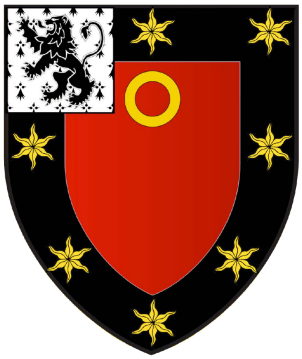
# Polydisperse chaperone proteins and the mechanisms by which they inhibit aggregation

Thesis presented for the degree of
Doctor of Philosophy
University of Oxford

St John's College



Olga Tkachenko

24 April 2018

# Abstract

The small heat-shock proteins (sHSPs) are ATP-independent molecular chaperones, found in all kingdoms of life. In humans their malfunction is associated with neurodegenerative disease, cataracts, myopathies and cancer. Their molecular mechanism is contentious as sHSPs are challenging experimental targets: they typically exist as large, polydisperse and exchanging oligomers. This heterogeneity is compounded by that of aggregating substrates.

Here I investigate the structure, dynamics and function of a major human sHSP, $\alpha$B-crystallin ($\alpha$BC), using nuclear magnetic resonance (NMR) and native mass spectrometry (MS). Following the introduction, Chapters 2 and 3 focus on structure and dynamics of $\alpha$BC. In Chapter 2, I elucidate an allosteric communication network between two interfaces of the core $\alpha$-crystallin domain within the oligomers. In Chapter 3, I show that the connectivity of oligomer vertices is consistent with polyhedral architectures, and present size measurements of individual $\alpha$BC stoichiometries within the polydisperse ensemble.

Chapters 4-7 link these insights to the chaperone function of $\alpha$BC. Chapter 4 presents a new methodology to derive kinetic mechanisms of aggregation, using the model protein $\alpha$-lactalbumin ($\alpha$Lac). I show that $\alpha$Lac aggregation proceeds along two pathways via an unstable dimeric nucleus. Chapter 5 derives a kinetic mechanism for the chaperone action of $\alpha$BC on $\alpha$Lac. Remarkably, $\alpha$BC disassembles the aggregation nuclei, and is in kinetic competition with $\alpha$Lac in large aggregates. In Chapter 6, I characterise the size and shape of these large $\alpha$BC-$\alpha$Lac complexes. Chapter 7 identifies $\alpha$BC residues affected by $\alpha$Lac, showing that the substrate engages a subset of the interactions regulating quaternary dynamics of $\alpha$BC.

Thus I have exploited the complementarity of NMR and native MS to obtain insight into function of $\alpha$BC, by studying its structure and dynamics from atomic-resolution to quaternary level. The results expand the arsenal of mechanisms by which molecular chaperones maintain cellular proteostasis.

*To my family - I love you to the moon and back.*

# Declaration

This thesis is a summary of research conducted at the Department of Chemistry, University of Oxford between September 2013 and October 2017. The work described here is wholly my own unless indicated below and in the text. It has not, neither in whole nor in part, been submitted for a degree, diploma or any other academic qualification at any other academic institution.

Olga Tkachenko
October 2017

I gratefully acknowledge the following collaborations. The code for analysing chemical exchange from titration data (**Chapter 2**), and the core of the chemical kinetics code for analysis of NMR data (**Chapters 4** and **5**) was written by Prof. Andrew Baldwin. Part of the investigation of the groove-tail crosslinks in $\alpha$B-crystallin (**Chapter 3**) was performed by Oliver Stach during his internship in September-December 2015 under my supervision. Dr Todd Mize kindly assembled the breakout cable device and assisted in performing the voltage crosstalk checks (**Chapter 3**). The work on detection of complexes between full-length $\alpha$B-crystallin (**Chapter 6**) was performed in collaboration with Dr Frances Kondrat. The pulse sequence for the DEST experiment (**Chapter 7**) was written by Gogulan Karunanithy.

# Acknowledgements

*The world is full of magic things, patiently waiting for our senses to grow sharper.*

William Butler Yeats (1865-1939)

*I will either find a way, or make one.*

Hannibal (247-183/181 BC)

Proteins can breathe. Sometimes a little, sometimes a lot, and sometimes these motions are key to molecular recognition, enzyme catalysis, molecular recycling and other remarkable protein functions that life is built on. I owe my predilection for protein dynamics and its study by nuclear magnetic resonance, one of the most complex and powerful atomic-resolution techniques, to Prof. Bill Broadhurst at the University of Cambridge. Prof. Matthias Mayer at the University of Heidelberg introduced me to the practical application of biological mass spectrometry and showed me how it too can report on dynamics, with direct access to complex composition and subunit exchange being very complementary to NMR.

So, I developed the idea of coupling NMR and MS in my PhD to interrogate the role of different levels of protein dynamics in a biological system. I thank Prof. Justin Benesch and Prof. Andrew Baldwin, experts in native MS and high molecular weight NMR, for co-supervision, which allowed me to follow through on this interest at Oxford. Together, we did cutting-edge work on a class of molecular chaperones – guardians for other proteins, whose malfunction is linked to many human diseases including neurodegeneration. I gratefully acknowledge the sources of funding that enabled this project: the Engineering and Physical Sciences Research Council, the Department of Chemistry and St John's College at Oxford.

A dear postdoc friend, Elias, said to me before I started, "A PhD is very hard. Be ready." Only now do I comprehend what he meant. I am deeply grateful to people who stood by me, who believed in me when I didn't, and who helped me through the pressures of uncertainty, failure and overwork. I stand stronger than before, with greater confidence, resourcefulness, enthusiasm for science and a vision for the future, because of you.

I thank Justin for his pragmatic and considerate supervision, giving me space to develop independence, and his incredible ability to make everything function smoothly. I thank Andy for hands-on, enthusiastic engagement and teaching me perseverance, focus and efficiency, without which the NMR and chemical kinetics project, our biggest endeavour, would not be complete. Both of you have challenged

me to *learn* from data and literature, given structure to my thinking and a mindset of following through on what I do. I am hugely grateful to you for your patience and your many lessons.

I thank Prof. Dame Carol Robinson for her care, and her wise counsel. I am grateful to my College Advisor Prof. Jason Schnell and the Director of Graduate Studies Prof. Grant Ritchie for their listening ear and pragmatic advice. I thank the College for providing a wonderful learning and living environment, which I will miss dearly. I am extemely grateful to my long-time mentors, Andrei Sergeevich Skitsyuk, Prof. Mark Carrington, Prof. Ernest Laue, Prof. Bill Broadhurst and Prof. Dame Jean Thomas for their continuing support, advice and encouragement. Mark, you will be happy to know I remember to always ask interesting questions.

I thank all Benesch, Baldwin and Robinson group members alongside whom I had the honour to work. Very special thanks go to Georg Hochberg, who invested much time in training me, and with whom I shared many weekends and late nights in the lab. You were always after the biological meaning of the experiments we were doing, which resonated with me, and your ingenuity and dedication is tremendous. I am grateful to Fran Kondrat for her insight and patience in collaborating with me, Weston Struwe and Tim Allison for valuable guidance on crosslinking and CCS projects, Todd Mize for technical expertise, and Oliver Stach for his hard work during his internship. I thank Justin, Andy, Gogs Karunanithy, Reid Alderson, Max Hantke, Iva Pritišanac, Tim Nott, Mark Agasid, and Jason Schnell for careful review of thesis chapters. I am grateful to Mike Barber, Iva Pritišanac, Henrik Müller, Elodie Limer, Ben Girling, Erik Marklund, Matteo Degiacomi, Fran Kondrat, Heidi Gastall, Chérine Bechara, Weston Struwe, Tim Allison, Dale Shepherd, Reid Alderson, Mike Marty, Gogs Karunanithy, David Miguel Dias, Shane Chandler, Miranda Collier, Joe Gault, Jason Yen, Idi Liko, Jon Hopper, Tim Nott, Mark Agasid and Max Hantke for your companionship, much appreciated help of many kinds, and heartfelt conversations, many of which have grown to lifelong friendships.

I am thankful to friends who were an inspiration and joy over these four years and patient listeners of my predicaments: Karrar, who has always met me with open arms, Alex, a brilliant scientist with almost equally brilliant jokes, and Alina, the provider of most enjoyable impromptu Romanian tutorials over lunch. I thank Kai Olaf, who introduced me to the seven habits, Manfred for encouraging me to think big, and Paul for his generosity, sense of humour and colossal composure.

Finally, I thank my dear partner Sergio, who has gone through the ups and downs with me, often literally on the innumerable coach journeys. Your enthusiasm for collaborative science, open-mindedness and love of adventure are just some of the things with which you inspire me. I am immensely grateful to my parents and grandparents: my mother Nataliya, my father Oleg, my grandmothers Mika and Tamara, and my grandfather Anatoliy. Without you, I would never be in Oxford or have the skills and drive to complete this PhD. I adore every one of you. In the loving memory of my grandfather, protector of the family and a passionate scientist - your heart lives on in mine.

# Abbreviations

**$\alpha$BC** $\alpha$B-crystallin

**$\alpha$Lac** $\alpha$-lactalbumin

**1D** one-dimensional

**2D** two-dimensional

**3D** three-dimensional

**ATP** adenosine triphosphate

**ATR** attennuated total reflectance

**CCS** collisional cross-section

**CD** circular dichroism

**CEST** chemical exchange saturation transfer

**CID** collision-induced dissociation

**CPMG** Carr-Purcell-Meiboom-Gill

**CPMG-RD** Carr-Purcell-Meiboom-Gill relaxation dispersion

**CSP** chemical shift perturbation

**c$\alpha$BC** core domain of $\alpha$B-crystallin

**DLS** dynamic light scattering

**DMPC** 1,2-dimyristoyl-sn-glycero-3-phosphocholine

**DTT** dithiothreitol

**EDTA** ethylenediaminetetraacetic acid

**EM** electron microscopy

**ESI** electrospray ionisation

**FID** free induction decay

**FTIR** Fourier transform infra-red

**HSQC** heteronuclear single-quantum correlation

**HX-MS** hydrogen exchange mass spectrometry

**ILV** isoleucine, leucine, valine

**IM** ion mobility

**IMMS** ion mobility mass spectrometry

**ITC** isothermal titration calorimetry

**LC-MS** liquid chromatography mass spectrometry

**LMA** Levenberg-Marquardt algorithm

**MALDI** matrix-assisted laser desorption ionisation

**MS** mass spectrometry

**NADPH** reduced nicotinamide adenine dinucleotide phosphate

**NMR** nuclear magnetic resonance

**SAXS** small-angle X-ray scattering

**SDS-PAGE** sodium dodecyl sulfate polyacrylamide gel electrophoresis

**SEC** size exclusion chromatography

**sHSP** small heat shock protein

**TEV** tobacco etch virus

**ThT** thioflavin T

**ToF** time-of-flight

# Contents


**1 Introduction** **14**
- 1.1 Dynamic self-assembly in biology . . . 14
- 1.2 Protein aggregation . . . 19
- 1.3 Small heat-shock protein chaperones . . . 24
- 1.4 $\alpha$B-crystallin . . . 27
- 1.5 Introduction to methods . . . 29
  - 1.5.1 Nuclear magnetic resonance . . . 30
  - 1.5.2 Multiparameter optimisation . . . 40
  - 1.5.3 Native mass spectrometry . . . 46
  - 1.5.4 Ion mobility . . . 51

**2 Allosteric communication between interfaces in $\alpha$BC** **56**
- 2.1 Introduction . . . 57
- 2.2 Materials and methods . . . 60
- 2.3 Results and discussion . . . 63
  - 2.3.1 Optimisation of cABC $^{15}$N backbone spectra for assignment . 63
  - 2.3.2 Assignment of the c$\alpha$ABC $^{15}$N$^{1}$H NMR spectrum in salt-free buffer . . . 69
  - 2.3.3 Assignment of the c$\alpha$BC $^{15}$N$^{1}$H NMR spectrum in aggregation buffer . . . 72
  - 2.3.4 Allosteric communication between interfaces in $\alpha$BC . . . 74
- 2.4 Conclusions . . . 90

**3 Structure and dynamics of full-length $\alpha$BC** **92**
- 3.1 Introduction . . . 93
- 3.2 Materials and methods . . . 96
- 3.3 Results and discussion . . . 100
  - 3.3.1 Crosslinking the intra-dimer interface . . . 100
  - 3.3.2 Crosslinking the inter-dimer interface . . . 107
  - 3.3.3 CCS measurements of individual stoichiometries of $\alpha$BC using post-IM CID . . . 116
- 3.4 Conclusions . . . 131

**4 Kinetic analysis of $\alpha$Lac aggregation** **133**
- 4.1 Introduction . . . 134
- 4.2 Materials and methods . . . 137
- 4.3 Results and discussion . . . 140
  - 4.3.1 Characterisation of $\alpha$Lac aggregation by light scattering . . . 140

4.3.2 $\alpha$Lac shows features of amyloid aggregation . . . 144
4.3.3 Following aggregation of $\alpha$Lac monomer by NMR . . . 146
4.3.4 Chemical kinetics modelling of the NMR data . . . 151
4.3.5 An aggregation mechanism for $\alpha$Lac . . . 160
4.4 Conclusions . . . 164

**5 Mechanism of $\alpha$BC chaperone action 165**
5.1 Introduction . . . 166
5.2 Materials and methods . . . 168
5.3 Results and discussion . . . 169
5.3.1 Chaperone action followed over time by NMR . . . 169
5.3.2 Chemical kinetics modelling of the nuclear magnetic resonance (NMR) data in the presence of chaperone . . . 174
5.3.3 Chaperone mechanisms for $\alpha$BC and c$\alpha$BC . . . 185
5.3.4 Limitations of the method . . . 192
5.4 Conclusions . . . 194

**6 Characterisation of chaperone-substrate complexes 195**
6.1 Introduction . . . 196
6.2 Materials and methods . . . 198
6.3 Results and discussion . . . 200
6.3.1 Overall sizes and shapes of particles . . . 200
6.3.2 Detection of complexes with low-resolution mass transport methods . . . 206
6.3.3 Characterisation of complexes with native MS . . . 213
6.4 Conclusions . . . 225

**7 Residue-resolution analysis of chaperone function 227**
7.1 Introduction . . . 228
7.2 Materials and methods . . . 230
7.3 Results and discussion . . . 233
7.3.1 The $\beta$4-$\beta$8 groove of $\alpha$BC is not involved in the interaction with the aggregating substrate . . . 233
7.3.2 Greater exposure of the N-terminus implicated in chaperone function by HX-MS . . . 236
7.3.3 Peak position and intensity variation in 2D HSQC spectra . . . 241
7.3.4 Intermediate exchange analysis with $\alpha$BC: CPMG-RD . . . 247
7.3.5 Slow exchange analysis: DEST . . . 253
7.4 Conclusions . . . 255

**8 Conclusions 258**

**References 261**

**Appendix A Supplementary figures and equations 296**

# Chapter 1

# Introduction

## 1.1 Dynamic self-assembly in biology

Self-assembly of molecules is one of the most beautiful phenomena in the natural world (**Figure 1.1**). In biology, it underpins the ability of organisms to build complex structures with emergent properties, respond to stimuli, and propagate themselves in the face of changing environmental conditions.

The research I describe in this thesis deals broadly with intermolecular self-assembly, that is association of two or molecules of the same kind via non-covalent interactions. This can be distinguished firstly from formation of covalent polymers, e.g. synthesis of proteins from amino acids, and secondly from non-covalent intramolecular folding of these polymers, such as folding of a single polypeptide chain. Although I focus on proteins, the concepts are fully applicable to other kinds of macromolecules as well as smaller non-polymeric molecules.

I aim to obtain a quantitative mechanistic understanding of molecular self-assembly, i.e. to elucidate the quantitative physical and chemical relationships between the structure and behaviour of the components on one hand and the functional properties of the assembled system on the other. A foundation for this is the treatment of interactions as dynamic equilibria. Consider the following example of a

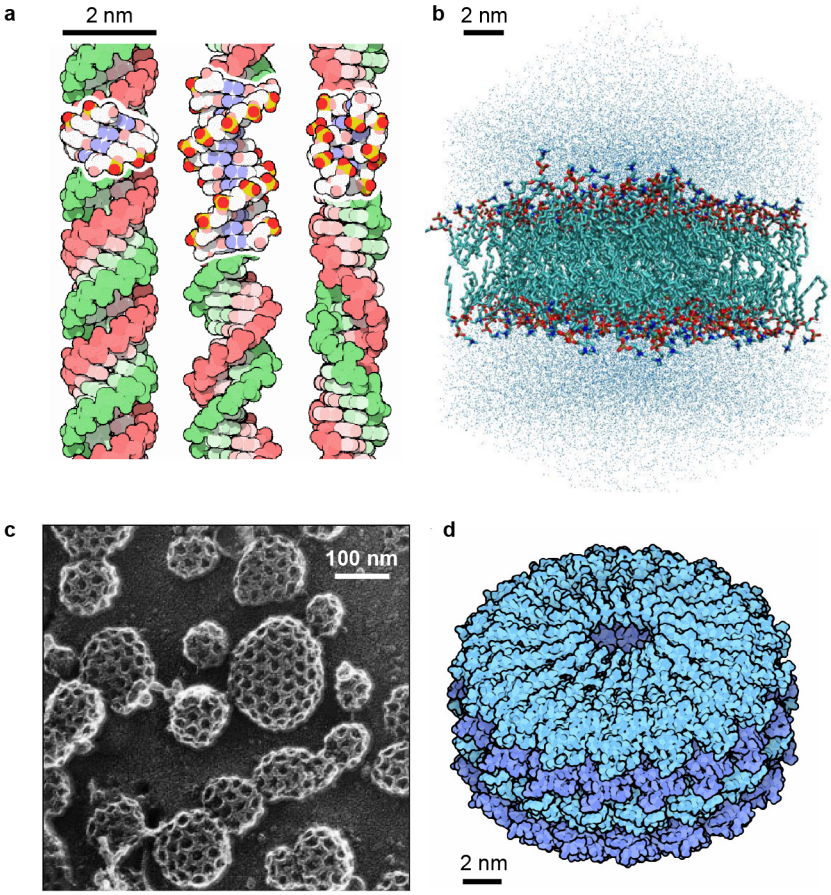



***Figure 1.1: Examples of molecular self-assembly in biology.*** ***a*** *Double-stranded DNA. From left to right, A, B and Z double helix forms (Goodsell [2001]).* ***b*** *The bilayer formed by the amphipathic phospholipid 1,2-dimyristoyl-sn-glycero-3-phosphocholine (DMPC) (Bessonov and Harauz [2010]).* ***c*** *A scanning electron microscopy image of cages formed by the triskelion-shaped protein clathrin (Heuser [1980]).* ***d*** *A coat protein disk of the tobacco mosaic virus, containing four rings of 17 monomer units (Goodsell [2009]).*

mechanism of influenza virus entry into a mammalian cell, which involves clathrin cages depicted in **Figure 1.1c**. It can be described as follows (Grove and Marsh [2011], Lakadamyali et al. [2004]). Initial interaction between haemagglutinin on the surface of the virus and glycoproteins on the surface of the target cell triggers a transmembrane signalling pathway in the cell that involves small GTPase proteins. It leads to recruitment of triskelion-shaped clathrin monomers that assemble into a cage with the help of adaptor proteins, with concomitant deformation of the membrane to form a vesicle inside the cage. Scission of the vesicle from the membrane occurs with involvement of the GTPase dynamin. Following uncoating of the vesicle and trafficking through the endosome pathway, the viral ribonucleoprotein material

is released into the cytosol.

Such a stepwise, deterministic description helps visualise the sequence of events that results in a given outcome. However, it must be remembered that every molecular step - an interaction or a conformational change - is a reversible equilibrium between two or more states. The forward and backward rates, which are influenced by the environmental conditions, define the position of each equilibrium and each steady state. To take the above example, not every interaction of a haemagglutinin molecule with the relevant glycoprotein will result in endocytosis. An individual binding event has a certain on- and off-rate and therefore residence time. Many binding events with a certain threshold residence time may be required for recruitment of a sufficient number of adaptor proteins, which themselves are in an equilibrium with the soluble pool and adaptors engaged elsewhere. Clathrin molecules continuously associate and dissociate; in the vesicle assembly stage the equilibrium is shifted towards association by the presence of a high concentration of adaptor proteins and in uncoating towards dissociation by the presence of uncoating proteins such as the ATPase Hsc70 (Merrifield and Kaksonen [2014]). Thus, in this case and more generally, the macroscopic processes in cells and organisms are based on a network of interconnected dynamic equilibria.

The kinetics and thermodynamics for each reaction are governed by the properties of the assembling units such as shape, charge and hydrophobicity, the concentration of each unit, the environmental conditions such as temperature, pH, presence and concentration of salts, and the resulting balance of interactions between the macromolecule and the solvent. A variety of regulator molecules also commonly influences the assembly and disassembly processes. In terms of free energy changes, there is a loss of translational and rotational entropy of the components upon formation of a complex, and a gain of solvent entropy due to the reduction in macromolecule surface area. On the other hand, the enthalpy change for both the macromolecules and the solvent, and local entropy contributions to the free energy change could be overall either positive or negative depending on the structure

and dynamics of the molecules in question. If the change in free energy is positive, the reaction must be coupled to another one so that the free energy of the system decreases overall for assembly to take place.

