Dynamics and function of small heat-shock proteins

Georg Hochberg

Pembroke College
University of Oxford

A thesis submitted for the degree of
Doctor of Philosophy in Systems Biology

Hilary 2015

Abstract

This thesis describes novel insights into the nature, dynamics and function of non-covalent protein complexes formed by the small heat-shock proteins (sHSPs), a family of important molecular chaperones. This work first provides insight into how conformational dynamics of a C-terminal motif influence the oligomeric distribution of a human sHSP. Further, this work interrogates the link between oligomeric structure and chaperone function in the same human sHSP by engineering and characterising functionally truncated constructs devoid of all oligomeric dynamics. I demonstrate that such truncated domains recapitulate the function of large and heterogeneous oligomers, dispelling the long-held belief that oligomeric structure and function are inextricably linked in sHSPs. Lastly, I show how small sHSPs in plants avoid forming hetero-oligomers between two functionally distinguishable paralogues that are co-expressed in the same tissue. I show how non-interfacial regions play a part in oligomeric selectivity through their influence on the structural dynamics of the conserved core domain of sHSPs. I also demonstrate that labile interfaces that are rapidly broken and re-formed in oligomers can provide oligomeric selectivity between the two proteins, whereas more stable but non-selective interfaces provide most of the thermodynamic stability of the oligomers formed by both proteins at equilibrium. Together these insights provide a framework for interrogating and understanding the underlying causes and functional importance of oligomeric structure and dynamics in biology.

Dynamics and function of small heat-shock proteins



Georg Hochberg

Pembroke College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy in Systems Biology

Hilary 2015

I dedicate this work to my teachers.

Acknowledgements

I am indebted to a large number of people for seeing me through this DPhil. In particular Gill Hilton for reviving my enthusiasm for science after synthetic chemistry had nearly extinguished it. Art Laganowsky for taking the time to make an actual scientist out of me, for embarrassing me in front of as many female members of the Chemistry Department and wider Oxford community as he could manage, getting me very drunk whenever I needed it, and always being a source of scientific inspiration. Eamonn Reading for the merry nights out of the lab and the hungover mornings in the lab. I would not have lasted the whole way without him. Dale Shepherd and Erik Marklund for joining forces with me on what became the most intellectually satisfying and enjoyable part of my DPhil. I thank Andrew Baldwin for always being a challenging mentor who was patient enough to teach protein dynamics to a plant biologist. I am indebted to Heath Ecroyd and Elizabeth Vierling for our discussions about the biology and evolution of small heat-shock proteins. I wish to apologise to Jon Hopper for my ill-conceived pranks, and thank him for his mass spectrometry advice. I thank Timothy Allison for his generous advice and help in molecular biology and crystallography, and his dry sense of humour. I thank Weston Struwe for his generous help on the SAXS beamline, and for preparing me mentally and culturally for America. I'm grateful for Fran Konradt's help with the α-crystallin work and her generously given advice. I am sorry Heidi Gastall had to put up with my childish sense of humour for so long, and am grateful for her cheerful presence throughout most of my DPhil. I thank Olya Tkachenko for the great time I had working with her, and our thought-provoking discussions. I am grateful to Todd Mize and Helena Hernandez for running such a well-organized lab despite my best efforts. I thank Matteo Degiacomi for his model building efforts and our discussions about protein assemblies. I am also indebted to all members of the Robinson, Baldwin, and Benesch groups that I have had the privilege to work with over the last three and a half years. I am grateful for my time at Pembroke College, in particular together with Josh Rosaler, Ben Jones, Meghan Campbell, Rens Krijgsman and Geoff Swenson. They ensured I created and erased a lot of lasting memories during my DPhil. I also thank Meelis Lootus and Himadri Mukhopadhyay for three indefensibly wild years under the same roof.

I want to thank Carol for constantly being an inspiring and supportive presence, and for the usually unannounced and very stimulating scientific grillings she subjected me to in her car or next to water fountains. Finally, I am in-

debted to Justin for being such an inspiring and supportive mentor. I am grateful he risked taking on a biologist who had never heard of native mass spectrometry as his first DPhil student, for extending that risk to giving me complete academic freedom, and for always doing his utmost to ensure I had all the support I could wish for. He was always a source of calm when I was anxious about my work and a source of urgency when I was too calm about it. He taught me how to think, talk, and write about my work. The completely coincidental circumstances that lead me to working in his lab make me all the more grateful for the wonderful time that I have had.

Statement of Originality

This work presents novel insights into the nature, structural causes, and functional importance of quaternary dynamics in the small heat-shock proteins (sHSPs). Each chapter is written to stand as a self-contained piece of work. While this introduces some repetition, it makes this thesis easier to read in several short rather than one long sitting. I hope the reader will forgive me for this indulgence.

Chapter 1 frames the problem in the broader context of quaternary structure, dynamics and function in biology and gives a brief technical description of the techniques used in this thesis.

Chapter 2 describes how a structural motif in the C-terminus of αB-crystallin, a human sHSP, influences the quaternary dynamics of αB-crystallin oligomers through an allosteric effect on the protomeric dimer. The work presented in this chapter was published as part of the following publication:
G. R. Hilton, **G. K. Hochberg**, A. Laganowsky, S. I. McGinnigle, A. J. Baldwin, and J. L. P. Benesch, "*C-terminal interactions mediate the quaternary dynamics of αB-crystallin*", Philos. Trans. R. Soc. Lond. B.Biol. Sci., vol. 368, no. 1617, p. 20110405, 2013.

Chapter 3 interrogates the functional role of quaternary dynamics in αB-crystallin. Truncated constructs of αB-crystallin engineered to be devoid of all quaternary dynamics are shown to be very potent chaperones. By comparison to an inactive, but structurally virtually identical paralogue it is shown that chaperone activity in αB-crystallin is encoded specifically in the structure of a simple immunoglobulin domain. The work presented in this chapter is published in:
G. K. Hochberg*, H. Ecroyd*, C. Liu, D. Cox, D. Cascio, M. R. Sawaya,M. P. Collier, J. Stroud, J. A. Carver, A. J. Baldwin, C. V. Robinson, D. S. Eisenberg, J. L. P. Benesch, and A. Laganowsky, "*The structured core domain of αB-crystallin can prevent amyloid fibrillation and associated toxicity.*", Proc. Natl. Acad. Sci. U.S.A., vol. 111, pp. E1562-E1570, Apr. 2014.
The functional assays in this chapter were carried out by Dr Heath

Ecroyd (*in vitro* assays, as a co-first author) and Dr Cong Liu (*in vivo* assays).The crystallography was carried out together with Dr Art Laganowsky, with help from Dr Dulio Cascio and Dr Michael Sawaya. Miranda Collier, Dezerae Cox, and Dr James Stroud contributed intellectually, although none of their data contributed directly to the work.

Chapter 4 attempts a synthesis of our current understanding of αB-crystallin's structural dynamics and their biological importance. The chapter also highlights how rapid quaternary dynamics can lead to erroneous measurements of stoichiometry using conventional structural biology techniques. This chapter is published in:
G. K. Hochberg and J. L. P. Benesch, "*Dynamical structure of αB-crystallin*", Prog. Biophys. Mol. Biol., 2014 vol. 115, no. 1, pp. 11-20, Mar. 2014

Chapter 5 explores how two functionally distinct plant sHSP paralogues avoid forming hetero-oligomers despite being co-expressed in the same tissue. The chapter explores which steps in oligomeric assembly are selective enough to ensure that formation of hetero-oligomers is prevented, and how residues that are not part of protein-protein interfaces can ensure such specificity through their influence on structural dynamics. The work in this chapter is a manuscript in preparation, with the following authors contributing:
G. K. Hochberg*, D. A. Shepherd*, E. G. Marklund*, I. Santhanagoplan, M. Degiacomi, A. Laganowksy, T. M. Allison, W. B. Struwe, E. Basha, N. Bopp, M. T. Marty, M. Tully, J. Doutch, A. J. Baldwin, Carol V. Robinson, E. Vierling and J. L. P. Benesch. "*Oligomeric incompatibility in two plant small heat-shock proteins.*"

I designed, cloned and purified all other constructs relevant to this chapter, with the exception of the HSP 18.1 cysteine mutant, which was provided by Indu Santhanagoplan. Dale Shepherd (co-first author) carried out subunit exchange experiments for HSP 18.1, the dimer titrations, and the crystallography for cHSP 18.1. Erik Marklund (co-first author) carried out the molecular dynamics simulations and analysed the potential of mean force simulations. Eman Basha provided intitial protein samples. Mark Tully, James Doutch and Weston Struwe assisted with SAXS experiments. The SAXS learning algorithm was designed and implemented by Matteo Degiacomi. Michael Marty developed software to extract relative abundances from mass spectra (since published elsewhere). Art Laganowsky and Tim Allison provided help with solving and refining the structures. Nathen Bopp provided a list of plant sHSPs used for phylogenetic analysis. All other work and analysis presented in

this chapter was carried out by me.

Abstract

This thesis describes novel insights into the nature, dynamics and function of non-covalent protein complexes formed by the small heat-shock proteins (sHSPs), a family of important molecular chaperones. This work first provides insight into how conformational dynamics of a C-terminal motif influence the oligomeric distribution of a human sHSP. Further, this work interrogates the link between oligomeric structure and chaperone function in the same human sHSP by engineering and characterising functionally truncated constructs devoid of all oligomeric dynamics. I demonstrate that such truncated domains recapitulate the function of large and heterogeneous oligomers, dispelling the long-held belief that oligomeric structure and function are inextricably linked in sHSPs. Lastly, I show how small sHSPs in plants avoid forming hetero-oligomers between two functionally distinguishable paralogues that are co-expressed in the same tissue. I show how non-interfacial regions play a part in oligomeric selectivity through their influence on the structural dynamics of the conserved core domain of sHSPs. I also demonstrate that labile interfaces that are rapidly broken and re-formed in oligomers can provide oligomeric selectivity between the two proteins, whereas more stable but non-selective interfaces provide most of the thermodynamic stability of the oligomers formed by both proteins at equilibrium. Together these insights provide a framework for interrogating and understanding the underlying causes and functional importance of oligomeric structure and dynamics in biology.

Contents

List of Figures **vii**

List of Tables **xi**

1 Protein quaternary structure, dynamics and function in biology **1**

1.1 Abstract 1

1.2 Protein quaternary structure in biology 2

1.2.1 Functional role of protein oligomers 2

1.2.2 Dynamics of protein oligomers 4

1.3 Small heat-shock proteins: A model for quaternary structure and dynamics relationships. 6

1.3.1 sHSP oligomers and their building blocks 6

1.3.2 Quaternary dynamics of sHSP oligomers 9

1.3.3 Regulation of sHSP quaternary dynamics 11

1.4 Biophysical techniques for the study of protein quaternary assemblies 12

1.4.1 X-ray crystallography 12

Reconstruction of atomic structures from X-ray diffraction data 13

Quaternary structure in crystal structures 15

1.4.2 Small angle X-ray scattering 15

1.4.3 Native mass spectrometry 16

Instrument configuration used in this thesis: the Q-TOF 17
Nano electro-spray ionization 17
Transmission of large ions 19
Mass analyzers for native MS 19
Gas-phase manipulation of ions 21
Ion mobility spectrometry 22
1.5 References 25

2 C-terminal interactions and dynamics of αB-crystallin 43
2.1 Abstract 43
2.2 Introduction 43
2.3 Results 48
2.3.1 The C-terminal extension of αB-crystallin modulates the binding of the tail to the core domain 48
2.3.2 IM-MS allows quantitative extraction of peptide binding affinities to the αB-crystallin core domain 49
2.3.3 Binding of the second peptide allosterically weakens the dimer interface 52
2.4 Discussion 53
2.5 Materials and methods 55
2.5.1 Measuring peptide binding to the αB-crystallin core domain 55
2.5.2 Ion-mobility mass spectrometry titration data analysis . 56
2.5.3 Determination of peptide binding affinity 56
2.5.4 Ligand binding model 58
2.5.5 Simplifying the model 59
2.5.6 Fitting the data to the model 60
2.5.7 Determination of errors in fitted binding affinity parameters 61
2.5.8 Binding model assessment 61

2.5.9 Adjustments due to non-specific association in the gas phase . . . 62
2.6 References . . . 63

3 The minimal chaperone unit of αB-crystallin 73
3.1 Abstract . . . 73
3.2 Introduction . . . 74
3.3 Results . . . 78
3.3.1 An alternative strategy to crystallise mammalian sHSPs 78
3.3.2 The core domain dimer of HSP27 has cysteines located about a two-fold axis . . . 80
3.3.3 An AP_{II} locked cABC dimer . . . 82
3.3.4 cABC predominantly populates the AP_{II} register . . . 84
3.3.5 The dynamics of $cABC_{E117C}$ under reducing conditions are equivalent to cABC . . . 85
3.3.6 cABC can prevent amorphous protein aggregation *in vitro* 87
3.3.7 cABC can prevent amyloid fibril formation *in vitro* . . . 90
3.3.8 The interaction of cABC with aggregating $A\beta_{1-42}$ reduces toxicity to cells . . . 91
3.4 Discussion . . . 92
3.5 Materials and Methods . . . 95
3.5.1 αBC, cABC and cHSP27 protein expression and purification . . . 95
3.5.2 αBC and HSP27 C-terminal peptide expression and purification . . . 97
3.5.3 Crystallization of cABC and cHSP27 in complex with their C-terminal peptides . . . 98
3.5.4 Structure determination and analysis . . . 98
3.5.5 Mass spectrometry . . . 99
3.5.6 NMR spectroscopy . . . 100

3.5.7 *In vitro* protein aggregation assays 101
3.5.8 Cell viability assays 102
3.6 References 103

4 Dynamical structure of αB-crystallin 115
4.1 Abstract 115
4.2 αB-crystallin: a refractory refractor 116
4.3 Coupled quaternary dynamics and polydispersity of αB-crystallin 117
4.3.1 The extent of polydispersity 118
4.3.2 Quantifying the rate at which αB-crystallin exchanges subunits 119
4.3.3 Fast subunit exchange can compromise measurements of polydispersity 120
4.3.4 Fast kinetics require faster measurements of the oligomeric distribution 122
4.4 Kinetic and allosteric determinants of the αB-crystallin ensemble. 124
4.4.1 The exchange model for the origin of αB-crystallin polydispersity 126
4.4.2 Allosteric communication between αB-crystallin interfaces 127
4.5 Dynamical structure of the αB-crystallin interfaces 128
4.5.1 The C-terminal tail cross-links dimers into oligomers . . 128
4.5.2 C-terminal transitions mediate oligomer disassembly . . 129
4.5.3 Variable binding of the C-terminal tail 131
4.5.4 Polymorphic dimeric interfaces 134
4.6 Integrative biology of αB-crystallin 135
4.6.1 Forging consensus structures of αB-crystallin oligomers 135
4.6.2 αB-crystallin and the dynamics-function paradigm . . . 137
4.7 References 139

5 Oligomeric incompatibility in two plant small heat-shock proteins 151

5.1 Abstract . . . 151

5.2 Introduction . . . 152

5.3 Results . . . 155

5.3.1 HSP 18.1 and HSP 17.7 do not hetero-oligomerise . . . 155

5.3.2 HSP 18.1 and HSP 17.7 have tetrahedral oligomeric architecture . . . 156

5.3.3 HSP 18.1 and HSP 17.7 display similar quaternary dynamics . . . 159

5.3.4 The dimer interface is incompatible, despite an extremely conserved structure . . . 163

5.3.5 The dimer interface is incompatible, despite an extremely conserved structure . . . 165

5.3.6 The formation of hetero-dimers depends on non-interfacial loops in the core domains . . . 165

5.3.7 Mutations in the loop affect hetero-dimerisation via the stability of monomers . . . 168

5.3.8 Dissociated monomers adopt distinct conformations . . . 170

5.3.9 The C-terminus provides specificity in the first of two higher-order assembly steps . . . 173

5.3.10 The N-terminus acts as a second, nonspecific step in assembly . . . 177

5.4 Discussion . . . 180

5.5 Materials and Methods . . . 183

5.5.1 Protein expression and purification . . . 183

5.5.2 Native (ion-mobility) mass spectrometry . . . 184

5.5.3 Subunit exchange of full-length proteins . . . 185

5.5.4 Extraction of dimer K_D values . . . 185

5.5.5 Crystallography . . . 187

5.5.6 Peptide binding experiments 187
5.5.7 Small angle X-ray scattering 187
5.5.8 SAXS learning algorithm 189
5.5.9 Molecular dynamics 189
5.5.10 Analysis of sequence conservation 191
5.6 References 191

6 Concluding remarks **203**

List of Figures

1.1 Quaternary structure in proteins 3
1.2 Small heat-shock proteins 7
1.3 Terminal interactions of sHSPs 8
1.4 X-ray crystallography 13
1.5 The Q-TOF instrument 17
1.6 The nanoESI ionization process 18
1.7 Collision-induced-dissociation of protein complexes 22
1.8 Ion-mobility spectrometry of protein complexes 23

2.1 Structural illustration of the C-terminus binding to the ACD . . 46
2.2 Peptide binding to cABC 48
2.3 IM-MS quantification of peptide binding 50
2.4 K_Ds extracted from IM-MS data 51
2.5 Fits of binding models to intensities extracted with CHAMP . . 52
2.6 CHAMP fits to experimental peptide-binding data 57

3.1 Structures of the αB and HSP27 core domains 77
3.2 Structural alignment of mammalian ACD X-ray and solid state NMR structures 79
3.3 Sequence differences between cABC and HSP27 mapped onto their structure 81
3.4 An AP_{II} locked dimer 82
3.5 The second dimer in the $cABC_{E117C}$ 83
3.6 cABC predominantly populates AP_{II}. 85

3.7 Quaternary dynamics of cABC and cABC$_{E117C}$ 86
3.8 Chaperone activity of mammalian sHSP core domains 88
3.9 cHSP27 is an ineffective chaperone 89
3.10 Surface properties of cABC$_{E17C}$ 90
3.11 Reduction of $A\beta_{1-42}$ cell toxicity by mammalian sHSP core domains 92
3.12 Hydrophobic patches on the structure of cABC, cABC$_{E117C}$, and cHSP27 94
3.13 Resonance assignment ambiguities cABC$_{E117C}$ 100

4.1 Measurements of the αB-crystallin ensemble 121
4.2 The kinetic model of αB-crystallin assembly and allosteric communication across interfaces 125
4.3 Allosteric effect of C-terminal binding to the ACD 130
4.4 Different registration states and flexibility of the ACD captured by crystallography 132
4.5 Dynamic movements of αB-crystallin in solution 136

5.1 HSP 18.1 and HSP 17.7 do not hetero-oligomerise 156
5.2 Quaternary architecture of HSP 18.1 and HSP 17.7 157
5.3 Validating the tetrahedral architecture for HSP 18.1 158
5.4 Quaternary dynamics of HSP 18.1 and HSP 17.7 160
5.5 Structure of the ACDs from HSP 17.7 and HSP 18.1 162
5.6 Estimating the free energy profile of dimer dissociation by umbrella sampling MD 164
5.7 Hetero-dimerisation depends on non-interfacial loops that affec . 166
5.8 Distinct conformations of the dissociated ACD monomers 169
5.9 Class-specific sequence conservation maps to differences in the dissociated monomer structures 171
5.10 Principal component analysis of chimeric ACDs 172

5.11 Binding of C-terminal peptides to ACDs reveals selectivity in c17.7 but not c18.1 173
5.12 Polydisperse assembly mediated by the C-terminus. 175
5.13 C-terminal selectivity prevents hetero-oligomers 176
5.14 The N-terminus acts as a non-specific bias towards 12mers . . . 178
5.15 Loop dynamics in the ACD dimers 181

List of Tables

3.1 Crystallographic table for the αB-crystallin and HSP27 core domains 96

5.1 Crystallographic table for the HSP 17.7 and HSP 18.1 ACDs . . 188

Chapter 1

Protein quaternary structure, dynamics and function in biology

1.1 Abstract

One of the central tenets of structural biology is that there is a direct relationship between the biological function of proteins and their 3-dimensional structure. This dogma has more recently been modified to reflect that the structure of proteins implicitly encodes the dynamic fluctuations they undergo at equilibrium, and that these fluctuations are often essential to their function. The small heat-shock proteins (sHSPs) are a family of molecular chaperones in which this link between structure, dynamics and function is thought to extend to the higher-order organization of individual protein subunits into complexes, or oligomers. sHSPs form large and dynamic assemblies that can inter-convert between different oligomeric and conformational states at equilibrium and these fuctuations are thought to be integral to their function. Understanding the structural mechanisms underpinning these dynamics, and how they relate to the function of sHSPs as molecular chaperones will help shed light on the general role of oligomeric organization and dynamics in biology.

1.2 Protein quaternary structure in biology

Protein structure is determined on four levels: the primary sequence of amino acids, their organization into secondary structure elements (helices, loops, and strands), the tertiary fold formed by the combination of secondary structure elements, and finally the arrangement of monomeric subunits through non-covalent interactions into a quaternary structure. The majority of proteins with known structures posses some degree of higher-order quaternary organization [1, 2], arranging their protomeric building blocks into either homomeric complexes containing only one type of subunit, or together with other proteins into heteromeric complexes. The sizes of the resulting structures can range from simple homomeric dimers [3, 4, 5] to gigantic assemblies the size of a eukaryotic organelle [6] (Fig. 1.1A). Although homomers represent the largest fraction of oligomers in databases of atomic structures, most proteins have several interaction partners in large-scale interactomes of both prokaryotes [7] and simple eukaroytes [8], at least some of which will represent stable hetero-oligomeric interactions [9]. Even though such estimates are not without their significant biases (especially if they are based on crystallographic data), all available data point towards an overwhelming abundance of oligomers in biology.

1.2.1 Functional role of protein oligomers

Homo-oligomerization is increasingly recognized as an important feature of protein regulation and signalling [10]. At the time of writing over 20% of proteins listed in the Brenda database of enzymes are annotated as oligomeric [11]. Homomeric interactions are more common in protein interaction networks than would be expected by chance [12], suggesting that there is a strong functional advantage in multimerization. Cellular functions that require large physical structures are often carried out by multimers, with viral capsids, clathrin cages and the nucelopore complex being examples. In such proteins, achieving large size through oligomerization rather than a single large polypeptide enables

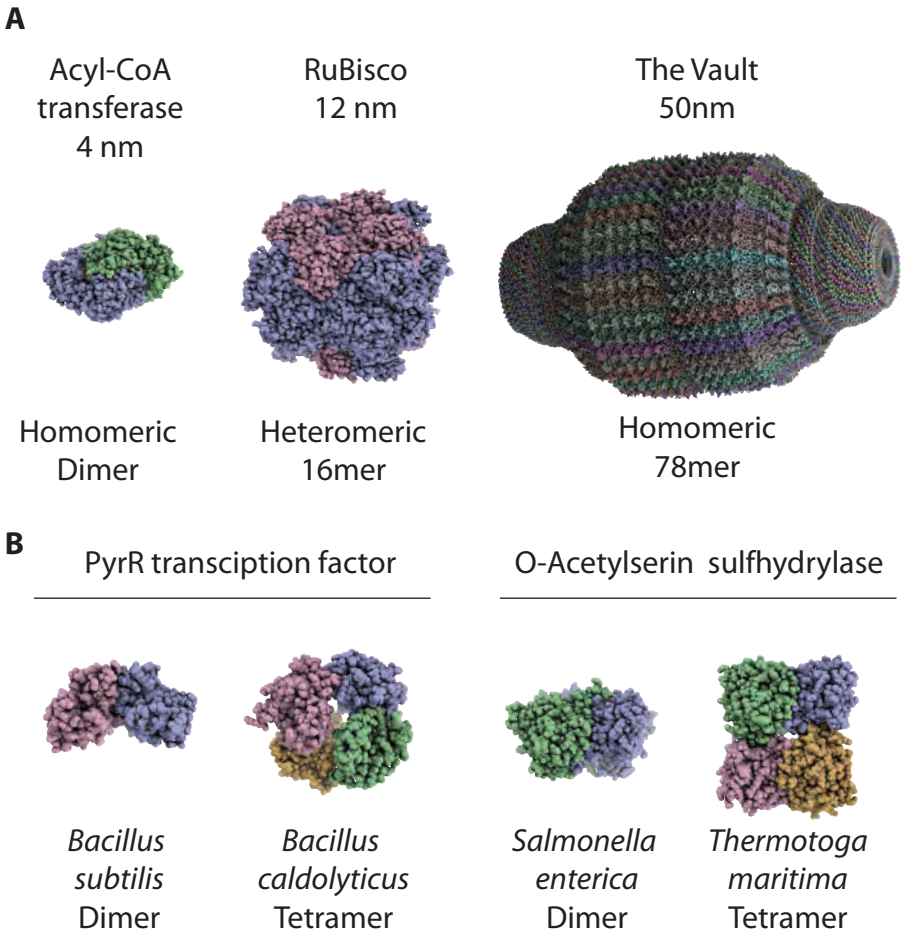


Figure 1.1: : Quaternary structure in proteins. A) Size range of quaternary architectures in proteins, ranging from homomeric dimers, as in the PLP-Dependent Acyl-CoA Transferase (left, PDB: 3HQT), to large hetermeric complexes like RuBisco (middle, PDB:1RLC) to huge assemblies like the eukaryotic vault (right, PDB: 4V60). Sizes indicate the maximum diameter as measured in the crystal structures. B) Stoichiometric variation in functionally identical proteins. Both the PyrR transcription factor and O-Acetylserin suldhydrylases can be found as either dimers or tetramers in different organisms.

higher folding and assembly efficiency, and is also a more efficient strategy to encode the structure genetically compared to a single very large protein [13]. Another functional advantage of an oligomeric architecture is homotropic allostery, in which the binding of a substrate to one subunit of an oligomer affects binding of substrate to the others (as opposed to heterotropic allostery, where binding of an effector at an allosteric site affects binding of a substrate at the active site) [13]. Allostery seems common in many protein-protein interactions

[14], with conformational change observed both at the oligomeric interfaces, and at parts of the complexes that are not forming connections between subunits.

In contrast to these well-studied examples of the advantages of oligomeric size, there are cases in which the situation is less clear. Often each subunit in an enzymatic oligomer has an active site, without clear signs of co-operativity between them. In such cases a quaternary assembly might simply act to increase the substrate encounter rate of the multimer by hiding parts of the protein that do not bind substrate on the inside of the complex [15]. In the rare cases where oligomeric proteins can be engineered to completely stop quaternary assembly, the monomeric subunits can sometimes still function [16] (also see chapter 3). Moreover, at least some protein families display variation in their stoichiometry despite conserved function (e.g. [17, 18, 19, 20] (Fig. 1.1B), with no clear trend towards either smaller or larger complexes being more efficient at their particular activity [21]. Although such data are still scarce, and do not cover any substantial phylogenetic breadth, they cast doubt over the functional significance of any one particular stoichiometry, even if oligomerization itself may be useful.

1.2.2 Dynamics of protein oligomers

Our understanding of quaternary protein dynamics still lags behind the study of dynamics at lower levels of protein organization. While, for example, significant inroads have been made over the last two decades in understanding the folding pathways of small protein monomers [22, 23, 24, 25], relatively little is known about the quaternary assembly process. Our current understanding relies almost exclusively on thermodynamic disassembly experiments, in which oligomeric complexes are disrupted in solution using organic solvents, pH, or extreme salt concentrations. In such experiments the resulting sub-complexes are analysed to infer the stoichiometry of kinetic assembly intermediates [26, 27]. Such experiments are likely biased towards what sub-complexes remain stable

under such harsh conditions, rather than true assembly intermediates. More direct measurements of the assembly process are, however, currently hampered by the technical difficulties involved in capturing transient and stoichiometrically heterogeneous interactions.

Once oligomers are formed, they can display complex dynamic behaviour on the quaternary level. This includes changing their stoichiometry in response to post-translational modification [28] or a change in solution conditions [29], continuously recycling their subunits via the process of subunit exchange [30], and even inter-converting between different oligomeric architectures at equilibrium [31]. All these dynamic changes are potentially linked to altered functional states of protein oligomers, but they are also the aspect of protein quaternary organisation that is the least well understood. In particular, the time-scales over which such behaviour takes place are largely unknown. At least some soluble proteins undergo seemingly rapid subunit exchange compared to typical protein half-lives [32, 33, 34], reaching equilibrium over a time-scale of minutes to hours, but for the vast majority of even well-studied complexes no information on subunit exchange rates is available. Similarly, subunit exchange in membrane proteins has been documented at least once [35], but seems extremely slow or absent in other cases [36].

The functional significance of such dynamics is currently still largely unclear, and it is tempting to speculate that they are biologically important. But establishing such a link between quaternary dynamics and function not only requires measuring such dynamics accurately (see section 1.3 and chapter 5 for discussions of the technical difficulties involved), but also their targeted manipulation without disturbing the protein structure in a major way (through e.g. truncation or covalent modification). Combining these two requirements is still painstakingly difficult for structural fluctuations on the seocndary and tertiary level, even in well-studied monomers [37], let alone large oligomeric proteins. Although significant progress has been made in the rational manipulation of stoichiometry [38, 39], essentially no tools are currently available to aid targeted

manipulation of quaternary dynamics, often leaving structural biologists without an unambiguous way to test their biological significance. These challenges have left protein quaternary dynamics comparatively unexplored, and place the field at the frontier of methodological development in structural biology.

1.3 Small heat-shock proteins: A model for quaternary structure and dynamics relationships.

sHSPs are a family of molecular chaperones whose canonical function it is to prevent other mis-folded proteins from forming non-functional and potentially harmful aggregates [40, 41]. They perform their function without binding or hydrolysis of ATP, and unlike the ATP-dependent chaperone machinery (GroEL/GroES, Hsp90/Hsp70) sHSPs do not re-fold their clients. Rather, they seem to prevent aggregation in its early stages or hold aggregates in a re-folding competent state until they transfer them to either other chaperones for re-folding or the proteasome for degradation [42]. It is not currently understood how they carry out their function on the molecular level, but it likely involves their quaternary architecture and the dynamic motions that sHSPs undergo on all levels of their protein organisation. While many other protein families, including most other chaperone families, are known to undergo fluctuations in side-chain conformations and tertiary organization as part of their function, sHSPs stand out in that they also display rapid and complex dynamics on the level of their oligomeric composition, with much evidence to suggest that these motions are integral to their activities *in vivo*.

1.3.1 sHSP oligomers and their building blocks

Most sHSPs, but certainly not all [44], form large oligomeric assemblies. Due to their inherently dynamic nature, sHSP oligomers are difficult to study by X-ray crystallography, and only a handful of structures have been solved of sHSP

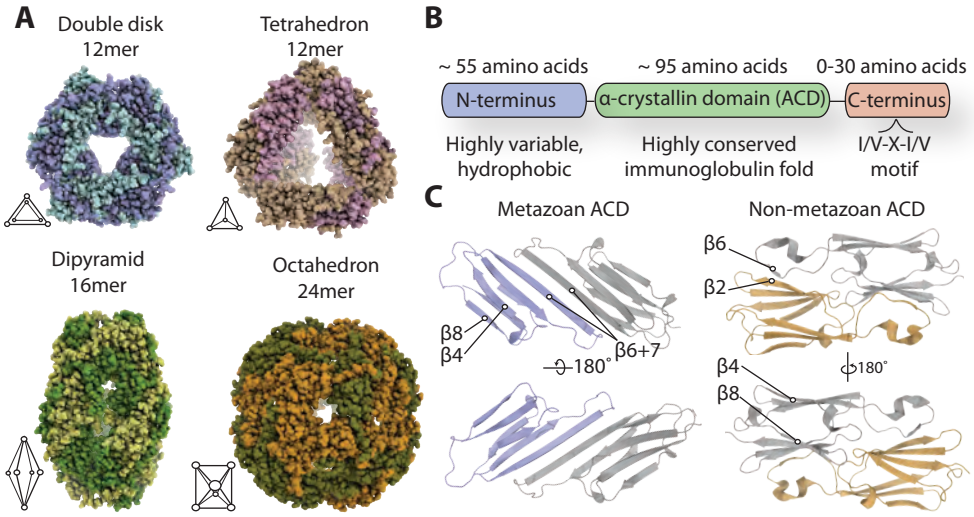


Figure 1.2: : Small heat-shock proteins A) Known quaternary architectures of sHSP oligomers. Top left: HSP 16.9 from *Triticum aestivum*. Top right: HSP 16.3 from *Mycobacterium tuberculosis* (EM reconstruction). Bottom left: HSP 16.0 from *Schizosaccharomyces pombe*. Bottom right: HSP 16.5 from *Methanococcus jannaschii*. Structures of chimeric sHSPs, such as [43], are omitted from this list. B) Primary sequence organization of sHSPs. C) Left: Structure of the non-metazoan ACD (PDB: 3GLA) from *Xanthomonas axonopodis* with the dimer interface between the β-6 and β-2 strand. Right: Metazoan ACD from human αB-crystallin (PDB: 2WJ7) with an isologous dimer interface along the fused β6+7 strand.

oligomers [45, 46, 47, 48]. Together with EM reconstructions [49, 50] these structures have revealed considerable variation in sHSP quaternary architecture as well as stoichiometry (Fig. 1.2A). All sHSP oligomers studied to date form polyhedral or disk-like oligomers, and the diversity of structures that have been observed stem from the inherent flexibility in different parts of its monomeric building blocks.

sHSPs are identified through the presence of a α-crystallin domain (Fig. 1.2B), a region of sequence rich in β-structure that adopts an immunoglobulin-type fold [51]. In general, the α-crystallin domain is found to dimerize, forming the basic building block of higher-order sHSP structure, in one of two ways [52, 53]. In the structures determined for mammalian sHSPs, extended β6+7 strands on each monomer pair-up in an anti-parallel orientation (Fig. 1.2C, left). Contrastingly, it appears that in plants and lower organisms the

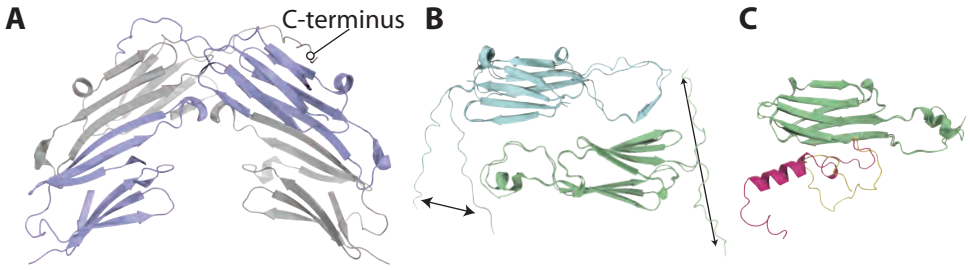


Figure 1.3: : Terminal interactions of sHSPs A) C-termini act to cross-link dimers into higher order structures, as shown for human αB-crystallin (PB: 3L1G). B) Within the same structure, C-termini can adopt very different orientations. Shown are alignments of sHSP monomers within the same structure, left from *Triticum aestivum* (PDB: 1GME), right from *Schizosaccharomyces pombe* (PDB: 3W1Z). D) Similar to C-terminus, the N terminus can adopt different conformations within the same oligomer. N-termini from two aligned monomers (also from *S. pombe*) are coloured in pink and yellow.

sHSP dimers are formed through reciprocal donation of the β6 strand, located in a loop, into the β-sandwich of the partner (Fig. 1.2C, right). While the latter dimerization mode has only been observed in one form, in studies on mammalian sHSPs three distinct states have been observed, brought about by shifts in register between the paired β6+7 strands [54] (see chapter 3 and 4 for a detailed discussion). These polymorphs, termed AP_I, AP_{II} and AP_{III} in order of decreasing overlap of anti-parallel β6+7 strands [55], have to date only been observed directly in X-ray structures obtained from truncated constructs. While it is not yet clear whether mammalian sHSPs can ratchet along the dimer interface in solution, there is evidence that the dimer interface is relatively weak and dynamic, with a dissociation constant on the order of a few micromolar [56] (see also chapter 2).

Typically, the α-crystallin domain is flanked by N-terminal and C-terminal regions, the lengths of which vary considerably across the sHSP family, and can contain further subdivisions (Fig. 1.2B). The C-terminal region is typically relatively short, and acts to bridge between dimers, such that a conserved I/V-X-I/V motif may bind into a groove between strands β4 and β8 of a neighbouring monomer (Fig. 1.3A). This cross-linking is enabled by a flexible C-terminal re-

gion, that appears to have complete freedom in the trajectory it takes towards its binding partner [53]. As a result, even within the same oligomer, monomers can adopt strikingly different C-terminal orientations [48, 46] (Fig. 1.3B).

The N-terminal region is the major source of variation in sHSP sequences across nature [57]. Unfortunately, structural and dynamical information on this part of the proteins is currently very sparse. The limited crystallographic data that does exist comes from sHSPs with a comparatively short N-terminal region, and is consistent with it being located within the central portion of the oligomer [48, 46]. Notably, the data also point towards structural heterogeneity, as not all monomers display the same orientations within the oligomers (Fig. 1.3C). Similarly, molecular dynamics investigations suggest that the N-termini can explore multiple different conformations [58, 59]. Nonetheless, the majority of the structural data that exists for the N-terminus points towards at least part of this region being involved in making inter-monomer contacts, and therefore contributing to the assembly of the oligomers. Complementing this data, hydrogen/deuterium exchange experiments reveal that N-terminal regions tend to be more solvent accessible than the majority of the sequence, suggesting a significant presence of dynamics at equilibrium, and revealing that these interfaces are relatively labile [60].

1.3.2 Quaternary dynamics of sHSP oligomers

The sHSPs therefore contain a number of inter-subunit contacts, including one that serves to construct the protomeric dimer. The collective behaviour of these interfaces determines the quaternary dynamics of sHSP oligomers. For example, all sHSP oligomers investigated to date exchange subunits at temperatures close to ambient for the organism they derive from [30, 61, 62, 63]. During subunit exchange in sHSPs, either monomers (in metazoa) [61, 62, 64] or monomers and dimers (in plant sHSPs) [30, 65] transiently dissociate from sHSP oligomers, only to rapidly associate back onto an oligomer shortly afterwards. The rate of exchange is a direct measure of the lability of the sHSP

assembly involved, while the proportion of the process associated with the movement of dimers and monomers provides information as to the relative ease of breaking the inter- and intra-dimer interfaces. Such quaternary dynamics, i.e. a fluctuation in the subunits comprising the assembled oligomeric forms, while not limited to the sHSPs, have been reported for relatively few other proteins. sHSPs that are expressed in the same compartment can also often exchange subunits amongst each other [30, 63, 66] resulting in an ensemble of hetero-oligomers with varying numbers of subunits from both proteins. One particularly notorious aspect of sHSP quaternary organization is that especially mammalian sHSPs often form polydisperse ensembles, populating several stoichiometries at once. αB-crystallin, an abundant human sHSP, populates stoichiomteries that range from 10 to just under 50 subunits per oligomer at equilibrium [67, 68]. Other sHSPs are monodisperse at ambient temperature, but become polydisperse at heat-shock temperatures [69].

The functional relevance of such dynamics might be that a hyper-variable binding surface on sHSP oligomers allows the fine-tuned binding of a large variety of clients [69], although this hypothesis remains to be proven. Polydisperse sHSPs that constitute major components of vertebrate eye-lenses have another, perhaps more obvious use for heterogeneous and dynamic assemblies. These proteins are present in the eye-lens at exceptionally high concentrations and must resist crystallization for nearly the whole of the organism's life-span [70]. Although this goes some way in explaining polydispersity, particularly in mammalian sHSPs, several non-lenticular sHSPs also assemble into similarly heterogeneous ensembles. As such, it remains to be seen whether the plasticity afforded by the variable sHSP quaternary structure is of general utility, in terms of binding targets, remodelling into stable sHSP:target complexes, both, or neither.

1.3.3 Regulation of sHSP quaternary dynamics

sHSP quaternary dynamics are highly sensitive to environmental conditions, a property thought to regulate their chaperone activity in response to various stresses. The temperature dependence of subunit exchange rate constants can be particularly acute [71], or involve multiple pathways [29, 72]. Moreover, other changes in solution conditions, such as a drop in pH, also appear to cause a similar cascade in the dynamics [73, 74]. That different modes of activation appear to function via a similar pathway to changes caused by heating, strongly provides evidence for the thermally induced changes observed being functional. This conclusion is perhaps unsurprising, considering that the sHSPs are, in the main, called into action under conditions of cellular thermal stress.

Dynamical changes caused by variations in primary structure have also been reported. For example, the R120G mutation of αB-crystallin is associated with chaperone hyper-activity [75], and exchanges subunits more rapidly than the wild type [76]. Many post-translational modifications (PTMs) of mammalian sHSPs have been observed, though not so in plants and lower organisms. While some of those reported for the α-crystallins are likely just an inevitable consequence of their longevity in the eye lens, other PTMs appear to be specific regulators of activity. Chief amongst these are phosphorylation, which is prevalent in several mammalian sHSPs, and in particular HSP27 and αB-crystallin. On each of these two proteins there are three major phosphosites that fall within the N-terminal portion of the protein. HSP27 undergoes dissociation into sub-olgomeric species upon phosphorylation [77, 78, 79], and, while more subtle, a destabilisation of the inter-subunit interfaces is observed for αB-crystallin [80, 81]. Post-translational modification, and phosphorylation in particular, is likely to play a key role in controlling sHSP activity in homeotherms. In other organisms, such as bacteria and plants, environmental temperature changes may be sufficient to regulate activity, taking advantage of the finely tuned quaternary dynamics of sHSPs.

The sHSPs emerge as a family of proteins in which the coupling between tertiary structure, quaternary dynamics and biological function seems particularly acute. This makes them an ideal system to understand how tertiary structure fluctuations propagate to the quaternary level [69], and how quaternary dynamics relate to biological function. In doing so care must be taken not to automatically interpret complex dynamical behaviour as biologically important: not all features of biological organisms and the molecules they are made of are necessarily adaptive [82, 83, 84]. With that caveat in mind, the sheer variety of quaternary architectures and dynamics in the sHSP family makes it a unique opportunity to study this still largely unexplored aspect of biochemistry.

1.4 Biophysical techniques for the study of protein quaternary assemblies

Studying the quaternary dynamics of protein complexes in general and sHSPs specifically still represents a formidable technical challenge. I will now give a brief description of the biophysical techniques used to study quaternary protein structure and dynamics, along with their particular advantages pertaining to the study of quaternary architecture.

1.4.1 X-ray crystallography

Most of our knowledge about protein quaternary structure derives from X-ray crystallography. X-ray crystallography is a diffraction based technique that, if successful, reveals the atomic structure of a protein. The principle of X-ray diffraction is simple: to produce diffraction, an object is exposed to light of a wavelength comparable to the object's size. Mathematically, diffraction produces the Fourier (or molecular) transform of the illuminated structure. The structure of the object can then be inferred by inverting the Fourier transform, either through use of a conventional lens, or by mathematical means.

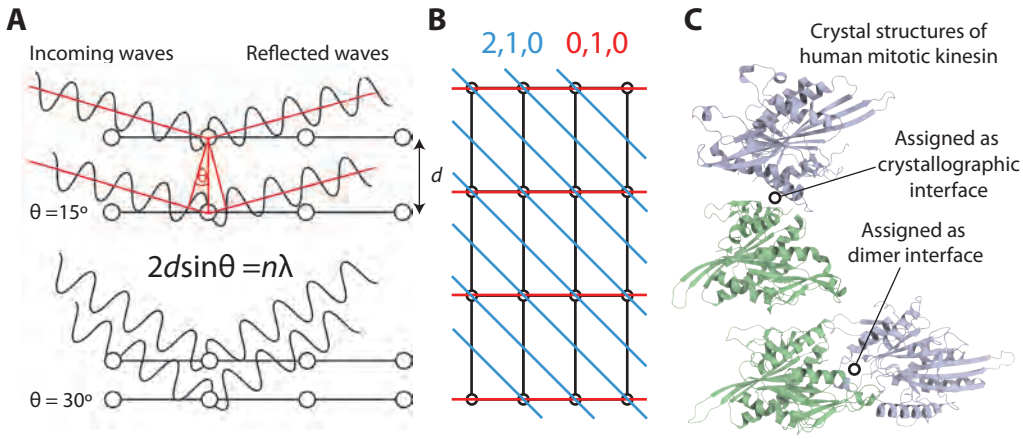


Figure 1.4: :X-ray crystallography A) Bragg's law. Atoms on two crystallographic planes will scatter in-phase if the additional length travelled by the second wave is an integer multiple of the wavelength. Shorter distances d between the planes require larger angles of incidence for Bragg's law to be fulfilled. B) Two sets of crystallographic planes drawn through a regular lattice of points. The Miller index (blue and red indices) denotes the number of unit cells a plane has to transverse before encountering another lattice point (z-dimension not shown). C) Ambiguous stoichiometry in crystal structures. Two crystal structures of the same component of human kinesin. In the left structure (PDB: 1QOB) the interface was correctly identified by the authors as crystallographic. The authors of the right structure (PDB: 1II6) assigned the structure of a virtually identical molecule as dimeric based on different crystal packing. Both proteins are assigned as monomeric in the PiQsi database.

Reconstruction of atomic structures from X-ray diffraction data

In order to even produce a measurable diffraction pattern, the signal from a single molecule must first somehow be amplified. In practice this is achieved by packing many identical copies of the molecule into a regular lattice, or crystal, in which the repeating unit is called the "unit cell". When a well-ordered crystal is exposed to X-rays of a suitable wavelength, the resulting diffraction pattern is not the full molecular transform, but rather a series of spots representing regularly spaced samples of the transform. This happens because a crystal amplifies the signal of the molecular transform at discrete points, owing to constructive interference of emitted light waves in only certain parts of the molecular transform.

Bragg's law describes the conditions necessary for constructive interference

in crystals [85]. It relies on the construction of sets of parallel planes that intersect the unit cell of the crystal at regular intervals, each plane being separated from the next by the distance d. Constructive interference occurs only if the additional length travelled by the second wave is a multiple of its wavelength. The additional length travelled by the second wave depends on the interplanar spacing d, and the angle of incidence, θ. The larger the d spacing, the larger the angle needs to be in order to generate a reflection (Fig. 1.4A). A Bragg peak generated in this way contains information about all components (every atom) present in the average unit cell. Each reflection is associated with a Miller index (Fig. 1.4B) that describes the orientation and spacing of the particular set of planes the reflection derives from relative to the unit cell [86].

To infer the atomic structure of a molecule from its X-ray diffraction pattern, the intensities from as many Bragg peaks as possible combined with their phases and Miller indices have to be measured. Together these three quantities (intensity, Miller index, and phase) describe a wave in Fourier space, whose peaks and troughs in real space correspond to areas of high and low electron density in the unit cell. An electron density map is generated by first summing all these waves in Fourier space, and then inverting the Fourier transform.

Phase information is not recorded in a typical crystallography experiment; the only quantities obtained are the amplitude (related to the intensity of a Bragg spot) and the frequency of the wave. The phase of each reflection can either be estimated using a known structure of a very similar protein (in a process called molecular replacement [87]), or measured experimentally by introducing heavy atoms into the protein crystal that perturb the diffraction signal in a predictable way. Such heavy atoms (Selenium, iodide, or heavy metals) lead to characteristic anomalous scattering in the diffraction experiment, that can be used to infer phases for each reflection directly [88, 89].

Quaternary structure in crystal structures

Fundamentally, the solvent content of a typical protein crystal (around 50% v/v, [90]) is not dissimilar to the water content of the cytoplasm (about 80%), such that one might reasonably expect protein quaternary structure to be preserved in the crystal phase. But because molecules have to be packed into crystals to produce measurable diffraction patterns, it is not always easy to determine which protein-protein contacts in a crystal structure represent quaternary contacts present in solution, and which are simply crystal packing contacts (Fig. 1.4C). Still the question of reliably finding oligomeric interfaces in crystal structures has attracted considerable computational attention, with approaches using the difference in surface-accessible area between a monomer within and outside of the oligomer [91], Voronoi cells [92], graph theory [93], and still in many cases requires manual intervention. Even when assigned manually, stoichiometries derived from crystal structures can be erroneous (Fig. 1.4C). An additional difficulty with determining the stoichiometry of protein complexes from crystal structures is the narrow and often unphysiological temperature range in which crystals are usually produced (between 4 and 25 °C), which can become problematic if the quaternary structure under study is particularly sensitive to temperature. On the other hand, X-ray structures are the most reliable source of information about subunit connectivities and the atomic contacts that make up interfaces. For that reason, the reliability of most other techniques for the measurement of stoichiometry is still evaluated based on data derived from crystal structures.

1.4.2 Small angle X-ray scattering

Small angle X-ray scattering (SAXS) is an emerging technique for the study of protein quaternary structure that, unlike X-ray crystallography, obtains structural information directly from molecules in dilute solution. X-rays are scattered elastically off proteins in the solution, and their intensity is recorded as a

function of scattering angle. The resulting intensity profile is usually expressed as a function of momentum transfer q given by $q = 4\pi sin\theta/\lambda$, where θ is the angle of incidence and λ is the wavelength [94]. SAXS data can be useful in several ways for the determination of quaternary architecture. First, the radius of gyration (R_g) and molecular mass can be calculated from a SAXS profile [95]. In the case of mixtures of oligomeric species in solution, SAXS can give fairly accurate estimates of their relative abundances when structures are known for all species involved [96]. When structures are not available, it can at least be used to determine which stoichiometry is most abundant [96].

In addition to providing stoichiometric information, SAXS data can be used to interrogate the higher-order structure of protein assemblies. Theoretical SAXS curves generated from low-resolution "bead-models" can be fitted iteratively to SAXS profiles to gain shape information up to a resolution of about 15 Å [96], although reconstructing a 3D structure from a 1D measurement means that usually many different models can produce similar SAXS profiles. This ambiguity in model interpretation is balanced by the very simple sample requirements (pure protein at 5mg/ml concentration) and fast through-put of a SAXS experiment, meaning that the technique is poised to become more important in the study of protein quaternary structure and dynamics [95].

1.4.3 Native mass spectrometry

Although still most frequently used for proteomics and small-molecule identification, over the last 20 years mass spectrometry (MS) has emerged as an invaluable technique to study protein quaternary structure and dynamics [97, 98, 99]. Drastic improvements in instrument design and operation have made it possible to ionize intact protein complexes, maintain their non-covalent interactions in the gas phase and measure their intact mass at high resolution [100]. So-called "native mass spectrometry" has since become an attractive tool for the study of quaternary dynamics [98]. It combines high mass resolution with very high

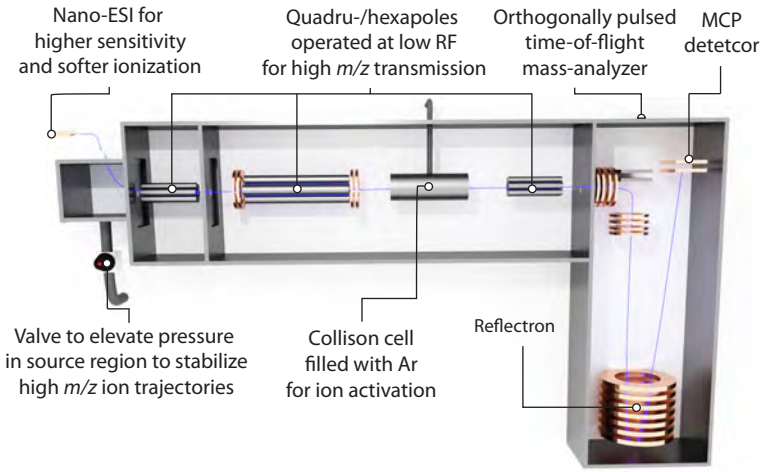


Figure 1.5: : The Q-TOF instrument as used in this thesis. Ions are produced via nanoESI, and then transmitted through an instrument consisting of a focusing hexapole, a resolving quadrupole, a collision cell, another focusing hexapole and finally a time-of-flight mass analyser fitted with a reflectron. Ions are detected on a microchannel plate (MCP) detector. Modifications for high mass transmission and analysis are indicated.

through-put, allowing the study of stoichiometrically heterogeneous protein assemblies [101, 102] and measuring quaternary dynamics of protein complexes in real time [101, 103].

Instrument configuration used in this thesis: the Q-TOF

In a typical MS experiment ions are first produced, then focused and transmitted to a mass analyser that separates them according to their mass-to-charge ratio (m/z), and subsequently detected in a way that preserves the ions' m/z separation. A typical Q-TOF instrument used for native MS includes those same steps, but with some crucial modifications that facilitate the ionization, transmission, and detection of large protein complexes [104] (Fig. 1.5).

