, alongside the R136W model presented here. Like R136W c27-2, the rat HSP20 structure was also determined to a high resolution (1.3 Å) and in the APIII state.

In their later paper[14], the authors postulated that the binding of some of the C-terminus to the rat HSP20 shared groove was mimicking the behaviour of the N-terminal region in the full-length protein. This may in turn modulate the assembly dynamics of the protein by favouring the dissociation of the dimer on binding of the N-terminus. This was motivated by implications of the N-terminal region in subunit exchange and oligomerisation of αA-crystallin, HSP27[18] and αB-crystallin[19], and they used it to rationalise the dissociation of HSP27 on phosphorylation due to tighter N-terminal binding. This suggestion does not quite account for the cross-dimer donation of the N-termini[19] (notably the three residues around the final phosphorylated serine on the N-terminus of αB-crystallin, p59S). However, in light of our observations of the donation of the N-terminus in the WT, T151I and T164A, and the same

behaviour of rat HSP20, the suggestion of N-termini binding to neighbouring dimers and the regulation of this by phosphorylation, is gaining speed.

The unmodelled density sits next to the L99 residue in the β3 strand. L99 forms part of the hydrophobic core of the monomer, sitting behind the sidechains of the β6+7 strand above (for example it is directly behind R140G). An L99M mutation was reported in a Pakistani family with distal HMN/CMT2 disease[20]. Only the homozygous son was affected, while the heterozygous father and sister were asymptomatic. This was the first report of an autosomal recessive mutation in HSP27 linked with distal HMN. It is possible that L99M is interrupting either the dimeric interface or the hydrophobic core (which extends from the shared groove through to the β4-β8 groove, where binding of the C-terminus occurs in oligomerisation of the FL). However, it should also be considered whether mutation is affecting binding in this shared groove as postulated here. It has been suggested that the variable age of onset of distal HMN/CMT2 may reflect the different influence of the associated mutations on the structural integrity of HSP27[21]. Indeed, it may be possible that L99M leads to an autosomal recessive trait (rather than the autosomal dominant mutations listed in Table 5 – 1) because of its more subtle effect on the HSP27 complex.

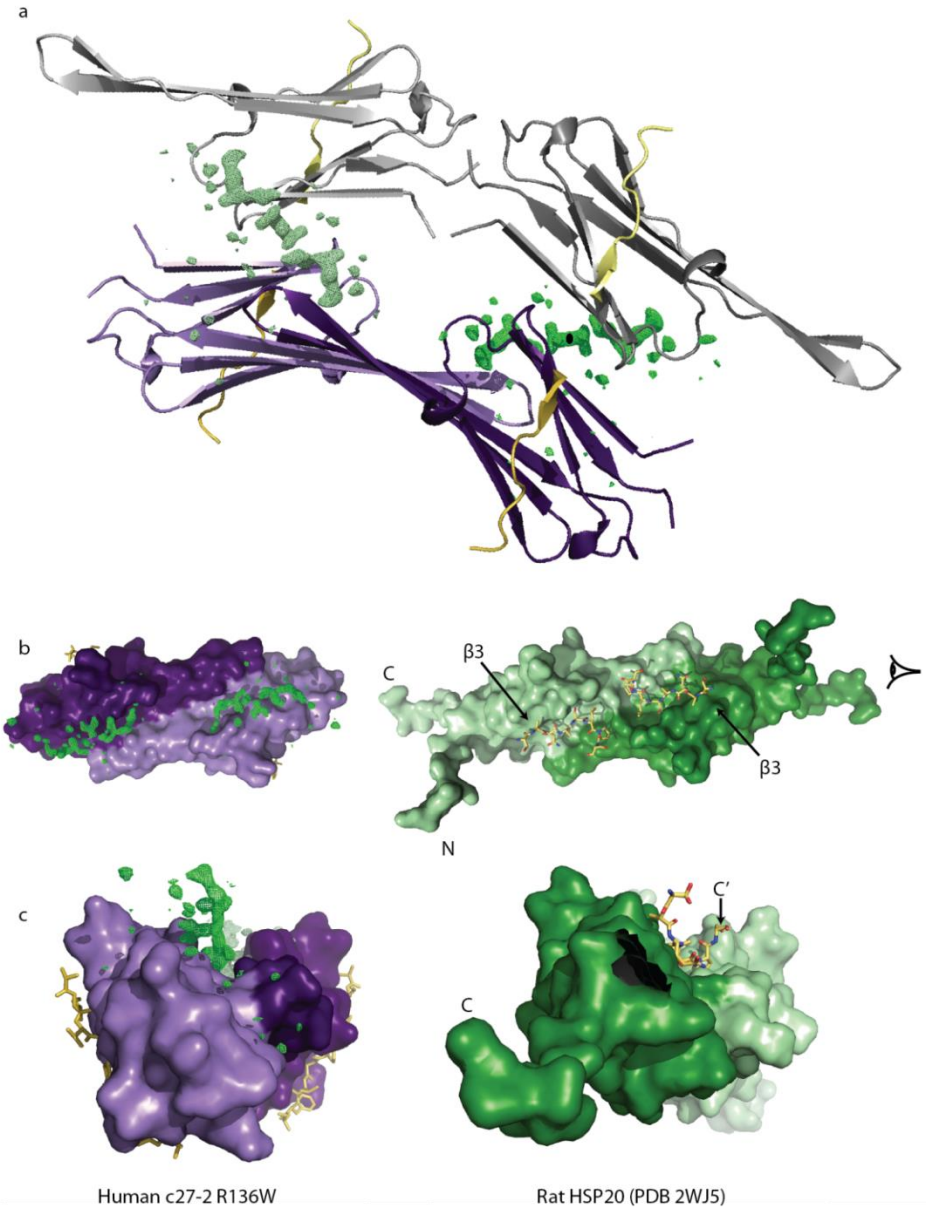


Figure 5 – 12 Unmodelled density in the R136W c27-2 data may be significant. (a) An F_o-F_c map of the final electron density and model (3 σ, bright green shows density that is not assigned). The density is in the proximity of the β3 strand of the R136W monomer (purple) and that of one of its symmetry partners (grey), and lies on a C_2 rotational symmetry axis perpendicular to the plane of the paper (black oval). The same configuration applies to all chains in the crystal lattice, as demonstrated here for the partner R136W monomer (light purple) and its C_2 related symmetry mate. (b) Surface representation of the R136W dimer in (a) with the shared groove facing upwards. The unmodelled density lies in a similar position to the C-termini within the shared groove of the rat HSP20 ACD (PDB 2WJ5, yellow stick representation shows the final eight resolved residues of two symmetry-related chains). When these dimers are viewed along the line of sight depicted, the two ACD chains are seen to form a groove in which the density or C-termini sit (c). In (c) the first ten residues from the N-terminal end of rat HSP20 have been removed for clarity.

2.3 Registration shifts of the intra-dimer interface

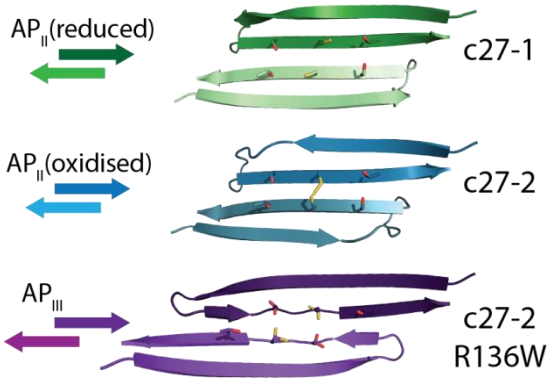


Figure 5 – 13 Three possible modes of interaction at the intra-dimer interface. In the APIII registration, the two monomers are shifted bilaterally to form a more extended β-strand interaction across the intra-dimer interface. The disulphide is thus unable to form.

Between the structures reported here and the previous structure of the core domain of HSP27 in the canonical dimer[12], there are three different observed modes of the intra-dimer interface (Figure 5 – 13). They comprise the reduced APII state of c27-1 (PDB 4MJH) where the cysteines are opposite one another but do not form the disulphide; the oxidised APII where the disulphide does form (Chapter 3); and the APIII state where the cysteines are not opposite one another in the interface and so do not form the disulphide despite being in oxidised conditions. This behaviour of the antiparallel registration states is well-documented for the core domain of αB-crystallin, for which crystal structures in all three registration states of the inter-dimer interface have been reported (Chapter 3 Introduction). However, this is the first time that the HSP27 intra-dimer interface has been observed in the APIII state and demonstrates an additional interaction site with variable behaviour, which can contribute to the plasticity of the HSP27 FL ensemble.

Chain A	Chain B	
	APII (WT)	**APIII (R136W)**
Main Chain – Main Chain		
H131		K141 – 2.9
Y133	K141 – 2.9, 3.3	T139 – 2.9, 3.1
S135	T139 – 3.0, 3.1	C137 – 2.9, 2.9
C137	C137 – 2.8, 2.9	S135 – 2.9, 2,9
T139	S135 – 2.9, 3.1	Y133 – 2.9, 3.1
K141	Y133 – 3.0	H131 – 2.9
Total no. of interactions	**9**	**10**
Side-chain - Side Chain		
E126	K141† – 2.9	
Q128		R140 – 3.2
D129	R140† – 2.9, 3.2	R140† – 3.0, 3.0
S135		C137 – 2.6*
C137	C137‡ – 2.6	S135 – 2.6*
R140	D129† – 2.7, 3.0	D129† – 3.0, 3.0
R140		Q128 – 3.2
K141	E126† – 3.1	(H_2O)
Total no. of interactions	**6**	**7**

Table 5 – 2 Inter-subunit interactions across the antiparallel interface. Interactions are listed between chain A and its interaction partner (chain B in the WT structure, or a symmetry mate for the R136W structure because this only has the monomer in the asymmetric unit). The length of each bond is indicated in Å after the identity of the interacting partner. Most of the interactions are hydrogen bonds, unless indicated otherwise. † - these interactions are at least partly electrostatic, ‡ - covalent bond, * - as discussed in Section 2.2.1, it is likely only one of these hydrogen bonds is present at one time.

As expected, the identity of the interactions across the interface varies depending on the registration states. These are detailed in Table 5 – 2 and Table 5 – 3 for the oxidised APII and APIII structures, and include hydrogen bonds, the covalent disulphide and salt bridges (Figure 5 – 14). Most notably, although the APIII registration cannot form the disulphide, it has slightly more extensive hydrogen-bonding between the main chains and side chains of the interface β-strands (Table 5 – 2). There is some rearrangement of the intra-chain interactions within the chains too (Table 5 – 3) which may facilitate the different interface binding mode, and possibly provide some entropy compensation.

APII (WT) Chain A	APII (WT) Chain B	APIII (R136W) Chain A*
Intra-chain and sidechain		
D129 – H_2O – R136		
(D129 – R136 is 4.4 Å)		
		R140 - Y142 – 2.8, 2.9
		R127 – Y133 - 2.8, 2.8
H103 – SO_4 – R127	H103 – SO_4 – R127	
(H103 – R127 is 4.4 Å)	(H103 – R127 is 4.6 Å)	
		Y142 – R140 - 2.8, 2.9
		Y133 - R127 –2.8, 2.8
Intra-chain		
D129 SC	H131 MC – 3.1	H131 MC – 2.9
	G132 MC – 3.3	
R127 SC	H103 MC – 3.1	E130 MC – 3.2†
R127 SC		D129 MC – 3.5†
H103 MC		W136 – 3.4†

Table 5 – 3 Intra-chain interactions that differ in the APII and APIII states observed. Again, distances are in Å. * - As there is only one monomer within the asymmetric unit for this model, the same intra-chain interactions are observed in both monomers of the dimer. SC – indicates from which residue the side chain is interacting with the main chain (MC) of its partner residue. † - indicates a possible interaction but one for which the geometry captured in the crystal structure does not suggest a strong interaction.

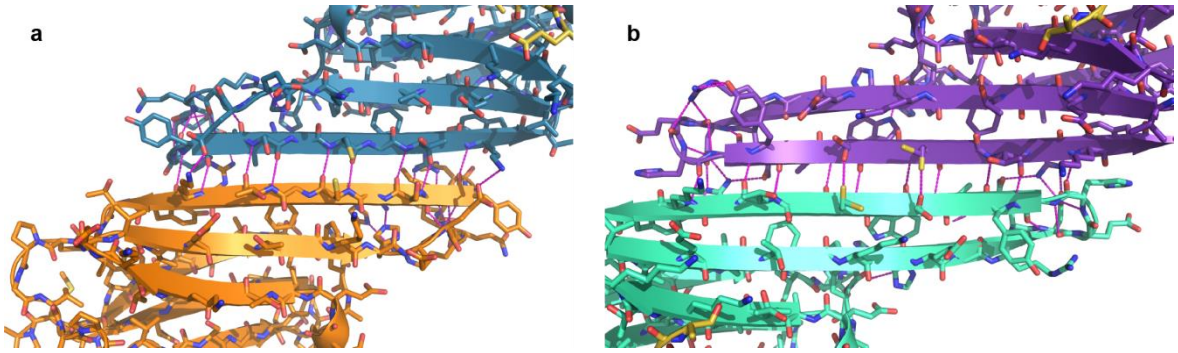
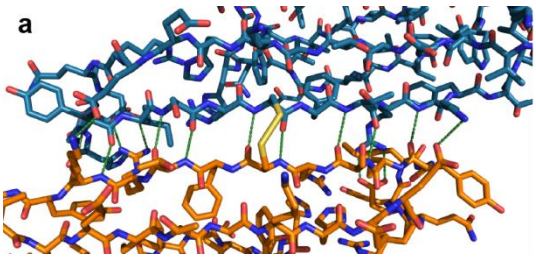


Figure 5 – 14 Interactions on residues near the dimer interface that vary with registration. (a) WT c27-2 chain A (blue) and chain B (orange). (b) c27-2(R136W) chain A (purple) and chain A symmetry mate (turquoise). The interactions were found with the PyMol 'find_polar_contacts' command with default parameters, and were manually curated for Table 5 – 2 and Table 5 – 3.

2.4 The intersubunit disulphide is rare and sterically strained

The different modes of the intra-dimer interface led us to question whether the disulphide acts as a 'redox switch' that on reduction allows the HSP27 to populate alternative states of the oligomer due to this interface plasticity. The sHSPs are highly dynamic proteins and the polydisperse nature of mammalian sHSPs is underpinned by their variable interaction at the inter-dimer and intra-dimer interfaces[22]. HSP27 is redox-sensitive[23], interacts with thiol-reductases[24,25] and its chaperone activity in response to oxidative stress is well-documented[26]. With its cross-subunit disulphide, perhaps HSP27 is able to regulate the behaviour of one of its interfaces in response to reducing or oxidising environments.

A redox sensor has been described in another molecular chaperone - HSP33 from *Bacillus substilis*[27] (this protein is not classified as part of the sHSP family[28]). This protein possesses two sets of neighbouring cysteines in its C-terminal region which together coordinate a zinc ion. It is only on oxidation of the cysteines to form disulphide bonds, and the concomitant release of the zinc ion and conformational change in the C-terminus, that the protein is able to access its active dimeric state[29]. So some molecular chaperones have evolved with redox-sensitive disulphide bonds to mediate their response to oxidative stress.



Please consult the hard-bound copy of this thesis deposited with the Bodleian Library, or the original figure(s) in the publication(s) referenced in the figure caption below.

Figure 5 – 15 a) The disulphide (yellow) in c27-2 WT bridges two registered cysteines that are main chain hydrogen bonded within anti-parallel β-strands. These β-strands lie on two different subunits (blue and orange). Hydrogen bonds are shown as green dashed lines. b) The nomenclature for the dihedral angles of the disulphide (Table 5 – 4), figure from Indu *et al.*[4].

This hypothesis gains further support in the work of Indu *et al.*, who characterised cross-strand disulphides within the PDB[4]. There they defined a cross-strand disulphide as one formed by two cysteines in a registered pair of antiparallel β-strands. The c27-2 disulphide meets these criteria (Figure 5 – 15), forming between two cysteines that are opposite one another in the β-strands and are main chain hydrogen-bonded. Indu and colleagues highlighted that cross-strand disulphides are very rare between hydrogen-bonded registered pairs; indeed they found only six occurrences in the entire PDB of which five proteins had high sequence identity (>90%). Moreover, the cross-strand disulphide observed in HSP27 not only bridges two β-strands, but two separate subunits. It also occupies unusual dihedral angles, especially for χ_1 (Table 5 – 4). Strikingly, it does not occupy the same area of conformational space observed for the other hydrogen-bonded cross-strand disulphides[4], and even somewhat deviates from the values observed in a structure of cABC engineered to mimic this HSP27 disulphide (Table 5 – 4). This suggests a distinct conformational character to those previously identified. Indu *et al.* have shown that in contrast to non-hydrogen bonding cross-strand disulphides, disulphides introduced to hydrogen-bonded pairs in two proteins are

destabilising[30,4], and can confer redox activity[4], which supports suggestions of cross-strand disulphides as redox switches[4].

Angle	c27-2	cABC(E117C)	Occurrence
χ_1	+79.4°	+65.4°	1.9 %
χ_2	-73.8°	-33.2°	18.6 %
χ_{ss}	+119.0°	+70.76°	12.3 %
χ_1'	+76.4°	+58.5°	1.9 %
χ_2'	-63.5°	-44.4°	21.3 %

Table 5 – 4 The dihedral angles of the cross-strand intra-subunit disulphide in c27-2(WT) and a disulphide-locked cABC(E117C) dimer[12], along with the occurrence of these angles in all naturally-occurring disulphides[4]. Dihedral angles defined in Figure 5 – 15.

3 Conclusion

The identification of numerous independent HSP27 mutations that lead to phenotypes characterised by muscle weakness and atrophy indicate the crucial role of HSP27 in healthy motor neurons. In this chapter, we have successfully introduced three of these mutations to the c27-2 co-crystallisation system to provide a structural rationale for their effect.

It has previously been suggested that the pathogenic effects of neuropathy-related mutations in HSP27 may be produced by different molecular mechanisms[13]. While the monomeric structures of each mutant c27-2 are very similar to those of the WT, the crystal structures shed light on the influence of these mutations, which may have varied influence on the behaviour of the protein.

We observed that the T164A mutation interrupts a hydrogen bonding network across the outward-facing surface of the minor β-sheet. As T164A mutation was found to reduce the

thermal stability of FL HSP27 and produce atypical oligomerisation with an increase in concentration[7], we suggested that this hydrogen bonding network may have an allosteric effect on the inter-subunit interfaces within the oligomer. The crystal structure shows very similar interface characteristics as the WT, at the intra-dimer (β6+7) and pseudo inter-dimer (CTP:β4-8 groove) interfaces as well as in the unravelling of the β2 and its association with neighbouring dimers in the crystal lattice, so the manner in which this mutation produces this allosteric effect remains enigmatic. However, the highly dynamic nature of the sHSPs, especially in the association and disassociation of the oligomers, means that the interfaces populated in solution may be more variable than those captured within the crystal lattice.

The T151I mutation will remove the additional stability afforded by sidechain hydrogen bonding to the 3_{10} helix between residues D149 - V153. This may introduce even more conformational flexibility to the loop between β7 and β8. As FL HSP27(T151I) was reported to exhibit a similar chaperone activity to the WT[13], we postulated that mutation in this loop may also have an allosteric effect on either the β6+7 or β8 strands it connects, both of which form inter-subunit interfaces.

c27-2(R136W) exhibited the greatest difference in structure as it formed the dimer in an APIII registration, such that the disulphide could not form across the intra-dimer interface. This was the first time that this interface register had been observed for HSP27. Importantly, the cysteine sidechain was observed in two conformations, one of which seems to bind a BME molecule. It will be important to ascertain whether c27-2(R136W) truly accesses the APIII registration to a different degree than the WT. The fact that this adduction remains despite

competition with the local cysteine partner and a very low concentration of the reducing agent after buffer exchange is remarkable in itself, but more importantly it is consistent with the lability of this disulphide bond which has an unusual character suggestive of a functional rather than structural disulphide[31]. A relatively weak disulphide would be able to respond to environmental fluctuations in redox potential, and so act as a 'redox switch'.

HSP27 is the only human sHSP that exhibits this intra-dimer disulphide, though plasticity at this interface is seen across the sHSP family, with the exchange of monomers within the dimeric building block from which oligomers are formed. For example, the oligomers of plant sHSPs, which exhibit the non-metazoan β2-β6 intra-dimer interface, have been seen to exchange monomers as well as the predominant dimer exchange[32], and the intra-dimer interface of metazoan sHSPs seems particularly labile with both monomer and dimers accounting for the exchange of αB-crystallin[22]. Indeed, the core domain of αB-crystallin exists in equilibrium between the monomeric and dimeric form[6], which can be biased to the dimer alone on introduction of a disulphide mimicking that in HSP27[12].

Our observation of the two registers of the intra-dimer interface of HSP27 thus suggest that the disulphide acts to regulate the lability of the intra-dimer interface, both in the relative register of the two monomers and in their dissociation and association at this interface. The unusual character of the disulphide, and the activity of the chaperone in oxidative stress conditions, in turn suggest that this disulphide is itself regulated by the cellular redox environment. In this way, it can act as a redox switch, changing the structural and dynamic behaviour of the core domain in response to the cellular environment.

4 Methods

4.1 Materials and reagents

All reagents were purchased from Sigma unless otherwise specified.

4.2 Constructs

Individual point mutations were introduced to c27-2 in the modified pET28a vector (6H-TEV-c27-2) using the QuikChange Lightning mutagenesis kit (Agilent) with primers purchased from IDT and designed using the Agilent QuikChange Primer Design server. Transformation was performed in XL1 or XL10 competent cells (Agilent) and cultures grown in Luria Broth (Fisher) with kanamycin (50 µg/ml) ready for harvest (MiniPrep kit, ThermoFisher). Successful mutation was determined by plasmid sequencing (SourceBioScience).

4.3 Protein expression and purification

Protein expression and purification were undertaken under the same conditions as c27-2 in Chapter 4: 1l cultures of BL21 (DE3) Gold cells (Agilent) were grown in Luria Broth (Fisher) with kanamycin (50 µg/ml) at 37°C until OD_{660} = 0.6-0.8, then induced with IPTG at 500 µM for three hours. Cells were harvested by centrifugation, washed in PBS and stored at -80°C. Cell pellets were resuspended and disrupted by sonication or microfluidisation in Buffer A (50 mM Tris, 300 mM NaCl, 20 mM imidazole, 5 mM BME, pH 8) supplemented with 1 EDTA-free protease inhibitor cocktail tablet (one per 1L pellet, Roche), followed by clarification by centrifugation. The first purification step was immobilised metal affinity chromatography (IMAC) conducted on a HisTrap HP column and Akta FPLC system (both GE Life Sciences)

in Buffer A with elution in Buffer B (50 mM Tris, 300 mM NaCl, 500 mM imidazole, 5 mM BME, pH 8), followed by a desalting column (HiPrep, GE Life Sciences) or overnight dialysis in Buffer DS (20 mM Tris, 150 mM NaCl, 20 mM imidazole, 5 mM BME, pH 8). Overnight cleavage with TEV protease (S219V mutant purified in-house, pRK793, Addgene plasmid #8827) was conducted at room temperature and followed by a reverse IMAC step in Buffer A and then size exclusion chromatography (SEC) in SEC Buffer (20 mM Tris, 150 mM NaCl, pH 8) on a Superdex 75 column 26/60 (GE Life Sciences).