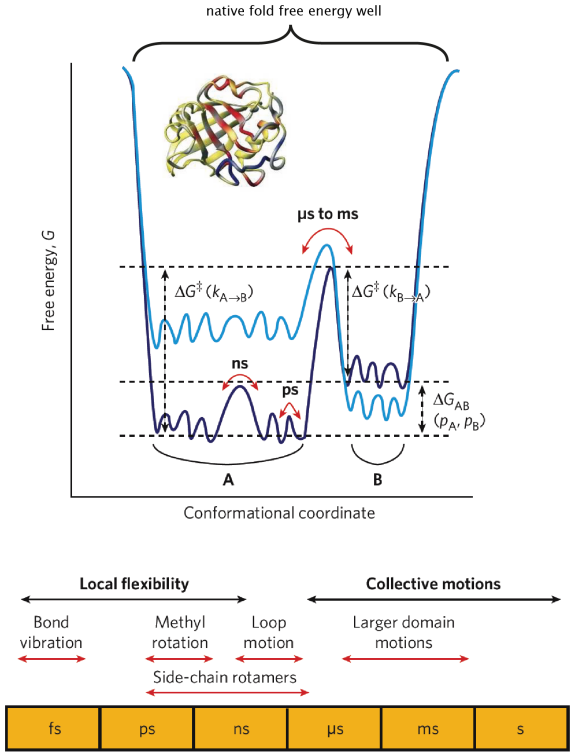



***Figure 1.2: An illustration of the dynamical landscape of a protein.*** *The free energy surface well corresponding to the native fold has fine structure corresponding to motions on various timescales. The free energy barriers, annotated here for the ensembles A and B, define the timescale of the interconversion between the different states.* $\Delta G^{\ddagger}$ *- free energy of activation,* $k_{A\to B}$ *- rate constant for conversion of A to B,* $k_{B\to A}$ *- rate constant for conversion of B to A,* $p_A, p_B$ *- populations in states A and B. Adapted from Henzler-Wildman and Kern [2007]).*

A further running thread in my thesis is the application of the concept of dynamic equilibrium to the individual macromolecular structures that interact to form complexes. Rather than being static building blocks, they are ensembles of structures that undergo internal motions on a range of timescales: for proteins, bond vibration, rotation, sometimes movement of parts of the polypeptide chain (**Figure 1.2**, Henzler-Wildman and Kern [2007]). The geometry of the molecule and the thermal energy available to it defines what movements are accessible. With external supply of energy, e.g. adenosine triphosphate (ATP) hydrolysis, larger movements are possible. For individual protein molecules, many examples have shown that motions can be important for function. One of the most elegant examples is the enzyme dihydrofolate reductase, which binds the substrate dihydrofolate and the cofactor reduced nicotinamide adenine dinucleotide phosphate (NADPH) to make the product tetrahydrofolate (Boehr et al. [2006]). The enzyme goes through several states as part of its catalytic cycle. It has been shown using NMR that in addition

to the ground conformation in each state the enzyme accesses a higher energy conformation that is similar to the ground state conformation of next state in the cycle. The rates of exchange with the higher-energy conformations correlate with the rates of catalysis. Thus the catalytic cycle of the enzyme is encoded in its dynamics.

In multisubunit assemblies, individual components can also undergo dynamics when they are part of a complex. This can be dependent on an input of energy, for example for ATP synthase. It uses the energy from a proton gradient across the membrane to power rotational motion, which is coupled to conformational changes in the ATP synthesis domains (Stock et al. [1999]). Proteins can also undergo fluctuations using only thermal energy. For example, in the archaeal proteasome, the N-termini of the barrel subunits exchange between the inside and outside of the antechamber of the barrel on the timescale of seconds (Religa et al. [2010]). The rate of proteolysis varies with the relative populations of these two conformations. The motion thus appears to play a gating role, regulating access to the catalytic sites of the proteasome. Finally, dynamics in a complex includes binding and unbinding, i.e. quaternary dynamics. In the dodecameric plant chaperone protein Hsp18.1, which is part of the proteostatic stress response system, subunit exchange is regulated by temperature: larger polydisperse oligomers form upon heat shock (Stengel et al. [2010]). The polydisperse forms, but not the dodecamer, are able to bind misfolded client proteins.

In summary, biological systems and self-assembly processes in particular should be considered as ensembles of conjugated equilibria which are sensitive to the surrounding conditions. The molecules that form part of an assembly are themselves dynamic entities undergoing molecular motions on different timescales.

## 1.2 Protein aggregation

A topical example of self-assembly on which I focus in this thesis is protein aggregation, i.e. formation of large ordered (e.g. amyloid) or amorphous structures from partially folded proteins. As I explain below, aggregation of proteins is interesting both from a biological and a medical point of view. Amyloid aggregates are fibres with an underlying $\beta$-sheet structure (**Figure 1.3a-b**); they produce a typical pattern in X-ray diffraction (**Figure 1.3c**) and can also be identified by green birefringence when stained with Congo red (Puchtler et al. [1962]), fluorescence of the dye thioflavin T (ThT) (Vassar and Culling [1959]) and high $\beta$-sheet content measured by circular dichroism (CD) or Fourier transform infra-red (FTIR) spectroscopy. Amorphous aggregates do not have a well-defined structure (e.g. **Figure 1.3d**).

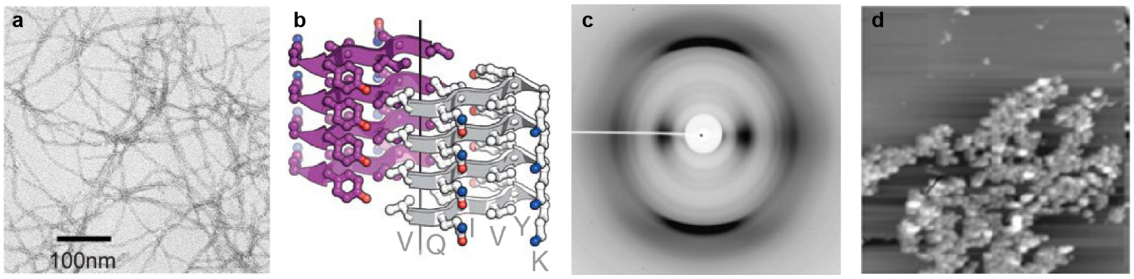



***Figure 1.3: Features of amyloid and amorphous aggregates.*** ***a*** *Amyloid fibres of a tau peptide observed in negative stain electron microscopy (EM) (Sawaya et al. [2007]).* ***b*** *$\beta$-sheet structure of an amyloid fibre formed by the tau peptide VQIVYK, with each strand lying perpendicular to the long fibre axis (Sawaya et al. [2007]).* ***c*** *The typical X-ray fibre diffraction fingerprint of amyloid aggregates (Sikorski et al. [2003]).* ***d*** *Amorphous aggregates of $\alpha$-synuclein visualised by atomic force microscopy (Uversky and Eliezer [2009]).*

**Figure 1.4** shows how different forms of aggregates fit in with other states in which a protein can exist. The ability to form amyloid is thought to be a phenomenon common to most proteins, although the conditions necessary for this vary widely (Chiti et al. [1999], Ramirez-Alvarado et al. [2000], Stefani and Dobson [2003]). Free energies of amyloid formation can be determined, for example, by adding variable amounts of denaturant and measuring the residual amount of free monomer of the protein of interest (Baldwin et al. [2011b]). In this analysis, it was

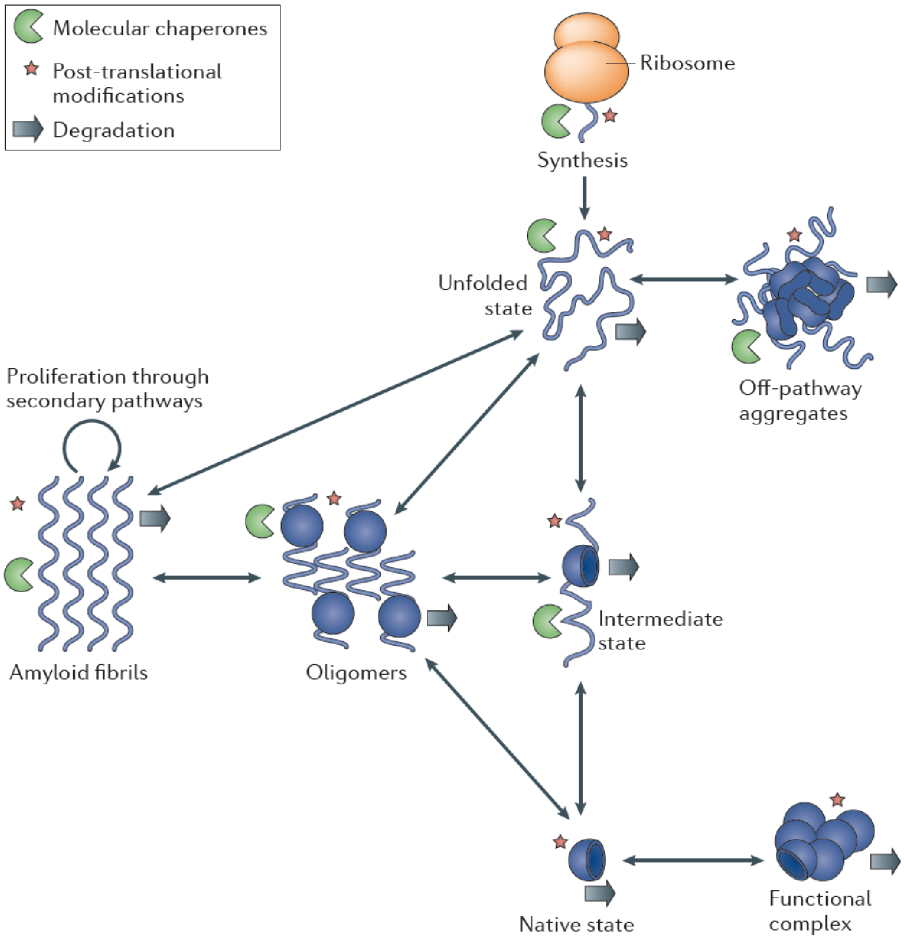



***Figure 1.4: Different states in the life cycle of a protein, from synthesis to degradation.*** *The state the protein tends to occupy depends on its sequence, post-translational modifications and environmental conditions, for example the presence of chaperones. Adapted from Knowles et al. [2014].*

found that, at protein concentrations typically found in cells, amyloid formation from the folded soluble form is thermodynamically favourable. In other words, the native state is only metastable with respect to aggregation and kinetic factors are important for regulating this transition. In accord with this view, proteomic analysis of human cells expressing model aggregating proteins showed that numerous unrelated proteins, especially regulators which are enriched in unstructured regions, can be co-opted into aggregates (Olzscha et al. [2011]).

A variety of organisms produce biologically functional aggregates, for example as biofilm components or virulence agents (DePas and Chapman [2012], Wickner et al. [2007]). However, most commonly, aggregate formation is a result of stress-

related disruption of the cellular proteostasis system, which is a complex network of molecular machinery responsible for correct synthesis, folding, and degradation of proteins. In humans, aggregation of specific proteins is associated with toxicity and cell death and implicated in a range of genetic neurodegenerative conditions such as Huntington's disease and spinocerebellar ataxia (Ross [2002]) and with debilitating non-neuronal disorders such as eye cataracts (Moreau and King [2012]) and type II diabetes (Hull et al. [2004]). Moreover, aggregation occurs in age-related degenerative disorders such as Alzheimer's and Parkinson's disease (Eisele et al. [2015]). With the ever increasing life expectancy and concomitant steep population ageing across the world, finding treatment for these diseases is becoming an increasingly acute problem. One of the biggest challenges of modern biomedical research is to understand how and why pathological protein aggregation happens, why it is toxic and how it can be prevented.

From a mechanistic point of view, aggregation proceeds through various steps involving different stoichiometries of the protein in question. Some of the mechanisms proposed for different proteins so far are shown in **Figure 1.5**. Sometimes a conformational rearrangement is needed to produce the aggregating form of the protein. This form associates to produce dimers, trimers and so on, in a process which is termed elongation. As with other assembly processes, each of these reactions is a reversible equilibrium with the outcome determined by the concentrations of every species involved and the solution conditions. Formation of a particular species may be rate-limiting. This species is then called the nucleus and its formation is termed primary nucleation. Mature aggregates can promote formation of nuclei on their surface, a process termed secondary nucleation, and provide reactive ends by fragmentation. Addition of these aggregates to soluble monomer, a phenomenon described as seeding, can therefore greatly speed up the aggregation process.

Which of the species in the aggregation pathway are the most harmful in disease is a topic of active investigation. In particular, it seems that pre-fibrillar oligomeric intermediates may be more toxic than the mature aggregates (e.g. Walsh

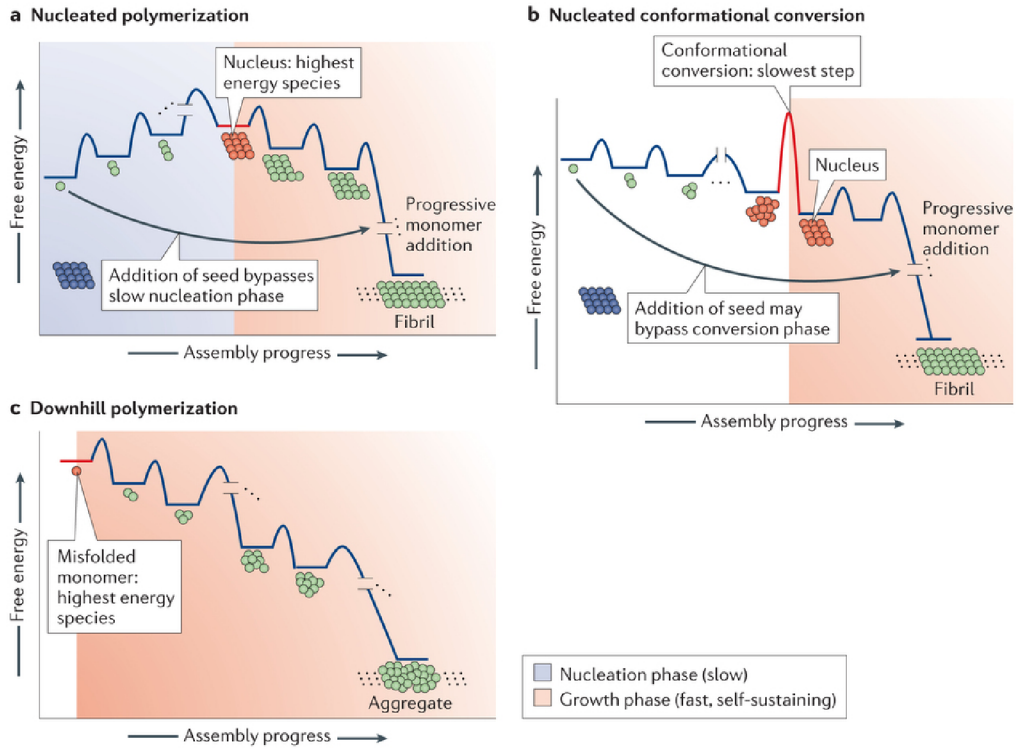



***Figure 1.5: Free energy diagrams for different mechanisms of protein aggregation.*** ***a*** *Commonly, association of smaller species to form the nucleus is rate-limiting (e.g. Oosawa and Kasai [1962], Jarrett and Lansbury [1992], Meisl et al. [2014].* ***b*** *Conformational conversion to form the nucleus is rate-limiting. This is the case for the yeast prion protein Sup35 (Serio et al. [2000]).* ***c*** *The formation or availability of the misfolded monomer is rate-limiting. For example, for transthyretin the dissociation of the native tetramer is rate-limiting for aggregation (Hammarstrom [2001]). Addition of seeds can bypass the rate-limiting nucleation requirement and speed up aggregation (Jarrett and Lansbury [1992], Lundmark et al. [2002], Knowles et al. [2009]). Adapted from Eisele et al. [2015].*

et al. [2002], Baglioni et al. [2006], Winner et al. [2011], Cremades et al. [2012]). The reasons suggested for this are the greater structural plasticity of oligomers compared to fibrils and greater exposure of residues that would normally be buried, either in the core of the soluble species or in the fibre, leading to spurious interactions with other macromolecules. In the study by Campioni et al. [2010], oligomers of the same protein formed under different solution conditions showed different cytotoxicity, with the more toxic oligomers being less tightly packed and more hydrophobic.

A variety of ideas for therapeutic intervention has been put forward to combat aggregation-related diseases (Knowles et al. [2014], Eisele et al. [2015]). Among the first strategies attempted is protein reduction, for example by modulating the

activity of secretase enzymes that produce the aggregation-prone form of A$\beta$, the peptide implicated in Alzheimer's disease (Yan and Vassar [2014]), or by gene silencing techniques such as RNA interference for transthyretin-related amyloidosis (Benson et al. [2010], Coelho et al. [2013]). Other approaches include protein stabilisation with small molecules (Coelho et al. [2012], Berk et al. [2013]), amyloid remodelling by polyphenols (Lorenzen et al. [2014]), or amyloid removal by development of catalytic antibodies with aggregate-degrading activity (Nishiyama et al. [2014]). Exploiting the protein quality-control network is another strategy, for example by overall small molecule-mediated enhancement of the heat shock response (e.g. Zhang and Sarge [2007]) or by targeting specific components of the network such as particular chaperones. For example, the chaperone Hsp70 reduces the toxicity of $\alpha$-synuclein *in vivo* and *in vitro* (Klucken et al. [2004]). The human extracellular chaperone clusterin has been shown to bind oligomeric forms of the A$\beta$ peptide and slow down its aggregation (Narayan et al. [2012]). In this thesis, I focus on small heat shock proteins, an important but less well studied family of chaperones implicated in a range of diseases including neurodegeneration. In addition to their potential as a target for the treatment of these diseases, their properties of ATP independence and polydispersity make them especially interesting from the point of view of mechanistic study. They are described in the next Section.

The complexity of the aggregation process and our limited understanding of the mechanistic links to pathogenicity make the development of treatments for aggregation-related diseases a challenge. The interlinked nature of the reactions involved mean that the effect of an intervention, such as a small molecule or a chaperone, may be more wide-ranging than expected and the resulting attenuation of pathogenicity is non-trivial to predict. The efforts to glean thorough, quantitative knowledge of aggregation mechanisms and the ability to predict their modulation, together with validation of these findings *in vivo*, present exciting opportunities to develop effective treatments for these complex diseases, and deepen our understanding of the functioning of the proteostasis network in cells and organisms.

## 1.3 Small heat-shock protein chaperones

The proteostasis machinery is responsible for promoting efficient folding of proteins and preventing their misfolding and pathological aggregation. It comprises several interlinked systems, such as the heat shock response, the ubiquitin-proteasome protein degradation system, autophagy and organellar unfolded protein responses. A central part of the proteostasis system are different families of molecular chaperones. A molecular chaperone can be defined as a protein that interacts with another protein to help it acquire its functional form or stabilise it without being present in its final structure (Hartl et al. [2011]). Chaperones can reduce or prevent aggregation of other proteins, in some cases resolubilise them from aggregates and refold them back into their functional form, or target them for degradation. Some of the major systems of eukaryotic chaperones are HSP70, HSP90 and HSP60.

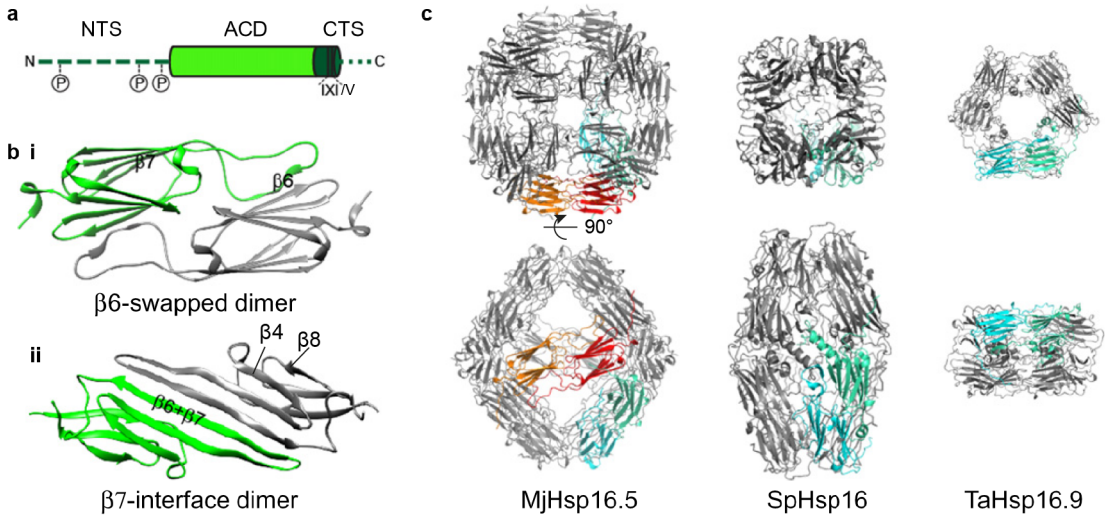



***Figure 1.6: The structure of small heat shock proteins (sHSPs).*** ***a*** *Schematic illustrating the alpha crystallin domain (ACD), the N-terminal and C-terminal sequences (NTS and CTS), the conserved IXI/V motif in the C-terminus and phosphorylation sites in the N-terminus (P).* ***b*** *The different types of intra-dimer interface characteristic of* ***(i)*** *sHSPs in bacteria, archaea and plants and* ***(ii)*** *of metazoa.* ***c*** *Crystal structures of some monodisperse sHSPs: Hsp16.5 from Methanocaldococcus jannaschii, Hsp16 from Schizosaccharomyces pombe and Hsp16.9 from Triticum aestivum. Adapted from Haslbeck and Vierling [2015].*

In my thesis I focus on the small heat-shock proteins (sHSPs). sHSPs are non-ATP-dependent promiscuous chaperones defined by the presence of the conserved $\alpha$-

crystallin domain flanked by flexible N- and C-terminal tails (**Figure 1.6a**, Haslbeck and Vierling [2015], Malmström et al. [2010]). They are present in all three domains of life: bacteria, archaea and eukaryotes. An sHSP gene in cyanobacterial phage has also been reported (Bourrelle-Langlois et al. [2016]). They are the least conserved chaperones, with the tails being especially variable (de Jong et al. [1998], Kriehuber et al. [2010]). Bacteria and archaea often have one or two sHSP genes, whereas metazoa and especially plants have considerably more; humans have 10 sHSP genes (Kappé et al. [2003]) and *Arabidopsis thaliana* has 24 (Siddique et al. [2008]). As many as 117 genes were recently identified in polyploid bread wheat (Wang et al. [2017]).