Nano electro-spray ionization

In native MS ions, are typically produced using a nano electro-spray ionization (nanoESI) source. As in conventional ESI, ions are produced by applying a high voltage across a capillary filled with analyte solution. At the tip of the

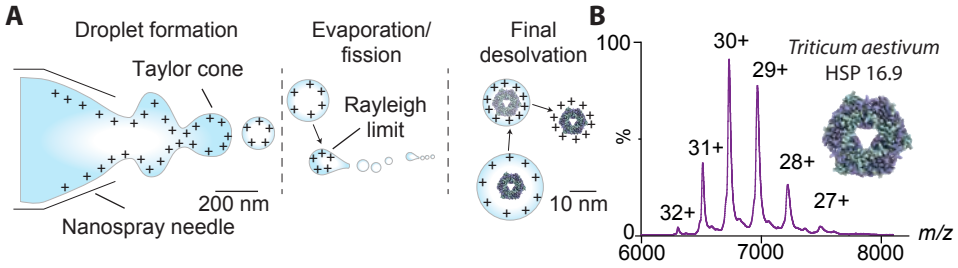


Figure 1.6: : Ionization of protein complexesby nanoESI ionization. A) Left: A high voltage applied across the nanoESI capillary leads to the formation of a Taylor cone at its tip from which charged droplets are expelled. Middle: Droplets undergo repeated cycles of evaporation, and, when the Rayleigh limit is reached, asymmetric droplet fission producing ever-smaller droplets. Right: The final droplet only contains one analyte after many evaporation/fission cycles. The residual solvent evaporates and a macromolecular ion is produced. B) A typical native mass spectrum of a dodecameric small heat-shock protein from wheat showing how nanoESI produces multiple charged ions. Each peak corresponds to a intact 12mer carrying a different number of positive charges.

capillary, charged droplets are expelled from a Taylor cone (Fig. 1.6A, left). The droplets subsequently shrink due to solvent evaporation, leading to an increase in the charge density on the droplet surface. The so-called Rayleigh limit is reached when the coulombic repulsion between the surface charges overcomes the surface tension of the droplet (Fig. 1.6A, middle). At this point asymmetric droplet fission occurs, producing a series of smaller drops that carry away the excess charge together with any solutes. For analytes the size of proteins, this process is repeated several times until only one analyte is left per droplet. The remaining solvent evaporates and the residual charges are transferred to the analyte, producing a macromolecular ion (Fig. 1.6A, right). The stochastic nature of this process produces ions that carry variable numbers of charges, resulting in a characteristic ion series for each single stoichiometry in a mass spectrum that can be used to determine the mass of analyte ions from their m/z ratio [105, 106] (Fig. 1.6B).

Compared to conventional ESI, nanoESI capillaries have a much smaller internal diameter than conventional ESI capillaries (1-2 μm compared to 25-50 μm). They consequently produce much smaller charged droplets with a higher

charge density. This has several advatages: to affect complete solvent evaporation, fewer droplet evaporation/fission cycles are needed, negating the need for the high interface temperatures and organic solvents used in conventional ESI to ensure fast enough solvent evaporation. This allows the use of aqueous solvents amenable to maintaining non-covalent interactions in solution prior to ionization. Also, any non-volatile salts present in solution are concentrated during droplet evaporation, leading to undesirable salt-adducts in mass spectra. Smaller initial droplets and fewer evaporation/fission cycles in nanoESI diminish this effect compared to ESI. Finally, the more efficient ionization during nanoESI increases sensitivity, and consequently allows for much lower flow rates (nL sec^{-1} compared to μL sec^{-1}).

Transmission of large ions

During ionization ions are accelerated in the droplet stream to velocities of several hundred meters per second regardless of their size [107]. The very large ions produced in a native MS experiment correspondingly possess substantially more kinetic energy than smaller ions produced in a conventional MS experiment. This excess kinetic energy makes it difficult to focus them on stable trajectories using conventional ion optics. For this reason instruments modified to transmit very large ions have elevated pressure in the initial stages of the vacuum [107]. Collisions of the ions and the buffer gas act to dissipate their kinetic energies into internal modes of vibration, leading to a dramatic increase in the transmission of high-mass ions in the first stages of the experiment.

Mass analyzers for native MS

Mass analysers are used in tandem in most modern native MS configurations: one low-resolution mass-analyser used for ion filtering prior to analysis, followed by a high-resolution mass-analyser. This design allows performing gas-phase experiments on a selected subset of ions. In Q-TOFs, quadrupoles are used

for ion filtering. A quadrupole consists of 4 parallel conductive rods. A radio-frequency (RF) time-dependent (AC) potential is applied between each of the two opposing pairs of rods, and overlaid with a time-independent DC potential. One pair of rods has a positive DC potential applied, the other a negative DC potential, and their AC potentials are 180° out of phase. Depending on the polarity, the AC potential of one pair of rods acts as a low-pass m/z filter, whereas the DC potentitial in the other pair acts as a high-pass m/z filter. In resolving mode, the AC and DC voltages are set such that only ions within a narrow window a few m/z wide have stable trajectories. In scanning mode the resolving window is moved linearly between different values during one scan, leading to the transmission of a larger window of m/z values. In RF-only mode the quadrupole effectively transmits all ions. The maximum mass that can be transmitted depends on the RF frequency, such that high mass instruments operate their quadrupoles at lower RFs than conventional quadrupoles (about 300 MHz compared to 700-1200 MHz) [108, 104].

For the final m/z separation, a time-of-flight (TOF) analyser is used in Q-TOFs because of its theoretically unlimited mass range [104]. In a modern TOF analyser, ions are collected in the pusher region, and then pulsed orthogonally into the analyser (Fig. 1.5). This orthogonal design allows using a TOF analyzer with a continuous ion beam as is produced by nanoESI [109, 110]. Ions are pulsed into a flight tube towards the detector, where the flight time can be related to their m/z. To correct for variation in the kinetic energy of ions of identical m/z, a reflectron (ion mirror) is placed at the bottom of the flight tube with a linear potential applied across it. Ions with higher kinetic energies penetrate deeper into the reflectron than ions with lower kinetic energies, resulting in much closer arrival times at the detector and correspondingly improved resolution in the resulting spectrum.

Gas-phase manipulation of ions

One of the most valuable tools for native mass spectrometrists is the ability to selectively dissociate subunits from a protein complex in the gas-phase. Typically this is done by accelerating ions into a collision cell, located in a Q-TOF between the quadrupole and the TOF analysers, that is filled with an inert gas (typically Argon, but Xenon and SF_6 are also used). During collision-induced-dissociation (CID), ions collide with the buffer gas, converting some of their kinetic energy into internal modes of vibration. This gradual increase in internal energy leads to the unfolding of usually a single subunit [111] (Fig. 1.7A). The increased surface area available on the unfolded subunit can accommodate charge from the protein complex, minimizing its overall coulombic repulsion. In a mutually reinforcing process, the excess charge acts to further unfold the subunit in a coulombically driven manner, which in turn attracts even more charge, eventually leading to the ejection of a highly charged monomer. In a mass spectrum, this process produces a characteristic series of peaks at low m/z corresponding to the highly charged ejected monomers, and another series at higher m/z than the original ion corresponding to the charge-stripped oligomer missing one subunit (Fig. 1.7B). Depending on the size and charge of the initial complex, and the available collision energy, this process can be repeated several times with the same ion.

CID can aid interpretation of a complex mass spectrum in several ways: first, collisional heating acts to strip residual salt and buffer adducts from the complex before a subunit is dissociated, resulting in better resolved peaks. Second, the masses from ejected monomers can be used to confirm the subunit composition of a protein complex, and, when used together with a quadrupole in resolving mode, to identify which subunits derive from which peaks in spectra that contain mixtures of proteins. Third, the lower average charge on the stripped oligomers acts to increase the spacing between individual peaks in an ion series. This increased peak spacing often allows the resolution of ion

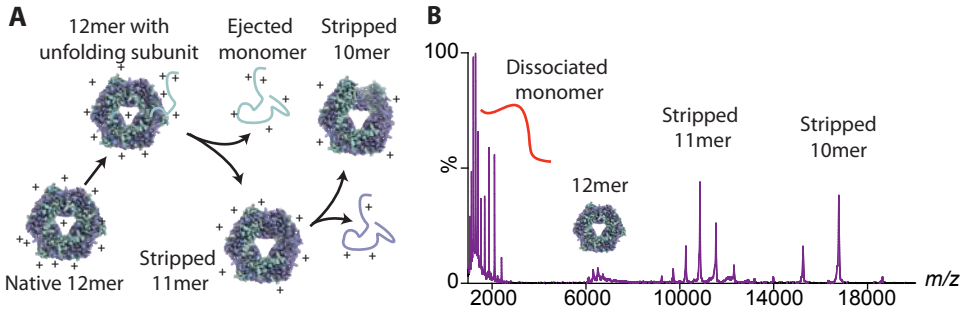


Figure 1.7: : Collision-induced-dissociation of protein complexes. A) Subunit ejection and charge reduction during CID. Collisions with a neutral gas gradually raise the internal energy of the complex. A single subunit unfolds in a coulombically driven process and attracts nearly half the charge on the original complex, before it is finally ejected as a highly charged monomer. It leaves behind a charge-stripped oligomer missing one subunit that can subsequently undergo another round of CID. B) A high-energy CID spectrum of HSP 16.9. Unfolded, ejected monomers appear at low m/z, whereas stripped oligomers shift to higher m/z relative to their parent ions.

series that overlap in the native oligomers when more than one stoichiometry is present in solution [112].

Ion mobility spectrometry

A relatively recent, but extremely powerful addition to the native MS tool-kit is ion mobility spectrometry (IMS). In IMS, ions are first bunched and then injected simultaneously into an ion mobility cell that is filled with a neutral buffer gas (Fig. 1.8A). In the simplest design (called a drift tube), a linear, low strength electric field is applied across a stack of ring electrodes along the length of the mobility cell [113]. The ions' movement along this field is opposed by collisions with the buffer gas which act to slow them down. Ions with larger collisional cross-sections (CCSs, in units of $Å^2$) experience more collisions, resulting in later arrival times at the detector for ions with larger CCS-to-z ratios (Fig. 1.8B). What is actually measured experimentally is the time each ion takes to transverse the drift tube (the arrival time). The CCS of a protein complex can be calculated directly from its arrival time in linear drift tube instruments using the Mason-Schempp equation, which relates the

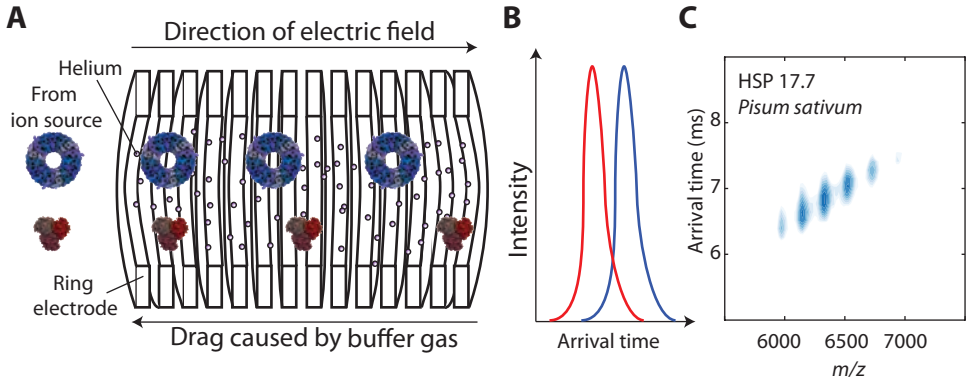


Figure 1.8: : Ion-mobility spectrometry of protein complexes A) Ions are first bunched and then injected into a mobility cell filled with a neutral buffer gas (usually helium). Collisions with the buffer gas oppose their movement along a weak electric field across the mobility cell. Assuming equal charge, the blue molecule has a larger CCS than the red protein, causing it to experience more collisions and moving more slowly through the mobility cell. B) Their different mobilities result different arrival times in a IMS experiment. C) IM-MS spectrum of HSP 17.7 from *Pisum sativum*. In IM-MS each ion has its m/z and arrival time measured. Ions with higher charge (lower m/z) have earlier arrival times than ions of the same mass with lower charge (higher m/z), because IMS separates according to the CCS-to-z ratio.

measured arrival time to the CCS based on kinetic theory [114]. There is currently no analytical solution for the relationship between arrival time and CCS in more complex experimental designs than linear drift tubes. When such devices are used to measure CCSs, they must be calibrated using standard protein complexes that have had their CCS measured using a linear drift tube [113]. Physically, the CCS measured by IM is related to the Stokes radius in solution, except that the CCS does not include the hydration shell of the protein under investigation [115]. A theoretical CCS can be calculated from atomic structures using a variety of approximations, the simplest (and fastest) of which computes the CCS as the rotationally averaged projected area of the structure [116].

Ion mobility can be coupled to a conventional MS experiment by analysing ions emerging from the drift cell in successive TOF experiments, thereby adding a second dimension of separation to the measurement. In such experiments each ion has its arrival time and m/z measured simultaneously (Fig. 1.8C).

The additional CCS dimension can be used resolve charge state series belonging to different analytes that are unresolvable in the m/z dimension. The high resolution in both the mobility and mass dimension makes this approach particularly attractive fro heterogeneous protein assemblies, where both different stoichiometries, and different architectures corresponding to the same stoichiometry can potentially be resolved. Alternatively, CCSs from IM-MS experiments can be used to distinguish between models of multiple plausible architectures for any given stoichiometry based on their theoretical CCSs [117, 31], and can be integrated with proteomics, chemical cross-linking MS and native MS to construct macromolecular models of protein complexes that are not amenable to crystallography [118].

1.5 References

[1] E. D. Levy, J. B. Pereira-Leal, C. Chothia, and S. A. Teichmann, "3D complex: a structural classification of protein complexes.," *PLoS Comput. Biol.*, vol. 2, Nov. 2006.

[2] K. Henrick and J. M. Thornton, "PQS: a protein quaternary structure file server.," *Trends Biochem. Sci.*, vol. 23, pp. 358–361, Sept. 1998.

[3] R. C. Kelly, M. E. Bolitho, D. A. Higgins, W. Lu, W.-L. L. Ng, P. D. Jeffrey, J. D. Rabinowitz, M. F. Semmelhack, F. M. Hughson, and B. L. Bassler, "The *Vibrio cholerae* quorum-sensing autoinducer CAI-1: analysis of the biosynthetic enzyme CqsA.," *Nat. Chem. Biol.*, vol. 5, pp. 891–895, Dec. 2009.

[4] I. García-Sáez, A. Párraga, M. F. Phillips, T. J. Mantle, and M. Coll, "Molecular structure at 1.8 Å of mouse liver class pi glutathione S-transferase complexed with S-(p-nitrobenzyl)glutathione and other inhibitors.," *J. Mol. Biol.*, vol. 237, pp. 298–314, Apr. 1994.

[5] J. A. Tainer, E. D. Getzoff, K. M. Beem, J. S. Richardson, and D. C. Richardson, "Determination and analysis of the 2 Å-structure of copper, zinc superoxide dismutase.," *J. Mol. Biol.*, vol. 160, pp. 181–217, Sept. 1982.

[6] H. Tanaka, K. Kato, E. Yamashita, T. Sumizawa, Y. Zhou, M. Yao, K. Iwasaki, M. Yoshimura, and T. Tsukihara, "The structure of rat liver vault at 3.5 Ångstrom resolution.," *Science*, vol. 323, pp. 384–388, Jan. 2009.

[7] S. V. Rajagopala, P. Sikorski, A. Kumar, R. Mosca, J. Vlasblom, R. Arnold, J. Franca-Koh, S. B. Pakala, S. Phanse, A. Ceol, R. Hauser, G. Siszler, S. Wuchty, A. Emili, M. Babu, P. Aloy, R. Pieper, and

P. Uetz, "The binary protein-protein interaction landscape of *Escherichia coli*," *Nat. Biotech.*, vol. 32, pp. 285–290, Mar. 2014.

[8] P. Uetz, L. Giot, G. Cagney, T. A. Mansfield, R. S. Judson, J. R. Knight, D. Lockshon, V. Narayan, M. Srinivasan, P. Pochart, A. Qureshi-Emili, Y. Li, B. Godwin, D. Conover, T. Kalbfleisch, G. Vijayadamodar, M. Yang, M. Johnston, S. Fields, and J. M. Rothberg, "A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*," *Nature*, vol. 403, pp. 623–627, Feb. 2000.

[9] T. Perica, J. A. Marsh, F. L. Sousa, E. Natan, L. J. Colwell, S. E. Ahnert, and S. A. Teichmann, "The emergence of protein complexes: quaternary structure, dynamics and allostery. Colworth Medal Lecture.," *Biochem. Soc. Trans.*, vol. 40, pp. 475–491, June 2012.

[10] K. Hashimoto, H. Nishi, S. Bryant, and A. R. Panchenko, "Caught in self-interaction: evolutionary and functional mechanisms of protein homooligomerization.," *Phys. Biol.*, vol. 8, pp. 035007+, June 2011.

[11] I. Schomburg, A. Chang, S. Placzek, C. Söhngen, M. Rother, M. Lang, C. Munaretto, S. Ulas, M. Stelzer, A. Grote, M. Scheer, and D. Schomburg, "BRENDA in 2013: integrated reactions, kinetic data, enzyme function data, improved disease classification: new options and contents in BRENDA," *Nucleic Acids Res.*, vol. 41, pp. D764–D772, Jan. 2013.

[12] I. Ispolatov, A. Yuryev, I. Mazo, and S. Maslov, "Binding properties and evolution of homodimers in protein-protein interaction networks.," *Nucleic Acids Res.*, vol. 33, no. 11, pp. 3629–3635, 2005.

[13] D. S. Goodsell and A. J. Olson, "Structural symmetry and protein function," *Annu. Rev. Biophys. Biomol. Struct.*, vol. 29, no. 1, pp. 105–153, 2000.

[14] L. S. Swapna, S. Mahajan, A. G. de Brevern, and N. Srinivasan, "Comparison of tertiary structures of proteins in protein-protein complexes with unbound forms suggests prevalence of allostery in signalling proteins.," *BMC Struct. Biol.*, vol. 12, pp. 6+, May 2012.

[15] M. Lynch, "Evolutionary diversification of the multimeric states of proteins.," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 110, pp. E2821–E2828, July 2013.

[16] J. L. Robertson, L. Kolmakova-Partensky, and C. Miller, "Design, function and structure of a monomeric ClC transporter.," *Nature*, vol. 468, pp. 844–847, Dec. 2010.

[17] A. Heine, J. M. Canaves, F. von Delft, L. S. Brinen, X. Dai, A. M. Deacon, M. A. Elsliger, S. Eshaghi, R. Floyd, A. Godzik, C. Grittini, S. K. Grzechnik, C. Guda, L. Jaroszewski, C. Karlak, H. E. Klock, E. Koesema, J. S. Kovarik, A. Kreusch, P. Kuhn, S. A. Lesley, D. McMullan, T. M. McPhillips, M. A. Miller, M. D. Miller, A. Morse, K. Moy, J. Ouyang, R. Page, A. Robb, K. Rodrigues, R. Schwarzenbacher, T. L. Selby, G. Spraggon, R. C. Stevens, H. van den Bedem, J. Velasquez, J. Vincent, X. Wang, B. West, G. Wolf, K. O. Hodgson, J. Wooley, and I. A. Wilson, "Crystal structure of O-acetylserine sulfhydrylase (TM0665) from *Thermotoga maritima* at 1.8 Å resolution.," *Proteins*, vol. 56, pp. 387–391, Aug. 2004.

[18] P. Burkhard, C. H. Tai, C. M. Ristroph, P. F. Cook, and J. N. Jansonius, "Ligand binding induces a large conformational change in O-acetylserine sulfhydrylase from *Salmonella typhimurium.*," *J. Mol. Biol.*, vol. 291, pp. 941–953, Aug. 1999.

[19] P. Chander, K. M. Halbig, J. K. Miller, C. J. Fields, H. K. Bonner, G. K. Grabner, R. L. Switzer, and J. L. Smith, "Structure of the

nucleotide complex of PyrR, the pyr attenuation protein from *Bacillus caldolyticus*, suggests dual regulation by pyrimidine and purine nucleotides.," *J. Bacteriol.*, vol. 187, pp. 1773–1782, Mar. 2005.

[20] D. R. Tomchick, R. J. Turner, R. L. Switzer, and J. L. Smith, "Adaptation of an enzyme to regulatory function: structure of *Bacillus subtilis* PyrR, a pyr RNA-binding attenuation protein and uracil phosphoribosyltransferase.," *Structure*, vol. 6, pp. 337–350, Mar. 1998.

[21] M. Lynch, "The evolution of multimeric protein assemblages.," *Mol. Biol. Evol.*, vol. 29, pp. 1353–1366, May 2012.

[22] P. Neudecker, P. Robustelli, A. Cavalli, P. Walsh, P. Lundström, A. Zarrine-Afsar, S. Sharpe, M. Vendruscolo, and L. E. Kay, "Structure of an intermediate state in protein folding and aggregation," *Science*, vol. 336, pp. 362–366, Apr. 2012.

[23] C. M. Dobson, "Protein folding and misfolding," *Nature*, vol. 426, pp. 884–890, Dec. 2003.

[24] D. M. Korzhnev, X. Salvatella, M. Vendruscolo, A. A. Di Nardo, A. R. Davidson, C. M. Dobson, and L. E. Kay, "Low-populated folding intermediates of Fyn SH3 characterized by relaxation dispersion NMR," *Nature*, vol. 430, pp. 586–590, July 2004.

[25] K. Lindorff-Larsen, S. Piana, R. O. Dror, and D. E. Shaw, "How fast-folding proteins fold," *Science*, vol. 334, pp. 517–520, Oct. 2011.

[26] E. D. Levy, E. Boeri Erba, C. V. Robinson, and S. A. Teichmann, "Assembly reflects evolution of protein complexes.," *Nature*, vol. 453, pp. 1262–1265, June 2008.

[27] J. A. Marsh, H. Hernández, Z. Hall, S. E. Ahnert, T. Perica, C. V. Robinson, and S. A. Teichmann, "Protein complexes are under

evolutionary selection to assemble via ordered pathways.," *Cell*, vol. 153, pp. 461–470, Apr. 2013.

[28] N. Jia-Lin Ma and D. F. Stern, "Regulation of the Rad53 protein kinase in signal amplification by oligomer assembly and disassembly.," *Cell cycle*, vol. 7, pp. 808–817, Mar. 2008.

[29] J. L. Benesch, J. A. Aquilina, A. J. Baldwin, A. Rekas, F. Stengel, R. A. Lindner, E. Basha, G. L. Devlin, J. Horwitz, E. Vierling, J. A. Carver, and C. V. Robinson, "The quaternary organization and dynamics of the molecular chaperone HSP26 are thermally regulated.," *Chem. Biol.*, vol. 17, pp. 1008–1017, Sept. 2010.

[30] F. Sobott, J. L. P. Benesch, E. Vierling, and C. V. Robinson, "Subunit Exchange of Multimeric Protein Complexes," *J. Biol. Chem.*, vol. 277, pp. 38921–38929, Oct. 2002.

[31] A. J. Baldwin, H. Lioe, G. R. Hilton, L. A. Baker, J. L. Rubinstein, L. E. Kay, and J. L. Benesch, "The polydispersity of αB-crystallin Is rationalized by an interconverting polyhedral architecture," *Structure*, vol. 19, pp. 1855–63+, 2011.

[32] S.-J. J. Cai and M. Inouye, "Spontaneous subunit exchange and biochemical evidence for trans-autophosphorylation in a dimer of *Escherichia coli* histidine kinase (EnvZ).," *J. Mol. Biol.*, vol. 329, pp. 495–503, June 2003.

[33] G. Chevreux, C. Atmanene, P. Lopez, J. Ouazzani, A. Van Dorsselaer, B. Badet, M.-A. A. Badet-Denisot, and S. Sanglier-Cianférani, "Monitoring the dynamics of monomer exchange using electrospray mass spectrometry: the case of the dimeric glucosamine-6-phosphate synthase.," *J. Am. Soc. Mass Spectrom.*, vol. 22, pp. 431–439, Mar. 2011.

[34] A. Y. Y. Park, S. Jergic, A. Politis, B. T. Ruotolo, D. Hirshberg, L. L. Jessop, J. L. Beck, D. Barsky, M. O'Donnell, N. E. Dixon, and C. V. Robinson, "A single subunit directs the assembly of the *Escherichia coli* DNA sliding clamp loader.," *Structure*, vol. 18, pp. 285–292, Mar. 2010.

[35] M. C. Leake, J. H. Chandler, G. H. Wadhams, F. Bai, R. M. Berry, and J. P. Armitage, "Stoichiometry and turnover in single, functioning membrane protein complexes.," *Nature*, vol. 443, pp. 355–358, Sept. 2006.

[36] Y.-C. Chang and J. U. Bowie, "Measuring membrane protein stability under native conditions," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 111, pp. 219–224, Jan. 2014.

[37] G. Bouvignies, P. Vallurupalli, D. F. Hansen, B. E. Correia, O. Lange, A. Bah, R. M. Vernon, F. W. Dahlquist, D. Baker, and L. E. Kay, "Solution structure of a minor and transiently formed state of a T4 lysozyme mutant," *Nature*, vol. 477, pp. 111–114, Sept. 2011.

[38] Y.-T. Lai, E. Reading, G. L. Hura, K.-L. Tsai, A. Laganowsky, F. J. Asturias, J. A. Tainer, C. V. Robinson, and T. O. Yeates, "Structure of a designed protein cage that self-assembles into a highly porous cube," *Nat. Chem.*, vol. 6, pp. 1065–1071, Dec. 2014.

[39] N. P. King, W. Sheffler, M. R. Sawaya, B. S. Vollmar, J. P. Sumida, I. André, T. Gonen, T. O. Yeates, and D. Baker, "Computational design of self-assembling protein nanomaterials with atomic level accuracy.," *Science*, vol. 336, pp. 1171–1174, June 2012.

[40] H. S. McHaourab, J. A. Godar, and P. L. Stewart, "Structure and mechanism of protein stability sensors: chaperone activity of small heat shock proteins.," *Biochemistry*, vol. 48, pp. 3828–3837, May 2009.

[41] M. Haslbeck, T. Franzmann, D. Weinfurtner, and J. Buchner, "Some like it hot: the structure and function of small heat-shock proteins," *Nat. Struct. Mol. Biol.*, vol. 12, pp. 842–846, Oct. 2005.

[42] S. Carra, P. Rusmini, V. Crippa, E. Giorgetti, A. Boncoraglio, R. Cristofani, M. Naujock, M. Meister, M. Minoia, H. H. Kampinga, and A. Poletti, "Different anti-aggregation and pro-degradative functions of the members of the mammalian sHSP family in neurological disorders.," *Philos. Trans. R. Soc. Lond. B. Biol. Sci.*, vol. 368, pp. 20110409+, May 2013.

[43] H. S. McHaourab, Y.-L. L. Lin, and B. W. Spiller, "Crystal structure of an activated variant of small heat shock protein Hsp16.5.," *Biochemistry*, vol. 51, pp. 5105–5112, June 2012.

[44] E. Basha, C. Jones, A. E. Blackwell, G. Cheng, E. R. Waters, K. A. Samsel, M. Siddique, V. Pett, V. Wysocki, and E. Vierling, "An unusual dimeric small heat shock protein provides insight into the mechanism of this class of chaperones.," *J. Mol. Biol.*, vol. 425, pp. 1683–1696, May 2013.

[45] K. K. Kim, R. Kim, and S.-H. Kim, "Crystal structure of a small heat-shock protein," *Nature*, vol. 394, pp. 595–599, Aug. 1998.

[46] R. L. van Montfort, E. Basha, K. L. Friedrich, C. Slingsby, and E. Vierling, "Crystal structure and assembly of a eukaryotic small heat shock protein.," *Nat. Struct. Biol.*, vol. 8, pp. 1025–1030, Dec. 2001.

[47] Y. Hanazono, K. Takeda, M. Yohda, and K. Miki, "Structural studies on the oligomeric transition of a small heat shock protein, StHsp14.0.," *J. Mol. Biol.*, vol. 422, pp. 100–108, Sept. 2012.

[48] Y. Hanazono, K. Takeda, T. Oka, T. Abe, T. Tomonari, N. Akiyama, Y. Aikawa, M. Yohda, and K. Miki, "Nonequivalence observed for the

16-meric structure of a small heat shock protein, SpHsp16.0, from *Schizosaccharomyces pombe.*," *Structure*, vol. 21, pp. 220–228, Feb. 2013.

[49] C. K. Kennaway, J. L. Benesch, U. Gohlke, L. Wang, C. V. Robinson, E. V. Orlova, H. R. Saibil, H. R. Saibi, and N. H. Keep, "Dodecameric structure of the small heat shock protein Acr1 from *Mycobacterium tuberculosis.*," *J. Biol. Chem.*, vol. 280, pp. 33419–33425, Sept. 2005.

[50] H. E. White, E. V. Orlova, S. Chen, L. Wang, A. Ignatiou, B. Gowen, T. Stromer, T. M. Franzmann, M. Haslbeck, J. Buchner, and H. R. Saibil, "Multiple distinct assemblies reveal conformational flexibility in the small heat shock protein Hsp26.," *Structure*, vol. 14, pp. 1197–1204, July 2006.

[51] P. Poulain, J.-C. Gelly, and D. Flatters, "Detection and Architecture of Small Heat Shock Protein Monomers," *PLoS One*, vol. 5, pp. e9990+, Apr. 2010.

[52] E. Basha, H. O'Neill, and E. Vierling, "Small heat shock proteins and α-crystallins: dynamic proteins with flexible functions.," *Trends Biochem. Sci.*, vol. 37, pp. 106–117, Mar. 2012.

[53] G. R. Hilton, H. Lioe, F. Stengel, A. J. Baldwin, and J. L. Benesch, "Small heat-shock proteins: paramedics of the cell.," *Top. Curr. Chem.*, vol. 328, pp. 69–98, 2013.

[54] A. R. Clark, C. E. Naylor, C. Bagnéris, N. H. Keep, and C. Slingsby, "Crystal structure of R120G disease mutant of human αB-crystallin domain dimer shows closure of a groove.," *J. Mol. Biol.*, vol. 408, pp. 118–134, Apr. 2011.

[55] A. Laganowsky, J. L. Benesch, M. Landau, L. Ding, M. R. Sawaya, D. Cascio, Q. Huang, C. V. Robinson, J. Horwitz, and D. Eisenberg, "Crystal structures of truncated αA and αB crystallins reveal structural

mechanisms of polydispersity important for eye lens function.," *Protein Sci.*, vol. 19, pp. 1031–1043, May 2010.

[56] S. Jehle, B. van Rossum, J. R. Stout, S. M. Noguchi, K. Falber, K. Rehbein, H. Oschkinat, R. E. Klevit, and P. Rajagopal, "αB-crystallin: a hybrid solid-state/solution-state NMR investigation reveals structural aspects of the heterogeneous oligomer.," *J. Mol. Biol.*, vol. 385, pp. 1481–1497, Feb. 2009.

[57] T. Kriehuber, T. Rattei, T. Weinmaier, A. Bepperling, M. Haslbeck, and J. Buchner, "Independent evolution of the core domain and its flanking sequences in small heat shock proteins," *FASEB J.*, vol. 24, pp. 3633–3642, Oct. 2010.

[58] S. Patel, E. Vierling, and F. Tama, "Replica exchange molecular dynamics simulations provide insight into substrate recognition by small heat shock proteins.," *Biophys. J.*, vol. 106, pp. 2644–2655, June 2014.

[59] S. D. Weeks, E. V. Baranova, M. Heirbaut, S. Beelen, A. V. Shkumatov, N. B. Gusev, and S. V. Strelkov, "Molecular structure and dynamics of the dimeric human small heat shock protein HSPB6.," *J. Struct. Biol.*, vol. 185, pp. 342–354, Mar. 2014.

[60] G. Cheng, E. Basha, V. H. Wysocki, and E. Vierling, "Insights into small heat shock protein and substrate structure during chaperone action derived from hydrogen/deuterium exchange and mass spectrometry.," *J. Biol. Chem.*, vol. 283, pp. 26634–26642, Sept. 2008.

[61] J. A. Aquilina, J. L. P. Benesch, L. L. Ding, O. Yaron, J. Horwitz, and C. V. Robinson, "Subunit exchange of polydisperse proteins: mass spectrometry reveals consequences of αA-crystallin truncation," *J. Biol. Chem*, vol. 280, pp. 14485–91+, 2005.

[62] A. J. Baldwin, H. Lioe, C. V. Robinson, L. E. Kay, and J. L. Benesch, "αB-crystallin polydispersity is a consequence of unbiased quaternary dynamics," *J. Mol. Biol.*, vol. 413, pp. 297–309+, 2011.

[63] E. Basha, C. Jones, V. Wysocki, and E. Vierling, "Mechanistic differences between two vonserved classes of small heat shock proteins found in the plant cytosol," *J. Biol. Chem.*, vol. 285, pp. 11489–11497, Apr. 2010.

[64] J. A. Aquilina, S. Shrestha, A. M. Morris, and H. Ecroyd, "Structural and functional aspects of hetero-oligomers formed by the small heat shock proteins αB-crystallin and HSP27.," *J. Biol. Chem.*, vol. 288, pp. 13602–13609, May 2013.

[65] A. J. Painter, N. Jaya, E. Basha, E. Vierling, C. V. Robinson, and J. L. Benesch, "Real-time monitoring of protein complexes reveals their quaternary organization and dynamics.," *Chem. Biol.*, vol. 15, pp. 246–253, Mar. 2008.

[66] P. J. van den Oetelaar, P. F. van Someren, J. A. Thomson, R. J. Siezen, and H. J. Hoenders, "A dynamic quaternary structure of bovine α-crystallin as indicated from intermolecular exchange of subunits.," *Biochemistry*, vol. 29, pp. 3488–3493, Apr. 1990.

[67] J. A. Aquilina, J. L. P. Benesch, O. A. Bateman, C. Slingsby, and C. V. Robinson, "Polydispersity of a mammalian chaperone: mass spectrometry reveals the population of oligomers in αB-crystallin," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 100, pp. 10611–10616, Sept. 2003.

[68] A. J. Baldwin, H. Lioe, G. R. Hilton, L. A. Baker, J. L. Rubinstein, L. E. Kay, and J. L. Benesch, "The polydispersity of αB-crystallin is rationalized by an interconverting polyhedral architecture.," *Structure*, vol. 19, pp. 1855–1863, Dec. 2011.

[69] F. Stengel, A. J. Baldwin, A. J. Painter, N. Jaya, E. Basha, L. E. Kay, E. Vierling, C. V. Robinson, and J. L. Benesch, "Quaternary dynamics and plasticity underlie small heat shock protein chaperone function," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 107, pp. 2007–12+, 2010.

[70] H. Bloemendal, W. de Jong, R. Jaenicke, N. H. Lubsen, C. Slingsby, and A. Tardieu, "Ageing and vision: structure, stability and function of lens crystallins.," *Prog. Biophys. Mol. Biol.*, vol. 86, pp. 407–485, Nov. 2004.

[71] A. J. Baldwin, G. R. Hilton, H. Lioe, C. Bagneris, J. L. Benesch, and L. E. Kay, "Quaternary dynamics of αB-crystallin as a direct consequence of localised tertiary fluctuations in the C-terminus," *J. Mol. Biol.*, vol. 413, pp. 310–20+, 2011.

[72] T. M. Franzmann, P. Menhorn, S. Walter, and J. Buchner, "Activation of the chaperone Hsp26 is controlled by the rearrangement of its thermosensor domain.," *Mol. Cell*, vol. 29, pp. 207–216, Feb. 2008.

[73] A. J. Baldwin, H. Lioe, C. V. Robinson, L. E. Kay, and J. L. P. Benesch, "αB-Crystallin Polydispersity Is a Consequence of Unbiased Quaternary Dynamics," *J. Mol. Biol.*, Aug. 2011.

[74] S. Jehle, P. Rajagopal, B. Bardiaux, S. Markovic, R. Kühne, J. R. Stout, V. A. Higman, R. E. Klevit, B.-J. J. van Rossum, and H. Oschkinat, "Solid-state NMR and SAXS studies provide a structural basis for the activation of αB-crystallin oligomers.," *Nat. Struct. Mol. Biol.*, vol. 17, pp. 1037–1042, Sept. 2010.

[75] M. P. Bova, O. Yaron, Q. Huang, L. Ding, D. A. Haley, P. L. Stewart, and J. Horwitz, "Mutation R120G in αB-crystallin, which is linked to a desmin-related myopathy, results in an irregular structure and defective chaperone-like function.," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 96, pp. 6137–6142, May 1999.

[76] M. Michiel, F. Skouri-Panet, E. Duprat, S. Simon, C. Férard, A. Tardieu, and S. Finet, "Abnormal assemblies and subunit exchange of αB-crystallin R120 mutants could be associated with destabilization of the dimeric substructure.," *Biochemistry*, vol. 48, pp. 442–453, Jan. 2009.

[77] D. Hayes, V. Napoli, A. Mazurkie, W. F. Stafford, and P. Graceffa, "Phosphorylation dependence of hsp27 multimeric size and molecular chaperone function.," *J. Biol. Chem.*, vol. 284, pp. 18801–18807, July 2009.

[78] E. T. McDonald, M. Bortolus, H. A. Koteiche, and H. S. Mchaourab, "Sequence, structure, and dynamic determinants of Hsp27 (HspB1) equilibrium dissociation are encoded by the N-terminal domain.," *Biochemistry*, vol. 51, pp. 1257–1268, Feb. 2012.

[79] T. Rogalla, M. Ehrnsperger, X. Preville, A. Kotlyarov, G. Lutsch, C. Ducasse, C. Paul, M. Wieske, A. P. Arrigo, J. Buchner, and M. Gaestel, "Regulation of Hsp27 oligomerization, chaperone function, and protective activity against oxidative stress/tumor necrosis factor α by phosphorylation.," *J. Biol. Chem.*, vol. 274, pp. 18947–18956, July 1999.

[80] J. A. Aquilina, J. L. Benesch, L. L. L. Ding, O. Yaron, J. Horwitz, and C. V. Robinson, "Phosphorylation of αB-crystallin alters chaperone function through loss of dimeric substructure.," *J. Biol. Chem.*, vol. 279, pp. 28675–28680, July 2004.

[81] J. Peschek, N. Braun, J. Rohrberg, K. C. C. Back, T. Kriehuber, A. Kastenmüller, S. Weinkauf, and J. Buchner, "Regulated structural transitions unleash the chaperone activity of αB-crystallin.," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 110, pp. 201308898+, Oct. 2013.

[82] M. Kimura, "Evolutionary rate at the molecular level.," *Nature*, vol. 217, pp. 624–626, Feb. 1968.

[83] J. L. King and T. H. Jukes, "Non-Darwinian evolution.," *Science*, vol. 164, pp. 788–798, May 1969.

[84] S. J. Gould and R. C. Lewontin, "The spandrels of San Marco and the Panglossian paradigm: a critique of the adaptationist programme.," *Proc. R. Soc. London, Ser. B*, vol. 205, pp. 581–598, Sept. 1979.

[85] W. H. Bragg and W. L. Bragg, "The Reflection of X-rays by Crystals," *Proc. Roy. Soc. London A*, vol. 88, pp. 428–438, July 1913.

[86] W. H. Miller, *A Treatise on Crystallography.* Cambridge: Deighton, 1839.

[87] R. A. Crowther and D. M. Blow, "A method of positioning a known molecule in an unknown crystal structure," *Acta Crystallogr.*, vol. 23, pp. 544–548, Oct. 1967.

[88] W. A. Hendrickson, J. L. Smith, and S. Sheriff, "Direct phase determination based on anomalous scattering.," *Methods Enzymol.*, vol. 115, pp. 41–55, 1985.

[89] W. A. Hendrickson and C. M. Ogata, "Phase Determination from Multiwavelength Anomalous Diffraction Measurements," *Methods Enzymol.*, vol. 276, pp. 494–523, 1997. Hendrickson97.

[90] M. Chruszcz, W. Potrzebowski, M. D. Zimmerman, M. Grabowski, H. Zheng, P. Lasota, and W. Minor, "Analysis of solvent content and oligomeric states in protein crystals–does symmetry matter?," *Protein Sci.*, vol. 17, pp. 623–632, Apr. 2008.

[91] J. Janin and F. Rodier, "Protein-protein interaction at crystal contacts.," *Proteins*, vol. 23, pp. 580–587, Dec. 1995.

[92] F. Cazals, F. Proust, R. P. Bahadur, and J. Janin, "Revisiting the Voronoi description of protein-protein interfaces.," *Protein Sci.*, vol. 15, pp. 2082–2092, Sept. 2006.

[93] C. Lesieur, *Oligomerization of Chemical and Biological Compounds*. InTech, 2014.

[94] J. Lipfert and S. Doniach, "Small-angle X-ray scattering from RNA, proteins, and protein complexes.," *Annu. Rev. Biophys. Biomol. Struct.*, vol. 36, no. 1, pp. 307–327, 2007.

[95] R. P. Rambo and J. A. Tainer, "Accurate assessment of mass, models and resolution by small-angle scattering," *Nature*, vol. 496, pp. 477–481, Apr. 2013.

[96] G. L. Hura, A. L. Menon, M. Hammel, R. P. Rambo, F. L. Poole, S. E. Tsutakawa, F. E. Jenney, S. Classen, K. A. Frankel, R. C. Hopkins, S.-J. J. Yang, J. W. Scott, B. D. Dillard, M. W. Adams, and J. A. Tainer, "Robust, high-throughput solution structural analyses by small angle X-ray scattering (SAXS).," *Nat. Methods*, vol. 6, pp. 606–612, Aug. 2009.

[97] G. R. Hilton and J. L. Benesch, "Two decades of studying non-covalent biomolecular assemblies by means of electrospray ionization mass spectrometry.," *J. R. Soc. Interface*, vol. 9, pp. 801–816, May 2012.

[98] J. L. Benesch and B. T. Ruotolo, "Mass spectrometry: come of age for structural and dynamical biology.," *Curr. Opin. Struct. Biol.*, vol. 21, pp. 641–649, Oct. 2011.

[99] A. J. Heck and R. H. Van Den Heuvel, "Investigation of intact protein complexes by mass spectrometry.," *Mass Spectrom. Rev.*, vol. 23, no. 5, pp. 368–389, 2004.

[100] H. Hernandez and C. V. Robinson, "Determining the stoichiometry and interactions of macromolecular assemblies from mass spectrometry," *Nat. Protocols*, vol. 2, pp. 715–726, Mar. 2007.

[101] A. J. Painter, N. Jaya, E. Basha, E. Vierling, C. V. Robinson, and J. L. P. Benesch, "Real-time monitoring of protein complexes reveals their quaternary organization and dynamics," *Chem. Biol.*, vol. 15, pp. 246–253+, 2008.

[102] J. A. Aquilina, J. L. P. Benesch, O. A. Bateman, C. Slingsby, and C. V. Robinson, "Polydispersity of a mammalian chaperone: mass spectrometry reveals the population of oligomers in αB-crystallin," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 100, pp. 10611–6+, 2003.

[103] G. Ben-Nissan and M. Sharon, "Capturing protein structural kinetics by mass spectrometry.," *Chem. Soc. Rev.*, vol. 40, pp. 3627–3637, July 2011.

[104] F. Sobott, H. Hernández, M. G. McCammon, M. A. Tito, and C. V. Robinson, "A tandem mass spectrometer for improved transmission and analysis of large macromolecular assemblies.," *Anal. Chem.*, vol. 74, pp. 1402–1407, Mar. 2002.

[105] J. B. Fenn, M. Mann, C. K. Meng, S. F. Wong, and C. M. Whitehouse, "Electrospray ionization for mass spectrometry of large biomolecules.," *Science*, vol. 246, pp. 64–71, Oct. 1989.

[106] M. Wilm and M. Mann, "Analytical Properties of the Nanoelectrospray Ion Source," *Anal. Chem.*, vol. 68, pp. 1–8, Jan. 1996.

[107] I. V. Chernushevich and B. A. Thomson, "Collisional cooling of large ions in electrospray mass spectrometry.," *Anal. Chem.*, vol. 76, pp. 1754–1760, Mar. 2004.

[108] B. E. Winger, K. J. Light-Wahl, R. R. Ogorzalek Loo, H. R. Udseth, and R. D. Smith, "Observation and implications of high mass-to-charge ratio ions from electrospray ionization mass spectrometry.," *J. Am. Soc. Mass Spectrom.*, vol. 4, pp. 536–545, July 1993.

[109] J. H. J. Dawson and M. Guilhaus, "Orthogonal-acceleration time-of-flight mass spectrometer," *Rapid Commun. Mass Spectrom.*, vol. 3, pp. 155–159, May 1989.

[110] J. Coles and M. Guilhaus, "Orthogonal acceleration a new direction for time-of-flight mass spectrometry: Fast, sensitive mass analysis for continuous ion sources," *Trends Anal. Chem.*, vol. 12, pp. 203–213, May 1993.

[111] J. L. P. Benesch, "Collisional activation of protein complexes: picking up the pieces," *J. Am. Soc. Mass. Spectrom.*, vol. 20, pp. 341–8+, 2009.

[112] J. L. P. Benesch, J. A. Aquilina, B. T. Ruotolo, F. Sobott, and C. V. Robinson, "Tandem mass spectrometry reveals the quaternary organization of macromolecular assemblies," *Chem. Biol.*, vol. 13, pp. 597–605+, 2006.

[113] M. F. Bush, Z. Hall, K. Giles, J. Hoyes, C. V. Robinson, and B. T. Ruotolo, "Collision cross sections of proteins and their complexes: a calibration framework and database for gas-phase structural biology," *Anal. Chem.*, vol. 82, pp. 9557–9565, Oct. 2010.

[114] E. A. Mason and H. W. Schamp, "Mobility of gaseous lons in weak electric fields," *Ann. Phys.*, vol. 4, pp. 233–270, July 1958.

[115] D. Hewitt, E. Marklund, D. Scott, C. Robinson, and A. Borysik, "A hydrodynamic comparison of solution and gas phase proteins and their complexes.," *The journal of physical chemistry. B*, vol. 118, pp. 8489–8495, July 2014.

[116] E. Marklund, M. Degiacomi, C. Robinson, A. Baldwin, and J. Benesch, "Collision cross sections for structural proteomics.," *Structure (London, England : 1993)*, vol. 23, pp. 791–799, Apr. 2015.

[117] B. T. Ruotolo, J. L. Benesch, A. M. Sandercock, S.-J. J. Hyung, and C. V. Robinson, "Ion mobility-mass spectrometry analysis of large protein complexes.," *Nat. Protoc.*, vol. 3, pp. 1139–1152, June 2008.

[118] A. Politis, F. Stengel, Z. Hall, H. Hernández, A. Leitner, T. Walzthoeni, C. V. Robinson, and R. Aebersold, "A mass spectrometry-based hybrid method for structural modeling of protein complexes.," *Nat. Methods*, vol. 11, pp. 403–406, Apr. 2014.

Chapter 2

C-terminal interactions and dynamics of αB-crystallin

2.1 Abstract

αB-crystallin is a highly dynamic, polydisperse small heat shock protein (sHSP) that can form oligomers ranging in mass from 200 to 800 kDa. Here we employ an ion-mobility mass spectrometry (IM-MS) approach to assess the role of the C-terminal tail in the self-assembly and dynamics of αB-crystallin. Titration experiments allow us to monitor the binding of peptides representing the C-terminus to the αB-crystallin core domain, and observe affinities to both monomeric and dimeric forms individually. We find that binding the second peptide equivalent to the core domain dimer is considerably more difficult than the first. This effect is significantly more pronounced for a peptide that includes a disordered and flexible C-terminal extension. Together these results reconcile seemingly contradictory observations about the structural role of the C-terminus and reveal how its interaction with the core domain contributes significantly to the quaternary dynamics of αB-crystallin oligomers.

2.2 Introduction

αB-crystallin is an oligomeric vertebrate small heat-shock protein (sHSP) with molecular chaperone activity [1]. Like other sHSPs [2, 3, 4, 5], αB-crystallin, interacts with destabilized client proteins in an ATP-independent manner that

prevents their aggregation [6], and facilitates their recovery by the downstream ATP-dependent molecular chaperone machinery [7]. Its activity extends to binding to mature amyloid fibrils, resulting in inhibition of fibril elongation [8, 9, 10, 11]. These functions, defined primarily *in vitro*, likely make αB-crystallin an integral part of the complex regulatory machinery ensuring cellular proteostasis [12, 13, 14]. *In vivo*, malfunction of αB-crystallin has been linked to numerous protein deposition diseases ranging from cataract formation [15] to motor neuropathies [16], and neurodegenerative disease [17].

Relatively little is known about the structure-function relationship in sHSPs. Despite being ubiquitous across all kindoms of life, high-resolution structural information about the assembled form of sHSPs is surprisingly scarce [2, 3, 4, 5]. This is primarily due to their dynamic and frequently polydisperse nature, properties that make high-resolution studies particularly challenging [18]. αB-crystallin assembles into an ensemble of interconverting oligomeric states spanning approximately 10-40 subunits [19, 20]. In recent years, considerable strides have been made in overcoming this heterogeneity and interrogating the structure of αB-crystallin. X-ray crystallography studies of N- and C- terminally truncated constructs of both αB-crystallin and its eye lens specific isoform αA [21, 22, 23, 24], and solid-state nuclear magnetic resonance spectroscopy (ss-NMR) of the full-length protein [25], have revealed the structure of a dimeric building block. This protomer is composed of a β-sandwich α-crystallin core that assembles through anti-parallel (AP) pairwise interactions between extended β6+7 strands (Fig. 2.1). This interface appears common to metazoan sHSPs, and has been observed in different registration states [21, 22, 23, 24], termed AP_I, AP_{II} and AP_{III} [23].

αB-crystallin dimers assemble into oligomers via interactions mediated, at least in part, by the terminal regions [15, 4]. The N-termini of αB-crystallin are clearly involved in stabilizing its oligomers, since truncation of this region results in a shift towards low molecular weight oligomers [23]. Structurally, the contacts made by the N-terminus are the least well understood, although the

limited available data indicate considerable structural heterogeneity [25]. The C-terminus has been observed to make reciprocal interactions between dimers in the related HSP 16.9 [26] and HSP 16.5 [27], via a conserved IXI motif in the C-terminus that binds into a hydrophobic groove between the β4 and β8 strands. The same inter-molecular interaction has also been observed in αB-crystallin in NMR experiments [28, 25], in X-ray structures of N-terminally truncated versions of αB-crystallin [23], and by FRET [29]. For this reason, this interaction has served as a spatial restraint in several recent models proposed for the structure of αB-crystallin oligomers [30, 31, 32].

Despite this C-terminal interaction being considered canonical in the sHSPs, its role in the assembly of αB-crystallin is contentious. Mutations in the conserved IXI region that disrupt oligomer formation in plants or prokaryotes do not produce suboligomeric stoichiometries in αB-crystallin [29]. The C-terminus of αB-crystallin further differs from its better understood homolgoues in having two distinct regions: the tail, which is comprised of a linker followed by a 9-residue palindromic sequence that includes the IXI motif, and a flexible extension that includes all downstream residues (Fig. 2.1A) [4]. The residues in the extension adopt a disordered conformation, and in NMR experiments tumble sufficiently fast to be observable despite the large size of αB-crystallin oligomers [33]. In ssNMR experiments performed at low temperature, the IXI motif was observed to be bound to the β4-β8 groove, similar to its conformation in a crystal structure of αB-crystallin lacking its N-terminus and the C-terminal extension (Fig. 2.1B) [23]. However, NMR experiments performed at physiological temperature and pH have revealed that under such conditions the IXI motif adopts a largely disordered conformation, and that, although held in close proximity to its binding site, it is not bound to the β4-β8 groove the vast majority of the time [28, 34, 35]. Instead, it undergoes a temperature dependent transition between being largely unbound at physiological temperature, and a mixture of bound and unbound at temperatures below 0 °C [35].

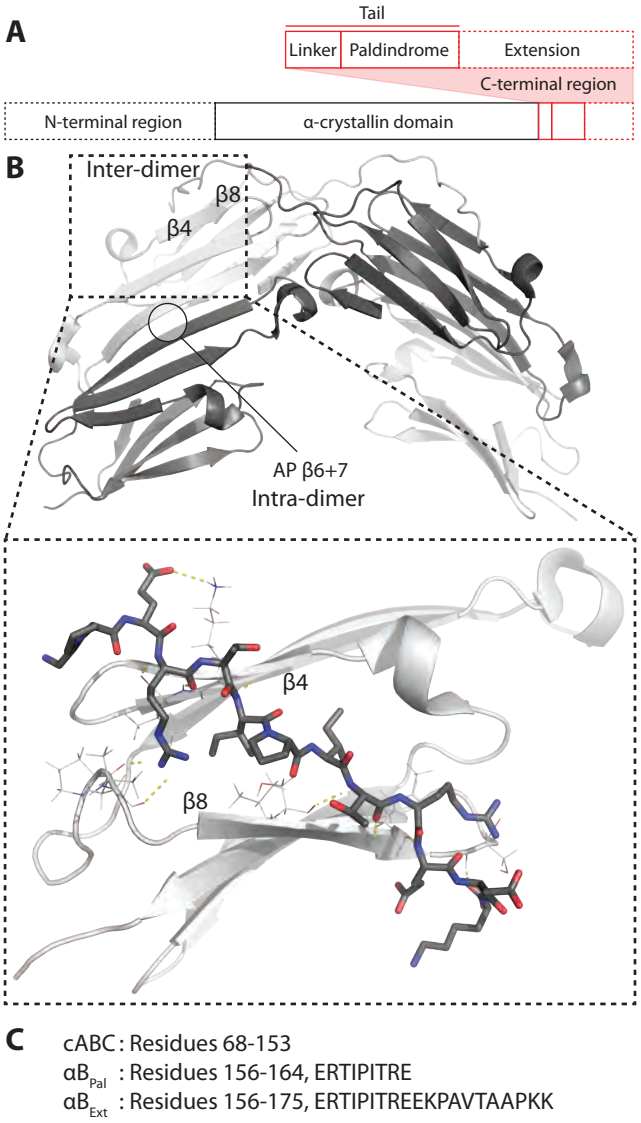


Figure 2.1: A) The primary sequence of αB-crystallin consists of N- and C-terminal regions flanking a core α-crystallin domain. The C-terminal region is itself split into three sections. A palindromic region of sequence, centred on an IXI motif that is highly conserved amongst sHSPs, is preceded by a linker to the α-crystallin domain. These two sections are often termed the C-terminal tail, and are followed by the extension, a region of sequence that is highly flexible. Atomic resolution information is lacking for this region, as well as the N-terminal region (dashed lines). B) Crystal structure of a truncated construct of αB-crystallin (PDB: 3L1G) illustrating the inter-molecular contacts made. Two types of dimer interface are formed: an intra-dimer interface between β6+7 strands, and an inter-dimer interface between the C-terminal tail and a hydrophobic groove separating the β4 and β8 strands on an adjacent monomer. The expansion highlights this latter interaction, showing the palindromic residues E156-E164 and the hydrogen bonds (yellow) made with the groove. C) The nomenclature for the truncated domain and C-terminal peptides under investigation here is tabulated.