4.4 Crystallisation trials

The proteins were screened at 14.8mg/ml in SEC buffer with 1.78 mM CTP using sitting-drop vapour-diffusion in 96-well plates and the commercial screens PactPremier, JCSG+, 'Crystal Screen', PegRx and Membrane Gold (this screen is marketed for the crystallisation of membrane proteins, but occasionally is successful for soluble protein crystals). The ratios of the well solution to protein solution were 0.1 µl: 0.2 µl, 0.1 µl: 0.1 µl and 0.2 µl: 0.1 µl with 80-90 µl of well solution in the reservoir. Initial hits were then optimised with hand-set hanging drop vapour-diffusion and sitting-drop vapour diffusion in 1:1, 1:2 and 2:1 ratio of the protein-peptide solution to well solution. The drops were 2-4.5 µl with a 1ml volume of the reservoir solution (the reservoir solutions which gave the most successful crystallisation are listed in Table 5 – 5). For R136W, the most successful crystal was harvested from the initial screen in 'Crystal Screen' in the 96-well plate, with drop size 0.2 µl in 1:1 protein: well solution. Crystallisation occurred at 20°C in all cases, and the crystals were then harvested

with brief transfer into the cryoprotectant listed. Typically crystals began to form within one week, and were fully formed within three weeks.

Construct	Reservoir Solution	Crystallisation Method	Cryoprotectant
c27-2(WT)	0.2M Ammonium Sulphate, 0.1M Sodium cacodylate trihydrate pH6.5, 22-24 % w/v PEG 8000	Sitting drop vapour diffusion	20% v/v glycerol
c27-2(R127W)	- 0.1 M Citric acid pH 3.5, 2.8-3.2 M Sodium chloride - 0.1 M Sodium acetate trihydrate pH 4.5, 2.8-3.2 M Sodium chloride - 0.2 M sodium iodide, 18-22 % w/v PEG 3350	Sitting drop vapour diffusion	20% v/v glycerol
c27-2(S135F)	0.2 M sodium citrate, 0.09-0.11 M Bis Tris propane, pH 6.5, 16-26 % w/v PEG 3350	Sitting drop vapour diffusion	20% v/v glycerol
c27-2(R136W)	0.1M SPG buffer pH4, 25 % w/v PEG 1500	Sitting drop vapour diffusion	20% v/v glycerol
c27-2(T151I)	0.1M MES monohydrate pH 6.5, 1.6M Magnesium sulfate heptahydrate	Hanging-drop vapour diffusion	Sodium formate (very briefly immersed)
c27-2(T164A)	0.2M Ammonium sulphate, 0.1M Sodium cacodylate trihydrate pH 6.5, 30 % w/v PEG 8000	Sitting drop vapour diffusion	20% v/v glycerol

Table 5 – 5 Crystallisation conditions for the c27-2 constructs. The conditions for R127W and S135F refer to those in which crystals were developed, even though the diffraction quality of the crystals was poor. For WT, R136W, T151I and T164A, the conditions listed correspond to the crystal which gave the best diffraction data and are presented in this chapter. The 20% v/v glycerol cryoprotectant was prepared with the well solution listed, whereas sodium formate (7M) was prepared in water.

4.5 Structure determination

Data for c27-2(R136W), c27-2(T151I) and c27-2(T164A) were collected on the ID29 beamline at the European Synchrotron Radiation Facility, Grenoble, France using a radiation

wavelength of 0.979 Å at 100 K. They were then processed using XDS and XSCALE[33]. The unit cell dimensions and most appropriate space groups are listed in Table 5 – 6 along with the key data collection and refinement statistics. Molecular replacement with the c27-2(WT) structure as a search molecule was conducted using Phaser[34].

Refinement was undertaken with Refmac[35] within the CCP4 interface[36]. Coot[37] was used to analyse the electron density maps and correct conformational abnormalities that were detected using the Coot and Molprobity[38] validation tools. Ramachandran Statistics within Coot classified 97.47 % of c27-2(R136W) in preferred regions, 2.53 % in allowed regions and 0 % as outliers.

For c27-2(T151I), this was 94.25 % in the preferred regions, 5.75 in allowed regions and 0 % as outliers. The residues modelled in the allowed regions are glycines, valines, glutamic acids, aspartic acid and threonine.

Lastly, for c27-2(T164A), this was 96.43 % in preferred regions, 2.98 % allowed and 0.6 % as outliers. The outlier corresponds to the S75 chain B, which lies on the turn into the β9 strand from the β8 strand. All the residues in the allowed regions of the Ramachandran plot are valines, aspartic acids and lysines.

Table 5 – 6 (overleaf) Data collection and refinement statistics for c27-2(R136W), c27-2(T151I) and c27-2(T164A). The values for c27-2(WT) were presented in Chapter 4, but are reproduced here for comparison. †This is the value for the mean ($I/\sigma(I)$). *The mean residue temperature factor B was calculated with the BAverage program within the CCP4[36] suite. **Mean B factor across residues of both protein chains. ‡(Chain A length + peptide, Chain B length). †† Mean B factor across residues of CTP alone. R_{meas}[39] is the multiplicity-independent measure of data quality where $R_{meas} = \frac{\sum_h \sqrt{\frac{n_h}{n_h-1}} \sum_i^{n_h} |\bar{I}_h - I_{h,i}|}{\sum_h \sum_i^{n_h} I_{h,i}}$ for observation i of reflection h with multiplicity n_h, intensity $I_{h,i}$ and average $\bar{I}_h$. R_{work} and R_{free} are defined in Chapter 1. V_M is the Matthew's coefficient (the unit cell volume divided by the mass of protein (and peptide) in the unit cell[40]) and was calculated using the Matthews_coef tool in CCP4[36], along with the solvent content of the crystal.

	c27-2	**R136W**	**T151I**	**T164A**
		Data collection		
Crystal system, space group	Tetragonal, $P\,4_3\,2_1\,2$ (#96)	Orthorhombic C 2 2 2_1 (#20)	Tetragonal $P\,4_3\,2_1\,2$ (#96)	Tetragonal $P\,4_3\,2_1\,2$ (#96)
Unit cell parameters a, b, c (Å) (α, β, γ (°))	56.06, 56.06, 166.29 (90, 90, 90)	34.172, 78.249, 69.080, (90, 90, 90)	52.760, 52.760, 178.210 (90, 90, 90)	55.71, 55.71, 166.72 (90, 90, 90)
Resolution range (Å)	28.69 – 2.25 (2.31-2.25)	23.20 - 1.45 (1.48-1.45)	50.59 – 3.10 (3.18-3.10)	28.63 – 2.90 (2.98-2.90)
No. of measurements	332431 (24472)	194192 (6810)	119281 (8823)	75093 (5751)
No. of unique reflections	13304 (932)	16793 (825)	5070 (344)	6328 (442)
Redundancy	25.0 (26.3)	11.6 (8.3)	23.5 (25.6)	11.9 (13.0)
Completeness (%)	99.9% (100%)	99.7% (97.2%)	99.8% (100.3%)	99.9 % (100.2%)
$I/\sigma(I)$	32.30 (2.24)	13.7 (1.9)†	21.43 (2.62)	14.17 (1.33)
Wilson B value (Å^2)	59.66	12.5	91.338	93.021
R_{meas}	0.077 (1.821)	0.161 (2.983)	0.131 (2.034)	0.131(2.638)
$CC_{1/2}$	1.000 (0.804)	0.998 (0.438)	0.999 (0.811)	0.999 (0.505)
		Refinement		
Resolution (Å)	28.69 – 2.25 (2.308-2.250)	39.12-1.45 (1.488-1.450)	50.59 – 3.10 (3.180 - 3.100)	28.63 – 2.90 (2.974-2.900)
R_{work}	0.2215	0.18188	0.25192	0.20634
No. reflections	12638	15926	4769	5688
R_{work} / R_{free}	0.2215 / 0.2645	0.18188/0.20691	0.25192/0.28670	0.20634/ 0.26854
No. atoms				
Protein	1297	629	1312	1285
Ligand/ion	73	65, 4 (BME)	73	73, 30 (GOL), 25 (SO4)
Water	34	37	0	0
R_{free}	0.2645	0.20691	0.28670	0.26854

Amino acids‡	88 + 9, 78	78 + 9	88 + 9, 83	88 + 9, 77
No. of water molecules	34	37	0	0
No. of others	5 sulphate, 4 glycerol	1 β-mercaptoethanol	0	5 sulphate, 5 glycerol
Mean B^* (Å^2)				
Protein	64.2**	21.8	105.6**	103.7**
Water	52.42	28.4	N/A	N/A
Other	78.9	26.1	122.0	123.6
	(SO4, CTP and glycerol)	(CTP and BME)	(CTP)	(SO4, CTP and glycerol)
Ligand/ion††	75.7	22.7	122.0	106.2
RMSD from target values				
Bond lengths (Å)	0.017	0.041	0.012	0.018
Bond angles (°)	1.855	3.134	1.733	1.840
Dihedral angles (°)	7.361	6.807	7.255	7.269
Ramachandran				
Most favoured (%)	94.67	97.47	94.25	96.43
Allowed (%)	4.73	2.53	5.75	2.98
Disallowed (%)	0.59	0.00	0	0.60
V_M (Å^3/Da)	3.16	2.12	3.00	3.14
Solvent content (%)	61.09	42.01	58.96	60.82

RMSD values were calculated with the 'align' tool in PyMol (The PyMOL Molecular Graphics System, Version 1.2r3pre, Schrödinger, LLC) and the images prepared with the same package.

4.6 F_o-F_c map

The F_o-F_c maps for the c27-2(R136W) structure in Figure 5 – 11 and Figure 5 – 12 were calculated using the Run FFT – Create Map program within the CCP4i suite. They are represented in PyMol within 4 Å of the BME molecule and 11 Å of residues 96-99 in the β3 and at a level of 0.3769 e.Å^{-3} and 0.3900 e.Å^{-3} respectively which corresponds to a cutoff of RMSD (σ) = 3 in Coot.

5 References

1. Ismailov, S. M. *et al.* A new locus for autosomal dominant Charcot-Marie-Tooth disease type 2 (CMT2F) maps to chromosome 7q11-q21. *Eur. J. Hum. Genet.* **9,** 646–650 (2001).

2. Evgrafov, O. V *et al.* Mutant small heat-shock protein 27 causes axonal Charcot-Marie-Tooth disease and distal hereditary motor neuropathy. *Nat. Genet.* **36,** 602–606 (2004).

3. Lin, K.-P. *et al.* The Mutational Spectrum in a Cohort of Charcot-Marie-Tooth Disease Type 2 among the Han Chinese in Taiwan. *PLoS One* **6,** e29393 (2011).

4. Indu, S., Kochat, V., Thakurela, S., Ramakrishnan, C. & Varadarajan, R. Conformational analysis and design of cross-strand disulfides in antiparallel β-sheets. *Proteins* **79,** 244–60 (2011).

5. Kijima, K. *et al.* Small heat shock protein 27 mutation in a Japanese patient with distal hereditary motor neuropathy. *J. Hum. Genet.* **50,** 473–476 (2005).

6. Hilton, G. R. *et al.* C-terminal interactions mediate the quaternary dynamics of αB-crystallin. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **368,** 20110405 (2013).

7. Chalova, A. S., Sudnitsyna, M. V, Strelkov, S. V & Gusev, N. B. Characterization of human small heat shock protein HspB1 that carries C-terminal domain mutations associated with hereditary motor neuron diseases. *Biochim. Biophys. Acta* **1844,** 2116–2126 (2014).

8. Joosten, R. P. *et al.* A series of PDB related databases for everyday needs. *Nucleic Acids Res.* **39,** D411–D419 (2011).

9. Kabsch, W. & Sander, C. Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22,** 2577–2637 (1983).

10. Bagnéris, C. *et al.* Crystal Structures of α-Crystallin Domain Dimers of αB-Crystallin and Hsp20. *J. Mol. Biol.* **392,** 1242–1252 (2009).

11. Laganowsky, A. *et al.* Crystal structures of truncated alphaA and alphaB crystallins reveal structural mechanisms of polydispersity important for eye lens function. *Protein Sci.* **19,** 1031–43 (2010).

12. Hochberg, G. K. A. *et al.* The structured core domain of αB-crystallin can prevent amyloid fibrillation and associated toxicity. *Proc. Natl. Acad. Sci. U. S. A.* **111,** E1562–70 (2014).

13. Almeida-Souza, L. *et al.* Increased monomerization of mutant HSPB1 leads to protein hyperactivity in Charcot-Marie-Tooth neuropathy. *J. Biol. Chem.* **285,** 12778–86 (2010).

14. Clark, A. R., Naylor, C. E., Bagnéris, C., Keep, N. H. & Slingsby, C. Crystal structure of R120G disease mutant of human αb-crystallin domain dimer shows closure of a groove. *J. Mol. Biol.* **408,** 118–134 (2011).

15. Rajagopal, P., Liu, Y., Shi, L., Clouser, A. F. & Klevit, R. E. Structure of the α-

crystallin domain from the redox-sensitive chaperone, HSPB1. *J. Biomol. NMR* 0–5 (2015). doi:10.1007/s10858-015-9973-0

16. Capponi, S. *et al.* HSPB1 and HSPB8 in inherited neuropathies: study of an Italian cohort of dHMN and CMT2 patients. *J. Peripher. Nerv. Syst.* **16,** 287–294 (2011).

17. Rajagopal, P. *et al.* A conserved histidine modulates HSPB5 structure to trigger chaperone activity in response to stress-related acidosis. *Elife* **4,** 1–21 (2015).

18. Bova, M. P., Mchaourab, H. S., Han, Y. & Fung, B. K.-K. Subunit Exchange of Small Heat Shock Proteins: Analysis of oligomer formation of αA-crystallin and HSP27 by fluorescence resonance energy transfer and site-directed truncations. *J. Biol. Chem.* **275,** 1035–1042 (2000).

19. Jehle, S. *et al.* Solid-state NMR and SAXS studies provide a structural basis for the activation of αB-crystallin oligomers. *Nat. Struct. Mol. Biol.* **17,** 1037–1042 (2010).

20. Houlden, H. *et al.* Mutations in the HSP27 (HSPB1) gene cause dominant, recessive, and sporadic distal HMN/CMT type 2. *Neurology* **71,** 1660–1668 (2008).

21. Ikeda, Y. *et al.* A clinical phenotype of distal hereditary motor neuronopathy type II with a novel HSPB1 mutation. *J. Neurol. Sci.* **277,** 9–12 (2009).

22. Hochberg, G. K. A. & Benesch, J. L. P. Dynamical structure of αB-crystallin. *Prog. Biophys. Mol. Biol.* **115,** 11–20 (2014).

23. Rajasekaran, N. S. *et al.* Human αB-Crystallin Mutation Causes Oxido-Reductive

Stress and Protein Aggregation Cardiomyopathy in Mice. *Cell* **130,** 427–439 (2007).

24. Arrigo, A.-P. & Gibert, B. Protein interactomes of three stress inducible small heat shock proteins: HspB1, HspB5 and HspB8. *Int. J. Hyperth.* **29,** 409–22 (2013).

25. Arrigo, A. P. Human small heat shock proteins: Protein interactomes of homo- and hetero-oligomeric complexes: An update. *FEBS Lett.* **587,** 1959–1969 (2013).

26. Arrigo, A.-P. Hsp27: Novel Regulator of Intracellular Redox State. *IUBMB Life (International Union Biochem. Mol. Biol. Life)* **52,** 303–307 (2001).

27. Ilbert, M. *et al.* The redox-switch domain of Hsp33 functions as dual stress sensor. *Nat. Struct. Mol. Biol.* **14,** 556–563 (2007).

28. Richter, K., Haslbeck, M. & Buchner, J. The heat shock response: life on the verge of death. *Mol. Cell* **40,** 253–66 (2010).

29. Janda, I. *et al.* The Crystal Structure of the Reduced, Zn2+-Bound Form of the B. subtilis Hsp33 Chaperone and Its Implications for the Activation Mechanism. *Structure* **12,** 1901–1907 (2004).

30. Chakraborty, K. *et al.* Protein Stabilization by Introduction of Cross-Strand Disulfides. *Biochemistry* **44,** 14638–14646 (2005).

31. Wouters, M. A., Fan, S. W. & Haworth, N. L. Disulfides as Redox Switches: From Molecular Mechanisms to Functional Significance. *Antioxid. Redox Signal.* **12,** 53–91 (2010).

32. Sobott, F., Benesch, J. L. P., Vierling, E. & Robinson, C. V. Subunit exchange of multimeric protein complexes: real-time monitoring of subunit exchange between small heat shock proteins by using electrospray mass spectrometry. *J. Biol. Chem.* **277,** 38921–38929 (2002).

33. Kabsch, W. XDS. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66,** 125–132 (2010).

34. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40,** 658–674 (2007).

35. Murshudov, G. N. *et al. REFMAC* 5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **67,** 355–367 (2011).

36. Winn, M. D. *et al.* Overview of the CCP4 suite and current developments. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **67,** 235–242 (2011).

37. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66,** 486–501 (2010).

38. Chen, V. B. *et al.* MolProbity : all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66,** 12–21 (2010).

39. Diederichs, K. & Karplus, P. A. Improved R-factors for diffraction data analysis in macromolecular crystallography. *Nat. Struct. Biol.* **4,** 269–275 (1997).

40. Matthews, B. W. Solvent content of protein crystals. *J. Mol. Biol.* **33,** 491–497 (1968).

Chapter 6: Redox Regulation of HSP27

1 Introduction

HSP27 is unusual among its paralogs, being the only human sHSP known to form a disulphide bond across the intra-dimer interface. This disulphide has been suggested to mediate the response of HSP27 to oxidative stress[1,2]. Functional disulphides are receiving greater recognition as a regulatory mechanism, whereby their oxidation and reduction moderate the function of the protein[3]. Despite the generally reducing nature of intracellular compartments, oxidised cysteines are both present in the cell and proposed to act as a sink for reactive oxygen species[4,5].

The HSP27 disulphide is particularly noteworthy as cross-strand disulphides are rare[6] and have been associated with redox activity. Specifically, their high torsional energy indicates a functional rather than structural role and they have been postulated to act as a 'redox switch', mediating a protein's redox regulation[3,7,8]. Indeed, cross-strand disulphides introduced to a model protein Top7 by mutation all showed mild redox activity[6]. Notably disulphides in hydrogen-bonded registered pairs, such as that found in c27-2, destabilised Top7 in chemical denaturation. In contrast, disulphides in non-hydrogen-bonded registered pairs were generally stabilising[6].

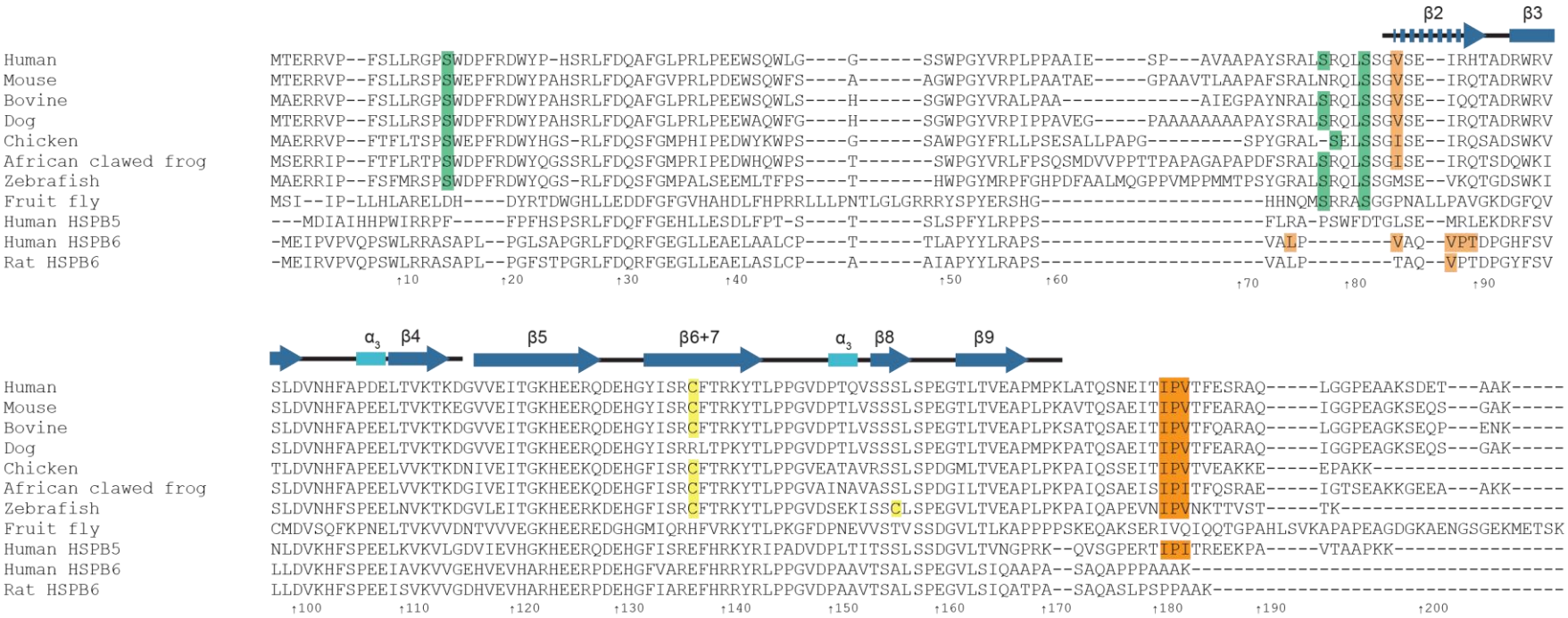


Figure 6 – 1 Sequence alignment of HSPB1 (HSP27) genes identified in a range of animals, along with αB-crystallin and HSP20. Human (homo sapiens, P04792); mouse (mus musculus, P14602); bovine (bos taurus, Q3T149); dog (canis lupus familiaris, P42929); chicken (gallus gallus, Q00649); African clawed frog (xenopus laevis; Q66KY8); zebrafish (danio rerio, Q5PR64); fruit fly (drosophila melanogaster, P02518; notably this is not considered an ortholog of vertebrate HSP27 despite being labelled as an HSP27 ortholog in Uniprot); as well as human HSPB5 (αB-crystallin, homo sapiens, P02511); human HSPB6 (HSP20, homo sapiens, O14558); and rat HSPB6 (rattus norvegicus, P97541). Cysteines (yellow) are observed in the equivalent position in the β6+7 strand in most vertebrate HSP27. The phosphorylatable serines in human HSP27 (green) and C-terminal IPV motif (orange) are largely conserved among its orthologs, except in fruit fly. The C-terminal IXI motif is also observed in the paralogous αB-crystallin but not in HSPB6. Rather, β4-β8 patching by the N-terminus was observed in HSPB6 core domain constructs[9,10] (residues with side chains buried into the groove in at least one conformer are shown in light orange), similar to that observed with V85 in c27-2 here. A hydrophobic at this point is observed in several of the orthologs (light orange). Secondary structure (arrow – β-strand; box – 3_{10} α-helix) was assigned using DSSP[11,12] on the chain A coordinates of the c27-2 model presented herein, except for the β2 strand which was disordered but had previously been observed bound to the β3 strand (in c27-1(4MJH)[13], dashed arrow). Alignment was performed using the Clustal Omega webserver[14,15].