The $\alpha$-crystallin domain is the defining feature of sHSPs (Haslbeck and Vierling [2015]). It is about 90 residues long and forms a $\beta$-sandwich with 6 or 7 strands arranged as two antiparallel sheets. The domain commonly forms dimers, which exist in two forms (**Figure 1.6b**): in plants, archaea and bacteria one of the strands ($\beta$6) swaps out to the neighbouring monomer, whereas in animals the interface is formed by the extended $\beta$6+7 strand from each monomer.

A distinguishing property of sHSPs is the ability to form oligomers of varying size. Some are monodisperse, such as the *Methanocaldococcus jannaschii* Hsp16.5 24-mer (Kim et al. [1998]), and others are polydisperse, such as human $\alpha$B-crystallin (Aquilina et al. [2003]). The dimer is a common substructure of the oligomers, with the N- and C-termini also important in intersubunit contacts. The structures determined so far for monodisperse sHSPs have polyhedral architectures (**Figure 1.6c**), with the dimers being a common substructure but the N- and C-termini also important for intersubunit contacts. In particular, the conserved IXI/V motif in the C-terminus (**Figure 1.6a**), where X is any amino acid, can interact with the groove formed by strands $\beta$4 and $\beta$8 on neighbouring monomers. The quaternary structure of sHSPs is dynamic, i.e. subunits exchange in and out of a given oligomer at measurable rates (Bova et al. [2002], Sobott et al. [2002]). In some organisms, e.g. bacteria and plants, sHSPs can be divided into evolutionary classes. sHSPs can

usually hetero-oligomerise with other sHSPs of the same or related species, but only if they are members of the same class (Studer and Narberhaus [2000], Bova et al. [2002], Bloemendal et al. [2004], Aquilina et al. [2013]).

sHSPs are mostly cytosolic proteins, with some able to translocate to the nucleus (van Montfort et al. [2002]). Plants express both cytosolic and various organelle-specific sHSPs. Some sHSPs are developmentally regulated, whereas others are present constitutively in different cell types under normal conditions. sHSPs are often upregulated under stress (Mansfield and Key [1987], Klemenz et al. [1991]) and can bind a variety of unfolding proteins at an early stage, slowing down their aggregation, as well as mature aggregates. While they cannot rescue already aggregated proteins, the bound substrate can be released and refolded *in vitro* in the presence of ATP-dependent chaperones (e.g. Lee and Vierling [2000]). It is still unclear how sHSPs bind their misfolded substrates. Crosslinking and other experiments have implicated the N-terminus (Ahrman et al. [2007], Aquilina and Watt [2007]), whereas others indicate that the N-termini remain dynamic in the presence of substrate (Cheng et al. [2008]). Studies with chimeric, mutant and truncated proteins have shown that for different substrates the $\alpha$-crystallin domain and the C-terminus can also be important (Basha et al. [2006], Treweek et al. [2007], Mainz et al. [2015]). The isolated $\alpha$-crystallin domains or even smaller fragments show significant chaperone activity (Feil et al. [2001], Bhattacharyya et al. [2006], Hochberg et al. [2014]). The quaternary structure and subunit exchange may play a role too. The subunit exchange rates and oligomeric distribution of a sHSP can vary depending on the environmental conditions such as pH or temperature and this can affect their chaperone activity. For example, the *Escherichia coli* sHSP IbpB oligomers dissociate to smaller species including dimers at heat shock temperatures and are concomitantly much better able to inhibit aggregation of the model substrate insulin (Jiao et al. [2005]). In contrast, pea HSP18.1 forms both large and small oligomers at heat shock temperature and it is the larger oligomers that bind the model substrate luciferase (Stengel et al. [2010]). The activity of sHSPs can also

be regulated by post-translational modifications, such as phosphorylation of the N-terminus in eukaryotes (e.g. Ecroyd et al. [2007]), which could act indirectly by affecting oligomeric size and subunit exchange rate or by directly affecting substrate binding.

In summary, sHSPs are a widespread family of stress response proteins which exist mostly as dynamic oligomers. The interrelation of sHSP structure and activity is still far from clear. This field therefore presents exciting opportunities to resolve some of the long-standing questions with the application of cutting-edge high resolution biophysical methods, which I explore in this thesis.

## 1.4 $\alpha$B-crystallin

$\alpha$B-crystallin ($\alpha$BC) is one of the most abundant human sHSPs and perhaps the most well studied, with interesting aspects of its structure and function subject to active research and discussion, as I detail below and in the individual Chapters. Despite the attention it has received, no consensus has been achieved on its structure, the mechanism of its chaperone function and the role of its various levels of dynamics. Therefore I chose to focus on this protein as my main target of study, addressing structural, dynamic and functional questions regarding $\alpha$BC alone and in combination with the model aggregating substrate $\alpha$-lactalbumin ($\alpha$Lac).

$\alpha$BC was discovered as a major component of the eye lens. No protein turnover occurs in differentiated lens cells, and total protein concentration reaches 450 mg/ml to give the required refractive index (Fagerholm et al. [1981]). The chaperone function of $\alpha$BC and other crystallins is important in preventing aggregation of lens proteins and maintaining lens transparency under these remarkable conditions (Horwitz [2003]). $\alpha$BC has since been found to be expressed in many tissues of the human body with particularly high levels in the muscle and the brain in addition to the eye lens (Iwaki et al. [1990], Lowe et al. [1992]) and to be stress-inducible (Kle-

menz et al. [1991]). Accordingly, its defects have been associated with cataracts, cardiomyopathies and neurodegenerative disease *in vivo* (Renkawek et al. [1994], Boncoraglio et al. [2012], Boelens [2014]) and shown to reduce aggregation of many substrates including A$\beta$ and $\alpha$-synuclein *in vitro* (Arrigo [2013]).

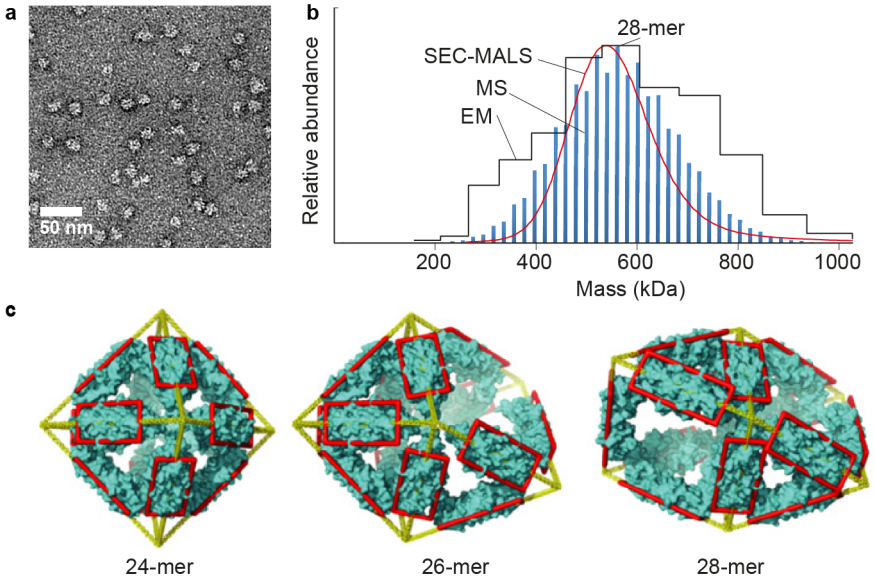



***Figure 1.7:*** ***$\alpha$B-crystallin is a polydisperse sHSP.*** ***a*** *Negative stain EM image of $\alpha$BC.* ***b*** *The oligomeric distribution of $\alpha$BC derived from different experimental data. SEC-MALS - size exclusion chromatography and multiangle light scattering, MS - mass spectrometry, EM - electron microscopy. Adapted from Hochberg and Benesch [2014]).* ***c*** *Suggested structural models for some of the $\alpha$BC oligomers. The cyan blocks represent crystal structures of $\alpha$BC core domain dimers. Red rectangles are placed on the cyan structures to help visualise their relative position and the overall geometry of the oligomers is schematically shown by yellow frames. Adapted from Baldwin et al. [2011c].*

$\alpha$BC appears as roughly spherical particles of a range of sizes, with an average radius of 9 nm in negative stain EM (**Figure 1.7a**, see also **Chapter 6**). Native mass spectrometry shows that it is polydisperse with oligomer size ranging between 10 and 40 and mean stoichiometry of about 28 (Aquilina et al. [2003], Baldwin et al. [2011d]), which is consistent with the EM data (**Figure 1.7b**). Both odd and even oligomers exist, although there is a preference for the even stoichiometries at neutral pH. Interestingly, the distribution is invariant with concentration over three orders of magnitude (Horwitz [2009]). This is consistent with a simple oligomerisation model where the microscopic association and dissociation rates of a monomer are independent of oligomer size (Baldwin et al. [2011d]), but the macroscopic dissociation rate

is linearly dependent on the stoichiometry. Monomer association is energetically favourable but at larger sizes is offset by the higher dissociation rate. This gives rise to a peaked distribution as in **Figure 1.7b**. Polydispersity has hampered the elucidation of structures for $\alpha$BC oligomers. The various proposed models are discussed in **Chapter 3**; briefly, $\alpha$BC oligomers are thought to most likely be polyhedral complexes with dimeric substructure, similar to the other sHSPs whose structure is known (see **Figure 1.6c**). The oligomers exchange subunits on the timescale of minutes in physiological conditions (Peschek et al. [2013], Baldwin et al. [2011d]) and can hetero-oligomerise with $\alpha$A-crystallin and Hsp27. Similarly to other members of the sHSP family, it is still enigmatic which structural and dynamic determinants of $\alpha$BC allow it to perform its functions, in particular as a chaperone, towards its many substrates.

Manipulation of the expression levels or function of $\alpha$BC has been proposed as a therapeutic strategy for various diseases where its chaperone function plays a central role (Reddy and Reddy [2015]). However, despite the biological and medical importance of sHSPs and $\alpha$BC in particular, much remains to be learnt about the basis of its structural plasticity and mechanism of action as a chaperone. $\alpha$BC thus presents an interesting object of study for me both from the point of view of biological self-assembly and of its functional interaction with substrate and partner proteins.

## 1.5 Introduction to methods

This section gives an introduction to the four main methods I employed in this thesis: nuclear magnetic resonance, computational optimisation, native mass spectrometry and ion mobility.

### 1.5.1 Nuclear magnetic resonance

NMR is one of the most information-rich techniques in biophysics, giving atomic-resolution insight into structure and dynamics of molecules. It relies on nuclear spin - a property of some isotopes including $^{1}$H, $^{2}$H, $^{13}$C, $^{15}$N, $^{31}$P and $^{19}$F. Nuclear spin is described by two quantum numbers: the nuclear spin quantum number $I$ and the magnetic quantum number $m$. A nucleus has $2I + 1$ non-degenerate energy levels in a magnetic field, each characterised by a unique value of $m$ which varies from $-I$ to $I$ in steps of 1. Thus a nucleus with $I = \frac{1}{2}$, such as a proton, has two energy levels with $m = +\frac{1}{2}$ and $m = -\frac{1}{2}$. They are commonly denoted as $\alpha$ and $\beta$. Here I present the basic quantum mechanical principles of the NMR experiment and give an overview of key phenomena that are important for the work presented in this thesis.

The behaviour of nuclear spin magnetisation can be understood by looking at how its quantum mechanical wavefunctions behave under the action of operators. Any wavefunction can be described as a linear combination of basis operators, in Dirac's bra-ket notation:

$$\Psi = \sum_{n} c_n \left|n\right\rangle \tag{1.1}$$

The basis functions $n$ could be the $x, y, z$ rotation operators $(\hat{I}_x, \hat{I}_y, \hat{I}_z, \hat{E})$, for example, and $c_n$ are the complex, time-dependent amplitude coefficients for the basis state $n$. For the isolated proton, a convenient choice of basis functions are the states $\alpha$ and $\beta$, so that we can write:

$$\Psi = c_\alpha \left|\alpha\right\rangle + c_\beta \left|\beta\right\rangle \tag{1.2}$$

In vector notation this wavefunction can be represented as:

$$\Psi = \begin{pmatrix} c_\alpha \\ c_\beta \end{pmatrix} \tag{1.3}$$

An operator acts on a wavefunction to produce another wavefunction. For example, $\hat{I}_z$ scales the functions $\alpha$ and $\beta$, while $\hat{I}_y$ and $\hat{I}_x$ interconvert them:

$$\hat{I}_z |\alpha\rangle = \frac{1}{2} |\alpha\rangle \qquad \hat{I}_x |\alpha\rangle = \frac{1}{2} |\beta\rangle \qquad \hat{I}_y |\alpha\rangle = \frac{1}{2} i |\beta\rangle$$
$$\hat{I}_z |\beta\rangle = -\frac{1}{2} |\beta\rangle \qquad \hat{I}_x |\beta\rangle = \frac{1}{2} |\alpha\rangle \qquad \hat{I}_y |\beta\rangle = \frac{1}{2} i |\alpha\rangle$$

An operator can be written in matrix representation as follows, using $\hat{I}_z$ as an example:

$$\hat{I}_z = \begin{pmatrix} \langle\alpha| \hat{I}_z |\alpha\rangle & \langle\alpha| \hat{I}_z |\beta\rangle \\ \langle\beta| \hat{I}_z |\alpha\rangle & \langle\beta| \hat{I}_z |\beta\rangle \end{pmatrix} = \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & -\frac{1}{2} \end{pmatrix} \tag{1.4}$$

Since the basis vectors are orthonormal, $\langle\alpha|\alpha\rangle = \langle\beta|\beta\rangle = 1$ and $\langle\alpha|\beta\rangle = \langle\beta|\alpha\rangle = 0$.

The key operator to consider in NMR is the Hamiltonian $\hat{H}$, which describes the total energy of a given system. For example, for spins in a static magnetic field along the z-axis, $\hat{H} = -\gamma B_0 \hbar \hat{I}_z = \omega_0 \hbar \hat{I}_z$, where $\gamma$ is the gyromagnetic ratio of the nucleus in question, $B_0$ is the strength of the magnetic field and $I_z$ is the unit operator in the direction of the field. $\omega_0$ is the corresponding rotational frequency in radians per second, termed the Larmor frequency. Working in atomic units, such that $\hbar = 1$, $\hat{H} = \omega_0 \hat{I}_z$. Another example of a Hamiltonian is that of a radiofrequency pulse, which can be used to manipulate magnetisation. In a convenient reference frame that rotates around the z-axis at the Larmor frequency, it is $\hat{H} = \frac{\pi}{2} \hat{I}_y$ for a hard 90° pulse along the y-axis. The evolution of a wavefunction over time under a given Hamiltonian is described by the time-dependent Schrödinger equation (with $\hbar = 1$):

$$i \frac{\partial}{\partial t} \Psi = \hat{H} \Psi \tag{1.5}$$

To understand what happens in an NMR experiment we need to solve this equation, i.e. find the form of $\Psi(t)$ under a given Hamiltonian.

A useful concept that simplifies the analysis is the so-called density matrix,

which has the form

$$\rho(t) = |\Psi(t)\rangle \langle\Psi(t)| \tag{1.6}$$

$\rho(t)$ is a square matrix, with an element for each combination of basis states. For example, for the spin-half wavefunction in **Equation 1.3**, the density matrix is:

$$\rho(t) = \begin{pmatrix} c_\alpha \\ c_\beta \end{pmatrix} \otimes \begin{pmatrix} c_\alpha^* & c_\beta^* \end{pmatrix} = \begin{pmatrix} c_\alpha c_\alpha^* & c_\alpha c_\beta^* \\ c_\beta c_\alpha^* & c_\beta c_\beta^* \end{pmatrix} \tag{1.7}$$

The diagonal elements give the probability of occupying the corresponding state, here $\alpha$ or $\beta$. The off-diagonal elements are complex and time-dependent, and describe evolution of superpositions of states. Thus the density matrix is a comprehensive description of the spin system at a given time $t$. For a macroscopic system of many spins, the density matrix is an ensemble average of the microscopic density operators: $\rho_{nm} = \overline{c_m^* c_n}$.

To understand how the spin system, i.e. the density matrix, will evolve over time, we differentiate the density operator as defined in **Equation 1.6** and substitute in expressions from the time-dependent Schrödinger equation (**Equation 1.5**) and its complex conjugate:

$$\frac{\partial}{\partial t} |\Psi\rangle = -i\hat{H} |\Psi\rangle \quad \text{and} \quad \frac{\partial}{\partial t} \langle\Psi| = i\hat{H} \langle\Psi| \tag{1.8}$$

Thus we obtain the Liouville-von Neumann equation:

$$\frac{\partial \rho}{\partial t} = -i(\hat{H}\rho - \rho\hat{H}) = -i[\hat{H}, \rho] \tag{1.9}$$

where the term in square brackets is the commutator of the two operators. This is the master equation, which describes the evolution of the density matrix under a given Hamiltonian.

In NMR, the Hamiltonian is usually time-independent for each section of the

experiment. In these cases the solutions to the Liouville-von Neumann equation are of the following form:

$$\rho(t) = e^{-i\hat{H}t}\rho(0)e^{i\hat{H}t} \tag{1.10}$$

This can be verified by differentiation to obtain the Liouville-von Neumann equation again. Thus, knowing the state of the system at time 0, $\rho(0)$, we can construct the system at time $t$, $\rho(t)$. This is a very important result. If multiple Hamiltonians are applied in a sequence, the solution can be easily extended:

$$\rho(t) = e^{-i\hat{H}_2t}e^{-i\hat{H}_1t}\rho(0)e^{i\hat{H}_1t}e^{i\hat{H}_2t} \tag{1.11}$$

The sets of exponentials left and right of $\rho(0)$ can be simplified before $\rho(t)$ is evaluated. This analysis avoids laborious computation of the effects of individual Hamiltonians.

The solution can be represented in another form, which helps to provide a more visual picture. Let us express the Hamiltonian in the form $\hat{H} = c\hat{O}$, where $c$ is a constant and $\hat{O}$ is the operator itself. For the static field Hamiltonian $\hat{H} = \omega_0\hat{I}_z$, $c = \omega_0$ and $\hat{O} = \hat{I}_z$. A further useful property is that the commutators of the Cartesian operators for spin-$\frac{1}{2}$, $\hat{I}_x$, $\hat{I}_y$ and $\hat{I}_z$, are tricyclic:

$$[\hat{I}_x, \hat{I}_y] = i\hat{I}_z \qquad [\hat{I}_y, \hat{I}_z] = i\hat{I}_x \qquad [\hat{I}_z, \hat{I}_x] = i\hat{I}_y \tag{1.12}$$

Using the above and taking the second derivative of $\rho(t)$ in **Equation 1.10** with respect to $t$ shows that it obeys the wave equation:

$$\rho(t) = -\frac{\partial^2\rho}{\partial t^2} = Acos(ct) + Bsin(ct) \tag{1.13}$$

The boundary conditions allow us to identify the factors $A$ and $B$:

$$A = \rho(0) \qquad B = \frac{\partial\rho}{\partial t}(0) = [\hat{O}, \rho(0)] \tag{1.14}$$

Therefore the full solution is:

$$\rho(t) = \rho(0)cos(ct) - [\hat{O}, \rho(0)]isin(ct) \tag{1.15}$$

Applying this to initial magnetisation of the form e.g. $\hat{I}_x$, with the static field Hamiltonian, we obtain:

$$\rho(t) = \hat{I}_x cos(\omega_0 t) - [\hat{I}_z, \hat{I}_x]isin(\omega_0 t) \tag{1.16}$$

$$= \hat{I}_x cos(\omega_0 t) + \hat{I}_y sin(\omega_0 t) \tag{1.17}$$

**Equation 1.17** says that transverse magnetisation in a static magnetic field will rotate in the x-y plane with the frequency $\omega_0$ (see **Figure 1.8a**). This frequency will be around $\gamma B_0$ but will deviate slightly from this value since the local field is modulated by e.g. the electrons surrounding the nucleus, and is referred to as the chemical shift frequency of that nucleus.

Thus, using the principles of quantum mechanics, we have arrived at an intuitive vector model description of NMR. Unlike the vector model, however, the density matrix approach is general and can be used for simulating spectra and incorporating more complex phenomena such as relaxation, coupling and chemical exchange in a robust way, as explained below.