At physiological temperatures, the IXI motif transitions between its unbound state and a scarcely populated bound state on the millisecond timescale [35]. The rate of these fluctuations correlates tightly with the rate of subunit exchange between αB-crystallin oligomers over a large range of solution conditions [28], indicating that they have some regulatory effect on αB-crystallin's quaternary dynamics. Similarly, truncation of the flexible extension reduces the rate of subunit exchange [36, 37], suggesting that this region, despite being largely disordered, is also involved in determining the quaternary dynamics of αB-crystallin oligomers. Mutations near the C-terminus lead to aberrant chaperone activity *in vitro* [29, 38], in cells [39, 40], and are associated with disease [41, 42, 43, 44], suggesting that the influence of the C-terminus on the quaternary αB-crystallin may be important for its chaperone function.

In this study we investigate the molecular details of the interaction between the α-crystallin C-terminal region and the β4-β8 groove in its core domain. We perform our investigations using nanoelectrospray mass spectrometry (MS) and ion mobility spectrometry (IM). Since its initial applications to the study of protein assemblies in the early 1990s [45], MS has matured as a structural biology approach to allow the accurate determination of oligomeric stoichiometry and dynamics [46, 47, 48]. We investigate the binding of the palindromic amino acids surrounding the IXI motif (^{156}ERTIPITRE^{164}), in the form of a peptide, to the isolated αB-crystallin core domain. Taking advantage of high-resolution IM-MS, we determine separate association constants for sequential peptide binding events to monomers and dimers for peptides with and without the flexible C-terminal extension. Using this approach we find that the C-terminal extension destabilises binding of the IXI motif, preventing it from forming the tightly bound conformations as observed in other sHSPs, and that the binding of two C-terminal peptides to a dimer weakens the dimer interface of the core domain. These results lend further support to the idea that the C-terminus is an subtle regulator of the quaternary dynamics of αB-crystallin oligomers.

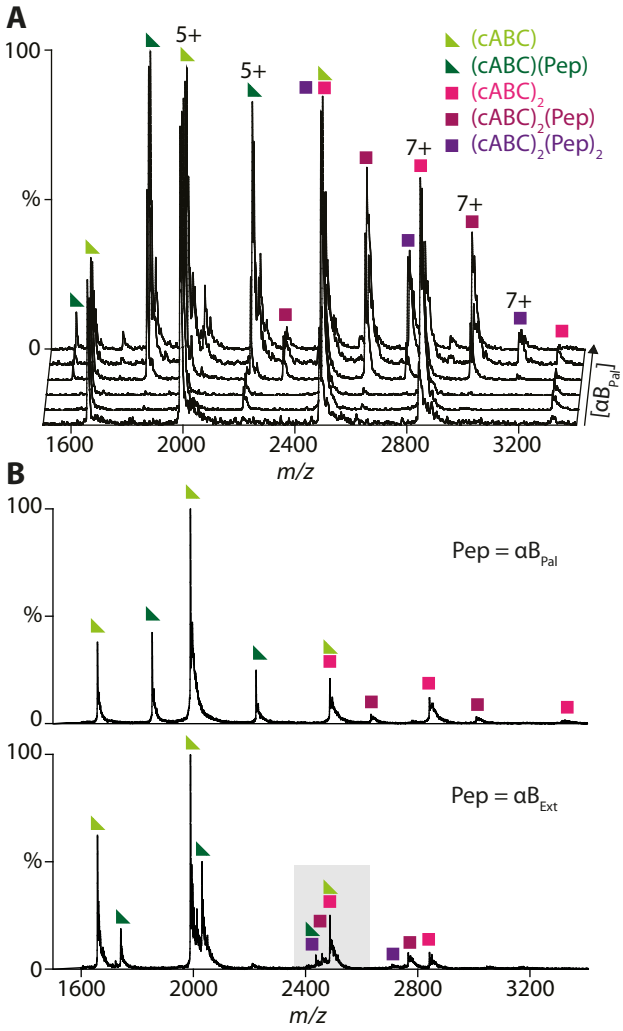


Figure 2.2: A) Nanoelectrospray mass spectra of 5 μM αB-crystallin core-domain, cABC (comprising residues 68-153) incubated with a peptide αB_{Pal}, representing the palindromic C-terminal tail (residues 156-164). The concentrations of αB_{Pal} are 4, 8, 16, 32, 64, 128 μM from front to back. Both monomeric (triangles) and dimeric (squares) cABC are visible and, as the concentration of αB_{Pal} is increased, multiple complexes with ligand appear. Up to one peptide is observed to bind the cABC monomer, whereas up to two are seen to bind the cABC dimer. B) Comparison of spectra obtained at a peptide concentration of 16 μM for αB_{Pal} (upper panel) and αB_{Ext} (lower panel), with the latter, longer peptide representing the combined tail and extension of αB-crystallin (residues 156-175). Binding is observed in both cases, but quantification of the relative abundances is hampered by the overlap of charge states from the different species (e.g. grey box).

2.3 Results

2.3.1 The C-terminal extension of αB-crystallin modulates the binding of the tail to the core domain

To examine the binding of the tail residues of αB-crystallin we investigated the binding of peptides to a truncated construct cABC, residues 68-153, in which the β4-β8 groove is necessarily unoccupied. We performed titration experiments in which we incubated cABC with either the palindromic peptide ERTIPITRE (αB_{Pal}), representing residues 156-164 of αB-crystallin or the peptide ERTIPITREEKPAVTAAPKK (αB_{Ext}), residues 156-175, the combined

C-terminal tail and extension, at a range of concentrations (Fig. 2.1C).

At the lowest concentration of peptide (4 μM), peaks are observed corresponding to cABC monomer and dimer (Fig. 2.2A). The preponderance of monomer is in line with a weak intra-dimer interface [49] with a K_D in the low μM range [23]. Upon increasing the concentration of peptide, peaks appear that correspond to the complexes formed between the protein and peptide at stoichiometries of $(\text{cABC})_1(\alpha\text{B}_{Pal})$, $(\text{cABC})_2(\alpha\text{B}_{Pal})$ and $(\text{cABC})_2(\alpha\text{B}_{Pal})_2$ (Fig. 2.2A). Notably, even at the highest peptide concentrations for which we were able to obtain mass spectra (128 μM peptide, more than a ten-fold excess), we never observe more than about half of the available sites are occupied. Analogous experiments with αB_{Ext} showed apparently similar levels of bound forms (Fig. 2.2B). Binding of the C-terminal peptide therefore appears to be relatively weak, in agreement with previous data [28, 34, 35].

2.3.2 IM-MS allows quantitative extraction of peptide binding affinities to the αB-crystallin core domain

While these MS data show qualitatively that the peptides bind to cABC they are challenging to interpret quantitatively due to the overlap of charge states from the different species, e.g. $(\text{cABC})^{4+}$ and $(\text{cABC})_2^{8+}$. This is particularly noticeable in the case of the cluster of peaks around 2470 *m/z* for cABC incubated with αB_{Ext} (boxed, Fig. 2.2B). In order to facilitate the deconvolution of these peaks we employed a hybrid IM-MS strategy. These experiments provide an orthogonal dimension of separation to *m/z* according to the ions' ability to traverse a tube of neutral gas at low pressure under the influence of a weak electric field. The transit time through the drift cell of the ions is proportional to their CCS and inversely proportional to charge [50].

An IM-MS spectrum of cABC with αB_{Ext} shows a series of charge states separated in both *m/z* and arrival time (Fig. 2.3A). The separation between monomer and dimer is dramatically improved relative to the one-dimensional MS experiment (Fig. 2.2). In order to analyse this spectrum we employed an

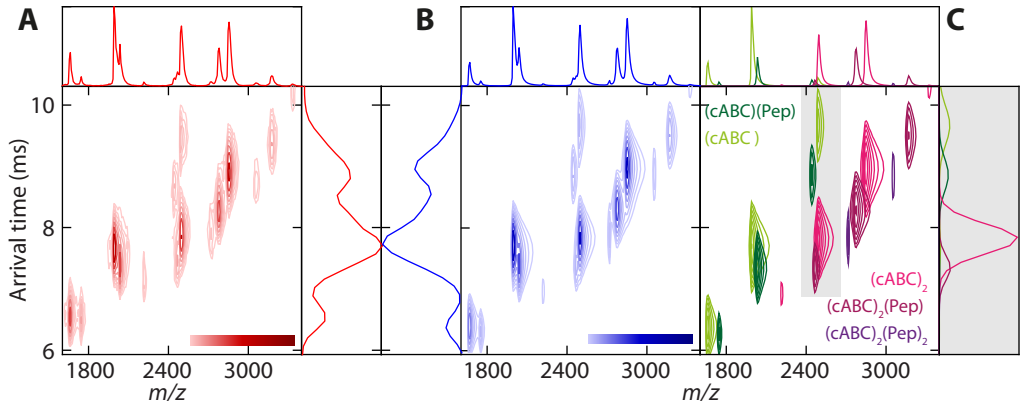


Figure 2.3: A) Two-dimensional IM-MS spectrum of cABC incubated with αB_{Ext} at a concentration of 32 μM, with the corresponding line-projections in the m/z and arrival time dimensions. The one-dimensional data are complex, and charge states can be seen to overlap in either dimension. They are however well separated in the two dimensional IM-MS spectra. B) The data are analysed by determining the best-fitting calculated spectrum using the CHAMP algorithm. The fit (blue) matches the experimental data (red) extremely well, and can therefore be used to obtain the relative contribution of the different species to the overall spectrum. C) A breakdown of the fitted spectrum showing the specific contributions of the monomeric (greens) and dimeric (magentas) species. The arrival time projection (right panel) shows the projected intensity of the peaks in the range 2350-2650 m/z (grey box), as in Fig. 2.2B. The complexity of this region emphasises the vital role of IM for obtaining robust abundances of the various species observed over the course of the titration.

extension of our spectrum calculation algorithm CHAMP [51], modified to accommodate the arrival time dimension [30]. By calculating different IM-MS spectra from candidate distributions of the molecular components in the spectrum, and comparing this to the data, a best fit spectrum is obtained (Fig. 2.3B, Fig. 2.6). The correspondence between data and fit is excellent, allowing us to deconvolve the relative contributions of the different stoichiometries to the overall spectrum (Fig. 2.3C). Using this approached we analysed our titration series in order to interrogate the concentration dependence of peptide binding (Fig. 2.6). At the lowest concentration of peptide (4 μM) peaks are visible corresponding to cABC monomers and dimers, and negligible binding of peptide is observed for either αB_{Pal} or αB_{Ext} (Fig. 2.4A, upper panels). At the highest concentration (128 μM) however, significant abundances of the bound forms are present (Fig. 2.4A, lower panels). It is notable how in the case of

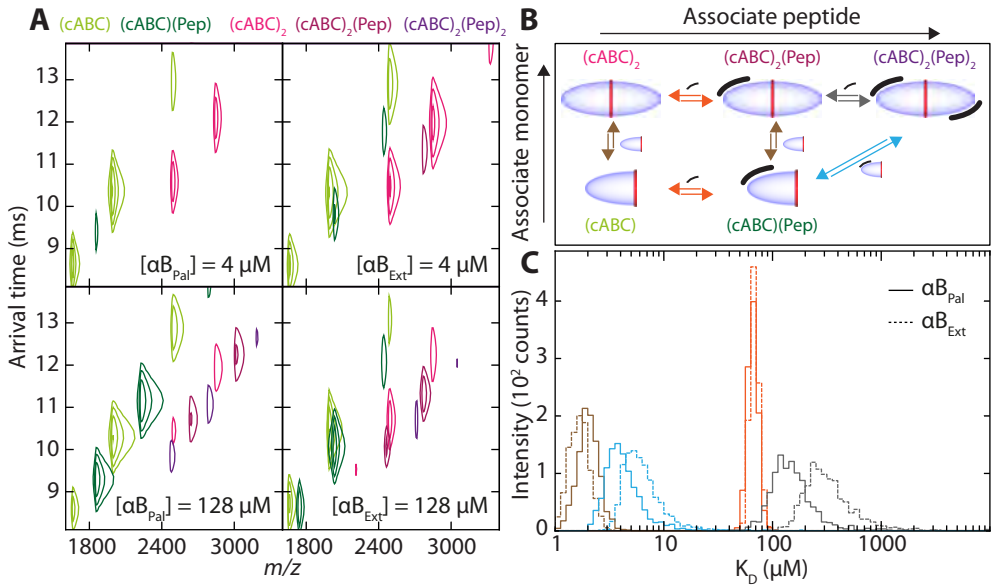


Figure 2.4: A) Representative IM-MS spectra of cABC incubated with either αB_{Pal} (left panels) or αB_{Ext} (right panels). At 4 μM peptide (upper panels) both cABC monomers and dimers are observed, with very little peptide binding, revealing that the K_D must be significantly larger than 1 μM. At 128 μM peptide (lower panels) both apo and holo forms are observed. Notably, the abundance of $(cABC)_2(Pep)_2$, the domain dimer with two peptides, is lower in the case of αB_{Ext} than in the case of αB_{Pal}, whereas the quantities of $(cABC)_1(Pep)_1$ are similar in both. B) Quantitative analysis of peptide binding. We analyse the MS titration data in terms of a ligand binding model invoking a total of six coupled equilibria with an increasing number of free parameters (see methods for further details). In the simplest model all sets of monomer/dimer equilibria (brown, grey) and peptide binding equilibria (orange, blue) are identical so that the system is controlled by only two independent parameters. Allowing the second peptide binding to have different binding affinity to the first binding event leads to a model specified by four affinities (orange, blue, brown and grey), determined from three independent fitting parameters. The inclusion of this additional parameter is statistically justified for the titrations with αB_{Pal} and αB_{Ext}, whereas models of increased complexity were found not to be. As described in the text, these data enable us to conclude that the second peptide binds with lower affinity than the first. C) A bootstrap analysis was performed to assess the uncertainty in our estimates of dissociation constants. The resulting histograms are shown for fitting the statistically justified four K_D model to data from αB_{Pal} (solid lines) and αB_{Ext} (dashed lines). The histograms reveal that a single peptide has a similar affinity to cABC in both cases (orange). Binding the affinity of binding the second peptide to the cABC dimer is lower for both peptides (grey). Moreover, the K_D of $(cABC)_2(Pep)_2$ is notably higher in the case of αB_{Ext} than αB_{Pal}, indicating weaker binding.

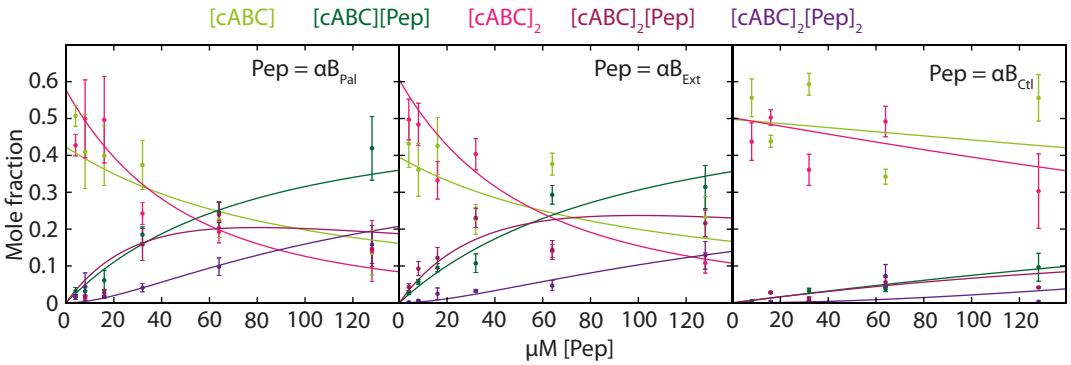


Figure 2.5: Peptide binding titrations fitted to the affinity model described above. In the case of αB_{Pal} (left) and αB_{Ext} (middle), using the three parameter model rather than the two parameter was justified (P-level $< 6\%$). Physically, the two parameter fit was poor as far less $[P_2L_2]$ was observed than the model would predict. The reduction in χ^2 stems from the three parameter model allowing for this, which enables us to conclude that binding the second peptide is significantly more difficult than binding the first. By contrast, no significant reduction in χ^2 was found moving from the two to the three parameter model in the case of the negative-control peptide αB_{Ctl} (right). In no instance was fitting using the four parameter model statistically justified.

αB_{Ext} significantly less $(cABC)_2(Pep)_2$ is observed than for αB_{Pal}.

2.3.3 Binding of the second peptide allosterically weakens the dimer interface

To interpret these differences we examined the abundances of the different molecular species in the context of a protein-ligand binding model (Fig. 2.4B). In order to adequately reproduce the data we had to employ a scheme that allowed the second peptide to bind a dimer with an equilibrium constant that differed to that of binding the first (Fig. 2.4B,C and Fig. 2.5). For both αB_{Pal} and αB_{Ext} we find that the K_D of binding the first peptide equivalent was found to be approximately 70 μM, with a dimer interface K_D of approximately 2 μM (orange and brown, respectively, Fig. 2.4B,C). Remarkably, the affinity for the second peptide equivalent we find to be substantially lower: 150 μM for αB_{Pal} and 300 μM for αB_{Ext} (grey, Fig. 2.4B,C) Furthermore, the dimer interface for $(cABC)_2(Pep)_2$ is slightly weaker (4 μM for αB_{Pal} and 6 μM for αB_{Ext}) than in

the other dimeric species. Taken together, our data demonstrate that binding two peptides to a dimer is substantially more difficult than binding just one, and has the concomitant effect of actively weakening the intra-dimer interface. This destabilisation effect of the second peptide is substantially greater for αB_{Ext} relative to αB_{Pal}. Extrapolated to the context of oligomers of full-length protein therefore, the concurrent binding of two C-termini would precipitate the dissociation of a monomer.

2.4 Discussion

Our study demonstrates that MS can be used to provide a highly detailed view of macroscopic quaternary structure and dynamics of αB-crystallin. By employing an IM-MS strategy to analyse titrations of cABC and C-terminal mimicking peptides, we were able to robustly quantify the different molecular species that co-exist in solution. This has allowed us to extract the K_Ds of peptide binding and, crucially, do so for both the monomeric and dimeric forms of cABC present at equilibrium. Interestingly, we observe that two sequential peptide binding events to $(cABC)_2$ are not equivalent. While binding of the first peptide is relatively weak, it is significantly more difficult to bind the second. This effect is particularly pronounced in the case of αB_{Ext}, which is larger than αB_{Pal}, and has a significant number of positively charge residues at the C-terminus. Both steric and coulombic interactions are therefore likely to play a role in decreasing the binding affinity of the second peptide. This is supported by observation that not only does truncation of the C-terminus reduce the rate of subunit exchange of the α-crystallins [36, 37], but also that mutating charged residues affects their self-assembly [38]. Binding of the second peptide has an additional interesting effect in that the intra-dimer interface is destabilised. This remarkable finding is consistent with the model derived by solution-state NMR that suggests that subunit exchange is facilitated by the C-terminus binding the β4+8 groove [28, 52].

It is interesting to note that even in the case of αB_{Ext}, we still readily observed binding of the peptide to the αB-crystallin core domain. This seemingly contradicts NMR experiments that find the C-terminus largely unstructured [35]. This is particularly puzzling since the concentration of the C-terminus localized near the β4-8 groove would be very high in αB-crystallin oligomers. Under such conditions, the C-terminus would be expected to be mostly bound even if its affinity for the β4-8 groove is relatively low. Our binding experiments with αB_{Ext} provide a possible explanation: the reduced affinity of a second peptide to the core domain likely results from steric clashing and coulombic repulsion between the residues in the extension of the two C-termini. In all current models of αB-crystallin oligomers the α-crystallin domains are arranged such that the β4-8 grooves from several monomers are in close proximity [30, 31, 32]. Steric clashes and culombic repulsion between the C-terminal extensions localised to their respective binding sites in αB-crystallin oligomers could therefore perhaps act to keep the IXI largely unbound and flexible.

In such a model the C-terminal extension ensures that the C-terminus remains largely unbound. When bound, it catalizes the dissociation of monomers through its allosteric effect on the dimer interface, thereby regulating the quaternary dynamics of αB-crystallin oligomers. This is in accordance with an alanine scan of the palindromic tail in the full-length protein, which, rather than leading to disassembly, lead to a shift towards more even numbered stoichiometries (see [53], in which the results presented here are published). Such a direct influence on the rate of subunit exchange would explain why mutant proteins with shorter extensions have reduced quaternary dynamics [36, 54]. At present, we are forced to speculate about the biological significance of such dynamics. Flexible C-termini and fast subunit exchange may be important to ensure αB-crystallin remains soluble at exceptionally high concentrations in the eye lens [55]. Similarly, the molecular chaperone activity of sHSPs may itself benefit from fast subunit exchange, if αB-crystallin oligomers primarily serve as a passive storage form for chaperone active subunits that become liberated

during subunit exchange [56]. In either case the allosteric coupling between binding of the C-terminus and the dimer interface provides a rationale as to why mutations of C-terminal residues lead to aberrant function *in vitro* [29, 38], in cells [39, 40], and have been identified in dilated cardiomyopathy [42, 43] and cataract [41, 44].

2.5 Materials and methods

Truncated αB-crystallin (residues 68-153, cABC) was expressed in *Escherichia coli* (*E. coli*) and purified as previously described [23, 57]. The peptide, ERTIPITRE, was expressed and purified from *E. coli* as described previously [58] whereas peptides ERTIPITREEKPAVTAAPKK and TIERPREIT were purchased from BioMatik (Canada).

2.5.1 Measuring peptide binding to the αB-crystallin core domain

All titration experiments were carried out at room temperature and allowed to equilibrate for a minimum of 30 min prior to measurement. Solutions of cABC were prepared at a final monomeric concentration of 4 μM (as determined using a bicinchoninic acid assay on a stock solution) in 200 mM ammonium acetate pH 6.9, whilst the final concentrations of the C-terminal peptides were varied. All C-terminal peptides were dissolved in water, with final concentrations of 4, 8, 16, 32, 64, 128 μM.

Titration data were collected on a Synapt G1 HDMS instrument (Waters, UK), modified to incorporate a linear drift tube [59]. Parameters were set to minimize activation of gas phase ions and maximize separation between monomeric and dimeric species in the arrival-time dimension. To ensure the preservation of the labile interaction between peptide and cABC in the mass spectrometer, and reduce the amount of alkali metal adducts, a small reservoir of acetonitrile was introduced into the source region of the mass spectrometer

[60]. The instrumental voltages used were: capillary, 1.4 V; sampling cone, 10 V; extraction cone, 1 V; trap, 5 V; bias 20 V; drift tube, 50 V. The gas flow rates were: trap (argon), 2.6 ml/min; drift tube (nitrogen), 50 or 20 ml/min; and the backing pressure was 3.8 mbar. Three spectra were recorded for each titration point, using a new needle each time to account for needle-to-needle variation.

2.5.2 Ion-mobility mass spectrometry titration data analysis

External calibration of the spectra was performed using MassLynx software (Waters UK Ltd), and the data was exported for analysis using our home-built algorithm CHAMP [51]. CHAMP was adjusted to incorporate the arrival-time dimension, along the lines we have described previously [30]. The CCS of each molecular species was considered to be invariant with charge state, an assumption which appears to be valid for protein assemblies in the absence of denaturant or organic solvent [59]. To ensure the stability of the protein assemblies in the gas phase, and CCSs close to that expected from their solution structure we examined them in a charge-reduced form [61], achieved here by a partial pressure of acetonitrile in the source region [60]. Individual IM-MS peaks were considered to have Gaussian profiles in the arrival time dimension [30], and asymmetric Lorenzian profiles in the mass-to-charge (m/z) dimension [62]. The width of the IM peaks were fitted by assuming the full-width-at-half-maximum scales linearly with drift time for ions of similar mobility [63]. Minimisation was performed using a combination of stimulated annealing and Levenberg-Marquart methods [51]. Examples of fits to raw data are shown in Figure 2.6.

2.5.3 Determination of peptide binding affinity

To determine the binding affinities of the C-terminal peptides to cABC we used the abundances extracted by CHAMP as described above. In order to obtain

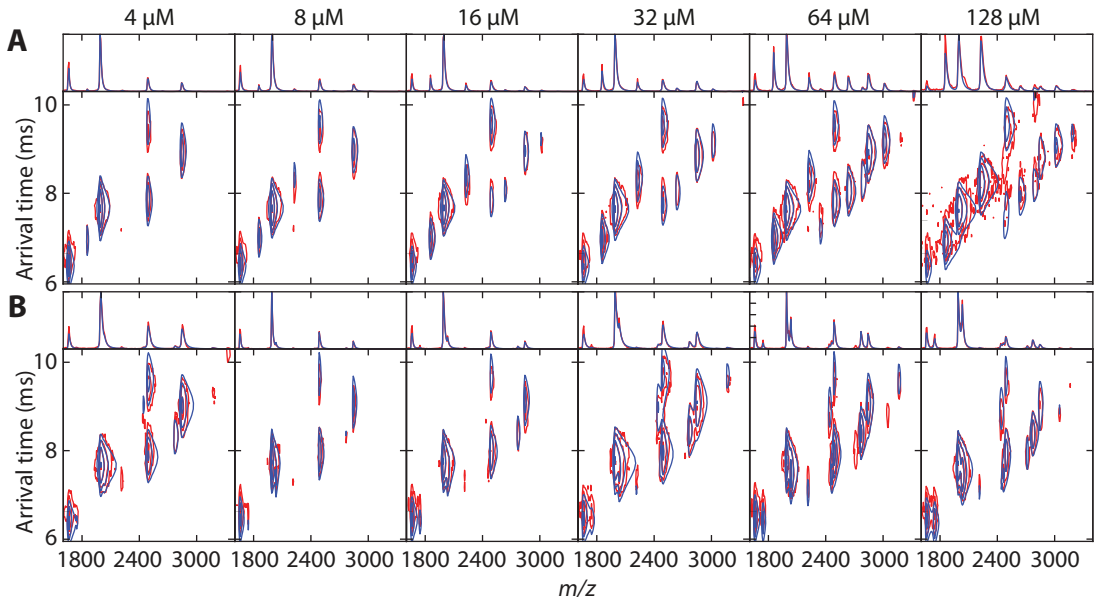


Figure 2.6: Comparison of both IM-MS experimental data (blue) and fits (red) obtained using the CHAMP algorithm, as described in the text, for titrations of A) αB_{Pal} and B) αB_{Ext} with cABC. The fitted spectra are in excellent agreement with the experimental data, as can be seen from the one-dimensional projections.

reliable binding affinities it is important to account for non-specific complexes [63]. These arise from non-specific association during the electrospray process, and can become significant at high protein or ligand concentrations [64]. To address this we performed two sets of control experiments to assess the extent of non-specific binding. In the first we obtained spectra of the scrambled C-terminal peptide TIERPREIT incubated at a range of concentrations with cABC, and in the second, the specific peptide ERTIPITRE with the unrelated protein cytochrome C. In both cases the abundance of protein-ligand complexes can be entirely ascribed to non-specific association. In this way the abundances of detected protein-ligand complexes can be corrected to reflect solely the specific associations [65]. Both of these control experiments were found to correspond to an apparent K_D on the order of mM. Thus adjusted, the experimental distribution of apo- and holo- cABC were examined in terms of a protein-ligand binding model (see below) in order to obtain the binding affinities.

2.5.4 Ligand binding model

In the absence of peptide ligands, L, the truncated αB-crystallin constructs, P, exist in equilibrium between monomeric, P_1 and dimeric, P_2, forms. Given that each molecule of P has the capacity to bind one molecule of L, and that oligomerisation rarely proceeds past the level of the dimerisation in these constructs, we have six species to consider, P_1, P_1L, P_2, P_2L, P_2L_2, and L. Interpreting binding data therefore requires consideration of six coupled equilibria and their corresponding association constants.

$$P_1 + L \rightleftharpoons P_1L, K_{A,0} = \frac{[P_1L]}{[P_1][L]} \tag{2.1}$$

$$P_2 + L \rightleftharpoons P_2L, K_{A,1} = \frac{[P_2L]}{[P_2][L]} \tag{2.2}$$

$$P_2L + L \rightleftharpoons P_2L_2, K_{A,2} = \frac{[P_2L_2]}{[P_2L][L]} \tag{2.3}$$

$$P_1 + P_1 \rightleftharpoons P_2, K_{A,3} = \frac{[P_2]}{[P_1]^2} \tag{2.4}$$

$$P_1L + P_1 \rightleftharpoons P_2L, K_{A,4} = \frac{[P_2L]}{[P_1L][P_1]} \tag{2.5}$$

$$P_1L + P_1L \rightleftharpoons P_2L_2, K_{A,5} = \frac{[P_2L_2]}{[P_1L]^2} \tag{2.6}$$

$$\tag{2.7}$$

There is some degeneracy in the association pathways in this scheme: $K_{A,2}K_{A,4} = K_{A,5}K_{A,0}$ and $K_{A,1}K_{A,3} = K_{A,4}K_{A,0}$. These equalities imply physically that weakening the monomer/dimer interface results in a decrease in species like P_2L and P_2L_2. The dimer interface free energy and the effective association constants for the dimer species are therefore closely linked. Since the free energies specified by these association constants are state functions (meaning that the free energies associated with the equilibria are independent of the route taken through the association scheme), we can completely define the system in terms of $K_{A,0}$, $K_{A,3}$, $K_{A,4}$ and $K_{A,5}$. $K_{A,2}$ and $K_{A,1}$ can subsequently be determined using the relations stated above. The total concentration of ligand

is given by:

$$[L_{tot}] = \sum_{i=1}^{2}\sum_{j=0}^{i} j[P_i L_j + [L] \tag{2.8}$$

$$= [L] + [P_1 L] + [P_2 L] + 2[P_2 L_2] \tag{2.9}$$

$$= [L] + K_{A,0}[L][P_1] + K_{A,4}K_{A,0}[P_1^2][L] + 2K_{A,5}K_{A,0}^2[P_1]^2[L]^2 \tag{2.10}$$

Which is quadratic in $[L]$. Solving for $[L]$ and taking the positive solution yields:

$$[L] = \frac{(1 + K_{A,0}[P_1] + K_{A,4}K_{A,0}[P_1]^2)}{2L_{Tot}} + \frac{\sqrt{(1 + K_{A,0}[P_1] + K_{A,4}K_{A,0}[P_1]^2)^2 + 8L_{Tot}K_{A,5}K_{A,0}^2[P_1]^2}}{2L_{Tot}} \tag{2.11}$$

The quantity of free peptide can be directly determined if the four equilibrium constants, total peptide concentration and quantity of free monomeric protein is known. Similarly, the total concentration of protein is given by:

$$[P_{Tot}] = \sum_{i=1}^{2}\sum_{j=0}^{i} i[P_i L_i] \tag{2.12}$$

$$= [P_i] + [P_i L] + 2([P_2] + [P_2 L] + [P_2 L_2]) \tag{2.13}$$

$$= [P_1] + K_{A,0}[P_1][L] + 2K_{A,3}[P_1]^2 + 2K_{A,0}K_{A,4}[P_1]^2[L] + 2K_{A,5}K_{A,0}^2[P_1]^2[L]^2 \tag{2.14}$$

Inserting the expression for $[L]$ into this expression gives an expression for $[P_{Tot}]$ in terms of $[P1]$ and the four equilibrium constants. In the binding experiments described in this chapter, the total quantity of protein has been experimentally determined, and so the equation can be solved numerically for $[P1]$ for any set of four trial equilibrium constants. With these values, all remaining concentrations $[P_i L_j]$ are trivially back-calculated.

2.5.5 Simplifying the model

The equilibrium scheme above implicitly allows for ligand binding to influence the strength of the dimer interface, and vice versa. For example, if $K_{A,5} > K_{A,3}$

the dimer interface is effectively strengthened by the binding of two ligands. This model is a function of four equilibrium constants and is the most complex considered in this study. We can make simplifying assumptions that result in models of reduced complexity.

If we make the assumption that $K_{A,3} = K_{A,4}$, which results in the further equality $K_{A,0} = K_{A,1}$, the simplified scheme states that a single ligand binds both monomers and dimers with equal affinity. This model still allows the second ligand to bind with a different affinity, and is specified by three free parameters rather than four. Finally, we can assume that $K_{A,3} = K_{A,4} = K_{A,5}$ then $K_{A,0} = K_{A,1} = K_{A,2}$. This model assumes that binding the ligand does not alter the dimer interface in any way, and that ligand binding is entirely independent of dimerization. This relatively simple model is specified by only two free parameters, and is the least complex of those tested in this study.

2.5.6 Fitting the data to the model

From the IM-MS experiments, we have experimental measurements of the signal intensity from each species, $I(P_iL_j)$, such that the total signal intensity is $I(Tot) = \sum_{i=1}^{2}\sum_{j=0} I(P_iL_j)$. For any given set of equilibrium constants, we can evaluate the following χ^2 function:

$$\chi^2 = \sum_{i=1}^{2}\sum_{j=0}^{i}((\frac{I(P_iL_j)}{I(Tot)})_{exp} - (\frac{(P_iL_j)}{P_{Tot}})_{calc})^2 \tag{2.15}$$

Where $I(P_iLj)$, $I(Tot)$ and $[P_{Tot}]$ are determined from experiment, and the set of $[P_iL_j]$ are obtained as a function of the four equilibrium constants, as described above. Assuming that signal intensities are proportional to the concentrations of the corresponding species in solution, the set of equilibrium constants that return the lowest χ^2 value are those that best describe the species present in solution.

2.5.7 Determination of errors in fitted binding affinity parameters

Multiple repeats of the samples were run on independent protein separations on different days in order to obtain error bars. Typically, 3-6 independent measurements were performed on each peptide / concentration combination. The error bars shown in the isotherms in figure 2.5 represent the standard error of the mean $\sigma_M = \sigma/\sqrt{N-1}$ where σ is the standard deviation of the repeated measurements. In order to obtain standard deviations and means for the fitted quantities from our non-linear fits, we employed a boot-strapping procedure. Briefly, each fit was repeated 1000 times by re-sampling with replacement n times from the data-points of each titration, where n is the total number of data-points in each dataset. Histograms of the K_As extracted from each round of bootstrapping were constructed and fitted to Gaussian distributions. Since there is only one 'true' value for each fitted K_A, these Gaussians describe the uncertainty in our estimate of the true K_A introduced by error in the measurement, which is assumed to be randomly distributed around the 'true' mean. Their standard deviation thus provide a numerical estimate of the error in our estimate of each affinity constant.

2.5.8 Binding model assessment

To discriminate between models of increasing complexity (different numbers of independent K_As), we employed F-tests. Beginning from the lowest level of model complexity, we determined whether the reduction in the χ^2 achieved by introducing another free K_A is justified given the resulting reduction in the degrees of freedom (ν, equal to number of datapoints - the number of free parameters). We could then assess the probability of whether the reduction in χ^2 in moving from a simple to a complex model is justified by the reduction in degrees of freedom by evaluating the P-level:

$$P = \beta_r\left(\frac{\nu_{complex}}{2}, \frac{\nu_{simple} - \nu_{complex}}{2}, \frac{\chi^2_{complex}}{\chi^2_{simple}}\right) \tag{2.16}$$

Where β_R is the regularised β function, defined as

$$\beta_R(x,y,z) = \frac{\Gamma(x+y)}{\Gamma(x)\Gamma(y)} \int_0^z t^{x-1}(1-t)^{y-1}dt \tag{2.17}$$

and Γ is the Gamma function,

$$\Gamma(z) = \int_0^{\inf} t^{z-1}e^{-t}dt \tag{2.18}$$

in the relevant $z > 0$ limit. A probability of less than 0.06 (6%), was taken to justify an increase in model complexity. If the probability was greater than 6%, then noise in the data is most likely responsible for the improved χ^2. In this case the simpler model was chosen for analysis.

2.5.9 Adjustments due to non-specific association in the gas phase

In a typical nanoelectrospray MS experiment, the protein will be at concentrations on the order of 10 μM. In the case where K_D is on the order of 100 μM or weaker, it is necessary to mix with ligand at concentrations that exceed 100 μM. Under these conditions of relatively high concentration, the probability of having a peptide and an αB-crystallin domain in the same droplet becomes appreciable, appearing in the spectrum as a low concentration bound species [66]. Such non-specific binding effects occlude the effects of specific binding present in the solution phase.

Considering the case of a binding equilibrium where both specific (present in solution) and non-specific (formed during the electrospray process) are detected, the equilibria and corresponding association constants are given by:

$$P + L \rightleftharpoons PL^{spec}, K_A^{spec} = \frac{[PL^{spec}]}{[P][L]} \tag{2.19}$$

$$P + L \rightleftharpoons PL^{non}, K_A^{non} = \frac{[PL^{non}]}{[P][L]} \tag{2.20}$$

The total protein concentration will be given by:

$$P_{Tot} = [P] + [PL^{spec}] + [PL^{non}] \tag{2.21}$$

$$= [P] + [P][L](K_A^{spec} + K_A^{non}) = [P] + K_A^{eff}[P][L] \tag{2.22}$$

This leads to a useful general result, $K_A^{eff} = K_A^{non} + K_A^{spec}$. The effective association constant is that which is determined through neglecting the effects of non-specific association. We arrive at an estimate of the required association constant through $K_A^{spec} = K_A^{eff} - K_A^{non}$. Providing we can estimate the magnitude of non-specific binding effects, they can be subtracted. When the association constant is more than an order of magnitude larger than the effects of non-specific binding, the correction is negligible. However, when the effective constant and the non-specific constant are on the same order of magnitude, the correction will be large.

In order to estimate the magnitude of effects due to non-specific association, 128μM of the αB_{Pal} peptide was added to the unrelated protein cytochrome C. In the resulting spectra, a bound state populated to approximately 10% state was detected. This corresponds to an effective K_D of 1 mM. In this study, this value is taken as an approximation for K_D^{non}. The association constants that describing peptide binding, $K_{A,3}$, $K_{A,4}$ and $K_{A,5}$ are susceptible to effects of this kind. Values reported in the manuscript for these constants have had non-specific effects subtracted. This correction is small for all K_D values presented in this work.

2.6 References

[1] J. Horwitz, "α-crystallin," *Exp. Eye Res.*, vol. 76, pp. 145–153, Feb. 2003.

[2] E. Basha, H. O'Neill, and E. Vierling, "Small heat shock proteins and α-crystallins: dynamic proteins with flexible functions.," *Trends Biochem. Sci.*, vol. 37, pp. 106–117, Mar. 2012.

[3] M. Haslbeck, T. Franzmann, D. Weinfurtner, and J. Buchner, "Some like it hot: the structure and function of small heat-shock proteins," *Nat. Struct. Mol. Biol.*, vol. 12, pp. 842–846, Oct. 2005.

[4] G. R. Hilton, H. Lioe, F. Stengel, A. J. Baldwin, and J. L. P. Benesch, "Small heat-shock proteins: paramedics of the cell," *Top. Curr. Chem.*, 2012.

[5] H. S. McHaourab, J. A. Godar, and P. L. Stewart, "Structure and mechanism of protein stability sensors: chaperone activity of small heat shock proteins.," *Biochemistry*, vol. 48, pp. 3828–3837, May 2009.

[6] J. Horwitz, "α-crystallin can function as a molecular chaperone.," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 89, pp. 10449–10453, Nov. 1992.

[7] U. Jakob, M. Gaestel, K. Engel, and J. Buchner, "Small heat shock proteins are molecular chaperones.," *J. Biol. Chem.*, vol. 268, pp. 1517–1520, Jan. 1993.

[8] S. L. Shammas, C. A. Waudby, S. Wang, A. K. Buell, T. P. Knowles, H. Ecroyd, M. E. Welland, J. A. Carver, C. M. Dobson, and S. Meehan, "Binding of the molecular chaperone αB-crystallin to Aβ amyloid fibrils inhibits fibril elongation.," *Biophys. J.*, vol. 101, pp. 1681–1689, Oct. 2011.

[9] C. A. Waudby, T. P. Knowles, G. L. Devlin, J. N. Skepper, H. Ecroyd, J. A. Carver, M. E. Welland, J. Christodoulou, C. M. Dobson, and S. Meehan, "The interaction of αB-crystallin with mature α-synuclein amyloid fibrils inhibits their elongation," *Biophys. J.*, vol. 98, pp. 843–51+, 2010.

[10] B. Raman, T. Ban, M. Sakai, S. Y. Pasta, T. Ramakrishna, H. Naiki, Y. Goto, and C. M. h. M. Rao, "αB-crystallin, a small heat-shock protein, prevents the amyloid fibril growth of an amyloid β-peptide and β2-microglobulin.," *Biochem. J.*, vol. 392, pp. 573–581, Dec. 2005.

[11] T. P. Knowles, W. Shu, G. L. Devlin, S. Meehan, S. Auer, C. M. Dobson, and M. E. Welland, "Kinetics and thermodynamics of amyloid formation

from direct measurements of fluctuations in fibril mass," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 104, pp. 10016–21+, 2007.

[12] F. U. Hartl, A. Bracher, and M. Hayer-Hartl, "Molecular chaperones in protein folding and proteostasis," *Nature*, vol. 475, pp. 324–32+, 2011.

[13] K. Richter, M. Haslbeck, and J. Buchner, "The heat shock response: life on the verge of death.," *Mol. Cell*, vol. 40, pp. 253–266, Oct. 2010.

[14] W. E. Balch, R. I. Morimoto, A. Dillin, and J. W. Kelly, "Adapting proteostasis for disease intervention," *Science*, vol. 319, pp. 916–9+, 2008.

[15] A. R. Clark, N. H. Lubsen, and C. Slingsby, "sHSP in the eye lens: crystallin mutations, cataract and proteostasis," *Int. J. Biochem. Cell Biol.*, 2012.

[16] A. Boncoraglio, M. Minoia, and S. Carra, "The family of mammalian small heat shock proteins (HSPBs): implications in protein deposit diseases and motor neuropathies.," *Int. J. Biochem. Cell Biol.*, vol. 44, pp. 1657–1669, Oct. 2012.

[17] H. Ecroyd and J. A. Carver, "Crystallin proteins and amyloid fibrils.," *Cell. Mol. Life. Sci.*, vol. 66, pp. 62–81, Jan. 2009.

[18] J. Horwitz, "α crystallin: the quest for a homogeneous quaternary structure.," *Exp. Eye Res.*, vol. 88, pp. 190–194, Feb. 2009.

[19] J. A. Aquilina, J. L. P. Benesch, O. A. Bateman, C. Slingsby, and C. V. Robinson, "Polydispersity of a mammalian chaperone: mass spectrometry reveals the population of oligomers in αB-crystallin," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 100, pp. 10611–10616, Sept. 2003.

[20] D. A. Haley, J. Horwitz, and P. L. Stewart, "The small heat-shock protein, αB-crystallin, has a variable quaternary structure.," *J. Mol. Biol.*, vol. 277, pp. 27–35, Mar. 1998.

[21] C. Bagnéris, O. A. Bateman, C. E. Naylor, N. Cronin, W. C. Boelens, N. H. Keep, and C. Slingsby, "Crystal structures of α-crystallin domain dimers of αB-crystallin and Hsp20," *J. Mol. Biol.*, vol. 392, pp. 1242–1252, Oct. 2009.

[22] A. R. Clark, C. E. Naylor, C. Bagnéris, N. H. Keep, and C. Slingsby, "Crystal structure of R120G disease mutant of human αB-crystallin domain dimer shows closure of a groove.," *J. Mol. Biol.*, vol. 408, pp. 118–134, Apr. 2011.

[23] A. Laganowsky, J. L. Benesch, M. Landau, L. Ding, M. R. Sawaya, D. Cascio, Q. Huang, C. V. Robinson, J. Horwitz, and D. Eisenberg, "Crystal structures of truncated αA and αB crystallins reveal structural mechanisms of polydispersity important for eye lens function," *Protein Sci.*, vol. 19, pp. 1031–43+, 2010.

[24] A. Laganowsky and D. Eisenberg, "Non-3D domain swapped crystal structure of truncated zebrafish αA crystallin," *Protein Sci.*, vol. 19, pp. 1978–84+, 2010.

[25] S. Jehle, P. Rajagopal, B. Bardiaux, S. Markovic, R. Kühne, J. R. Stout, V. A. Higman, R. E. Klevit, B.-J. J. van Rossum, and H. Oschkinat, "Solid-state NMR and SAXS studies provide a structural basis for the activation of αB-crystallin oligomers.," *Nat. Struct. Mol. Biol.*, vol. 17, pp. 1037–1042, Sept. 2010.

[26] R. L. van Montfort, E. Basha, K. L. Friedrich, C. Slingsby, and E. Vierling, "Crystal structure and assembly of a eukaryotic small heat shock protein.," *Nat. Struct. Biol.*, vol. 8, pp. 1025–1030, Dec. 2001.

[27] K. K. Kim, R. Kim, and S.-H. Kim, "Crystal structure of a small heat-shock protein," *Nature*, vol. 394, pp. 595–599, Aug. 1998.

[28] A. J. Baldwin, G. R. Hilton, H. Lioe, C. Bagneris, J. L. Benesch, and L. E. Kay, "Quaternary dynamics of αB-crystallin as a direct consequence of localised tertiary fluctuations in the C-terminus," *J. Mol. Biol.*, vol. 413, pp. 310–20+, 2011.

[29] S. Y. Pasta, B. Raman, T. Ramakrishna, and M. Rao Ch, "The IXI/V motif in the C-terminal extension of α-crystallins: alternative interactions and oligomeric assemblies," *Mol. Vis.*, vol. 10, pp. 655–62+, 2004.

[30] A. J. Baldwin, H. Lioe, G. R. Hilton, L. A. Baker, J. L. Rubinstein, L. E. Kay, and J. L. Benesch, "The polydispersity of αB-crystallin Is rationalized by an interconverting polyhedral architecture," *Structure*, vol. 19, pp. 1855–63+, 2011.

[31] N. Braun, M. Zacharias, J. Peschek, A. Kastenmuller, J. Zou, M. Hanzlik, M. Haslbeck, J. Rappsilber, J. Buchner, and S. Weinkauf, "Multiple molecular architectures of the eye lens chaperone αB-crystallin elucidated by a triple hybrid approach," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 108, pp. 20491–6+, 2011.

[32] S. Jehle, B. S. Vollmar, B. Bardiaux, K. K. Dove, P. Rajagopal, T. Gonen, H. Oschkinat, and R. E. Klevit, "N-terminal domain of αB-crystallin provides a conformational switch for multimerization and structural heterogeneity.," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 108, pp. 6409–6414, Apr. 2011.

[33] J. A. Carver and R. A. Lindner, "NMR spectroscopy of α-crystallin. Insights into the structure, interactions and chaperone action of small heat-shock proteins," *Int. J. Biol. Macromol.*, vol. 22, pp. 197–209+, 1998.

[34] A. Ghahghaei, A. Rekas, J. A. Carver, and R. C. Augusteyn, "Structure/function studies of dogfish α-crystallin, comparison with bovine α-crystallin," *Mol. Vis.*, vol. 15, pp. 2411–20+, 2009.

[35] A. J. Baldwin and L. E. Kay, "Measurement of the signs of methyl ^{13}C chemical shift differences between interconverting ground and excited protein states by R(1ρ): an application to αB-crystallin.," *J. Biomol. NMR*, 2012.

[36] J. A. Aquilina, J. L. P. Benesch, L. L. Ding, O. Yaron, J. Horwitz, and C. V. Robinson, "Subunit exchange of polydisperse proteins: mass spectrometry reveals consequences of αA-crystallin truncation," *J. Biol. Chem*, vol. 280, pp. 14485–91+, 2005.

[37] L. S. Kallur, A. Aziz, and E. C. Abraham, "C-Terminal truncation affects subunit exchange of human αA-crystallin with αB-crystallin," *Mol. Cell. Biochem.*, vol. 308, pp. 85–91+, 2008.

[38] T. M. Treweek, H. Ecroyd, D. M. Williams, S. Meehan, J. A. Carver, and M. J. Walker, "Site-directed mutations in the C-terminal extension of human αB-crystallin affect chaperone function and block amyloid fibril formation," *PLoS One*, vol. 2, pp. e1046+, 2007.

[39] V. H. Hayes, G. Devlin, and R. A. Quinlan, "Truncation of αB-crystallin by the myopathy-causing Q151X mutation significantly destabilizes the protein leading to aggregate formation in transfected cells," *J. Biol. Chem.*, vol. 283, pp. 10500–12+, 2008.

[40] H. Zhang, N. S. Rajasekaran, A. Orosz, X. Xiao, M. Rechsteiner, and I. J. Benjamin, "Selective degradation of aggregate-prone CryAB mutants by HSPB1 is mediated by ubiquitin-proteasome pathways," *J. Mol. Cell. Cardiol.*, vol. 49, pp. 918–30+, 2010.

[41] R. R. Devi, W. Yao, P. Vijayalakshmi, Y. V. Sergeev, P. Sundaresan, and J. F. Hejtmancik, "Crystallin gene mutations in Indian families with inherited pediatric cataract," *Mol. Vis.*, vol. 14, pp. 1157–70+, 2008.

[42] N. Inagaki, T. Hayashi, T. Arimura, Y. Koga, M. Takahashi, H. Shibata, K. Teraoka, T. Chikamori, A. Yamashina, and A. Kimura, "αB-crystallin mutation in dilated cardiomyopathy," *Biochem. Biophys. Res. Commun.*, vol. 342, pp. 379–86+, 2006.

[43] A. Pilotto, N. Marziliano, M. Pasotti, M. Grasso, A. M. Costante, and E. Arbustini, "αB-crystallin mutation in dilated cardiomyopathies: low prevalence in a consecutive series of 200 unrelated probands," *Biochem. Biophys. Res. Commun.*, vol. 346, pp. 1115–7+, 2006.

[44] L. J. Takemoto, "Changes in the C-terminal region of αA crystallin during human cataractogenesis," *Int. J. Biochem. Cell Biol.*, vol. 29, pp. 311–5+, 1997.

[45] G. R. Hilton and J. L. P. Benesch, "Two decades of studying non-covalent biomolecular assemblies by means of electrospray ionization mass spectrometry," *J. R. Soc. Interface*, 2012.

[46] J. L. Benesch and B. T. Ruotolo, "Mass spectrometry: come of age for structural and dynamical biology," *Curr. Opin. Struct. Biol.*, vol. 21, pp. 641–9+, 2011.

[47] A. J. Heck, "Native mass spectrometry: a bridge between interactomics and structural biology," *Nat. Methods*, vol. 5, pp. 927–33+, 2008.

[48] N. Morgner and C. V. Robinson, "Linking structural change with functional regulation-insights from mass spectrometry," *Curr. Opin. Struct. Biol.*, vol. 22, pp. 44–51+, 2012.

[49] S. Jehle, B. van Rossum, J. R. Stout, S. M. Noguchi, K. Falber, K. Rehbein, H. Oschkinat, R. E. Klevit, and P. Rajagopal, "αB-Crystallin: A Hybrid Solid-State/Solution-State NMR Investigation Reveals Structural Aspects of the Heterogeneous Oligomer," *J. Mol. Biol.*, vol. 385, pp. 1481–1497+, 2009.

[50] B. T. Ruotolo, J. L. P. Benesch, A. M. Sandercock, S. J. Hyung, and C. V. Robinson, "Ion mobility-mass spectrometry analysis of large protein complexes," *Nat. Protoc.*, vol. 3, pp. 1139–52+, 2008.

[51] F. Stengel, A. J. Baldwin, M. F. Bush, G. R. Hilton, H. Lioe, E. Basha, N. Jaya, E. Vierling, and J. L. Benesch, "Dissecting heterogeneous molecular chaperone complexes using a mass spectrum deconvolution approach," *Chem. Biol.*, vol. 19, pp. 599–607+, 2012.

[52] A. J. Baldwin, P. Walsh, D. F. Hansen, G. R. Hilton, J. L. P. Benesch, S. Sharpe, and L. E. Kay, "Probing dynamic conformations of the high-molecular-weight αB-crystallin heat shock protein ensemble by NMR spectroscopy," *J. Am. Chem. Soc.*, vol. under review, 2012.

[53] G. R. Hilton, G. K. Hochberg, A. Laganowsky, S. I. McGinnigle, A. J. Baldwin, and J. L. Benesch, "C-terminal interactions mediate the quaternary dynamics of αB-crystallin.," *Philos. Trans. R. Soc. Lond. B. Biol. Sci.*, vol. 368, May 2013.

[54] H. Hernández and C. V. Robinson, "Determining the stoichiometry and interactions of macromolecular assemblies from mass spectrometry," *Nat. Protoc.*, vol. 2, pp. 715–26+, 2007.

[55] A. Tardieu, "Eye lens proteins and transparency: from light transmission theory to solution X-ray structural analysis.," *Annu. Rev. Biophys. Biophys. Chem.*, vol. 17, pp. 47–70, 1988.

[56] M. P. Bova, L. L. Ding, J. Horwitz, and B. K. Fung, "Subunit exchange of αA-crystallin.," *J. Biol. Chem.*, vol. 272, pp. 29511–29517, Nov. 1997.

[57] S. Meehan, T. P. Knowles, A. J. Baldwin, J. F. Smith, A. M. Squires, P. Clements, T. M. Treweek, H. Ecroyd, G. G. Tartaglia, M. Vendruscolo, C. E. Macphee, C. M. Dobson, and J. A. Carver, "Characterisation of amyloid fibril formation by small heat-shock chaperone proteins human αA-, αB- and R120G αB-crystallins," *J. Mol. Biol.*, vol. 372, pp. 470–84+, 2007.

[58] A. Laganowsky, C. Liu, M. R. Sawaya, J. P. Whitelegge, J. Park, M. Zhao, A. Pensalfini, A. B. Soriaga, M. Landau, P. K. Teng, D. Cascio, C. Glabe, and D. Eisenberg, "Atomic view of a toxic amyloid small oligomer," *Science*, vol. 335, pp. 1228–31+, 2012.