The cysteine on the β6+7 strand is conserved among HSP27 orthologs in vertebrates (Figure 6 – 1), indicating that the thiol side chain plays an important role in HSP27 structure-function. It has seemingly been lost from HSP27 in dog so it not essential, but as the β6+7 strand formed the intra-dimer interface in all metazoan sHSP structures solved to date[16–18] it is likely that this cysteine can form an intra-dimer disulphide in the HSP27 ortholog complexes too. This implies that the cross-strand disulphide has not been lost in selection despite the entropic cost of its formation and its possibly sterically strained nature, suggesting that it influences the efficacy of the protein.

HSP27 is well-established as a 'redox chaperone[19]'. HSP27 over-expression protected against oxidative stress[20] and conversely oxidative stress is associated with HSP27 down-regulation[21]. On the other hand, reductively stressed cells also demonstrated significant overexpression of the mouse HSP27 ortholog in transgenic mouse hearts (a 25-fold increase in the insoluble fraction relative to the non-stressed control)[19]. The protein was up-regulated to a much greater degree than HSP70 or HSP90 chaperones (less than 5-fold increase, with a similar modest increase for all three chaperones in the soluble fraction), indicating a preference for HSP27 chaperone in redox-compromised cells. The reductive stress resulted from the expression of mutant αB-crystallin (R120G) which also exemplifies the interconnected nature of the sHSP network. In terms of the protein itself, oxidation diminished HSP27 chaperone activity *in vitro* yet increased thermoresistance[1].

In Chapter 5, I described the effect of disease-related mutations on the structure of the HSP27 core domain and presented a rationale for the effect of mutation on the local environment of

the residue. The most striking difference was an alternative AP registration state on introduction of the R136W mutation with concomitant absence of the disulphide.

It has been suggested the registration of cABC and the sHSPs may regulate function[22,13]. However WT cABC exhibited a very similar CCS and ^{1}H-^{15}N HSQC spectrum to an APII-locked form of cABC, produced by introducing the equivalent cysteines to those in HSP27 and air oxidation, and showed equivalent chaperone activity[13]. This indicated that cABC predominantly exists in the APII form, and the ability of the WT to access the API and APIII dimeric states (or the monomeric form) does not ameliorate its activity. Similarly, SAXS data implied that the core domain of HSP27 mostly exists in the APII dimeric form, and could be disrupted to the APIII form on lowering the pH[23]. It is therefore possible that some disease-linked mutations interrupt the otherwise preferred and active APII dimeric form.

In contrast to c27-2(R136W), T151I and T164A c27-2 mutants exhibited little structural difference to the WT (Chapter 5). Almeida-Souza *et al.* suggested that T151I may have a different pathogenesis to R136W (and R127W and S135F), based on the different phenotypes seen in motor neuropathy patients carrying the former or latter mutations[24].

An R127L mutation reduced the heat stress tolerance of HSP27 in primary fibroblasts of Charcot-Marie-Tooth disease patients[25]. However, stable overexpression of R127W and S135F HSP27 FL in SH-SY5Y neuronal cells produced an increase in thermotolerance relative to the WT, and seemed to be better molecular chaperones *in vivo*, as did R136W to a lesser extent[24]. All three mutants bound their client proteins more strongly than the WT, suggesting neuropathy may be mediated by aberrant binding to a neuron-specific partner.

Notably, heat shock produced an increase in the reduced HSP27 WT population, which returned to basal levels within 24 hours[24]. Moreover, Western blots on SH-SY5Y cell lysates showed R127W and S135F FL HSP27 formed fewer oxidised dimers than the WT. This experiment did not control for oxidation in air, however the differential behaviour of the mutants indicates that they may affect the formation of this bond[24].

Based on the unusual nature of the intra-dimer disulphide, the response of HSP27 to oxidative stress, and the evidence indicating that at least some mutations cause differential formation of the disulphide and access to the AP states, we hypothesised that the disulphide moderates the activity of the protein, and that mutation interrupts the formation of this bond.

In this chapter, I detail the work undertaken to test our hypotheses, broadly: the disulphide can form within the cell; the disulphide affects HSP27 dynamics and activity; reduction and mutation interrupts the formation of the disulphide such that the APIII state can be populated more frequently. The investigations discussed herein have given us a broad insight on the nature of this disulphide and its role in disease.

2 Results and Discussion

2.1 The disulphide forms within the cell

To understand the significance of the disulphide bond within the cell, Western blots against endogenous HSP27 in human cell lysates were undertaken (Figure 6 – 2). Initial immunoblots with anti-HSP27 antibody on human embryonic kidney (HEK) 293T cells and human foreskin

fibroblasts (HFF) were prepared without iodoacetamide (IAM) or trichloroacetic acid (TCA) precipitation, to determine whether any dimeric population could be observed (Figure 6 – 2a). A mixture of monomeric and dimeric HSP27 has been found in human fibroblasts previously[25]. As the disulphide cross-links two subunits, the presence of a dimer detected in denaturing SDS-PAGE would represent a disulphide-containing population. Without reducing agent, there is a clear band corresponding to the dimer in the HFF sample whereas the HEK cells contained very little dimer, as did both samples on reduction (demonstrating that the dimeric band constitutes the oxidised species). All four samples had a substantial monomer population. The different degree of oxidation in the two cell types indicates that the dimer observed in the HFF cells is not an artefact of air oxidation, as the HEK sample had an equal exposure to air. This indicates that the disulphide is likely forming within the HFF cell.

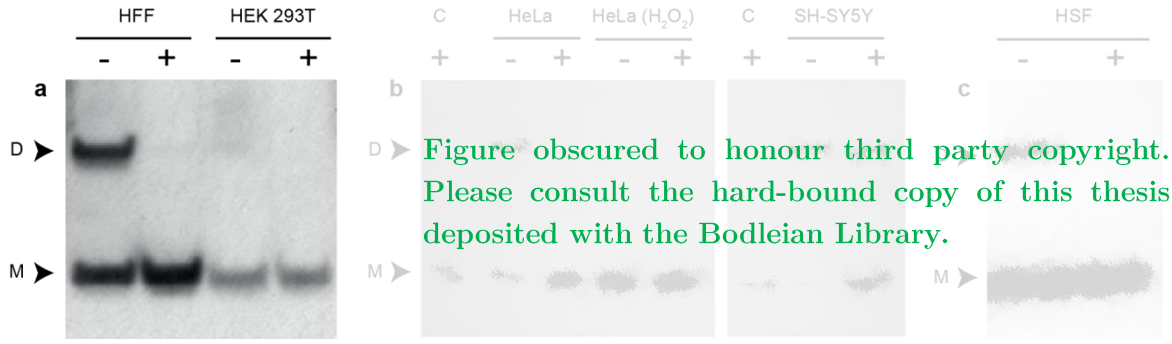


Figure 6 – 2 Immunoblots show differential oxidation of HSP27 endogenously expressed in various human cell lysates. (a) Initial results in HEK 293T and HFF cells showed differing oxidation profiles for cell lines treated under the same conditions. SDS-PAGE of cell lysates (15.0-15.5 µg loaded) with and without reducing agent (DTT, '+' and '-' respectively) and blotted with anti-HSP27 antibody reveals two species corresponding to the HSP27 monomer and disulphide-bonded dimer (M and D respectively). HEK 293T cells contain little HSP27 dimer whereas HFF show a significant population of oxidized dimers, which disappear on incubation with DTT. (b and c) TCA acid precipitation followed by treatment with IAM confirmed the presence of the dimer *in vivo* in HeLa cells (10 µg loaded), SH-SY5Y ((22 µg) cells and HSF, which again decreased in the presence of reducing agent (BME). Oxidative stress induced by hydrogen peroxide in HeLa cells (b) caused HSP27 to be almost exclusively reduced. 'C' – purified HSP27 control, 15 ng. The Western blots in (b) and (c), which were two independent experiments, were performed by Anthea Rote, University of Wollongong.

This result was then recapitulated with HeLa, SH-SY5Y neuroblastoma and human skin fibroblasts (HSF) (Figure 6 – 2b, c). TCA acid precipitation and capping with IAM were employed to capture the thiols in their cellular-state. Both HeLa and fibroblast lysates showed a significant oxidised dimer population which was reduced in the presence of reducing agent. The HSP27 population of the SH-SY5Y cells was mainly oxidised, so much so that there was still a substantial population of disulphide-locked dimers after addition of reducing agent. This further demonstrated the differential oxidation within different cell lines, which may arise from cell growth stage as well as the particular cell type. Finally, oxidative stress induced by application of hydrogen peroxide to HeLa cells showed a decrease in oxidised HSP27 (Figure 6 – 2b). This surprising result highlighted that the response of HSP27 to oxidative stress may follow a complex mechanism, with other contributing factors other than the oxidative potential of the environment.

These experiments illustrate that the disulphide is relevant in cellular FL HSP27, and that the presence of the disulphide may be modulated by multiple factors *in vivo*, including growth stage and cell type as well as extraneous oxidative stress. We next investigated whether breaking this disulphide could lead the protein to occupy the additional registration states at the dimeric interface.

2.2 Using IM-MS to probe the conformation of the core domain: solution-mediated effects

To investigate whether the different antiparallel registration states (Chapter 5) were observable outside the constraints of the crystal lattice, we employed IM-MS. This is discussed in detail in Chapter 1, but briefly, in IM-MS an additional ion mobility cell is introduced to

the mass spectrometer. Ions are separated according to their drift time through this cell, which depends on their mobility K and ultimately their collision cross section (CCS). Thus, the CCS of a particular ion can be calculated from its drift time through the mass spectrometer. In the TriWave IM cell used here, the CCS is determined from the drift time of the ion using a calibration curve with calibrants of known CCS and mass[26].

As the R136W structure was determined from crystals grown in pH 4 solution and the WT c27-2 crystals were formed in pH 6.5 solution, it was important to determine whether the different AP states observed could be a result of the different pH conditions. The pH difference could affect the protonation state of the sidechains of aspartic acid, glutamic acid, histidine and cysteine[†] (pK_a of 3.7 – 8.1) which are involved in hydrogen bonding across the interface.

Interruption of this hydrogen bonding network on lowering the pH could thus account for the difference in AP states. As the sHSPs are responsive to pH[27], movement to the APIII state may thus be part of the protective activity of the WT protein rather than a result of the R136W mutation. Indeed, Baranova *et al.* have previously suggested that low pH may increase the propensity for the APIII state in the HSP27 core domain[23]. This prediction followed SAXS

[†] Disulphide bond formation is favoured at basic pH[61] because they are formed via the thiolate ion. Therefore if the cysteine thiols are protonated, formation of the disulphide bond between any free cysteines would be hindered. All of the proteins were prepared at pH 8 with the final stages in air oxidising conditions. We thus assume the thiols have reached equilibrium between the oxidised (disulphide) and reduced (thiol) forms before crystallisation. Thus any later difference in disulphide population due to pH would be due to the interruption of this equilibrium.

experiments on a core domain construct of HSP27: at pH 5.1 the data was better modelled by the rat HSP20 core domain dimer in APIII than the cABC dimer in APII, despite the cABC model being adequate for the HSP27 construct at neutral pH[23].

Another reason that the R136W might be in the APIII state is because of the adducted BME. All the proteins were purified with the same protocol and BME concentration. However, it could be that the small population with BME adducted were the only ones that could form crystals for the R136W variant. As the BME prevents formation of the disulphide bond, it may allow the dimer to access the APIII interface, which has more hydrogen bonds across it than the APII state (Table 5 – 2). If this is the case, we would also expect to find the dimer in the APIII state on reduction of the disulphide bond.

In order to test these two hypotheses, we measured the CCS of the WT c27-2 at pH 4 and 6.5, or under oxidising or reducing conditions. We hypothesised that if the different registration state was a result of reversible modification rather than the mutation, the WT c27-2 should also be able to access the APIII in the same conditions.

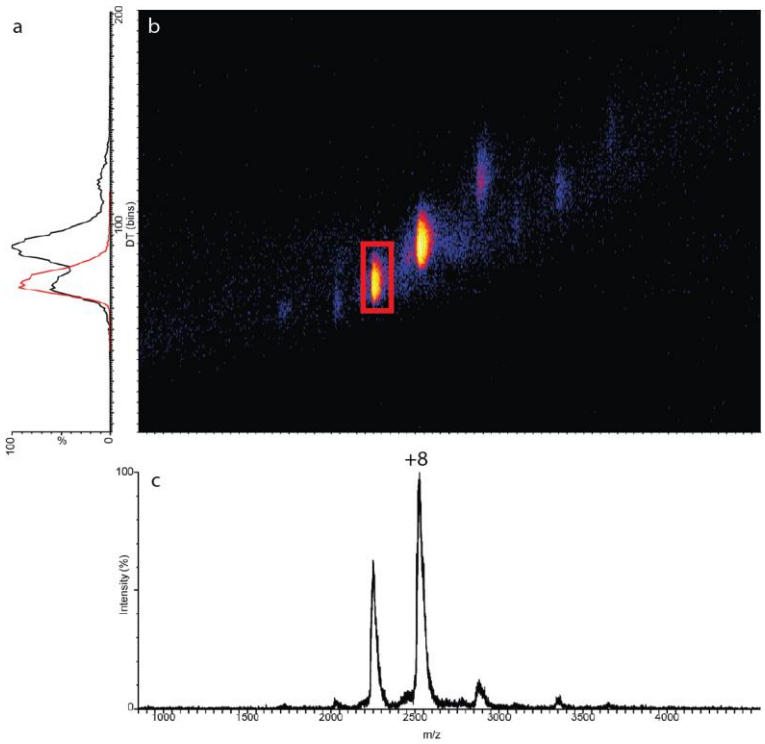


Figure 6 – 3 Drift time extraction for the +9 charge state of the c27-2 WT oxidised dimer. IM-MS data can be displayed as a 2D plot (b) incorporating the 1D mass spectrum (c) and 1D drift time profile (a, black trace) for all the charge states accessed by the protein. The drift time for each individual peak can be extracted, for example the dimer +9 charge state (b, red box) to determine the drift time for all ions within a particular peak (a, red trace). The drift time of a particular charge state can be used to determine its CCS.

Figure 6 – 3 shows the types of data generated in this experiment. Panel (a) shows the result of the ion mobility separation for the oxidised c27-2 WT dimer. The profile of ion intensity with bin number for c27-2 WT is shown in black. There are three obvious peaks, with those at a higher charge having a lower DT. When all the drift time bins are combined, the corresponding mass spectra detected for each bin are also combined, resulting in the familiar mass spectrum for c27-2 WT.

Therefore, this experiment results in a two-dimensional spectrum, as shown in Figure 6 – 3b, with the drift time dimension on the y-axis and mass spectrum on the x-axis. Normalised intensity is shown on a log scale with blue indicating low abundance and yellow very high

relative abundance. The four most abundant peaks correspond to the +10 to +7 charge states of the dimer, and there are also two lowly abundant charge states of monomer and tetramer, as discussed in Chapter 4. As drift time depends on charge, each charge state has a different drift time. It is possible to extract the drift times for each individual charge state, as demonstrated for the +9 charge state (Figure 6 – 3b). From this, the CCS of the protein can be calculated[26,28].

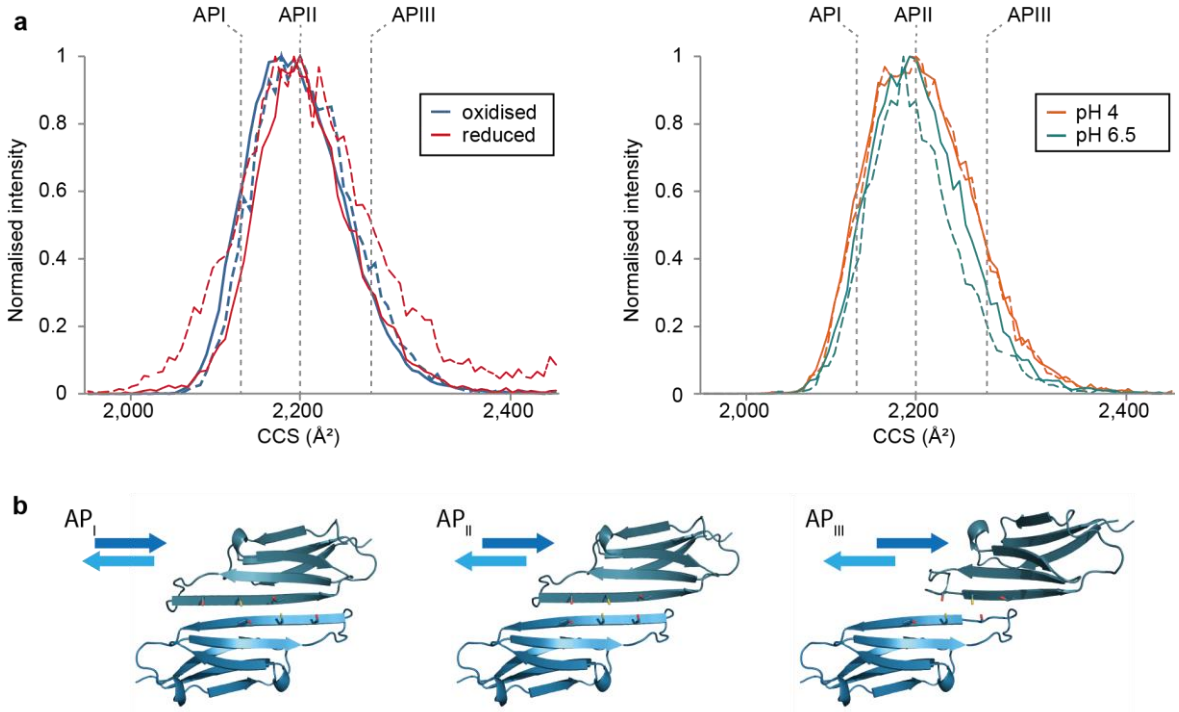


Figure 6 – 4 IM-MS shows the core domain is predominantly in the APII register when oxidized at pH4, pH6.5 and pH6.9, and when reduced. (a) The experimental drift time profiles for the +9 charge state of WT c27-2 at 20 µM when oxidized or reduced, or at pH 4 and pH 6.5, plotted on a CCS scale. Dashed lines indicate repeats. Vertical lines indicate the estimated CCS values for the API, APII and APIII structures shown in (b). The solution conditions do not change the population of the distribution of states.

To compare the c27-2(WT) protein in different conditions, the individual drift time profiles for the +9 charge state were extracted (Figure 6 – 4a). Spectra were recorded on 20 µM of the oxidised and reduced c27-2 WT at pH 6.9, and the air oxidised c27-2 at pH 6.5 and 4. These are shown on a scale of CCS, and the expected CCS values of the API, II and III states are

indicated by vertical lines (estimated using the PA algorithm, see Methods), on the simulated API, APII and APIII WT structures (Figure 6 – 4b, see Methods for the construction of these models). This estimated that the APIII would be 103.5% the size of the APII. As we determined that the large majority of the 'oxidised' c27-2 sample at pH 7 was indeed oxidised (Section 2.3) and thus necessarily in the APII dimeric form, we scaled the CCS values so that the estimated APII CCS value aligned with the experimental value determined for the oxidised c27-2 sample. This discrepancy in the estimated and calculated CCS values is mainly due to the removal of the β2 strand from the API-III structural models, so that the simulated value underestimates the true value (see Methods for further discussion).

The drift time profiles of samples in each solution condition are not noticeably different, indicating that there has not been a significant redistribution of the population amongst the different AP registrations. The reduced sample drift times appear more variable as the dimer is much lower in abundance when reduced than in the oxidised samples, but there is not a reproducible change in the width of the profile or in the position of the centroid. Furthermore, as the majority of the oxidised sample must be in the APII state, none of the solution conditions has allowed c27-2 WT to access the API and APIII states in a significantly different manner. This result agrees with SAXS data on a reduced and oxidised HSP27 core domain (with E125A and E126A mutations) which did not produce any noticeable change in the shape of the SAXS curve, suggesting that the disulphide did not produce a significant change on the shape of the core domain[23].

The CCS for the individual charge states for each condition were calculated using the Pulsar software[29], which fits a Gaussian to the individual drift time profiles, and uses the centroid of the peak to calculate the CCS from a calibration curve (Figure 6 – 5). The estimated difference between APII and APIII is much larger than the variation in the data and the variation between the different conditions, thus for all four charge states, the main population of c27-2 is in APII.

The variation in calculated CCS with charge state results from a discrepancy in the mobilities of the calibrants versus c27-2, leading to a slight overestimation of the CCS at charge states higher than z = 9, and conversely an underestimation at charge states lower than z = 9 (Figure 6 – 5). Unfortunately, there are not native-like calibrants available that occupy the same region of the calibration plot as c27-2 (*i.e.* that show the same mobility behaviour) and for which CCS values have been published. However, the c27-2 +9 charge state lies within the mobility range of the calibrants, and from this it is clear that reduction or change in pH does not cause the CCS to vary significantly from that expected for APII. Moreover, each of the other charge states shows a similar small variation in CCS with solution conditions, which is comparable to the variation in the repeats, again demonstrating that the APII remains the dominantly occupied state.

While it is possible that there is a small population of API and APIII that this experiment cannot isolate, it is clear that pH and reduction have not led to a large population of APIII dimers. This suggests that the crystallisation of the APIII R136W is not a result of the

difference in pH or the absence of the disulphide bond inhibited by the adducted BME, and leaves open the possibility that the R136W occupies the APIII more than the WT.

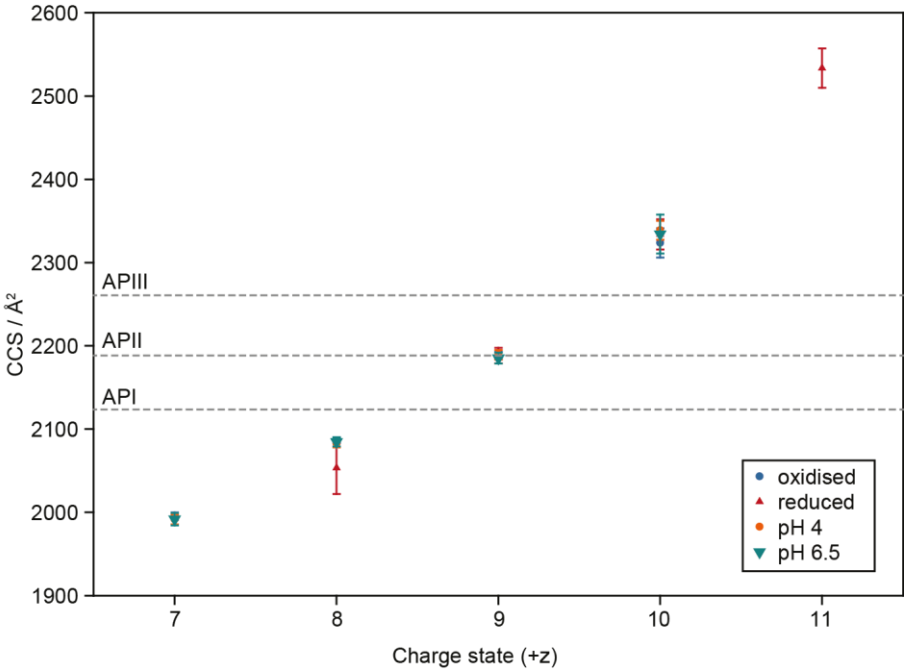


Figure 6 – 5 Calculated CCS values for all charge states in each condition. The variation in CCS between solution conditions is much smaller than the estimated difference between the API, II and III states (dashed line, from Figure 6 – 4b). Error bars show mean ± s.d. for the CCS calculated from two repeats at three different wave heights.