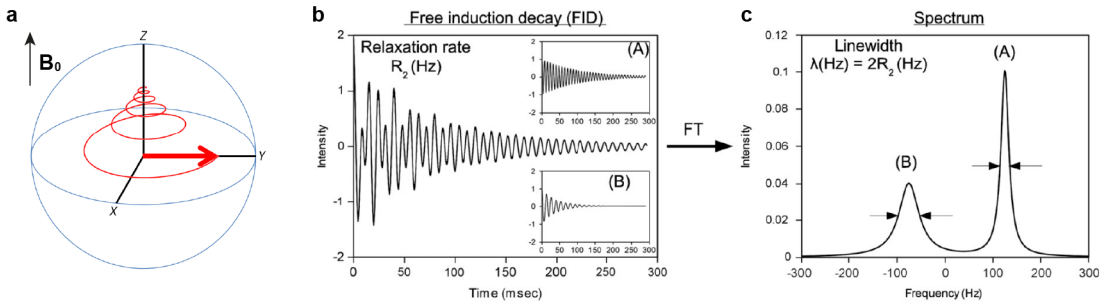



***Figure 1.8: The NMR signal.*** ***a*** *An illustration of the precession of magnetisation, modulated by relaxation towards the equilibrium value along the z-axis.* ***b*** *The free induction decay with two underlying frequencies (A) and (B).* ***c*** *The corresponding Fourier transform spectrum. Adapted from Kleckner et al. [2012].*

**Relaxation**

Relaxation is the return of the magnetisation to its equilibrium state. This happens via transitions between states caused by random local magnetic fields of the appropriate frequency. The local magnetic field is influenced by phenomena such as dipolar coupling and chemical shift, which can be anisotropic, i.e. dependent on the orientation of the molecule in the external magnetic field. These interactions are modulated by molecular tumbling in solution, leading to random fluctuations in the local field. Unlike transitions caused by pulses, where each spin experiences the same magnetic field, in relaxation the variation in local magnetic field is stochastic. The net effect is modulation of the Hamiltonian by a relaxation constant. Different magnetisation terms have different relaxation properties depending on frequencies that are relevant to them. The relaxation constant of transverse terms of the form $\hat{I}_x$ and $\hat{I}_y$ is denoted $R_2$, and of longitudinal terms such as $\hat{I}_z$ as $R_1$. Relaxation can be incorporated into the density operator approach as follows. Using the Liouville-von Neumann equation and the commutator relations in **Equation 1.12**, we can write out the derivatives of the single spin Cartesian basis operators under the free precession Hamiltonian, $\hat{H} = \omega_0 \hat{I}_z$, in vector form:

$$\frac{d}{dt}\begin{pmatrix}\hat{I}_x\\ \hat{I}_y\\ \hat{I}_z\end{pmatrix} = \begin{pmatrix}0 & \omega_0 & 0\\ -\omega_0 & 0 & 0\\ 0 & 0 & 0\end{pmatrix}\begin{pmatrix}\hat{I}_x\\ \hat{I}_y\\ \hat{I}_z\end{pmatrix} \tag{1.18}$$

These are known as the Bloch equations. Including transverse and longitudinal relaxation, we can write:

$$\frac{d}{dt}\begin{pmatrix}\hat{I}_x\\ \hat{I}_y\\ \hat{I}_z\\ \hat{E}\end{pmatrix} = \begin{pmatrix}-R_2 & \omega_0 & 0 & 0\\ -\omega_0 & -R_2 & 0 & 0\\ 0 & 0 & -R_1 & M_0 R_1\\ 0 & 0 & 0 & 0\end{pmatrix}\begin{pmatrix}\hat{I}_x\\ \hat{I}_y\\ \hat{I}_z\\ \hat{E}\end{pmatrix} \tag{1.19}$$

where $M_0$ is the equilibrium magnetisation. **Figure 1.8a** illustrates the precession of initially transverse magnetisation modulated by relaxation. The transverse component decays to zero while the longitudinal component returns to its equilibrium value $M_0$.

Signal detection has its own Hamiltonian. Put simply, some (but not all) transverse terms can be detected. Thus, by analogy to the form of the result in **Equation 1.17**, the signal looks like a sinusoid wave damped by the exponential transverse relaxation term $e^{-R_2 t}$ - this is the free induction decay (**Figure 1.8b**). For an ensemble, it is a superposition of signals from each nucleus. Fourier transformation of this signal allows deconvolution of the different frequencies to produce the NMR spectrum (**Figure 1.8c**). The nucleus-specific relaxation constants determine the width of each peak - the faster the relaxation, the broader the peak.

### Scalar coupling

Scalar coupling allows transfer of magnetisation between nuclei, either of the same element or of different elements. It is key in heteronuclear NMR, e.g. for the backbone assignment experiments outlined in the next Chapter. The scalar coupling Hamiltonian between two nuclei, e.g. $^1$H and $^{15}$N, is of the form $\hat{H} = 2\pi J \hat{I}_z \hat{S}_z$, where $J$ is the coupling constant and $\hat{I}_z \hat{S}_z$ is the Kronecker product of the $\hat{I}_z$ and $\hat{S}_z$ operators. While the matrix representation of $\hat{I}_z$ and $\hat{S}_z$ is 2x2 for a single spin (see **Equation 1.4**), $\hat{I}_z \hat{S}_z$ is a 4x4 matrix, reflecting the various combinations of $\alpha$ and $\beta$ states for the two spins.

### Chemical exchange

Chemical exchange refers to interconversion between different chemical or physical states of the molecule. The states could be intra- or intermolecular, for example a conformational change or binding to an interaction partner. Its effect on the position and intensity of peaks depends on the relative frequency of the exchange process and

on the difference in chemical shifts of the two states. In addition, the effect depends on the relaxation properties of the states involved. For two states, termed ground ($G$) and excited ($E$), we can represent this process using the Bloch equations, for simplicity only considering the $\hat{I}_x$ and $\hat{I}_y$ terms:

$$\frac{d}{dt}\begin{pmatrix}\hat{I}_x^G\\ \hat{I}_y^G\\ \hat{I}_x^E\\ \hat{I}_y^E\end{pmatrix} = \begin{pmatrix}-R_2^G - k^{GE} & \omega_0^G & k^{EG} & 0\\ -\omega_0^G & -R_2^G - k^{GE} & 0 & k^{EG}\\ k^{GE} & 0 & -R_2^E - k^{EG} & \omega_0^E\\ 0 & k^{GE} & -\omega_0^E & -R_2^E - k^{EG}\end{pmatrix}\begin{pmatrix}\hat{I}_x^G\\ \hat{I}_y^G\\ \hat{I}_x^E\\ \hat{I}_y^E\end{pmatrix} \tag{1.20}$$

where $k^{GE}$ and $k^{EG}$ represent exchange rates from ground to excited state and vice versa.

Different exchange regimes and their manifestations in the spectrum are illustrated in **Figure 1.9**. It shows exchange trajectories between states A and B for several spins in the experiment. If the timescale of exchange is much slower than the recording time, each spin is observed either at the frequency $\omega_0^A$ or $\omega_0^B$, giving rise to two distinct peaks in the spectrum. In the fast exchange limit, each spin flips between the states A and B many times during the measurement time, so that effectively a population average frequency is observed. However, when the timescale of exchange and detection are comparable, the precession frequency is ill-defined, giving rise to a broad smear in the Fourier transform spectrum. In the Carr-Purcell-Meiboom-Gill (CPMG) experiment, a pulse train is used to refocus spin magnetisation and sharpen the peak (Carr and Purcell [1954], Meiboom and Gill [1958]). Variation of the effective transverse relaxation with the pulse frequency can be used to derive the exchange rate constants, populations and the difference in chemical shifts between the exchanging states.

The effects on the spectrum are also dependent on the relative relaxation rates in the two states, as illustrated in **Figure 1.10**. This is especially relevant for the large proteins I tackle in the thesis. There is a wide range of pulse sequences in NMR

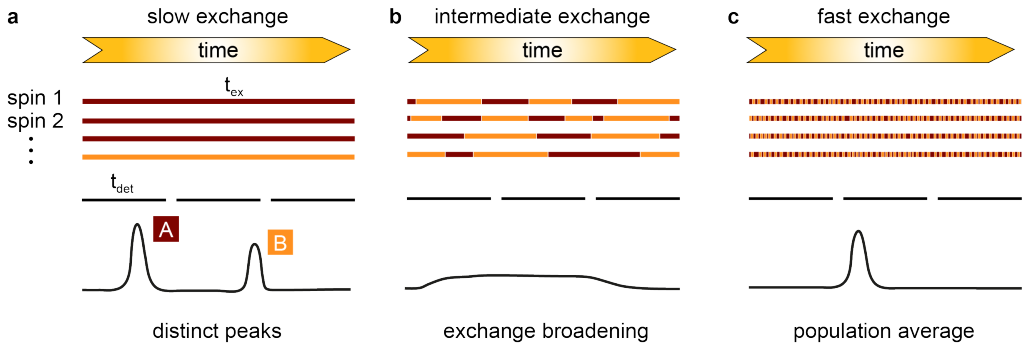



***Figure 1.9: The origin of exchange broadening.*** *$t_{ex}$ - timescale of exchange, $t_{det}$ - timescale of signal detection.* ***a*** *In slow exchange, $t_{ex} << t_{det}$ and each spin is observed either in state A or state B, therefore two distinct peaks corresponding to the two states are detected.* ***b*** *In intermediate exchange, $t_{ex}$ is comparable to $t_{det}$ and the chemical shift precession frequency is ill-defined for each spin. The Fourier transform produces a broad range of low intensity frequencies in the spectrum.* ***c*** *In fast exchange, $t_{ex} >> t_{det}$ and each spin is observed to precess at the population average frequency. A single peak is seen in the spectrum.*

designed to study exchange at different timescales. Some of these experiments are mentioned in **Figure 1.10**; the ones most relevant to this thesis are CPMG and dark state exchange saturation transfer (DEST).

A final NMR element worth mentioning here is pulsed field gradients. In gradients, a field is applied whose strength varies across the sample, for instance from top to bottom. In particular, gradients enable measurement of diffusion coefficients by NMR (Stejskal and Tanner [1965], Sinnaeve [2012]): a defocusing gradient, a 180° pulse and the equivalent refocusing gradient is applied to the sample (**Figure 1.11**). Slowly diffusing species will have an intense signal since they receive the same amount of energy from the two gradients and are effectively refocused. Species that have diffused away from their initial position will have a weak signal since the fields they experience in the two gradients are not quite identical and the refocusing is less effective. Measurement of signal intensity as a function of gradient strength enables calculation of a diffusion coefficient.

In summary, NMR is a technique that allows us to obtain information about the environment of the nuclei in the molecule, for example their connectivity, the size and motion of the species they are part of, the local dynamics they are undergoing

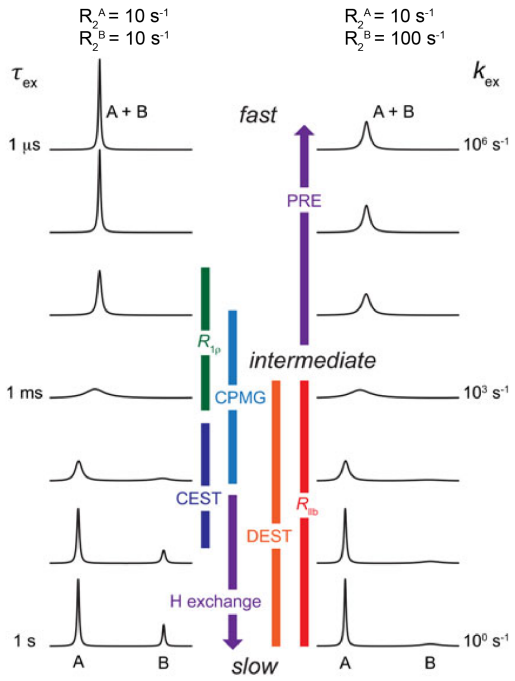



***Figure 1.10: Illustration of the effects of chemical exchange on the NMR spectrum.*** *Simulated spectra for a spin that can exist either in state A or in state B are shown. From top to bottom, the exchange rate varies from* $10^6\,s^{-1}$ *to* $10^0 s^{-1}$*, with the equivalent variation in exchange lifetime* ($\tau_{ex}$) *also shown. Left column: the* $R_2$ *constants of the two states are similar at* $10\,s^{-1}$*. Right column:* $R_2$ *of state B* ($100\,s^{-1}$) *is much greater than that of state A* ($10\,s^{-1}$)*. In the slow exchange limit, the two peaks are well separated. In the fast exchange limit, a population-weighted average peak is observed. In the intermediate regime, the peaks are broadened. If* $R_2$ *of the second state is high, the average signal is subject to additional broadening in the intermediate and fast regime. The coloured ranges denote the applicability of various experiments for the study of chemical exchange. CPMG - Carr-Purcell-Meiboom-Gill relaxation dispersion, PRE - paramagnetic relaxation enhancement, CEST - chemical exchange saturation transfer, DEST - dark state exchange saturation transfer,* $R_{ilb}$ *- lifetime line broadening. Note that CEST and DEST pulse sequences are identical. Adapted from Anthis and Clore [2015].*

and interaction with other nuclei. It can be understood by looking at the evolution of the density matrix under the relevant Hamiltonians. The quantum mechanical approach to the technique allows us to simulate pulse sequences designed to probe different aspects of molecular behaviour, and to fit the data with models with an understanding of the underlying contributions of various physical effects.

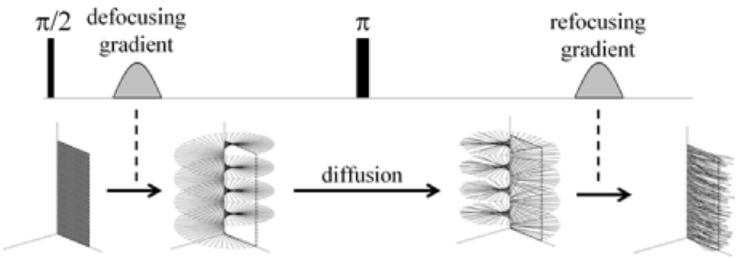



***Figure 1.11: Illustration of the basic NMR diffusion experiment.*** *The top panel shows the pulse sequence, with black bars denoting hard pulses. The bottom panel shows the effects on transverse magnetisation along the tube, aligned with the z-axis. For slowly diffusing species, the refocusing by the second gradient nearly cancels out the effect of the defocusing gradient and their signal intensity is high. Quickly diffusing species experience different field strengths during the two gradients, leading to loss of signal intensity. Note that for proteins additional pulses are applied so that magnetisation is longitudinal during the diffusion delay. This is necessary for adequate sensitivity, since for large molecules longitudinal relaxation is much slower than transverse relaxation. Adapted from Sinnaeve [2012].*

### 1.5.2 Multiparameter optimisation

Optimisation refers to finding a set of parameters that maximises or minimises a given function. This is possible when the number of datapoints is greater than the number of parameters to be optimised, i.e. the problem is overdetermined. Consider a simple example of a set of data that conform to a linear relationship, for instance, ThT fluorescence as a function of aggregate concentration in **Figure 1.12**. We seek a line that fits the data best, i.e. for which the distance between each point and the line is minimal. The line is of the form $y = p_1x + p_2$. The function to be minimised is then the sum of the squares of the deviations from the line:

$$\begin{aligned} F &= (y_1^{meas} - y_1^{pred})^2 + (y_2^{meas} - y_2^{pred})^2 + ... \\ &= (y_1^{meas} - p_1x_1 - p_2)^2 + (y_2^{meas} - p_1x_2 - p_2)^2 + ... \end{aligned}$$

The parameters to be optimised, or the degrees of freedom, are the gradient $p_1$ and the intercept $p_2$, and their possible values constitute the parameter search space. The search space can be constrained, e.g. the scaling parameter $p_1$ cannot be negative. The function F is the scoring or cost function.

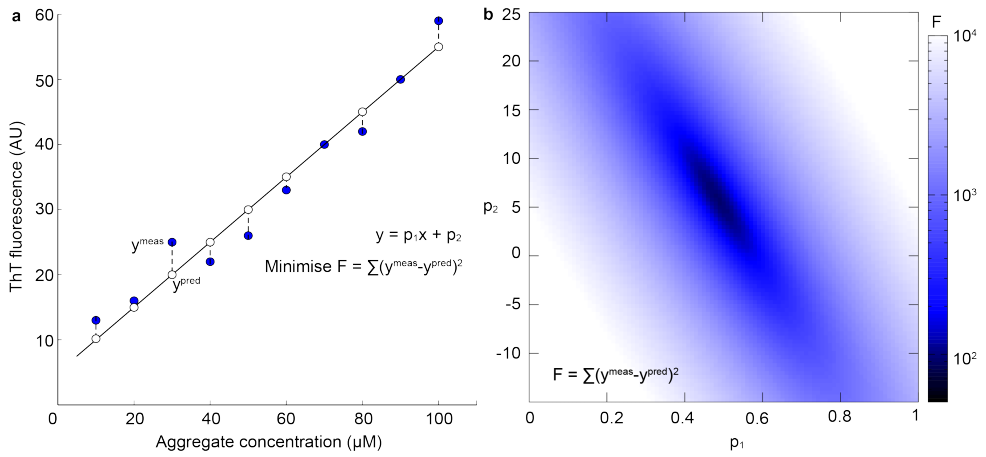



***Figure 1.12: Parameter space and cost function in linear least squares fitting.*** ***a*** *Example dataset and a line fitting the data, illustrating the measured and predicted values of $y$.* ***b*** *The surface of the cost function as a function of the parameters $p_1$, the gradient, and $p_2$, the intercept. The optimal values in this case are 0.5 and 5.0, respectively.*

In this case, the minimisation can be done analytically, by differentiating $F$ as a function of $p_1$ and $p_2$ and setting these derivatives to zero. This approach is possible in cases where the combination of parameters is linear, i.e. a sum rather than a product or ratio. It is termed linear least squares optimisation.

If the relationship is a non-linear combination of parameters, e.g. a Gaussian distribution of the form $y = p_1 \frac{(x-p_2)^2}{2p_3^2} + p_4$, the analytical approach cannot be used. When $F$ is differentiated with respect to each parameter, the resulting equations will be functions of other parameters, so they will not have closed solutions. In non-linear least squares optimisation, initial values of the parameters are chosen and then refined iteratively. In **Figure 1.13**, this corresponds to moving from the point $(p_1, p_2)_{init}$ through a series of intermediate points to $(p_1, p_2)_{min}$.

If the cost function has only one minimum, the solution can be found regardless of where the algorithm was initialised. However, such cases are rare and in practice it is important to choose initial parameters that are as close as possible to the optimal solution. Suitable initial guesses can be determined on the basis of exploring the parameter space and knowledge of the physical constraints on the problem (e.g. a rate constant of a bimolecular reaction usually cannot be higher than the diffusion

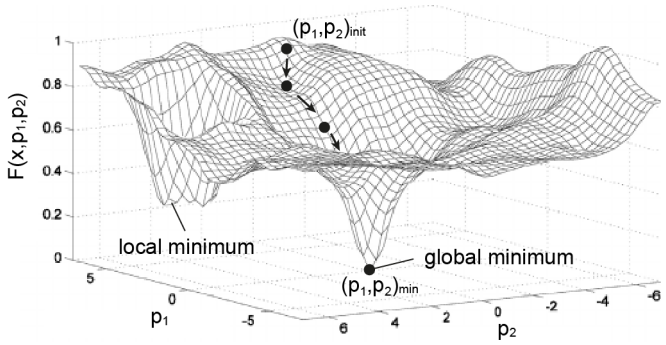



***Figure 1.13: An illustration of non-linear least squares fitting for a system with two parameters,*** $p_1$ ***and*** $p_2$***.*** *We start from a given point on the cost function surface* $F(x, p_1, p_2)$ *(initial conditions,* $(p_1, p_2)_{init}$*) and move in steps towards the minimum (*$(p_1, p_2)_{min}$*) using a certain recurrence relation, which depends on the algorithm used. Adapted from Cuevas et al. [2013].*

limit).

There are a range of strategies for the trajectory to take from the initial guess (Press et al. [2007]). In the gradient descent method of minimisation, steps proportional to the negative of the gradient at the given point are taken. In maximisation, the equivalent would be gradient ascent where the steps are proportional to the gradient. Note that the differentiation to calculate the gradient is performed with respect to the parameters as opposed to the dependent variables. The recurrence relation, i.e. the relation between successive values of $p$, is as follows:

$$p_{i+1} = p_i - \gamma_i \nabla F(x, p_i) \tag{1.21}$$

Here the scaling parameter $\gamma_i$ is chosen using one of several existing strategies, e.g. using a Wolfe line search (Wolfe [1969]). The gradient descent method performs well in bowl-shaped functions, but tends to zig-zag and therefore converge slowly in narrow valley-shaped functions.

In the Gauss-Newton method, the step size is dependent on the second derivative of $F(x, p)$, i.e. it is a second-order method. The recurrence relation is:

$$p_{i+1} = p_i - \frac{\nabla F(x, p_i)}{\nabla^2 F(x, p_i)} \tag{1.22}$$

The matrix of the second derivatives of $F(x, p)$ is also called the Hessian. The method can be slow if many parameters need to be optimised because of the need to invert the Hessian matrix.

The Levenberg-Marquardt method, also known as the damped least squares method, combines the above algorithms. Its recurrence relation is as follows:

$$p_{i+1} = p_i - \frac{\nabla F(x, p_i)}{\nabla^2 F(x, p_i) - \gamma_i \mathbf{I}} \tag{1.23}$$

The value of $\gamma_i$ is adjusted in each step of the algorithm. If the surface of the cost function is steep, a smaller value of $\gamma_i$ can be used, making the algorithm more similar to the Gauss-Newton method. If the surface is shallow, $\nabla^2 F(x, p_i)$ will be small and $\gamma_i$ will be dominant, and the method will operate more like gradient descent.

The above methods will find the nearest minimum, i.e. the local minimum, but not necessarily the overall, or global minimum in the search space if the cost function has a complex shape (see **Figure 1.13**). A thorough grid search is usually prohibitively time-consuming as a means of definitively locating the global minimum. Therefore, heuristic approaches for varying the parameters are used, which often incorporate rules that allow stochastic variation. They are sometimes combined with local optimisation techniques and termed meta-heuristics. The heuristic algorithms are more efficient than a grid search but are not guaranteed to find the best solution.

A simple meta-heuristic for avoiding local minima is to allow random variation of the parameters between successive runs of the local minimiser (a basin hopping strategy, Olson et al. [2012]). For cases where the parameters are discrete, the parameters already tested can be stored in memory and not used in subsequent runs, in a method termed the tabu search (Glover and Laguna [1997]). This is more difficult to implement for cases where parameters are continuous rather than discrete variables.

Another meta-heuristic is the simulated annealing technique (Kirkpatrick and Gelatt [1983]). The concept comes from the process used to cool glass or metals. When the material is cooled slowly, the atoms form a perfect, lowest-energy crystalline arrangement. If it is cooled too fast, it often ends up in an amorphous higher-energy state. The key is therefore allowing sufficient time for exploration of other states before committing to a minimum. This is accomplished by a gradually decreasing temperature variable which is equivalent to the probability of accepting significantly worse solutions. At high temperature, coarse variation in the cost function landscape is explored, and at low temperature, the search is mostly restricted to fine variation in the best local minimum.

A further variety of a meta-heuristic, the family of genetic algorithms, is based on the analogy of evolution of populations in biology (Goldberg [1989]). In these algorithms, populations of solutions are improved with successive iterations by selection of a subpopulation of best "parent" solutions and recombination of parameters from this subset to produce "offspring". The parameters are allowed to mutate. Like the tabu search, this strategy is better suited for discrete rather than continuous variables.