[59] M. F. Bush, Z. Hall, K. Giles, J. Hoyes, C. V. Robinson, and B. T. Ruotolo, "Collision cross sections of proteins and their complexes: a calibration framework and database for gas-phase structural biology," *Anal. Chem.*, 2010.

[60] J. T. Hopper, K. Sokratous, and N. J. Oldham, "Charge state and adduct reduction in electrospray ionization-mass spectrometry using solvent vapor exposure," *Anal. Biochem.*, vol. 421, pp. 788–90+, 2012.

[61] K. Pagel, S. J. Hyung, B. T. Ruotolo, and C. V. Robinson, "Alternate dissociation pathways identified in charge-reduced protein complex ions," *Anal. Chem.*, vol. 82, pp. 5363–72+, 2010.

[62] F. Sobott, J. L. P. Benesch, E. Vierling, and C. V. Robinson, "Subunit Exchange of Multimeric Protein Complexes," *J. Biol. Chem.*, vol. 277, pp. 38921–38929+, 2002.

[63] E. N. Kitova, A. El-Hawiet, P. D. Schnier, and J. S. Klassen, "Reliable determinations of protein-ligand interactions by direct ESI-MS

measurements. Are we there yet?," *J. Am. Soc. Mass Spectrom.*, vol. 23, pp. 431–41+, 2012.

[64] J. L. P. Benesch, B. T. Ruotolo, D. A. Simmons, and C. V. Robinson, "Protein complexes in the gas phase: technology for structural genomics and proteomics," *Chem. Rev.*, vol. 107, pp. 3544–67+, 2007.

[65] J. Sun, E. N. Kitova, W. Wang, and J. S. Klassen, "Method for distinguishing specific from nonspecific protein-ligand complexes in nanoelectrospray ionization mass spectrometry," *Anal. Chem.*, vol. 78, pp. 3010–8+, 2006.

[66] L. A. Lane, B. T. Ruotolo, C. V. Robinson, G. Favrin, and J. L. P. Benesch, "A Monte Carlo approach for assessing the specificity of protein oligomers observed in nano-electrospray mass spectra," *Int. J. Mass Spectrom.*, vol. 283, pp. 169–177, June 2009.

Chapter 3

The minimal chaperone unit of αB-crystallin

3.1 Abstract

Mammalian sHSPs form polydisperse and dynamic complexes with target proteins, serving as a first line of defence in preventing their aggregation into either amorphous deposits or amyloid fibrils. Their apparently broad target specificity makes sHSPs attractive for investigating how to tackle disorders of protein aggregation. The two most abundant sHSPs in human tissue are αB-crystallin (αBC) and HSP27; here we present high-resolution structures of their core domains (cABC, cHSP27), each in complex with a segment of their respective C-terminal regions. We find that both truncated proteins dimerise, and while this interface is labile in the case of cABC, in cHSP27 the dimer can be cross-linked by an inter-monomer disulfide linkage. Using cHSP27 as a template we have designed an equivalently locked cABC to enable us to investigate the functional role played by oligomerisation, disordered N- and C-termini, subunit exchange and variable dimer interfaces in αBC. We have assayed the ability of the different forms of αBC to prevent protein aggregation in vitro, and remarkably find that cABC has chaperone activity comparable to the full-length protein, even when monomer dissociation is restricted through disulphide linkage. Furthermore, cABC is a potent inhibitor of amyloid fibril formation, and, by slowing the rate of its aggregation, effectively reduces the

toxicity of amyloid-β peptide to cells. Overall we present a small chaperone unit together with its atomic coordinates that potentially enable the rational design of more effective chaperones and amyloid inhibitors.

3.2 Introduction

The proteome is inherently metastable [1, 2], and therefore the cell is required to actively maintain protein homeostasis (or proteostasis) through the balancing of a multitude of biochemical pathways [3]. The breakdown of this steady state can lead to a variety of diseases, many of which are characterised by the aggregation and deposition of misfolded proteins [4]. Molecular chaperones, proteins that act to prevent improper polypeptide associations, are crucial components of the cellular proteostasis machinery [5, 6]. They include the small heat-shock proteins (sHSPs), which are found in organisms across all branches of the tree of life and play an important role in preventing protein misfolding and aggregation [7, 8]. In general the sHSPs are capable of intercepting destabilized targets [9], and either: holding them in a refolding competent state; preventing them from aggregating into unrecoverable deposits; or directing them towards degradation [10]. αB-crystallin (αBC) is an abundant mammalian sHSP, the expression of which is constitutive in most human tissues and upregulated in a variety of pathological disorders [11]. The chaperone activity of αBC has been established for over two decades [12], and it is associated with amyloid fibril deposits *in vivo* that are characteristic of protein misfolding diseases including Alzheimer's and Parkinson's diseases [13, 14, 15].

Proteins enter the amyloid cascade from their native state and form insoluble fibrils via various intermediates, including oligomeric forms [16, 17]. Both amyloid fibrils and oligomers are harmful to cells, however the latter appear to be more toxic [18]. αBC has been shown to mitigate amyloid toxicity to cells in culture [19], to interact directly with amyloid oligomers *in vitro* [20], and to prevent the fibrillation of a variety of targets [21, 22, 23, 24, 25]. Additionally

αBC has been shown to bind to mature amyloid-β peptide ($A\beta_{1-42}$) [22, 24], α-synuclein [23, 25], and apolipoprotein C-II (apoC-II) fibrils [26], apparently coating them and preventing their elongation [21]. An understanding of how αBC carries out these activities has been hindered by the structural and dynamical complexities of this chaperone. αBC, as is typical for most metazoan sHSPs, consists of a dimeric building block that assembles via terminal interactions into a polydisperse ensemble [8]. In αBC, these oligomers range from approximately 10 to 40 subunits and readily interconvert via the exchange of monomers [27], in a process that facilitates the formation of heterooligomers between different sHSPs [28]. While several new models have recently been developed for αBC oligomers, a consensus as to their quaternary structure remains to be reached [8, 29].

The sequence of αBC can be divided into an immunoglobulin-like α-crystallin domain (ACD) that mediates dimerization, and is flanked by N- and C-terminal regions that are poorly conserved and variable in length between sHSPs [7, 8]. Various regions of the αBC sequence have been implicated as potential binding sites, but there is little consensus [30, 31, 32, 33, 34]. Moreover, it even remains unclear whether the polydisperse oligomeric ensemble of αBC, the oligomeric dissociation that mediates subunit exchange, or remodelling of the dimer interface is responsible for chaperone function [8]. Models have been proposed wherein sub-oligomeric forms, which are in equilibrium with the assembled state, are the active chaperoning unit. This idea is based on the observation that solution conditions that accelerate subunit exchange also lead to increased chaperone activity [35]. Conversely, other studies have demonstrated that αBC can still function as a chaperone despite being cross-linked as an oligomer [36], and that mutations that slow its subunit exchange kinetics do not necessarily diminish its activity as a chaperone *in vitro* [37].

These apparent conflicts likely stem from the intrinsic heterogeneity of αBC and the difficulty in assessing the interplay between its structural and dynamical aspects [8, 29]. Here we have gained insight into the chaperone activity of

αBC, and the prevention of protein aggregation in general, by establishing a minimal, yet chaperone-active, unit of αBC that is suitable for structural studies. We drew inspiration from recent structures of the ACD that have revealed the αBC dimer interface to be composed of paired β6+7 strands [38, 39, 40, 41]. Interestingly, their anti-parallel interaction has been crystallized in two different registers, termed AP_I and AP_{II} [41], with a third (AP_{III}) having been found for the related HSP20 [38]. The three registrations have different amounts of buried surface area in the dimer interface, resulting from monomers being progressively shifted outward relative to one another in AP_{II} and AP_{III} relative to AP_I. We have engineered a construct comprising exclusively the core domain of αBC (cABC), which allowed us to test the relationship between oligomeric state, subunit exchange, registration state, and chaperone function of αBC.

Our experimental strategy combines X-ray crystallography with native mass spectrometry (MS) and nuclear magnetic resonance spectroscopy (NMR) to examine the structure of cABC. Using a novel approach to crystallise the protein we show that cABC can populate three different registers, and that the AP_{II} form is predominantly populated in solution. Through comparison with a new structure of the HSP27 core domain (which we also present here), we have engineered a cysteine mutant of cABC ($cABC_{E117C}$) that can be locked into a dimeric state in the AP_{II} registration. This protein is therefore unable to exchange monomers under oxidizing conditions. We show that both cABC and $cABC_{E117C}$ can strongly inhibit amorphous aggregation, amyloidogenesis and amyloid toxicity, all as effectively as the wild-type protein. Together our data demonstrate that the core domain of αBC is responsible for its potent molecular chaperone function, which is retained regardless of stoichiometry or registration state of the dimer.

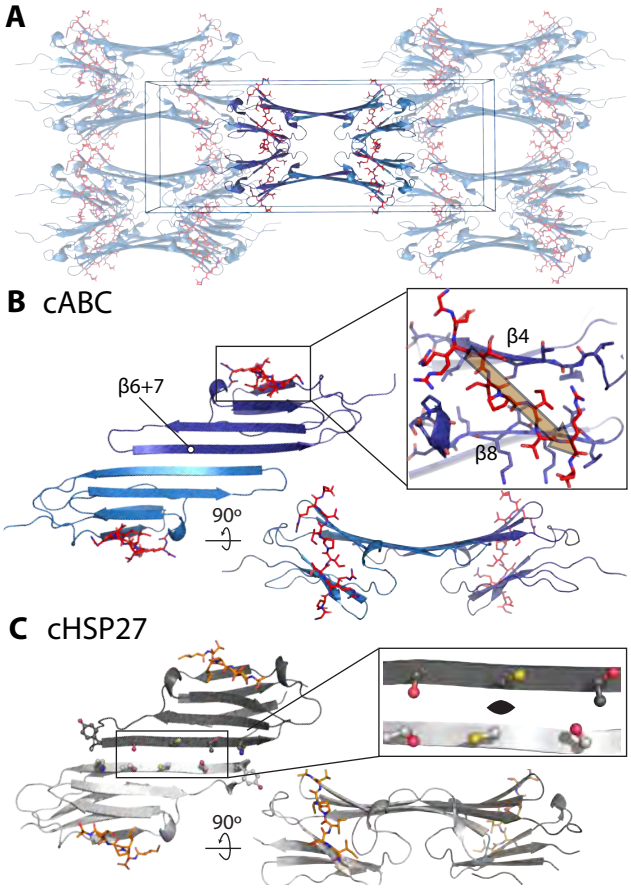


Figure 3.1: Crystal structures of cABC and cHSP27. A) cABC crystallizes as a dimer of dimers, with one C-terminal peptide of sequence ERTIPITRE (red) bound to each monomer. B) Two immunoglobulin-like cABC monomers assemble into a dimer through pairwise and anti-parallel interactions between extended β6+7 strands. In this structure the dimer is found in the AP_{III} register. The palindromic C-terminal peptide binds to a hydrophobic groove between β4 and β8 strands, in an anti-parallel direction to the β8 strand (inset, direction N to C illustrated by yellow arrow). C) The crystal structure of cHSP27 reveals a dimer similar to cABC, rich in β sheet structure and with C-terminal peptides bound. cHSP27 however is in the AP_{II} register, with C137 (thiol coloured in yellow) located about a two-fold axis at the dimer interface (inset). C137 is reduced in the structure due to the presence of reductant during crystallisation, but can readily be oxidised (Fig. 3.7). Both structures presented here were solved by Art Lanagowksy.

3.3 Results

3.3.1 An alternative strategy to crystallise mammalian sHSPs

In a novel approach to crystallise mammalian ACDs, we designed a system composed of a core unit of αBC (residues 68-153, cABC), and a peptide mimicking its C-terminal region (residues 156-164, ERTIPITRE), to avoid the runaway domain swapping that resulted in a polymer-like crystal array in our previous study [41]. Co-crystallisation of cABC and the peptide produced crystals that led to structure determination at 1.35 Å (Table 3.1). The structure of cABC reveals a crystal packing of tetrameric units, assembled essentially as a dimer of dimers (Fig. 3.1A). One peptide is bound to each cABC monomer in an orientation anti-parallel to the β8 strand (Fig. 3.1B), the inverse of that in our previous structure of αBC (41). This bi-directionality is enabled by the palindromic nature of the peptide [42], and is consistent with the binding observed in NMR experiments [43]. The two monomers form a dimer interface in which they are slightly bent relative to each other (Fig. 3.1A), reminiscent of a structure of αBC obtained by means of solid-state NMR [40] (Fig. 3.2). The angle between monomers is however larger in our structure, reflecting a flatter interface, and we also observe no intrinsic twist of the monomers (Fig. 3.2A). Notably, we find that the dimer is in the AP_{III} register, a state previously only observed in the structure of HSP20 [38]. A comparison of the registers for αBC reveals the AP_{III} structure to have 685.1 $Å^2$ of buried surface area, compared to 694.1 in AP_{II} (PDB ID: 2WJ7) and 820.7 $Å^2$ in AP_I (3LIG). Together with structures of truncated αBC constructs in the AP_I (41) and AP_{II} states [38, 39, 40], this reveals that αBC can populate multiple dimer interface registers that differ in their exposed surface area.

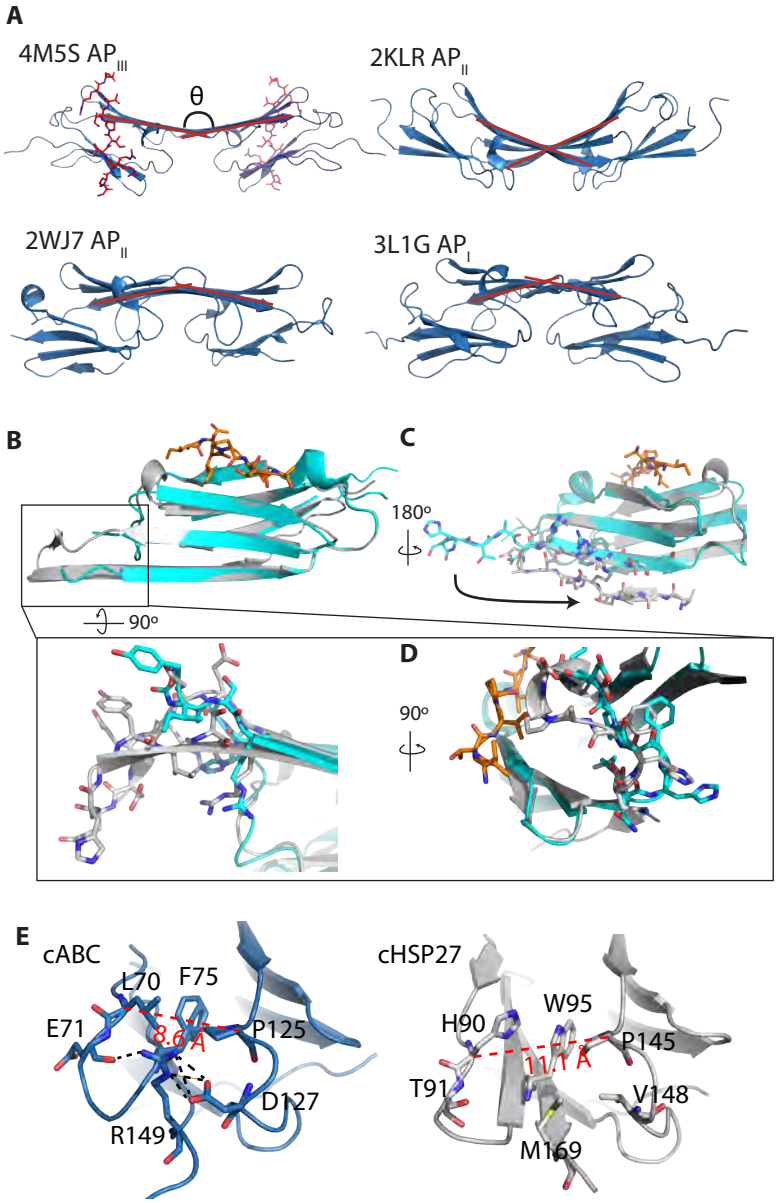


Figure 3.2: A) Comparison of X-ray and solid-state NMR structures of sHSP ACD structures. The solid-state NMR structure of αBC (PDB ID: 2KLR) has a bent dimer interface, which is partly mirrored in our AP_{III} structure (4M5S). AP_{II} (2WJ7) and AP_{I} (3L1G) crystal structures show flatter dimer interfaces. Red lines indicate the angle between the two β-strands forming the dimer interface. B) Overlay of our cHSP27 structure (4MJH, grey, peptide in orange) with a previous structure that crystallized as a crystallographic hexamer (3Q9P, cyan). Inset is the region around the β5-β6+7 loop that is ordered in our structure by a longer β6+7 β-strand that forms part of the dimer interface. C) The longer N-terminus in our construct allows formation of the β2 strand. D) Differences in the β3-β4 loop between our structure and 3Q9P. E) Repacking in the β-sandwich in cHSP27 compared to cABC. In aligning cABC to cHSP27 we find that F75 is replaced by the bulkier W95, L70 by H90 which disrupts van der Waals interactions in the core, and R149 by M169 which abolishes salt-bridges to both sides of the β-sandwich (specifically to D127 and carbonyl of E71).

3.3.2 The core domain dimer of HSP27 has cysteines located about a two-fold axis

In order to interrogate the potential role of these different registers we drew inspiration from HSP27, a related human sHSP that has a cysteine residue (C137), located within the β6+7 strand, that can readily be oxidised [44]. By employing our co-crystallisation strategy we succeeded in determining the structure for the HSP27 ACD (residues 86-169, cHSP27) in complex with its C-terminal peptide (residues 179-185, ITIPVTF) (Table 3.1). We find that cHSP27 forms a canonical dimer, in an AP_{II} register, and with the C-terminal peptide bound in an anti-parallel orientation relative to the β8 strand (Fig. 3.1C). Our structure superimposes well with a previous one in which monomers assembled into a non-canonical hexamer within the crystal lattice (backbone RMSD = 0.45 Å^2 comparing monomers, Fig. 3.2B), and the registration state is consistent with small-angle X-ray scattering data [45]. The structures are slightly different around the loop between β5 and β6+7 strands, which contained several point mutations in the previous structure and was disordered. We find this region to be ordered in an extended conformation along the β6+7 sheet of the dimer interface. This can be explained by our construct including additional N-terminal residues, allowing the formation of a β2 strand consistent with structures obtained for other metazoan ACDs (Fig. 3.2C). Differences in the primary sequences of cABC and cHSP27 are distributed over the entire domain, with slightly more disparity observed in flexible loops between successive β-strands (Fig. 3.3A). While the monomeric fold is nearly identical in the two proteins (Fig. 3.3B), in cHSP27 the β-sandwich is splayed outward by ≈2.5 Å. This can be attributed to the combined effect of several substitutions (Fig. 3.2E). The most striking difference, however, is found for residues located next to C137 on the β6+7 strand in HSP27. In cABC there is a charge network of two histidines (H101 and H119) and two glutamates (E99 and E117) that implies pH sensitivity (Fig. 3.3B,C) [39, 40]. In cHSP27, these histidines are re-

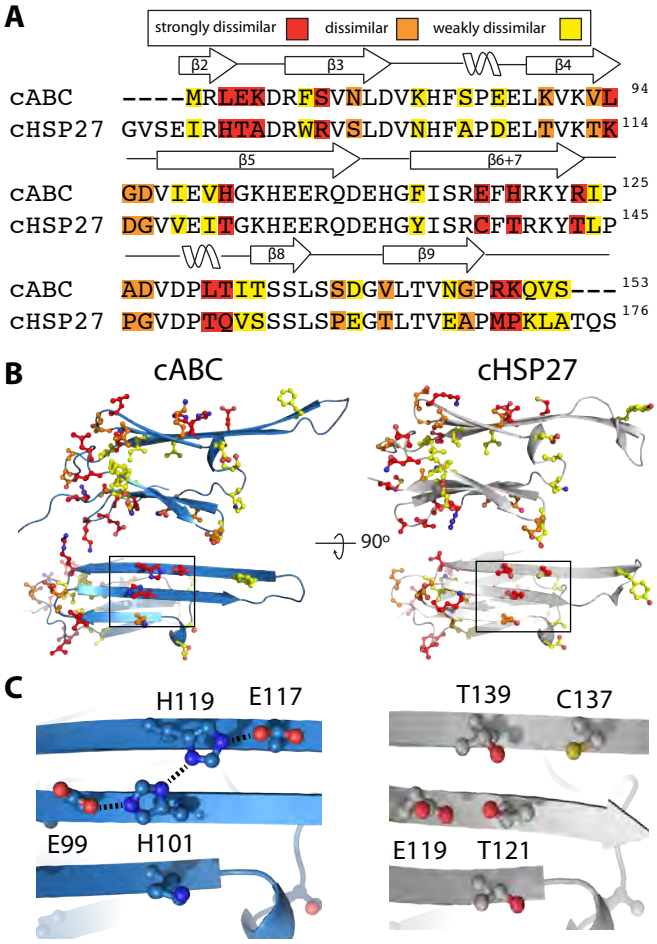


Figure 3.3: Structural differences between cHSP27 and cABC. A) Sequence alignment of the two domains coloured to highlight differences in amino-acid composition (strongly dissimilar, red; dissimilar, orange; weakly dissimilar, yellow, see methods). The two domains are clearly highly similar (54.7 % identity), though differences are distributed throughout the sequence. B) Mapping the disparate residues on the cABC (left) and cHSP27 (right) structures (colouring as in A) reveals them to be spread over the entire structure. C) An expansion of the boxed region in B reveals a charge network formed by residues E99, H101, E117, and H119 in cABC (left) that is absent in cHSP27 (right) where the equivalent residues are E119, T139, T121 and C137.

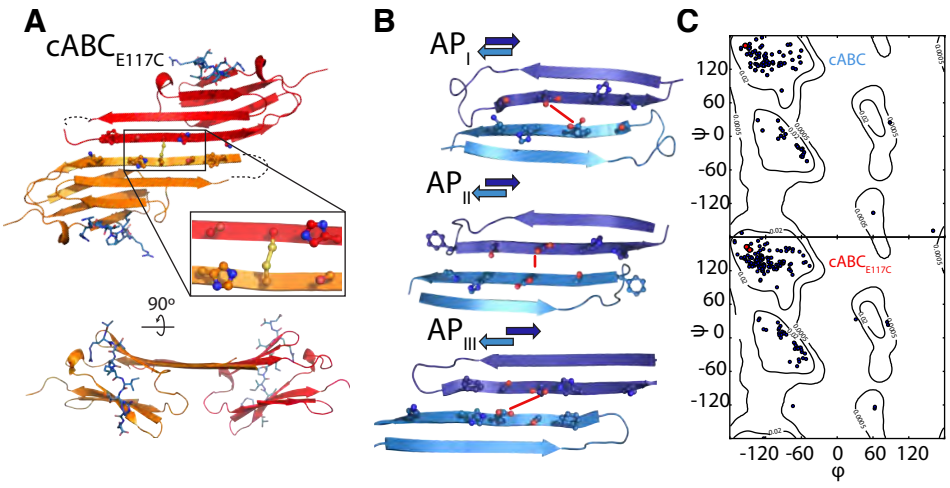


Figure 3.4: A) Crystal structure of cABC$_{E117C}$ in which the introduced cysteine acts to lock the domain into a dimer in the AP$_{II}$ register. The disulphide bond formed between two monomers is shown in the inset. The overall structure of this engineered ACD is closely similar to that of cABC. B) The three registers of the αBC dimer observed by X-ray crystallography: AP$_I$ (PDB ID: 3L1G), AP$_{II}$ (2WJ7), and AP$_{III}$ (present work, 4M5S). Red lines indicate the vector between α-carbons of residues E117 on the two monomers, which is located on a two-fold axis in AP$_{II}$. The distance between two modelled cysteines at position 117 is close enough for disulphide bond formation only in AP$_{II}$ (9.13 Å between the two thiols in AP$_I$, 0.93 Å in AP$_{II}$, and 6.5 Å in AP$_{III}$). C) C117 is not strained in cABC$_{E117C}$. Ramachandran plots for cABC (left) and cABC$_{E117C}$ (right). Red dots correspond to residue 117, which has normal β-sheet ϕ and ψ angles. Numbers on contour lines correspond to probabilities. The χ1 (65.37), χ2 (-33.22),χ3 (70.76), χ2', (-44.41), and χ1' (58.45) dihedral angles along the disulphide bond are within the range of unstrained cysteines.

placed by threonines and E117 by C137, resulting in a disruption of the charge network (Fig. 3.3B,C) . Moreover, our structure reveals that C137 is located on a central two-fold axis about the dimer interface, and is in a conformation compatible with disulphide bond formation. These differences are consistent with the observation that the αBC dimer interface is pH sensitive [27, 40], and HSP27 activity shows redox dependence [46].

3.3.3 An AP$_{II}$ locked cABC dimer

Motivated by the cHSP27 structure we engineered a cysteine into cABC at position E117, which aligns to C137 in HSP27 (Fig. 3.3A), and is located

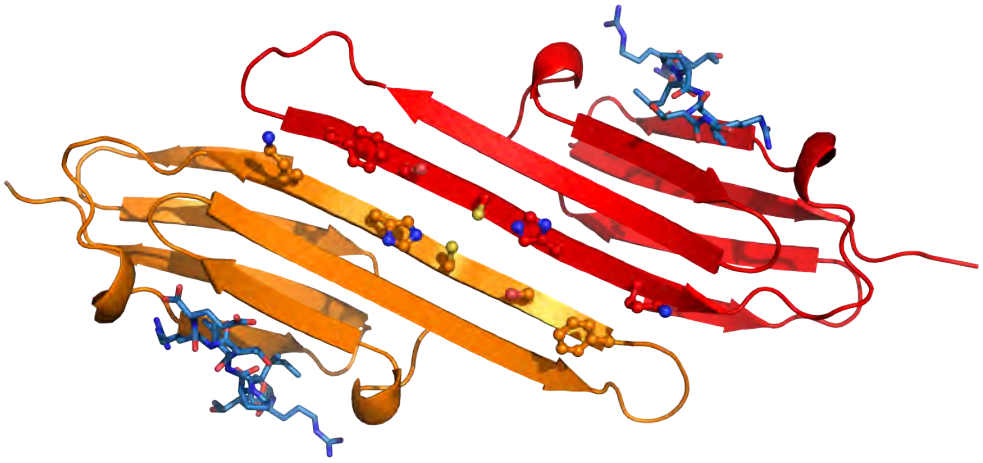


Figure 3.5: Crystal structure of the second cABC$_{E117C}$ dimer in the AU. In this dimer the two cysteines are reduced, and the loops between β5 and β6+7 are ordered.

about the two-fold axis in the AP$_{II}$ register (Fig. 3.1C). We crystallized this cABC$_{E117C}$ construct using our co-crystallisation method (Table 3.1), revealing a fold very similar to cABC and comparable curvature of the dimer interface (Fig. 3.4A). However, in one of the two dimers comprising the asymmetric unit, this dimer was covalently locked by a disulphide bond between monomers rendering it unable to access registers other than AP$_{II}$ (Fig. 3.4B). In the locked dimer, the loops between the β5 and β6+7 strands are disordered in the crystal lattice, but are ordered in the other dimer (Chains E and G) (Fig. 3.5). There is near-perfect agreement in side-chain positions between cABC and cABC$_{E117C}$ (backbone and side-chain RMSD = 0.33 Å^2). The two structures only differ significantly in that the β5 and β6+7 strands are elongated in the AP$_{III}$ register of cABC. The disulphide bond in the cABC$_{E117C}$ structure does not lead to unusual torsion angles in the $\beta-$sheet around the bond, nor in the dihedral with the bond (Fig 3.4C). While the introduction of strain into anti-parallel β-sheets by disulphide bonds between register-paired cysteines has been proposed as a mechanism of redox regulation [47], our cABC$_{E117C}$ structure hints that redox regulation in HSP27 is achieved via another mechanism.

3.3.4 cABC predominantly populates the AP_{II} register

To assess whether $cABC_{E117C}$ adopts the same monomeric fold as cABC in solution we recorded ^{1}H-^{15}NHSQC NMR spectra for both cABC and $cABC_{E117C}$ under oxidising conditions. In both cases, well-dispersed resonances were observed indicative of a folded structure and consistent with data for similar constructs [43, 48]. For most cross-peaks in the cABC spectrum, there is a corresponding cross-peak for oxidized $cABC_{E117C}$ (Fig. 3.6A). Additional cross-peaks are visible in the $cABC_{E117C}$ spectrum that likely result from differences in subunit exchange dynamics between cABC and $cABC_{E117C}$ [49]. To assess the magnitude of any changes we determined the change in cross-peak position in both the ^{1}H and ^{15}N dimension (chemical-shift perturbation, CSP) between cABC and $cABC_{E117C}$ based on a previous assignment [48] (Fig. 3.6B). In most cases the CSPs are < 0.16 ppm, which is considerably smaller that those observed in experiments probing the binding of ligands to αBC [43]. The two largest CSPs (≈0.7 ppm) are E117 (the site of mutation) and H119, which is hydrogen-bonded to E117 in our cABC structure. Mapping these small CSPs onto the structure of cABC reveals them to cluster around the site of mutation (Fig. 3.6B), as expected due to the change in chemical environment caused by the substitution of side-chains. The monomeric folds of cABC and $cABC_{E117C}$ in solution are therefore extremely similar, despite $cABC_{E117C}$ being locked into the AP_{II} registration.

To investigate which register cABC populates at equilibrium, we employed ion mobility spectrometry (IM), a technique that reports on the overall size of the molecule in terms of a rotationally averaged collisional cross section (CCS) [50]. Theoretical calculations based on the crystal structures of the different forms (Fig. 3.4B) predict AP_I and AP_{III} to differ from AP_{II} in CCS by -4.5 % and 15.8 %, respectively (Fig. 3.6C, dashed lines). Comparison of IM data for cABC and $cABC_{E117C}$ under oxidizing conditions however reveals no noticeable differences in CCS distributions (Fig. 3.6C), even though the disulphide-locked

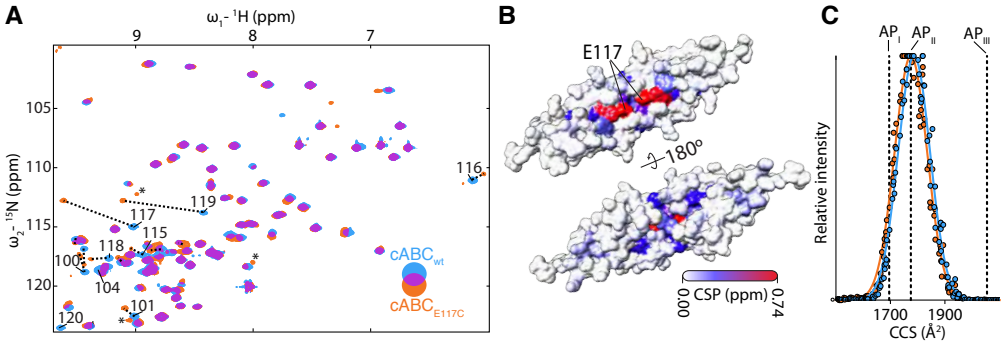


Figure 3.6: cABC predominantly populates AP_{II}. A) ^{1}H-^{15}N-HSQC spectra of cABC (blue) $cABC_{E117C}$ (orange) acquired under identical and oxidizing solution conditions. Overlap of peaks is represented by purple. Peaks with significant shifts in the mutant are labelled on the plot. Dotted lines indicate peak movement in the mutant compared to cABC, and asterisks unassigned peaks. The spectra overlay very well, indicating that the fold of the two proteins is very similar. B) A heat map of CSP projected onto the structure of cABC (largest CSP, red; lowest, white) reveals that the most significant changes in chemical shift are observed near C117. C) IM measurements of the cABC (blue) and $cABC_{E117C}$ (orange) 7+ charge state (see Fig.3.7E,F) under oxidizing conditions reveal very similar CCS distributions. Dashed lines indicate anticipated CCS values of AP_I, AP_{II}, and AP_{III}. As the CCS distribution of cABC matches that of $cABC_{E117C}$, which is fixed as an AP_{II} dimer, we can infer that cABC preferentially populates the AP_{II} register.

$cABC_{E117C}$ is unable to access either AP_I or AP_{III}. This, combined with the similarity of monomeric folds of cABC and $cABC_{E117C}$, indicates that, outside of a crystal lattice, cABC exists predominantly in the AP_{II} register.

3.3.5 The dynamics of $cABC_{E117C}$ under reducing conditions are equivalent to cABC

In order to interrogate the dynamics of cABC we performed nanoelectrospray MS measurements under conditions that preserve non-covalent interactions in vacuum [51]. A mass spectrum obtained for cABC features two charge-state distributions, centred on 2000 m/z and 2500 m/z, which partially overlap and correspond to monomers and dimers, respectively (Fig. 3.7A). At this concentration (8 μM, based on the molar mass of a monomer) a substantial amount of monomer is observed. Increasing the concentration of cABC to 32 μM results in an increasing abundance of dimers (Fig. 3.7B), consistent with a K_D

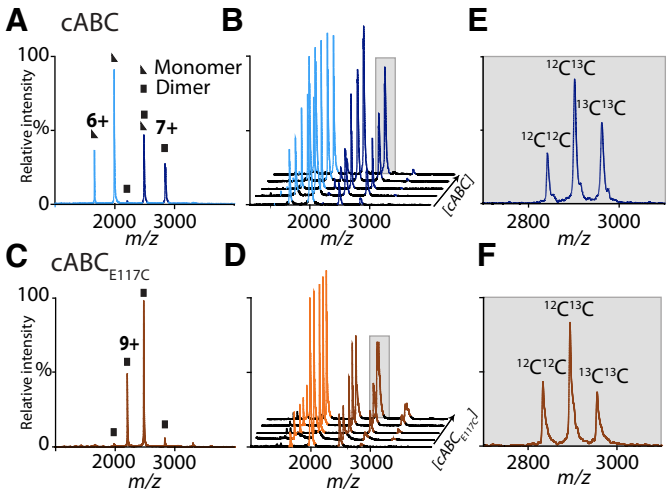


Figure 3.7: : Quaternary dynamics of cABC and $cABC_{E117C}$. A) Nanoelectrospray MS of cABC reveals two charge state distributions corresponding to an equilibrium of monomers (triangles) and dimers (squares). B) A titration series of cABC reveals an increase in the abundance of dimer (dark blue) relative to monomer (light blue) as the concentration is increased (2, 4, 8, 16 and 32 μM, front to back). Spectra are normalized such that the most intense peak in each spectrum is equal to 100%. The shaded area indicates the 7+ charge state of the dimer. C) A mass spectrum of $cABC_{E117C}$, obtained under the same conditions as for cABC in A, reveals the exclusive presence of dimers due to disulphide bond formation. D) In the presence of reductant $cABC_{E117C}$ reverts to a concentration-dependent equilibrium of monomers (orange) and dimers (brown) (concentrations as in B) E) A mass spectrum of unlabelled cABC mixed with ^{13}C-labelled cABC, acquired as soon as possible, and focussing solely on the 7+ charge state (shaded in C). Both homo- and heterodimers are observed, indicating rapid subunit exchange. F) An equivalent experiment to that in E carried out with $cABC_{E117C}$ under reducing conditions demonstrates fast subunit exchange as observed for cABC.

on the order of a few μM [52]. $cABC_{E117C}$ under oxidizing conditions only forms dimers, in line with the disulphide bond formation between monomers observed in the crystal structure (Fig. 3.7C). Upon the addition of reductant, $cABC_{E117C}$ immediately reverts to a monomer-dimer equilibrium with a K_D comparable to that of cABC (Fig. 3.7D).

To characterize the kinetics of the monomer-dimer equilibrium in cABC, we performed experiments wherein mass spectra of cABC and a ^{13}C-labelled equivalent were acquired as soon as possible after mixing at 4 °C. Examination of the region of the spectrum corresponding to dimers reveals $^{12}C^{12}C$- and

$^{13}C^{13}C$-homodimers, and $^{12}C^{13}C$-heterodimers of cABC. Notably, these three dimeric forms are observed at a 1:1:2 ratio (Fig. 3.7E), as would be expected statistically upon complete equilibration of this equimolar mixure. As such, complete exchange is reached on a timescale faster than the dead-time of the experiment. This corresponds to a lower limit for the off-rate constant of a monomer from the dimer of 0.1 s^{-1} , approximately five orders of magnitude faster than a monomer dissociating from an oligomer of αBC at the same temperature [27]. The equivalent experiment performed for $cABC_{E117C}$ under oxidising conditions shows no subunit exchange, even after prolonged incubation, as anticipated due to the covalent linkage between homodimers. In the presence of reductant, $cABC_{E117C}$ exchanges subunits as rapidly as cABC (Fig. 3.7D, F). Combined, these results demonstrate that $cABC_{E117C}$ is a redox-sensitive protein that exists as a stable cross-linked dimer under oxidising conditions, and immediately reverts to a monomer-dimer equilibrium with extremely fast subunit exchange dynamics, comparable to those of cABC, upon the addition of reductant.

3.3.6 cABC can prevent amorphous protein aggregation *in vitro*

The cABC system represent an excellent means to probe the chaperone activity of the core domain itself, and to study the importance of the monomer-dimer equilibrium and its associated subunit-exchange dynamics. In order to investigate this, we assayed the ability of cABC, $cABC_{E117C}$ and full-length αBC to protect three very different targets: α-lactalbumin (α-lac); κ-casein; and the amyloid β-peptide ($A\beta_{1-42}$). These polypeptides differ in their amino-acid sequence, molar mass, mechanism and rate of aggregation, and morphology of the resulting aggregates. We optimised the chaperone:target molar ratios such that our assays exhibited measurable aggregation of the targets, in order to be sensitive to small differences in chaperone activity.

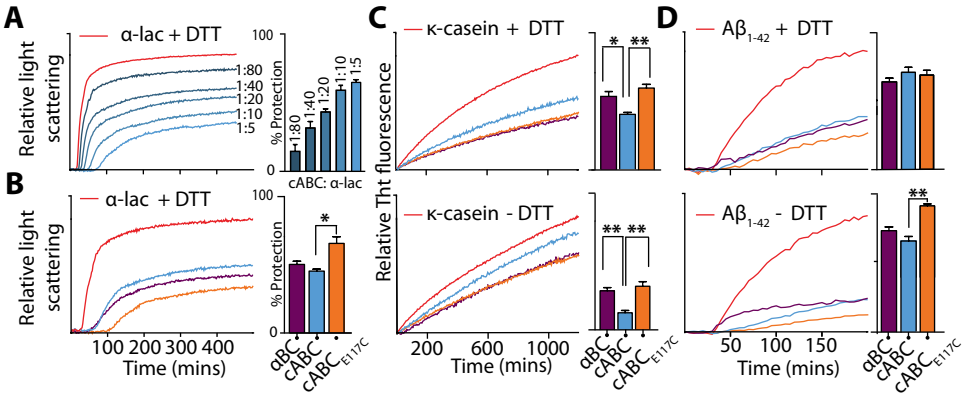


Figure 3.8: *In vitro* chaperone activity of ACDs. A) Dose-dependent inhibition of reduction-induced α-lac aggregation by cABC at pH 7. Molar ratios are indicated as cABC:α-lac. Even at sub-stoichiometric quantities of cABC, significant chaperone activity is observed. B) Comparing different constructs of αBC at a molar ratio of 1:5 (chaperone:α-lac) reveals cABC to be equivalently potent to the full-length protein. cABC_{E117C} is slightly more effective than cABC, likely due to altered surface properties around the introduced cysteine (Fig. 3.9). C) Assaying κ-casein fibrillation under reducing (upper) and oxidising (lower) conditions in the presence of our αBC constructs at a molar ratio of 1:2 (chaperone: κ-casein) demonstrates that they are capable of slowing aggregation. cABC is less efficient than the full-length protein in both conditions, while cABC_{E117C} is more effective than cABC, mirroring the data in B. D) The αBC constructs are also potent inhibitors of $A\beta_{1-42}$ fibrillation at a molar ratio of 1:20 (chaperone:$A\beta_{1-42}$). Under reducing conditions (upper), all three constructs perform equivalently, while under oxidising conditions (lower), cABC_{E117C} is more effective than cABC. This suggests that locking into an AP_{II} dimer improves chaperone action in this assay. In all cases (A-D), a representative aggregation assay is shown as well as the percentage protection. Error bars correspond to mean $\pm$ standard error of the mean, with n=3. * $p < 0.05$, ** $p < 0.01$. The assays in this figure were carried out by Heath Ecroyd.

We first monitored the ability of the different constructs to inhibit the reduction-induced amorphous aggregation of α-lac at 37 °C. In the absence of chaperone, after a lag phase of about 20 minutes, we observed a rapid increase in apparent absorbance due to light scattering (Fig. 3.8A, red), indicating the aggregation of α-lac [53]. Upon addition of cABC, we found a considerable delay to the onset and reduction in the rate of α-lac aggregation (Fig 3.8A, blue). The extent of protection conferred was strongly dependent on the amount of chaperone, though significant protection ($p < 0.01$) was observed even at molar ratios as low as 1:80 (cABC:α-lac). This indicates that cABC affords po-

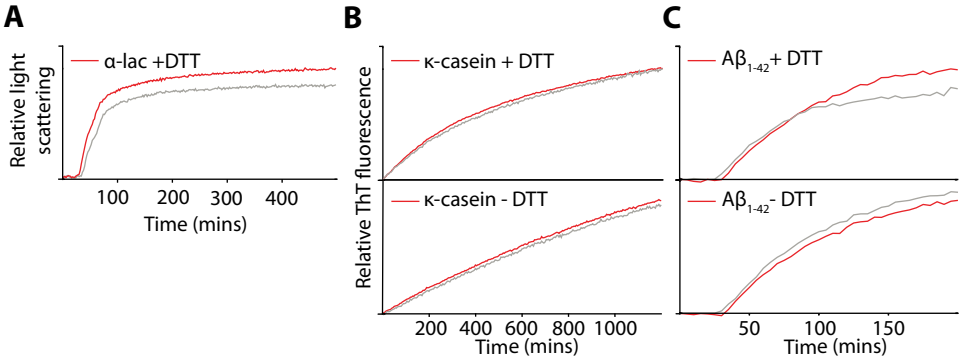


Figure 3.9: Chaperone activity of cHSP27, equivalent to data presented for the αBC constructs in Fig. 3.8. Aggregation traces without chaperone are shown in red, while those with chaperone are shown in grey. cHSP27 is ineffective in protecting against the aggregation of A) α-lac, B) κ-casein, and C) $A\beta_{1-42}$. Molar ratios are the same as in the cABC experiments (see Fig. 3.8).

tent protection against amorphous protein aggregation, at sub-stoichiometric amounts, and in a dose-dependent manner.

In order to quantify the relative efficacy of protection relative to full-length αBC and cABC$_{E117C}$, we performed a comparative assay at a ratio of 1:5 (chaperone:α-lac). All three chaperones significantly delayed and slowed α-lac aggregation (Fig. 3.8B). Remarkably, comparison of cABC and full-length αBC reveals no significant difference in their activity. Notably cHSP27, despite being of similar fold to cABC, displayed minimal chaperone activity in this and other assays (Fig. 3.9). As such it can be regarded as a negative control, and demonstrates the specificity of the chaperone action observed for the constructs of αBC.

In this assay, we find therefore that the ACD in αBC is entirely sufficient for chaperone function (Fig. 3.8B). Notably, we observe that cABC$_{E117C}$ is more effective at preventing aggregation of α-lac than cABC. It is important to consider that the presence of DTT in this assay means that the disulphide bond in cABC$_{E117C}$ is reduced (Fig. 3.7F), so this increased activity cannot be ascribed to the exclusive presence of an AP$_{II}$-locked dimer. Instead we observe that cABC$_{E117C}$ displays increased binding, relative to cABC, of the hydropho-

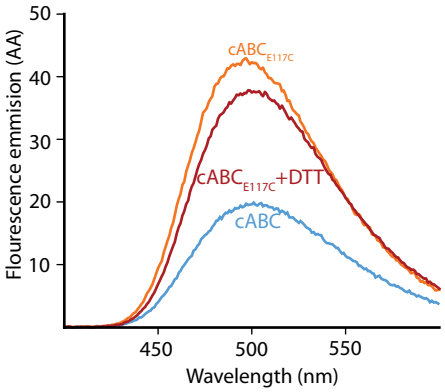


Figure 3.10: Mutating E117 in cABC to a cysteine increases the exposed hydrophobic surface area of cABC$_{E117C}$. A) bis-ANS binding to core domain forms of αBC reveals that cABC$_{E117C}$ displays higher bis-ANS fluorescence than cABC under both oxidizing and reducing conditions, suggesting an increased exposed hydrophobic surface in cABC$_{E117C}$.

bic probe bis-ANS (4,4'-dianilino-1,1'-binaphthyl-5,5'-disulfonic acid), revealing an increased solvent-accessible hydrophobic surface area (Fig. 3.10). The slight differences in structure between cABC and cABC$_{E117C}$ revealed by our NMR experiments (Fig. 3.6A) provide one possible rationale for the difference in chaperone activity.

3.3.7 cABC can prevent amyloid fibril formation *in vitro*

We next tested the ability of cABC to prevent amyloid fibril formation using two model systems: κ-casein and $A\beta_{1-42}$. In these experiments, fibril formation is assayed by monitoring the characteristic red-shift in fluorescence of the dye thioflavin T (ThT) upon binding to amyloid structures. In the absence of chaperone, κ-casein forms fibrils relatively slowly (over a period of $>$ 20 hours), without a significant lag phase (Fig. 3.8C). The timescale of this aggregation is not significantly altered with the presence (upper panel) or absence (lower panel) of the reductant DTT. At a molar ratio of 1:2 (chaperone: κ-casein), αBC slows κ-casein fibril formation under both oxidising and reducing conditions (Fig. 3.8C). At an equivalent ratio, cABC also retards amyloidogenesis, and only slightly less effectively than the full-length protein. Notably, cABC$_{E117C}$ is more effective at slowing κ-casein aggregation than cABC. This improved activity is observed under both oxidising and reducing conditions, in other words independent of whether the chaperone is locked into AP$_{II}$ dimers. This suggests that the difference in activity in this case is due to the altered

surface properties of cABC$_{E117C}$ compared to cABC (Fig. 3.6A, Fig. 3.10), as observed in our α-lac assay (Fig. 3.8B).

In order to test our constructs in preventing the aggregation of a target that forms fibrils more rapidly, we assayed their ability to inhibit $A\beta_{1-42}$ amyloidogenesis. In the absence of chaperone, this peptide aggregates over a period of approximately 3 hours, both in the presence and absence of DTT (Fig. 3.8D). At a molar ratio 1:20 (chaperone:$A\beta_{1-42}$), and in both conditions, we found that all three constructs dramatically slowed aggregation. Notably, no significant difference was observed for cABC relative to the full-length protein (Fig. 3.8D). Under reducing conditions (upper panel), cABC and cABC$_{E117C}$ have equivalent chaperone ability in this assay. However, under oxidising conditions (lower panel), cABC$_{E117C}$ is more effective at inhibiting $A\beta_{1-42}$ fibril formation than cABC. Therefore, locking cABC into an AP$_{II}$ dimer makes it a more potent chaperone in this assay.

3.3.8 The interaction of cABC with aggregating $A\beta_{1-42}$ reduces toxicity to cells

To test whether the inhibitory effect of cABC on $A\beta_{1-42}$ amyloid formation would result in a decrease in toxicity we performed assays on cultured cells. Specifically we monitored the viability of HEK293, HeLa, and PC12 cells upon the addition of pre-incubated $A\beta_{1-42}$ (Fig. 3.11). In all three cell lines, the addition of $A\beta_{1-42}$ resulted in a decrease in viability to about 40%, indicating toxicity of the aggregating peptide. Addition of solutions of $A\beta_{1-42}$ that had been incubated in the presence of chaperone showed a reduction in this toxicity. Notably, not all cell lines responded equivalently, with both cABC and cABC$_{E117C}$ being more effective at rescuing HEK293 and HeLa cells than PC12 cells. While the reasons for this are not known, it is clear that the protective effect was more pronounced with increasing ratios of chaperone:$A\beta_{1-42}$, demonstrating a dose-dependency in all cases. This mirrors data obtained for full-length αBC [19, 20], and reveals that the ACD is sufficient to slow the

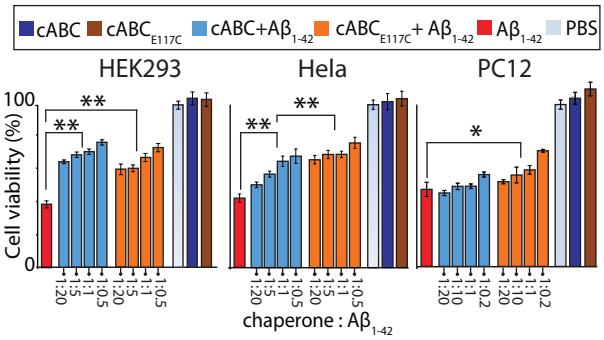


Figure 3.11: Cell viability assays upon addition of $A\beta_{1-42}$ to HEK293, HeLa, and PC12 cells. $A\beta_{1-42}$ on its own (red) causes a decrease in viability to 40 %, while cABC (dark blue), cABC$_{E117C}$ (brown), or buffer (grey) have no effect. $A\beta_{1-42}$ incubated in the presence of increasing amounts of cABC (blue) and cABC$_{E117C}$ (orange) is less toxic than in the absence of chaperone. Molar ratios indicated are chaperone:$A\beta_{1-42}$, with an $A\beta_{1-42}$ concentration of 0.5 μM in all experiments (except the controls without $A\beta_{1-42}$). In all cases a clear concentration dependence to chaperone protection is observed. Error bars correspond to mean ± standard error of the mean, with n=4. * p< 0.05. The data presented in this figure were produced by Cong Liu.

rate of $A\beta_{1-42}$ aggregation and thereby reduce the toxicity of the aggregating mixture. Interestingly, in these assays that are performed in the absence of reductant, we detect a clear increase in potency for cABC$_{E117C}$ relative to cABC in HeLa cells, and in PC12 only cABC$_{E117C}$ displayed statistically significant potency. This implies that locking cABC into an AP$_{II}$ dimer, which improves its efficacy in slowing the rate of $A\beta_{1-42}$ *in vitro* (Fig. 3.8D, lower), results in an associated reduction in toxicity of the aggregating mixture.

3.4 Discussion

Here we have found that the core domain of αBC is a potent inhibitor of both amorphous and fibrillar aggregation, and amyloid-associated toxicity. Previous reports describing alternatively truncated forms of αBC have shown limited chaperone activity towards other target proteins [41, 54]. However, our construct performed comparably to full-length αBC with regards to protecting α-lac and κ-casein from aggregation at sub-stoichiometric levels. Remarkably,

our amyloidogenesis and cell toxicity assays of $A\beta_{1-42}$ places cABC among the most potent known inhibitors of $A\beta_{1-42}$ on a molar basis [55]. The finding that the ACD can prevent amyloid formation and toxicity is surprising, and raises questions about the role of its flanking and relatively unstructured N- and C-terminal regions, and necessity of the oligomeric form of αBC.

It is important to recognise that our assays report primarily on the ability of the ACD to delay aggregation, and not on the stability of any complexes formed, nor their interactions with the cellular disaggregation, refolding, or degradation machinery [10, 56]. Similarly, we are probing only the chaperone function of αBC against selected targets, when it is known that the chaperone has other important roles, including as a major structural component of the eye lens, which may require oligomerisation [57]. It is interesting to note that chaperone activity of dimeric sHSPs has been reported from various organisms, including HSP17.7 from *Deinococcus radiodurans* [58], HSP18.5 from *Arabidopsis thaliana* [59], and human HSP20 [60]. Although these chaperones do not assemble into oligomers, they still have intact N- and C-termini. Our results clearly demonstrate that the ACD itself can be sufficient for chaperone activity, and is not simply a passive building-block of sHSP oligomers.

We have shown that, against all target proteins tested, cABC_{E117C} locked into an AP_{II} dimer was at least as active as cABC. This demonstrates that dissociation into monomers, subunit exchange dynamics, or accessing other dimer registration states (AP_{I} or AP_{III}) is not a pre-requisite for the chaperone function of αBC. This observation is consistent with measurements made on an equivalent E117C mutation in full-length αBC, which had activity comparable to the wild-type in preventing rhodanese from aggregating [44]. Furthermore, in all our experiments the concentration of chaperone used was close to the K_D of the cABC dimer interface, meaning that a sizeable proportion of the chaperone was monomeric [52]. In the case of α-lac and $A\beta_{1-42}$ this did not impair chaperone activity, suggesting that the monomeric ACD is itself chaperone-active. This is consistent with the observation of monomers being the exchang-

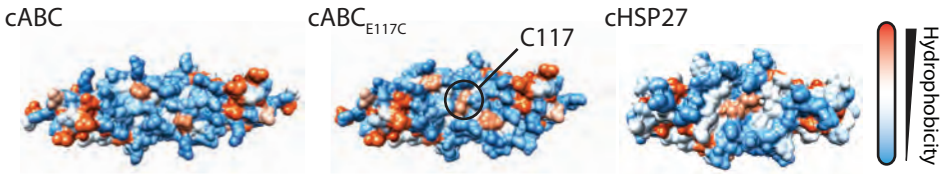


Figure 3.12: Molecular surface of cABC, cABC$_{E117C}$, and cHSP27 coloured according to Kyte-Doolittle hydrophobicity [61]. cABC has few hydrophobic patches, even fewer than cHSP27.

ing unit in αBC [27]. Taken together with the observation that full-length αBC is active, including when cross-linked to prevent its dissociation [36], a picture emerges wherein the ACD is chaperone-active, and its binding surfaces accessible, irrespective of its quaternary state.