2.3 The disulphide stabilises the intra-dimer interface and significantly slows subunit exchange

The unusual nature of the interface disulphide, and the further intrigue fuelled by the alternative registers observed herein, led us to ask what role the disulphide plays within the SE of the protein. The sHSPs are highly dynamic proteins[30–34], able to exchange monomers and dimers between the larger oligomers. MS is well placed to monitor the successive exchange of subunits between sHSP assemblies and the data can be used to determine the kinetic parameters characterising exchange[35–39]. Moreover, the presence of the disulphide *in*

vivo (Section 2.1) raises enquiry into the lability of this bond – how easily does the bond break?

Exchange of subunits between sHSP oligomers is dependent on several factors. Both monomers, dimers or a combination of the two have been identified as the exchanging species[36,35,37,40]. Their exchange depends on dissociation at the inter-dimer interfaces (and the intra-dimer interface in the case of monomers). They must then associate with an association-competent partner, which could be the oligomer they have just vacated, or a 'new' partner. When SE has been investigated in detail, dissociation has been found to be rate limiting[30,33,35]. The nature of this exchange is susceptible to regulation, for example pH[31] or phosphorylation[27]. The disulphide bridging the inter-subunit interface of HSP27 could be an additional mechanism of regulation, with the oxidised dimer requiring the disulphide to break before SE can occur.

The core domain again represented a useful system to deconvolve the role of these different factors in HSP27. Here we employed successive steps to elucidate the role of the disulphide in the dynamic exchange of the protein. First, we used MS to monitor the exchange of the core domain dimers when the protein was fully reduced, we then followed the exchange of the oxidised core domain with MS before finally using our observations to rationalise the exchange of the full length protein observed with fluorescence resonance energy transfer (FRET).

MS can be used to follow SE in solution with use of an isotopically-labelled equivalent. sHSP expressed in cells grown in media with natural isotope abundance ('unlabelled' protein) will have a lower mass than the same protein expressed with the sole carbon and nitrogen sources

as ^{13}C and ^{15}N, for example. The protein will comprise the same sequence, and should be identical in non-covalent assembly and dynamics (pre-equilibration mitigates any co-translational folding variability that may have occurred in the different media). Thus when the proteins are mixed in solution, any dissociation and association of subunits occurring in the sample should result in oligomers constituting a combination of unlabelled and labelled subunits, with a concomitant variation in mass which can be analysed by native MS in the gas phase. This assumes that the exchange is indiscriminate of the isotopic composition of the subunits. The kinetic-isotope effect on formation of hydrogen bonds in β-sheets is small when labelled with deuterium[41], so the kinetic isotope effect on formation of the HSP27 intra-dimer β-sheet with ^{15}N and ^{13}C labelling is assumed to be negligible, as is the secondary kinetic isotope effect of the C_β on the breaking and forming of the disulphide bond (the kinetic isotope effect is greatest for hydrogen as ^{1}H to ^{2}H isotopic substitution has the largest proportional mass increase). Monitoring SE in this way is well established[13,36,38].

To characterise the SE of the non-covalent dimer, a time course was first conducted with 17.6 μM monomer of labelled and unlabelled c27-2 WT and 100mM DTT (an excess of 5700 reducing agent to disulphide). This high excess was employed following analysis of SDS-PAGE of the sample, which appeared to retain a small population of non-reduced dimers even with a 10700 excess. On reduction, the protein populates both monomers and non-covalent dimers (Figure 6 – 6).

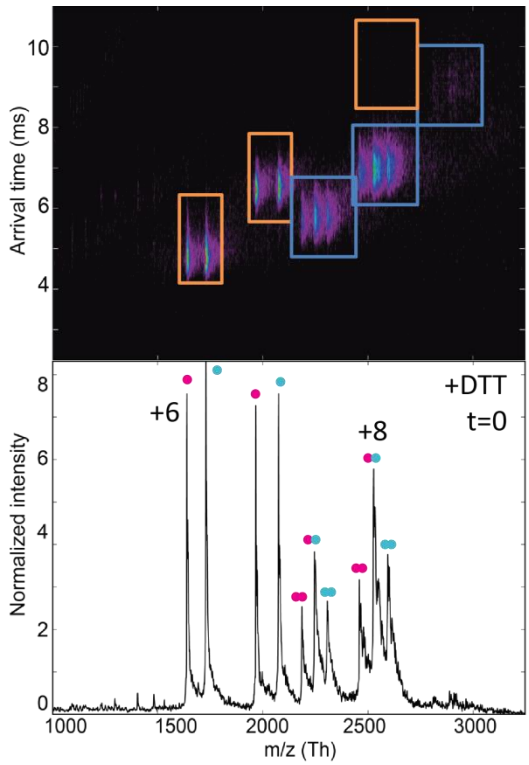


Figure 6 – 6 Ion mobility confers an extra dimension for separation. When reduced, both unlabelled (pink) and labelled (blue) c27-2(WT) occupy the monomer form as well as non-covalent dimers (lower panel, single and double circles respectively). Subunits exchange rapidly, forming new homodimers and heterodimers. IM separation is required to accurately identify the contribution of monomers (yellow boxes) and homodimers to the +8 and +10 dimer charge states (upper panel). This spectrum is representative of all the spectra collected at 't = 0' in the presence of 100 mM DTT with 18 μM each of labelled and unlabelled c27-2.

Here, IM-MS was employed for separation due to its extra drift time dimension. The need for this separation is demonstrated in Figure 6 – 6. In the lower panel is a mass spectrum of the +DTT sample at t = 0 minutes. This is typical of the spectra recorded for the +DTT sample at any value of t (the high number of peak adducts results from the high DTT concentration). There are multiple charge state distributions corresponding to either monomers or dimers of the labelled and unlabelled protein. There is an additional charge state envelope due to the heterodimer formed between one labelled and one unlabelled subunit. Any heterodimer can only be formed by SE, whereas the labelled or unlabelled homodimers may or may not have undergone exchange. It is possible that the +4 monomer peaks and +8 homodimer peaks

overlap (or any +n monomer and +2n dimer peaks), complicating assignment of the relative abundances of the heterodimers and homodimers. When IM-MS is employed, the monomer and dimer distributions are also separated by their drift time at any particular m/z value (yellow and blue boxes, Figure 6 – 6, upper panel). Thus IM-MS facilitated the isolation of individual species underlying the mass spectrum, the abundance of which could be extracted from the area under the peaks. The population of monomers that would overlap with dimers is very subtle, and so the use of IM-MS ensured that the analysis correctly assigned the abundance of each species, and also allowed all available charge states to be used in the analysis (rather than just those with an odd-numbered charge where there is not overlap with monomer peaks). This improved the robustness of the abundance measurements.

When exchange between the dimers has reached equilibrium, they will populate a 1:2:1 ratio of unlabelled homodimer: heterodimer: labelled homodimer (assuming an equal concentration of unlabelled and labelled homodimers are mixed). In the +DTT sample, this ratio is already observed for the t = 0 spectra and thus the dimers are fully equilibrated within the dead time of the experiment (between mixing and recording the mass spectrum). In a conservative estimate of two minutes for the dead time, this would correspond to a rate coefficient for the depletion of homodimers of 0.35 min^{-1} or 0.006 s^{-1}. This represents the minimum value, as the system could have reached equilibrium much earlier than two minutes after mixing.

Next, a time course was collected to follow the SE of the oxidised protein at 37°C (Figure 6 – 7a and b). In contrast to the reduced protein, the SE proceeded slowly and was not complete after ten hours, consistent with the oxidised cysteine mutant of cABC which did not exchange

after prolonged incubation[13]. Before mixing, the labelled and unlabelled protein was first oxidised by exposure to air, most notably in a dialysis cassette for two days. Unlike the time course with the reduced sample, the protein was allowed to exchange in PBS pH 7 rather than ammonium acetate pH 7. Previously, when exchange of oxidised c27-2 in ammonium acetate was followed with MS, deterioration of the spectra occurred before significant exchange had occurred. This deterioration is presumably due to the incubation of the protein at 37°C for a prolonged period of time, which is not unexpected in a non-native purified state especially when the protein lacks its natural N- and C-termini. However, when the experiment was undertaken in PBS, exchange occurred slightly faster. Thus, though samples above ten hours of incubation do not provide well-resolved spectra, the first ten hours allow observation of the majority of the equilibration process.

At t = 0, the air oxidised c27-2 populates homodimers almost exclusively (Figure 6 – 7a orange peaks with pink double circles as in Figure 6 – 6). Over time the heterodimer (green peak, blue circles) increases in abundance and the homodimers become less abundant as their subunits exchange to form homodimers and heterodimers. After ten hours (t = 600 minutes), the oxidised sample has not yet fully equilibrated, but is approaching the 1:2:1 ratio of the equilibrated sample in the presence of DTT.

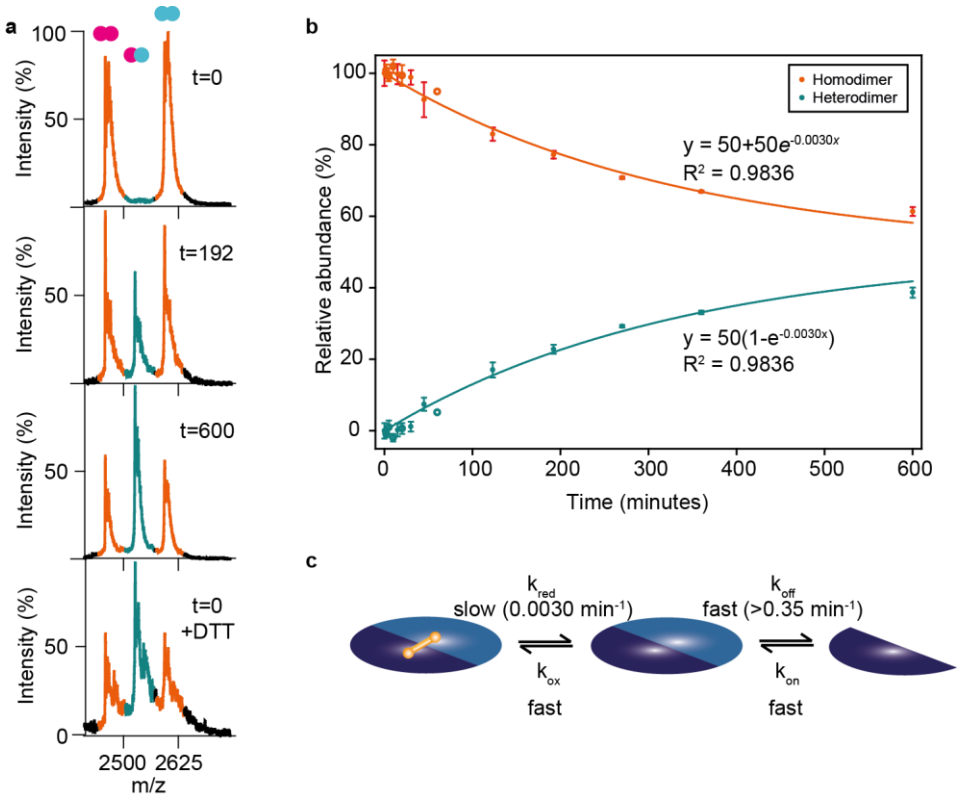


Figure 6 – 7 The breaking of the disulphide bond is the rate determining step of SE of the core domain. (a) SE between unlabelled and labelled homodimers (orange) leads to a gradual increase in the heterodimer population (green), with concomitant decay in the homodimer population. At t=600, the system is approaching equilibrium and resembles the distribution in the reduced sample at t=0 ('+DTT', 100 mM DTT). 18 µM each of unlabelled and labelled sample were mixed. t – the time equilibrated at 37°C in minutes. Only the +9 charge state is shown for clarity, though each dimer charge state was identified with IM-MS and used to calculate the dimer populations in (b). (b) The change in abundance of the homodimers and heterodimer in the oxidised sample over time can be modelled by a simple exponential decay, from which the first order rate coefficient can be extracted. Open circles are additional points not used to calculate the regression line. (c) Comparison of the SE of reduced and oxidised c27-2(WT) indicates that breaking of the disulphide bond is rate determining, and the speed of each step can be deduced.

It should be noted that in the t = 0 spectrum, the baseline is raised slightly at the position of the heterodimer peak. This may arise from a low population of subunits with reduced cysteines that are able to exchange immediately on mixing, though it is not possible to assign an area that meaningfully describes the abundance of the exchanged species. However, as the subsequent SE of the remainder of the sample proceeded so slowly in comparison to the +DTT sample, it was clear that any exchanged population at t = 0 must already have been reduced. This was an important observation as it demonstrated that in the *in vitro*, air-

oxidised WT sample there were very few dimers with a reduced disulphide. This is supported by an NMR experiment by Rajagopal *et al.* on an HSP27 core domain dimer comprising residues 80 – 176. The authors found that addition of hydrogen peroxide to an air-oxidised solution did not change the HSQC spectrum of the core domain, indicating that the disulphide forms spontaneously and comprehensively under air oxidation[2]. This allowed us to confidently assign the WT c27-2 dimer in the APII registration (Section 2.2 and 2.7). Additionally, as this small peak was not reflective of the SE kinetics of the oxidised species, its area was removed from the area of the heterodimer for all time points.

The UniDec software for deconvolution of mass spectra[42] was employed to extract the peak areas from the 2D spectra collected over time, across all charge states for the dimer. The free monomer abundance remained very low over the whole time course in the absence of DTT. This implies that formation of the disulphide is fast under air oxidation (as any reduced subunits would populate both monomers and non-covalent dimers (Figure 6 – 6)). So the breaking of the disulphide is the rate limiting step, followed by fast SE (i.e. fast dissociation and association of the dimer) and subsequent fast formation of the disulphide in the new dimer (Figure 6 – 7c).

The areas of the two homodimer peaks in each charge state were combined to give the total homodimer abundance at each time point, and the peak areas were then normalised so that the population of homodimers at t = 0 was 100% (Figure 6 – 7b). Regression was employed with an exponential decay model for the homodimers (which initially decrease rapidly in concentration) and an exponential rise to maximum for the heterodimers (which should rise

from zero rapidly as the homodimers begin to exchange). The system was modelled assuming that the ratio of labelled:unlabelled protein is mixed in a 1:1 ratio, where in fact the mean extracted peak area at t=0 for the labelled species was 56.4 ± 1.7 % and for the unlabelled species was 43.6 ± 2.2 %. However, the data were well-described by the simpler model, such that when the system has reached equilibrium it is distributed in a 1:2:1 ratio of unlabelled homodimer: heterodimer: labelled homodimer, and the curves approach an asymptote at 50%:

$$\textbf{Total homodimer: } \boldsymbol{y = 50(1 + e^{-bx})} \quad (1)$$

$$\textbf{Total heterodimer: } \boldsymbol{y = 50(1 - e^{-bx})} \quad (2)$$

where y is the relative dimer abundance and x is the time in minutes[‡].

The rate limiting step in the SE of the dimers is the breaking of the disulphide bond because SE occurs rapidly when the disulphide is not present. This bond breaking was anticipated to exhibit first order kinetics, *i.e.* the dimer is the only species in the rate equation. A plot of ln(total homodimer abundance) against time is linear, as expected for a first-order reaction. Thus in the equations above, b is equal to the first order rate coefficient k, where k was found to be 0.0030 min^{-1}. This describes the decay rate of oxidised c27-2 and in turn the rate of the reduction of the disulphide.

The value of k is an order of magnitude slower than the dissociation rate coefficient of a monomer of full-length αB-crystallin at pH 7 and 37°C[27], which requires simultaneous

[‡] In Figure 6 – 7b, the open circles are additional points that were not used to calculate the regression line in order that the regression was performed on more evenly distributed data. However, they are clearly in agreement with the trend (when the regression was performed on the whole dataset, R^2 was 0.9786 and k was 0.0029 min^{-1} for both curves).

dissociation at three interaction sites at least (one intra-dimer and two inter-dimer interfaces). The slower dissociation of c27-2 subunits could be due to two factors: the disulphide confers a great degree of stability to the HSP27 intra-dimer interface which is not available to cysteine-lacking αB-crystallin; or the intra-dimer interface is normally weakened within the full-length oligomer due to the distal binding of the C-termini, but is artificially strengthened in the absence of this interaction. The inverse relationship between the strength of the intra-dimer and inter-dimer interfaces has been characterised for αB-crystallin[36]. However, it also clear that the disulphide does greatly strengthen the interface of the HSP27 core domain relative to the αB-crystallin core domain (which exchanges within the dead time of the experiment[13]).

Having characterised the effect of the disulphide on the exchange at the dimer interface, we next turned to FRET to isolate the effect of the disulphide on dimer exchange within the FL oligomers. This was employed to monitor SE of FL HSP27 with and without DTT. In this experiment, fluorescently-labelled acceptor- and donor- HSP27 are mixed and allowed to come to equilibrium. The acceptor and donor labels come into close proximity as HSP27 exchanges subunits. Excitation of the donor at 485 nm causes it to emit light which is absorbed by the acceptor molecule (i.e. FRET occurs between them) which in turn causes the acceptor to emit at 670 nm. Equilibration thus establishes a maximum in the fluorescence intensity of light emitted by the acceptor. Unlabelled HSP27 FL is then mixed with the fluorescent mixture. As the unlabelled oligomers equilibrate with the labelled ones via exchange of subunits, the donor and acceptor labels are less likely to be in close proximity to one another. Hence, FRET efficiency diminishes and the emission detected at 670 nm reduces over time (Figure 6 – 8).

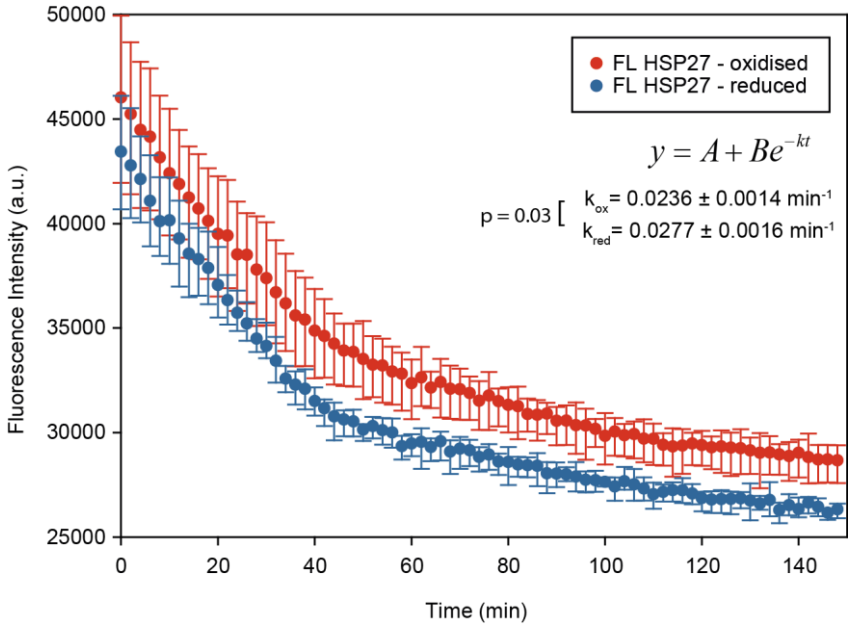


Figure 6 – 8 SE of FL HSP27 in the absence or presence of 20 mM DTT. The intensity of emitted light from FRET decays over time as the labelled oligomers exchange with unlabelled oligomers. The difference in SE between the two redox conditions is much less marked than for the core domain. Symbols show the mean and error bars show the range of n=3 for each condition. Each independent curve was fitted according to the equation shown, and the mean transfer coefficients k_{ox} and k_{red} were significantly different. Values are mean ± s.d. of n=3 repeats for each condition. FRET SE exchange experiments were conducted by Heath Ecroyd, University of Wollongong.

The data of each independent repeat were fitted to a curve according to the equation,

$$Fluorescence\ Intensity = A + Be^{-kt} \quad (3)$$

where k is the transfer rate coefficient. A and B are constants quantifying the intensity when the curve plateaus and the overall drop in intensity across the curve, respectively. This is comparable to analysis employed for sHSPs elsewhere[34].

This yielded a value for the transfer rate coefficient of 0.0236 ± 0.0014 min^{-1} for the oxidised sample and 0.0277 ± 0.0016 min^{-1} for the reduced sample. These values are significantly different with p = 0.0288 (two-tailed unpaired t-test), and the reduced protein exchanges faster than the oxidised protein as anticipated from the c27-2 experiment. The reduced sample can exchange subunits by both monomer and dimer exchange, whereas the oxidised sample is

restricted to dimer exchange alone (Figure 6 – 9). The transfer rate coefficient for the oxidised sample thus reflects the rate of dimer exchange, and the c27-2 SE monitored by MS above gives us information on interruption of the dimer interface. The reduced sample can undergo dissociation at both the inter-dimer and intra-dimer interfaces, permitting both the possible dissociation of dimers seen in the FL oxidised sample, and the dissociation of monomers involving fast dissociation at the intra-dimer interface observed in the reduced core domain. The FRET signal thus decreases more rapidly, but there is likely also an energy compensation between the intra-dimer and inter-dimer interactions, as observed in αB-crystallin[36]. In this way, the oxidised sample would have a stronger intra-dimer interface due to the disulphide which destabilizes the inter-dimer interface and encourages dissociation of a dimer, and the reduced sample has a weaker intra-dimer interface which is compensated with a stronger interaction between the C-termini and the binding groove[36], such that both monomers and dimers can exchange.

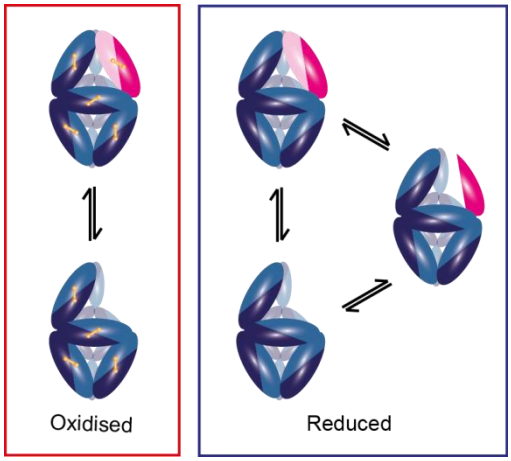


Figure 6 – 9 SE between FL oligomers comprises dimer exchange exclusively for the oxidised sample (red box), while both dimers and monomers could be the exchanging species in the reduced sample (blue box). Although the transfer rate coefficients are similar in the two conditions, the strength of the interfaces within the oligomers may vary substantially. While shown as one multimer here for clarity, in reality, HSP27 occupies an ensemble of oligomers.

The k values are approximately equal to that measured with FRET for FL αA-crystallin SE and half the value observed for HSP27 SE with αA-crystallin when HSP27 was necessarily reduced by the addition of the donor and acceptor tags to the cysteine[34]. The larger transfer rate coefficient between two different sHSPs may reflect faster exchange kinetics for cross-exchange between two sHSPs rather than exchange within oligomers of the same sHSP species. Alternatively, it could partly belie a population of oxidised subunits within our reduced sample that the chemically modified HSP27 could not access.