Particle swarm optimisation (Kennedy and Eberhart [1995]) is a recently invented method that is also based on a population of solutions ("particles"). The particles move around in the search space with certain velocities and share knowledge about the search space with each other. The velocity of each particle depends on the distance from its own best solution so far and the distance from the global best solution achieved by any particle in the swarm. In addition, the particles have inertia which increases over time and leads to exploration of the optimal area locally towards the end of optimisation process.

An important consideration in each of the above is the definition of convergence, i.e. rules that determine when a sufficiently good solution has been found. Strategies for convergence criteria include defining the acceptably small value of the

cost function, stopping when no further improvement in fitness is obtained after a certain defined number of steps, or when variation in the parameters is below a defined threshold. Another approach is to start several parallel runs and stop when they all converge to similar values of parameters. Like the choice of initial conditions, the selection of appropriate convergence criteria relies on knowledge of the problem at hand.

In summary, finding the best parameters to describe a given system involves optimisation of the cost function in the parameter space. If the problem is linear with respect to the parameters, the problem can be solved analytically. For non-linear problems, a number of minimisation algorithms and meta-heuristics have been developed. Factors that influence which algorithm to choose include the number of parameters, any knowledge of the shape of the cost function surface in parameter space and the effort required to calculate the cost function and its various derivatives. Here I use the Levenberg-Marquardt algorithm in combination with the basin hopping metaheuristic, which is easy to implement and performs at a practical speed for the multiparameter aggregation mechanisms, unlike the biological metaheuristics. The choice of initial conditions and convergence criteria were tailored to the system under study, as explained in more detail in **Chapter 4**.

### Model comparison

Often, several alternative models describing a given system are available and a choice of the most appropriate model must be made. For example, this is the case for various hypothetical aggregation and chaperone action mechanisms in **Chapters 4** and **5**. The decision can be made objectively using statistical principles. In general, a model with a larger number of parameters is expected to make a better fit to the data. The statistical significance of this improvement can be evaluated using an F-test. A sum of $\nu$ independent variables, such as a number of independent experimental measurements, gives a $\chi^2$ distribution. The F-distribution is a ratio

of two $\chi^2$ distributions, for example for a given simple and more complex models, weighted by their respective degrees of freedom. The F-statistic is calculated as:

$$F = \frac{\chi^2_{simple} - \chi^2_{complex}}{\frac{\nu_{simple} - \nu_{complex}}{\nu_{complex}} \chi^2_{complex}} \tag{1.24}$$

where $\nu_{complex}$ and $\nu_{simple}$ are the number of degrees of freedom, i.e. data points minus the number of parameters. To judge whether the difference in the $\chi^2$ is significant, a confidence interval $P$ can be calculated, giving the probability of the more complex model giving a fit of at least this quality by chance:

$$P = \beta_R\left(\frac{\nu_{complex}}{2}, \frac{\nu_{simple} - \nu_{complex}}{2}, \frac{\chi^2_{complex}}{\chi^2_{simple}}\right) \tag{1.25}$$

where $\beta_R$ is the regularised incomplete $\beta$ function, defined as:

$$\beta_R(x, y, z) = \frac{\Gamma(x+y)}{\Gamma(x)\Gamma(y)} \int_0^z t^{x-1}(1-t)^{y-1} dt \tag{1.26}$$

where the gamma function $\Gamma(z) = \int_0^z t^{z-1} e^{-t} dt$. A value $P$ of less than a certain threshold, e.g. 0.05, is interpreted as the more complex model being statistically justified. Note that the P-level is undefined if the number of degrees of freedom is equal for the two models ($\nu_{simple} = \nu_{complex}$). In this case the model with a lower $\chi^2$ should be accepted.

### 1.5.3 Native mass spectrometry

MS is a technique by which chemical species can be characterised in the gas phase based on their mass and charge properties. It relies on manipulation of gaseous ions with electric fields and gas pressure. In native MS, non-covalent interactions within a macromolecular assembly are preserved. This is illustrated in **Figure 1.14a** for haemoglobin. Energy can be applied to the sample, causing the complexes to break up into constituent parts and therefore reveal their composition. Exploiting isotopic

labelling also allows measurement of solution subunit exchange rates of a multimeric assembly (**Figure 1.14b**).

An MS experiment has at least three stages: ionising the sample, resolving the ions in the gas phase and detecting the resolved ions. One of the prerequisites for the development of native MS was the invention of soft ionisation methods: matrix-assisted laser desorption ionisation (MALDI) and electrospray ionisation (ESI) (Tanaka et al. [1988], Fenn et al. [1989]). MALDI works by firing a laser at the sample deposited in a matrix on a metal target. In ESI, a thin jet of sample solution is sprayed from a capillary under an electric field (**Figure 1.15**). Several models for ionisation of the analyte from droplets exist. In the ion evaporation model (Iribarne and Thomson [1976], Nguyen and Fenn [2007]), the field on the

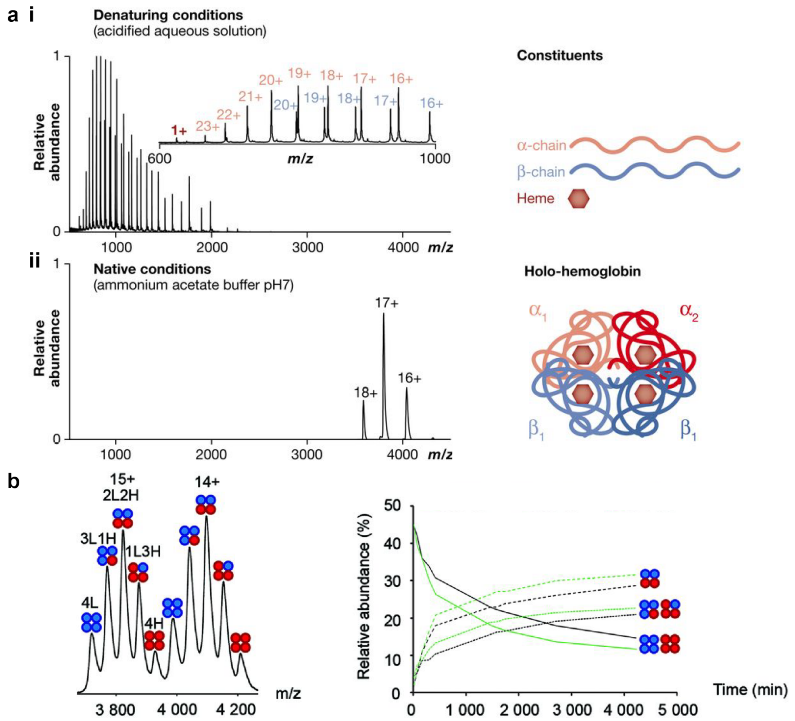



***Figure 1.14: Example applications of native mass spectrometry (MS). a*** *Denaturing versus native mass spectrometry.* ***i*** *Denaturing spectrum of haemoglobin, in which the component α and β chains and the haem appear as distinct series of charge state peaks.* ***ii*** *Native spectrum of haemoglobin showing a single charge distribution corresponding to the intact holo tetramer. Adapted from Lössl et al. [2016].* ***b*** *Illustration of measurement of subunit exchange rates of tetrameric transthyretin by mixing samples of isotopically light (L, $^{12}C^{14}N$) and heavy (H, $^{13}C^{15}N$) samples and monitoring the intensity of mixed oligomer peaks. Adapted from Marcoux et al. [2015].*

surface of a droplet due to charge density acts to eject analyte molecules, which carry multiple charges away with them. In the charged residue model (Dole et al. [1968], Mack et al. [1970], Fernandez De La Mora [2000]), the solvent evaporates completely, with the charges left on the naked analyte. Recently a model which combines both effects has been put forward (Hogan et al. [2009]).

ESI is much more amenable to native MS as the sample is sprayed in solution, and since ESI confers multiple charges there is no problem of the singly charged high-mass ions being outside of the m/z range of the instrument.

Other important features of the instrument necessary for native mass spectrometry are the use of a mass analyser with a sufficiently high upper limit in m/z, mild voltages throughout the instrument and higher source pressure (Hernandez and Robinson [2007], Leney and Heck [2017]). The most common analyser for native MS work is the time-of-flight analyser, which theoretically has no upper limit on m/z. Its basic components are the accelerator region and the time-of-flight region (**Figure 1.16**). In the former, all ions are accelerated by a voltage $V$, acquiring equivalent kinetic energy. If $q$, $m$ and $v$ are the charge, mass and velocity of an ion, and $q = ze$, where $e$ is the charge of an electron and $z$ is an integer, the energy acquired by the

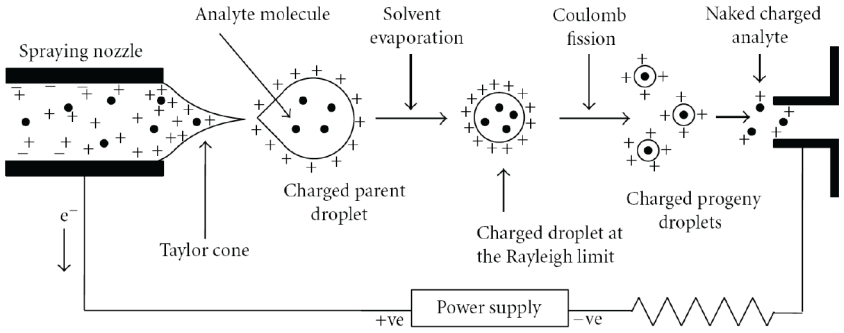



***Figure 1.15: An illustration of the electrospray ionisation process.*** *Application of a voltage to the spraying nozzle, or needle, leads to formation of charged droplets containing solvent and analyte molecules. The solvent gradually evaporates and the size of the droplet decreases. When a critical threshold of charge density with respect to surface tension is reached (the Rayleigh limit) the droplet explodes into smaller droplets. The process is repeated. In the charged residue model of analyte ionisation, eventually most of the solvent evaporates, leaving the charged analyte behind. Adapted from Banerjee and Mazumdar [2012].*

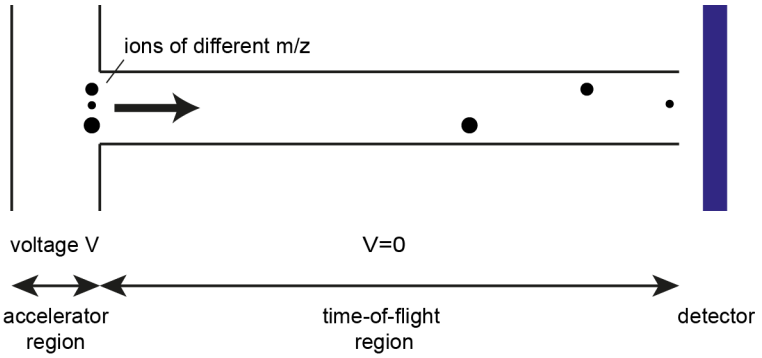



***Figure 1.16: The time-of-flight analyser.*** *In the accelerator region, all ions are given equivalent kinetic energy by the accelerating voltage. Since their masses and charges are different, they acquire different velocities. In the field-free time-of-flight region, the ions separate according to their velocity and reach the detector at different times.*

ions is

$$qV = zeV = \frac{mv^2}{2} \tag{1.27}$$

The expression for velocity $v$ then is:

$$v = \sqrt{2eV\frac{z}{m}} \tag{1.28}$$

The ions then enter the drift tube region which is field-free, and are separated according to their velocity. If the length of the drift tube is $L$, an ion will take the time $t = \frac{L}{v}$ to reach the detector. Substituting the expression for $v$ from **Equation 1.28**, we have

$$t^2 = \frac{m}{z}(\frac{L^2}{2eV}) \tag{1.29}$$

Thus measurement of $t$ and knowledge of instrumental parameters $L$ and $V$ allows calculation of m/z. Ions with higher m/z will arrive at the detector later than ions with low m/z.

Another important analyser is the quadrupole. In simple terms, it consists of four rods to which oscillating radiofrequency (RF) voltages and direct current (DC) offset are applied (**Figure 1.17a**). Neighbouring rods have opposite polarities. Under the influence of the field, which is described by the Mathieu equations (De Hoffmann and Stroobant [2012]), ions have sinusoid trajectories. Depending on

their m/z and the exact voltage pattern applied, some ions have stable trajectories, allowing them to reach the detector, and others have unstable trajectories, colliding with the parts of the analyser. Solutions of the Mathieu equations give regions in voltage space in which ions of different m/z have stable trajectories (**Figure 1.17b**). In RF-only mode, all ions are transmitted, i.e. the quadrupole acts as an ion guide. However, if DC and RF voltages are set to certain values, ions of desired m/z can be selected. In combination with further analysis using e.g. a time-of-flight analyser this is referred to as tandem MS or MSMS.

In native mass spectrometry, the conditions in most parts of the instrument are mild in order to preserve the molecules intact. However, it is often useful to be able to disrupt the assemblies, for example to analyse their composition. To achieve this, special elements termed collision cells can be included in an instrument. In a collision cell, an inert gas such as argon is supplied at adjustable pressure. High voltage can be applied across the cell. At sufficiently high pressure and voltage ions undergo energetic collisions with the gas atoms. The collisions gradually increase their internal energy of the molecules to the point where they fragment. This is termed collision-induced dissociation (CID). If the analyte is a non-covalent assembly, some of the subunits dissociate. **Figure 1.18** shows an example for transthyretin. Usu-

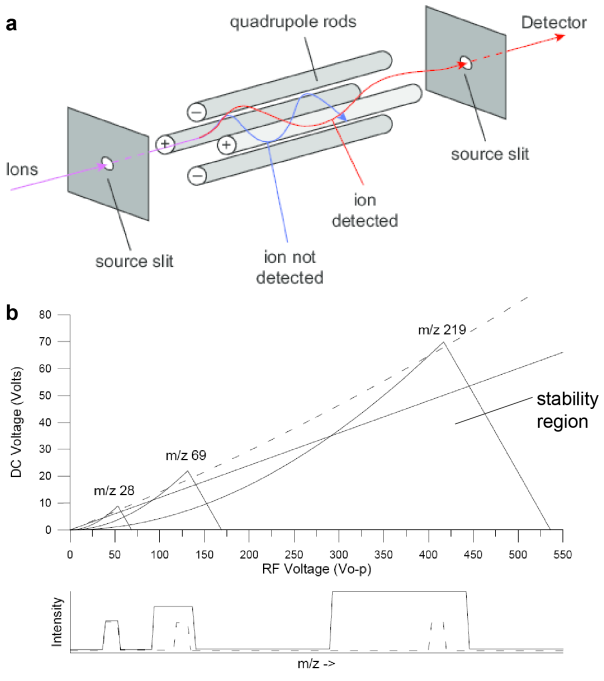



***Figure 1.17: The quadrupole analyser.*** ***a*** *Schematic of the arrangement of the quadrupole rods and illustration of a stable or an unstable ion trajectory.* ***b*** *Illustration of stability regions in DC/RF voltage space for ions of different m/z and example scan lines (solid line and dashed curve) with the peak widths they produce (lower panel). With voltages varied as per the dashed curve, higher resolution and uniform peak widths across the m/z range can be achieved. Adapted from Kaklamanos et al. [2016].*

ally asymmetric dissociation occurs, i.e. a tetramer dissociates into a monomer and a trimer, although symmetric dissociation has also been observed (Pagel et al. [2010]). CID is a popular ion activation method due to the control that it gives over the energy supplied to the analyte. Many other ways to fragment complexes, such as electron transfer, surface- or laser-induced dissociation, have also been implemented and have uses in specific applications (Brodbelt [2016]).

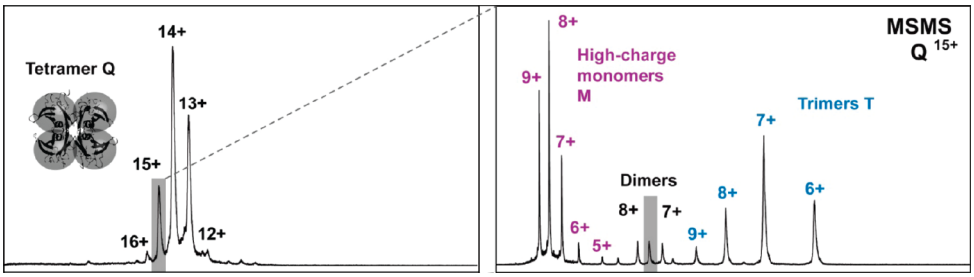



***Figure 1.18: Collision-induced dissociation.*** *An example intact spectrum of transthyretin tetramer is shown on the left and MSMS CID spectrum of the 15+ charge state is shown on the right. Asymmetric dissociation is prominent, with disproportionately highly charged monomers released, leaving stripped trimers behind. Some dimers arising from symmetric dissociation are also observed. Adapted from Pagel et al. [2010].*

### 1.5.4 Ion mobility

With native mass spectrometry, we can obtain not only stoichiometry information, but also structural information in combination with ion mobility (IM). In an IM cell, the ions move through an inert gas such as nitrogen at relatively high pressure under a weak electric field (**Figure 1.19a**). The field is too weak for any fragmentation to occur. Instead, the collisions act to slow down the ion. Among other factors, the number of collisions depends on the size and shape of the ion. IM allows determination of the ion's rotationally averaged collision cross-section, a quantity that reports on its size and shape. Comparison of CCS measurements with values calculated from X-ray or NMR structures for the same proteins shows that they are very well correlated (Benesch and Ruotolo [2011]). Measurement of the CCS can separate different conformers or oligomeric states of a macromolecule. The CCS can be used as a restraint in building structural models. Predicted CCS values can be

calculated from existing or candidate structures of the macromolecule using various algorithms, such as projection approximation, and compared to experimental values. This way incorrect models can be ruled out and the most likely models deduced. The extra dimension in combination with mass spectrometry can also allow analysis of more complex mixtures. For example, **Figure 1.19b** shows an experiment in which binding of a peptide ($Pep$) to a domain of the protein $\alpha$B-crystallin ($\alpha B_{Dom}$) is monitored. $\alpha B_{Dom}$ can bind one peptide. The domain can also dimerise ($(\alpha B_{Dom})_2$) and the dimer can bind up to two peptides ($(\alpha B_{Dom})_2(Pep)$ and $(\alpha B_{Dom})_2(Pep)_2$). The resulting MS and IM spectra (horizontal and vertical projections in **Figure 1.19b i**) have overlapping peaks which are resolved in the 2D IMMS form. The spectrum can be computationally fitted with the knowledge of the mixture components. Such a fitted spectrum is shown in **Figure 1.19b ii**, and in **Figure 1.19b iii** the peaks are coloured by species to illustrate the assignment. Note that each species gives rise to several peaks as it appears in several charge states with their characteristic

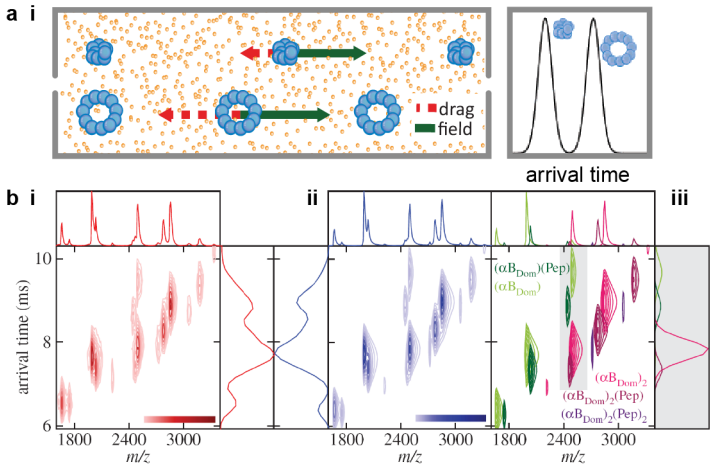



***Figure 1.19: Ion mobility mass spectrometry (IMMS). a*** *Schematic of an ion mobility drift tube, in which the ions move through a bath of buffer gas under the influence of a weak electric field. The macromolecular factors influencing the arrival time are ion charge and collisional cross-section (CCS). Different conformers or oligomeric states can be resolved by their arrival time. From Benesch et al. [2007].* ***b i*** *An example 2D IMMS spectrum of a mixture of a protein domain ($\alpha B_{Dom}$) and a peptide it binds ($Pep$). The domain can dimerise, which is denoted as $(\alpha B_{Dom})_2$.* ***ii*** *The computationally fitted spectrum based on the predicted relative abundances of the various species.* ***iii*** *The same spectrum highlighting the contributions of different species. The ion mobility dimension helps to resolve the protein monomer, dimer and the various peptide-bound forms. Adapted from Hilton et al. [2013].*

m/z and arrival times. Kinetic and thermodynamic parameters of the interaction can be extracted from relative abundance of the different species over the course of a titration.

The Mason-Schamp equation (Revercomb and Mason [1975]) relates the CCS to the drift time of the ion, $t_D$, its charge $z$ and various constants:

$$CCS = \frac{\sqrt{18\pi}}{16}\frac{ze}{\sqrt{k_b T}}\sqrt{\frac{1}{m_i}+\frac{1}{m_g}}\frac{760}{P}\frac{T}{273.2}\frac{1}{N}\frac{t_D E}{L} \tag{1.30}$$

Here $e$ is the charge on an electron, $k_b$ is Boltzmann constant, $T$ is temperature, $m_i$ is the mass of the ion, $m_g$ is the mass of the buffer gas molecule, $P$ is pressure is the cell, $N$ is the buffer gas number density, $E$ is the electric field and $L$ is the length of the cell.

Some of the newest instruments operate using travelling waves (**Figure 1.20**). These are produced by wave-like switching of voltages in ring electrodes in the different cells of the instrument. Depending on the speed of the wave and the height of the barrier, the ions can either travel in the wave valleys (regions of low electric potential) or skip over the waves. Travelling waves improve transmission of the ions as well as resolution in IMMS.