Hydrophobic interactions are often hypothesized as the likely mode of chaperone - target interactions [5, 6]. Interestingly, cABC has few exposed hydrophobic patches, and has fewer than the structurally similar, yet inactive cHSP27 (Fig. 3.12). Although we did measure an increase in exposed hydrophobic surface area in cABC$_{E117C}$ relative to cABC (Fig. 3.10), we did not detect major structural distortions in the monomeric fold due to the mutation in either our NMR or X-ray crystallography experiments. It is possible therefore, that the increased chaperone activity of cABC$_{E117C}$ may be due to residue 117 being proximal to the binding site of cABC, and mutation to cysteine resulting in favourable changes to the protein surface. Alternatively, we have established previously that the dimer interface and the β4-β8 groove that accommodates the C-terminal peptide are allosterically coupled: strengthening of one results in the weakening of the other [27, 52]. As such, it may be that stabilising the dimer interface through the covalent linkage may increase the accessibility of the groove as a potential site for target binding.

While cABC is an effective chaperone against the aggregation of all the targets assayed here, there nonetheless remain differences in effectiveness of protection compared to full-length αBC. In the case of α-lac, the core is entirely sufficient for chaperone function, whereas with κ-casein it is somewhat

less active than full-length αBC. The effect of the E117C mutation resulted in an increase in potency against α-lac and κ-casein regardless of its redox state, which could be rationalised by slight differences in surface hydrophobicity (Fig. 3.10). Interestingly, cABC$_{E117C}$ was also more potent against $A\beta_{1-42}$ aggregation than cABC, but only when the disulphide bond locked cABC$_{E117C}$ into a dimeric form. Combined, this suggests there are subtle differences in the way the ACD interacts with different targets, perhaps through separate parts of its molecular surface. This is consistent with observations of different modes of action for amorphous and amyloidogenic aggregation [62, 63], and chaperone activity for peptides from different regions of αBC [31, 32].

It is important to note that sequence differences between cABC and the chaperone inactive cHSP27 are distributed all over the surface, and do not cluster around either the dimer interface or the β4-β8 groove (Fig. 3.3). The small number of differences in primary amino acid sequence therefore does not point to any particular patch of the domain that renders cABC active and cHSP27 inactive. Irrespective of the exact location of the binding site on cABC, the similarity of the cABC and cHSP27 structures suggests that specific residues mediate the interaction between cABC and the targets assayed here. Importantly, the system and approaches we have developed will enable us to probe structurally this important function of sHSP. Understanding how such a small folded protein inhibits amyloid toxicity will potentially be of great use in the search for effective biotherapeutic strategies for diseases associated with amyloid fibril formation.

3.5 Materials and Methods

3.5.1 αBC, cABC and cHSP27 protein expression and purification

Human αBC, cABC (residues 68-153) and cHSP27 (residues 84-176) were cloned, expressed and purified as described previously [41, 52], with the excep-

	cABC	cABC$_{E117C}$	cHSP 27
Crystal parameters			
Space group	I222	C+2	$P2_1$
Cell dimensions			
a,b,c (Å)	35.38, 47.86, 122.84	117.01, 46.77, 77.71	31.28, 48.96, 57.54
$\alpha, \beta, \gamma(^\circ)$	90, 90, 90	90, 90, 90	90, 99.95, 90
Molecules in A.U.[a]	2	8	4
Data collection			
Synchrotron	APS	APS	In-House[b]
Beamline	24-ID-C	24-ID-E	DSE-1.2
Wavelength (Å)	0.979	0.979	1.542
Resolution (Å)	19.6-1.37	69-2.0	37-2.6
Reflections observed/ unique	167150 / 21884	87187 / 25505	38626 / 5331
Completeness (%)	97.5 (96.0)[c]	97.6 (97.52)	99.0 (89.5)
R_{mrgd-F} (%)[d]	4.5 (54.6)	6.9 (43.4)	7.8 (53.7)
$I/\sigma I$	25.13 (3.33)	14.10 (3.5)	17.21 (3.2)
Refinement			
Resolution (Å)	19.6-1.37	68.8-2.0	37-2.6
R_{work}(%)[e]	15.6 (17.4)	19.3 (25.3)	23.4 (27.7)
R_{free}(%)[f]	18.8 (23.9)	24.4 (31.9)	26.6 (34.9)
No. of non H-atoms			
Protein	823	2993	1397
Non-protein	79	117	1
Avg. B- factors	24.2	40.1	50.5
R.m.s deviation			
Bond length (Å)	0.016	0.008	0.012
Bond angle (°)	1.64	1.24	0.995

Table 3.1: Crystallographic table for the ACD constructs. a. A.U. = asymmetric unit. b. Data were collected using an FRD rotating anode generator with an R-AXIS HTC imaging plate detector. c. Values in parentheses correspond to the highest resolution shell. d. $R_{mrgd-F} = \sum|(AI(h,P) - AI(h,Q)|)/(0.5 * \sum AI(h,P) + AI(h,q))$ where $AI = (\sqrt{I}$ if $I >= 0$ or $-\sqrt{I}$ if $I < 0)$ as described previously (3). e. $R_{work} = \sum|F_o - F_c|/\sum F_o$. e. R_{free} calculated from 5% of the data.

tion that in the His-tag buffers phosphate was replaced by TRIS. Truncated proteins were expressed with N-terminal TEV protease cleavable His-tags and purified by using nickel-affinity chromatography. The N-terminal His-tag was removed with TEV protease and the protein buffer exchanged using either a Superdex 75 gel filtration column (GE Healthcare) or two 5 mL HiTrap desalting columns (GE Healthcare) connected in series. For subunit exchange experiments, cABC was expressed in M9 minimal media containing ^{13}C-labelled D-glucose as the only carbon source and purified as described above. For NMR experiments, cABC and cABC$_{E117C}$ were expressed in M9 minimal media containing ^{15}N ammonium chloride as the only nitrogen source.

3.5.2 αBC and HSP27 C-terminal peptide expression and purification

The C-terminal peptides (αBC residues 156-164, ERTIPITRE; HSP27 residues 179-185, ITIPVTF) were either purchased (Biomatik or Genscript), or expressed recombinantly and purified as described previously [64]. Briefly, an expression construct containing an N-terminal His-tagged maltose binding protein followed by a TEV protease cleavage site in pET15b (Novagen) was amplified by PCR using a T7 forward primer and reverse primer with a 5 overhang containing the DNA sequence for the C-terminal peptide followed by a stop codon and XhoI restriction site. The subsequent PCR product was digested with NdeI and XhoI (New England Biolabs) followed by ligation using a Quick Ligation kit (New England Biolabs) into pET28b. The fusion construct was expressed and psurified using nickel-affinity chromatography followed by TEV protease cleavage, leaving an additional N-terminal glycine, and purification using reverse-phase high performance liquid chromatography. Peptide fractions were verified by means of MS, and stored in desiccant jars at -20 °C.

3.5.3 Crystallization of cABC and cHSP27 in complex with their C-terminal peptides

Crystals were grown in hanging-drop plates (VDX, Hampton Research) at room temperature. A final mixture containing 8-10 mg/mL monomer supplemented with 2-3 fold molar excess of the appropriate C-terminal peptide (10 mM stock, in water) was prepared in X-tal buffer (100 mM sodium chloride, 20 mM TRIS, pH 8.0). Crystals of cABC with recombinant C-terminal peptide grew in 0.1 M SPG (succinic acid, dihydrogen phosphate, glycine) buffer pH 6.0 and 25 % PEG 1500. Crystals of $cABC_{E117C}$ with synthetic C-terminal peptide grew in 0.085 M MES pH 6.5, 0.17 M ammonium sulphate, 25.5 % PEG 5000 MME, and 15 % glycerol. A final mixture containing 9 mg/mL cHSP27 supplemented with 2-fold molar excess of the synthetic C-terminal peptide (10mM stock, in water) prepared in Xtal buffer supplemented with 1mM DTT was crystallized in Crystal Screen (Hampton Research) condition no. 6 (0.2 M magnesium chloride, 0.1M TRIS pH 8.5, and 30 % PEG 4,000). With the exception of $cABC_{E117C}$, crystals were cryo-protected in mother-liquid solution containing 20 % glycerol. All crystals were flash-frozen in liquid nitrogen.

3.5.4 Structure determination and analysis

X-ray diffraction data were processed using XDS [65], and structures determined by molecular replacement using Phaser [66], followed by automated model building using Phenix [67]. Models were built using Coot [68] and refined with REFMAC [69], Phenix [67], and Buster [70]. Structures of cABC, $cABC_{E117C}$, and cHSP27 have been deposited in the PDB with accession codes 4M5S, 4M5T, 4MJH respectively.

Buried surface areas in the dimer interface were calculated using the PISA server [71]. 3L1G and 2WJ7 were used for calculations for AP_I and AP_{II}. Sequence alignments were generated using the ClustalW2 server [72]. Amino acids were coloured according to their score in the Gonnet Pam250 matrix gen-

erated by the ClustalW2 server (< 0: red; $=< 0.5$: orange, > 0.5: yellow), and fully conserved residues were not coloured (Fig. 3.3A). Surface hydrophobicity was calculated and rendered using the UCSF Chimera [73].

3.5.5 Mass spectrometry

MS measurements were carried out on a IM QToF (Synapt G1, Waters) modified to incorporate a linear drift tube [74]. Nanoelectrospray mass spectra were acquired under conditions optimised for the preservation of noncovalent interactions essentially as described previously [52] and with the following instrument parameters: sample cone 10 V, extraction cone 5 V, trap 10 V, trap gas (argon): 2.6 ml/min, drift-cell pressure (helium) 2.14 Torr. Unless stated otherwise, samples were analysed at monomeric concentrations of 10-20 μM in 200 mM ammonium acetate pH 6.9. DTT, to a final concentration of 5 mM, was added where stated. For subunit exchange experiments, ^{13}C-labelled and unlabelled protein was mixed at equimolar ratios at 4 °C and analysed by MS as soon as possible after mixing (<1 min). CCSs were measured by mixing oxidized cABCE11C and ^{13}C-labelled cABC, and then recording spectra at nine drift voltages in the range 50-200 V. At each drift voltage, arrival times were fitted to normal distributions, and the transport time (T_0) of ions from the exit of the drift-tube region to the time-of-flight analyzer obtained [74]. CCS distributions were obtained from the arrival time (T_a) distributions of the 7+ charge state at each drift voltage V_D using $CCS = 4.482 * 10^{-3}(T_a - T_o)V_D z T P^{-1}$ where z is the charge of the ion, T is the temperature in K, P the pressure in Torr, and the constant amalgamates the length of the drift tube with the necessary components of the Mason-Schamp equation [75, 74]. The data from all drift voltages were combined and fitted globally to a normal distribution. Theoretical CCSs of AP_I-AP_{III} were calculated from the cABC crystal structure using CCSCalc (Waters) with a gas radius of 1 Å and tolerance of 0.1%, and normalized to the experimentally determined CCS value of AP_{II} for comparison.

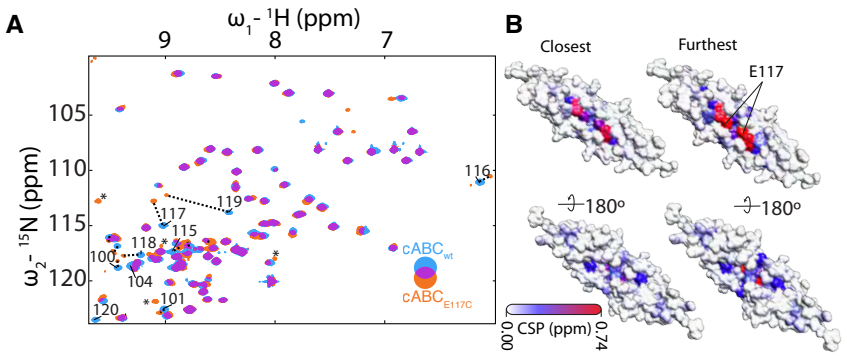


Figure 3.13: A) A small number of resonances were difficult to assign in cABC$_{E117C}$ because of several peaks in cABC$_{E117C}$ plausibly mapping to a peak in cABC. To deal with this ambiguity, two putative assignment lists for cABC$_{E117C}$ were produced. The first assumes that the resonance position in cABC$_{E117C}$ is furthest from that in cABC, and the second assumes that the resonance position in cABC$_{E117C}$ is closest to that in the cABC. The CSP values obtained in both cases are essentially identical. In the text, we show the assignment list assuming cABC$_{E117C}$ is furthest from cABC. Here, we show the assignment list assuming cABC$_{E117C}$ is nearest WT. B) The CSP values when mapped onto the structure of cABC are essentially indistinguishable.

3.5.6 NMR spectroscopy

NMR experiments were performed on a 600 MHz spectrometer (Varian) equipped with a 5 mm z-axis gradient triple-resonance probe. ^{15}N-labelled cABC and cABC$_{E117C}$ were prepared in 25 mM sodium phosphate, 2 mM EDTA, 2 mM sodium azide, pH 7.5 and concentrated to 500 μM using amicon concentrators (Millipore, UK). 2D ^{15}N-^{1}H HSQC correlation spectra were recorded with 8 scans per transient with acquisition times (t1, t1) of (65.8, 135.1) ms, over (608, 100) complex points and a 1 s relaxation delay, for a total acquisition time of 26 min. All spectra were analysed using NMRPipe [76] and Sparky (T. D. Goddard and D. G. Kneller, University of California, San Francisco). Peaks were ascribed to particular residues based on the previous assignment of a similar construct [77]. To measure chemical shift perturbations (CSPs) in both dimensions, the chemical shift difference in the nitrogen dimension was divided by 5 to account for the distribution of shifts recorded in the BMRB. In the few cases where several peaks could not be clearly assigned in the cABC$_{E117C}$ spectrum, the furthest peak was used for CSP calculations in order not to

under-estimate the CSP. If the closest peak was used instead, the pattern of CSPs was very similar, although the CSPs were slightly smaller (Fig. 3.13A, B).

3.5.7 *In vitro* protein aggregation assays

The aggregation and precipitation of the target proteins, reported via either ThT fluorescence or light-scattering assays, was monitored by using sealed 96-microwell plates and Fluostar Optima plate reader (BMG Labtechnologies). The amorphous aggregation of α-lac (Sigma-Aldrich, 100 μM), incubated at 37 °C in 50 mM phosphate buffer containing 100 mM NaCl and 5 mM EDTA, pH 7.0, was initiated by the addition of DTT (20 mM). Chaperones were added at the molar ratios stated. Aggregation was monitored by measuring the change in apparent absorbance due to light scattering at 340 nm, which was negligible in the absence of α-lac. The formation of amyloid fibrils by target proteins was monitored using an in situ ThT-binding assay [63]. κ-casein (Sigma-Aldrich, 50 μM) was incubated at 37 °C in 50 mM phosphate buffer (pH 7.4), and $A\beta_{1-42}$ (Anaspec, 10 μM) at 37 °C in PBS (pH 7.4). Chaperones and DTT (20 mM) were added where stated at a molar ratio of 1:2 (chaperone:κ-casein) or 1:20 (chaperone:$A\beta_{1-42}$). Samples were incubated with 10 μM ThT and the fluorescence levels measured with a 440/490 nm excitation/emission filter set. The change in ThT fluorescence in the absence of either κ-casein or $A\beta_{1-42}$ was negligible for each assay. In all cases, the relative ability of each chaperone to prevent aggregation was evaluated by comparing the apparent absorbance/ThT fluorescence at the end of each assay, as previously described [78]. Data are reported as mean SEM (n = 3) and were analysed by one-way ANOVA and Tukeys post-hoc test. To assay exposed hydrophobicity, samples were prepared at 2.5 μM in PBS and bis-ANS was added to a final concentration of 10 μM. Fluorescence spectra were recorded as previously described [62].

3.5.8 Cell viability assays

Cell viability was measured by using a CellTiter 96 aqueous non-radioactive cell proliferation assay kit (Promega G4100). PC-12 (ATCC cat. no. CRL-1721), HeLa and HEK293 cell lines were used to assess the inhibition of our αBC constructs on $A\beta_{1-42}$ toxicity. HeLa and HEK293 cells were cultured in DMEM medium with 10% fetal bovine serum. PC12 cells were cultured in ATCC-formulated RPMI 1640 medium (ATCC) with 10% heat-inactivated horse serum and 5% fetal bovine serum. All cells were maintained at 5% CO2 at 37 °C. For cell viability measurements, HeLa, HEK293 and PC-12 cells were plated out in 96-well plates (Costar) at 10000 cells/well, and cultured for 20 h at 37 °C in 5% CO2 before adding different incubation mixtures. These were prepared to a final concentration of 5 μM $A\beta_{1-42}$ monomer in PBS (pH 7.4), in the presence or absence of chaperone as indicated, and incubated at 37 °C for 18 h. 10 L of incubated sample was added to each well containing 90 L cell suspension. After 24 h incubation, 15 L dye solution was added to each well, and the cells incubated for a further 4 h before the addition of 100 L solution/stop mix to each well. After incubation at room temperature for 12 h, the absorbance of each well was measured at 570 nm, with the background absorbance recorded at 700 nm. The final data were normalized by using the buffer-treated cell as 100% viability and 0.2% SDS-treated cell as 0% viability. Data are reported as mean SEM (n = 4) and the titration series were analysed globally using an unpaired Students T-test. Specifically, the plateau of viability at high ratios of chaperone: $A\beta_{1-42}$ was determined by fitting the data to an exponential rise to maximum [79], and compared to the absence of chaperone altogether.

3.6 References

[1] A. J. Baldwin, H. Lioe, G. R. Hilton, L. A. Baker, J. L. Rubinstein, L. E. Kay, and J. L. Benesch, "The polydispersity of αB-crystallin is rationalized by an interconverting polyhedral architecture.," *Structure*, vol. 19, pp. 1855–1863, Dec. 2011.

[2] H. Olzscha, S. M. Schermann, A. C. Woerner, S. Pinkert, M. H. Hecht, G. G. Tartaglia, M. Vendruscolo, M. Hayer-Hartl, F. U. Hartl, and R. M. Vabulas, "Amyloid-like aggregates sequester numerous metastable proteins with essential cellular functions.," *Cell*, vol. 144, pp. 67–78, Jan. 2011.

[3] E. T. Powers and W. E. Balch, "Diversity in the origins of proteostasis networks–a driver for protein function in evolution.," *Nat. Rev. Mol. Cell Biol.*, vol. 14, pp. 237–248, Apr. 2013.

[4] W. E. Balch, R. I. Morimoto, A. Dillin, and J. W. Kelly, "Adapting proteostasis for disease intervention," *Science*, vol. 319, pp. 916–9+, 2008.

[5] F. U. Hartl, A. Bracher, and M. Hayer-Hartl, "Molecular chaperones in protein folding and proteostasis," *Nature*, vol. 475, pp. 324–32+, 2011.

[6] K. Richter, M. Haslbeck, and J. Buchner, "The heat shock response: life on the verge of death," *Mol. Cell*, vol. 40, pp. 253–66+, 2010.

[7] E. Basha, H. O'Neill, and E. Vierling, "Small heat shock proteins and α-crystallins: dynamic proteins with flexible functions," *Trends Biochem. Sci.*, vol. 37, pp. 106–17+, 2012.

[8] G. R. Hilton, H. Lioe, F. Stengel, A. J. Baldwin, and J. L. P. Benesch, "Small heat-shock proteins: paramedics of the cell," *Top. Curr. Chem.*, 2012.

[9] H. S. McHaourab, J. A. Godar, and P. L. Stewart, "Structure and mechanism of protein stability sensors: chaperone activity of small heat shock proteins.," *Biochemistry*, vol. 48, pp. 3828–3837, May 2009.

[10] S. Carra, P. Rusmini, V. Crippa, E. Giorgetti, A. Boncoraglio, R. Cristofani, M. Naujock, M. Meister, M. Minoia, H. H. Kampinga, and A. Poletti, "Different anti-aggregation and pro-degradative functions of the members of the mammalian sHSP family in neurological disorders.," *Philos. Trans. R. Soc. Lond. B. Biol. Sci.*, vol. 368, pp. 20110409+, May 2013.

[11] C. Garrido, C. Paul, R. Seigneuric, and H. H. Kampinga, "The small heat shock proteins family: the long forgotten chaperones.," *Int. J. Biochem. Cell Biol.*, vol. 44, pp. 1588–1592, Oct. 2012.

[12] J. Horwitz, "α-crystallin can function as a molecular chaperone.," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 89, pp. 10449–10453+, 1992.

[13] A. Boncoraglio, M. Minoia, and S. Carra, "The family of mammalian small heat shock proteins (HSPBs): Implications in protein deposit diseases and motor neuropathies," *Int. J. Biochem. Cell Biol.*, 2012.

[14] H. Ecroyd and J. A. Carver, "Crystallin proteins and amyloid fibrils," *Cell. Mol. Life. Sci.*, vol. 66, pp. 62–81+, 2009.

[15] H. H. Kampinga and C. Garrido, "HSPBs: small proteins with big implications in human disease.," *Int. J. Biochem. Cell Biol.*, vol. 44, pp. 1706–1710, Oct. 2012.

[16] F. Chiti and C. M. Dobson, "Protein misfolding, functional amyloid, and human disease.," *Annu. Rev. Biochem.*, vol. 75, no. 1, pp. 333–366, 2006.

[17] D. Eisenberg and M. Jucker, "The Amyloid State of Proteins in Human Diseases," *Cell*, vol. 148, pp. 1188–1203, Mar. 2012.

[18] C. Haass and D. J. Selkoe, "Soluble protein oligomers in neurodegeneration: lessons from the Alzheimer's amyloid β-peptide," *Nat. Rev. Mol. Cell Biol.*, vol. 8, pp. 101–112, Feb. 2007.

[19] M. M. Wilhelmus, W. C. Boelens, I. Otte-Höller, B. Kamps, R. M. de Waal, and M. M. Verbeek, "Small heat shock proteins inhibit amyloid-β protein aggregation and cerebrovascular amyloid-β protein toxicity.," *Brain Res.*, vol. 1089, pp. 67–78, May 2006.

[20] B. Mannini, R. Cascella, M. Zampagni, M. van Waarde-Verhagen, S. Meehan, C. Roodveldt, S. Campioni, M. Boninsegna, A. Penco, A. Relini, H. H. Kampinga, C. M. Dobson, M. R. Wilson, C. Cecchi, and F. Chiti, "Molecular mechanisms used by chaperones to reduce the toxicity of aberrant protein oligomers.," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 109, pp. 12479–12484, July 2012.

[21] T. P. Knowles, W. Shu, G. L. Devlin, S. Meehan, S. Auer, C. M. Dobson, and M. E. Welland, "Kinetics and thermodynamics of amyloid formation from direct measurements of fluctuations in fibril mass," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 104, pp. 10016–21+, 2007.

[22] B. Raman, T. Ban, M. Sakai, S. Y. Pasta, T. Ramakrishna, H. Naiki, Y. Goto, and C. M. h. M. Rao, "αB-crystallin, a small heat-shock protein, prevents the amyloid fibril growth of an amyloid β-peptide and β2-microglobulin.," *Biochem. J.*, vol. 392, pp. 573–581, Dec. 2005.

[23] A. Rekas, C. G. Adda, J. Andrew Aquilina, K. J. Barnham, M. Sunde, D. Galatis, N. A. Williamson, C. L. Masters, R. F. Anders, C. V. Robinson, R. Cappai, and J. A. Carver, "Interaction of the molecular chaperone αB-crystallin with α-synuclein: effects on amyloid fibril formation and chaperone activity.," *J. Mol. Biol.*, vol. 340, pp. 1167–1183, July 2004.

[24] S. L. Shammas, C. A. Waudby, S. Wang, A. K. Buell, T. P. Knowles, H. Ecroyd, M. E. Welland, J. A. Carver, C. M. Dobson, and S. Meehan, "Binding of the molecular chaperone αB-crystallin to Aβ amyloid fibrils inhibits fibril elongation," *Biophys. J.*, vol. 101, pp. 1681–9+, 2011.

[25] C. A. Waudby, T. P. Knowles, G. L. Devlin, J. N. Skepper, H. Ecroyd, J. A. Carver, M. E. Welland, J. Christodoulou, C. M. Dobson, and S. Meehan, "The interaction of αB-crystallin with mature α-synuclein amyloid fibrils inhibits their elongation," *Biophys. J.*, vol. 98, pp. 843–51+, 2010.

[26] K. J. Binger, H. Ecroyd, S. Yang, J. A. Carver, G. J. Howlett, and M. D. W. Griffin, "Avoiding the oligomeric state: αB-crystallin inhibits fragmentation and induces dissociation of apolipoprotein C-II amyloid fibrils," *FASEB J.*, vol. 27, pp. 1214–1222, Mar. 2013.

[27] A. J. Baldwin, H. Lioe, C. V. Robinson, L. E. Kay, and J. L. Benesch, "αB-crystallin polydispersity is a consequence of unbiased quaternary dynamics," *J. Mol. Biol.*, vol. 413, pp. 297–309+, 2011.

[28] J. A. Aquilina, S. Shrestha, A. M. Morris, and H. Ecroyd, "Structural and functional aspects of hetero-oligomers formed by the small heat shock proteins αB-crystallin and HSP27.," *J. Biol. Chem.*, vol. 288, pp. 13602–13609, May 2013.

[29] S. P. Delbecq and R. E. Klevit, "One size does not fit all: the oligomeric states of B crystallin.," *FEBS Lett.*, vol. 587, pp. 1073–1080, Apr. 2013.

[30] J. A. Aquilina and S. J. Watt, "The N-terminal domain of αB-crystallin is protected from proteolysis by bound substrate," *Biochem. Biophys. Res. Commun.*, vol. 353, pp. 1115–1120, Feb. 2007.

[31] J. Bhattacharyya, E. G. Padmanabha Udupa, J. Wang, and K. K. Sharma, "Mini-αB-crystallin: a functional element of αB-crystallin with

chaperone-like activity.," *Biochemistry*, vol. 45, pp. 3069–3076, Mar. 2006.

[32] J. G. Ghosh, M. R. Estrada, and J. I. Clark, "Interactive domains for chaperone activity in the small heat shock protein, human αB crystallin.," *Biochemistry*, vol. 44, pp. 14854–14869, Nov. 2005.

[33] S. Narayanan, B. Kamps, W. C. Boelens, and B. Reif, "αB-crystallin competes with Alzheimer's disease β-amyloid peptide for peptide-peptide interactions and induces oxidation of Aβ-Met35.," *FEBS Lett.*, vol. 580, pp. 5941–5946, Oct. 2006.

[34] K. K. Sharma, H. Kaur, and K. Kester, "Functional elements in molecular chaperone α-crystallin: identification of binding sites in αB-crystallin.," *Biochem. Biophys. Res. Commun.*, vol. 239, pp. 217–222, Oct. 1997.

[35] M. P. Bova, L. L. Ding, J. Horwitz, and B. K. Fung, "Subunit exchange of αA-crystallin.," *J. Biol. Chem.*, vol. 272, pp. 29511–29517, Nov. 1997.

[36] R. C. Augusteyn, "Dissociation is not required for α-crystallin's chaperone function.," *Exp. Eye Res.*, vol. 79, pp. 781–784, Dec. 2004.

[37] J. A. Aquilina, J. L. P. Benesch, L. L. Ding, O. Yaron, J. Horwitz, and C. V. Robinson, "Subunit exchange of polydisperse proteins: mass spectrometry reveals consequences of αA-crystallin truncation," *J. Biol. Chem*, vol. 280, pp. 14485–91+, 2005.

[38] C. Bagnéris, O. A. Bateman, C. E. Naylor, N. Cronin, W. C. Boelens, N. H. Keep, and C. Slingsby, "Crystal structures of α-crystallin domain dimers of αB-crystallin and Hsp20," *J. Mol. Biol.*, vol. 392, pp. 1242–1252, Oct. 2009.

[39] A. R. Clark, C. E. Naylor, C. Bagneris, N. H. Keep, and C. Slingsby, "Crystal structure of R120G disease mutant of human αB-crystallin domain dimer shows closure of a groove," *J. Mol. Biol.*, vol. 408, pp. 118–34+, 2011.

[40] S. Jehle, P. Rajagopal, B. Bardiaux, S. Markovic, R. Kuhne, J. R. Stout, V. A. Higman, R. E. Klevit, B. J. van Rossum, and H. Oschkinat, "Solid-state NMR and SAXS studies provide a structural basis for the activation of αB-crystallin oligomers," *Nat. Struct. Mol. Biol.*, vol. 17, pp. 1037–42+, 2010.

[41] A. Laganowsky, J. L. Benesch, M. Landau, L. Ding, M. R. Sawaya, D. Cascio, Q. Huang, C. V. Robinson, J. Horwitz, and D. Eisenberg, "Crystal structures of truncated αA and αB crystallins reveal structural mechanisms of polydispersity important for eye lens function," *Protein Sci.*, vol. 19, pp. 1031–43+, 2010.

[42] A. Laganowsky and D. Eisenberg, "Non-3D domain swapped crystal structure of truncated zebrafish αA crystallin," *Protein Sci.*, vol. 19, pp. 1978–84+, 2010.

[43] S. P. Delbecq, S. Jehle, and R. Klevit, "Binding determinants of the small heat shock protein, αB-crystallin: recognition of the IxI motif," *EMBO J.*, vol. 31, pp. 4587–4594, Dec. 2012.

[44] E. Mymrikov, O. Bukach, A. Seit-Nebi, and N. Gusev, "The pivotal role of the β7 strand in the intersubunit contacts of different human small heat shock proteins," *Cell Stress Chaperon.*, vol. 15, no. 4, pp. 365–377, 2010.

[45] E. V. Baranova, S. D. Weeks, S. Beelen, O. V. Bukach, N. B. Gusev, and S. V. Strelkov, "Three-dimensional structure of α-crystallin domain

dimers of human small heat shock proteins HSPB1 and HSPB6.," *J. Mol. Biol.*, vol. 411, pp. 110–122, Aug. 2011.

[46] A.-P. Arrigo, "Hsp27: Novel Regulator of Intracellular Redox State," *IUBMB Life*, vol. 52, pp. 303–307, Dec. 2001.

[47] S. Indu, V. Kochat, S. Thakurela, C. Ramakrishnan, and R. Varadarajan, "Conformational analysis and design of cross-strand disulfides in antiparallel β-sheets.," *Proteins*, vol. 79, pp. 244–260, Jan. 2011.

[48] S. Jehle, B. van Rossum, J. R. Stout, S. M. Noguchi, K. Falber, K. Rehbein, H. Oschkinat, R. E. Klevit, and P. Rajagopal, "αB-crystallin: a hybrid solid-state/solution-state NMR investigation reveals structural aspects of the heterogeneous oligomer.," *J. Mol. Biol.*, vol. 385, pp. 1481–1497, Feb. 2009.

[49] S. Jehle, B. van Rossum, J. R. Stout, S. M. Noguchi, K. Falber, K. Rehbein, H. Oschkinat, R. E. Klevit, and P. Rajagopal, "αB-Crystallin: A Hybrid Solid-State/Solution-State NMR Investigation Reveals Structural Aspects of the Heterogeneous Oligomer," *J. Mol. Biol.*, vol. 385, pp. 1481–1497+, 2009.

[50] B. T. Ruotolo, J. L. Benesch, A. M. Sandercock, S.-J. J. Hyung, and C. V. Robinson, "Ion mobility-mass spectrometry analysis of large protein complexes.," *Nat. Protoc.*, vol. 3, pp. 1139–1152, June 2008.

[51] J. L. Benesch and B. T. Ruotolo, "Mass spectrometry: come of age for structural and dynamical biology," *Curr. Opin. Struct. Biol.*, vol. 21, pp. 641–9+, 2011.

[52] G. R. Hilton, G. K. Hochberg, A. Laganowsky, S. I. McGinnigle, A. J. Baldwin, and J. L. Benesch, "C-terminal interactions mediate the

quaternary dynamics of αB-crystallin.," *Philos. Trans. R. Soc. Lond. B. Biol. Sci.*, vol. 368, May 2013.

[53] J. A. Carver, R. A. Lindner, C. Lyon, D. Canet, H. Hernandez, C. M. Dobson, and C. Redfield, "The interaction of the molecular chaperone α-crystallin with unfolding α-lactalbumin: a structural and kinetic spectroscopic study.," *J. Mol. Biol.*, vol. 318, pp. 815–827, May 2002.

[54] I. K. Feil, M. Malfois, J. Hendle, H. van Der Zandt, and D. I. Svergun, "A novel quaternary structure of the dimeric α-crystallin domain with chaperone-like activity.," *J. Biol. Chem.*, vol. 276, pp. 12024–12029, Apr. 2001.

[55] T. Härd and C. Lendel, "Inhibition of Amyloid Formation," *J. Mol. Biol.*, vol. 421, pp. 441–465, Aug. 2012.

[56] K. Liberek, A. Lewandowska, and S. Zietkiewicz, "Chaperones in control of protein disaggregation.," *EMBO J.*, vol. 27, pp. 328–335, Jan. 2008.

[57] J. Piatigorsky and G. Wistow, "The recruitment of crystallins: new functions precede gene duplication.," *Science*, vol. 252, pp. 1078–1079, May 1991.

[58] A. Bepperling, F. Alte, T. Kriehuber, N. Braun, S. Weinkauf, M. Groll, M. Haslbeck, and J. Buchner, "Alternative bacterial two-component small heat shock protein systems.," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 109, pp. 20407–20412, Dec. 2012.

[59] E. Basha, C. Jones, A. E. Blackwell, G. Cheng, E. R. Waters, K. A. Samsel, M. Siddique, V. Pett, V. Wysocki, and E. Vierling, "An unusual dimeric small heat shock protein provides insight into the mechanism of this class of chaperones.," *J. Mol. Biol.*, vol. 425, pp. 1683–1696, May 2013.

[60] O. V. Bukach, A. S. Seit-Nebi, S. B. Marston, and N. B. Gusev, "Some properties of human small heat shock protein Hsp20 (HspB6).," *Eur. J. Biochem.*, vol. 271, pp. 291–302, Jan. 2004.

[61] J. Kyte and R. F. Doolittle, "A simple method for displaying the hydropathic character of a protein.," *J. Mol. Biol.*, vol. 157, pp. 105–132, May 1982.

[62] M. Kulig and H. Ecroyd, "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.*, vol. 448, pp. 343–352, Dec. 2012.

[63] H. Ecroyd, T. Koudelka, D. C. Thorn, D. M. Williams, G. Devlin, P. Hoffmann, and J. A. Carver, "Dissociation from the Oligomeric State Is the Rate-limiting Step in Fibril Formation by κ-Casein," *J. Biol. Chem.*, vol. 283, pp. 9012–9022, Apr. 2008.

[64] A. Laganowsky, C. Liu, M. R. Sawaya, J. P. Whitelegge, J. Park, M. Zhao, A. Pensalfini, A. B. Soriaga, M. Landau, P. K. Teng, D. Cascio, C. Glabe, and D. Eisenberg, "Atomic view of a toxic amyloid small oligomer," *Science*, vol. 335, pp. 1228–31+, 2012.

[65] W. Kabsch, "Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants," *J. Appl. Crystallogr.*, vol. 26, pp. 795–800, Dec. 1993.

[66] A. J. McCoy, R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni, and R. J. Read, "Phaser crystallographic software.," *J. Appl. Crystallogr.*, vol. 40, pp. 658–674, Aug. 2007.

[67] P. D. Adams, R. W. Grosse-Kunstleve, L. W. W. Hung, T. R. Ioerger, A. J. McCoy, N. W. Moriarty, R. J. Read, J. C. Sacchettini, N. K. Sauter, and T. C. Terwilliger, "PHENIX: building new software for

automated crystallographic structure determination.," *Acta Crystallogr. D Biol. Crystallogr.*, vol. 58, pp. 1948–1954, Nov. 2002.

[68] P. Emsley, B. Lohkamp, W. G. Scott, and K. Cowtan, "Features and development of Coot.," *Acta Crystallogr. D Biol. Crystallogr.*, vol. 66, pp. 486–501, Apr. 2010.

[69] G. N. Murshudov, A. A. Vagin, and E. J. Dodson, "Refinement of macromolecular structures by the maximum-likelihood method.," *Acta Crystallogr. D Biol. Crystallogr.*, vol. 53, pp. 240–255, May 1997.

[70] O. S. Smart, T. O. Womack, C. Flensburg, P. Keller, W. Paciorek, A. Sharff, C. Vonrhein, and G. Bricogne, "Exploiting structure similarity in refinement: automated NCS and target-structure restraints in BUSTER.," *Acta Crystallogr. D Biol. Crystallogr.*, vol. 68, pp. 368–380, Apr. 2012.

[71] E. Krissinel and K. Henrick, "Inference of Macromolecular Assemblies from Crystalline State," *J. Mol. Biol.*, vol. 372, pp. 774–797, Sept. 2007.

[72] M. A. Larkin, G. Blackshields, N. P. Brown, R. Chenna, P. A. McGettigan, H. McWilliam, F. Valentin, I. M. Wallace, A. Wilm, R. Lopez, J. D. Thompson, T. J. Gibson, and D. G. Higgins, "Clustal W and Clustal X version 2.0.," *Bioinformatics*, vol. 23, pp. 2947–2948, Nov. 2007.

[73] E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, and T. E. Ferrin, "UCSF Chimera–a visualization system for exploratory research and analysis.," *J. Comput. Chem.*, vol. 25, pp. 1605–1612, Oct. 2004.

[74] M. F. Bush, Z. Hall, K. Giles, J. Hoyes, C. V. Robinson, and B. T. Ruotolo, "Collision cross sections of proteins and their complexes: a

calibration framework and database for gas-phase structural biology," *Anal. Chem.*, vol. 82, pp. 9557–9565, Oct. 2010.

[75] B. T. Ruotolo, J. L. P. Benesch, A. M. Sandercock, S. J. Hyung, and C. V. Robinson, "Ion mobility-mass spectrometry analysis of large protein complexes," *Nat. Protoc.*, vol. 3, pp. 1139–52+, 2008.

[76] F. Delaglio, S. Grzesiek, G. W. Vuister, G. Zhu, J. Pfeifer, and A. Bax, "NMRPipe: a multidimensional spectral processing system based on UNIX pipes.," *J. Biomol. NMR*, vol. 6, pp. 277–293, Nov. 1995.

[77] S. Jehle, B. S. Vollmar, B. Bardiaux, K. K. Dove, P. Rajagopal, T. Gonen, H. Oschkinat, and R. E. Klevit, "N-terminal domain of αB-crystallin provides a conformational switch for multimerization and structural heterogeneity," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 108, pp. 6409–14+, 2011.

[78] H. Ecroyd and J. A. Carver, "The effect of small molecules in modulating the chaperone activity of αB-crystallin against ordered and disordered protein aggregation.," *FEBS J.*, vol. 275, pp. 935–947, Mar. 2008.

[79] W. Slob, "Dose-Response Modeling of Continuous Endpoints," *Toxicol. Sci.*, vol. 66, pp. 298–312, Apr. 2002.

Chapter 4

Dynamical structure of αB-crystallin

4.1 Abstract

The human small heat-shock protein αB-crystallin is an extremely difficult molecule to study, with its inherent structural dynamics posing unique challenges to all biophysical and structural biology techniques. Here we highlight how the polydispersity and quaternary dynamics of αB-crystallin are intrinsically inter-twined, and how this can impact on measurements of the oligomeric distribution. We show that, in spite of these difficulties, considerable understanding of the varied fluctuations αB-crystallin undergoes at equilibrium has emerged in the last few years. By reporting on data obtained from a variety of biophysical techniques, we demonstrate how the αB-crystallin solution ensemble is governed by molecular motions of varying amplitude and time-scales spanning several orders of magnitude. We describe how these diverse measurements are being used to construct an integrated view of the dynamical structure of αB-crystallin, and highlight areas that require further interrogation. With its study continuing to motivate refinement of experimental techniques, and the development of new approaches to combine the hybrid datasets, we conclude that αB-crystallin continues to represent a paradigm for dynamical biology.

4.2 αB-crystallin: a refractory refractor

αB-crystallin is a member of the small heat-shock protein (sHSP) family of molecular chaperones [1, 2, 3]. It is expressed in particularly high abundance in the eye lens, but is found throughout the human body [4]. In line with the canonical chaperone function of the sHSPs, αB-crystallin can bind partially unfolded proteins and prevent their accumulation into irrecoverable and potentially toxic aggregates [5, 6, 7]. *In vitro*, this activity strongly inhibits the formation of both amorphous and fibrillar aggregates of a variety of model targets [8]. Consequently, αB-crystallin is associated with insoluble protein deposits in several important amyloid diseases, including Alexanders, Alzheimers and Parkinsons, as well as variants leading to myofibrillar myopathy, desmin-related cardiomyopathy, and dilated cardiomyopathy [9, 8]. These observations indicate that αB-crystallin is an important member of the human protein homeostasis network [10], where it acts not to refold misfolded proteins nor disaggregate protein deposits, but as an ATP-independent overflow reservoir that sequesters destabilized proteins in situations that might otherwise overwhelm the ATP-dependent chaperone machinery [11, 12].

In the human eye lens, αB-crystallin remains soluble over the organisms lifespan at very high concentrations (>300 mg/mL of total lens protein) [13], providing a refractive index sufficient for vision, while concomitantly inhibiting the formation of protein aggregates that would otherwise scatter light [14, 15]. A number of mis-sense mutations in αB-crystallin lead to cataract, revealing its importance for eye-lens transparency [16], although a homozygous knock-out in mice [17] and a severe truncation in humans [18] do not lead to cataract formation, possibly because the more abundant αA-crystallin can compensate for the loss of αB-crystallin in the eye lens. This suggests that congenital cataract formation represents a deleterious gain of function by αB-crystallin. Accordingly, it evolved by acquiring properties in order to remain soluble while performing its functions in the eye lens [19], with the corrolary that its does not crystal-

lize, even at extremely high concentrations and over long time frames [14, 15]. Chief amongst these characteristics is that αB-crystallin does not have a single defined quaternary structure, but rather populates a wide range of stoichiometries at equilibrium. This polydispersity is also apparent for mammalian sHSPs not present in the eye-lens (and may prove to be the rule rather than the exception across the whole sHSP family), and has inevitably frustrated attempts at structure determination of αB-crystallin oligomers by crystallographic means. Similarly, other structural biology techniques such as nuclear magnetic resonance spectroscopy (NMR), electron microscopy (EM), or small angle X-ray scattering (SAXS), either do not directly report on quaternary architecture, or require knowledge of oligomeric populations for accurate model building. The question of quaternary assembly therefore lies at the heart of any discussion on the solution structure of αB-crystallin, and a clear understanding of its stoichiometry at equilibrium is a prerequisite for the construction of candidate structures. Although details have begun to emerge over the past decade, the complete structure of αB-crystallin remains elusive and one of the most coveted targets in structural biology.

4.3 Coupled quaternary dynamics and polydispersity of αB-crystallin

An understanding of the quaternary assembly of αB-crystallin rests on two key observations: the protein exists as an ensemble of oligomeric states, and these states are capable of inter-conversion. An accurate biophysical description of αB-crystallin, therefore requires quantitative knowledge of two fundamental properties: the oligomeric distribution populated by αB-crystallin at equilibrium, and the rate of its recycling via subunit exchange. As we demonstrate below, these quantities are intrinsically inter-twined, with the consequence that subunit exchange can profoundly influence measurements of αB-crystallin stoichiometries.

4.3.1 The extent of polydispersity

αB-crystallins polydispersity was noted almost fifty years ago in analytical ultracentrifugation (AUC) experiments [20], nearly three decades earlier than the first unequivocal observations of subunit exchange [21, 22]. Much of the research in intervening years had been taken to suggest that the population of αB-crystallin oligomers does not represent a chemical equilibrium. The reason for this deduction, which we now know to be inaccurate, can be readily understood in terms of the experimental time-scale and the rate of subunit exchange. For instance, different average molecular weights were obtained when AUC was performed on aliquots from either the front or the tail of a size-exclusion chromatography (SEC) peak [23]. However, the SEC separation and AUC analysis was performed at 4 °C, at which temperature it actually takes weeks to re-establish equilibrium after fractionation. This phenomenon has been exploited to isolate oligomeric sub-populations of αB-crystallin by using SEC at low temperature, and submit them to downstream analysis. In this way EM data were obtained that revealed substantially different sizes of spherical particles for different fractions [24]. Two decades later, cryo-EM studies with subsequent single-particle analysis resulted in poorly defined globular shapes for the oligomers, further confirming αB-crystallins heterogeneity [25], but stopped short of detailed reconstructions due to the need for classification strategies that can accurately separate individual stoichiometries [26].

The advent of nanoelectrospray mass spectrometry (MS) techniques that allow the interrogation of intact protein complexes in the gas-phase [27] enabled the first measurements of αB-crystallin that resolved individual stoichiometries [28]. The resulting distributions revealed an ensemble of oligomers containing between 10 and 50 subunits, centred approximately on a 28mer (Fig. 4.1A). Interestingly, the ability to resolve individual stoichiometries demonstrated that both oligomers with odd and even numbers of subunits are present, with the latter slightly more abundant [29]. The presence of odd-numbered oligomers

was unexpected as it revealed that monomers play an important role in the assembly of αB-crystallin, while sHSPs had in general been thought to be constructed exclusively from dimeric building blocks [30]. Notably, though displaying a much higher resolution of separation, the MS data are in good agreement with multi-angle laser light scattering (MALS) and EM measurements, when effort is made to size-sort the oligomers, respectively by means of SEC or in silico (Fig. 4.1A). However, using similar solution-based techniques, other researchers have arrived at apparently more homogeneous distributions, even leading to calls in recent years to question whether αB-crystallin is polydisperse at all. This led to the unsatisfying situation that the technique and experimental conditions used appeared to strongly influence the measurement αB-crystallins quaternary assembly.

4.3.2 Quantifying the rate at which αB-crystallin exchanges subunits

It is our contention that subunit exchange lies at the heart of many of the inconsistencies in estimates of the oligomeric distribution populated by αB-crystallin and, conversely, can therefore be used to resolve them. Such quaternary dynamics had been hypothesised in the late 1970s [24], but were uncovered experimentally only much later by using isoelectric focusing. Through exploiting the different charges of αA- and αB-crystallin in solution, the hetero-oligomerisation of the two pure isoforms was observed [21], and the evolving mixture monitored [22]. This discovery provided an immediate explanation for several previously puzzling observations: the re-equilibration of the oligomeric distribution after fractionation [31], the charge micro-heterogeneity of α-crystallin [24], and gradual shifts in its molecular weight during ageing [32]. Importantly, the fact that αB-crystallin exchanges subunits and is co-expressed with αA-crystallin in the eye lens, and with other sHSPs in non-lenticular tissues, means that oligomers consisting solely of αB-crystallin subunits probably do not exist at significant abundances *in vivo* [33]

The rate of subunit exchange was revealed more precisely a few years later by using more sensitive fluorescence techniques, in which homo-oligomers of αA-crystallin labelled with either a fluorescence donor or acceptor we found to reach an equilibrium of hetero-oligomers within an hour at 37 °C [34]. Further refinement in the ability to monitor subunit exchange came from anion exchange chromatography [35], and MS experiments [36], both of which have the advantage of not requiring a potentially disruptive bound label. The attraction of the MS experiment is that, by combining iso-structural samples (differentiated through metabolic labelling), the precise change in mass can be monitored, without being obscured by potential conformational effects. Measured in this way, the subunit exchange rate of αB-crystallin, in ammonium acetate buffer over the pH range 5-9 and at temperatures between 24-50 °C, was found to vary by four orders of magnitude [37].

4.3.3 Fast subunit exchange can compromise measurements of polydispersity

The rate of subunit exchange is a crucial parameter for αB-crystallin oligomer distributions, because it can indirectly affect measurements of the size of αB-crystallin oligomers. As described in section 2.1, early AUC data were erroneously interpreted to conclude that the different oligomeric states of αB-crystallin are not at equilibrium, because the timescale of the experiments was far shorter than the rate of oligomeric inter-conversion. Accordingly, the converse situation must also hold true: if the rate of subunit exchange is faster than the timescale of the experiment, then unanticipated effects may occur.

To illustrate this effect we have obtained SEC traces for a modified version of αB-crystallin (αB3D), in which three serines have been mutated to aspartic acids to mimic phosphorylation [42]. This protein populates a polydisperse ensemble centred on a 28mer that is broadly similar to the wild type but with a reduced preference for even-numbered oligomers [40], and exchanges subunits an order of magnitude faster than the wild-type [43]. At 4 °C αB3D elutes as

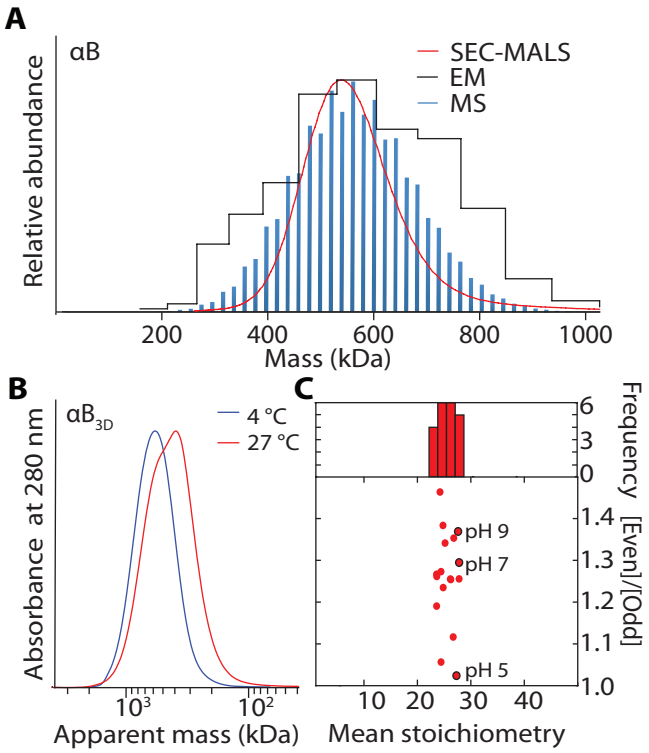


Figure 4.1: A) Stoichiometry of αB-crystallin as measured by means of different biophysical techniques. MS (blue bars) resolves individual stoichiometries, showing a slight preference for even-numbered oligomers in the distribution at pH 7. These MS data are in good agreement with low-resolution measurements from SEC-MALS [38] (red line), and negative-stain EM [39] (black histogram). B) The influence of temperature on a SEC trace obtained for a αB-crystallin triple phosphomimic mutant (αB3D) known to rapidly exchange subunits. The experiment was performed on a gel-filtration column (10/30 Superdex 200, GE Healthcare), prepared as described previously [40]. At 27 °C the elution time is longer than at 4 °C, due to much faster subunit exchange. This falsely gives the impression of a shift in molar mass of the oligomers, which actually remains unaltered [40]. C) Plot of the mean stoichiometry for a large number of αB-crystallin distributions measured by means of MS [29, 37, 41]. The mean number of subunits comprising αB-crystallin oligomers remains confined to a narrow range between around 24 to 28 subunits, whereas there is significant variation in the ratio of even- to odd-numbered oligomers in the distributions. Circled data-points correspond to wild-type αB-crystallin at pH 5, 7, and 9, respectively.

a single broad but symmetric peak corresponding to a mass of approximately 600 kDa on a calibrated SEC column (Fig. 4.1B, red trace), consistent with published SEC-MALS data [40]. Performing the experiment at 27 °C results in an appreciable change in the SEC trace, with the peak having shifted to longer elution times and being notably asymmetric (Fig. 4.1B, blue trace). This change in elution time translates into an apparent mass decrease in excess of 150 kDa (>25%), yet mass spectra obtained after SEC at both temperatures are identical. This seeming conflict is readily resolved by noting that during SEC at 27 °C subunit exchange is much more rapid than at 4 °C. At the higher temperature, the dissociation rate of subunits from the αB3D oligomers is large enough that they undergo a significant number of dissociation and re-association steps over the 25 min SEC run, each of which acts to retard them. Cumulatively this acts to causes a displacement, broadening, and skewing of the elution profile, which could easily be misinterpreted as a shift in oligomeric state to smaller stoichiometries [44, 45, 46, 47].

In general, therefore, techniques based on mass transport, such as SEC, AUC (both sedimentation velocity and sedimentation equilibrium experiments) and native gels, will give potentially misleading results if the rate of oligomer dissociation is comparable to or faster than the timescale of the measurement. The biophysical concept and mathematical formulation of this phenomenon is well established, even if not very accessible [48, 49]. However, together with the sensitivity of the αB-crystallin subunit exchange rate to solution conditions [50, 37] and mutation [41, 35, 43], neglect of this effect provides an explanation for at least some of the discrepancies in oligomeric state in the literature.

4.3.4 Fast kinetics require faster measurements of the oligomeric distribution

Nanoelectrospray MS, while also a mass-transport approach, can be thought of as a desolvation-jump experiment: proteins are delivered from solution into

the vacuum of the instrument in a rapid process during which individual protein molecules are isolated from each other on the order of milliseconds [51]. The speed at which protein molecules are isolated from one another effectively freezes the solution-phase oligomeric distribution, and allows for accurate quantification of the stoichiometries that are populated at equilibrium [38]. In this way, the effect on stoichiometry caused by perturbations known to dramatically increase the subunit exchange rate of αB-crystallin can be readily assessed. We have collated here a series of oligomeric distributions for αB-crystallin, including the wild-type protein in different solution conditions [37], and a range of point mutants [29, 41]. It is apparent that in all these cases the average number of subunits comprising the oligomeric ensemble remains approximately constant (Fig. 4.1C). Functionally, this might be important to maintain oligomeric integrity during stress conditions, or to ensure that post-translational modifications accumulated during ageing do not lead to aggregation and, as a result, opacity of the eye lens.