If we assume that dissociation of a subunit is again the rate limiting step in SE, the transfer rate coefficients measured with FRET demonstrate a similar value to the decay coefficients determined with the high mass resolution of native mass spectrometry for a subunit of FL αB-crystallin[27] (approximately 0.036 min^{-1}) or FL αB-crystallin with reduced HSP27[39] (0.071 min^{-1}). These comparisons highlight a remarkable feature of these dynamic proteins – despite a significant 'molecular staple' in the form of the HSP27 disulphide, the protein is still able to exchange subunits on the same order as its sHSP counterparts in the cell. This must certainly facilitate their inter-exchange with one another and suggests that HSP27, with its redox-sensitive cross-subunit disulphide, would maintain the same degree of higher oligomer plasticity when the cell is under oxidative stress. This may maintain its ability to respond quickly under stress, while the degree of oxidation of the dimers themselves may moderate the particular stoichiometries adopted and the nature of its chaperone response.

2.4 Solution NMR spectroscopy reveals multiple populations of the interface in solution

Residue-level information on the behaviour of a protein in different solution conditions can be achieved with HSQC NMR experiments (Chapter 4). Therefore HSQC spectra were collected on 1 mM of ^{15}N-labelled c27-2 under oxidising and reducing conditions, and each of the individual peaks was assigned to the corresponding residue of the protein. On reduction there was a wide-scale decrease in the relative peak intensity (Figure 6 – 10), which was particularly marked along the β5, β5-6 loop and β6+7 strand. The reduction in intensity is indicative of peak broadening and chemical exchange on the microsecond to millisecond timescale. This chemical exchange could be between the free monomer and reduced dimer forms observed in MS (Figure 6 – 6) but could also include alternate registration states.

A map of the CSPs between peaks in the oxidised and reduced ^{1}H-^{15}N HSQC spectra (Figure 6 – 10 lower panel) shows the largest chemical shifts (or disappearance of the peak due to line broadening) for residues 134 - 138, which are the residues surrounding the C137 of the disulphide. Notably, G122 and K123 have a large CSP as they lie on the β5 strand behind C137. The G122 residue forms main chain hydrogen bonds to F138 on the interface, and though K123 does not bond any of the interface residues itself, its immediate neighbours do. The E126 peak also moves by 0.25ppm. Its side chain forms a salt bridge across the top of the extended β-sheet to K141I of the other monomer in the c27-2(WT) dimer structure, but the same E126 faces downwards and may be able to form an intra-chain hydrogen bond with H124 in the APIII structure. The IM-MS (Figure 6 – 4) and MS data (Figure 6 – 6) demonstrated that the reduced dimer does not access the APIII state more than the oxidised

form but does show significant dissociation on reduction, so we interpret this CSP (and the widespread peak attenuation across the protein) as reporting on monomer-dimer exchange rather than a change in registration.

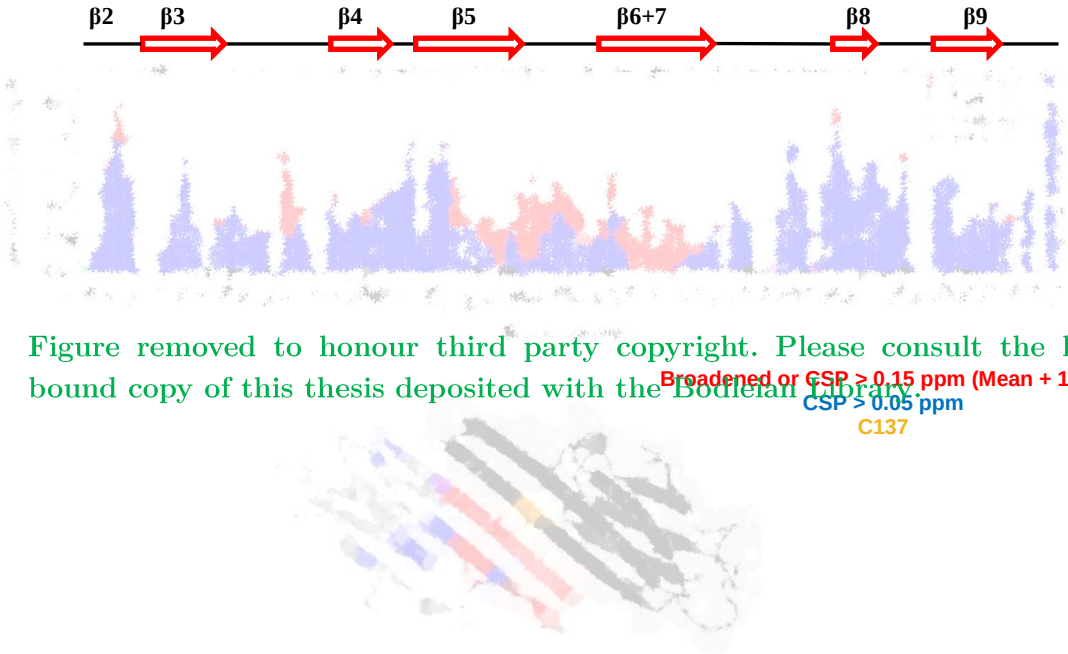


Figure 6 – 10 Reduction of c27-2 causes wide-scale moderation of ^{15}N HSQC peaks. (a) The relative peak heights of each residue of c27-2 vary on reduction, with peaks in the central region most notably affected (peaks 134 – 137 disappeared on reduction so are not marked here). (b) Chemical shift perturbations on reduction, mapped onto one chain of the 4MJH structure (the other chain is shown in black for clarity). Reduction causes large CSPs in the β6+7 and β5 strand (red), with some change in the chemical environment of residues even further from the disulphide (blue). T. Reid Alderson performed the HSQC experiment and assignment, and generously produced the figure.

There are smaller CSPs in other parts of the protein (blue), for example the loop between the β3 and β4 strands. This includes the H103 residue, which interacts with residues near the interface of WT c27-2 such as R127 and W136 (Table 5 – 3). F104 also displays a CSP, possibly due to its close proximity to H103 or because of a putative pi-pi interaction with F138. Notably, there are only subtle CSPs in the β4-β8 groove, implying that any allosteric effect of the disulphide on this inter-dimer interface (formed between the β4-8 groove and a

neighbouring C-terminus) could either be mediated by the C-terminus or involve only a small change in chemical environment. This could be investigated further by measuring the K_d for binding of the CTP to the core domain in oxidising or reducing conditions or with the cysteine mutated, to ascertain whether the disulphide truly has an impact on this interface.

The solution spectra clearly indicate the disulphide has a marked effect on the environment of many of the residues of the interface and surrounding sites. This demonstrates that reduction interrupts the core domain far more than simply removal of a disulphide; the protein is able to access other chemical states. This implies that subunits in the FL would also experience a significantly different environment when reduced or oxidised, which adds an extra opportunity for regulation to this particular sHSP.

2.5 Disulphide formation affects chaperone behaviour

Having demonstrated that the intra-subunit disulphide has a marked effect on the solution-phase dynamics of c27-2, we then questioned whether there is any difference in activity between reduced and oxidised HSP27.

This was investigated by monitoring the aggregation of creatine phosphokinase (CPK) at 43°C with and without FL HSP27 in the absence or presence of reducing agent. Aggregation was detected by the change in absorbance at 340 nm, which occurs due to light scattering by protein aggregates[43,44]. The aggregation was followed for five hours (Figure 6 – 11).

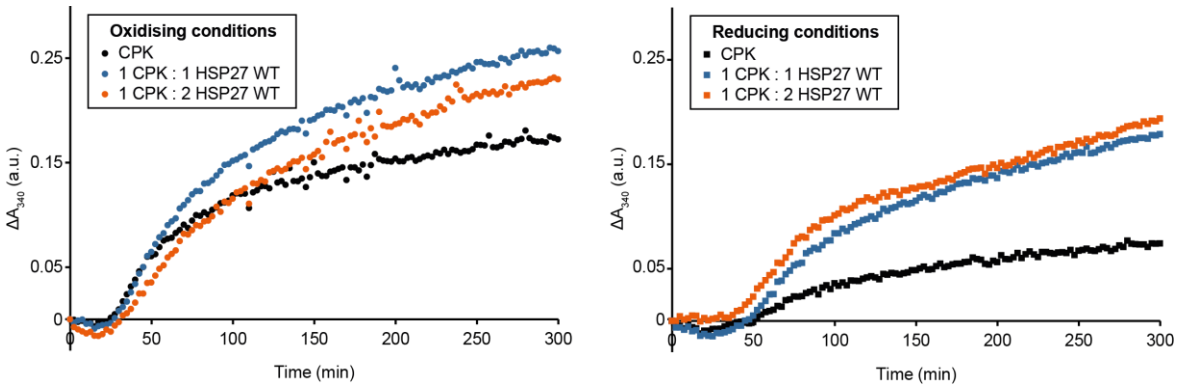


Figure 6 – 11 The redox state of HSP27 changes its chaperoning of CPK. Aggregation assays monitoring the change in absorbance over time of 50 µM CPK in the presence of stoichiometric amounts of FL HSP27 (50 µM or 100 µM) in the absence or presence of 20 mM DTT. In both redox conditions, addition of HSP27 leads to an increase in light scattering, though there is a greater proportional increase in the absorbance under reducing conditions. Chaperone assays were performed by Anthea Rote, University of Wollongong. Each point shows the mean ΔA_{340} of duplicate wells, and all traces are representative of five independent repeats.

In the absence of chaperone, the light scattering of CPK begins to increase as the protein spontaneously aggregates at 43°C over time, after a lag phase of approximately 25 minutes. When CPK aggregation is heat-induced in the presence of 20mM DTT, there is a smaller overall increase in light scattering, possibly due to the reduction of disulphide bonds within CPK which leads to different morphologies of the aggregates it forms under heat denaturation.

On addition of FL HSP27, light scattering is increased in both oxidising and reducing conditions. This is in contrast to the chaperone response typically observed in these assays, where addition of chaperone leads to a smaller increase in absorbance inferring a decrease in the formation of large protein aggregates[43,44]. The increased signal is not simply an additive effect of independent aggregation of CPK and HSP27 because oxidised HSP27 is thermostable up to 68°C[1,39]. However, a similar increase in light scattering has been observed with molecular chaperones before[45], thus this observation is likely also indicative of chaperone

action of the sHSP, which can co-aggregate with substrate in amorphous and fibrillar aggregates[17,46]. For example, sHSP co-aggregation has been observed in HSPB6, where the dimeric WT was compared to numerous constructs carrying deletions in the N-terminus[47]. Incubation of reduced hen egg white lysozyme in the presence of the chaperones consistently produced an increase in light scattering when compared to the reduced substrate alone. Moreover, the constructs that had been found to be the most effective chaperones for insulin and alcohol dehydrogenase produced the highest increase in scattering with the lysozyme[47], implying that is possible to detect a co-aggregation protective mechanism with these assays.

Under air oxidation, where the disulphide bonds of both the CPK substrate and the FL HSP27 oligomers remain intact, there is a smaller proportional increase in light scattering compared to the same assay under reducing conditions (Figure 6 – 11, Table 6 – 1). The negative values for the relative protection reflect the increase in light scattering in the presence of chaperone (Eqn. (7)). While the absolute increase in absorbance at the end point is similar, the significant difference in protection relative to the CPK control suggests that the reduced HSP27 may co-aggregate in a larger stoichiometric ratio than its oxidised counterpart (Table 6 – 1). Addition of 50 or 100 µM HSP27 were not statistically different in the increased light scattering they produced, suggesting specific ratios in the HSP27:CPK complexes. Together these suggest that the effect of the chaperone is different under oxidising and reducing conditions.

Therefore, the redox state of the HSP27 cysteines has a marked effect on the dynamics of the intra-dimer interface with a possible concomitant influence on distal interaction sites.

Chaperone behaviour appears to be moderated by the presence of the disulphide. Moreover, both redox states are present *in vivo* and vary with cell type. Taken together with the rare occurrence of cross-strand disulphides and the redox response of FL HSP27, our data thus support an essential role of this bond in the function of the WT protein.

	% Protection		p-value
	Oxidised	**Reduced**	
50 µM HSP27	-28 ± 16	-134 ± 45	0.0011***
100 µM HSP27	-31 ± 9	-145 ± 38	0.0002***

Table 6 – 1 Protection afforded to 50 µM CPK by 50 µM and 100 µM FL HSP27 in oxidising or reducing conditions. Reduced HSP27 gives a proportional increase in light scatter that is five-fold that of oxidised HSP27. There is not a significant difference between the protection induced by 50 µM or 100 µM HSP27. Percentage protection values are calculated according to Eqn.(7) from chaperone assays as in Figure 6 – 11 and are shown as mean ± s.d., where n = 5 for each condition. P-values are from an unpaired two-tailed t-test, ***p<0.05,. Each time point in each n = 5 repeat was sampled in duplicate.

2.6 Native MS of the core domain shows S135F oxidises less readily

The data described herein suggest that the intra-subunit disulphide has a regulatory role in the function of the protein. For the T151I and T164A c27-2 constructs discussed in Chapter 5, mutation had only a subtle effect on the conformations they assumed in the crystal lattice, and even the R136W had a similar monomeric form to the WT. We thus hypothesised that the deleterious effect of mutation may lie in the formation of this disulphide bond and the consequential alteration of the chaperoning ability of the protein. This may affect the chaperone response of the protein under other forms of cellular stress as well as redox imbalance.

To investigate whether mutation has an effect on the assembly of the HSP27 core domain, we collected mass spectra on all the c27-2 mutants described in Chapter 5 (Figure 6 – 12). These proteins had been fully reduced before buffer exchange and equilibration in air at 37 °C for two hours to control for any oxidation variation induced in the purification.

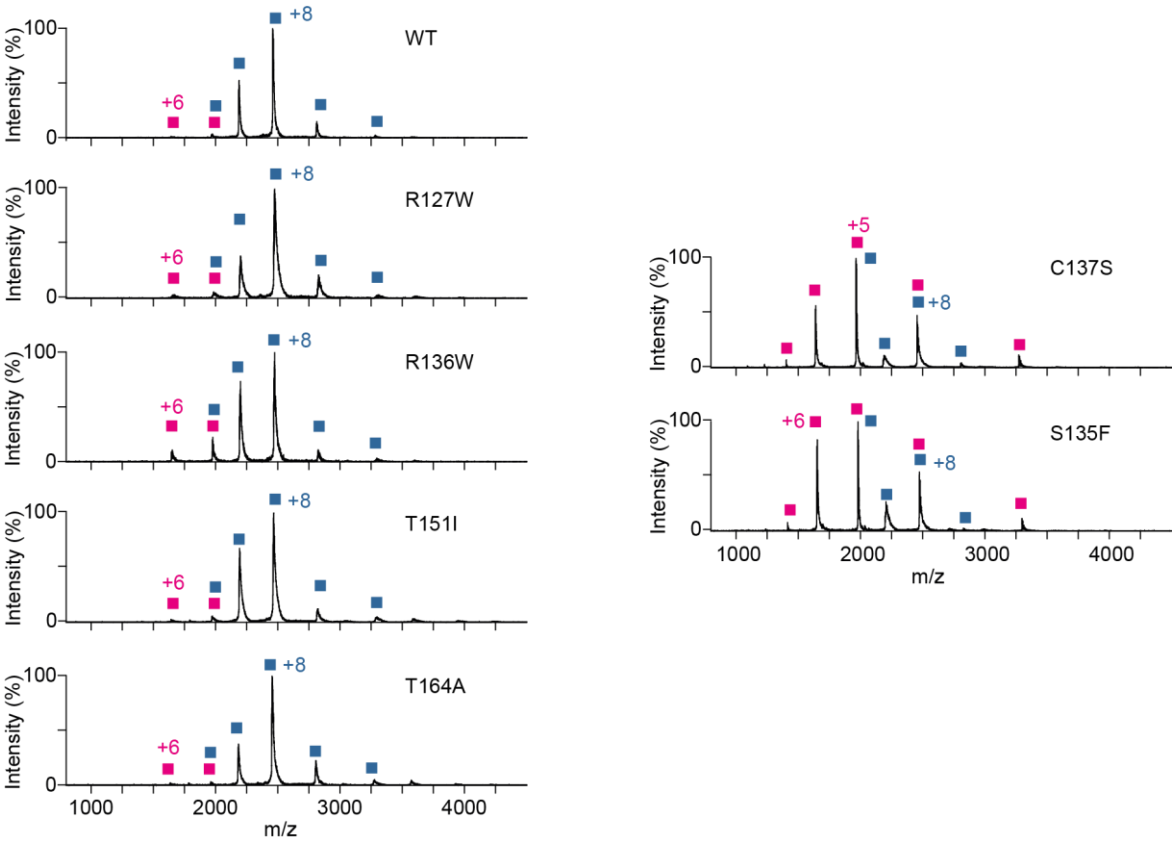


Figure 6 – 12 The monomer-dimer equilibrium of the core domain constructs. MS of 20 µM of each construct collected after reduction, buffer exchange and equilibration in air at 37 °C for two hours. Four of the mutants show a similar abundance of dimers (blue squares) as the WT. c27-2(S135F) displays a marked increase in the abundance of monomers (pink squares), with a similar equilibrium distribution to the core domain with a C137S mutation which cannot form disulphide-locked dimers. The charge of the most abundant charge state is indicated.

The WT c27-2 forms mainly dimers, with a negligible amount of monomer (Chapter 4). When C137 is mutated to a serine, which is unable to form the disulphide bond, the mass spectrum shows a different distribution of charge states. Those peaks corresponding to a monomer (pink squares) are much more intense than in the WT spectrum so a larger number of charge states

are detectable (z = +3 to +7 vs z=+6, +5 in the WT). There is a concomitant decrease in the peaks corresponding to the dimers (blue squares), with only a small contribution from the dimer to the overlapped peak at m/z = 1964 Th. This is a non-covalent dimer as the intra-dimer disulphide cannot form. The spectrum is very similar to that seen in the presence of reducing agent (without the peak broadening due to adduction of reducing agent (Figure 6 – 6)), and demonstrates the propensity of the core domain to form non-covalent dimers. The monomers and dimers are likely rapidly exchanging with one another at equilibrium, as was demonstrated for the reduced WT sample (Section 2.3). This monomer-dimer equilibrium of the core domain is very similar to cABC which does not naturally contain the disulphide[13].

Most of the neuropathy-associated mutants show a very similar spectral profile to the WT, with the dimer species being the major component. There is a slight variation in the relative abundance of the different charge states within the dimer charge state envelope; this degree of variability between spectra is not unusual and results from variation in the nESI charging process due to slightly different capillary diameter and distance. Thus these mutants can dimerise as readily as the WT under these conditions. Both covalent and non-covalent dimers could contribute to the dimer peaks, but in the WT the monomer population increased substantially when the disulphide was not present (Figure 6 – 6). Thus there could only be a different ratio of non-covalent and disulphide-locked dimer in each sample if the mutants had a greater ability to form the non-covalent dimer than the WT. This is unlikely for R136W c27-2 which has a greater steric bulk at the interface.

Strikingly, the spectrum of S135F c27-2 is distinctly different from the other mutants with a much higher abundance of monomers, akin to the C137S c27-2 spectrum. This indicates that it occupies the monomer much more frequently than the WT. The air-oxidised WT dimer is almost completely disulphide-locked, as demonstrated with the SE experiment (Section 2.3). Assuming the system is at equilibrium, the S135F c27-2 disulphide is thus weaker than in the WT.

Hence at least one of the mutants has a weaker disulphide bond than the WT. It was therefore important to investigate further whether the other mutants show an altered redox potential to the WT. Additionally, it was necessary to test whether the mutants abnormally occupy the API and APIII registrations, as suggested by the c27-2(R136W) crystal structure.

2.7 Using IM-MS to probe the conformation of the core domain: effects of mutation

R136W c27-2 did not dimerise significantly differently to the WT (Figure 6 – 12) suggesting that it could have a similar reduction potential. However, it was observed in the APIII state in the crystal (Chapter 5). It was thus important to verify whether c27-2(R136W) accesses the APIII state to a different degree than the WT, or if this was artificially enhanced in crystallisation. Additionally, as c27-2(S135F) had a greater reduced population (Figure 6 – 12), it was possible that this also could access the APIII to a higher degree than the WT.

In order to determine this, we measured the CCS of the mutated core domain using IM-MS, as described in Section 2.2. We hypothesised that if the APIII state was favoured by the R136W or S135F mutation, we would see a shift in the drift time distribution reflecting this

larger CCS. Additionally, we measured the drift times of all the mutants and C137S control to identify whether any of the other mutations interrupted the antiparallel populations of the dimer.

Theoretical CCS values were calculated with the PA for the crystal structures of R136W, T151I and T164A c27-2, and the API-III models of the WT in Figure 6 – 4b (Figure 6 – 13a). The values were calculated from PDB files that were all truncated to the same degree with any objects outside of the main chain and peptide removed. This ensured that any differences in the estimated CCS were due to the different registration states, rather than the number of resolved residues in the PDB files. If the crystal structures in Chapter 5 are representative of the population of the mutants, we would thus expect the CCS of R136W to be larger than the oxidised WT, and T151I and T164A to be similar to the WT.

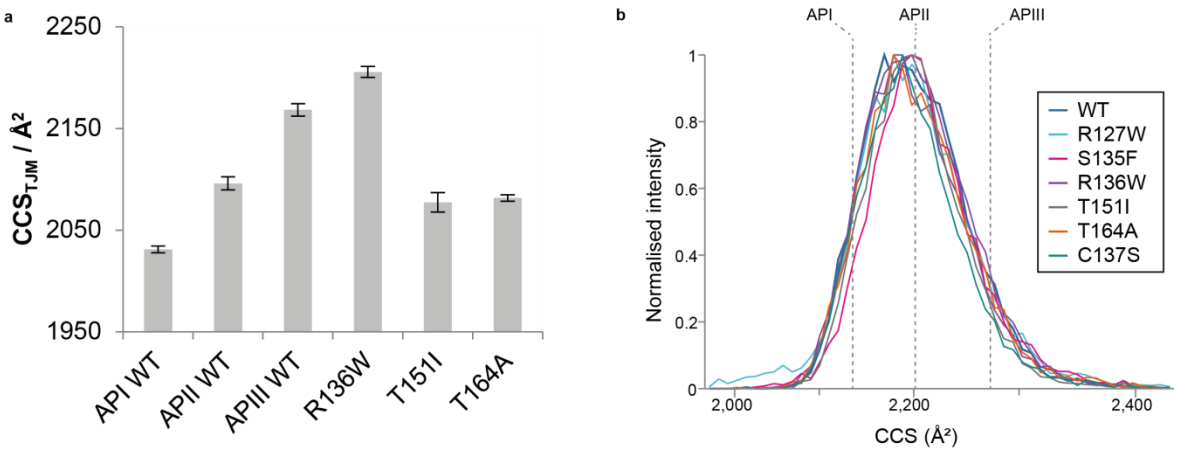


Figure 6 – 13 The mutants do not populate the AP states differently to WT (a) Estimated CCS values for the API, II and III WT, R136W (APIII), T151I and T164A (APII) c27-2. All values were calculated with the PA through IMPACT[18]. CCS$_{TJM}$ indicates the value expected from the Trajectory Method, based on the value determined from the PA. Error bars are ±1 s.d. on three runs. (b) Drift time profiles for the +9 charge state for each oxidized mutant at a wave height of 5.5 V and pH 6.9. Vertical lines indicate the expected drift time for the WT in the API-III states. All mutants mainly occupy the APII registration, without any obvious change in the population of the API and APIII states.