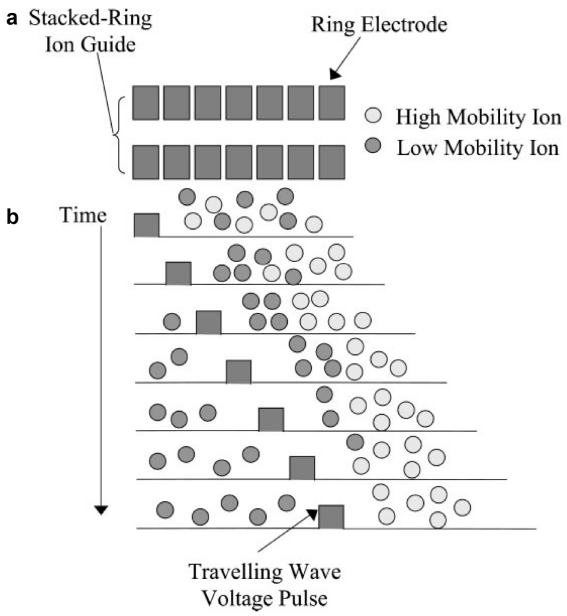



***Figure 1.20: Ion mobility with travelling waves.*** ***a*** *Top cross-section view of a cell with ring electrodes.* ***b*** *Effect of the waves on analyte ions. DC pulses are applied to the electrodes in turn, resulting in periodic wave-like electric field. The field pushes all ions forward, but ions of lower mobility tend to roll over the waves more frequently and are therefore retarded. The height and speed of the waves can be adjusted to achieve optimum separation for a specific analyte mixture. Adapted from Giles et al. [2004].*

In a travelling wave instrument, the electric field in the drift cell is no longer constant. Mobility separation occurs by differential rollover over the waves. Empirically, while in a linear drift tube $CCS \propto zt_D$, in a travelling wave instrument, $CCS \propto zt_D^a$. The exponent $a$ depends on the exact wave conditions and is typically less than 1. The above equation can no longer be used to calculate the CCS directly. Instead, calibration is performed with proteins of known CCS previously analysed on a linear drift tube instrument. The calibrants must belong to the same macromolecular class, since denatured and native proteins, peptides, lipids, nucleic acids and salts behave differently in IMMS (Fenn and McLean [2008], Bush et al. [2010], Hines et al. [2016]).

**Figure 1.21** shows a schematic of the Waters Synapt G1 instrument used for most experiments in this thesis. It shows how the different elements described here - the quadrupole, the collision cells, the ion mobility cell and the time-of-flight (ToF) analyser - can be combined in a configuration that enables a range of experiments to be performed. Most notably, the ability to select a narrow m/z range in the quadrupole is very useful for working with polydisperse proteins. The ability to perform CID in the transfer cell after the ion mobility measurement is indispensable

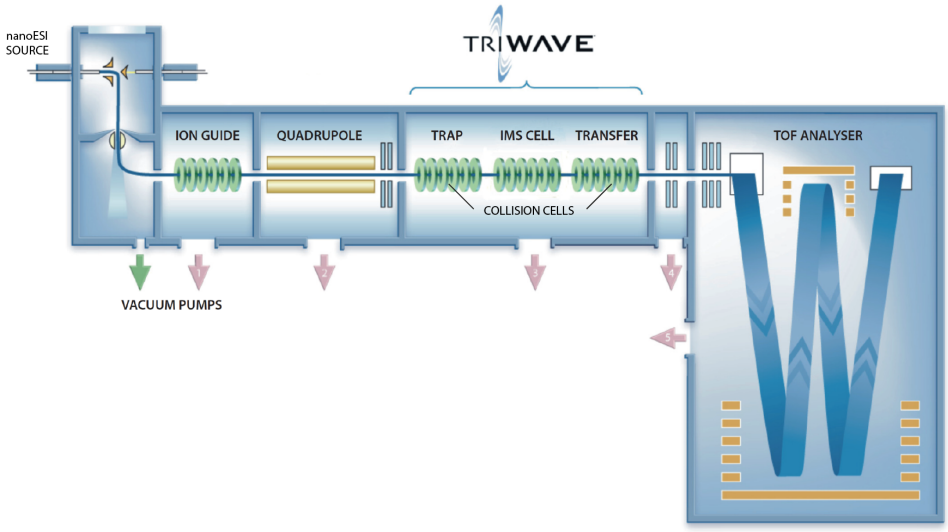



***Figure 1.21: A schematic of the Waters Synapt G1 instrument.*** *ESI - electrospray ionisation, IMS - ion mobility separation, TOF - time of flight. From www.waters.com.*

for the determination of CCS of the individual stoichiometries of polydisperse $\alpha$BC, as described in **Chapter 3**.

In conclusion, manipulation of gaseous ions with mass spectrometry and ion mobility allows elucidation of masses, stoichiometries, subunit exchange rates and structural information on native macromolecular ions. The great resolving power of native MS and IM makes these techniques especially useful for the study of polydisperse proteins. Moreover, the insight they provide into quaternary structure, dynamics and stoichiometry of interactions is highly complementary to NMR, which reports on the ensemble average of the stoichiometries present but can deliver information on the structure, dynamics and interactions of a protein at atomic resolution. These methods thus represent a powerful combination to study the structure and function of the polydisperse small heat shock protein $\alpha$BC. I describe their application to answer different questions about the structure and function of $\alpha$BC in this thesis, starting with the mechanism of crosstalk between interfaces in the oligomers in the next Chapter.

# Chapter 2

# Allosteric communication between interfaces in $\alpha$BC

*In $\alpha$BC oligomers, monomers contact each other via several sets of interfaces. One of them is formed by the $\beta$6+7 strands in the core crystallin domain dimer. Another involves the C-terminus binding into a groove on a neighbouring third crystallin domain. Recently, energy compensation has been found to take place between these interfaces, i.e. information transfer from one interface to the other must be occurring across the domain. This Chapter describes mechanistic investigation of this allostery by NMR. After initial characterisation and backbone NMR spectrum assignment of the isolated core domain construct of $\alpha$BC, core domain of $\alpha$B-crystallin (c$\alpha$BC), I present NMR data from titration experiments of C-terminal peptides to c$\alpha$BC. I use the chemical shift and intensity perturbations to trace the communication network linking the two interfaces and show that it is activated in at least two steps.*

## 2.1 Introduction

Several interactions occur between subunits within $\alpha$BC oligomers. The most notable are those (a) between core crystallin domains within a dimer, (b) between a C-terminus of one monomer and the $\beta$4-$\beta$8 groove of another monomer and (c) between the N-termini of different monomers. Little is known about the latter interaction, since the arrangement of the N-termini appears very heterogeneous and truncated constructs containing the N-termini suffer from insolubility (Jehle et al. [2009], Bagnéris et al. [2009], Laganowsky et al. [2010], Haslbeck and Vierling [2015]). The former two interfaces are better understood (**Figure 2.1**). This Chapter deals with the mechanism of crosstalk between them.

Various structures of the core crystallin domain dimer of $\alpha$BC are available (Bagnéris et al. [2009], Laganowsky et al. [2010], Jehle et al. [2010], Clark et al. [2011], Hochberg et al. [2014]). The intra-dimer interface consists of residues 110-120, which form a long strand denoted $\beta$6+7 (**Figure 2.1a-b**). In different conditions, different registers of this interface have been observed, termed API, APII and APIII (**Figure 2.1c i**). In solution, the APII register appears dominant (Hochberg et al. [2014], **Figure 2.1c ii**).

The C-terminus of a monomer can bind into the $\beta$4-$\beta$8 groove on another monomer, distinct from the one with which it interacts via the $\beta$6+7 strand (Jehle et al. [2010], Laganowsky et al. [2010]). This is termed the inter-dimer interface. **Figure 2.1a** shows this interaction in a construct with the the N-terminus truncated, and **Figure 2.1b** illustrates the interaction of a peptide with part of the C-terminus sequence with the grooves in a shorter c$\alpha$BC construct. However, NMR studies have shown that at physiological temperatures in solution the C-termini are mostly unbound (Baldwin et al. [2012, 2011a]). In fact, the timescale of two C-terminus binding events is comparable to the timescale of subunit dissociation (Baldwin et al. [2011a], Hilton et al. [2013]). This suggests that, paradoxically, C-terminus binding may precipitate subunit release.

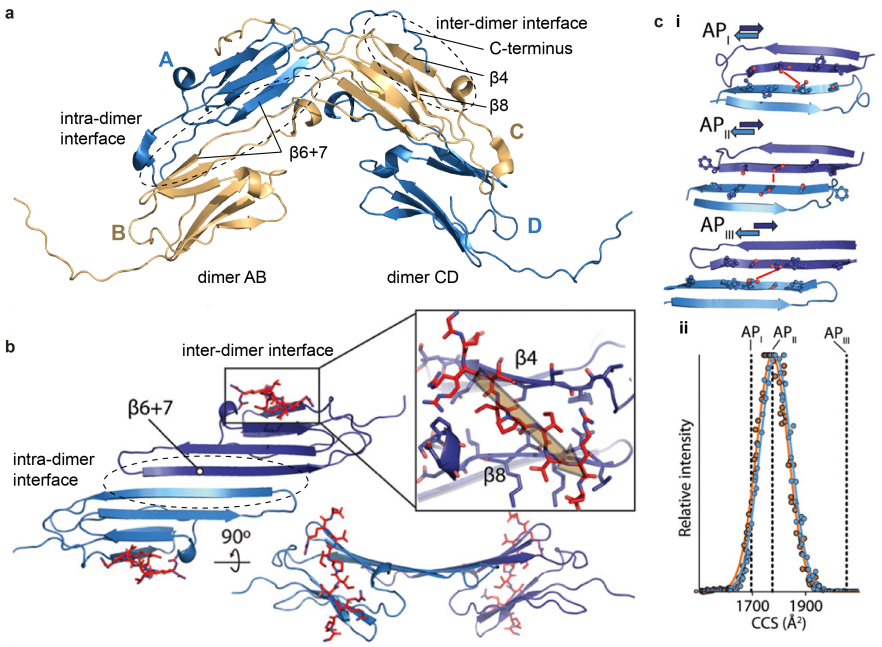



***Figure 2.1: Subunit interfaces in αBC.*** ***a*** *A crystal structure of the construct $\alpha BC_{68-162}$ (PDB ID 3L1G, Laganowsky et al. [2010]), showing four monomers A-D. They are arranged into two dimers, AB and CD, with contact along the β6+7 strands. This is termed the intra-dimer interface. Monomers A and C are also in contact through their C-termini swapping into the respective β4-β8 grooves, forming the inter-dimer interface.* ***b*** *A crystal structure of the isolated core domain dimer, $\alpha BC_{68-153}$, with monomers in purple and blue, and the C-terminal peptide ERTIPITRE, residues 156-164, shown as sticks (PDB ID 4M5S, Hochberg et al. [2014]). The β-strands involved in intermolecular interaction are annotated.* ***c i*** *An illustration of the three registers of the intra-dimer interface, API, APII and APIII.* ***ii*** *CCS measurement of the +7 charge state of wild-type cαBC (blue) and oxidised cαBC E117C mutant (orange), which is locked into the APII register. The distributions overlap and differ from computational predictions for API and APIII (dashed lines). This suggests that the wild-type protein preferentially populates the APII register in solution. Adapted from Hochberg et al. [2014].*

The following observations may account for this paradox. Native MS measurements on point mutants of αBC together with C-terminal peptide-cαBC binding assays have shown that there is energetic compensation between the interfaces: the weaker the intra-dimer interface, the stronger the inter-dimer interaction, and vice versa (**Figure 2.2**, Baldwin et al. [2011d], Hilton et al. [2013]). For example, in **Figure 2.2a**, the first peptide binds the dimer with a $K_D$ of 70 $\mu$M whereas the second peptide binds only with an affinity of about 105 $\mu$M. Similarly, the intra-dimer interface $K_D$ is 2 $\mu$M in the absence of C-terminal peptides and 4 $\mu$M in the presence of both peptides, with distinct bootstrap distributions (**Figure 2.2a ii**). This implies

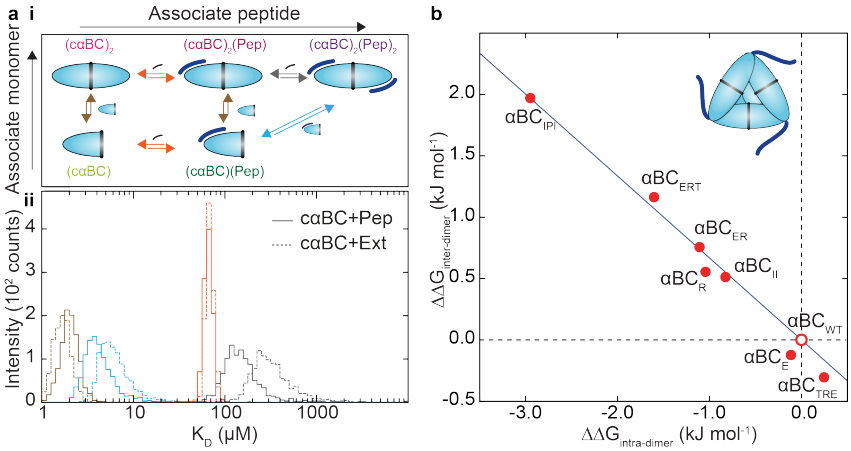



***Figure 2.2: Energetic compensation between interfaces in αBC. a*** *Native MS experiments can be used to monitor the binding of the C-terminal peptide (Pep) to cαBC monomer and dimer in the network of equilibria depicted in* ***i****. The $K_D$ values for the various interactions are shown in* ***ii****. Notably, the intra-dimer interface becomes weaker in the presence of peptide (compare the brown and blue distributions), and binding of a second peptide is less thermodynamically favourable than the binding of the first (compare grey and orange distributions). The solid and dashed lines correspond to a short (residues 156-164) and extended (residues 156–175) version of the C-terminal peptide respectively.* ***b*** *Analysis of oligomeric distributions of full-length αBC mutants allows calculation of association free energy for each interface ($\Delta G_{inter-dimer}$ and $\Delta G_{intra-dimer}$). The quantity $\Delta\Delta G$ is the difference between $\Delta G$ values for each mutant relative to the wild-type. Plotting $\Delta\Delta G_{inter-dimer}$ versus $\Delta\Delta G_{intra-dimer}$ reveals a negative correlation between these values, i.e. the stronger one interface, the weaker the other and vice-versa. Adapted from Hilton et al. [2013].*

that there is transfer of information between the two sides of the molecule across the $\beta$-sandwich structure. In other words, the molecule shows allostery, defined as the process by which proteins transmit the effect of binding at one site to another, often distal, functional site, allowing for regulation of activity (Motlagh et al. [2014]).

In the experiments described in this Chapter, I aimed to elucidate the mechanism of this allosteric communication between interfaces in $\alpha$BC with NMR. I exploited the small size of the isolated core domain dimer, which allowed me to observe it by solution-state NMR, and used C-terminal peptides to mimic the interactions on the inter-dimer interface. First, I present the characterisation of the NMR properties of c$\alpha$BC and an assignment of its backbone heteronuclear single-quantum correlation (HSQC) spectrum. This is followed by the analysis of titrations of two C-terminal peptide variants to the core domain, wild-type and the mutant

R157H, linked to dilated cardiomyopathy (Boncoraglio et al. [2012]). I show that the network of chemical shift and intensity perturbations propagates from the $\beta$4-$\beta$8 groove to the intra-dimer interface. The mutation in the C-terminal peptide does not affect the structure of the communication network but does affect the kinetic and thermodynamic parameters of its engagement.

## 2.2 Materials and methods

**c$\alpha$BC expression and purification** Human $\alpha$BC (residues 68–153) was expressed in M9 minimal medium containing $^{15}$N ammonium chloride as the only nitrogen source (characterisation and peptide titration experiments) or in deuterated M9 minimal medium with $^{15}$N ammonium chloride and $^{13}$C glucose (assignment experiments) and purified as described previously (Laganowsky et al. [2010]). Briefly, c$\alpha$BC with the N-terminal tobacco etch virus (TEV) protease-cleavable hexahistidine tag was expressed from an isopropyl $\beta$-D-1-thiogalactopyranoside (IPTG)-inducible pET28 vector in *Escherichia coli* BL21 (Agilent) for 3 hours (LB) or overnight (M9) at 37°C. The protein was purified by nickel affinity chromatography. The His tag was removed with TEV protease, and the protein buffer was exchanged using a Superdex 75 gel filtration column (GE Healthcare).

**NMR experiments** NMR experiments were performed on a 600 MHz spectrometer (Varian) equipped with a 5-mm z-axis gradient triple-resonance probe. $^{15}$N-labeled c$\alpha$BC was prepared in either 25 mM sodium phosphate (pH 7.5 unless stated otherwise), 2 mM ethylenediaminetetraacetic acid (EDTA), 2 mM sodium azide (salt-free buffer), or in 50 mM sodium phosphate, 100 mM sodium chloride, 2 mM EDTA (aggregation buffer) with 5% $D_2O$ and concentrated to 200 $\mu$M, unless stated otherwise, using Amicon concentrators (Millipore). Data were acquired at 15°C unless stated otherwise. All spectra were analyzed using NMRPipe (Delaglio et al. [1995]) and Sparky (Lee et al. [2015]).

**HSQC experiments** Two-dimensional (2D) $^{15}$N-$^{1}$H HSQC spectra were recorded with 16 scans per transient with acquisition times (t1, t2) of (23.7, 56.8) ms, over (100, 512) points and a 1 s relaxation delay. The spectra were processed with a Gauss-Lorentz window function, zero-filling once and linear prediction in the indirect dimension.

**Diffusion experiments** One-dimensional (1D) $^{1}$H constant-time pulsed field gradient $^{15}$N-edited diffusion experiments (Dehner and Kessler [2005]) were recorded with a diffusion delay of 0.1 s and gradient strengths of 4.0, 21.5, 32.0, 36.9, 42.5, 47.5, 52.0, 56.1 and 60 $\mathrm{Gcm}^{-1}$. The number of points was 576 and the acquisition time was 64.0 ms. The intensity variation at each point in the spectrum was fitted to the Stejskal-Tanner equation (Stejskal and Tanner [1965]). The average translational diffusion coefficient and its error were determined by fitting a Gaussian curve to the distribution of diffusion coefficient values at each ppm point across the spectrum.

**1D CPMG** 1D $^{1}$H $^{15}$N-edited Carr-Purcell-Meiboom-Gill (CPMG, Hansen et al. [2008]) experiments were performed on a sample of c$\alpha$BC at 200 $\mu$M in salt-free buffer (see section **NMR experiments** above). A constant time relaxation delay of 30 ms was employed with two CPMG frequencies, 67 Hz and 1000 Hz. In the $^{1}$H dimension, the acquisition time was 64.0 ms and the number of points was 616.

**Assignment experiments** Triple-resonance HNCA, HNCO and HN(CA)CO experiments (Yamazaki et al. [1994]) were acquired as a data matrix of 576×30×15 (HNCA, HN(CA)CO) or 649×40×15 (HNCO) points with acquisition times of 64.0, 6.67, and 12.4 ms (HNCA, HN(CA)CO), 72.1, 28.6 and 10.7 ms (NHCO) in the $^{1}$H, $^{13}$C and $^{15}$N dimensions, respectively; 32 (HNCA, HN(CA)CO) or 8 (HNCO) transients were acquired per FID.

**Peptide titrations** The peptides ERTIPITRE (pWT) and EHTIPITRE (pR157H) were purchased from Biomatik. The lyophilised peptides were diluted in salt-free buffer (see section **NMR experiments** above). The pH of the solution was adjusted to 7.5. A stock at a concentration of 70 mM was prepared and added in 0.5-2 $\mu$l aliquots to c$\alpha$BC, which was at a starting concentration of 200 $\mu$M and volume of 500 $\mu$l. For the titration with pWT, the tube was weighed before and after addition of peptide for more accurate correction of concentrations.

**Titration peak shape fitting** For fitting of 2D titration data, regions of the spectrum corresponding to each peak were extracted and each residue was simulated under a two-state chemical exchange model using Bloch-McConnell equations (Baldwin [2014], McConnell [1958]) implemented in Python. A Gauss-Lorentz peak shape was used with $R_2$ fixed to 15 $s^{-1}$ for both states, initial $\Delta\tilde{\omega}$ were taken from a 1D fit with an assumption of fast exchange and initial values of $K_D$ and $k^+$ describing chemical exchange were set manually. Optimisation of $K_D$ and $k^+$ was performed using Levenberg-Marquardt non-linear least squares protocol (Press et al. [2007]) as implemented in the Python SciPy library.

**CSP network analysis** The c$\alpha$BC structure with the PDB ID 2W7J, in APII register, (Bagnéris et al. [2009]) and the set of chemical shift perturbations (CSPs) for the pWT titration were used. Code was written in Python to extract all heteroatom positions for the backbone and the sidechains from the PDB file and create a library of pairwise interatomic distances between above a given threshold (4 Å, see main text). Where multiple connections were found between two residues, the shortest one was retained. A PyMOL script was written to connect the atoms in question by cylinders, with each end of the cylinder also coloured by the CSP of the backbone amide of that residue. Connections were shown only if both residues had a CSP>0.045. Alternative representations were coded where the cylinders could be drawn between the C$\alpha$ or the C$\beta$ atoms of the same residue. Classification of

connections into clusters A and B was performed based on relative proximity to the nearest loop region.

## 2.3 Results and discussion

### 2.3.1 Optimisation of cABC $^{15}$N backbone spectra for assignment

An assignment of the $^{15}N^{1}H$ HSQC solution NMR spectrum of c$\alpha$BC has previously been carried out by the Klevit group (Jehle et al. [2009], residues 64-152, including a point mutation N146D). I sought to confirm the assignments for our construct and buffer conditions. Before acquiring assignment data, I optimised conditions to observe the highest number of well-defined peaks.