This data set reveals one parameter that does vary significantly however: the ratio of even- to odd-numbered oligomers. For instance, at pH 7 this [Even]/[Odd] ratio is 1.3, but at pH 5 is decreased by 22% (Fig. 4.1C). This reveals the dimeric sub-structure of the oligomers [29] and can be interpreted in terms of a modulation of the strength of the dimer interface [37]. This also provides a rationale post hoc as to why αB-crystallin isolated from eye lenses at low temperatures [52] does not display the same preference for even-numbered oligomers as the recombinant protein [29]. Namely, notwithstanding any covalent modifications they may cause [53], the high concentrations of denaturant that are typically used in the chromatographic separation of αB- from αA-crystallin and post-translationally modified forms stimulate complete dissociation of the oligomers and unfolding of the constituent monomers [54]. If equilibrium is not re-established thereafter, for instance if refolding and re-assembly is carried out at low temperature, this can cause altered chaperone activity [55]. This experiment demonstrates how it is necessary to not only

consider the dissociation and association kinetics relative to the timescale of the experiment, but also whether the distribution is already at equilibrium, or if it may be kinetically trapped. Despite these caveats, it is reassuring to note that when measurements of stoichiometry and subunit exchange rate are combined, many of the apparent conflicts in the literature can be resolved. This has led to an emerging consensus of αB-crystallin populating a range of stoichiometries between 10 and 50 subunits, with subunit exchange kinetics that are highly dependent on solution conditions and modification of the protein.

4.4 Kinetic and allosteric determinants of the αB-crystallin ensemble.

Ever since the polydispersity of αB-crystallin was established, many qualitative models have been put forward in an attempt to rationalise its size distribution. The majority of these models have sought explanations from a structural standpoint: for instance, a micellar quaternary architecture for αB-crystallin oligomers, because micelles themselves exhibit size heterogeneity [21, 56]. Although such structural hypotheses may provide a rationale as to why αB-crystallin is polydisperse, they make few predictions as to the nature of the distribution and its response to physical perturbation. As our knowledge of the equilibrium properties of αB-crystallin has evolved, and in particular our understanding that polydispersity and quaternary dynamics are linked, more predictive models can be constructed. In this way, a quantitative biophysical model was formulated to estimate the shape of the distribution from purely kinetic considerations, with minimal assumptions as to the proposed structures of the oligomers [37].

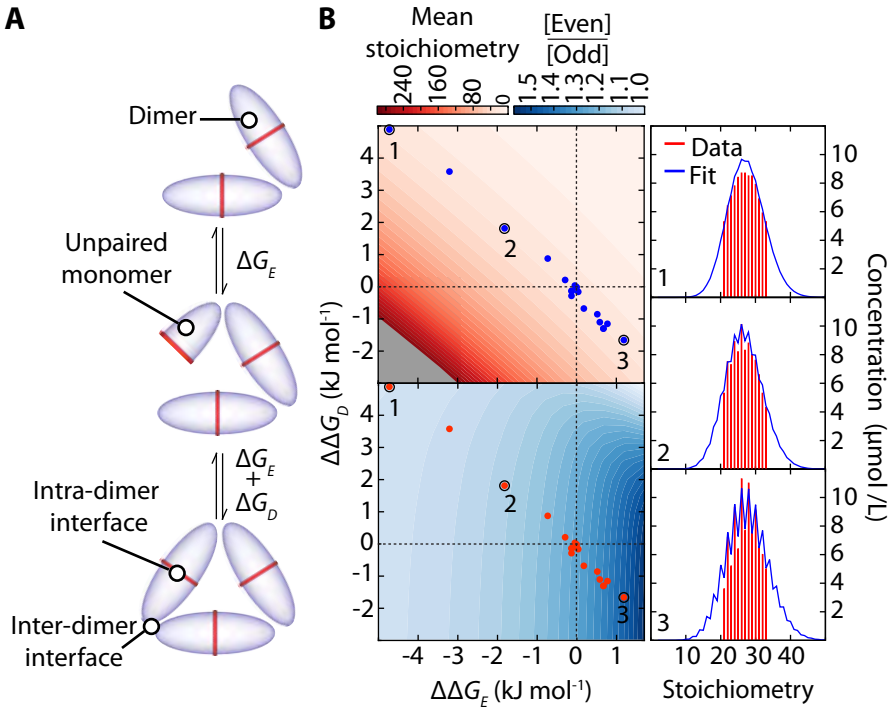


Figure 4.2: Allosteric interaction between two interfaces in αB-crystallin. A) Schematic diagram of the free energies involved in monomer dissociation. In an even-numbered oligomer, two interfaces must break for monomer dissociation: the inter-dimer interface (ΔG_E) and the intra-dimer interface (ΔG_D). In an odd-numbered oligomer only the latter has to break for oligomer dissociation, releasing an unpaired monomer. For simplicity, the schematic is drawn for a hypothetical hexamer, but holds for any oligomer. B) The effect of varying the strength of $\Delta\Delta G_E$ and $\Delta\Delta G_D$ on the average stoichiometry of αB-crystallin oligomers (upper panel), and the ratio between even- and odd-numbered oligomers in theoretical αB-crystallin distributions (lower panel). $\Delta\Delta G_D$ and $\Delta\Delta G_E$ are calculated relative to the wild-type distribution at 37 °C, pH 7 (dashed lines). Negative values indicate a strengthening of the relevant interface, while positive values indicate weakening. Data points are from fits to experimental data obtained by varying the pH [37], or introducing mutations into the C-terminal tail [41], or to mimic phosphorylation [29]. The data points lie on a straight line, keeping the average oligomeric state constant and varying the even/odd ratio through a tight negative correlation between $\Delta\Delta G_E$ and $\Delta\Delta G_D$. The right-hand panels show the data (red bars) and fits (blue lines) for the three data points highlighted in the contour plots.

4.4.1 The exchange model for the origin of αB-crystallin polydispersity

The exchange model is based on the fundamental hypothesis put forward in the 1970s that existing aggregates [oligomers] could be in continuous equilibrium with their monomeric subunits allowing random exchange of polypeptide chains between aggregates [oligomers] [24]. In this way the distribution can be described as an equilibrium of many coupled exchange reactions, each consisting of oligomers of a particular stoichiometry gaining or losing monomeric subunits. The biophysical framework of the model essentially relies on three simple premises [37]: 1) the association (on) rate constant for a monomer binding to an oligomer is invariant with stoichiometry. In other words, a free monomer is equally likely to associate with oligomers of all sizes in the distribution, in a manner that depends solely on the concentrations of the two partners. 2) The dissociation (off) rate of each monomer in any oligomer is directly proportional to the stoichiometry of the oligomer. In other words, all the monomers in the polydisperse ensemble have the same probability of dissociating, but larger oligomers are more likely to lose a monomer at any given time simply because they comprise more. These assumptions (1 and 2) combine, when the on-rate is faster than the off-rate, to make both very small and very large oligomers short lived, naturally leading to a bell-shaped distribution with defined maximum.

This is in accordance with low-resolution measurements of the polydisperse ensemble that provide a mean stoichiometry and associated width of the distribution. However, as described in Section 2.4, this bell-curve is not necessarily smooth: at many conditions (including the wild-type protein at pH 7) oligomers containing an even number of subunits are more abundant than their odd counterparts (Fig. 4.1C). This observation led to the introduction of an additional assumption into the exchange model: 3) monomers make two sets of contacts, one to form a dimer (i.e. intra-molecular to the dimer, ascribed the free energy ΔG_D) and another between dimers, such that they build the oligomers (inter-

molecular to the dimer, ΔG_E). The oligomers are consequently assembled such that all monomers make the maximum number of contacts to minimize both free energies (Fig. 4.2A).

The exchange model allows the shape of the distribution of αB-crystallin ensembles to be fully described in terms of just these two quantities. Physically, this requires that the conformation of a monomer must be similar across all different sizes of oligomers, a proposition that is validated by both solid- and solution-state NMR data that find single peaks for most residues despite the heterogeneous ensemble [57, 58, 59]. Another prediction of the model is that the average oligomer size should be approximately constant over a wide concentration range, an observation that has been made by means of AUC [60, 61] and SEC [62] in the range $10^{-6} - 10^{0}$ g/mL. Importantly, the exchange model does not assign the interactions to particular regions of the protein, but relies solely on conceptual definitions. In general the model, despite its simplicity, reproduces the available data extremely well, and provides an important biophysical link between the quaternary dynamics and polydispersity of αB-crystallin.

4.4.2 Allosteric communication between αB-crystallin interfaces

The practical utility of the exchange model lies in its ability to quantify changes in the thermodynamics and kinetics of αB-crystallin upon perturbation, for instance through change in solution conditions, post-translational modification, or mutation. Here we have interrogated our set of oligomeric distributions of αB-crystallin (Fig. 4.1C) to extract the free energies of the inter- and intra-dimer interfaces, and compared this to the wild-type protein at standard conditions (Fig. 4.2B). Changes significantly greater than the thermal energy in the strength of these interfaces are observed. Most strikingly, however, is that there is a clear negative linear correlation between $\Delta\Delta G_E$ and $\Delta\Delta G_D$. In other words, weakening of one interface strengthens the other, and vice versa.

This is a remarkable phenomenon, as it demonstrates that despite their separation, there is a flow of energy, i.e. allostery, between these two sites. The shift in the relative contributions of ΔG_E and ΔG_D to the overall stability of the oligomers then produces changes in the ratio even- to odd-numbered oligomers, with larger contributions of ΔG_D leading to a stronger bias towards the former. A notable consequence of this allosteric communication is that the sum of the two free energies (ΔG_E+ΔG_D) remains approximately constant, with the corollary that the overall stoichiometry does not change significantly upon perturbation of the interfaces that are responsible for oligomerisation (Fig. 4.1C).

4.5 Dynamical structure of the αB-crystallin interfaces

One difficulty in delineating the mechanism of allostery is that it is not fully understood how precisely the αB-crystallin dimers are held together in the oligomers (i.e. what defines ΔG_D structurally). Indeed, it appears that there are contributions from at least both the N- and C-terminal regions of αB-crystallin. While it has only been possible to get relatively sparse direct structural information about the former due to its conformational heterogeneity [58, 63], considerable insight has been obtained as to the structure and dynamics of the latter.

4.5.1 The C-terminal tail cross-links dimers into oligomers

The crystal structures of a construct of αB-crystallin truncated of its N-terminal region revealed the C-terminal tail to bind onto an adjacent monomer, effectively amounting to cross-links between dimers [64]. This inter-monomer interaction was validated in the full-length protein by means of solid-state NMR experiments [58], and has been replicated through the development of a system

composed of the αB-crystallin core domain (ACD) dimer together with a peptide mimicking its palindromic C-terminal region. This reductionist system has proven very useful, both as a means for improving crystallographic resolution [65], and assaying the affinity of the C-terminal interaction [66, 41].

The details of the contacts made at this interface have proven extremely interesting. The structural data demonstrate that the tail can bind diagonally across one side of the β-sandwich, forming hydrogen bonds with the β4 and β8-strands, with a conserved IXI motif docking into a hydrophobic groove between them. Interestingly, solution state NMR experiments showed that, within the oligomers at physiological temperature, the IXI motif is bound in this way with a population of only about 2%. The remainder of the time it is detached, but largely remains localised to the vicinity of the groove by the upstream interactions the tail makes with the β-strands [57]. Interestingly, the proportion of bound versus unbound states is strongly temperature dependent [67], and can be correlated very well with the rate of monomer dissociation from the oligomer [57], suggesting a regulatory function [58].

4.5.2 C-terminal transitions mediate oligomer disassembly

As such, somewhat counter-intuitively, binding of the IXI is linked to dissociation of a monomer [67]. Reconciliation of this apparent paradox comes from experiments in which the binding of a palindromic peptide, containing the IXI from the C-terminus (Fig. 4.3A), to the ACD was quantified. The ACD populates an equilibrium of monomers and dimers at low micromolar concentrations (Fig. 4.3B), so the advantage of a MS-based experiment over more traditional titration approaches is that dissociation constants can be obtained for each co-populated binding stoichiometry [41]. In this manner we found that the ACD dimer binds a first peptide equivalent with a dissociation constant (KD) of approximately 70 μM, and a second at a substantially higher value (Fig. 4.3B, upper panel). The average of these K_Ds is in close accord with the ensemble

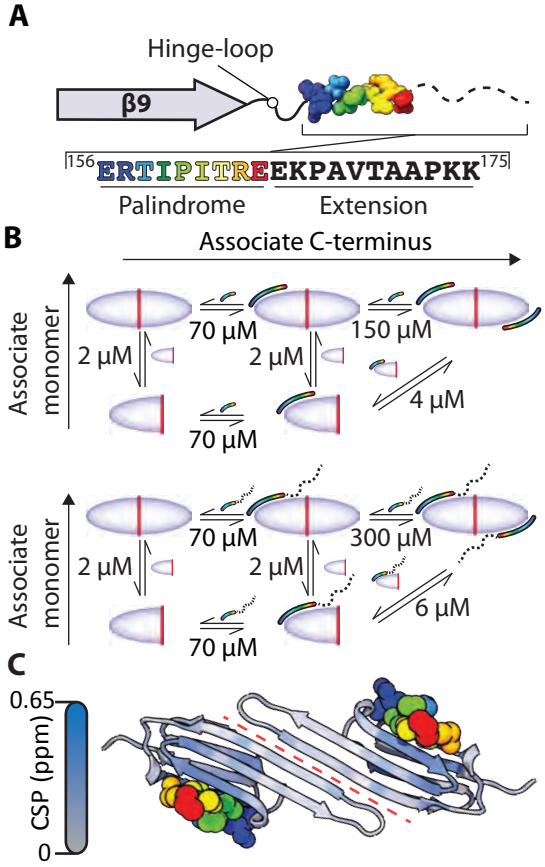


Figure 4.3: Allosteric effect of C-terminal binding to the ACD. A) Primary structure of the C-terminus. A flexible hinge loop connects the palindromic section of the C-terminal tail with the $\beta 9$ strand. The sequence (residues 165-175) after the palindrome is an unstructured extension containing several charged. B) Allosteric binding of the C-terminus to the core domain of αB-crystallin. The upper panel shows the six coupled equilibra involved in monomers and dimers interconverting, and binding the palindromic peptide from the C-terminus (up to one equivalent per monomer). The lower panel shows the same set of equilibria, but for a peptide that includes both the palindrome and the flexible extension. Values under the arrows represent K_Ds obtained from ion-mobility MS experiments in which all the different species can be separated and quantified [41]. Binding of the second peptide to a dimer is substantially more difficult than binding of the first, with concomitant weakening of the dimer interface. This indicates an allosteric effect upon binding, and effect is more pronounced in the longer peptide (lower panel). C) Chemical shift perturbations induced by peptide binding to the ACD, as observed by means of NMR [66]. The majority of the changes are observed near the peptide-binding interface, but propagate from here all the way down to the dimer interface, potentially hinting at a pathway for allosteric communication.

measurement obtained by means of NMR [66]. The discrepancy in affinity for binding the first and second peptide equivalent is much increased if the peptide incorporates the extension (residues 165-175) of the C-terminal tail (Fig. 4.3A), which is known to be highly dynamic [68], yielding a K_D that is four times higher for the second binding event (Fig. 4.3B, lower panel). This can be rationalised by steric and coulombic repulsions between the flexible C-terminal extensions, which would be expected to be even more significant in the context of the αB-crystallin oligomers where many monomers are inevitably in close proximity [6].

Concomitant to the difficulty in binding multiple peptide equivalents, we observed that the dimer interface itself was considerably weaker when two, rather than fewer, peptides were bound (Fig. 4.3B). The net effect of this system of coupled equilibria is that binding of two peptides to a dimer effectively drives it towards dissociation into monomers [6]. The binding of the C-terminal region to the β4 and β8 strands, and groove between them, therefore has an allosteric effect on the dimer interface of αB-crystallin, in agreement with the observations made for the full-length protein (Fig. 4.2B). Residue-level details as to this phenomenon come from similar experiments performed using NMR, in which binding of the palindromic peptide induces chemical shift perturbations in spectra of the ACD [66]. Even though these perturbations are relatively small, they propagate from the peptide-binding site all the way to the dimer interface (Fig. 4.3C). This could conceivably hint at the allosteric pathway through which energy is transferred to the dimer interface upon binding of the C-terminal tail.

4.5.3 Variable binding of the C-terminal tail

The αB-crystallin C-terminus has the capacity for considerable variation in its interaction with the ACD. Notably, two different inter-molecular orientations across the β-sandwich have been observed (Fig. 4.4A), running either parallel or antiparallel to the β8 strand, for both αB- and αA-crystallin [65, 64](Fig.

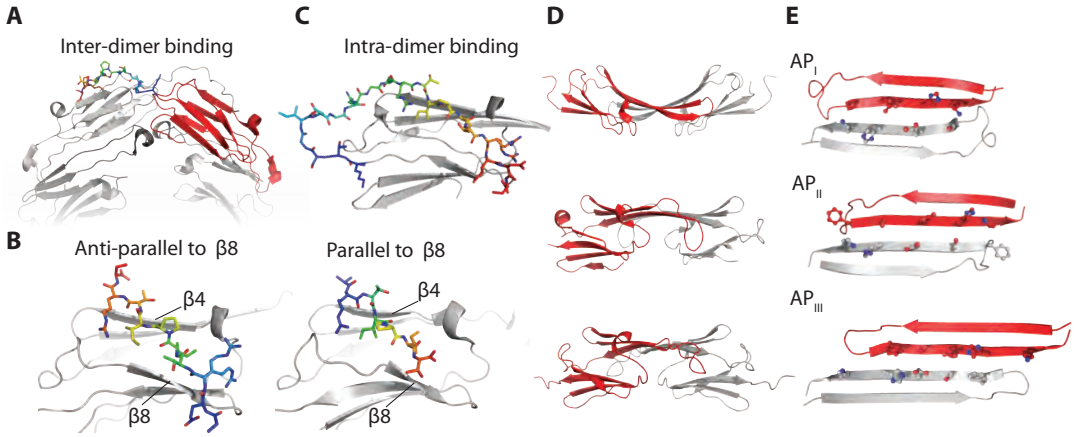


Figure 4.4: Different registration states and flexibility of the ACD captured by crystallography. A) Inter-dimer contacts made by the C-terminus in human αB-crystallin (PDB: 3L1G), with the C-terminus of a monomer binding to the ACD of a monomer in an adjacent dimer. Colouring of the C-terminal tail, in the N- to C-terminus direction is from blue to red. B) Bi-directional binding of the C-terminal tail to the core domain is facilitated by the palindromic sequence of the peptide (colouring as in A), allowing pseudo-equivalent interactions in either orientation. C) Intra-dimer tail-binding in the closely related αA-crystallin (zebrafish, PDB: 3N3E) is enabled by the C-terminus of a monomer looping back to attach onto its own ACD. D) Curvature of the β-sheets forming the dimer interface is observed in different structures of the ACD. The most bent structure is observed in a solid-state NMR model of the full-length protein (upper panel, human, PDB: 2KLR), and less bent conformations observed in crystallographic studies of truncated constructs (middle and lower panels, both human, PDB: 3L1G, 2WJ7). E) Different registration states of the human αB-crystallin ACD, labelled AP_I to AP_{III} (respectively, PDB: 3L1G, 2WJ7, 4M5S), have been observed in crystallographic studies. The AP_{II} state is that which is predominantly populated in solution, but the possibility remains that transient excursions to the other registers may play a functional role.

4.4B). This bi-directional binding has also been observed in solution, with NMR measuring the three possible combinations of two peptides binding to a dimer approximately at the 1:2:1 ratio (both parallel : one parallel and one antiparallel : both antiparallel) that would be expected statistically if both binding modes were equally favourable [66].

Notwithstanding amino-acid chirality, this phenomenon can be explained by the palindromic sequence of the peptide leading to pseudo-equivalent interactions in either direction. However, it is unlikely that the two orientations are equally favourable in full-length αB-crystallin, due to conformational re-

strictions on the tail made by the hinge-loop that connects it to the ACD [69]. This has been demonstrated by alanine scanning of the palindromic sequence, which revealed that mutation of ^{156}ERT^{158} to three alanines had a five-fold larger destabilisation effect on the interface than equivalent mutation of ^{162}TRE^{164} [41]. Another intriguing hypothesis as to conformations adopted by the C-terminus comes from a structure of zebrafish αA-crystallin, in which the tail binds to the same site, but intra-molecularly by looping back onto itself [70] (Fig. 4.4C). This binding mode has not yet been observed directly for αB-crystallin but, even when noting that the zebrafish αA-crystallin has a hinge loop three amino acids longer than human αB-crystallin, is geometrically feasible.

In order to relate these crystallographic data to our understanding of full-length αB-crystallin dynamics we can consider spectroscopic data that indicate that the most populated (ground) arrangement of the C-terminal tail is inter-molecular [71, 58, 72]. Oligomer dissociation then necessitates the breaking of two C-terminal interactions, as any given monomer donates one tail and accommodates another. Accordingly, subunit exchange rates are predicted very well by coincident transitions of two C-termini from their ground state to an excited one [71]. While it is clear that in this excited state the IXI is bound, whether this is inter- or intra-molecular is not known [67].

As a means to reconcile the currently available data, we put forward the following hypothesis. In its ground state the C-terminal tail makes an inter-molecular contact with the IXI localized over, but unbound to, a binding pocket on an adjacent monomer. This state is in equilibrium with two lowly populated ones, in both of which the IXI is bound to the hydrophobic groove, but one is inter-molecular, and the other intra-molecular (Fig. 4.5, middle panel). In the event of two tails on neighbouring monomers transitioning to the intra-molecular configuration a monomer is released, catalysed by the allosteric communication between the two interfaces. An alternative rationale could be that two coincident transitions from the ground state to the inter-molecular IXI

bound state could also lead to dissociation of the monomer if they caused a conformational distortion that compromised other inter-dimer contributions to ΔG_E made, for instance, by the N-terminus [67]. In either case, the C-terminal dynamics would effectively rate-limit the subunit exchange of oligomers.

4.5.4 Polymorphic dimeric interfaces

The dynamic nature of αB-crystallin appears not to be confined to transitions of its terminal regions, but also to its protomeric building block, the ACD dimer. When, after decades of frustrating crystallographers, the structure of the ACD was finally solved it presented the community with several surprises that had been hinted at previously [73, 59, 74]. Specifically, the dimer interface of αB-crystallin (and other mammalian sHSPs) is markedly different to that from lower organisms, in that its β6 and β7 strands combine into one long β6+7 strand that pairs in an antiparallel orientation [75]. Notably there is substantial variation of angle between the two β6+7 strands of dimers in the different atomic-resolution structures of αB-crystallin [76, 65] (Fig. 4.4D). Furthermore, the registration state of this dimer interface varies between different structures, with three different registers having been observed for αB-crystallin [75, 65, 64]. These are shifted relative to one another by two residues in the β-sheet and are termed AP_I to AP_{III} [64], from the most compact to the most extended, respectively (Fig. 4.4E).

Whether all three registration states are populated in solution, or are merely crystallographic artefacts, is an intriguing question. Solid state NMR experiments of αB-crystallin oligomers only gave evidence for the APII state [58], while both NMR [66, 59] and ion mobility MS [65] investigations also failed to detect alternative registers. Nonetheless, the possibility of lowly populated, rapidly exchanging registration states cannot be ruled out, especially in the context of the oligomer. What is clear however is that the dimer interface is weak, with a K_D on the order of a few micromolar (Fig. 4.3B), and that the ACD dimer exchanges subunits extremely rapidly, with an off-rate $>10^{-1}$ s^{-1}

at 4 °C [65]. This suggests that no large rearrangements are necessary for two monomers to assemble into a dimer, and paint the picture of a dynamic and labile dimer interface. Indirect evidence that such flexibility may be important comes from studying the interaction of αB-crystallin with aggregating target proteins. We recently found that, by using an engineered disulphide bond [65], the ACD of αB-crystallin inhibits the aggregation of the amyloidogenic peptide $A\beta_{1-42}$ somewhat more effectively if it is covalently locked into AP_{II}, suggesting that the exact conformation of the dimer interface may yet prove to be relevant to chaperone function.

4.6 Integrative biology of αB-crystallin

In this review we have attempted to paint the dynamical landscape of αB-crystallin. A picture has emerged in which we have to consider fluctuations varying in both timescale and amplitude. These include, but are unlikely to be limited to: the subunit exchange of oligomers mediated by their dissociation (rate constants at physiological pH and temperature ranges: $10^{-1} > k > 10^{-2}$ s^{-1}); the breaking of the dimer interface in oligomer ($k > 10^{1}$ s^{-1} at 4 °C); and transitions of the C-terminal tails ($10^{2} > k > 10^{1}$ s^{-1}) (Fig. 4.5). All these dynamics appear to be key features of αB-crystallin, but do they, together with our knowledge of the distribution of stoichiometries, and atomic structures of the ACD dimer, allow us to construct useful models of the assembled protein?

4.6.1 Forging consensus structures of αB-crystallin oligomers

Fundamentally, any structural model should satisfy as many experimental observations as possible, but ideally also provide a rationale for previously unexplained aspects of the protein under investigation. Similarly, it is of course desirable that the model provides hypotheses for further testing. Over the

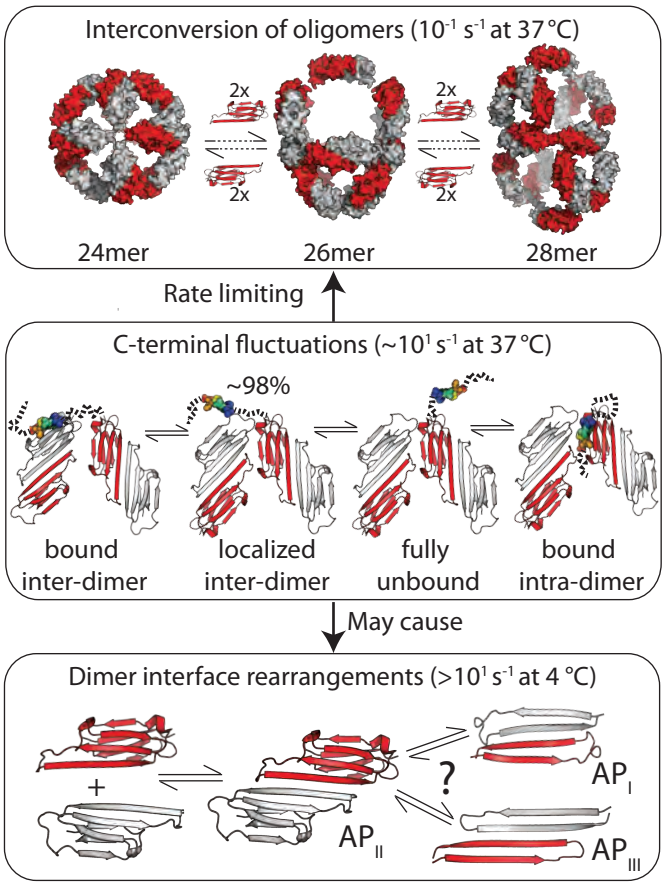


Figure 4.5: A model of how the dynamic motions of αB-crystallin at three different time-scales are inter-related. The C-terminus is localized to an adjacent dimer with the IXI unbound for the majority of time, but converts on the millisecond time scale into a bound conformation that can be either inter- or intra-molecular (middle panel). This tail-binding may induce distortions in the dimer interface that lead to rearrangements including breaking of the dimer interface or registration shifts (lower panel). Together these two effects determine the rate of subunit exchange between higher-order oligomers, which is ultimately rate-limited by C-terminal fluctuations (upper panel). For clarity, only pseudo-atomic models of even-numbered oligomers are shown, although subunit exchange proceeds via the dissociation of monomers.

last decades many structural models for αB-crystallin oligomers have been proposed, and these have evolved in sophistication as ever more experimental data has been obtained. With the advent of high-resolution structures of the ACD [75, 77, 64], three contrasting pseudo-atomic models of the oligomers have been proposed [39, 78, 63]. In all three the modelling strategy centred on combining spatial restraints obtained on the oligomeric states with an estimate of the oligomeric distribution, in order to assemble the protomeric dimers into oligomers. However, using experimental data from different techniques, the resulting three models differ considerably, and all have their merits and deficiencies [79, 6].

The reasons for this lack of consensus lie in the difficulties that the intrinsic polydispersity of αB-crystallin poses, with successive oligomeric states varying by only about 1% in diameter. Traditional structural biology approaches typically provide an ensemble average and, in cases where at least some separation is possible, it is currently difficult if not impossible to achieve it at the resolution required for studying individual stoichiometries of αB-crystallin. Because of this, the three pseudo-atomic models in the literature rely on (differing) assumptions of symmetry and architecture to aid the deconvolution of the experimental data, leading to an inevitable bias in the final models. The challenge for the future is to integrate all the current structural data on the oligomers, from ion mobility MS [39], SAXS [63], and EM [80], as well as orthogonal approaches, to determine unbiased consensus structures.

4.6.2 αB-crystallin and the dynamics-function paradigm

Over the past few years significant progress has been made in describing the structure and dynamics of the αB-crystallin ensemble, stemming directly from exciting technological advances in the bioanalytical approaches used. While all the ensuing data have not yet been reconciled into a fully coherent picture, we hope to have shown here how further progress crucially depends on an clear understanding of αB-crystallins complex dynamics that, if neglected, can lead

to measurements that do not reflect the equilibrium state of the polydisperse ensemble. This principle will certainly be equally important to other areas of structural biology, as increasingly complex and dynamic protein assemblies are investigated.

The emerging view of a dynamic ensemble of αB-crystallin oligomers poses important questions about its functional significance. Subunit exchange has been characterized in other proteins, including the β-crystallins of the eye lens [81, 82], but in much less detail than in the sHSPs. Similarly, polydisperse yet globular quaternary architectures are not well documented outside of the sHSP family, so analogies to better-understood protein families cannot be easily drawn. It has long been assumed that these properties must be integral to αB-crystallin chaperone function, but recently some evidence is emerging that this relationship may not be straightforward. For instance, αB-crystallin can function efficiently as a chaperone in forms that either cannot oligomerise beyond dimers [65], or cross-linked into large oligomers that cannot dissociate [83]. The relationship between the characteristic quaternary dynamics and structure of αB-crystallin and its cellular functions represent the next frontier of research in the field, where a particular challenge will be to relate the *in vitro* data to the situation in the cell. The data that have recently emerged on αB-crystallin, and we have described here, will provide excellent springboard for these investigations, and as a means to probe the emerging dynamics-function paradigm in structural biology.

4.7 References

[1] J. Horwitz, "α-crystallin can function as a molecular chaperone.," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 89, pp. 10449–10453, Nov. 1992.

[2] U. Jakob, M. Gaestel, K. Engel, and J. Buchner, "Small Heat-Shock Proteins Are Molecular Chaperones," *J. Biol. Chem.*, vol. 268, pp. 1517–1520+, 1993.

[3] P. J. Muchowski, J. A. Bassuk, N. H. Lubsen, and J. I. Clark, "Human αB-crystallin. Small heat shock protein and molecular chaperone.," *J. Biol. Chem.*, vol. 272, pp. 2578–2582, Jan. 1997.

[4] J. Lowe, H. McDermott, I. Pike, I. Spendlove, M. Landon, and R. J. Mayer, "α B crystallin expression in non-lenticular tissues and selective presence in ubiquitinated inclusion bodies in human disease.," *J. Pathol.*, vol. 166, pp. 61–68, Jan. 1992.

[5] E. Basha, H. O'Neill, and E. Vierling, "Small heat shock proteins and α-crystallins: dynamic proteins with flexible functions.," *Trends Biochem. Sci.*, vol. 37, pp. 106–117, Mar. 2012.

[6] G. R. Hilton, H. Lioe, F. Stengel, A. J. Baldwin, and J. L. Benesch, "Small heat-shock proteins: paramedics of the cell.," *Top. Curr. Chem.*, vol. 328, pp. 69–98, 2013.

[7] H. S. McHaourab, J. A. Godar, and P. L. Stewart, "Structure and mechanism of protein stability sensors: chaperone activity of small heat shock proteins.," *Biochemistry*, vol. 48, pp. 3828–3837, May 2009.

[8] H. Ecroyd and J. A. Carver, "Crystallin proteins and amyloid fibrils.," *Cell. Mol. Life. Sci.*, vol. 66, pp. 62–81, Jan. 2009.

[9] A. Boncoraglio, M. Minoia, and S. Carra, "The family of mammalian small heat shock proteins (HSPBs): implications in protein deposit

diseases and motor neuropathies.," *Int. J. Biochem. Cell Biol.*, vol. 44, pp. 1657–1669, Oct. 2012.

[10] W. E. Balch, R. I. Morimoto, A. Dillin, and J. W. Kelly, "Adapting proteostasis for disease intervention," *Science*, vol. 319, pp. 916–9+, 2008.

[11] F. U. Hartl, A. Bracher, and M. Hayer-Hartl, "Molecular chaperones in protein folding and proteostasis," *Nature*, vol. 475, pp. 324–332, July 2011.

[12] K. Richter, M. Haslbeck, and J. Buchner, "The heat shock response: life on the verge of death.," *Mol. Cell*, vol. 40, pp. 253–266, Oct. 2010.

[13] A. Tardieu, "Eye lens proteins and transparency: from light transmission theory to solution X-ray structural analysis.," *Annu. Rev. Biophys. Biophys. Chem.*, vol. 17, pp. 47–70, 1988.

[14] H. Bloemendal, W. de Jong, R. Jaenicke, N. H. Lubsen, C. Slingsby, and A. Tardieu, "Ageing and vision: structure, stability and function of lens crystallins.," *Prog. Biophys. Mol. Biol.*, vol. 86, pp. 407–485, Nov. 2004.

[15] J. Horwitz, "Proctor Lecture. The function of α-crystallin.," *Invest. Ophthalmol. Vis. Sci.*, vol. 34, pp. 10–22, Jan. 1993.

[16] A. R. Clark, N. H. Lubsen, and C. Slingsby, "sHSP in the eye lens: crystallin mutations, cataract and proteostasis," *Int. J. Biochem. Cell Biol.*, 2012.

[17] J. P. Brady, D. L. Garland, D. E. Green, E. R. Tamm, F. J. Giblin, and E. F. Wawrousek, "αB-crystallin in lens development and muscle integrity: a gene knockout approach.," *Invest. Ophthalmol. Vis. Sci.*, vol. 42, pp. 2924–2934, Nov. 2001.

[18] K. M. Forrest, S. Al-Sarraj, C. Sewry, S. Buk, S. V. Tan, M. Pitt, A. Durward, M. McDougall, M. Irving, M. G. Hanna, E. Matthews,

A. Sarkozy, J. Hudson, R. Barresi, K. Bushby, H. Jungbluth, and E. Wraige, "Infantile onset myofibrillar myopathy due to recessive CRYAB mutations.," *Neuromuscul. Disord.*, vol. 21, pp. 37–40, Jan. 2011.

[19] J. Piatigorsky and G. Wistow, "The recruitment of crystallins: new functions precede gene duplication.," *Science*, vol. 252, pp. 1078–1079, May 1991.

[20] A. Spector, "The soluble proteins of the lens.," *Invest. Ophthalmol.*, vol. 4, pp. 579–591, Aug. 1965.

[21] J. A. Thomson and R. C. Augusteyn, "On the structure of α-crystallin: construction of hybrid molecules and homopolymers.," *Biochim. Biophys. Acta*, vol. 994, pp. 246–252, Feb. 1989.

[22] P. J. van den Oetelaar, P. F. van Someren, J. A. Thomson, R. J. Siezen, and H. J. Hoenders, "A dynamic quaternary structure of bovine α-crystallin as indicated from intermolecular exchange of subunits.," *Biochemistry*, vol. 29, pp. 3488–3493, Apr. 1990.

[23] A. Spector, L. K. Li, R. C. Augusteyn, A. Schneider, and T. Freund, "α-Crystallin. The isolation and characterization of distinct macromolecular fractions.," *Biochem. J.*, vol. 124, pp. 337–343, Sept. 1971.

[24] R. J. Siezen, J. G. Bindels, and H. J. Hoenders, "The quaternary structure of bovine α-crystallin," *Eur. J. Biochem.*, vol. 91, pp. 387–396, Nov. 1978.

[25] D. A. Haley, J. Horwitz, and P. L. Stewart, "The small heat-shock protein, αB-crystallin, has a variable quaternary structure.," *J. Mol. Biol.*, vol. 277, pp. 27–35, Mar. 1998.

[26] D. A. Haley, J. Horwitz, and P. L. Stewart, "Image restrained modeling of αB-crystallin.," *Exp. Eye Res.*, vol. 68, pp. 133–136, Jan. 1999.

[27] G. R. Hilton and J. L. Benesch, "Two decades of studying non-covalent biomolecular assemblies by means of electrospray ionization mass spectrometry.," *J. R. Soc. Interface*, vol. 9, pp. 801–816, May 2012.

[28] J. A. Aquilina, J. L. P. Benesch, O. A. Bateman, C. Slingsby, and C. V. Robinson, "Polydispersity of a mammalian chaperone: mass spectrometry reveals the population of oligomers in αB-crystallin," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 100, pp. 10611–6+, 2003.

[29] J. A. Aquilina, J. L. Benesch, L. L. L. Ding, O. Yaron, J. Horwitz, and C. V. Robinson, "Phosphorylation of αB-crystallin alters chaperone function through loss of dimeric substructure.," *J. Biol. Chem.*, vol. 279, pp. 28675–28680, July 2004.

[30] R. L. van Montfort, E. Basha, K. L. Friedrich, C. Slingsby, and E. Vierling, "Crystal structure and assembly of a eukaryotic small heat shock protein.," *Nat. Struct. Biol.*, vol. 8, pp. 1025–1030, Dec. 2001.

[31] R. J. Siezen and E. A. Owen, "Physicochemical characterization of high-molecular-weight -crystallin subpopulations from the calf lens nucleus," *Biochim. Biophys. Acta*, vol. 749, pp. 227–237, Dec. 1983.

[32] G. J. Bessems, H. J. Hoenders, and J. Wollensak, "Variation in proportion and molecular weight of native crystallins from single human lenses upon aging and formation of nuclear cataract.," *Exp. Eye Res.*, vol. 37, pp. 627–637, Dec. 1983.

[33] E. V. Mymrikov, A. S. Seit-Nebi, and N. B. Gusev, "Heterooligomeric complexes of human small heat shock proteins.," *Cell Stress Chaperon.*, vol. 17, pp. 157–169, Mar. 2012.

[34] M. P. Bova, L. L. Ding, J. Horwitz, and B. K. Fung, "Subunit exchange of αA-crystallin.," *J. Biol. Chem.*, vol. 272, pp. 29511–29517, Nov. 1997.

[35] M. Michiel, F. Skouri-Panet, E. Duprat, S. Simon, C. Férard, A. Tardieu, and S. Finet, "Abnormal assemblies and subunit exchange of αB-crystallin R120 mutants could be associated with destabilization of the dimeric substructure.," *Biochemistry*, vol. 48, pp. 442–453, Jan. 2009.

[36] J. A. Aquilina, J. L. P. Benesch, L. L. Ding, O. Yaron, J. Horwitz, and C. V. Robinson, "Subunit Exchange of Polydisperse Proteins," *J. Biol. Chem.*, vol. 280, pp. 14485–14491, Apr. 2005.

[37] A. J. Baldwin, H. Lioe, C. V. Robinson, L. E. Kay, and J. L. P. Benesch, "αB-Crystallin Polydispersity Is a Consequence of Unbiased Quaternary Dynamics," *J. Mol. Biol.*, Aug. 2011.

[38] J. L. Benesch and B. T. Ruotolo, "Mass spectrometry: come of age for structural and dynamical biology.," *Curr. Opin. Struct. Biol.*, vol. 21, pp. 641–649, Oct. 2011.

[39] A. J. Baldwin, H. Lioe, G. R. Hilton, L. A. Baker, J. L. Rubinstein, L. E. Kay, and J. L. Benesch, "The polydispersity of αB-crystallin is rationalized by an interconverting polyhedral architecture.," *Structure*, vol. 19, pp. 1855–1863, Dec. 2011.

[40] H. Ecroyd, S. Meehan, J. Horwitz, J. A. Aquilina, J. L. Benesch, C. V. Robinson, C. E. Macphee, and J. A. Carver, "Mimicking phosphorylation of αB-crystallin affects its chaperone activity.," *Biochem. J.*, vol. 401, pp. 129–141, Jan. 2007.

[41] G. R. Hilton, G. K. Hochberg, A. Laganowsky, S. I. McGinnigle, A. J. Baldwin, and J. L. Benesch, "C-terminal interactions mediate the quaternary dynamics of αB-crystallin.," *Philos. Trans. R. Soc. Lond. B. Biol. Sci.*, vol. 368, May 2013.

[42] H. Ito, K. Kamei, I. Iwamoto, Y. Inaguma, D. Nohara, and K. Kato, "Phosphorylation-induced change of the oligomerization state of αB-crystallin.," *J. Biol. Chem.*, vol. 276, pp. 5346–5352, Feb. 2001.

[43] J. Peschek, N. Braun, J. Rohrberg, K. C. C. Back, T. Kriehuber, A. Kastenmüller, S. Weinkauf, and J. Buchner, "Regulated structural transitions unleash the chaperone activity of αB-crystallin.," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 110, pp. 201308898+, Oct. 2013.

[44] J. R. Cann, E. J. York, J. M. Stewart, J. C. Vera, and R. B. Maccioni, "Small zone gel chromatography of interacting systems: Theoretical and experimental evaluation of elution profiles for kinetically controlled macromolecule-ligand reactions," *Anal. Biochem.*, vol. 175, pp. 462–473, Dec. 1988.

[45] G. A. Gilbert and Jenkins, "Boundary Problems in the Sedimentation and Electrophoresis of Complex Systems in Rapid Reversible Equilibrium," *Nature*, vol. 177, pp. 853–854, May 1956.

[46] F. J. Stevens, "Analysis of protein-protein interaction by simulation of small-zone size exclusion chromatography. Stochastic formulation of kinetic rate contributions to observed high-performance liquid chromatography elution characteristics," *Biophys. J.*, vol. 55, pp. 1155–1167, June 1989.

[47] D. J. Winzor and H. A. Scheraga, "Studies of Chemically Reacting Systems on Sephadex. I. Chromatographic Demonstration of the Gilbert Theorys," *Biochemistry*, vol. 2, pp. 1263–1267, Nov. 1963.

[48] J. A. Cann and W. B. Goad, *Interacting macromolecules: the theory and practice of their electrophoresis, chromatography, and ultracentrifugation.* Academic Press, 1970.

[49] L. W. W. Nichol and D. J. Winzor, *Migration of interacting systems.* Oxford University Press, 1972.

[50] J. A. Aquilina, S. Shrestha, A. M. Morris, and H. Ecroyd, "Structural and functional aspects of hetero-oligomers formed by the small heat shock proteins αB-crystallin and HSP27.," *J. Biol. Chem.*, vol. 288, pp. 13602–13609, May 2013.

[51] B. T. Ruotolo and C. V. Robinson, "Aspects of native proteins are retained in vacuum.," *Curr. Opin. Chem. Biol.*, vol. 10, pp. 402–408, Oct. 2006.

[52] J. A. Aquilina, J. L. P. Benesch, O. A. Bateman, C. Slingsby, and C. V. Robinson, "Polydispersity of a mammalian chaperone: mass spectrometry reveals the population of oligomers in αB-crystallin," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 100, pp. 10611–10616, Sept. 2003.

[53] D. Volkin, H. Mach, and Middaugh, "Degradative covalent reactions important to protein stability," *Mol. Biotechnol.*, vol. 8, no. 2, pp. 105–122, 1997.

[54] P. J. van den Oetelaar and H. J. Hoenders, "Electrophoretic titration curve in 6 M urea of the bovine eye lens protein α-crystallin.," *J. Chromatogr.*, vol. 411, pp. 507–509, Dec. 1987.

[55] J. L. Benesch, M. Ayoub, C. V. Robinson, and J. A. Aquilina, "Small heat shock protein activity is regulated by variable oligomeric substructure.," *J. Biol. Chem.*, vol. 283, pp. 28513–28517, Oct. 2008.

[56] M. T. Walsh, A. C. Sen, and B. Chakrabarti, "Micellar subunit assembly in a three-layer model of oligomeric α-crystallin.," *J. Biol. Chem.*, vol. 266, pp. 20079–20084, Oct. 1991.

[57] A. J. Baldwin, G. R. Hilton, H. Lioe, C. Bagnéris, J. L. P. Benesch, and L. E. Kay, "Quaternary Dynamics of αB-Crystallin as a Direct Consequence of Localised Tertiary Fluctuations in the C-Terminus," *J. Mol. Biol.*, vol. 413, pp. 310–320, Oct. 2011.

[58] S. Jehle, P. Rajagopal, B. Bardiaux, S. Markovic, R. Kühne, J. R. Stout, V. A. Higman, R. E. Klevit, B.-J. J. van Rossum, and H. Oschkinat, "Solid-state NMR and SAXS studies provide a structural basis for the activation of αB-crystallin oligomers.," *Nat. Struct. Mol. Biol.*, vol. 17, pp. 1037–1042, Sept. 2010.

[59] S. Jehle, B. van Rossum, J. R. Stout, S. M. Noguchi, K. Falber, K. Rehbein, H. Oschkinat, R. E. Klevit, and P. Rajagopal, "αB-crystallin: a hybrid solid-state/solution-state NMR investigation reveals structural aspects of the heterogeneous oligomer.," *J. Mol. Biol.*, vol. 385, pp. 1481–1497, Feb. 2009.

[60] J. Loutas, A. Stevens, G. J. Howlett, and R. C. Augusteyn, "The effect of concentration on the structure of α-crystallin.," *Exp. Eye Res.*, vol. 62, pp. 613–620, June 1996.

[61] J. Peschek, N. Braun, T. M. Franzmann, Y. Georgalis, M. Haslbeck, S. Weinkauf, and J. Buchner, "The eye lens chaperone α-crystallin forms defined globular assemblies.," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 106, pp. 13272–13277, Aug. 2009.

[62] J. Horwitz, "α crystallin: the quest for a homogeneous quaternary structure.," *Exp. Eye Res.*, vol. 88, pp. 190–194, Feb. 2009.

[63] S. Jehle, B. S. Vollmar, B. Bardiaux, K. K. Dove, P. Rajagopal, T. Gonen, H. Oschkinat, and R. E. Klevit, "N-terminal domain of αB-crystallin provides a conformational switch for multimerization and

structural heterogeneity.," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 108, pp. 6409–6414, Apr. 2011.

[64] A. Laganowsky, J. L. Benesch, M. Landau, L. Ding, M. R. Sawaya, D. Cascio, Q. Huang, C. V. Robinson, J. Horwitz, and D. Eisenberg, "Crystal structures of truncated αA and αB crystallins reveal structural mechanisms of polydispersity important for eye lens function.," *Protein Sci.*, vol. 19, pp. 1031–1043, May 2010.

[65] G. K. Hochberg, H. Ecroyd, C. Liu, D. Cox, D. Cascio, M. R. Sawaya, M. P. Collier, J. Stroud, J. A. Carver, A. J. Baldwin, C. V. Robinson, D. S. Eisenberg, J. L. Benesch, and A. Laganowsky, "The structured core domain of αB-crystallin can prevent amyloid fibrillation and associated toxicity.," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 111, pp. E1562–E1570, Apr. 2014.

[66] S. P. Delbecq, S. Jehle, and R. Klevit, "Binding determinants of the small heat shock protein, αB-crystallin: recognition of the IxI motif," *EMBO J.*, vol. 31, pp. 4587–4594, Dec. 2012.

[67] A. J. Baldwin, P. Walsh, D. F. Hansen, G. R. Hilton, J. L. Benesch, S. Sharpe, and L. E. Kay, "Probing dynamic conformations of the high-molecular-weight αB-crystallin heat shock protein ensemble by NMR spectroscopy.," *J. Am. Chem. Soc.*, vol. 134, pp. 15343–15350, Sept. 2012.

[68] J. A. Carver, J. A. Aquilina, R. J. Truscott, and G. B. Ralston, "Identification by 1H NMR spectroscopy of flexible C-terminal extensions in bovine lens α-crystallin.," *FEBS Lett.*, vol. 311, pp. 143–149, Oct. 1992.

[69] T. M. Treweek, A. Rekas, M. J. Walker, and J. A. Carver, "A quantitative NMR spectroscopic examination of the flexibility of the

C-terminal extensions of the molecular chaperones, αA- and αB-crystallin," *Exp. Eye Res.*, vol. 91, pp. 691–699, Nov. 2010.

[70] A. Laganowsky and D. Eisenberg, "Non-3D domain swapped crystal structure of truncated zebrafish αA crystallin.," *Protein Sci.*, vol. 19, pp. 1978–1984, Oct. 2010.

[71] A. J. Baldwin, G. R. Hilton, H. Lioe, C. Bagneris, J. L. Benesch, and L. E. Kay, "Quaternary dynamics of αB-crystallin as a direct consequence of localised tertiary fluctuations in the C-terminus," *J. Mol. Biol.*, vol. 413, pp. 310–20+, 2011.

[72] S. Y. Pasta, B. Raman, T. Ramakrishna, and M. Rao Ch, "The IXI/V motif in the C-terminal extension of α-crystallins: alternative interactions and oligomeric assemblies," *Mol. Vis.*, vol. 10, pp. 655–62+, 2004.

[73] I. K. Feil, M. Malfois, J. Hendle, H. van Der Zandt, and D. I. Svergun, "A novel quaternary structure of the dimeric α-crystallin domain with chaperone-like activity.," *J. Biol. Chem.*, vol. 276, pp. 12024–12029, Apr. 2001.

[74] H. A. Koteiche and H. S. Mchaourab, "Folding pattern of the -crystallin domain in A-crystallin determined by site-directed spin labeling," *J. Mol. Biol.*, vol. 294, pp. 561–577, Nov. 1999.

[75] C. Bagnéris, O. A. Bateman, C. E. Naylor, N. Cronin, W. C. Boelens, N. H. Keep, and C. Slingsby, "Crystal structures of α-crystallin domain dimers of αB-crystallin and Hsp20," *J. Mol. Biol.*, vol. 392, pp. 1242–1252, Oct. 2009.

[76] A. R. Clark, C. E. Naylor, C. Bagnéris, N. H. Keep, and C. Slingsby, "Crystal structure of R120G disease mutant of human αB-crystallin domain dimer shows closure of a groove.," *J. Mol. Biol.*, vol. 408, pp. 118–134, Apr. 2011.

[77] A. J. Baldwin and L. E. Kay, "Measurement of the signs of methyl ^{13}C chemical shift differences between interconverting ground and excited protein states by R(1ρ): an application to αB-crystallin.," *J. Biomol. NMR*, 2012.

[78] N. Braun, M. Zacharias, J. Peschek, A. Kastenmuller, J. Zou, M. Hanzlik, M. Haslbeck, J. Rappsilber, J. Buchner, and S. Weinkauf, "Multiple molecular architectures of the eye lens chaperone αB-crystallin elucidated by a triple hybrid approach," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 108, pp. 20491–6+, 2011.

[79] S. P. Delbecq and R. E. Klevit, "One size does not fit all: the oligomeric states of B crystallin.," *FEBS Lett.*, vol. 587, pp. 1073–1080, Apr. 2013.

[80] N. Braun, M. Zacharias, J. Peschek, A. Kastenmüller, J. Zou, M. Hanzlik, M. Haslbeck, J. Rappsilber, J. Buchner, and S. Weinkauf, "Multiple molecular architectures of the eye lens chaperone B-crystallin elucidated by a triple hybrid approach.," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 108, pp. 20491–20496, Dec. 2011.

[81] J. F. Hejtmancik, P. T. Wingfield, C. Chambers, P. Russell, H. C. Chen, Y. V. Sergeev, and J. N. Hope, "Association properties of βB2- and βA3-crystallin: ability to form dimers.," *Protein Eng.*, vol. 10, pp. 1347–1352, Nov. 1997.

[82] K. Srivastava, R. Gupta, J. M. Chaves, and O. P. Srivastava, "Truncated human βB1-crystallin shows altered structural properties and interaction with human βA3-crystallin.," *Biochemistry*, vol. 48, pp. 7179–7189, Aug. 2009.

[83] R. C. Augusteyn, "Dissociation is not required for α-crystallin's chaperone function.," *Exp. Eye Res.*, vol. 79, pp. 781–784, Dec. 2004.

Chapter 5

Oligomeric incompatibility in two plant small heat-shock proteins

5.1 Abstract

Oligomerisation is ubiquitous in proteins across the domains of life. The large numbers of different protein complexes that co-exist in a cell require there to be considerable specificity in oligomeric interfaces to prevent the formation of non-productive complexes. This is particularly challenging between proteins that share a conserved fold from a gene duplication event. Using a combination of experimental and computational structural biology approaches we examine two functionally distinct, paralogous plant small heat-shock proteins (sHSPs) that originated from a gene duplication event more than 400 million years ago. We show that, despite having the same quaternary architecture and similar subunit exchange dynamics at equilibrium, they are unable to form hetero-oligomers. We find that the highly conserved α-crystallin domain is incompatible between the two proteins despite having a virtually identical structure. Using a set of chimeric proteins we demonstrate that non-interfacial loop residues drive this incompatibility through their effect on the structure of transiently populated monomers. We further show, as dimers are organised into higher-order oligomers, how the interfaces responsible for oligomeric selectivity are weak and rapidly broken in solution. Non-selective interfaces provide most of the thermo-

dynamic stability to the oligomers formed at equilibrium. Together our results demonstrate how the types of oligomers formed in a mixture is determined by stoichiometries and conformations that are only lowly populated.