All of the drift times were measured on the air-oxidised form at pH 6.9 (Figure 6 – 13b); those shown are representative of the data collected at wave heights of 5.5, 6 and 7 V, for the +8, +9, +10 charge states for 20 μM of monomeric protein. In this figure, the CCS scale on the x-axis is only accurate for the WT, as the slight deviation in mass for each of the mutants means that their CCSs scale slightly differently with drift time, as the relationship between CCS and drift time is mass dependent. However, the maximum mass difference is only 0.6%, so this is a useful approximation.

As discussed above, the drift time profile for the +9 charge state of the oxidised WT is a Gaussian whose centroid necessarily describes the CCS of the APII state, as the WT c27-2 sample is almost fully oxidised (Figure 6 – 13b). There is not a noticeable difference between the drift times of any of the mutants, including S135F and R136W, which all mainly populate the APII state (Figure 6 – 13b). Each mutant may have a different proportion of covalent and non-covalent dimer (Section 2.6), and the APIII and API states will only be accessible for the non-covalent dimer[23]. The C137S control, which can only form the dimer non-covalently, did not result in a different population of the AP states either. Thus the cross-subunit disulphide does not seem to affect the equilibrium distribution among the registration states. This is in keeping with cABC, which though lacking any cysteines was found to populate the APII state in the same relative proportion as a fully oxidised disulphide mutant[13].

The Gaussian drift time profile results from several experimental factors (e.g. the initial ion packet introduced to the IM drift tube is of non-negligible length, an individual ion's path through the drift tube is stochastic, ions are imperfectly transmitted after the IM cell), which

may fully describe the spread of the ion drift times. However, it is possible that all seven constructs access the API and APIII states a small proportion of the time such that they cannot be deciphered due to the resolution of the drift time profiles. Nonetheless, we can surmise that there is not a substantial difference in the distribution of the dimer register between the constructs.

Different behaviour was not observed between R136W and the WT, so the APIII state seen in the R136W crystal does not seem to be an effect of the mutation, nor an artefact of the pH of crystallisation or of a different degree of oxidation (Section 2.2). Therefore, it seems likely that the APIII state captured in the R136W crystal is only very lowly populated in solution, possibly transiently. It may have been artificially enhanced in the crystallisation process, possibly induced by the steric hindrance of the BME bound to some of the cysteines. Alternatively a small population of native APIII states may have been enriched simply because they were able to form an ordered crystal.

This experiment was performed under air oxidation to replicate the conditions of the crystallisation. It would also have been useful to perform the experiment with reducing agent, to probe whether there was any different population of the AP states for the non-covalent dimer with mutation (or indeed to probe whether c27-2(R136W) accesses the APIII at low pH even though the WT does not). Reduction did not noticeably enhance the API and APIII population of WT c27-2 by IM (Figure 6 – 4), and ^{1}H-^{15}N HSQC spectra demonstrated a dynamic system with interexchange between at least two states. However, in light of the high number of hydrogen bonds maintained at the interface in APIII (Table 5 – 2), the frequent

observation of these alternate registrations for cABC, and the direct consequence on the formation of this seemingly significant disulphide, the data presented here still allow for a small, likely transient population of the alternate registrations.

Moreover, the more important question must be raised as to the effect of these mutations. Each of the five mutations have been identified in motor neuropathy and are autosomal dominant, so mutation either causes a drastic detrimental effect or haploinsufficiency (*i.e.* a loss of function in the mutated protein for which the WT population cannot compensate). The crystal structures and IM data presented here reveal subtle structural changes, so what is the cause of their altered behaviour from the WT which causes such an altered phenotype?

2.8 Motor neuropathy mutations affect the redox potential of the HSP27 disulphide

Following the monomer-dimer equilibria captured with native MS, we hypothesised that the mutants respond differently to oxidation and access the reduced form to different degrees. This would have a concomitant effect on chaperone activity, as observed for the full length protein (Section 2.4).

We developed a simple assay to characterise the redox behaviour of the mutants in solution. This gave reproducible results and was sensitive to the differential redox behaviour of the proteins. It comprised denaturing SDS-PAGE of the mutant core domains after incubation in different reducing environments at 37°C. The intensities of the disulphide-bound dimer band and monomer band were used to calculate the relative population of protein in the monomeric (reduced) state. The reducing environments were produced using a different ratio of reduced

L-glutathione (GSH_{red}) to oxidised L-glutathione ($GSSG_{ox}$), with a final glutathione concentration of 10 mM for 30 µM protein monomer. Densitometry of the dimeric and monomeric bands was quantitative as Coomassie Blue staining produces a band intensity proportional to the amount of protein for the concentrations used here[49] (here *ca.* 1.8 µg was applied per lane).

The results with the glutathione-mediated reduction potential are presented in both panels of Figure 6 – 14. The ratio of GSSG:GSH increases along the x-axis, and so at each point the protein is subjected to a more oxidising environment during the equilibration period. The relative monomer abundance (y-axis) reports on the intensity of the monomeric SDS-PAGE band (which runs as a peptide chain of *ca.* 10 kDa) as a proportion of the total intensity in the monomeric and dimeric bands (the latter reproducibly appears *ca.* 25 kDa). As the SDS-PAGE is run under denaturing conditions, any dimer detected must be covalently linked.

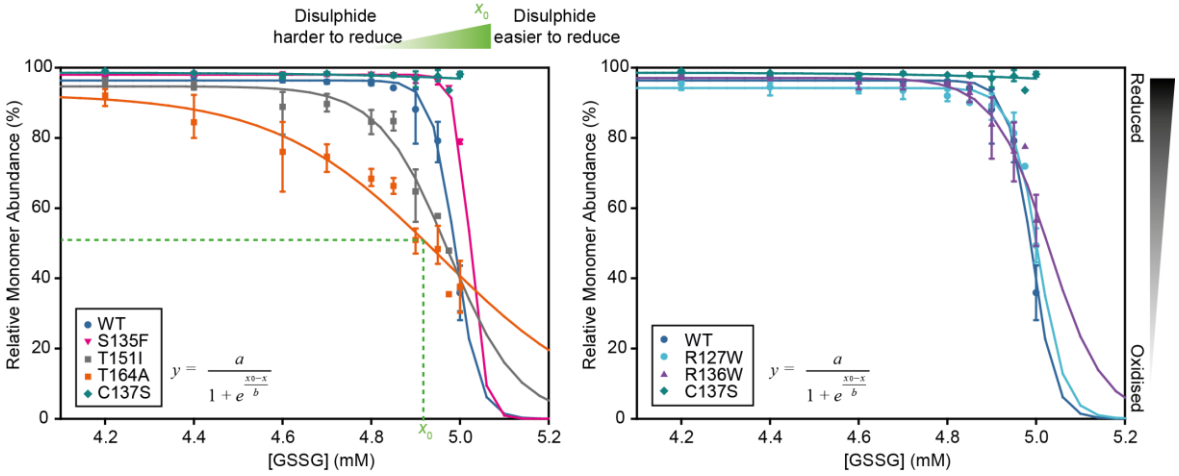


Figure 6 – 14 Quantifying the reduction potential of the core domain and its mutants. The relative abundance of monomers in samples incubated at different reducing potentials provided by a $GSSG_{ox}$:GSH_{red} ratio (denoted by the $GSSG_{ox}$ concentration here). Monomer and disulphide-bound dimers were quantified by densitometry on their denaturing SDS-PAGE gels. The data were fitted according to the equation shown. Each point reflects the mean relative monomer abundance detected at a particular $GSSG_{ox}$:GSH_{red} value and the error bars indicate the range.

For the WT core domain, at [GSSG] < 4.8 mM, the protein is solely monomeric (the values are not 100% here due to slight variation in the local background noise of the bands, even after background subtraction). At values above 4.8 mM the protein shows large decreases in the relative monomer abundance reflecting the appearance of the dimer band and disulphide bonding of some of the population.

To interpret the nature of the oxidation, a sigmoidal curve was fitted to the data according to:

$$y = \frac{a}{1 + e^{\frac{x_0 - x}{b}}} \tag{4}$$

Where b characterises the steepness of the curve; x_0 is the turning point or point of inflexion of the sigmoid; and a is the height of the curve as it approaches its maximum and minimum at large |x| values.

The maximum value of the $GSSG_{ox}$:GSH_{red} value is 5:0 as:

$$[GSH_{tot}] = [GSH_{red}] + 2[GSSG_{ox}] = 10\ \text{mM} \tag{5}$$

10 mM of GSH_{tot} was chosen because it is representative of an intracellular concentration[50], though it is the $GSSG_{ox}$:GSH_{red} ratio that determines how much the protein will be reduced[51].

The sigmoid was thus fitted to all the experimental points from 0mM to 5mM and three additional artificial points of 0% monomer abundance at 10mM GSSG concentration to mimic full oxidation at higher (less negative) reducing potentials.

The C137S (green diamonds), which is unable to form the disulphide, retains high monomeric concentration for all the experimental $GSSG_{ox}$ concentrations. The C137S data points were

also fitted with a sigmoidal curve but without the restraint of 0% monomer abundance at 10mM, as the protein could never be induced to form dimers.

The T164A c27-2 (orange squares) shows deviation from 100% monomeric abundance at lower [$GSSG_{ox}$] values than the WT. This shallower curve implies it is redox sensitive over a wider window. It also has a lower value of x_0, the turning point of the curve which characterises the translation of the curve on the x-axis. Thus a protein with a lower x_0 is easier to oxidise (and harder to reduce).

T151I c27-2 shows similar behaviour to T164A but to a lesser degree. On the other hand, S135F only begins to decrease at 4.975 mM GSSH and drops steeply to 5 mM. Thus the fitted curve has higher x_0 and higher b, demonstrating that S135F is harder to oxidise and responds to the redox environment over a smaller window.

Plots for R136W and R127W c27-2 are shown in the right panel of Figure 6 – 14, which show similar behaviour to the WT. These proteins were investigated further with square wave voltammetry (please see below) in order to investigate this comparison in more detail.

Each individual point could be used to calculate the redox potential of each core domain according to the Nernst equation[51]. However, it seems that the 200V potential applied across the SDS-PAGE gel itself acts to reduce some of the protein, as a significant monomeric band is observed even in oxidised samples that are shown to have a very small reduced component by SE (as in Figure 6 – 7a). Thus, it may be possible to correct the relative monomeric abundance based on the amount of reduction seen on the SE samples. Certainly the artificial points at 10 mM GSSG and 0% monomer contribution currently underestimate the apparent

monomeric abundance which would be seen when fully oxidised. However, if absolute quantification of the redox potential is required, a gel shift or mass shift assay[52] is more appropriate, but these would rely on complete capping of free-thiols with an alkylating agent and equal ionisation efficiency of the oxidised and capped-reduced samples in MS analysis. As the aim of this experiment was to identify a shift in redox potential from that of the WT rather than absolute values, the protocol presented above represents a more direct approach without the use of toxic alkylating agents. This takes advantage of the inter-subunit position of the disulphide such that reduction causes a mass shift difference of one protein subunit.

The most notable difference in redox behaviour is between T164A, T151I and WT. These threonine mutations are distal to the intra-dimeric interface and so their influence suggests some allosteric interplay with the intra-dimer interface, which has a larger effect than the side chains of the interface itself (such as the S135F and R136W mutations).

It is possible that the assay is not as sensitive to an improved propensity to be reduced as to be oxidised. This could be due to the upper limit of the $GSSG_{ox}/GSH_{red}$ ratio and the proximity of the turning points of WT, R127W, R136W and S135F to this value, such that the curve is less characterised around the turning point than T151I and T164A. Despite its limitations, this biochemical assay clearly demonstrates that mutation affects the redox behaviour of the core domain.

These observations broadly agree with the results discussed in Section 2.4 from MS of the species (after full reduction, buffer exchange and air oxidation). There S135F was more easily reduced than the other constructs, with a much larger monomer contribution that was similar

to the C137S control. The native MS measurement did not distinguish between the redox behaviour of the other mutants and WT at equilibrium, which are all principally oxidised in air. As such, SDS-PAGE redox characterisation has provided an additional layer of insight.

To better characterise the effect of the R127W and R136W mutations on the redox potential of the core domain, square wave voltammetry was performed on the WT, R136W and R127W. In this experiment, the protein is drop-cast onto a working electrode, which is placed in a bath of PBS at pH 7. An electrical potential difference is applied and measured between a saturated calomel reference electrode and the working electrode. The technique uses a series of potential pulses and heights applied on a linear time scale. There are two measurements made at each pulse, which are on the forward and backward cycle. The current is measured at the working electrode at the end of each of these forward and backward pulses, and the difference between these two measurements of the current is plotted against the applied potential to produce the voltammogram (Figure 6 – 15). Peaks are indicative of redox processes as they occur where there is the biggest difference between the current measured in the forwards and backwards sweeps. This is where the protein itself is responding to the electrical potential applied and gives a signal corresponding to the relative reduction potential of the disulphide.

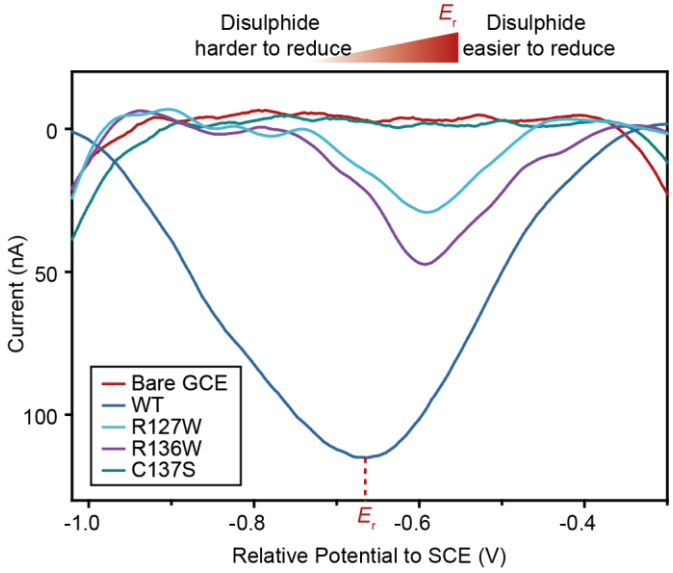


Figure 6 – 15 Square wave voltammetry elucidates the reduction potential E_r of the WT, R127W and R136W c27-2. Both R127W and R136W produce the greatest change in current at a less negative potential than the WT, indicating that the WT disulphide is formed more readily than in these mutants. The applied potential is measured relative to a saturated calomel reference electrode. Traces shown are representative of n=8 repeats. Square wave voltammetry was conducted by Patricia Lee, University of Oxford.

The voltammogram traces show the resulting difference in current on applying this variable waveform potential to the different core domain constructs, and are representative of at least eight repeats undertaken for each construct. Unfortunately, signal was not seen for T164A, T151I and S135F. This is most likely due to their dissolution from the electrode, rather than any specific difference in redox behaviour. The voltammograms of R136W and R127W show peaks at a less negative relative reduction potential E_r (Table 6 – 2) than the WT, indicating that they are both easier to reduce than the WT. As expected, C137S does not give any signal; this negative control gives a very similar profile to the bare glassy carbon electrode.

The results in this section demonstrate that the neuropathy-associated mutations each affect the reduction potential of the disulphide bond. Alongside the difference in *in vitro* chaperone

behaviour observed for the WT FL HSP27 (Section 2.5), this suggests that the deleterious effect of mutation may be mediated through the formation of the disulphide bond.

	Peak potential† (V vs. SCE)	p-value
WT	-0.633 ± 0.014	
R127W	-0.587 ± 0.036	<0.005***
R136W	-0.579 ± 0.011	<0.0001***

Table 6 – 2 Summary of disulphide reduction peak potentials (E_r) of proteins with different mutations. †Mean ± standard deviation, n=8. P-values are for the mean peak potential compared to that of the WT. The means for the two mutants were not significantly different. ***p<0.05.

2.9 Mutations affect the change in chaperone activity of the full length on reduction

In order to characterise whether mutation modulates HSP27 molecular chaperone activity in response to redox state, FL HSP27 mutants were investigated with the CPK aggregation assay (Section 2.4). The CPK aggregation was followed in the presence of one or two molar equivalents of FL HSP27 (50 µM or 100 µM monomeric concentration, Figure 6 – 16a, b and Figure 6 – 16c, d respectively) carrying one of the point mutations, again at 43°C for five hours.

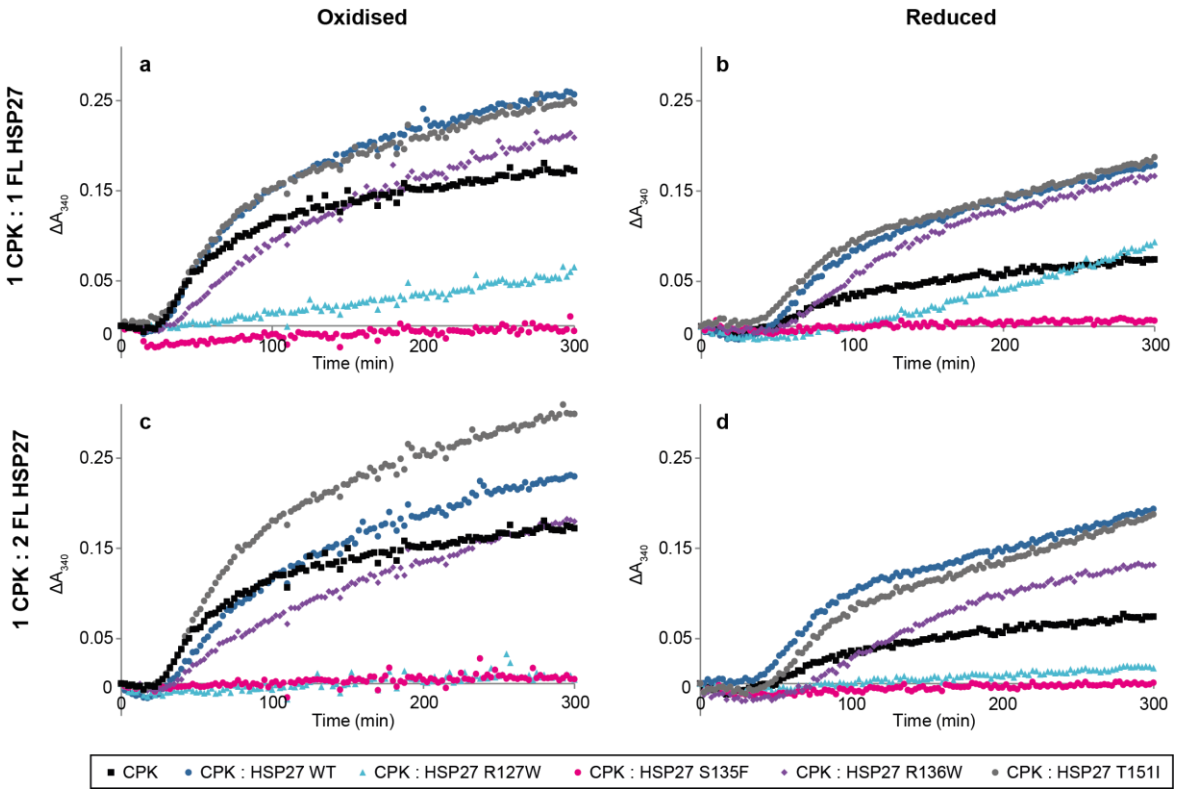


Figure 6 – 16 The effect of reduction of FL HSP27 mutants on their inhibition of CPK aggregation. ***In vitro*** **assays monitor the change in absorbance over time of 50 µM CPK in the presence of 50 µM or 100 µM FL HSP27, either under air oxidation or with 20 mM DTT. The mutants each show a different chaperone effect with both increased and decreased light scattering observed. The form of the curves relative to one another and the CPK control changes on reduction. Aggregation assays were performed by Anthea Rota, University of Wollongong. Each point shows the mean ΔA_{340} of duplicate wells, and all traces are representative of five independent repeats.**

T151I shows similar behaviour to the WT, where the presence of HSP27 produced an increase in light scattering relative to the CPK alone, which may reflect co-aggregation of HSP27 with substrate as a chaperoning mechanism. R136W FL HSP27 also produces increased scattering, but to a lesser extent. R127W and S135F cause diminished light scattering relative to the CPK alone. As light scattering typically reports on the presence of large aggregates[43], this is indicative of smaller aggregates or a reduction in aggregate formation. There are several competing effects in this assay such that the chaperone activity of the different mutants *in vivo* cannot be ranked with certainty (*e.g.* chaperone activity being exhibited in co-aggregation as well as decreasing aggregation; the difference in the degree of aggregation of

CPK alone in the presence or absence of DTT). Indeed, HSP27 carrying R127W, S135F or R136W mutations was previously found to be hyperactive in forming sHSP:substrate complexes with numerous client proteins in the cell[24]. However, considering that WT HSP27 by its very nature exhibits sufficient chaperone activity, we are able to discuss our results in relation to the differential activity observed for WT HSP27 between reducing and oxidising conditions.

That is, by comparing aggregation under oxidising and reducing conditions we were able to monitor the differential chaperone behaviour with and without the disulphide. We could thus probe the effect of mutation on the contribution of the disulphide (Figure 6 – 17). The WT shows a marked difference in light scattering on reduction (Section 2.5). Similarly, R136W and T151I HSP27 FL have a significantly larger increase in light scattering in the presence of DTT than when oxidised, for both the 1:1 and 1:2 stoichiometric ratios. Again, the fact that there is not a significantly higher protection in the 1:2 substrate:chaperone ratio condition indicates that the increased light scattering is not simply due to independent aggregation of the HSP27.

HSP27 R127W also shows a more positive degree of protection when oxidised than when reduced. When it is used at 50 µM, it appears to show a more pronounced difference between the two redox conditions than at 100 µM. This is likely due to the lower CPK aggregation curve under DTT, as the level of light scattering in each redox condition is similar. This highlights a limitation of interpreting the aggregation curves at a fixed point, though this established practice allows for direct comparison of multiple conditions. Nonetheless, it is clear

that reduction again affects the chaperone behaviour of the protein, though to a less pronounced effect than the WT.