#### Temperature

To optimise the acquisition temperature, I recorded a series of $^{15}N^{1}H$ HSQC spectra at 10-degree intervals between 10-50°C. The spectra for 10-40°C are shown in **Figure 2.3**. The peaks shift along straight lines, and some also change in intensity or disappear altogether at high or low temperature. There are several temperature-dependent phenomena that can affect the appearance of NMR spectra: higher rate of tumbling, hydrogen bonding, exchange of protons with water, and any protein-specific dynamic processes, such as conformational changes or interactions. The extent of peak movement is thought to report on the presence and strength of hydrogen bonds, with the exposed amides showing stronger upfield shifts with temperature (Cierpicki et al. [2002], Baxter and Williamson [1997]) and may also reflect increased thermal motion (Hong et al. [2013]). A faster rate of tumbling would be expected to reduce peak width at higher temperature. On the other hand, increasing rates of solvent exchange with temperature would result in broadening, as could confor-

mational exchange. For cαBC, most peaks sharpen at high temperature, but some broaden (see arrows in **Figure 2.3**). CPMG experiments described below suggest that minor conformational exchange occurs at these temperatures. Strong broadening is therefore more likely to result from solvent exchange. Moreover, at 50°C, a dramatic loss of peaks occurred (see **Figure A.1** in the **Appendix**). Upon return to 20°C, the number of peaks and their intensity was restored (**Figure A.1**). Possible explanations for this could be very high rates of solvent exchange, reversible unfolding or oligomerisation of the protein at high temperature. In any case the higher temperatures were not suitable for assignment.

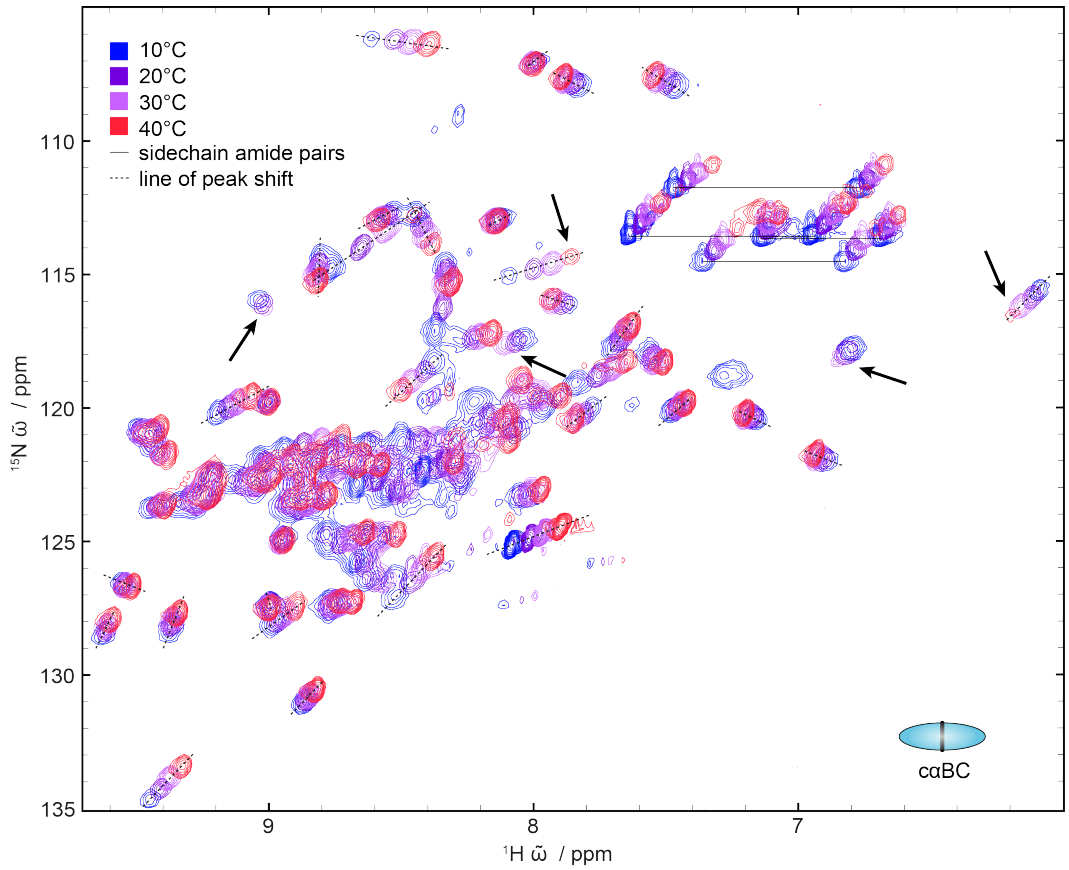



***Figure 2.3: A series of $^{15}N^{1}H$ HSQC spectra of cαBC as a function of temperature.*** *The pairs of peaks corresponding to sidechain amides in asparagines and glutamines are connected with solid lines. Temperature-dependent linear peak shifts are highlighted with dashed lines for some of the peaks. Some of the peaks that broaden at high temperature are marked with arrows.*

The number of resolved peaks was highest at 10°C and 20°C (see **Figure 2.4**). The spectrum at the intermediate temperature, 15°C, showed further minor improvements in resolution. With 78 resolved peaks, 15°C was selected as the

optimal temperature for assignment.

| Temp. / pH | 10ºC | 15ºC | 20ºC | 30ºC | 40ºC | 50ºC |
|---|---|---|---|---|---|---|
| 6.8 | | 39 | | | | |
| 7.0 | | 77 | | | | |
| 7.5 | 76 | **78** | 77 | 65 | 63 | 2 |
| 8.0 | | 70 | | | | |
| 8.5 | | 73 | | | | |

***Figure 2.4:*** *Number of resolved peaks in the $^{15}N^{1}H$ HSQC spectra of cαBC at various temperatures and pH values. The optimum conditions are highlighted in grey.*

### pH

pH generally affects the rate of exchange of the amides with the solvent, as well as electrostatic interactions with a protein. It is known to influence the monomer-dimer equilibrium in cαBC (Rajagopal et al. [2015]). As the next step in optimisation of conditions for assignment, I varied pH between 6.8-8.5 at 15°C. The results are shown in **Figure 2.5**. Fewer chemical shift changes were observed compared to the temperature series, but there were variations in intensity and resolution of peaks. The quality of the spectra deteriorated especially dramatically with lower pH. No peaks appeared in the unfolded region of the spectrum (around 8 ppm in $^{1}$H dimension). Thus the protein does not denature but may aggregate at this lower pH. The number of resolved peaks at each pH is documented in **Figure 2.4**. pH 7.5 was chosen as optimal.

The spectrum under optimised conditions (see **Figure 2.9**) contained 78 of the expected 82 backbone amide peaks (87 residues minus four prolines and the rapidly exchanging N-terminal amide). Four sidechain amide pairs were seen, as expected from two asparagines and two glutamines in the sequence.

### Concentration-dependent effects

Assignment spectra need to be acquired at high protein concentration (about 1 mM) since they are three-dimensional and the pulse sequences are necessarily longer. Any peak broadening, which could come from conformational exchange or aggre-

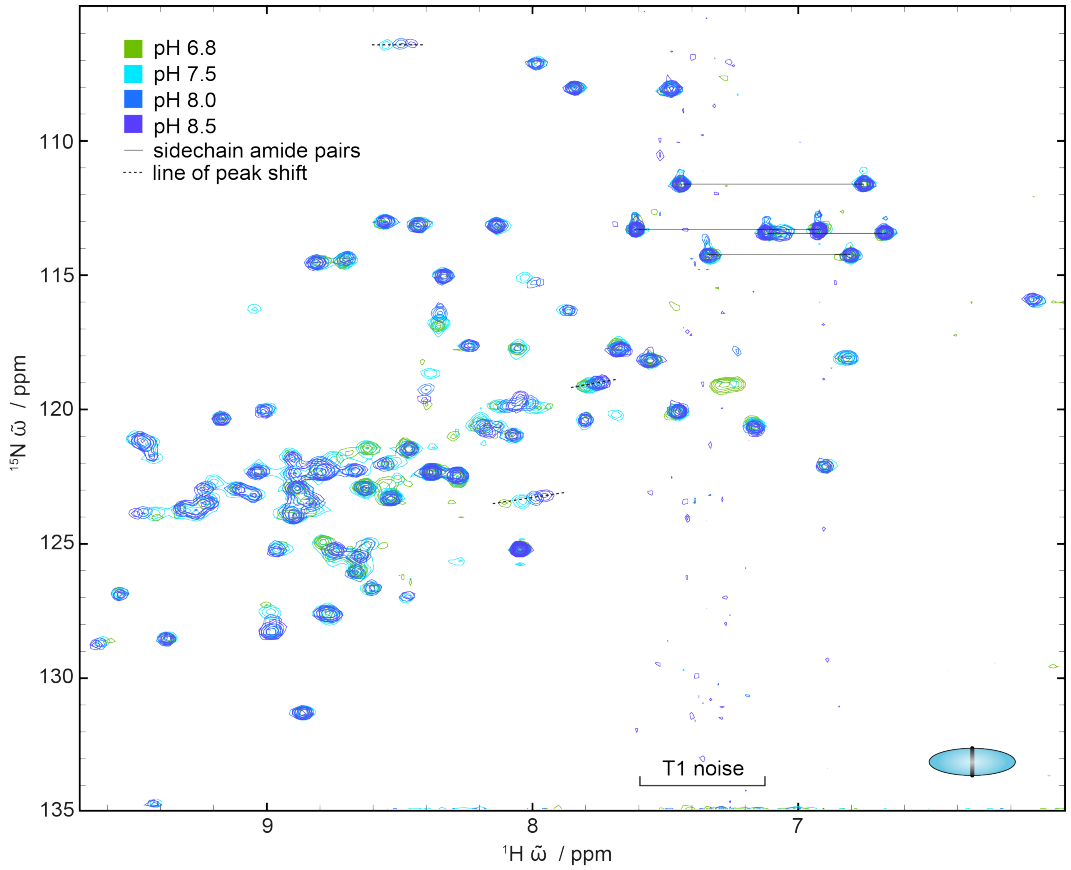



***Figure 2.5: A series of $^{15}N^{1}H$ HSQC spectra of cαBC as a function of pH.*** *The pairs of peaks corresponding to sidechain amides in asparagines and glutamines are connected with solid lines. pH-dependent linear peak shifts are highlighted with dashed lines for some of the peaks. T1 noise is an artifactual ridge of noise that appears where an intense peak is present (Mehlkopf et al. [1984]).*

gation at this concentration will hinder assignment. To check whether significant concentration-dependent broadening and/or oligomerisation was occurring with cαBC, I compared integrals of $^{15}N^{1}H$ spectra and translational diffusion coefficients as a function of protein concentration. The results are shown in **Figure 2.6**. These measurements were performed in aggregation buffer at pH 7.0 and at 37°C.

Multiple $K_D$ values have been reported for the intra-dimer interface at pH 7.5: 2±1 μM and 2±2 μM at 25°C, measured by native MS and isothermal titration calorimetry (ITC) respectively (Hilton et al. [2013], Rajagopal et al. [2015]), and 36±2 μM at 37°C, measured by ITC by Rajagopal et al. [2015]. Using the latter value, at 200 μM total concentration of protein, about 70 μM would be monomeric, i.e. about 35%. At 800 μM total concentration, about 150 μM, or 20%, would be

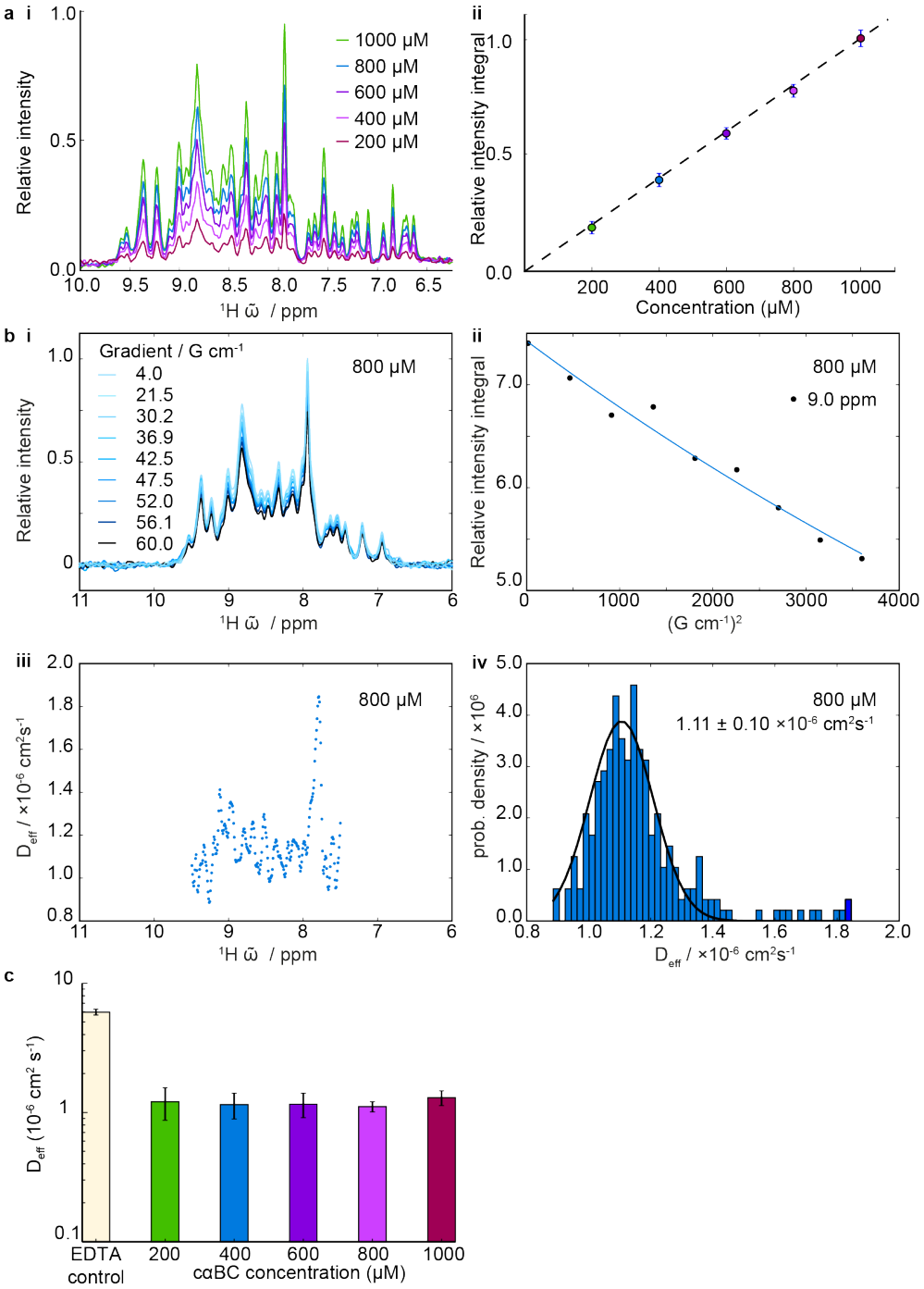



***Figure 2.6: Tests for concentration-dependent oligomerisation.*** ***a i*** *1D* $^{15}N^{1}H$ *HSQC spectra of cαBC as a function of concentration.* ***ii*** *The integrals under the spectra in* ***i*** *as a function of total protein concentration.* ***b*** *Illustration of the analysis of the diffusion experiment.* ***i*** $^{15}N^{1}H$ *HSQC spectra as a function of gradient strength. The intensity of the spectrum decreases with increasing gradient.* ***ii*** *The exponential fit of the integral versus gradient strength squared at one ppm point (9.0 ppm).* ***iii*** *Values of the effective diffusion coefficient obtained by the type of analysis shown in* ***ii****.* ***iv*** *The histogram of the* $D_{eff}$ *values in* ***iii*** *and the mean and standard deviation obtained by fitting a Gaussian curve to this data.* ***c*** *Comparison of diffusion coefficients for each cαBC concentration tested. EDTA is included as a relative scale reference.*

monomeric. Depending on relative ratios of the exchange rates and chemical shift differences between peaks in the monomeric and dimeric states, the monomer-dimer exchange could fall into the intermediate regime at higher total protein concentration (see **Section 1.5.1**). An expected relative change in diffusion coefficient for going from a 10 kDa monomer to a 20 kDa dimer is about 20 %, but since the relative proportion of the two changes little in this concentration range, significant changes in diffusion coefficient are not expected. However, formation of significant amounts of higher order oligomers would lead to nonlinear variation of the integral of the spectrum with concentration and lower diffusion coefficients at higher concentrations.

**Figure 2.6a** shows the spectral integral analysis. The integrals were clearly highly linear with concentration in the range tested. The translational diffusion analysis is presented in **Figure 2.6b-c**, with an example data analysis sequence in panel **b** and the diffusion coefficients determined at each concentration in panel **c**. There was no significant variation of the diffusion coefficients with concentration. Therefore signal broadening due to oligomerisation is unlikely to be an issue at the concentrations required for assignment.

### Exchange broadening

To determine whether c$\alpha$BC underwent any significant intermediate exchange, which would broaden the peaks and hinder assignment, I carried out a series of scouting 1D CPMG experiments at 10, 20, 30 and 40°C. The experiments consisted of three acquisitions: a reference 1D $^{15}$N$^{1}$H HSQC, without any CPMG block, and two experiments with a CPMG block with a 30 ms relaxation delay and pulse frequencies of either 67 Hz or 1000 Hz. Any chemical exchange broadening arising from chemical exchange on the $\mu$s-ms timescale would be more effectively refocused at the higher CPMG pulse frequency. Therefore any intensity increase in the high-frequency experiment relative to the low-frequency one would signal the presence of exchange

broadening.

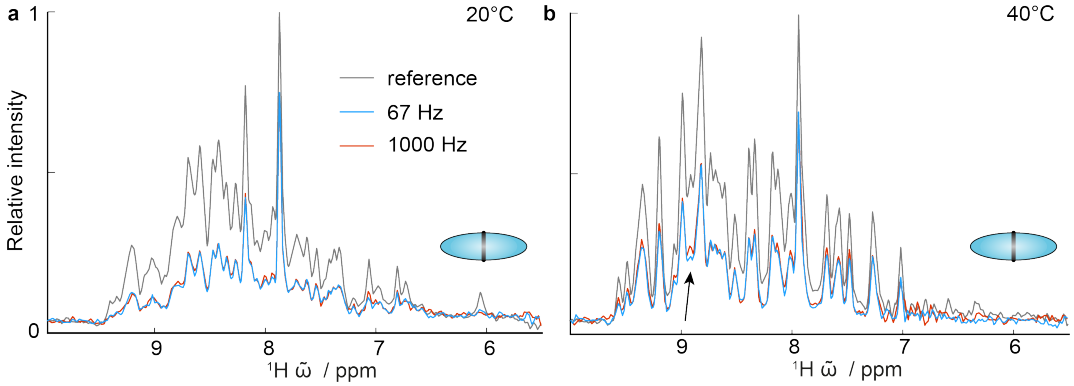



***Figure 2.7: Scouting 1D CPMG experiments on cαBC. a** 20°C, **b** 40°C. The reference spectrum (black) contained no CPMG block. The blue and red spectra are the results of experiments with a CPMG constant time block and CPMG pulse frequency as indicated.*

I carried out the experiment at four temperatures: 10°C, 20°C, 30°C and 40°C. The data for 20°C and 40°C are shown in **Figure 2.7**. At 20°C, no noticeable changes in intensity occurred when more refocusing pulses were applied (**Figure 2.7 a**), thus ruling out significant chemical exchange on the intermediate timescale. The data at 10°C were similar. At 40°C, weak changes in intensity were seen (**Figure 2.7 b**), e.g. for the peak indicated by an arrow. Thus some minor exchange might be taking place. The spectrum at 30°C was similar with a weaker difference in intensity of the indicated peak. This temperature range is close to physiological for this human chaperone, therefore this intermediate exchange, although minor, might have functional significance. It is investigated further in 2D CPMG experiments presented in **Chapter 7**. However, for the purposes of assignment at 15°C, peak broadening due to intermediate exchange clearly did not present a hurdle.

### 2.3.2 Assignment of the cαABC $^{15}N^{1}H$ NMR spectrum in salt-free buffer

Assignment experiments exploit magnetisation transfer between nuclei via scalar coupling to make links between atoms in adjacent residues. HNCA, HNCO and HN(CA)CO spectra were used for assignment (Yamazaki et al. [1994]). HN(CO)CA

and HCCH TOCSY experiments were also recorded but were of insufficient sensitivity to be useful.