5.2 Introduction

Most proteins of known atomic structure self-assemble into homo-oligomers [1, 2]. Such self-interactions range from simple dimerisation to assembly into very large oligomers, and play a key part in cellular function and regulation. Homo-oligomerization can serve as the basis of homotropic allostery (or cooperativity)[3], encode large physical structures in a genetically efficient way, increase protein stability [4], augment substrate affinity through multivalent binding, and amplify substrate encounter rates [5]. These varied advantages are a possible rationale for why self-interactions are disproportionately enriched in interactomic maps of the proteome [6], and highlight how the ubiquity of homomers in biology is important in many evolutionary processes.

Gene duplication is the most common form of proteome expansion [7], and it is profoundly affected by homo-oligomeric interactions. When a gene encoding an oligomeric protein undergoes duplication, both daughter genes will necessarily produce oligomeric proteins with the ability to co-assemble. Most newly evolved oligomeric proteins therefore inherit from a gene duplication event a set of quaternary interactions with themselves, their ancestor, and all of its ancestor's interaction partners. After duplication, one of the two daughter genes may be lost or evolve a novel function [8], or both daughter genes subdivide the original function amongst each other [9]. In either case, retaining the ability of the two paralogues to co-assemble may be advantageous if the sub-functions need to be present in the same protein complex [10]. Otherwise, the effects of the duplication event can be neutral, or even deleterious, if co-assembly interferes with either of the newly evolved (sub-) functions.

In eukaryotes, the ability to co-assemble is often conserved amongst close homologues, such that functional protein complexes tend to be comprised of interacting paralogues [11], [12]. The alternative, in which co-assembly is actively prevented, is less well understood, but substantial indirect evidence points towards its importance. The prevention of harmful non-specific interactions between proteins that are not part of a functional protein complex exerts significant selective pressure on protein interaction networks [13], providing an upper limit for both the number of protein-coding genes, and the number of proteins that can safely co-exist in a single compartment.

For two paralogues not to co-assemble, their oligomeric interfaces need to become incompatible. The evolution of this property is balanced against maintaining the ability to interact with important binding partners, a constraint that significantly slows the rate of evolution of residues in protein interfaces [14, 15]. Homo-oligomeric binding modes are usually highly conserved within protein families [16], meaning that paralogues rarely evolve entirely new interfaces to avoid co-assembly. If the ability to oligomerise is not lost altogether in one of the daughter genes after duplication [17], it is likely these factors combine to constrain the evolutionary trajectories that lead to oligomeric incompatibility between paralogous protein oligomers. Data on protein families where this process has occurred are, however, still scarce and experimental validation of this hypothesis has not yet been possible.

The small heat-shock proteins (sHSPs) are a family of molecular chaperones in which gene duplications have produced both non-obligatory co-assembly of oligomeric paralogues [18, 19, 20] and paralogues that have lost the ability to co-assemble [21]. In the main, sHSPs act to intercept destabilised target proteins before they form non-functional and potentially harmful insoluble aggregates, holding them in a refolding-competent state [22]. The sHSPs role under conditions of stress as an aggregation buffering system requires them to interact with a very broad range of targets [23]. Probably for this reason, organisms that are particularly prone to large temperature fluctuations, such as plants,

often express several paralogous sHSPs per compartment and tissue [24]. Co-localised sHSPs can form hetero-oligomers both during assembly and through the process of subunit exchange, by which sHSP oligomers continuously recycle their subunits [25, 26]. The resulting ensemble of hetero-oligomers carrying variable numbers of different sHSP subunits has been theorised to further widen the substrate profile of sHSPs by providing a highly variable and multivalent binding surface on sHSP oligomers [23].

On the other hand, some sHSPs have evolved to prevent co-assembly with their paralogues as part of their function. An ancient gene duplication event over 400 million years ago resulted in two functionally distinct classes of cytosolic sHSPs in land plants and algae [21, 27, 28]. Within a single species, each class typically has several members resulting from subsequent gene duplications, which are able to form hetero-oligomers with members of only their own class, but not with members of the other. Members of the same class can however hetero-oligomerise even across species that are not closely related [21, 25, 26]. This indicates that plant sHSPs have to satisfy two highly evolutionarily conserved demands: the ability to hetero-oligomerise with paralogues of their own class, whilst avoiding hetero-oligomerisation with paralogues of the other. These two opposing constraints make plant sHSPs an excellent system to understand how evolution has encoded, structurally and dynamically, the selection for particular interactions amongst paralogues.

Here we investigate the assembly of two sHSPs from garden pea, HSP18.1 and HSP17.7 from cytosolic classes I and II, respectively, that have been shown not to co-assemble despite being expressed in the same compartment [21, 27]. We apply a combination of X-ray crystallography, native ion-mobility mass spectrometry (IM-MS), small angle X-ray scattering (SAXS), and molecular dynamics (MD) simulations applied to both wild-type sHSPs and a number of chimeric constructs to demonstrate how these two proteins sort their subunits into segregated oligomers without fundamentally remodelling their

oligomeric interfaces. We uncover how residues that are not part of an interface can drive association specificity by changing the conformation of dissociated monomers. In addition, we demonstrate how in a hierarchical assembly pathway of oligomerization, labile interactions can act efficiently to sort subunits into different oligomeric states even though they contribute minimally to the thermodynamic stability of the resulting assemblies. Our results represent a radically novel framework for understanding protein oligomers, shifting the focus away from stable contacts between subunits and towards the properties of transiently populated stoichiometries and labile interfaces that can dictate which oligomers are populated at equilibrium.

5.3 Results

5.3.1 HSP 18.1 and HSP 17.7 do not hetero-oligomerise

To understand how oligomeric paralogues can become structurally incompatible, we have investigated HSP 18.1 and HSP 17.7, two cytosolic plant sHSPs (class I and class II, respectively) from *Pisum sativum* (garden pea). These sHSPs originated from a gene duplication event and have subsequently lost the ability to co-assemble, despite 33.5 % sequence identity and 56.5 % sequence similarity [21]. Figure 5.1A shows mass spectra of both proteins acquired under conditions that preserve non-covalent interactions in the gas phase [29]. Both spectra reveal single charge series centred at 6300 m/z (HSP 18.1, top) and 6700 m/z (HSP 17.7, middle) corresponding to monodisperse dodecamers [21]. When HSP 17.7 and HSP 18.1 are mixed, no stable hetero-oligomers containing subunits from both HSP 17.7 and HSP 18.1 form, even when incubated for 16 hours at 40 °C (Fig. 5.1A, bottom). When both proteins were mixed and incubated in 30 % isopropanol, both dissociated into largely dimers, the protomeric building block of all sHSP oligomers (Fig. 5.1B, top). Upon dilution of the organic solvent, both proteins immediately reassembled into 12mers, but did so without forming any hetero-oligomers (Fig. 5.1B, bottom). HSP 18.1 and

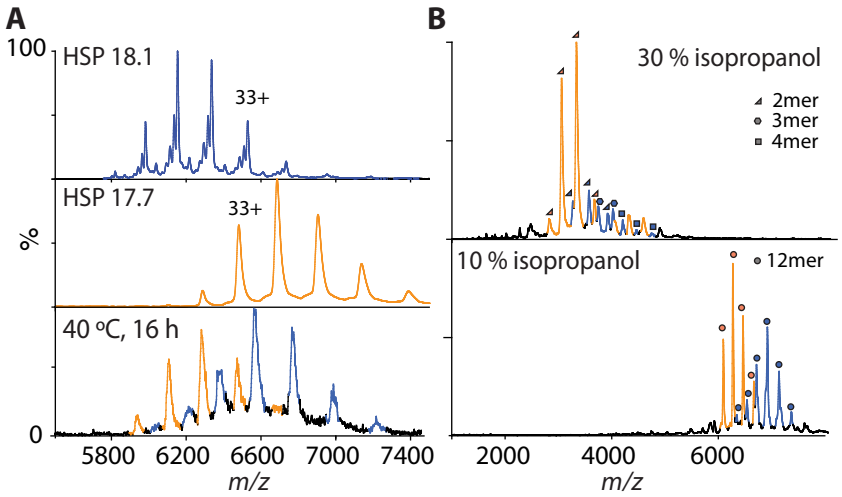


Figure 5.1: A) Native mass spectra of HSP 18.1 (top) and HSP 17.7 (middle) showing single charge-state distributions for both proteins corresponding to intact 12mers in both cases. The number indicates charge state of the annotated peak. When both proteins are mixed (bottom) and the incubated for 16 h at 40 °C no new peaks appear indicating that no hetero-oligomers form. B) Top: Solution disruption of HSP 17.7 and HSP 18.1 oligomers in 30 % isopropanol. HSP 17.7 disassembles into dimers, whereas HSP 18.1 disassembles into mainly dimers as well as slightly larger species. Bottom: If the organic solvent is diluted down both proteins reassemble into 12mers within minutes in the presence of one another without forming hetero-oligomers.

HSP 17.7 therefore neither co-assemble, nor form stable hetero-oligomers via subunit exchange once assembled.

5.3.2 HSP 18.1 and HSP 17.7 have tetrahedral oligomeric architecture

The structural incompatibility between HSP 18.1 and HSP 17.7 could arise from different quaternary structures. Two different architectures are known for dodecameric sHSPs: double rings consisting of two stacked rings of hexamers [30] and tetrahedra [31]. Circles are a third possibility, and although they have not been observed directly in sHSPs, the necessary connectivity linking each subunit to two others has been observed in a crystal structure of a truncated sHSP [32], and proposed as a substructure of higher-order oligomers [33]. In order to establish the most likely architecture for HSP 18.1 and HSP 17.7, we

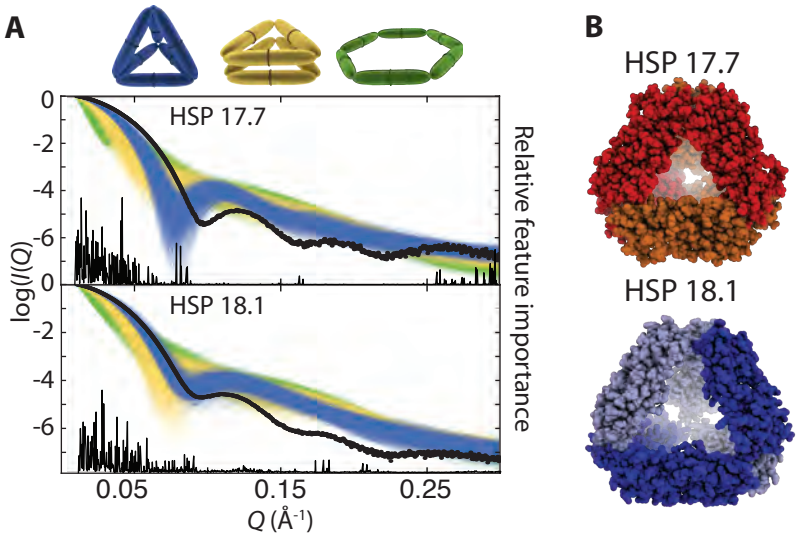


Figure 5.2: A) Experimental SAXS curves (black) of HSP 17.7 (top) and HSP 18.1 (bottom) overlaid on SAXS curves calculated for theoretical models of tetrahedral (blue), double ring (yellow) and circle (green) oligomers. Black lines correspond to the relative importance of features along the theoretical curves used by a learning algorithm trained to distinguish between candidate architectures. Both proteins are classified as tetrahedral by this algorithm. B) Representative tetrahedral models for HSP 18.1 and HSP 17.7. The classification algorithm used here was designed and implemented by Matteo Degiacomi.

constructed 6000 random geometrically feasible models, 1000 for each topology and each protein, and calculated theoretical SAXS curves for all. We used this dataset to train a learning algorithm to distinguish between the different architectures based on their SAXS signals. On a test set the algorithm achieved a classification accuracy of 99.3 % for models of HSP 17.7 and 95.0 % for HSP 18.1. We next determined experimental SAXS profiles for both HSP 17.7 and HSP 18.1 and challenged our learning algorithm with these data (Fig. 5.2A,B). Both proteins are classified as tetrahedral, and theoretical SAXS curves calculated from tetrahedral models show on average the highest cross-correlation coefficients to the experimental data for both proteins (0.95 $\pm$ 0.02 for HSP 17.7 compared to 0.92 $\pm$ 0.01 (double rings) and 0.93$\pm$0.02 (circles), and 0.98 $\pm$ 0.01 for HSP 18.1, compared to 0.91 $\pm$ 0.01 (double rings) and 0.96 $\pm$ 0.02 (circles)).

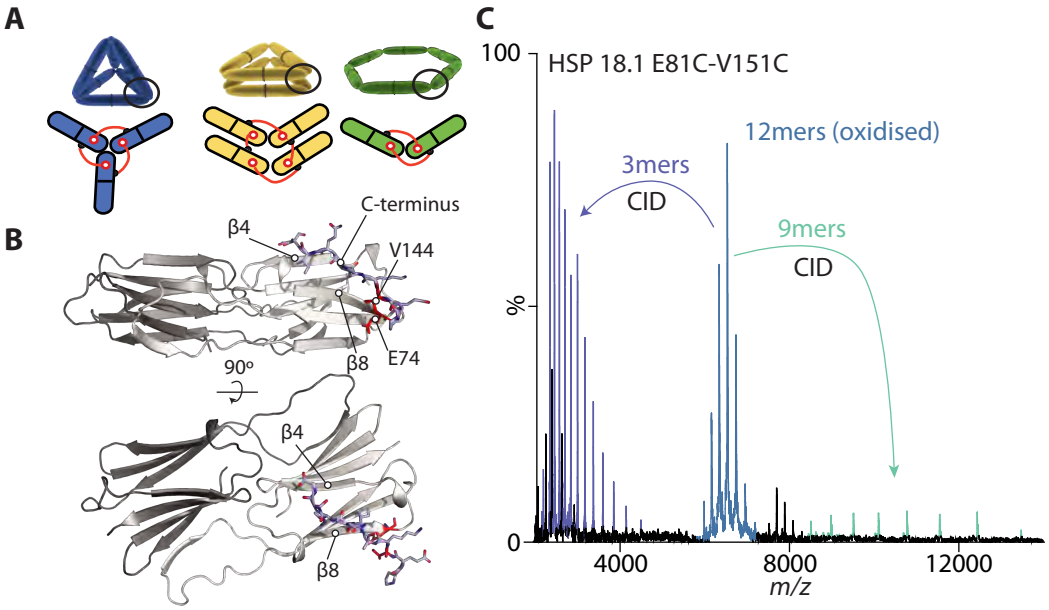


Figure 5.3: A) Different connectivities at vertices (red circles) between monomers in the three candidate architectures. In sHSP oligomers each monomer binds with its C-terminus (red line) to an adjacent monomer and also receives a C-terminus from another monomer. This leads to a connectivity of three in tetrahedra, four in double rings, and two in circles. B) In the homologous class I HSP 16.9 from wheat (PDB: 1GME) the C-terminus (blue) from an adjacent monomer (not shown) binds into the β4-β8 groove (grey). Residues E74 and V144 (red, corresponding to E81 and V151 in HSP 18.1) are suitably close to allow two cysteines to form a disulfide bond when engineered into those positions. C) The mutant E81C-V151C in HSP 18.1 can form a covalent disulphide bond between the C-terminus of one monomer and its binding site on another. It still forms a monodisperse dodecamer, but when the oxidised protein is subjected to gas-phase dissociation, unfolded but covalently linked trimers are expelled from dodecamers (purple charge series at low m/z), leaving behind charge-stripped nonamers (green charge series at high m/z), consistent only with a tetrahedral architecture.

To account for the lower classification accuracy of our algorithm for HSP 18.1, we sought orthogonal validation of its tetrahedral geometry experimentally. We made use of the different connectivity between subunits in circles, double rings and tetrahedra. In sHSPs, each monomer is connected to an adjacent monomer in a neighbouring dimer via its C-terminal arm, while it also receives a C-terminus itself. How many monomers are linked up in this way differs between architectures: in tetrahedra it is three, compared to four in double rings, and two in circles (Fig. 5.3A). We made use of this distinction

by engineering two cysteines based on homology to HSP 16.9 [30] into HSP 18.1, at positions V151 in the C-terminal arm, and E81 near its binding site (Fig. 5.3B). This mutant protein (HSP 18.1cys) also forms a monodisperse dodecamer. When we measured its gas-phase collision cross-section (CCS) using ion mobility MS, we found it to have the same CCS as the wild-type protein, under both oxidizing and reducing conditions (HSP 18.1 wt: 8754 ± 47 Å^2, oxidized HSP 18.1cys: 8743 ± 39 Å^2, reduced HSP 18.1cys: 8763 ± 32 Å^2), indicating that the mutant retains the architecture of the wild-type protein. When HSP 18.1cys is subjected to collision-induced dissociation (CID) in the gas-phase, it dissociates only into highly charged, covalently linked trimers and charge-stripped nonamers (Fig. 5.3C). This observation can only be explained by the exclusive population of a tetrahedral architecture. We conclude, therefore, that both HSP 18.1 and HSP 17.7 assemble into tetrahedral dodecamers. As a result, their inability to hetero-oligomerise is not due to gross architectural incompatibilities, but must instead arise directly at their oligomeric interfaces.

5.3.3 HSP 18.1 and HSP 17.7 display similar quaternary dynamics

To determine the level at which the two proteins are incompatible, we performed subunit exchange experiments to measure the stoichiometries and rates at which sub-complexes dissociate transiently from native oligomers at equilibrium. We incubated a 3:1 mixture of ^{12}C- and ^{13}C-labelled (light and heavy, respectively) HSP 17.7 dodecamers at 40 °C, and monitored the evolving mixture at different time-points by means of MS. Within 5 minutes, peaks appear corresponding to dodecamers containing both heavy and light subunits, in a sequential series of incorporations of two subunits at a time (Fig. 5.4A). At longer time-points, additional peaks are observed for every charge state, each corresponding to dodecamers carrying different numbers, both odd and even of heavy and light subunits, with their final relative intensities following a binomial distribution determined by the overall ratio of ^{12}C- to ^{13}C-labelled HSP

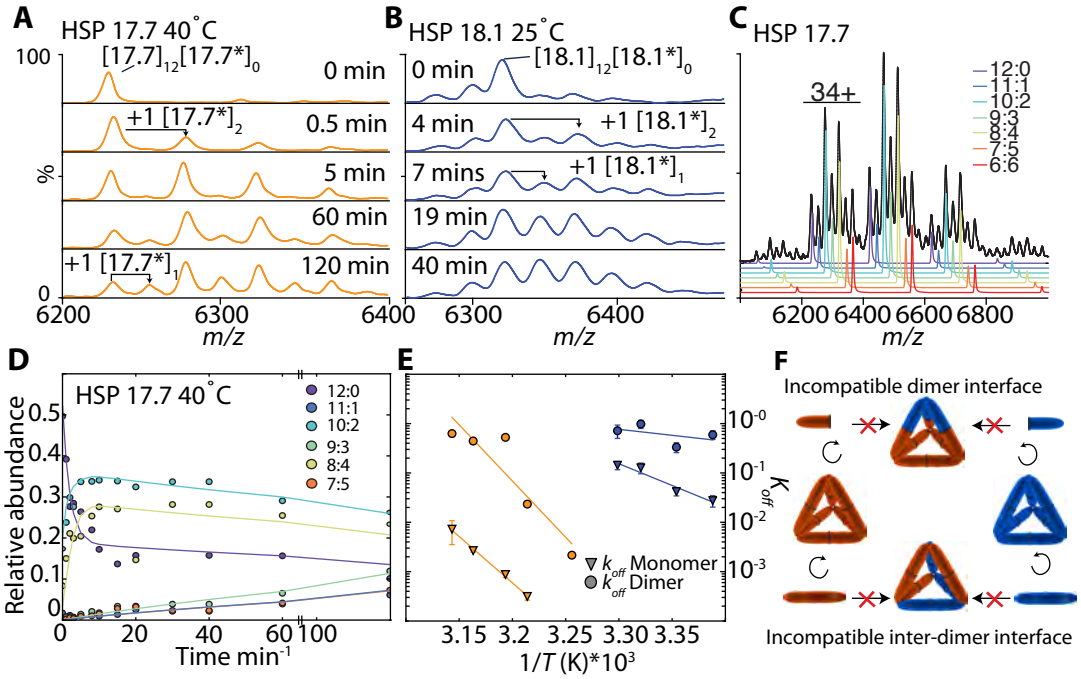


Figure 5.4: A) Time-resolved spectra of a subunit exchange reaction between ^{12}C and ^{13}C-labelled (17.7*) HSP 17.7 dodecamers mixed at a ≈ 3:1 ratio at 40 °C. Peaks shown correspond to the 34+ charge state B) Same experiment as in A) for 33+ charge-state of HSP 18.1 mixed with ^{13}C-labelled protein (18.1*) at a 5:1 ratio. C) Subunit exchange spectrum of HSP 17.7 (40 °C, 120 min) deconvolved to show the relative intensities of all exchanging species. Ratios indicate number of ^{12}C:^{13}C subunits in a dodecamer. For clarity only species up to a ratio of 6:6 are shown. D) Kinetic fit of the time course in A), assuming concurrent monomer and dimer exchange. Only fits for oligomers up to a 7:5 ^{12}C:^{13}C ratio are shown. E) Arrhenius plots for monomer (triangles) and dimer (circles) dissociation constants in HSP 17.7 (orange) and HSP 18.1 (blue). F) Formation of hetero-oligomers must be prevented at two levels: at the dimer interface to prevent hetero-oligomerisation via monomer exchange, and at the inter-dimer interface, to prevent hetero-oligomerisation via dimer exchange. The subunit exchange spectra of HSP 18.1 were collected by Dale Shepherd.

17.7 in the incubation mixture. When we performed an equivalent experiment with HSP 18.1, we observed a similar pattern of exchange behaviour, albeit at a significantly faster rate, even at 25 °C (Fig. 5.4B). This indicates that dodecamer exchange is mediated in both proteins by the transfer of monomers and dimers, the latter at a faster rate than the former.

In order to quantify the exchange processes, we performed subunit exchange experiments at a range of different temperatures and extracted the relative abundances of all species in the exchanging mixtures at each timepoint (Fig. 5.4C). To obtain rate constants we fitted the resulting intensities to a kinetic exchange model describing the subunit exchange of 12mers via competing monomer and dimer dissociation from 12mers (Fig. 5.4D). The dissociation rates in both proteins show marked temperature dependence, with subunit exchange being faster at higher temperatures (Fig. 5.4E), and both monomer and dimer dissociation from HSP 17.7 and HSP 18.1 display Arrhenius behaviour. Despite undergoing the same subunit exchange process, the activation energies and pre-exponential factors differ between the two proteins by over an order of magnitude for dimer dissociation (817 kJ mol^{-1} and 18 s^{-1} for HSP 17.7; 38 kJ mol^{-1} and 1 s^{-1} for HSP 18.1), and a small factor for monomer dissociation (640 kJ mol^{-1}, 9 s^{-1} for HSP 17.7; 148 kJ mol^{-1} and 3 s^{-1} for HSP 18.1). The larger pre-exponential factors in HSP 17.7 partially compensate for its much higher activation energy, such that at heat-shock temperatures (about 42 °C for pea) both proteins exchange dimers at relatively similar rates. Since at heat-shock temperatures both proteins therefore transiently populate monomers and dimers, both the inter- and intra-dimer interfaces must be incompatible between the two proteins for no hetero-oligomers to form (Fig. 5.4F).

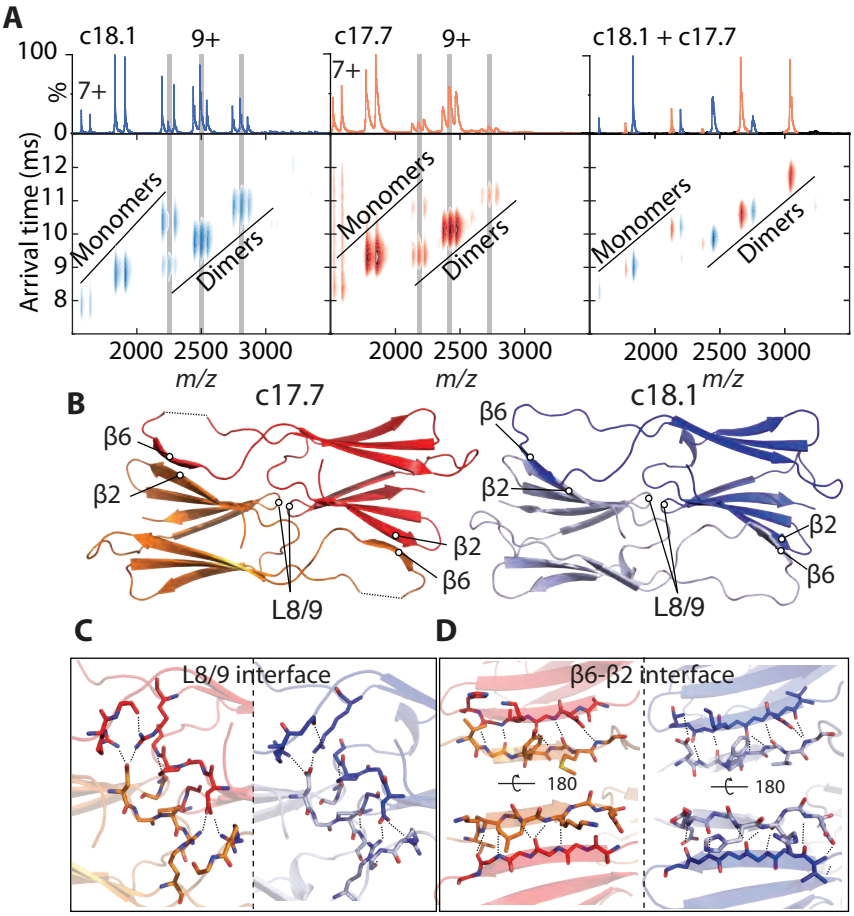


Figure 5.5: A) Subunit exchange dynamics of plant ACDs followed by IM-MS. Top: mass spectra of core domains mixed with ^{13}C labelled protein (left and middle) and a mixture of the two different core domains. Both core domains form monomers and dimers at low μM concentrations. Grey bars indicate peaks corresponding to dimers comprising a heavy and a light subunit as a result of subunit exchange. Exchange is extremely rapid for homo-dimers, but does not occur between class I and class II domains (right). B) Crystal structures of c17.7 (left) and c18.1 (right). The two domains have an identical β-sandwich fold in which two separate interfaces connect monomers into dimers. One interface centred on the β8-β9 loop (L8/9) between the β-sandwiches, and another formed by a strand exchange between the β6 and β2 strands. The crystal structure of c81.1 was solved by Dale Shepherd. C) The L8/9 interface, which consists largely of side chain interactions that are entirely conserved between the two domains. D) The β6-β2 interface consists largely of backbone-interactions with less conserved side-chains.

5.3.4 The dimer interface is incompatible, despite an extremely conserved structure

In sHSPs dimerisation is mediated by the α-crystallin domain (ACD), whereas oligomers are formed by inter-dimer contacts made by the C- and N terminal regions. To confirm that the dimer interface is incompatible between HSP 17.7 and HSP 18.1, we examined the ACDs of the two proteins (c17.7, corresponding to residues 48-140, c18.1, corresponding to residues 50-142). Both ACDs exist as mixtures of monomers and dimers at low μM concentrations, similar to our observations on mammalian ACDs [34] (Fig. 5.5A). When mixed with ^{13}C-labelled equivalents, both exchange monomeric subunits extremely rapidly, resulting in the formation of dimers incorporating one heavy and one light subunit within seconds. No hetero-dimers form when c17.7 and c18.1 are mixed (Fig. 5.5A, right), even when left to equilibrate at 40 °C for 16 hours. In our experiment, this corresponds to an increased stability of the homo-dimers of $>$9 kJ mol^{-1} compared to the hetero-dimer.

One possible cause for the absence of hetero-dimers could be fundamentally different dimer interfaces in the two proteins, as have been observed in the sHSP family [23]. To test this hypothesis, we solved the structures of c18.1 and c17.1 by X-ray crystallography, to 1.8 Å and 2.7 Å, respectively (table 5.1). Both dimers display an immunoglobulin fold comprised of 8 β-strands (numbered 2-9) very similar to other non-metazoan sHSPs (Fig. 5.5B). Strands 2-5 and 7-9 form a β-sandwich, whereas β6 is separated from the β-sandwich through two loops. In both structures, two separate interfaces combine to form the dimer. The first is formed by the β-sandwiches of two monomers centred on the β8-β9 loop (L8/9) mediated largely by intermolecular salt bridges that are fully conserved between HSP 17.7 and HSP 18.1 (Fig. 5.5C). The second interface is formed by a strand exchange between the β6 and β2 strands, and is mediated by backbone interactions, and some side-chain interactions that are not fully conserved (Fig. 5.5D). Upon close inspection neither interface

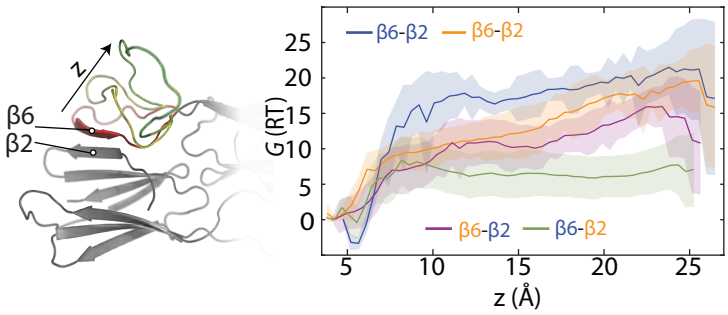


Figure 5.6: The free energy profile generated by gradually removing β6 from β2 along reaction coordinate z in MD simulations for the c17.7 homo-dimer (orange) and c18.1 homo-dimer (blue). In a simulated hetero-dimer this force is significantly reduced for the c18.1 β6/c17.7 β2 interaction (green), and only slightly lower than the homo-dimers for the c17.7 β6/c18.1 β2 interaction (purple). Shaded area corresponds to the standard error. The simulation and analysis in this figure were carried out by Erik Marklund.

provides an obvious source of incompatibility.

To test whether the lack of obvious differences between our structures is a result of having crystallized the molecules in frustrated conformations, we conducted 2-μs MD simulations of the dimers. The structures were remarkably stable during the simulations, with an average backbone RMSD over the final 200 ns of 1.7 Å (c17.7) and 1.4 Å (c18.1). We also constructed a hetero-dimer of c17.7 and c18.1 by structural alignment and simulated it for the same period of time. The hetero-dimer was similarly stable over the entire simulation (1.5 Å RMSD), confirming that there are no clear steric incompatibilities in the two structures. Furthermore, the number of inter-monomer hydrogen bonds formed on average during the simulations for the three dimers is essentially equivalent (c18.1: 20.2 ± 0.1; c17.7: 20.9 ± 0.1, hetero-dimer: 20.8 ± 0.1), and insufficient to explain the observed difference in stability between the homo- and hetero-dimers.

5.3.5 The dimer interface is incompatible, despite an extremely conserved structure

To test for more subtle structural incompatibilities, we performed steered MD simulations for each dimer for a total 6 μs in which we gradually detached the β6 strand from the β2 strand. From this, we calculated the free-energy profile in the form of a potential of mean force along this reaction coordinate using umbrella sampling (Fig. 5.6). In each case we observe a steep increase in the free energy along the first 7 Å of the reaction coordinate corresponding to the gradual breaking of β2-β6 interface. This is followed by a plateau in all four cases (two homo-dimers, and two hetero-dimers), indicating that in this region along the reaction coordinate the loop is relatively free to move and no interfaces are being broken. The height of this plateau differs between the simulations: while we find no significant difference in strength between the homomeric interactions in HSP 18.1 and HSP 17.7 (blue and orange curves), the heteromeric interactions are less stable. Specifically, we find that a HSP 17.7 β6 and a HSP 18.1 β2 strand combination is only very slightly easier to pull apart than the homomeric interactions, while removing a HSP 18.1 β6 strand from a HSP 17.7 β2 strand is significantly easier. Despite the similarity in the structures, and number of hydrogen bonds formed, probing them with umbrella sampling reveals that a hetero-dimeric interaction is easier to break apart, suggesting it to be less favourable.

5.3.6 The formation of hetero-dimers depends on non-interfacial loops in the core domains

To understand what makes a hetero-dimer easier to dissociate, we designed a set of chimeric core domains and tested their ability to dimerise with one another by means of MS. We divided the core domain sequence into three sections based on our atomic structures: the β-sandwich, which includes the L8/9 interface and the β2 part of the β2-β6 interface (residues 48-87, 109-140 in HSP 18.1, 50-89; 101-142 in HSP 17.7); the β6 strand (residues 100-105 in HSP 18.1,

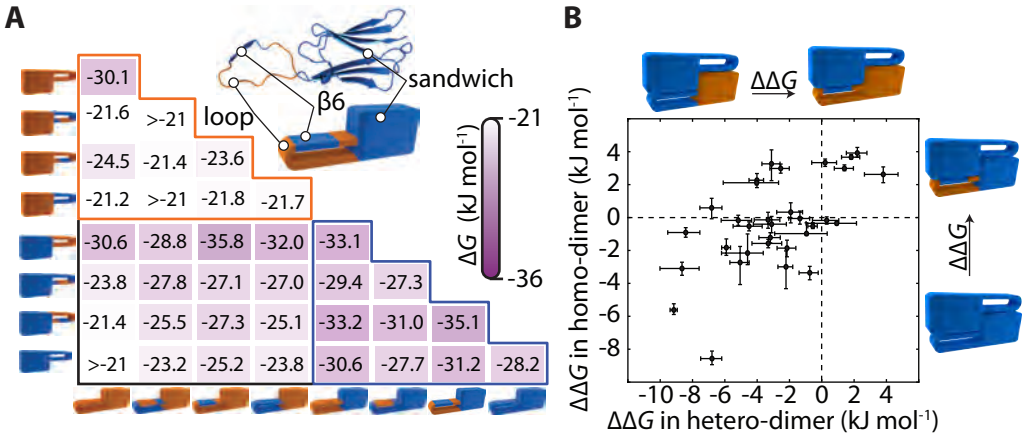


Figure 5.7: A) Free energy of dimer formation for a set of chimeric core domains that have either their β6 strand or the non-bonding portion of the loop swapped between c17.7 and c18.1. Free energies below -21 kJ mol^{-1} are coloured as white. All chimeras with a β-sandwich from c18.1 dimerise readily (lower right triangle), wheras all chimeras that have β-sandwich from c17.7 (upper left triangle) have substantially reduced free energies of dimer formation, even when just the non-bonding parts of the loop are substituted. Hetero-dimers between chimeras with different β-sandwiches are readily formed. Spectra used for this figure were collected by Dale Shepherd. B) Correlation plot for $\Delta\Delta G$ values between swaps that result in the same β6-β2 contacts in dimers comprised of two different β-sandwiches (x-axis) and dimers that only have one type of β-sandwich (y-axis). Error bars are the summed standard deviation of the individual ΔG values. The poor correlation suggests that the permutations change the $\Delta\Delta G$ also through their effect on dissociated monomers.

102-107 in HSP 17.7), and the non-bonding parts of the loop positioning β6 on β2 (residues 88-99, 105-108 in HSP 18.1, 90-101, 107-110 in HSP 17.7) (Fig. 5.7A). We next performed titration IM-MS titration experiments to measure the ability to form dimers for all 36 combinations of wild-type and chimeric domains. For each titration we manually collected 12 IM-MS spectra at varying concentrations of one core domain whilst keeping the concentration of the other constant, resulting in a total of 432 spectra to measure all interactions.

From analysis of IM-MS spectra, all constructs that have a β-sandwich derived from HSP 18.1 (Fig. 5.7A, lower right triangle) dimerise with one another, with favourable free energies of dimerization ($\Delta G_{Dimer} = G_{Dimer} - 2G_{Monomer}$) in the range of -27 to -36 kJ mol^{-1} (corresponding to K_D values in the low μM range). Remarkably, this is regardless of whether one or both

monomers have a swapped β6 strand or swapped non-bonding parts of the loop. In fact, exchanging only the non-bonding parts of the loop in both monomers to HSP 17.7 sequence increases the stability on average by ≈7 kJ mol^{-1}. There is a small free energy penalty for swapping the β6 strand from the HSP 18.1 sequence to that in HSP 17.7, associated with a loss of about 0.5-1 kJ mol^{-1} per monomer. The HSP 18.1 β-sandwich can therefore form productive contacts with β6 strands from both classes, and can dimerise regardless from which class its loop is derived.

Dimers between constructs that have β-sandwiches derived from HSP 17.7 (Fig. 5.7A, upper left triangle) do not tolerate changing their corresponding loop residues in the same manner. Whereas the wild-type domain has a ΔG_{Dimer} of -30 kJ mol^{-1}, all dimers involving chimeras with HSP 18.1 loops have dramatically reduced stabilities. Notably, even when only the non-bonding parts of the loop are switched we do not observe any dimer (in our experiments, this corresponds to ΔG_{Dimer} > -21 kJ mol^{-1}, or K_D > 200 μM). Only when the β6 strand is swapped to the HSP 18.1 sequence does the resulting chimera still dimerise measurably, but with a substantially reduced stability (≈-24.5 kJ mol^{-1}) compared to the wild-type domain (of -30 kJ mol^{-1}). The non-bonding loop residues of c17.7 therefore have a stabilizing effect on both dimers relative to the non-bonding residues of c18.1, and both kinds of β-sandwich make more favourable contacts with β6 strands from their own class. These results are consistent with our MD simulations, in which a HSP 18.1 β6 strand and loop were much easier to remove from the HSP 17.7 β-sandwich than in the wild-type dimer. In order for both wild-type domains to have similar ΔG_{Dimer} values, the effect of the non-bonding parts of the loop (more stable in c17.7, less stable in c18.1) must be balanced by opposing contributions from the β-sandwich, suggesting that the two ACDs partition their thermodynamic stability differently between their respective parts.

5.3.7 Mutations in the loop affect hetero-dimerisation via the stability of monomers

To understand how these stability differences might prevent hetero-dimerization we obtained ΔG_{Dimer} values for all possible dimers with two different β-sandwiches. (Fig. 5.7A, lower left rectangle). Remarkably, most combinations produced hetero-dimers with ΔG_{Dimer} values between -23 and -36 kJ mol^{-1}. Swapping the entire loop in wild-type HSP 18.1 from its own sequence to the HSP 17.7 residues produces stable hetero-dimers, regardless of the composition of the loop fused to the HSP 17.7 β-sandwich (Fig. 5.7A, top row of lower left rectangle). This is consistent with our earlier observation that the HSP 18.1 β-sandwich tolerates any type of loop, whereas the HSP 17.7 β-sandwich only dimerises readily when fused to its own β6 strand and non-bonding residues. The unfavourable interaction between wild-type HSP 18.1 and HSP 17.7 core domains is therefore to a large extent due to the inability of the HSP 17.7 β-sandwich to form a stable interaction with the β6 strand in an HSP 18.1 loop. However, relatively stable hetero-dimers can also be formed if the non-bonding parts of the loop in HSP 17.7 are switched to the HSP 18.1 sequence (Fig. 5.7A, columns 2 and 4 in rectangle). This is remarkable, because the same permutation always leads to a decrease in stability in dimers involving the same β-sandwich in both monomers.

To understand the origins of this effect, we compared the change in ΔG_{Dimer} for each swap ($\Delta\Delta G_{swap} = \Delta G_{Dimer^*} - \Delta G_{Dimer}$, where $Dimer^*$ contains a single swap relative to $Dimer$) in a dimer comprised of two different β-sandwiches to the equivalent swap in a dimer comprised of two identical β-sandwiches such that the β-sandwich facing the altered loop remains unchanged (Fig. 5.7B). Remarkably, we found that the $\Delta\Delta G_{swap}$ for dimers with two kinds of β-sandwiches correlates poorly with $\Delta\Delta G_{swap}$ for dimers with just one kind of β-sandwich ($R^2 = 0.35$). The discrepancies between these quantities cannot be the result of different contacts in the $\beta6/\beta2$ interface, since those are

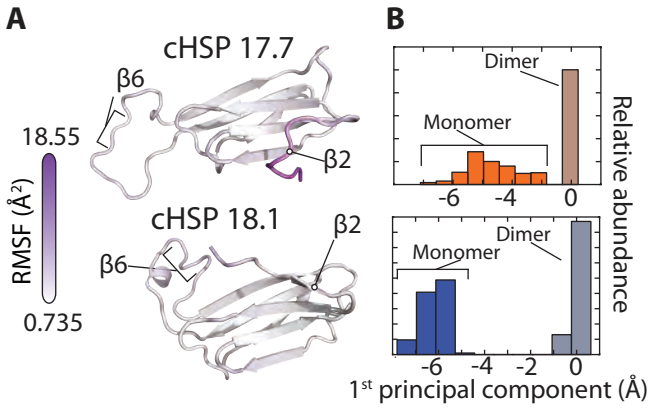


Figure 5.8: A) Cluster-average structures of the last μs of simulations of c17.7 and c18.1 monomers coloured according to the per-residue root-mean square fluctuation. In c17.7 the loop surrounding β6 stays in a conformation similar to its conformation in the dimer, whereas the β2 strand becomes highly mobile in the monomer. In 18.1 the loop forms two short helices and bends back over the β-sandwich, which is more rigid in c18.1 compared to c17.7 B) Principal component analysis showing that both simulated monomers are distinguishable from their conformations in dimers. In c17.7 the first principle component describes largely movements in β6 and some loop deformations, whereas for c18.1 the first principle component largely describes the rearrangement of the loop. The simulations used for the analysis in this figure were carried out by Erik Marklund.

held constant in these comparisons. Instead, the poor correlation suggests that the thermodynamic effect of the permutations is also strongly dependent on the exact sequence composition of the resulting chimeric monomer, regardless of what contacts it makes at the dimer interface. Specifically, the effect of the permutation differs depending on which β-sandwich the mutated loop is fused to. Given the extremely similar structure of the dimers, such behaviour most likely results from the permutations affecting the relative stability of dissociated monomers, and thereby affect the ΔG_{Dimer} independently of which contacts are made at the dimer interface. We therefore hypothesise that our loop permutations might alter the structure, and consequently the stability of dissociated monomers.

5.3.8 Dissociated monomers adopt distinct conformations

To determine the structure of transiently populated monomers, we carried out 2 μs MD simulations for both wild-type domains and all chimeras in their monomeric state. Interestingly, we observed that the two wild-type monomers adopt distinct conformations (Fig. 5.8A). In the simulation of monomeric c18.1, while its β-sandwich remains in a conformation similar to the dimer, the loop surrounding β6 folds back onto the β-sandwich and forms two short helices between residues 96-99 and what constitutes the β6 strand in the dimer (residues 102-105). In c17.7 the loop stays in a more extended state similar to its conformation in the dimer, whereas there is marked conformational flexibility in its β2 strand, which partially detaches from the β-sandwich during our simulation and shows a very high root-mean-square fluctuation compared to the rest of the molecule ($\approx$18 Å^2 compared to $\approx$1 Å^2 for residues on the opposite side of the β-sandwich). When we calculated the first principal component describing the transition from the dimeric to the monomeric conformation we found that the monomeric conformation of both molecules is consistently distinguishable from the dimeric conformation over the last μs of the simulation (Fig. 5.8B).

To ascertain whether these differences are likely to be fundamental properties of class I and class II core domains we collated 211 sequences of class I plant sHSPs and 89 of class II sHSPs and interrogated which residues are specifically conserved in only one class but not in the combined set of class I and class II sequences. When we mapped the results onto the monomer structures, we found that, in general, residues in the non-bonding parts of the loops are conserved specifically in each class (Fig. 5.9). In particular, the residues forming the two short helices in the c18.1 monomer are highly conserved in class I ACDs. In class II ACDs (corresponding to c17.7) several residues in the non-bonding part of the loop C-terminal to β6 are highly conserved, as well as residues in the highly dynamic β2 strand. The conformation and dynamics of

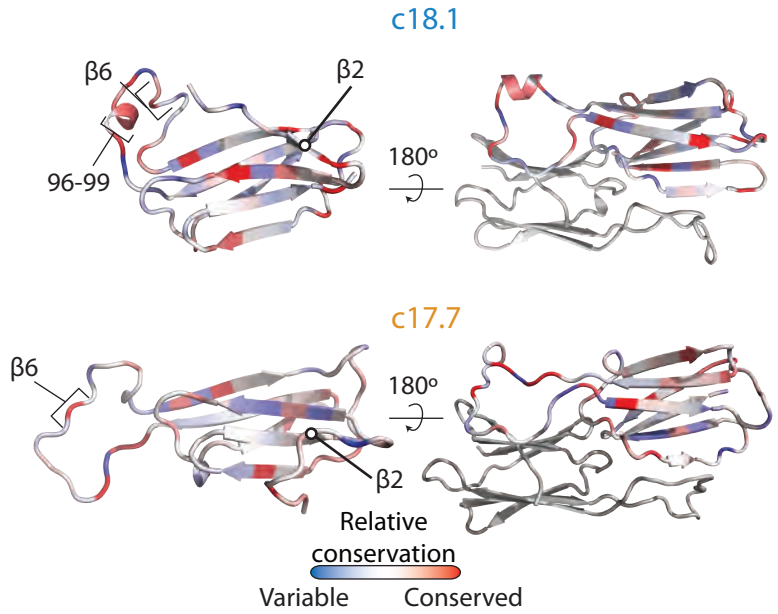


Figure 5.9: Sequence conservation specific to each class mapped onto the monomer (left) and dimer structures (right). Red indicates residues that are conserved in one class, but not the other. Blue indicates residues that are more variable in one class than in the other. Residues that are equally conserved in both classes are coloured white. All structures shown are cluster averages of their respective simulations. In c18.1 the residues forming the short helix in residues 96-96 are highly conserved. In c17.7 residues C-terminal to β6 in the non-bonding parts of the loop are highly conserved, as well as residues in the highly mobile β2 strand.

the monomeric ACD are therefore likely features that are consistently different between the two classes.

To understand how swapping parts of the loop between classes perturbs these conserved conformations, we calculated the first principal component describing the transition between the conformations populated by the two wild-type monomers, and projected trajectories of each chimera along the corresponding eigenvector. Using the first principal component as a proxy for structural similarity, we compared how similar each chimera is to either wild-type monomer (Fig. 5.10). In general, monomers with swapped loops become more similar to the monomer the swap corresponds to. Mutating the non-bonding parts of the loop makes the largest difference, and when the whole loop is swapped the resulting chimera is more similar to the wild-type monomer cor-

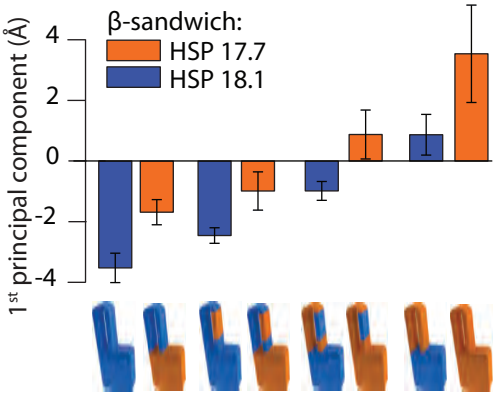


Figure 5.10: Principal component analysis of MD simulations of chimeric monomers. Chimeric trajectories are projected along the eigenvector corresponding to the first principal component describing the difference between the wild-type monomers. Mutation of β6 and/or the non-bonding parts of the loop in c18.1 to the c17.7 sequence (blue bars) makes the resulting chimera more similar to wild-type c17.7 and vice versa. Constructs that have the same loop but different β-sandwiches have distinguishable conformations due to the different behaviour of their β-sandwiches.

responding to its loop, than to the wild-type monomer corresponding to its β-sandwich. When we compared monomers that share the same loop but have different β-sandwiches, we found that they are always significantly different. The difference in their first principal component is in all cases $\approx$2 Å, indicating that any two loop-matched chimeras differ consistently in the behaviour of their β-sandwiches. Overall, dissociated monomers from HSP 17.7 and HSP 18.1 therefore have a significantly altered structure, and mutating just loop residues interconverts these partially. The remaining difference is due to the different behaviour of the β-sandwich, which is more dynamic in the HSP 17.7 monomer, and helps explain the poor correlation in figure 5.7B. Although the swapped loops of the two monomers in each comparison make the same contacts in the dimer interface, the two monomers behave differently in their dissociated states. When the balance between the stability conferred to monomers by the loop and β-sandwich present in the wild-type molecule is upset by mutation, this can lead to formation of hetero-dimers independently of any direct effects on the dimer interface. Overall hetero-dimerisation is therefore prevented by

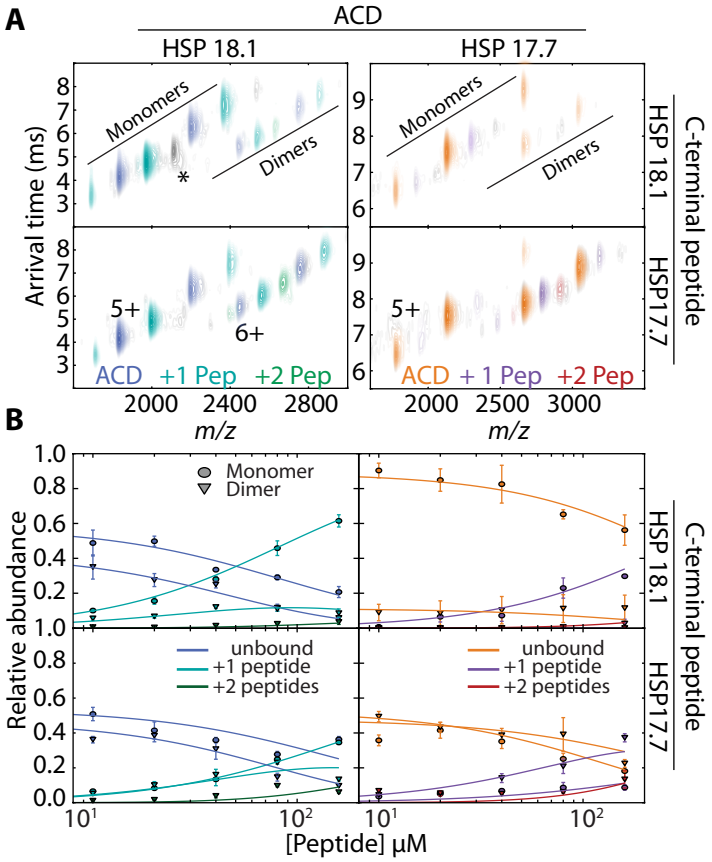


Figure 5.11: A) Binding of C-terminal peptides to core domains of HSP 18.1 and HSP 17.7 measured by IM-MS to resolve overlap between monomers and dimers. c18.1 (left column) can bind C-terminal peptides from both proteins with similar affinity. cHSP 17.7 binds the HSP 18.1 peptide only weakly (top right), and binds its own peptide only as a dimer with an affinity similar to the 18.1 ACD-18.1 peptide interaction. Spectra shown are 10 μM core domain and 80 μM peptide. Starred peak denotes an impurity in the peptide stocks. B) Titration curves for C-terminal peptides binding to ACDs corresponding to the spectra in A). For binding constants, see text. Error bars correspond to the standard deviation of three repeats.

an asymmetric partitioning of stability into different parts of the ACD, which manifests itself clearly in the different structures of dissociated monomers.

5.3.9 The C-terminus provides specificity in the first of two higher-order assembly steps

While these experiments on the isolated ACDs provide a rationale for the absence of hetero-dimerisation, they do not explain the absence of hetero-dodecamers comprised of homo-dimers. We next determined which inter-

faces prevent the co-assembly of HSP 17.7 and HSP 18.1 dimers into hetero-oligomers. We consider two possible oligomeric interfaces in turn: the C-terminal region that contains an IXI motif, highly conserved within sHSPs, and the heterogeneous N-terminus. In sHSPs studied to date the C-terminal region makes oligomeric contacts by binding into the hydrophobic β4-β8 groove in the ACD of an adjacent dimer.

To test whether C-termini from both proteins could bind to each other's ACD, we performed titration experiments using the ACDs from HSP 17.7 and HSP 18.1 and peptides mimicking their respective C-terminal regions. We quantified binding of peptides to monomers and dimers by IM-MS to obtain binding constants for binding of a single peptide to monomers, and binding of up to two peptides to dimers (Fig. 5.11). Figure 5.11 shows representative spectra collected at 8:1 molar excess of peptide for each of the 4 combinations. For the ACD from HSP 18.1 mixed with its own peptide we observe peaks for binding of peptides to monomers as well as dimers (Fig. 5.11A, top left). Single peptides bind to monomers and dimers with similar affinity (Fig. 5.11B, top left, K_D of 47 μM and 77 μM for monomers and dimers, respectively). The second peptide binds to dimers with considerably lower affinity (578 μM). c17.7 binds its own peptide only weakly as a monomer (K_D of 407μM) (Fig. 5.11A,B, bottom right), but binds as a dimer the first and second peptide with affinities comparable to c18.1 (101 μM and 407μM for first and second binding event). When c18.1 is incubated with the C-terminus from HSP 17.7 (Fig. 5.11A,B, bottom left) this peptide binds c18.1 monomers with a slightly lower affinity than its own peptide (111 μM), whereas dimers bind the first and second peptide with affinities roughly equal to its own peptide (84 μM and 345 μM for the first and second binding event, respectively). c18.1 can therefore bind both peptides with similar affinity. When we incubated c17.7 with peptide from HSP18.1 (Fig. 5.11A,B, top left), however, we only observed peaks corresponding to peptides binding to monomers (K_D of 267 μM) and essentially no binding to dimers, corresponding to K_D values in the mM range.