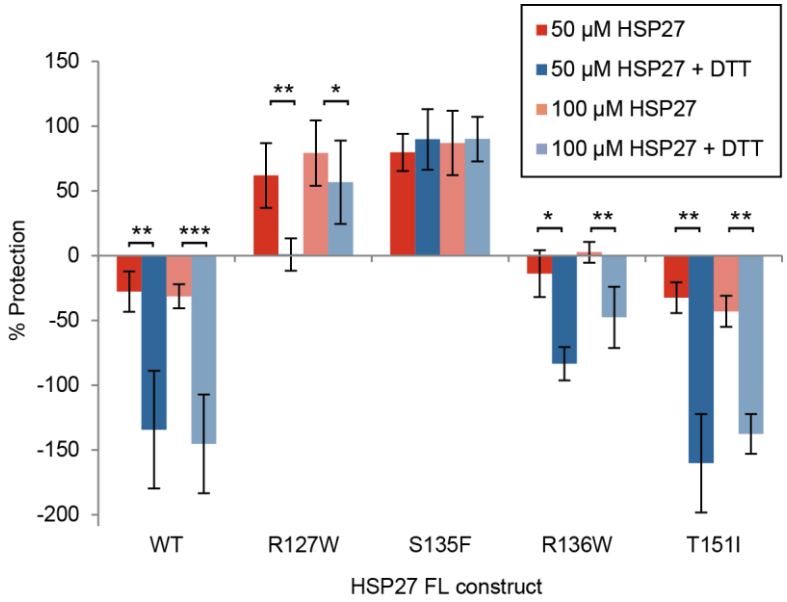


Figure 6 – 17 The effect of redox state on chaperone activity varies with mutation. Percentage protection conferred by FL HSP27 constructs on 50 µM heat-denatured CPK in the absence or presence of 20 mM DTT. Error bars correspond to mean ± s.d. percentage protection from n = 5 repeats for each condition. * P < 0.05, ** P < 0.01, *** P < 0.001, two-tailed unpaired t-test between oxidised and reduced conditions.

Strikingly, HSP27 S135F confers a similar percentage protection in each condition. It therefore seems to lack the response to the redox environment exhibited by the WT (or the other mutants to some degree), identified here by the effect of oxidation on its chaperone behaviour. S135F HSP27 is harder to oxidise than the WT (Figure 6 – 14), and so will respond less readily to oxidative stress in the cell. This was also demonstrated by a similar monomer population to the disulphide-lacking mutant (Figure 6 – 12) within the core domain on air oxidation. However, S135F HSP27 FL showed only a small population of reduced subunits under air oxidation with denaturing SDS-PAGE, similar to the WT and other mutants (Anthea Rote, manuscript in preparation), so a significant proportion of the sample will be

oxidised here. Therefore, the oxidised population present in this sample seems to exhibit similar chaperone behaviour to the reduced population (both of which almost completely suppress light scattering and formation of large aggregates), whereas the two populations appear to act differently in the WT and other mutants. The interrupted disulphide may also explain why crystallisation trials were unsuccessful – the mutation interrupts this disulphide adding heterogeneity to the sample which makes formation of ordered packing harder. Taken together, our results show that the intra-dimer disulphide is indeed interrupted by the introduction of phenylalanine at the S135 site, and furthermore, that the response of the HSP27 chaperone is mediated by the disulphide in response to oxidative stress.

It is clear that mutation is affecting the response of the chaperone to oxidation. The T151I and T164A mutations strengthen the disulphide such that it is easier to oxidise than the WT (Figure 6 – 14), exerted through allostery from their distal positions (Chapter 5). The T151 HSP27 FL shows a similar change in light scattering on oxidation to the WT, so the deleterious effect of mutation may be that it is able to assume the altered behaviour of the oxidised population more readily. The S135F, R136W and R127W mutations all weaken the intra-subunit disulphide (Section 2.8), all likely mediated by the extra steric bulk and the hydrophobic nature of the substitution. This can either disrupt the intra-dimer interface directly (R136W and S135F) or interrupt a hydrogen bond network involving residues at the interface as well as elsewhere (R127W, see Chapter 5). The precise position and chemical nature of the side chain may ultimately determine the degree of interruption on the disulphide, and concomitant response to the redox environment.

The polydispersity of the mammalian sHSPs relies on the plasticity of their interactions within their oligomers, and they are also highly dynamic with constant exchange of subunits at equilibrium. This dynamic behaviour is purportedly important in their chaperone function, and will depend on the presence of the disulphide, as demonstrated in Section 2.2. The redox potential of the disulphide varies when the different mutations are present (Section 2.8), and the chaperone activity of each mutant has a different response to the redox state.

3 Conclusion

The data in this thesis illustrate that neuropathy-indicated point mutations introduce subtle changes to the structure of the core domain of HSP27. Its redox activity and the unusual character of the inter-subunit disulphide led us to postulate that the disulphide acts as a 'redox switch' that regulates the activity of the protein in response to oxidative stress.

This hypothesis was supported in several ways: observation of the variable redox state of the inter-subunit disulphide *in vivo*; an *in vitro* chaperone assay which demonstrated that FL HSP27 had a different effect on the aggregation of CPK under oxidising and reducing conditions; MS and NMR experiments which established that the reduced protein readily accesses the monomeric state and undergoes chemical exchange on the microsecond to millisecond timescale; MS characterisation of the dynamics of SE showed that the disulphide bond greatly reduced the rate of SE of the core domain.

Observation of the disulphide-bound dimer in cell lysates indicated that the disulphide bond is relevant even within the reducing cytoplasmic environment. Moreover, the varying degree of

oxidation in different cell types not only gave us confidence that the experiment was reporting on a true cellular distribution (as it is not an artefact of the protocol); it also indicated that the degree of oxidation could be a response to the particular cellular environment.

In the SE experiment, we found that the exchange of monomers within the c27-2(WT) dimer is several orders of magnitude slower when air oxidised than when reduced. This was monitored by the appearance of heterodimers in the MS spectra. However, the disulphide bond did not have the same effect on the formation of FL HSP27 hetero-oligomers. Rather, exclusive dimer exchange, or combined exchange of monomers and dimers, gave relatively similar decay coefficients for the oxidised and reduced HSP27 FL respectively.

Together with the unusual registration state of the dimeric interface in the R136W structure and the proximity of several of these mutations to the C137 residue, this led us to hypothesise that the major effect of the point mutations is on the inter-subunit interaction. This could either be via accessing the alternate AP registration states or moderation of the SE mechanism, and both could be mediated by an interruption of the disulphide bond on mutation which would be accompanied by a change in its redox potential.

To test whether mutation caused the protein to access different registration states at the intra-dimer interface, we used IM-MS to detect any difference in AP population in the reduced and oxidised protein and with each of the mutations. However, we did not observe a significant redistribution among the registration states in any of the samples including the reduced WT and C137S, which lacked disulphides, or in S135F which likely contained a

significant population of non-covalent dimers. This strongly suggested that the deleterious effect of mutation is not caused by a modified population of the AP states.

To investigate whether the mutations affect the ease with which the disulphide can form in a particular redox environment, we used an SDS-PAGE assay and square wave voltammetry and found that each mutation does indeed have an effect on the redox potential of the disulphide. The distal T151I and T164A mutations both caused the cysteines to be more easily oxidised than in the WT c27-2 and responded over a greater redox potential window; the S135F, R136W and R127W which are on the β6+7 or β5 strand all made the cysteines harder to oxidise. Extrapolating this finding to the FL HSP27, mutation would affect the population of disulphide-bound dimers within the oligomers, and the rate at which this population increases as the oxidative environment in the cell changes. As discovered with the WT HSP27, the SE of FL oligomers is complex and involves an allosteric network of subunit interfaces. Interrupting this network by affecting the population of covalently-bound dimeric interfaces will consequently affect the strength of the inter-subunit interface between the β4-8 groove and the C-terminus[36]. It may also have an allosteric effect on the β2-β3 interaction that I have suggested is moderated by phosphorylation in Chapter 4 as well as other possible interactions within the N-terminus and with substrate. This is supported experimentally with the CPK aggregation assay, where each mutation affected the response of FL HSP27 to CPK aggregation. Moreover, reduction had a varying effect on the activity of each mutant, implying that the variable response of the disulphide in each mutant had a direct consequence on its chaperone behaviour.

In particular, we found that the S135F mutant did not display differential chaperone behaviour on oxidation, and that it was hardest to oxidise its disulphide bond in solution. R136W and R127W, which also lie close to the intra-dimer interface, were slightly easier to oxidise and did respond to oxidation in the *in vitro* chaperone assay, but showed different chaperoning behaviour to the WT. On the other hand, T151I had a similar effect on light scattering but was easier to oxidise than the WT. Therefore, the work presented here has begun to tease apart the regulatory effect of the disulphide and its variability on mutation from a structural, dynamical and functional perspective.

A logical next step to this work would be to measure the effect of mutation on the SE of the FL, ideally with the mass resolution afforded by MS[36,39]. This could be further delineated with a peptide binding experiment (Chapter 4) between the β2 peptide and CTP to the c27-2 in the presence of mutation. These experiments could be used to ascertain whether there is any cross-talk between the β2-β3 interaction (and phosphorylation at this site), the dimeric interface (and the presence of the disulphide) and the C-terminal inter-dimer interaction, which is consequently moderated by mutation. More generally, further investigation of the behaviour of the disulphide *in vivo* and identification of the HSP27 species involved in chaperoning under oxidative stress would be welcome insight into the behaviour of this complex chaperone and the deleterious effect of mutation.

In this chapter, we have therefore characterised the behaviour of the disulphide and its effect on the dynamics and activity of the protein, and presented our hypothesis that a principal effect of mutation is on the formation of the disulphide and consequently the activity of the

chaperone. This unusual molecular staple appears to be an important mechanism in the regulation of the HSP27 chaperone.

4 Methods

4.1 Materials and reagents

All reagents were purchased from Sigma unless otherwise specified.

4.2 Construct preparation

Individual point mutations were introduced to c27-2 in the modified pET28a vector (6H-TEV-c27-2) using the QuikChange Lightning mutagenesis kit (Agilent) with primers purchased from IDT and designed using the Agilent QuikChange Primer Design server. Transformation was performed in XL1 or XL10 competent cells (Agilent) and cultures grown in Luria Broth (Fisher) with kanamycin (50 µg/ml) ready for plasmid harvest (MiniPrep kit, ThermoFisher). Successful mutation was determined by plasmid sequencing (SourceBioScience).

4.3 Protein expression and purification

Protein expression and purification were undertaken under the same conditions as c27-2(WT) in Chapter 3: 1l cultures of BL21 (DE3) Gold cells (Agilent) were grown in Luria Broth (Fisher) with kanamycin (50 µg/ml) at 37°C until OD_{660} = 0.6-0.8, then induced with IPTG at 500 µM for three hours. Cells were harvested by centrifugation, washed in PBS and stored at -80°C. Cell pellets were resuspended and disrupted by sonication or microfluidisation in Buffer A (50 mM Tris, 300 mM NaCl, 20 mM imidazole, 5 mM BME, pH 8) supplemented

with 1 EDTA-free protease inhibitor cocktail tablet (one per 1l pellet, Roche), followed by clarification by centrifugation. The first purification step was immobilised metal affinity chromatography (IMAC) conducted on a HisTrap HP column and Akta FPLC system (both GE Life Sciences) in Buffer A with elution in Buffer B (50 mM Tris, 300 mM NaCl, 500 mM imidazole, 5 mM BME, pH 8), followed by a desalting column (HiPrep, GE Life Sciences) or overnight dialysis in Buffer DS (20 mM Tris, 150 mM NaCl, 20 mM imidazole, 5 mM BME, pH 8). Overnight cleavage with TEV protease (S219V mutant purified in-house, pRK793, Addgene plasmid #8827) was conducted at room temperature and followed by a reverse IMAC step in Buffer A and then size exclusion chromatography (SEC) in SEC Buffer (20 mM Tris, 150 mM NaCl, pH 8) on a Superdex 75 column 26/60 (GE Life Sciences).

4.4 Western blots

The HEK cells were a gift from Hsin-Yung Yen, and were maintained in DMEM media containing 10% FBS. The HFF cells were a gift from Bonnie van Wilgenburg. Both were washed and resuspended in PBS, lysed in lysis buffer (20 mM sodium phosphate buffer, pH 7.4, 150 mM NaCl, 10% v/v glycerol, 1% v/v NP40), in the presence of protease inhibitor (Roche). The cell lysate was harvested by centrifugation at 12000 rpm after incubation at 4 °C for 15 minutes. The cell lysates in the presence or absence of 4 % w/v DTT were then analysed by immunoblotting using standard techniques. Membranes were blocked with SeaBlock Reagent (Salmon Plasma, Merck Millipore) for 1 h at room temperature and then incubated with the primary antibody (anti-HSP27, Abcam PLC #ab2790, 1 in 1000 dilution) at room temperature for 1 h. The blots were washed four times (10 minutes) with TBS-T

(TBS containing 0.1 % Tween-20) before incubation in the peroxidase-conjugated secondary antibody (anti-mouse HRP, Sigma, 1 in 20000 dilution) for 1 h at room temperature and washed in TBS-T. HSP27 was detected using a chemiluminescent substrate (Thermo Scientific Pierce, catalogue number 32209). Monomeric and dimeric HSP27 was identified by comparison with a MW ladder (Invitrogen SeeBluePlus2).

For Figure 6 – 2b and c, cells were washed and resuspended in PBS before precipitation in ice-cold trichloroacetic acid (TCA) on ice followed by centrifugation and incubation of the pellet for 10 min in denaturing buffer (6 M Urea, 200 mM Tris-HCl pH 8.5, 10 mM EDTA, 0.5 % w/v SDS, supplemented with 100 mM iodoacetamide) and further centrifugation. The supernatant was analysed in the presence or absence of BME by immunoblotting with standard techniques as previously described[44].

4.5 Ion mobility mass spectrometry

The CCS measurements in Sections 2.2 and 2.7 were calculated from drift times recorded on a Synapt G1 Q-ToF mass spectrometer with a TriWave travelling wave ion mobility cell[53] (Waters, Manchester), operated for transmission of high m/z species[54] without collisionally activating the protein[26,28].

All samples were prepared to a final concentration of 20 µM in 200 mM ammonium acetate with reduced (100 mM DTT) c27-2 WT at pH 6.9 and air oxidised c27-2 WT at pH 6.9, pH 4 and pH 6.5, and air oxidised c27-2 carrying each of the mutations. Cytochrome C and β-lactalbumin A were used at 20-40 µM as calibrants for the CCS, and caesium iodide at 100 µg/ml was used as a mass calibrant.

Spectra were collected in ion mobility mode with the following instrument parameters: 1.4 kV capillary voltage; 10 V sample cone; 4 V extraction cone; 5 V trap; 4 V transfer; 2 $mlmin^{-1}$ flow and 0.03 mbar pressure of the trap and transfer gas (argon); 250 ms^{-1} IMS wave velocity; 5.5 - 7 V IMS wave height; 22 $mlmin^{-1}$ flow and 0.5 mbar pressure of the IMS drift gas (nitrogen); 4.3 mbar backing pressure, 1.4 $msTh^{-1/2}$ enhanced duty cycle delay coefficient, 90 µs pusher interval.

CCS values were then calculated using the Phaser software[29] and the known CCS value of the calibrants in nitrogen[55].

4.5.1 Estimating the CCS values of API, APII and APIII c27-2

The CCS values for c27-2 in the API-APIII registrations were estimated using the Projection Approximation[56,57] through the IMPACT implementation software[48] (release 0.9.1). To do so, models of the WT c27-2 lacking the β2 were constructed in the API-APIII states. The unmodified c27-2 and R136W models could not be used for the estimation as they exclude one or both of the N-terminal regions respectively. Thus, any difference in CCS due to registration shift would be masked by the difference in CCS with a structure containing the unravelled β2 strand and one that did not. The API model was constructed by aligning each chain of the c27-2 structure to the crystal structure of cABC in API (PDB: 3L1G) and the APIII model was constructed by aligning the c27-2 chains with the R136W structure (i.e. chain A and B aligned with chain A of R136W and its symmetry mate), both using the 'super' command in PyMol on the entire chain with default parameters.

Of course, the calculated CCS (CCS_{calc}) determined with IMPACT underestimates the true value as the model does not include the unstructured β2 strand. This is consistently true for the API-APIII models. To address this, as the experimental value for the oxidised protein must fully represent the APII state, all the AP estimates in Figure 6 – 5 and Figure 6 – 13 have been translated on the CCS axis to reflect this, with the assumption that the difference in the experimental and theoretical values of the CCS of the APII would be the same for the API and APIII CCS values.

For the CCS estimates of the mutants in Figure 6 – 13, IMPACT was employed on residues 10-87 for each of the c27-2 WT and mutant PDB files, which only contained the main chain and peptides. This section was chosen because most of the chains contained these residues (except one of the TI51I monomers, which had one additional residue missing).

Values depicted are the CCS_{TJM} values that IMPACT calculates from CCS_{PA} according to:

$$\mathbf{CCS_{TJM}} = 0.843 \times (\mathbf{CCS_{PA}})^{1.051} \quad (6)$$

4.6 Monitoring SE of c27-2

'Unlabelled' c27-2 was expressed and purified as described in Section 4.3. Purified 'labelled' ^{15}N,^{13}C-c27-2 was kindly donated by T.R. Alderson. BL21 (DE3) Gold E.Coli cells transformed with the pET28 6H-TEV-c27-2 plasmid were grown in M9 minimal medium with ^{15}N ammonium chloride and ^{13}C-D-glucose as the sole nitrogen and carbon source respectively. They were induced with IPTG, harvested and purified as described in Section 4.3.

For the reduced time course, unlabelled and labelled samples were dialysed into 200 mM ammonium acetate (pH 7) for two days at RT. 82.0 µM of unlabelled c27-2 and 114 µM labelled c27-2 were separately incubated with 100 mM DTT at 37°C for one hour.

For the time course with oxidised c27-2, unlabelled and labelled samples were dialysed into phosphate buffered saline ('PBS', pH 7) for two days at RT. They were pre-equilibrated separately at 37°C for one hour.

For both the oxidised and reduced time courses, the unlabelled and labelled c27-2 were mixed together in a 1:1 ratio (18 µM monomeric concentration of each) and incubated at 37°C (except for the 't=0', which was removed before incubation). Aliquots were removed over time and the exchange arrested by flash freezing (−196°C, liquid nitrogen).

Before MS analysis, the samples in PBS were exchanged into 200 mM ammonium acetate (pH 7) using BioSpin P-30 columns (BioRad) at 4°C. All samples were kept at 4°C. Mass spectra were collected on a Waters Synapt G1 in ion mobility mode with the following parameters: 1.4 kV capillary voltage; 50 V sample cone; 4 V extraction cone; 10 V trap; 4 V transfer; 2 $mlmin^{-1}$ flow and 0.03 mbar pressure of the trap and transfer gas (argon); 250 ms^{-1} IMS wave velocity; 7 V IMS wave height; 22 $mlmin^{-1}$ flow and 0.5 mbar pressure of the IMS drift gas (nitrogen); 4.1-4.2 mbar backing pressure, 0.05 mbar pressure in the quadrupole. Spectra were calibrated in Waters MassLynx software using a CsI calibrant spectrum.

Spectra were analysed using the UniDec[42] software with its IM-MS data analysis toolkit, and peak areas extracted with the Data Collector facility. Curve fitting was conducted with SigmaPlot.

4.7 Monitoring SE of FL HSP27

FL HSP27 was expressed and purified as previously described[39,58], and labelled with either Alexa 488 or Alexa 647 dyes. The two labelled samples were mixed in a 1:1 ratio to a final monomeric concentration of 25 µM and incubated at 37°C for 2 hours whereby the system had equilibrated and gave the maximal fluorescence intensity. The equilibrated labelled sample was then mixed 1:9 with unlabelled protein (25 µM final concentration) and SE was followed by loss of fluorescence intensity over time. The excitation and emission wavelengths were 485 nm and 670 nm respectively, with a filter bandwidth of 10 nm. Experiments were performed on a BMG Polarstar plate-reader. The FRET signal was measured every 2 min for 148 min, with n=3 repeats in both the absence and presence of 20 mM DTT. Curve fitting was performed in SigmaPlot.

4.8 NMR spectroscopy

c27-2 samples isotopically labelled uniformly with ^{15}N alone were prepared by growing BL21 (DE3) Gold E.Coli cells transformed with the pET28 6H-TEV-c27-2 plasmid in M9 minimal medium. ^{15}N ammonium chloride was used as the sole nitrogen source with 4 g/L of ^{12}C-glucose. The cells were induced with IPTG, harvested and purified as described in Section 4.3 before buffer exchange into 30 mM NaH_2PO_4, 2 mM EDTA, 2 mM NaN_3, pH 7.0 (NMR buffer) on a Superdex S75 26/60 or 10/300 (GE Life Sciences).

Uniformly ^{13}C and ^{15}N labelled c27-2 for the 3D resonance experiments required to assign the spectra were produced similarly.

All experiments were recorded on an 11.7 T or a 14.1 T Varian-Inova spectrometer equipped with a 5 mm *z*-axis gradient triple resonance, room temperature probe, and the spectra were processed with NMRPipe[59] and visualized in Sparky (T. D. Goddard and D. G. Kneller, SPARKY 3, University of California, San Francisco). The 2D ^{1}H-^{15}N HSQC spectra for c27-2 under reduced (5 mM BME) and oxidised conditions were recorded with eight scans and 128 increments at 298 K.

4.9 Creatine phosphokinase aggregation assays

FL HSP27 and single residue mutants were expressed and purified as previously described[39,58]. Assays were run with 50 µM CPK and 50 µM FL HSP27 (WT or carrying one of the mutations) in 50 mM phosphate buffer, pH 7.4, with 100 mM NaCl with or without 20 mM DTT. Aggregation was induced by incubating at 43°C for 5 hours and the absorbance at 340 nm measured on a FLUOstar OPTIMA microplate reader (BMG Technologies) with readings every 2 or 2.5 min.

The percentage inhibition for each chaperone was calculated according to:

$$\%\ Inhibition = \frac{\Delta\bar{A}_{340}(\text{CPK alone}) - \Delta\bar{A}_{340}(\text{CPK} + \text{HSP27})}{\Delta\bar{A}_{340}(\text{CPK alone})} \times 100\% \qquad (7)$$

Where $\Delta\bar{A}_{320}$ is the average of the three final ΔA_{320} readings (i.e. after the light scattering has plateaued).

4.10 Native MS

The spectra in Section 2.4 were collected on c27-2 as prepared in Section 4.3, but with further incubation with 400 fold excess of DTT for one hour before buffer exchange on a desalting

column into 200 mM ammonium acetate pH 6.9 and equilibration at 37 °C for two hours. The spectra were collected on 20 µM of sample using the gold-plated capillaries prepared in-house[60] on a Synapt G1 mass spectrometer in positive mode and ion-mobility mode (Waters, UK).

The spectra were collected with the following parameters: 1.4 kV capillary voltage; 150 V sample cone; 4 V extraction cone; 5 V trap; 4 V transfer; 2 $mlmin^{-1}$ flow and 0.03 mbar pressure of the trap and transfer gas (argon); 250 ms^{-1} IMS wave velocity; 6 V IMS wave height; 22 $mlmin^{-1}$ flow and 0.5 mbar pressure of the IMS drift gas (nitrogen); 0.05 mbar in the quadrupole and ion guides.

4.11 Measuring relative redox potential with glutathione-mediated reduction potential and SDS-PAGE

Each of the WT and mutant core domains was prepared as detailed in Section 4.3. The samples were then buffer exchanged into PBS pH 7 and equilibrated at 37 °C for more than 1 hour.

30 µM of each protein was incubated with a particular glutathione ratio for >2 hours at 37°C. The glutathione ratios were established using reduced L-glutathione (MP Biomedicals) and oxidised L-glutathione (AppliChem) in PBS pH 7.0 to a final total free glutathione concentration of 10mM, from 0 mM $GSSG_{ox}$ (10 mM GSH_{red}) to 5mM $GSSG_{ox}$ (0 mM GSH_{red}). The samples were then analysed with SDS-PAGE (Invitrogen LDS sample buffer and SeeBluePlus2 MW ladder) in MES buffer (Fisher) at 200V using precast 4-12% bis-tris gels (NuPAGE Novex, Invitrogen). The gels were stained overnight with InstantBlue (Expedeon), destained with deionised water and imaged (Bio-Rad Gel-Doc system). Densitometry was

performed in Bio-Rad ImageLab, and analysed in Microsoft Excel. SigmaPlot was used to perform non-linear regression with the sigmoidal model presented in Section 2.8, using a least squares minimisation (see SigmaPlot 13.0 User Manual). The sigmoidal curve was fitted on all individual data points to take account of some $GSSG_{ox}/GSH_{red}$ points that only have two or one replicates.

4.12 Measuring redox potential with square wave voltammetry

Each of the WT and mutant core domains was prepared as detailed in Section 4.3. The samples were then buffer exchanged into PBS pH 7.

A three-electrode system was employed comprising a saturated calomel reference electrode (SCE, Hach Lange, UK), a platinum mesh counter electrode (99.99 %, Goodfellow, UK), and a glassy carbon electrode (surface area 0.071 cm^2, CH Instruments, USA) partially submerged in nitrogen-purged bath solution (PBS pH 7.0 supplemented with 0.1 M potassium chloride electrolyte) within a sealed three-neck round bottom flask. This rested in a temperature controlled bath at (20 ± 1) °C in a Faraday cage. Square wave voltammetry measurements were conducted with a computer-controlled potentiostat, PGSTAT 101 (ECO-chemie, NL) and the following parameters: 25 Hz frequency; 5 mV wave amplitude; 4 mV step potential.

The bare glassy carbon electrode was polished (1.0, 0.3, and 0.05 µm alumina (Buehler, DE) slurry sequentially on a polishing pad), rinsed and dried thoroughly prior to drop-casting the protein suspension to give a surface coverage *ca.* 4.2 x 10^{-7} $mol.cm^{-2}$. The electrode was dried under flowing nitrogen and used immediately. Eight repeats were carried out for each protein.

5 References

1. Chalova, A. S., Sudnitsyna, M. V, Semenyuk, P. I., Orlov, V. N. & Gusev, N. B. Effect of disulfide crosslinking on thermal transitions and chaperone-like activity of human small heat shock protein HspB1. *Cell Stress Chaperones* **19,** 963–72 (2014).

2. Rajagopal, P., Liu, Y., Shi, L., Clouser, A. F. & Klevit, R. E. Structure of the α-crystallin domain from the redox-sensitive chaperone, HSPB1. *J. Biomol. NMR* 0–5 (2015). doi:10.1007/s10858-015-9973-0

3. Mallikarjun, V., Clarke, D. J. & Campbell, C. J. Cellular redox potential and the biomolecular electrochemical series: A systems hypothesis. *Free Radic. Biol. Med.* **53,** 280–288 (2012).

4. Held, J. M. & Gibson, B. W. Regulatory Control or Oxidative Damage? Proteomic Approaches to Interrogate the Role of Cysteine Oxidation Status in Biological Processes. *Mol. Cell. Proteomics* **11,** R111.013037 (2012).

5. Cai, Z. & Yan, L.-J. Protein Oxidative Modifications: Beneficial Roles in Disease and Health. *J. Biochem. Pharmacol. Res.* **1,** 15–26 (2013).

6. Indu, S., Kochat, V., Thakurela, S., Ramakrishnan, C. & Varadarajan, R. Conformational analysis and design of cross-strand disulfides in antiparallel β-sheets. *Proteins* **79,** 244–60 (2011).

7. Haworth, N. L., Feng, L. L. & Wouters, M. A. High torsional energy disulfides:

relationship between cross-strand disulfides and right-handed staples. *J. Bioinform. Comput. Biol.* **4,** 155–68 (2006).

8. Wouters, M. A., George, R. A. & Haworth, N. L. 'Forbidden' disulfides: their role as redox switches. *Curr. Protein Pept. Sci.* **8,** 484–95 (2007).

9. Weeks, S. D. *et al.* Molecular structure and dynamics of the dimeric human small heat shock protein HSPB6. *J. Struct. Biol.* **185,** 342–354 (2014).

10. Bagnéris, C. *et al.* Crystal Structures of α-Crystallin Domain Dimers of αB-Crystallin and Hsp20. *J. Mol. Biol.* **392,** 1242–1252 (2009).

11. Kabsch, W. & Sander, C. Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22,** 2577–2637 (1983).

12. Joosten, R. P. *et al.* A series of PDB related databases for everyday needs. *Nucleic Acids Res.* **39,** D411–D419 (2011).

13. Hochberg, G. K. A. *et al.* The structured core domain of αB-crystallin can prevent amyloid fibrillation and associated toxicity. *Proc. Natl. Acad. Sci. U. S. A.* **111,** E1562–70 (2014).

14. Sievers, F. *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7,** 1–6 (2011).

15. Goujon, M. *et al.* A new bioinformatics analysis tools framework at EMBL-EBI.

Nucleic Acids Res. **38,** W695–9 (2010).

16. Hilton, G. R., Lioe, H., Stengel, F., Baldwin, A. J. & Benesch, J. L. P. Small Heat-Shock Proteins : Paramedics of the Cell. *Top. Curr. Chem.* **328,** 69–98 (2013).

17. Haslbeck, M. & Vierling, E. A First Line of Stress Defense: Small Heat Shock Proteins and Their Function in Protein Homeostasis. *J. Mol. Biol.* **427,** 1537–1548 (2015).

18. Fleckenstein, T. *et al.* The Chaperone Activity of the Developmental Small Heat Shock Protein Sip1 Is Regulated by pH-Dependent Conformational Changes. *Mol. Cell* 1–12 (2015). doi:10.1016/j.molcel.2015.04.019

19. Rajasekaran, N. S. *et al.* Human αB-Crystallin Mutation Causes Oxido-Reductive Stress and Protein Aggregation Cardiomyopathy in Mice. *Cell* **130,** 427–439 (2007).

20. Rogalla, T. *et al.* Regulation of Hsp27 Oligomerization, Chaperone Function, and Protective Activity against Oxidative Stress/Tumor Necrosis Factor by Phosphorylation. *J. Biol. Chem.* **274,** 18947–18956 (1999).

21. Yan, L.-J. *et al.* Mouse heat shock transcription factor 1 deficiency alters cardiac redox homeostasis and increases mitochondrial oxidative damage. *EMBO J.* **21,** 5164–5172 (2002).

22. Laganowsky, A. *et al.* Crystal structures of truncated alphaA and alphaB crystallins reveal structural mechanisms of polydispersity important for eye lens function. *Protein Sci.* **19,** 1031–43 (2010).

23. Baranova, E. V *et al.* Three-dimensional structure of α-crystallin domain dimers of human small heat shock proteins HSPB1 and HSPB6. *J. Mol. Biol.* **411,** 110–22 (2011).

24. Almeida-Souza, L. *et al.* Increased monomerization of mutant HSPB1 leads to protein hyperactivity in Charcot-Marie-Tooth neuropathy. *J. Biol. Chem.* **285,** 12778–86 (2010).

25. Ylikallio, E. *et al.* Truncated HSPB1 causes axonal neuropathy and impairs tolerance to unfolded protein stress. *BBA Clin.* **3,** 233–242 (2015).

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

27. Baldwin, A. J., Lioe, H., Robinson, C. V., Kay, L. E. & Benesch, J. L. P. αB-Crystallin Polydispersity Is a Consequence of Unbiased Quaternary Dynamics. *J. Mol. Biol.* **413,** 297–309 (2011).

28. Kondrat, F. D. L., Struwe, W. B. & Benesch, J. L. P. in *Struct. Proteomics High-Throughput Methods* (ed. Owens, R. J.) **1261,** 349–371 (Springer New York, 2015).

29. Allison, T. M. *et al.* Quantifying the stabilizing effects of protein–ligand interactions in the gas phase. *Nat. Commun.* **6,** 8551 (2015).

30. Basha, E., O'Neill, H. & Vierling, E. Small heat shock proteins and α-crystallins: dynamic proteins with flexible functions. *Trends Biochem. Sci.* **37,** 106–117 (2012).

31. Jehle, S. *et al.* αB-Crystallin: A Hybrid Solid-State/Solution-State NMR Investigation Reveals Structural Aspects of the Heterogeneous Oligomer. *J. Mol. Biol.* **385,** 1481–1497 (2009).

32. Mymrikov, E. V, Seit-Nebi, A. S. & Gusev, N. B. Heterooligomeric complexes of human small heat shock proteins. *Cell Stress Chaperones* **17,** 157–69 (2012).

33. Baldwin, A. J. *et al.* Quaternary dynamics of αB-crystallin as a direct consequence of localised tertiary fluctuations in the C-terminus. *J. Mol. Biol.* **413,** 310–20 (2011).

34. Bova, M. P., Mchaourab, H. S., Han, Y. & Fung, B. K.-K. Subunit Exchange of Small Heat Shock Proteins: Analysis of oligomer formation of αA-crystallin and HSP27 by fluorescence resonance energy transfer and site-directed truncations. *J. Biol. Chem.* **275,** 1035–1042 (2000).

35. Sobott, F., Benesch, J. L. P., Vierling, E. & Robinson, C. V. Subunit exchange of multimeric protein complexes: real-time monitoring of subunit exchange between small heat shock proteins by using electrospray mass spectrometry. *J. Biol. Chem.* **277,** 38921–38929 (2002).

36. Hilton, G. R. *et al.* C-terminal interactions mediate the quaternary dynamics of αB-crystallin. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **368,** 20110405 (2013).

37. Painter, A. J. *et al.* Real-Time Monitoring of Protein Complexes Reveals their Quaternary Organization and Dynamics. *Chem. Biol.* **15,** 246–253 (2008).

38. Benesch, J. L. P. *et al.* The Quaternary Organization and Dynamics of the Molecular

Chaperone HSP26 Are Thermally Regulated. *Chem. Biol.* **17,** 1008–1017 (2010).

39. Aquilina, J. A., Shrestha, S., Morris, A. M. & Ecroyd, H. Structural and functional aspects of hetero-oligomers formed by the small heat shock proteins αB-crystallin and HSP27. *J. Biol. Chem.* **288,** 13602–9 (2013).

40. Delbecq, S. P., Rosenbaum, J. C. & Klevit, R. E. A Mechanism of Subunit Recruitment in Human Small Heat Shock Protein Oligomers. *Biochemistry* **54,** 4276–4284 (2015).

41. Krantz, B. A. *et al.* Understanding protein hydrogen bond formation with kinetic H/D amide isotope effects. *Nat. Struct. Biol.* **9,** 458–463 (2002).

42. Marty, M. T. *et al.* Bayesian Deconvolution of Mass and Ion Mobility Spectra: From Binary Interactions to Polydisperse Ensembles. *Anal. Chem.* **87,** 4370–4376 (2015).

43. Haslbeck, M. & Buchner, J. in *Stress Responses Methods Protoc.* (ed. Christine M. Oslowski) **1292,** 39–51 (Springer New York, 2015).

44. Kulig, M. & Ecroyd, H. The small heat-shock protein αB-crystallin uses different mechanisms of chaperone action to prevent the amorphous versus fibrillar aggregation of α-lactalbumin. *Biochem. J.* **448,** 343–352 (2012).

45. Bova, M. P. *et al.* Mutation R120G in B-crystallin, which is linked to a desmin-related myopathy, results in an irregular structure and defective chaperone-like function. *Proc. Natl. Acad. Sci.* **96,** 6137–6142 (1999).

46. Cox, D., Carver, J. A. & Ecroyd, H. Preventing α-synuclein aggregation: The role of the small heat-shock molecular chaperone proteins. *Biochim. Biophys. Acta - Mol. Basis Dis.* **1842,** 1830–1843 (2014).

47. Heirbaut, M., Beelen, S., Strelkov, S. V. & Weeks, S. D. Dissecting the Functional Role of the N-Terminal Domain of the Human Small Heat Shock Protein HSPB6. *PLoS One* **9,** e105892 (2014).

48. Marklund, E. G., Degiacomi, M. T., Robinson, C. V., Baldwin, A. J. & Benesch, J. L. P. Collision Cross Sections for Structural Proteomics. *Structure* **23,** 791–799 (2015).

49. De St. Groth, S. F., Webster, R. G. & Datyner, A. Two new staining procedures for quantitative estimation of proteins on electrophoretic strips. *Biochim. Biophys. Acta* **71,** 377–391 (1963).

50. Maher, P. The effects of stress and aging on glutathione metabolism. *Ageing Res. Rev.* **4,** 288–314 (2005).

51. Sevier, C. S. & Kaiser, C. A. Formation and transfer of disulphide bonds in living cells. *Nat. Rev. Mol. Cell Biol.* **3,** 836–847 (2002).

52. Li, F. *et al.* Redox active motifs in selenoproteins. *Proc. Natl. Acad. Sci. U. S. A.* **111,** 6976–6981 (2014).

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

54. Campuzano, I. & Giles, K. in *Nanoproteomics Methods Protoc.* (eds. Toms, S. A. & Weil, R. J.) **790,** 57–70 (Humana Press, 2011).

55. Bush, M. F. *et al.* Collision cross sections of proteins and their complexes: a calibration framework and database for gas-phase structural biology. *Anal. Chem.* **82,** 9557–65 (2010).

56. Mack, E. Average cross-sectional areas of molecules by gaseous diffusion methods. *J. Am. Chem. Soc.* **47,** 2468–2482 (1925).

57. von Helden, G., Hsu, M. T., Gotts, N. & Bowers, M. T. Carbon cluster cations with up to 84 atoms: structures, formation mechanism, and reactivity. *J. Phys. Chem.* **97,** 8182–8192 (1993).

58. Jovcevski, B. *et al.* Phosphomimics Destabilize Hsp27 Oligomeric Assemblies and Enhance Chaperone Activity. *Chem. Biol.* **22,** 186–195 (2015).

59. Delaglio, F. *et al.* NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **6,** 277–293 (1995).

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

61. Trivedi, M. V, Laurence, J. S. & Siahaan, T. J. The role of thiols and disulfides on protein stability. *Curr. Protein Pept. Sci.* **10,** 614–25 (2009).

Summary

The role of the sHSPs in a variety of cellular processes, along with their dynamic, highly effective, ATP-independent response, has demanded their structural and functional characterisation. For example, HSP27 is involved in an array of pathways such as cytoskeletal maintenance, tumour growth, protein degradation, preventing redox imbalance and protein aggregation, which may all be mediated by its chaperone activity[1]. In this thesis, I have described investigations undertaken to gain greater structural insight into the regulation of HSP27 by oxidation and phosphorylation.

In Chapter 3, I described attempts to crystallise HSP27, which were ultimately successful by employing co-crystallisation of the core domain and CTP. We observed the canonical dimer with the inter-subunit disulphide bond intact, providing a structural reference for the behaviour of this covalent bond. Additionally, we observed surprising behaviour of the β2 strand, which was extended to interact with two neighbours in the crystal lattice, rather than forming part of the minor β-sheet or being unresolved as it had been in previous models. This behaviour was particularly notable for two reasons: the N-terminal GV motif resided in the β4-β8 groove of a neighbour, which was vacant of the CTP usually found at this position; and in the FL sequence two phosphorylation sites lie immediately before this unravelled region. Phosphorylation interrupts oligomerisation and chaperone performance of HSP27 (as well as being implicated in sHSP regulation throughout the eukaryotic sHSPs[2]).

Therefore, we conducted a number of experiments to probe the behaviour of the β2 region and test whether our observations were replicated in solution. Firstly, MD simulations using the previously published structure where the β2 was hydrogen-bonding to the β3 showed that the β2 was rapidly forming and breaking hydrogen bonds with the β3 and could fully dissociate after 200 ns of the simulation, but was not so unstable as to fully dissociate more rapidly or in every simulation. This result was recapitulated in solution where NMR generalised order parameters showed that the β2 was not completely disordered in solution but does demonstrate mobility on the picosecond timescale. However, native MS showed that the dimer was by far the major species in the oxidised c27-2 sample, indicating that the runaway oligomers observed in the crystal structure were not present in solution. Further, an MS competition assay involving the dimer and peptides mimicking the IPV-containing region of the C-terminus or the β2-containing region showed that the identity of the β2-pep did not affect the binding of the CTP, implying that the β2-WT or β2-pS82 were not competing with the CTP. However, it did indicate that both the β2-WT and β2-pS82 have a higher affinity for the dimer than a scrambled peptide of the same region. This was progressed with solution NMR which showed that at higher fold concentrations, the β2-pS82 elicited larger CSPs and greater attenuation of the peak intensity than β2-WT, indicating greater binding. Our ongoing work has indicated that the β2-pep may be competing with the binding of the integral β2 of c27-2, and that binding causes a greater number of subunits to occupy the monomer stoichiometry. Taken together, this work has indicated a structural pathway for the influence of phosphorylation at sites 78 and 82.

In Chapter 5, I described crystallisation of c27-2 carrying single residue mutations identified in motor neuropathies. Three proteins crystallised successfully, namely the c27-2(T151I), c27-2(T164A) and c27-2(R136W). The first two structures appeared very similar to the WT, replete with the extended β2 and, through comparison with the WT, it was possible to postulate the structural implications of the mutations. In contrast, the β2 of the R136W mutant was unresolved, and both β4-8 grooves in the dimer were occupied by the CTP. Notably, the c27-2(R136W) dimer occupied the APIII registration state at the dimer interface, which has not been directly observed for HSP27 before. Concurrently, the intra-dimer disulphide bond was not formed. There were also two regions of additional electron density, one of which was consistent with a BME molecule attached to one of the C137 side chains, and the other which was reminiscent of similar density seen in previous crystal structures, perhaps suggestive of an interaction interface at this site.

The alternate registers observed for c27-2(WT) and R136W, along with the unusual nature of the intra-dimer disulphide, led us to postulate that the reported redox response of HSP27 may be mediated by the disulphide, and that mutation may interrupt the formation of this bond. We therefore undertook a series of investigations to test our hypotheses. Firstly, we found that the disulphide is formed *in vivo*, and was formed to a variable degree in the lysates of several human cell lines, indicating that it is formed in response to the intra-cellular environment. We next tested to see whether the crystallisation conditions (specifically, pH and obligatory reduction induced by the BME molecule) could have led the dimer to access the APIII state, using IM-MS. We found that the major population of the c27-2 dimer was APII under the different conditions, suggesting mutation may have been the cause of the different registration

state. We also found that the disulphide had a significant stabilising effect on the intra-dimer interface, such that the rate of decay of oxidised c27-2 homodimers in a sample containing isotopically labelled and unlabelled subunits was two orders of magnitude slower than in a reduced sample. In the FL protein, exchange of either dimers alone, or both dimers and monomers together, resulted in a decay rate of the same order for the oxidised and reduced protein respectively. There was only a slightly enhanced exchange rate on reduction, which is probably due to an allosteric network of interfaces in the FL HSP27 oligomer. We have therefore characterised the effect of oxidation on both the intra- and inter-dimer interfaces. A ^{15}N-^{1}H HSQC experiment on c27-2 showed that it occupies additional chemical states on removal of the disulphide, providing a residue-level description of the monomer-dimer exchange observed with MS. We found that the presence of the disulphide affects the chaperone activity of FL HSP27 against CPK aggregation, which together with our *in vivo* and dynamical investigations, indicated that the disulphide exerts regulation on HSP27 behaviour.

In light of this, we began to investigate the behaviour of the mutants, postulating that mutation may interrupt the formation of the disulphide. Native MS on oxidised samples showed that most of the mutants occupied the dimer stoichiometry, with a very similar profile to the WT. Notably, S135F showed a very similar distribution to the disulphide-lacking C137S mutant, with both monomers and dimers present. This suggested that the c27-2(S135F) mutant had a greater reduced population than the WT or other mutants after equilibration, allowing some of the non-covalent dimers to dissociate into monomers. Notably c27-2(R136W) did not exhibit a significantly larger monomeric population, implying that it wasn't as

reduced as c27-2(S135F). We therefore turned to IM-MS to investigate whether the APIII register observed in the c27-2(R136W) structure was significantly populated in the ensemble, but found that all the mutants displayed a similar CCS distribution to the APII oxidised WT. Therefore, the APIII is only lowly and transiently populated in solution, and does not appear to be differentially accessed by the different mutants. This suggested that the effect of mutation was not to alter the population of the intra-dimer registers. Rather, it could affect the monomer-dimer distribution, which would also be consistent with the differences in the strength of the disulphide we had begun to delineate. To characterise this fully, we employed a biochemical assay and square wave voltammetry, which showed that c27-2 T151I and T164A are both more easily oxidised than the WT, and c27-2 R127W, R136W and S135F are harder to oxidise. The latter three mutations are all either situated on the intra-dimer interface or just behind it, so introduction of large, hydrophobic groups will place strain on the extended β-sheet interface or interrupt stabilising hydrogen-bonding networks involving the sidechains. This was supported by chaperone assays *in vitro* where all three of these mutants showed different chaperone behaviour to the WT. Moreover, c27-2(S135F), which had the disulphide that was hardest to oxidise, did not exhibit the change in chaperone behaviour on oxidation that was observed for the WT, which we believe shows the redox switch role of the disulphide has been interrupted.

Taken together, this thesis has detailed structural, functional and dynamical investigations into the regulation of HSP27, and the change in this modulation exerted by mutation. Our data suggest two mechanisms of regulation in HSP27 in addition to the modulation of stoichiometry and SE by temperature, pH and substrate load. Specifically, I have suggested

that phosphorylation at S82 affects the behaviour of the dynamic β2-region; and the chaperone response of HSP27 to redox state is mediated by the redox switch disulphide, whose behaviour is interrupted on mutation.

Moreover, we would expect these two sites to exert influence on one another and the inter-dimer interaction involving the C-terminus, within the intricate system of allostery observed within sHSP oligomers. This will be the focus of work going forward, to better characterise these complex and often enigmatic proteins.

The study of sHSPs over the last three decades has increasingly revealed a dynamic and complex network, modulated by the cellular environment or PTM. Their stoichiometry, SE and possibly even shape can be modulated by pH, temperature, oxidation and phosphorylation (amongst other factors). In addition, the regulation of their expression, frequent co-expression in the same cellular compartment, and their ability to co-assemble to form hetero-oligomers, makes for a staggering number of possible assemblies that are able to act to maintain proteostasis[3]. This heterogeneity may facilitate the sHSPs' collective response to a wide range of partially unfolded substrates, both in their role within native processes or under a variety of cellular stresses[2]. Here, I hope I have provided greater detail into the regulation and response of one of the most prevalent sHSPs in the body, and by extrapolation, on this essential and highly complex network.

References

1. Arrigo, A.-P. & Gibert, B. Protein interactomes of three stress inducible small heat

shock proteins: HspB1, HspB5 and HspB8. *Int. J. Hyperth.* **29,** 409–22 (2013).

2. Haslbeck, M. & Vierling, E. A First Line of Stress Defense: Small Heat Shock Proteins and Their Function in Protein Homeostasis. *J. Mol. Biol.* **427,** 1537–1548 (2015).

3. Hilton, G. R., Lioe, H., Stengel, F., Baldwin, A. J. & Benesch, J. L. P. in *Top. Curr. Chem.* (ed. Jackson, S.) **328,** 69–98 (Springer Berlin Heidelberg, 2013).