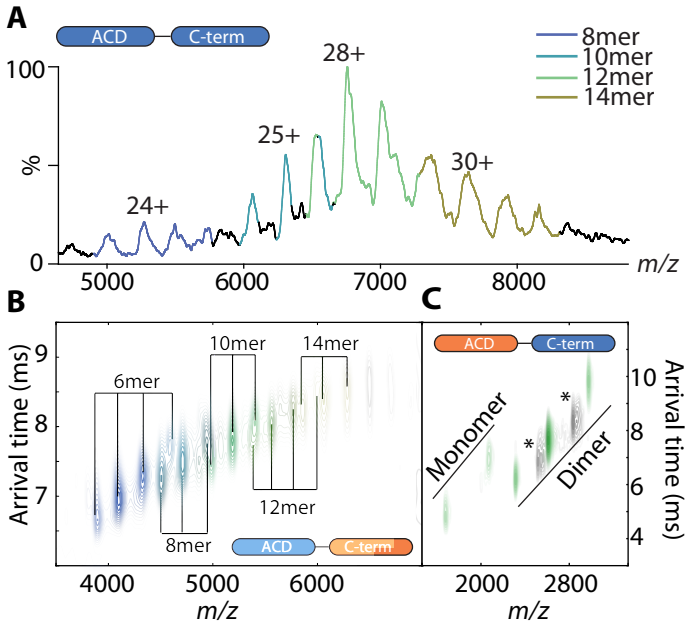


Figure 5.12: A) A construct consisting of the HSP 18.1 ACD fused to its own C-terminus populates stoichiometries ranging from 8mers to 14mers in steps of dimers. B) Oligomeric distribution formed by an ACD from HSP 18.1 fused to the C-terminus from HSP 17.7 as resolved by IM-MS. Even stoichiometries ranging from 6mers to at least 14mers are resolvable. C) The opposite chimera to the example in B) does not form higher order oligomers. Starred peaks denote an impurity.

Whereas the ACD of c18.1 is promiscuous in its interaction with C-termini, c17.7 therefore shows selectivity towards its own C-terminus.

To determine whether this unilateral selectivity is sufficient to prevent co-assembly we analysed constructs that consist of an ACD and a complete C-terminal region, but lack an N-terminus. An ACD from HSP 18.1 fused to its own C-terminus readily oligomerises at μM concentrations, and forms a polydisperse ensemble of even-numbered oligomers comprising 8-14 subunits (Fig. 5.12A). An ion mobility mass spectrum of the HSP 18.1 ACD fused to the C-terminus from HSP 17.7 reveals even-numbered stoichiometries ranging from 6 to at least 14 subunits, demonstrating that the 18.1 ACD can form similarly productive contacts with C-termini from both proteins, consistent with our peptide binding data (Fig. 5.12B). The opposite is true for HSP 17.7: the HSP 17.7 core domain fused to a HSP 18.1 C-terminus forms only

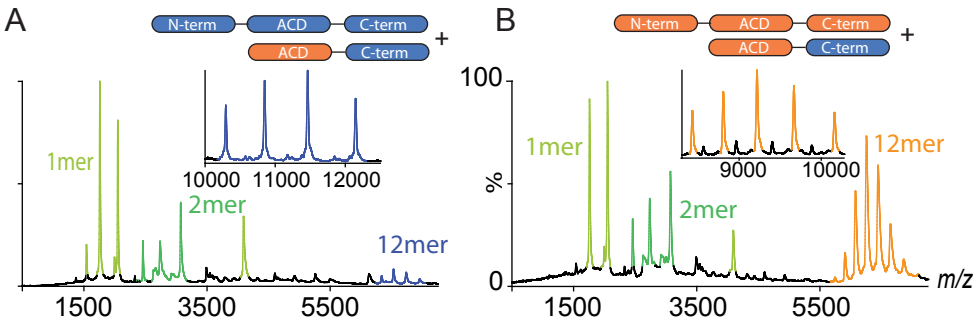


Figure 5.13: A) Mixture of full-length HSP 18.1 (blue peaks) and an assembly incompetent chimera consisting of an ACD from HSP 17.7 and a C-terminus from HSP 18.1. Starred peak indicates a CID product of the chimeric dimers. No hetero-oligomers form when both proteins are mixed and incubated for 40 minutes at 40 °C. Inset: stripped oligomers produced by CID of peaks corresponding to HSP 18.1 reveal no incorporation of chimeric subunits. B) Same experiment as in A), but using HSP 17.7 (orange peaks) instead of HSP 18.1. Similarly, no uptake of chimeric subunits into the 12mers is observed. Inset: stripped oligomers produced by CID of peaks corresponding to HSP 18.1 reveal no incorporation of chimeric subunits

monomers and dimers (Fig. 5.12C).

When this construct is mixed with full-length oligomers of HSP 18.1 (Fig. 5.13A) and incubated at 40 °C for 40 minutes we do not observe peaks corresponding to hetero-oligomers formed between the chimeric and the full-length protein. Since the chimera C-terminus is compatible with the HSP 18.1 ACD, the inability of the chimera to receive the C-terminus from HSP 18.1 prevents the formation of hetero-oligomers. Similarly, if the same chimera is mixed with HSP 17.7 full-length oligomers and incubated at 40 °C for 40 minutes no hetero-oligomers form (Fig. 5.13B), demonstrating that in this case the inability to donate its C-terminus to HSP 17.7 subunits prevents hetero-oligomerization. These results demonstrate that for hetero-oligomerisation to occur, each subunit has to be able to both receive and donate C-termini, meaning that the incompatibility of the HSP 17.7 ACD and the HSP 18.1 C-terminus is entirely sufficient to prevent hetero-oligomerization via the exchange of dimers.

5.3.10 The N-terminus acts as a second, nonspecific step in assembly

To understand the role of the N-terminus in assembly and subunit sorting, we engineered a construct comprising an HSP 18.1 N-terminus, HSP 17.7 ACD and HSP 18.1 C-terminus. Based on our previous results, this construct cannot form a productive interaction between its C-terminus and ACD. Figure 5.14A (bottom) shows that although the sample contains several impurities (black peaks), we could detect a charge series (green peaks) corresponding to a dimer formed by the chimera. We confirmed the stoichiometry of this chimera by selectively subjecting its 12+ charge state to CID (Fig. 5.14A, middle inset), resulting in the dissociation of the dimer into both highly charged monomers at low m/z and charge-stripped monomers at higher m/z. The monomers contain several N-terminal truncations (Fig. 5.14A, top inset, corresponding to the first 6, 7 and 8 residues missing), presumably because the N-terminus is fully accessible to cellular proteases in this non-assembling chimera. This construct nonetheless demonstrates that in the absence of a compatible C-terminus/ACD interaction, the N-terminus does not provide any assembly competence.

Based on these results we reasoned that the role of the N-terminus in assembly might be to bias an initially polydisperse ensemble of stoichiometries constructed through C-terminal interactions towards a monodisperse 12mer. To test this hypothesis we mixed HSP 18.1 full-length oligomers with a construct of HSP 18.1 that lacks an N-terminus but still forms a polydisperse ensemble via its C-terminus. When we mixed the two proteins we observed a broadening of peaks in the region where HSP 18.1 12mers are usually found (6000-7000 m/z, Fig. 5.14B), indicating incorporation of chimeric subunits into native 12mers. When this region of the spectrum is subjected to CID (Fig. 5.14B, inset), we observe charge-stripped oligomers resulting from the loss of a single subunit between 10000 and 13000 m/z. In this way we could resolve stripped 11mers resulting from 12mers that have exchanged up to 3 wild-type subunits

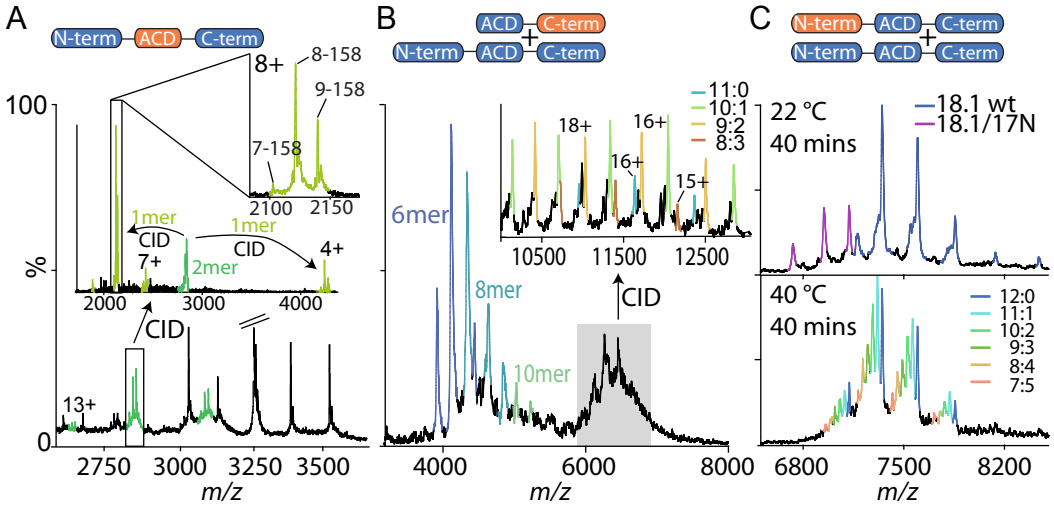


Figure 5.14: A) Mass spectrum of a full-length chimera comprised of an N terminus from HSP 18.1, an ACD from HSP 17.7 and a C-terminus from HSP 18.1. Green peaks correspond to dimers, showing that despite containing an N-terminus this construct is unable to oligomerise due to its incompatible C-terminus/ACD combination. Black peaks are the result of an impure sample. Middle inset: CID of the 12+ charge state of the dimer (boxed in bottom spectrum) resulting in dissociation of monomers. Top inset: 8+ charge state of the dissociated monomers, showing that monomers have suffered several N-terminal truncations. Peak labels correspond to residues remaining in the major truncated species. B) Uni-directional exchange between wild-type HSP 18.1 and its own ACD fused to the C-terminus of HSP 17.7 (see Fig. 5.12B). Coloured peaks correspond to chimeric homo-oligomers. The shaded signals correspond to peaks that have become unresolvable as a result of subunit exchange between the chimera and wild-type subunits. Inset: Stripped oligomers produced by subjecting the shaded peaks to CID. The resulting charge series correspond to stripped wild-type oligomers that have taken up at least 3 chimeric subunits. Ratios refer to number of wild type: chimeric subunits. C) Promiscuity of N-terminal contacts. Top: mixture of an N-terminal chimera of HSP 18.1 and HSP 18.1 wild type. The chimera forms a monodisperse 12mer at room temperature. Bottom: hetero-oligomers form readily after incubation at 40 °C. Ratios refer to number of wild type: chimeric subunits.

for chimeric subunits. In the native (non-CID) spectrum there are no additional peaks in the region corresponding to the smaller chimeric homo-oligomers (6mers to 10mers), indicating that no full-length subunits have exchanged into these oligomers. This one-directional exchange of truncated subunits onto full-length oligomers therefore confirms our hypothesis that subunits containing an N-terminus have a strong preference for residing in dodecamers. From the rela-

tive intensities of the 12mers in the truncated chimera and wild-type HSP 18.1 we can estimate that the relative stability of a single subunit within a 12mer compared to a dissociated monomer differs by 13 kJ mol^{-1} when all subunits in the dodecamer contain an N-terminus, compared to a dodecamer formed entirely by a subunits missing the N-terminus (≈-17 kJ mol^{-1} compared to ≈-4 kJ mol^{-1}). The N-terminus therefore acts as a thermodynamic bias converting an initially polydisperse ensemble constructed via C-terminal interactions into a monodisperse assembly, and dominates the thermodynamic stability of 12mers at equilibrium.

To understand if this process also acts to sort subunits, we replaced the N-terminus of HSP 18.1 with the sequence from HSP 17.7 and mixed this construct with wild-type HSP 18.1 oligomers at a 1:3 ratio. Figure 5.14C (top) shows that both proteins form monodisperse 12mers that do not exchange subunits appreciably over 25 minutes at room temperature. When we incubated the mixture for 40 minutes at 40°C and analysed it by native MS, we observed a distribution of peaks for each charge state, indicating that subunit exchange had occurred between the two proteins (Fig. 5.14C, bottom). The distribution of peaks corresponding to different numbers of chimeric and wild-type subunits per 12mer is centred on 11:1 (wild type: chimera), and closely corresponds to the binomial distribution expected for complete equilibration of a 3:1 mixture. This indicates that hetero-oligomers in this case are not significantly less favourable than homo-oligomers, demonstrating that the contacts made by the N-termini in HSP 17.7 and HSP 18.1 are not specific to each class. Subunit exchange between the N-terminal chimera and HSP 18.1 is slower than homomeric exchange in HSP 18.1 (which exchanges rapidly at 22 °C), indicating that breaking the N-terminal contacts must be the rate-limiting step in subunit exchange, at least in the case of the HSP 17.7 N-terminus. Even though N-terminal contacts therefore dominate the thermodynamic stability and subunit-exchange rate of the oligomers at equilibrium, they do not provide any specificity in sorting class I and class II subunits into separate assemblies.

Hetero-oligomerisation via dimer exchange is therefore prevented exclusively during the first of two assembly steps, through contacts made between the C-terminus and the ACD.

5.4 Discussion

Here, by investigating in depth the paralogous sHSPs HSP 17.7 and HSP 18.1, we have shown how two proteins with near identical structures manage to sort their subunits into distinct oligomers without possessing fundamentally different interfaces. We have found that dimerisation is extremely sensitive to non-interfacial parts of the protein sequence via their effect on the structure of dissociated subunits, and that interfaces that make a minor contribution to the overall thermodynamic stability can completely drive assembly specificity in oligomers. These findings are extremely surprising considering current theories about the assembly and quaternary dynamics of protein oligomers, and cast an entirely new light on what determines the equilibrium properties of oligomeric assemblies.

We have demonstrated that, despite the inability of the wild type and truncated forms of HSP 17.7 and HSP 18.1 to co-assemble, hetero-dimers are remarkably easy to engineer, including by the modification of non-bonding parts of the molecule. All of our chimeric constructs had at least some ability to hetero-dimerise, meaning that the thermodynamic properties of the wild-type core domains are delicately balanced to ensure both molecules homo-dimerise, with appropriate stability, but yet do not form hetero-dimers. We have shown that this balance can be upset directly or indirectly: by either modifying the stability of the hetero-dimer, or of one of the monomers.

Our extensive MD investigations provide evidence for this dual effect, by revealing differences in the conformations populated by the monomers, and ease with which the β6-β2 interaction can be broken. This can be seen clearly in the case of the HSP 18.1 loop which both in the monomer (Fig. 5.8A) and

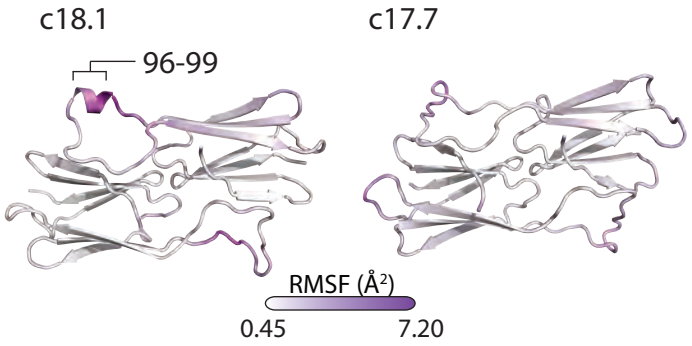


Figure 5.15: Loop dynamics in the ACD dimers. In HSP 18.1 residues 96-99 form a short helix in our simulations that is highly dynamic and introduces strain into the β2-β6 interface. The same region in c17.7 is less dynamic.

dimer (Fig. 5.15) has a propensity to form a short helix along residues 96-99. In the dimer, this helix formation occurs at the expense of straining the β6-β2 interaction. In the c17.7 monomer, on the other hand, the loop conformation resembles the dimeric conformation, and therefore does not introduce as much strain, leading to more stable dimers. To produce stable homo-dimers that do not hetero-dimerise in the wild-type molecules, the effect of the loop conformation (destabilising in c18.1, stabilising in c17.7) is balanced by an opposing effect of the β-sandwich that the loop binds to (stabilising in c18.1, destabilising in c17.7), such that the $\Delta G_{Homo-dimer}$ is similar for both. In the wild-type hetero-dimer these effects no longer cancel on the side of the HSP 18.1 loop and β6 strand binding to the c17.7 β-sandwich, producing an unstable hetero-dimer. Hetero-dimerization is consequently avoided by asymmetric partitioning thermodynamic stability into different parts of the ACDs in HSP 18.1 and HSP 17.7. This difference manifests itself most clearly in the structure and dynamics of the dissociated monomers, and is largely determined by the identity of the non-bonding parts of the loop surrounding the β6 strand.

Though our study details only one group of proteins in which evolution has encoded assembly specificity in non-interfacial regions, it likely represents a more general phenomenon. This is in line with the observation that dynamic

and flexible regions of proteins are more evolvable than rigid parts [35], and also that mutations in non-interfacial regions of protein homomers have previously been reported to drive the evolution of changes in oligomeric state by mimicking ligand-induced allosteric transitions [36]. Other examples of close homologues in which hetero-dimerisation is avoided despite sharing near identical ground-state structures include zinc-finger transcription factors [37], G-protein coupled receptors [38], and Toll-like receptors [39]. The opposite effect, in which homo-dimerisation is avoided in favour of hetero-dimers that form using the same interface, has also been described [40, 41]. The need to avoid co-assembly with proteins that share similar interfaces is therefore likely a general phenomenon in biology, and especially so when novel oligomers evolve by gene duplication. Whether avoiding such deleterious interactions generally involves asymmetric partitioning of stability in the relevant interfaces analogous to what we described here, or whether entirely different mechanisms remain to be discovered will undoubtedly be an exciting field of future research.

In our study of the full-length, dodecameric, sHSPs, we have found that the avoidance of heteromerisation is mediated exclusively by the thermodynamically weaker of two oligomeric interfaces. This is possible because assembly proceeds in two steps, beginning with the weaker interface in which the C-terminal region of one monomer binds ACD of another, and cannot proceed if this first step is blocked. This mechanism contradicts the theory that oligomers generally form their interfaces in order of their thermodynamic stability during assembly [42], although which route is taken may vary between protein families. The sequential assembly scheme described here, however, provides a physical mechanism for theories describing the evolution of transitions between stoichiometries that are often observed in protein families [5]. A polydisperse first assembly step already provides the necessary variation in stoichiometry for evolution to act on, which subsequently only needs to alter the geometry of the contacts made in the second assembly step (mediated in the case presented

here by the N-termini) to lead to a potentially dramatic shift in the equilibrium stoichiometry [43]. In sHSPs, this mechanism provides a rationale for the high sequence diversity in the N-termini [44], the large number of different stoichiometries different sHSPs populate at equilibrium [23], and the prevalence of polydispersity in the sHSP family [45]. It would also link the order in which interfaces form during assembly of sHSPs to the order in which they evolved, as has been proposed for a number of other protein complexes [42, 46].

Our detailed investigation of the facile self-assembly, and avoidance of co-assembly, of two archetypal sHSPs has fundamental implications for the study of protein oligomers in general. We have shown that the types of oligomers formsed at equilibrium by a mixture of subunits cannot be understood by simply considering the static contacts made at their interfaces. Instead, by relying on a range of biophysical tools that cover very different time and length scales, our results paint a complex picture in which assembly specificity is determined by stoichiometries and conformations that are barely detectable at equilibrium. This view represents a radical deviation from established ideas that attempt to understand oligomer assembly largely in terms of the size and strength of the interfaces formed at equilibrium. The large number of different paralogous protein oligomers in cells suggests that many other proteins may display similar behaviour to ensure only the right type of oligomer is formed. Although here we have only studied two proteins from one family, the interdisciplinary approach developed in this work puts this exciting new dimension of protein quaternary dynamics firmly within our methodological grasp.

5.5 Materials and Methods

5.5.1 Protein expression and purification

All core domains and constructs missing N-termini were codon-optimised for expression in *E. coli*, synthesised (IDT, Belgium) and cloned into a modified

pET 28b vector containing a TEV protease cleavable N-terminal his-tag using the BamHI and XhoI restriction sites using an InFusion kit (Clontech, CA). Proteins were expressed and purified as described previously [34]. Purified protein was concentrated, snap-frozen and stored at -80 °C until use. Wild-type HSP 17.7 and HSP 18.1 were purified as described previously [21]. All full-length proteins were cloned into pET 28b using the NdeI and BamHI sites to produce an untagged gene-product and were purified using the same protocol as for the wild-type proteins [21], but taking different ammonium sulfate cuts and adjusting the ion exchange buffers as appropriate.

5.5.2 Native (ion-mobility) mass spectrometry

Native MS experiments were performed on Q-Tof2 instruments (Waters, Wilmslow, UK), modified for analysis of large protein ions [29]. Parameters for all experiments were: capillary voltage: 1.3-1.7 V. Cone voltage: 10-80 V, extractor voltage: 10-20V, collision voltage: 10-80 V for non-activating spectra and 80-200 V for CID. The collision cell was pressurized with Argon at around 35 μbar. All ion mobility mass spectra were recorded on a modified first generation Synapt HDMS instrument (Waters, Wilmslow, UK) fitted with either with a travelling wave IM device, or a radially-confining drift tube [47]. For core domain spectra and peptide binding experiments a vial of isopropanol was introduced into the source region for charge reduction of the generated ions [48]. Settings for all experiments involving core domains were as previously described [49], and collisional cross sections were measured as described in [47]. ^{13}C and ^{13}C^{15}N -labelled protein was produced by growing cells in M9 media supplemented with ^{13}C glucose as the only carbon source, and either unlabelled or ^{15}N-labelled ammonium chloride. Labelled protein was purified as described above.

5.5.3 Subunit exchange of full-length proteins

Unlabelled and labelled samples in 200 mM ammonium acetate were equilibrated to the desired temperature using a water bath for at least 20 minutes before mixing. Labelled and unlabelled samples were mixed and incubated at the desired temperature in a water bath. At the desired time-points, aliquots were taken and snap-frozen in liquid nitrogen. Samples were kept in liquid nitrogen until analysis at room temperature. Spectra for individual time-points were fit using UniDec, a modified version of a previous deconvolution algorithm [50], to extract relative intensities [51]. Kinetic traces were fitted to a model describing concurrent subunit exchange of monomers and dimers based on a pervious model [52] implemented in Python by Andrew Baldwin. Fits were bootstrapped 1000 times to obtain means and standard deviations of all fitted parameters.

5.5.4 Extraction of dimer K_D values

Dimer titrations were performed by keeping the concentration of one ACD at 10μ and varying the concenrtatio of the other between 1 and 60 μM in 12 steps. K_D values for homo-dimers were calculated by extracting relative intensities of monomers and dimers from spectra using UniDec, and fitting the relative amount of dimer to the expression:

$$[D] = \frac{(4M_0 + K_D - \sqrt{8M_0K_D + K_D^2}}{8} \tag{5.1}$$

where [D] is the molar fraction of dimer and M_0 is the total protein concentration. The molar fraction of hetero-dimer AB formed between any two monomers A and B at given concentrations of both monomers was calculated by providing an initial guess for the concentration of free monomers [A] and

[B] of both components and evaluating the system:

$$[A_2] = \frac{[A]^2}{K_{D1}} \tag{5.2}$$

$$[B_2] = \frac{[B]^2}{K_{D2}} \tag{5.3}$$

$$[AB] = \frac{2[A][B]}{K_{D3}} \tag{5.4}$$

The factor of 2 in the hetero-dimer equilibrium adjusts the K_{D3} for the fact that there are more reaction partners in a hetero-dimeric reaction, and consequently twice as many collisions can occur potentially leading to hetero-dimer formation compared to homo-dimer formation. This term corresponds to a mixing entropy whose contribution to the $\Delta G_{Heterodimer}$ is equal to $-RTln(2)$ [53]. Since this term would even be added to the ΔG of hetero-dimers where both subunits interact in the heterodimer exactly like in the homo-dimers (leading to a lower K_D in the heterodimer) we removed this constant contribution in our analysis. This allowed us to evaluate the stability of the dimer interface across all homo- and heterodimers tested in terms of just the contacts made in the dimers.

The correct concentrations $[A]$ and $[B]$ for a given set of K_D values and total concentrations of A and B was found by calculating the total concentration of A and B as:

$$A_{tot} = [A] + 2[A_2] + [AB] \tag{5.5}$$

$$B_{tot} = [B] + 2[B_2] + [AB] \tag{5.6}$$

and minimizing the difference between the calculated A_{tot} and B_{tot} and the experimental A_{tot} and B_{tot} by varying $[A]$ and $[B]$. K_D values for the hetero-dimers (K_{D3}) were determined by fixing K_{D1} and K_{D2} to the values determined in titrations of homo-dimers. At each step of the fit the relative amounts of $A, B, A2, B2$ and AB were found using the procedure outlined above. Two additional parameters were introduced as multipliers of the experimental total concentrations of A and B at each titration point to account for different ionization and transmission efficiencies of the two types of monomers, and the

error in their respective concentrations. The ratio of these two parameters was never above 5 or below 0.2, well within the range previously reported in the literature for ligand binding experiments [54]. Fits were bootstrapped 1000 times to obtain means and standard deviations of all fitted quantities.

5.5.5 Crystallography

c17.7 was crystallized in 200 mM $CaCl_2$, 100 mM sodium cacodylate pH 6.5, 40 % (v/v) polyethylene glycol (PEG) 200 in hanging drop plates at room temperature. c18.1 crystallised in 200 mM Li_2SO_4, 100 mM Tris-HCl, pH8.5, 30 % (w/v) PEG 400 in hanging drop plates at room temperature. Crystals of both c17.7 and c18.1 were flash-frozen without addition of cryoprotectant. Diffraction images were processed using XDS. Structures were solved by molecular replacement using the HSP 16.9 dimer (PDB: 1GME) for c18.1 and c18.1 for c17.7 as search models. Manual refinement and validation was carried out in Phenix [56].

5.5.6 Peptide binding experiments

Peptides were obtained from Biomatik (Canda). Peptide binding experiments were carried out and analysed as described previously [49]. The concentration of the core domain was kept at 10 μM in each experiment. Peptide concentrations were 10, 20, 40, 80 and 160 μM.

5.5.7 Small angle X-ray scattering

SAXS data were collected at the B21 bending magnet instrument at Diamond Light Source (Harwell, UK). Samples were prepared in 200 mM ammonium acetate to a concentration of 5 mg/ml. Protein and corresponding buffer solutions were exposed to the beam in a 1.6 mm diameter quartz capillary at 15 °C. The sample capillary was held in vacuum and subjected to a cleaning cycle between each measurement. A Pilatus 2M two-dimensional detector was used to collect 180 frame exposures of 1 s from each sample and the corresponding buffer. The

	c17.7	c18.1
Crystal parameters		
Space group	C121	$P3_221$
Cell dimensions		
a,b,c (Å)	48.86, 82.69, 88.90	89.06, 89.06, 142.90
$\alpha, \beta, \gamma(^\circ)$	90, 92.108, 90	90, 90, 120
Molecules in A.U.[a]	3	6
Data collection		
Synchrotron	In-house	Diamond
Beamline		IO4
Wavelength (Å)	1.542	0.843
Resolution (Å)	35.02-2.63	77.13-1.85
Reflections observed/ unique	36842 / 10426	566725 / 56563
Completeness (%)	98.5 (93.7)[c]	99.9 (99.9)
$I/\sigma I$	7.34 (1.95)	35.12 (1.48)
Refinement		
Resolution (Å)	35.02-2.63	77.13-1.85
$R_{work}(\%)^e$	21.5	20.3
$R_{free}(\%)^f$	26.4	24.5
No. of non H-atoms		
Protein	2047	4532
Non-protein	1	425
Avg. B- factors	67.76	52.57
R.m.s deviation		
Bond length (Å)	0.015	0.017
Bond angle (°)	1.85	1.746

Table 5.1: Crystallographic table for the HSP 17.7 and HSP 18.1 ACDs. a. A.U. = asymmetric unit. b. Data were collected at 100K using a Rigaku Superbright rotating anode diffractometer equipped with a Saturn 944x detector. c. Values in parentheses correspond to the highest resolution shell. d. $R_{mrgd-F} = \sum|(AI(h,P) - AI(h,Q)|)/(0.5 * \sum AI(h,P) + AI(h,q))$ where $AI = (\sqrt{I}$ if $I >= 0$ or $-\sqrt{I}$ if $I < 0)$ as described previously [55]. e. $R_{work} = \sum|F_o - F_c|/\sum F_o$. e. R_{free} calculated from 5% of the data.

detector was placed at 3.9 m from the sample, giving a useful Q-range of 0.012 $Å^{-1} < 0.4$ $Å^{-1}$. Two dimensional data reduction consisted of normalisation for beam current and sample transmission, radial sector integration, background buffer subtraction and averaging. Each frame was inspected manually and discarded if signs of radiation damage were apparent. Data scaling, merging and Guinier analysis were performed in PRIMUS [57].

5.5.8 SAXS learning algorithm

For both HSP 17.7 and HSP 18.1, we generated an ensemble of 1000 random tetrahedral, 1000 double ring and 1000 circle assemblies using a method similar to one that was previously described [58]. Briefly, models were assembled using a two-step assembly modelling strategy aimed at minimizing the amount of missing atoms in models. We first produced a random clash-free assembly consisting just of our crystal structures of the ACDs (residues 50-142 for HSP 17.7 and 48-140 for HSP 18.1) on a scaffold corresponding to the appropriate architecture. N-termini were subsequently modeled by obtaining a homologous sequence using the PSIPRED webserver (PDB: 1BUC for HSP 17.7, PDB: 3HHV for HSP 18.1), and packing it into the core of the structure so that no clashes are produced and respecting the same symmetry as for the ACDs. A Random Forest learning algorithm [59] was trained on SAXS signals simulated between $Q = 0$ and 0.3 $Å^{-1}$ using CRYSOL [60] for a random set of 800 of the resulting structures. The remaining 200 models per topology were used for testing the accuracy of the learning algorithm. Classification accuracy was determined as 99.33 % for HSP 17.7 and 95 % for HSP 18.1. In all cases the classification errors were exclusively between double rings and circles.

5.5.9 Molecular dynamics

All simulations were carried out using the Gromacs software [61, 62] at 300 K using periodic boundary conditions. The crystallographic models for c17.7 and c18.1 were used with the Amber99sb-ildn forcefield to yield monomeric and

homo-dimeric systems for MD simulation. Dimers and monomers were put in a periodic cell corresponding to a rhombic dodecahedron with a shortest distance of 12 Å between the protein and the box edge. The proteins were solvated in Tip4p-ew water [63] and NaCl was added to a concentration of 0.154 M. All systems were subjected to steepest descent minimisation followed by 100 ps NVT (canonical ensemble) simulation at 300 K with position restraints applied to all heavy atoms in the protein, using a force constant of 1000 kJ mol^{-1} nm-2, and using the v-rescale thermostat [64] with a time constant of 200 fs. A 1 ns NVT simulation without position restraints allowed for initial relaxation, followed by 1 ns of NPT (isothermal-isobaric ensemble) simulation with Berendsen pressure coupling [65] to equilibrate the pressure. Dimers and monomers were simulated in the NPT ensemble, using a Parrinello-Rahman barostat [66] for 2 μs each. Chimeric monomers were equilibrated and run as described above, producing one 2 μs trajectory per chimera. All simulations in this study employed the Lincs algorithm [67, 68] to constrain bond lengths, SHAKE [69] to keep water molecules rigid, and virtual interaction sites [70] for hydrogens, allowing for leap-frog integration [71] of the forces in 4 fs time steps. The particle mesh Ewald method [72] was used for electrostatic interaction outside of a 10 Å cut-off and shifted Lennard-Jones interactions were used to model Van der Waals interactions within 10 Å.

A hetero-dimer was constructed from the c17.7 and c18.1 models by first superimposing one a c17.7 monomer onto the dimeric c18.1 model, then subjecting the resulting dimer to energy minimisation. To relax the dimeric structure, two subsequent simulations with position restraints applied to the protein were conducted. The first was run for 100 ps, using the heavy atoms from the starting configuration as reference coordinates for the restraints. The second was run for 25 ns, using the backbone atoms from dimeric c18.1 as reference coordinates for restraining the inserted c17.7 monomer in order to allow for the side chains to relax into the dimeric structure while keeping the backbone in

place. NVT and NPT relaxation and 2 μs of NVT production simulation were then carried out, following the methods described for the homo-dimers.

For umbrella sampling simulations a force was applied between the respective centre of mass of the backbone atoms of the strand and the hydrogen-bonding residues in the sheet to separate the two parts over 200 ns. These pulling simulations were used to yield starting structures for 21 umbrella-sampling simulations per interface along the reaction coordinate defined by the strand-sheet distance. Each umbrella simulation was run for 500 ns.

The number of hydrogen bonds formed in the dimer trajectories was calculated using Gromacs version 5.0.1 and setting 3 Å donor-acceptor distance and 20° angles between donor and acceptor as cut-offs. Principal component analysis and computation of RMSFs were also carried out in Gromacs version 5.0.1, using only the backbone α-carbons for all calculations.

5.5.10 Analysis of sequence conservation

Conservation scores were calculated by first inferring a phylogenetic tree for 89 class II sequence and 211 class I sequences that were aligned using the T-coffee server [73]. The tree was calculated using the core domain and C-terminal parts of the sequence in the MrBayes program [74]. Normalised conservation scores using this tree were calculated using the ConSurf [75] server for the whole tree and the sub-trees containing either just class I or just class II sequences. To eliminate sequence conservation signal derived from residues conserved in both families the conservation scores calculated for the whole tree were subtracted from the scores calculated for the individual families.

5.6 References

[1] K. Henrick and J. M. Thornton, "PQS: a protein quaternary structure file server," *Trends Biochem. Sci.*, vol. 23, no. 9, pp. 358–61, 1998.

[2] E. D. Levy, J. B. Pereira-Leal, C. Chothia, and S. A. Teichmann, "3D complex: a structural classification of protein complexes," *PLoS Comput. Biol.*, vol. 2, no. 11, p. e155, 2006.

[3] D. S. Goodsell and A. J. Olson, "Structural symmetry and protein function," *Annu. Rev. Biophys. Biomol. Struct.*, vol. 29, pp. 105–53, 2000.

[4] S. Bershtein, W. Mu, and E. I. Shakhnovich, "Soluble oligomerization provides a beneficial fitness effect on destabilizing mutations," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 109, no. 13, pp. 4857–62, 2012.

[5] M. Lynch, "Evolutionary diversification of the multimeric states of proteins," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 110, no. 30, pp. E2821–8, 2013.

[6] I. Ispolatov, A. Yuryev, I. Mazo, and S. Maslov, "Binding properties and evolution of homodimers in protein-protein interaction networks," *Nucleic Acids Res.*, vol. 33, no. 11, pp. 3629–35, 2005.

[7] S. A. Teichmann, J. Park, and C. Chothia, "Structural assignments to the *Mycoplasma genitalium* proteins show extensive gene duplications and domain rearrangements," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 95, no. 25, pp. 14658–63, 1998.

[8] S. Ohno, *Evolution by gene duplication*. Berlin, New York,: Springer-Verlag, 1970.

[9] M. Lynch and A. Force, "The probability of duplicate gene preservation by subfunctionalization," *Genetics*, vol. 154, no. 1, pp. 459–73, 2000.

[10] G. C. Finnigan, V. Hanson-Smith, T. H. Stevens, and J. W. Thornton, "Evolution of increased complexity in a molecular machine," *Nature*, vol. 481, no. 7381, pp. 360–4, 2012.

[11] A. J. Reid, J. A. Ranea, and C. A. Orengo, "Comparative evolutionary analysis of protein complexes in *E. coli* and yeast," *BMC Genomics*, vol. 11, p. 79, 2010.

[12] J. B. Pereira-Leal, E. D. Levy, C. Kamp, and S. A. Teichmann, "Evolution of protein complexes by duplication of homomeric interactions," *Genome Biol.*, vol. 8, no. 4, p. R51, 2007.

[13] M. E. Johnson and G. Hummer, "Nonspecific binding limits the number of proteins in a cell and shapes their interaction networks," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 108, no. 2, pp. 603–8, 2011.

[14] H. B. Fraser, A. E. Hirsh, L. M. Steinmetz, C. Scharfe, and M. W. Feldman, "Evolutionary rate in the protein interaction network," *Science*, vol. 296, no. 5568, pp. 750–2, 2002.

[15] W. Qian, X. He, E. Chan, H. Xu, and J. Zhang, "Measuring the evolutionary rate of protein-protein interaction," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 108, no. 21, pp. 8725–30, 2011.

[16] J. E. Dayhoff, B. A. Shoemaker, S. H. Bryant, and A. R. Panchenko, "Evolution of protein binding modes in homooligomers," *J. Mol. Biol.*, vol. 395, no. 4, pp. 860–70, 2010.

[17] A. Wagner, "The yeast protein interaction network evolves rapidly and contains few redundant duplicate genes," *Mol. Biol. Evol.*, vol. 18, no. 7, pp. 1283–92, 2001.

[18] J. A. Aquilina, S. Shrestha, A. M. Morris, and H. Ecroyd, "Structural and functional aspects of hetero-oligomers formed by the small heat shock proteins αB-crystallin and HSP27," *J. Biol. Chem.*, vol. 288, no. 19, pp. 13602–9, 2013.

[19] E. Franck, O. Madsen, T. van Rheede, G. Ricard, M. A. Huynen, and W. W. de Jong, "Evolutionary diversity of vertebrate small heat shock proteins," *J. Mol. Evol.*, vol. 59, no. 6, pp. 792–805, 2004.

[20] J. A. Thomson and R. C. Augusteyn, "On the structure of α-crystallin: construction of hybrid molecules and homopolymers," *Biochim. Biophys. Acta*, vol. 994, no. 3, pp. 246–52, 1989.

[21] E. Basha, C. Jones, V. Wysocki, and E. Vierling, "Mechanistic differences between two conserved classes of small heat shock proteins found in the plant cytosol," *J. Biol. Chem.*, vol. 285, no. 15, pp. 11489–97, 2010.

[22] H. S. McHaourab, J. A. Godar, and P. L. Stewart, "Structure and mechanism of protein stability sensors: chaperone activity of small heat shock proteins," *Biochemistry*, vol. 48, no. 18, pp. 3828–37, 2009.

[23] G. R. Hilton, H. Lioe, F. Stengel, A. J. Baldwin, and J. L. Benesch, "Small heat-shock proteins: paramedics of the cell," *Top. Curr. Chem.*, vol. 328, pp. 69–98, 2013.

[24] E. R. Waters, B. D. Aevermann, and Z. Sanders-Reed, "Comparative analysis of the small heat shock proteins in three angiosperm genomes identifies new subfamilies and reveals diverse evolutionary patterns," *Cell Stress Chaperones*, vol. 13, no. 2, pp. 127–42, 2008.

[25] A. J. Painter, N. Jaya, E. Basha, E. Vierling, C. V. Robinson, and J. L. Benesch, "Real-time monitoring of protein complexes reveals their quaternary organization and dynamics," *Chem. Biol.*, vol. 15, no. 3, pp. 246–53, 2008.

[26] F. Sobott, J. L. Benesch, E. Vierling, and C. V. Robinson, "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.*, vol. 277, no. 41, pp. 38921–9, 2002.

[27] M. Kirschner, S. Winkelhaus, J. M. Thierfelder, and L. Nover, "Transient expression and heat-stress-induced co-aggregation of endogenous and heterologous small heat-stress proteins in tobacco protoplasts," *Plant J.*, vol. 24, no. 3, pp. 397–411, 2000.

[28] E. R. Waters and E. Vierling, "The diversification of plant cytosolic small heat shock proteins preceded the divergence of mosses," *Mol. Biol. Evol.*, vol. 16, no. 1, pp. 127–39, 1999.

[29] H. Hernández and C. V. Robinson, "Determining the stoichiometry and interactions of macromolecular assemblies from mass spectrometry," *Nat. Protoc.*, vol. 2, no. 3, pp. 715–26, 2007.

[30] R. L. van Montfort, E. Basha, K. L. Friedrich, C. Slingsby, and E. Vierling, "Crystal structure and assembly of a eukaryotic small heat shock protein," *Nat. Struct. Biol.*, vol. 8, no. 12, pp. 1025–30, 2001.

[31] C. K. Kennaway, J. L. Benesch, U. Gohlke, L. Wang, C. V. Robinson, E. V. Orlova, H. R. Saibil, and N. H. Keep, "Dodecameric structure of the small heat shock protein Acr1 from *Mycobacterium tuberculosis*," *J. Biol. Chem.*, vol. 280, no. 39, pp. 33419–25, 2005.

[32] A. Laganowsky, J. L. Benesch, M. Landau, L. Ding, M. R. Sawaya, D. Cascio, Q. Huang, C. V. Robinson, J. Horwitz, and D. Eisenberg, "Crystal structures of truncated αA and αB crystallins reveal structural mechanisms of polydispersity important for eye lens function," *Protein Sci.*, vol. 19, no. 5, pp. 1031–43, 2010.

[33] N. Braun, M. Zacharias, J. Peschek, A. Kastenmuller, J. Zou, M. Hanzlik, M. Haslbeck, J. Rappsilber, J. Buchner, and S. Weinkauf, "Multiple molecular architectures of the eye lens chaperone αB-crystallin elucidated by a triple hybrid approach," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 108, no. 51, pp. 20491–6, 2011.

[34] G. K. Hochberg, H. Ecroyd, C. Liu, D. Cox, D. Cascio, M. R. Sawaya, M. P. Collier, J. Stroud, J. A. Carver, A. J. Baldwin, C. V. Robinson, D. S. Eisenberg, J. L. Benesch, and A. Laganowsky, "The structured core domain of αB-crystallin can prevent amyloid fibrillation and associated toxicity," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 111, no. 16, pp. E1562–70, 2014.

[35] N. Tokuriki and D. S. Tawfik, "Protein dynamism and evolvability," *Science*, vol. 324, no. 5924, pp. 203–7, 2009.

[36] T. Perica, Y. Kondo, S. P. Tiwari, S. H. McLaughlin, K. R. Kemplen, X. Zhang, A. Steward, N. Reuter, J. Clarke, and S. A. Teichmann, "Evolution of oligomeric state through allosteric pathways that mimic ligand binding," *Science*, vol. 346, no. 6216, p. 1254346, 2014.

[37] A. S. McCarty, G. Kleiger, D. Eisenberg, and S. T. Smale, "Selective dimerization of a C2H2 zinc finger subfamily," *Mol. Cell*, vol. 11, no. 2, pp. 459–70, 2003.

[38] S. C. Prinster, C. Hague, and R. A. Hall, "Heterodimerization of G protein-coupled receptors: specificity and functional significance," *Pharmacol. Rev.*, vol. 57, no. 3, pp. 289–98, 2005.

[39] K. Takeda and S. Akira, "Toll-like receptors in innate immunity," *Int. Immunol.*, vol. 17, no. 1, pp. 1–14, 2005.

[40] T. D. Halazonetis, K. Georgopoulos, M. E. Greenberg, and P. Leder, "c-Jun dimerizes with itself and with c-Fos, forming complexes of different DNA binding affinities," *Cell*, vol. 55, no. 5, pp. 917–24, 1988.

[41] E. K. O'Shea, R. Rutkowski, r. Stafford, W. F., and P. S. Kim, "Preferential heterodimer formation by isolated leucine zippers from fos and jun," *Science*, vol. 245, no. 4918, pp. 646–8, 1989.

[42] E. D. Levy, E. Boeri Erba, C. V. Robinson, and S. A. Teichmann, "Assembly reflects evolution of protein complexes," *Nature*, vol. 453, no. 7199, pp. 1262–5, 2008.

[43] T. Perica, C. Chothia, and S. A. Teichmann, "Evolution of oligomeric state through geometric coupling of protein interfaces," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 109, no. 21, pp. 8127–32, 2012.

[44] T. Kriehuber, T. Rattei, T. Weinmaier, A. Bepperling, M. Haslbeck, and J. Buchner, "Independent evolution of the core domain and its flanking sequences in small heat shock proteins," *FASEB J.*, vol. 24, no. 10, pp. 3633–42, 2010.

[45] D. A. Haley, M. P. Bova, Q. L. Huang, H. S. McHaourab, and P. L. Stewart, "Small heat-shock protein structures reveal a continuum from symmetric to variable assemblies," *J. Mol. Biol.*, vol. 298, no. 2, pp. 261–72, 2000.

[46] J. A. Marsh, H. Hernandez, Z. Hall, S. E. Ahnert, T. Perica, C. V. Robinson, and S. A. Teichmann, "Protein complexes are under evolutionary selection to assemble via ordered pathways," *Cell*, vol. 153, no. 2, pp. 461–70, 2013.

[47] M. F. Bush, Z. Hall, K. Giles, J. Hoyes, C. V. Robinson, and B. T. Ruotolo, "Collision cross sections of proteins and their complexes: a calibration framework and database for gas-phase structural biology," *Anal. Chem.*, vol. 82, no. 22, pp. 9557–65, 2010.

[48] J. T. Hopper, K. Sokratous, and N. J. Oldham, "Charge state and adduct reduction in electrospray ionization-mass spectrometry using solvent vapor exposure," *Anal. Biochem.*, vol. 421, no. 2, pp. 788–90, 2012.

[49] G. R. Hilton, G. K. Hochberg, A. Laganowsky, S. I. McGinnigle, A. J. Baldwin, and J. L. Benesch, "C-terminal interactions mediate the

quaternary dynamics of αB-crystallin," *Philos. Trans. R. Soc. Lond. B. Biol. Sci.*, vol. 368, no. 1617, p. 20110405, 2013.

[50] M. T. Marty, H. Zhang, W. Cui, M. L. Gross, and S. G. Sligar, "Interpretation and deconvolution of nanodisc native mass spectra," *J. Am. Soc. Mass Spectrom.*, vol. 25, no. 2, pp. 269–77, 2014.

[51] D. A. Shepherd, M. T. Marty, K. Giles, A. J. Baldwin, and J. L. P. Benesch, "Combining tandem mass spectrometry with ion mobility separation to determine the architecture of polydisperse proteins," *Int. J. Mass Spectrom.*, p. in press, 2015.

[52] A. J. Baldwin, H. Lioe, C. V. Robinson, L. E. Kay, and J. L. Benesch, "αB-crystallin polydispersity is a consequence of unbiased quaternary dynamics," *J. Mol. Biol.*, vol. 413, no. 2, pp. 297–309, 2011.

[53] J. Dam, T. Rose, M. Goldberg, and A. Blondel, "Complementation between dimeric mutants as a probe of dimer-dimer interactions in tetrameric dihydrofolate reductase encoded by r67 plasmid of *E. coli*," *Journal of Molecular Biology*, vol. 302, pp. 235–250, Sept. 2000.

[54] V. Gabelica, N. Galic, F. Rosu, C. Houssier, and E. De Pauw, "Influence of response factors on determining equilibrium association constants of non-covalent complexes by electrospray ionization mass spectrometry," *J. Mass Spectrom.*, vol. 38, no. 5, pp. 491–501, 2003.

[55] W. Kabsch, "Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants," *J. Appl. Crystallogr.*, vol. 26, pp. 795–800, 1993.

[56] P. D. Adams, P. V. Afonine, G. Bunkoczi, V. B. Chen, I. W. Davis, N. Echols, J. J. Headd, L. W. Hung, G. J. Kapral, R. W. Grosse-Kunstleve, A. J. McCoy, N. W. Moriarty, R. Oeffner, R. J. Read, D. C. Richardson, J. S. Richardson, T. C. Terwilliger, and P. H. Zwart,

"PHENIX: a comprehensive Python-based system for macromolecular structure solution," *Acta Crystallogr. D Biol. Crystallogr.*, vol. 66, no. Pt 2, pp. 213–21, 2010.

[57] M. V. Petoukhov, D. Franke, A. V. Shkumatov, G. Tria, A. G. Kikhney, M. Gajda, C. Gorba, H. D. Mertens, P. V. Konarev, and D. I. Svergun, "New developments in the program package for small-angle scattering data analysis," *J. Appl. Crystallogr.*, vol. 45, no. Pt 2, pp. 342–350, 2012.

[58] A. J. Baldwin, H. Lioe, G. R. Hilton, L. A. Baker, J. L. Rubinstein, L. E. Kay, and J. L. Benesch, "The polydispersity of αB-crystallin is rationalized by an interconverting polyhedral architecture," *Structure*, vol. 19, no. 12, pp. 1855–63, 2011.

[59] L. Breiman, "Random Forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[60] D. Svergun, C. Barberato, and M. H. J. Koch, "CRYSOL - A program to evaluate x-ray solution scattering of biological macromolecules from atomic coordinates," *J. Appl. Crystallogr.*, vol. 28, pp. 768–773, 1995.

[61] S. Pronk, S. Pall, R. Schulz, P. Larsson, P. Bjelkmar, R. Apostolov, M. R. Shirts, J. C. Smith, P. M. Kasson, D. van der Spoel, B. Hess, and E. Lindahl, "GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit," *Bioinformatics*, vol. 29, no. 7, pp. 845–854, 2013.

[62] D. Van der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, and H. J. C. Berendsen, "Gromacs: Fast, flexible, and free," *J. Comput. Chem.*, vol. 26, no. 16, pp. 1701–1718, 2005.

[63] H. W. Horn, W. C. Swope, J. W. Pitera, J. D. Madura, T. J. Dick, G. L. Hura, and T. Head-Gordon, "Development of an improved four-site

water model for biomolecular simulations: TIP4P-Ew," *J. Chem. Phys.*, vol. 120, no. 20, pp. 9665–78, 2004.

[64] G. Bussi, D. Donadio, and M. Parrinello, "Canonical sampling through velocity rescaling," *J. Chem. Phys.*, vol. 126, no. 1, 2007.

[65] H. J. C. Berendsen, J. P. M. Postma, W. F. Vangunsteren, A. Dinola, and J. R. Haak, "Molecular-dynamics with coupling to an external bath," *J. Chem. Phys.*, vol. 81, no. 8, pp. 3684–3690, 1984.

[66] M. Parrinello and A. Rahman, "Polymorphic transitions in single-crystals - a new molecular-dynamics method," *J. Appl. Phys.*, vol. 52, no. 12, pp. 7182–7190, 1981.

[67] B. Hess, "P-LINCS: A parallel linear constraint solver for molecular simulation," *J. Chem. Theory Comput.*, vol. 4, no. 1, pp. 116–122, 2008.

[68] B. Hess, H. Bekker, H. J. C. Berendsen, and J. G. E. M. Fraaije, "LINCS: A linear constraint solver for molecular simulations," *J. Comput. Chem.*, vol. 18, no. 12, pp. 1463–1472, 1997.

[69] J. P. Ryckaert, G. Ciccotti, and H. J. C. Berendsen, "Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes," *J. Comput. Phys.*, vol. 23, no. 3, pp. 327–341, 1977.

[70] K. A. Feenstra, B. Hess, and H. J. C. Berendsen, "Improving efficiency of large time-scale molecular dynamics simulations of hydrogen-rich systems," *J. Comput. Chem.*, vol. 20, no. 8, pp. 786–798, 1999.

[71] R. W. Hockney and J. W. Eastwood, *Computer simulation using particles*. Bristol England ; Philadelphia: A. Hilger, special student ed., 1988.

[72] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. G. Pedersen, "A smooth particle mesh ewald," *J. Chem. Phys.*, vol. 103, no. 19, pp. 8577–8592, 1995.

[73] H. McWilliam, W. Li, M. Uludag, S. Squizzato, Y. M. Park, N. Buso, A. P. Cowley, and R. Lopez, "Analysis Tool Web Services from the EMBL-EBI," *Nucleic Acids Res.*, vol. 41, no. Web Server issue, pp. W597–600, 2013.

[74] F. Ronquist and J. P. Huelsenbeck, "MrBayes 3: bayesian phylogenetic inference under mixed models," *Bioinformatics*, vol. 19, no. 12, pp. 1572–4, 2003.

[75] H. Ashkenazy, E. Erez, E. Martz, T. Pupko, and N. Ben-Tal, "ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids," *Nucleic Acids Res.*, vol. 38, no. Web Server issue, pp. W529–33, 2010.

Chapter 6

Concluding remarks

This thesis has provided novel insights into the nature and functional significance of quaternary dynamics in sHSPs. The findings are surprising on many levels: the complex quaternary architecture and dynamics of αB-crystallin described in part in chapters 2 and 4 does not seem to be necessary for its activity as a chaperone, at least in relatively simple aggregation systems. Nevertheless, the prevalence of such dynamics in the sHSP family unequivocally points towards their importance. Perhaps the significance of oligomeric state, polydispersity and subunit exchange are therefore only partly to be found in direct effects on activity, and rather might lie in the ability to regulate which types of subunits co-assemble into hetero-oligomers. The intricately balanced oligomeric incompatibility between HSP 17.7 and HSP 18.1 certainly points strongly in this direction. Whatever the ultimate reason for its existence may be, the complexity of the dynamical behaviour uncovered in this work casts a new light onto the nature of oligomeric assemblies in biology. The types of oligomers formed in mixtures of subunits depends on more than just the geometry of their respective interfaces, meaning that understanding why certain complexes do not form at equilibrium is as important as studying the ones that do.

The approaches developed in this thesis demonstrate how much the study of protein quaternary structure and dynamics benefits from a combination of biophysical techniques. Especially low-resolution techniques, like IM-MS and

SAXS, that can cope with heterogeneous samples and fast dynamics are well suited to tackle the particular challenges associated with the study of protein complexes. Combining the use of such low-resolution techniques on the level of quaternary organization with atomic-resolution techniques like X-ray crystallography, NMR, and MD on the level of individual subunits will help tackle one of the biggest structural challenges in this developing field: to understand the link between the structural fluctuations on the secondary and tertiary level of individual subunits and the self-assembly of protein complexes. The work presented in this thesis demonstrates that these levels can be linked in very counter-intuitive ways, involving non-interfacial parts of the sequence and very transiently populated conformations. As it is only now becoming technically possible to interrogate protein quaternary dynamics at the level of detail required to uncover such links, the years lying ahead promise to be exciting ones for studying this fundamental aspect of cellular biochemistry.