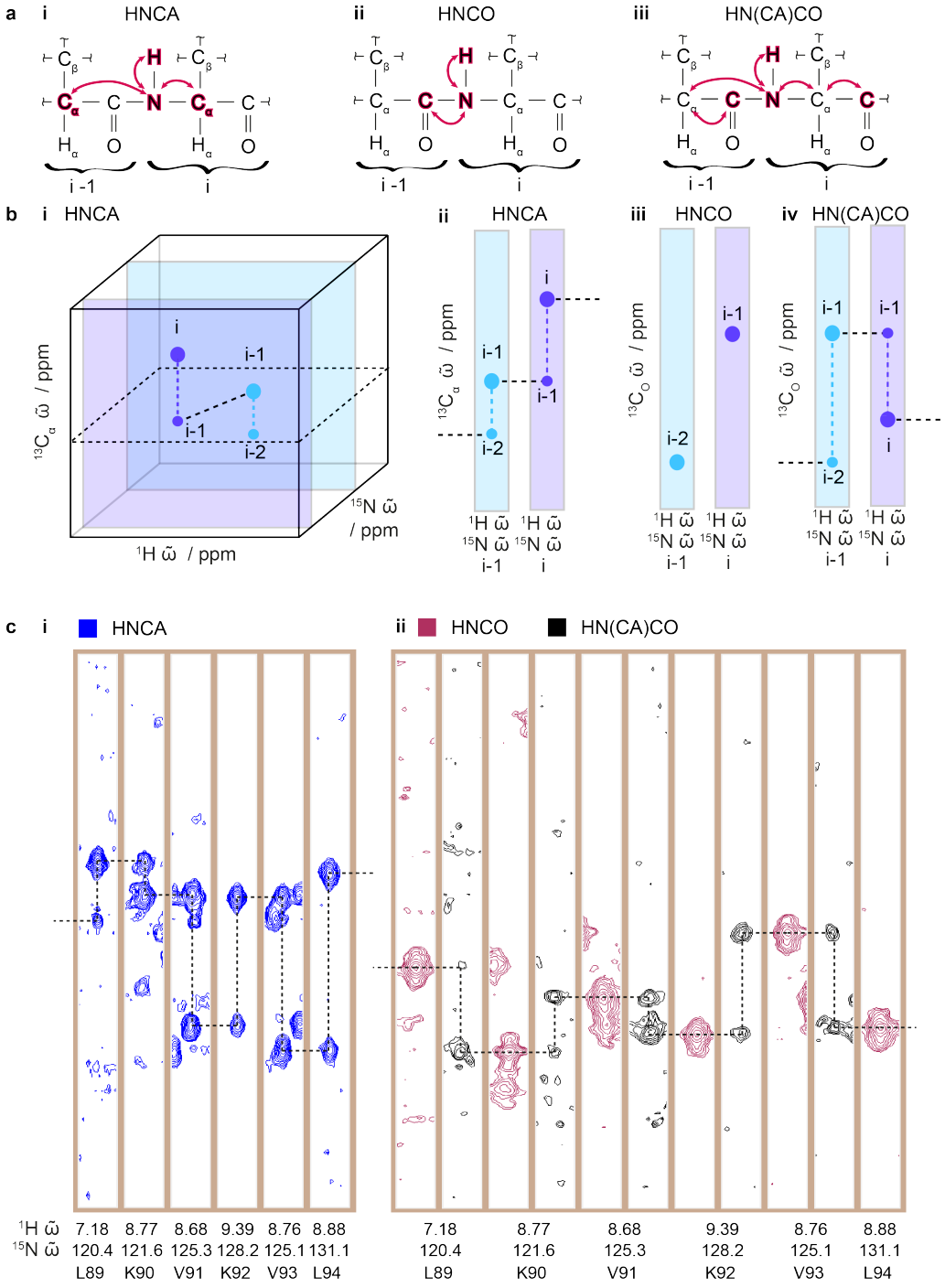



***Figure 2.8: The backbone assignment process.*** ***a*** *Schematic of the magnetisation transfer in the experiments used for assignment.* ***b i*** *Illustration of the relationship between peaks for neighbouring residues in the HNCA experiment.* ***ii*** *Schematic strip plots from an HNCA experiment corresponding to the diagram in* ***i.*** ***iii-iv*** *Analogous schematic strip plots for the HNCO and HN(CA)CO experiments.* ***c*** *Example series of experimental strip plots for a six-residue region in cαBC.*

**Figure 2.8a** illustrates the magnetisation correlations in each experiment I used. The HNCA experiment correlates the amide resonances with the $C_\alpha$ resonances of its own ($i$) and preceding ($i-1$) residue. The HNCO experiment correlates amides with the preceding carbonyl carbon, and the NH(CA)CO correlates the amide resonance with its own and preceding carbonyl carbon. There are thus two $C_\alpha$ peaks per amide in the HNCA experiment, one carbonyl carbon peak in the NHCO experiment and two carbonyl carbon peaks per amide in the HN(CA)CO experiment.

**Figure 2.8b** illustrates the assignment process. The spectra with two carbon resonances per amide can be used to build chains of resonances, since each carbon resonance appears twice: at its own $^1$H and $^{15}$N chemical shift and at the $^1$H and $^{15}$N chemical shifts of the preceding amino acid. The HNCO spectrum can be used to confirm which of the two resonances in the HN(CA)CO spectrum belongs to residue $i-1$ rather than residue $i$. An example of a chain shown in **Figure 2.8c**. The chains can be placed on the sequence of the protein using the characteristic chemical shifts of $C_\alpha$ atoms, documented in the BMRB database. For example, glycines have a distinctively high $C_\alpha$ chemical shift. Prolines do not possess an amide proton and therefore do not appear in the spectra. Breaks in the chains should therefore correspond to the locations of prolines in the sequence. The existing assignment (Jehle et al. [2009]) was used for confirmation in weak cases, where appropriate.

The assigned spectrum is shown in **Figure 2.9**. Of the 82 non-proline residues (also excluding the N-terminus, which usually has very high proton exchange rates with water and is therefore not visible), 80 were assigned. The remaining two peaks were undetectable in the spectra. **Figure 2.9b** shows the assigned and unassigned residues on the dimeric crystal structure of c$\alpha$BC.

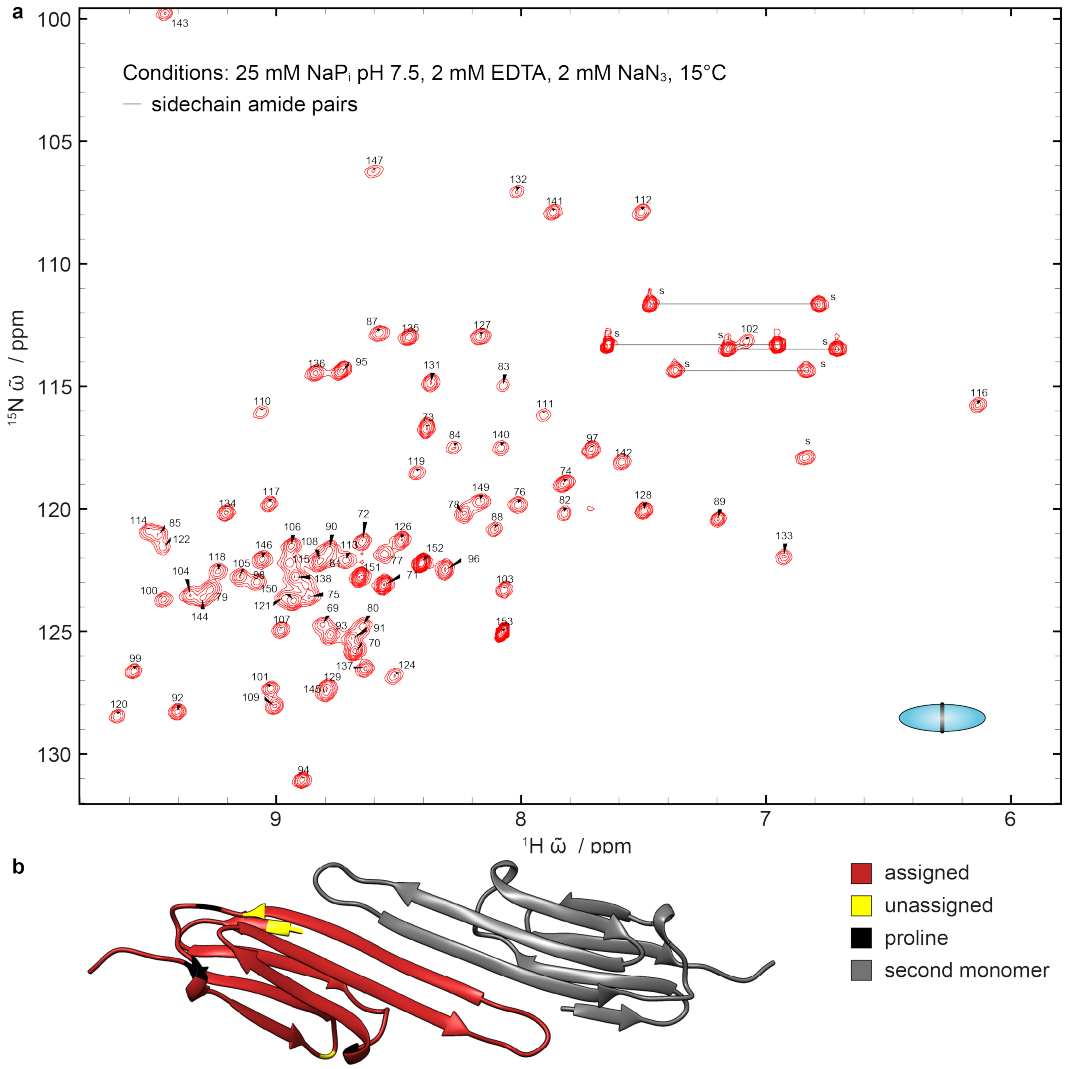



***Figure 2.9:*** *Backbone assignment of cαBC in salt-free buffer.* ***a*** *The assigned spectrum at 15°C. s - sidechain.* ***b*** *Extent of assignment illustrated on the cαBC dimer structure.*

### 2.3.3 Assignment of the cαBC $^{15}N^{1}H$ NMR spectrum in aggregation buffer

A large part of the work in this thesis concerns the function of αBC and cαBC as a chaperone in aggregation reactions. Aggregation assays required very specific buffer conditions to work reproducibly in a practical kinetic regime, in particular the presence of 100 mM salt and a temperature of 37°C. Therefore, I transferred as many assignments as possible from salt-free buffer to aggregation buffer conditions (see **Section 2.2** and **Figure 2.10** for detailed buffer composition).

For isolated peaks, I was able to transfer assignments easily. To be able to assign peaks in overlapped regions, I recorded an HNCA experiment in aggregation buffer and used it as described in **Figure 2.8** to deduce connectivities between residues and link the unassigned peaks to those where assignment was unambiguous.

16 peaks were missing in these conditions, significantly more compared to the salt-free buffer, which could be due to enhanced solvent exchange or conformational exchange. In addition, 13 well-defined peaks could not be unambiguously assigned even with the help of the HNCA. However, the remaining 53 peaks could be success-

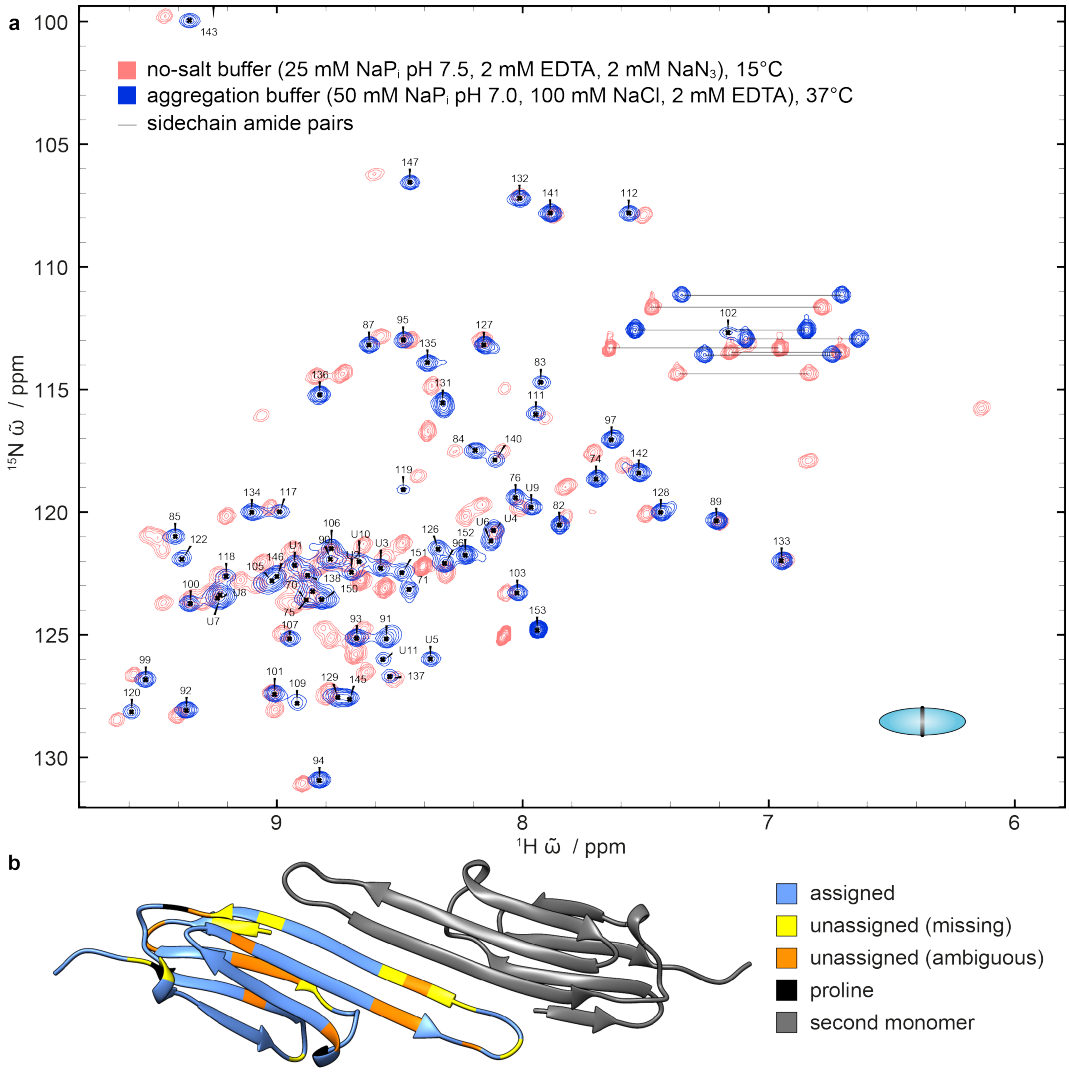



***Figure 2.10: Backbone assignment of cαBC in aggregation buffer.*** ***a*** *Assigned spectrum in aggregation buffer at 37°C, overlaid with the spectrum in salt-free buffer.* ***b*** *Extent of assignment in aggregation buffer illustrated on the cαBC dimer structure.*

fully assigned. The assignment together with mapping on the structure is shown in **Figure 2.10**.

### 2.3.4 Allosteric communication between interfaces in $\alpha$BC

As described in the introduction to this Chapter, previous native MS work has elucidated the existence of energetic compensation between the intra-dimer and inter-dimer interfaces in $\alpha$BC (see **Figures 2.1** and **2.2** in **Section 2.1**). Armed with the assignment of the backbone HSQC spectrum of c$\alpha$BC, I proceeded to explore the mechanism of communication between these interfaces.

The inter-dimer interaction can be mimicked by adding a peptide corresponding to part of the sequence of the $\alpha$BC C-terminus, which is known to bind to the core domain (**Figure 2.11a**, Delbecq et al. [2012], Hilton et al. [2013], Hochberg et al. [2014]). I used NMR to follow perturbations in residues across the core domain during the course of titration of two variants of the peptide, wild-type with the sequence ERTIPITRE and a mutant with the sequence EHTIPITRE. The mutation R157H occurs naturally in $\alpha$BC and is associated with dilated cardiomyopathy (Boncoraglio et al. [2012]). A different mutation in the same position, R157A, has been investigated *in vitro* and is known to result in a weaker inter-dimer interaction, with concomitant strengthening of the intra-dimer interaction (Hilton et al. [2013], see also **Figure 2.2**). I chose to use the natural mutant alongside the wild-type peptide, which should help discern key features of the communication network across c$\alpha$BC by showing which properties are essential for its operation.

In these experiments, c$\alpha$BC was at 200 $\mu$M and peptide was added at increasing concentrations in 11 (pWT) or 9 (pR157H) steps up to 2800 $\mu$M, i.e. 14-fold excess. Backbone HSQC spectra of c$\alpha$BC were recorded at each peptide concentration. Peaks in each spectrum in the presence of peptide were assigned as far as possible taking into account movement patterns throughout the titration. The overlay of the titration data for pWT is shown in **Figure 2.11b**.

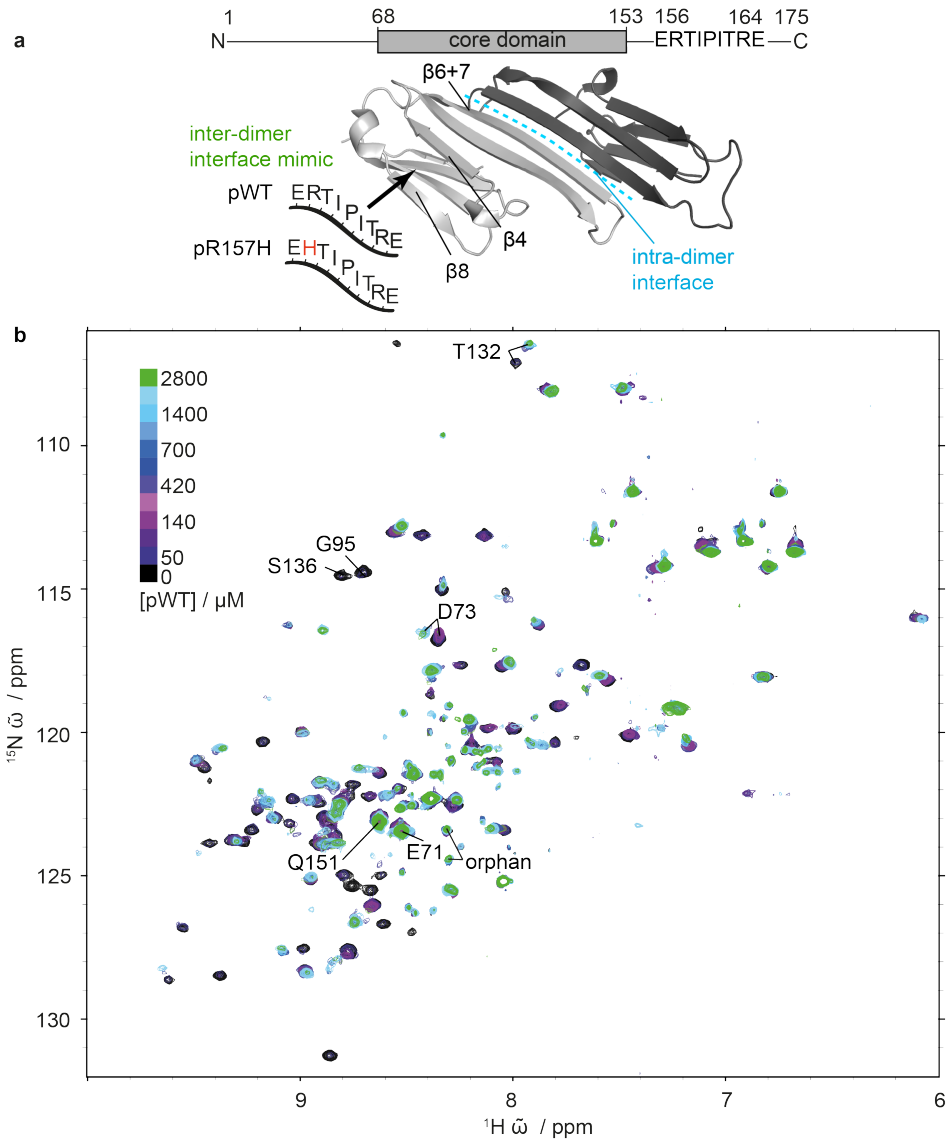



***Figure 2.11: Titration of C-terminus peptide to wild-type cαBC. a i*** *Schematic representation of αBC showing the boundaries of the core domain and the region of the C-terminus mimicked by the peptides.* ***ii*** *Illustration of peptide binding to the cαBC dimer in the groove formed by the strands β4 and β8. The sequence of the wild-type (pWT) and mutant (pR157H) peptides is shown.* ***b*** *Overlay of the titration spectra for pWT.*

Significant changes including peak movement and broadening occurred in the spectrum upon addition of increasing concentrations of peptide. Dramatic loss of intensity to noise level was observed for 22 peaks (14 for pR157H). Five peaks (six for pR157H) appeared at higher concentrations of peptide but could not be unambiguously assigned (e.g. orphans in **Figure 2.11b**). Nevertheless, evidently their number was not enough to account for all lost peaks. The behaviour of cαBC in the presence of pWT and pR157H was overall similar. Nevertheless, there were

some noticeable differences in peak positions and intensities. **Figure 2.12** shows an overlay of cαBC at 2100 μM of either peptide with annotation of the most prominent differences.

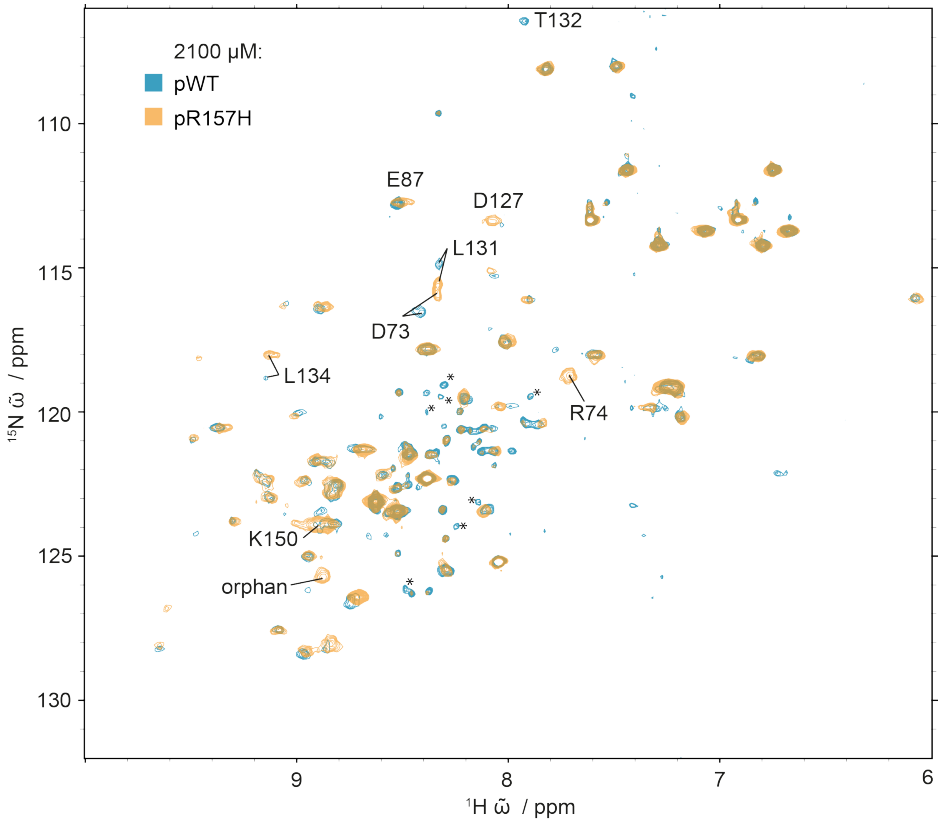



***Figure 2.12: Comparison of HSQC spectra of cαBC in the presence of pWT and pR157H.*** *Some of the peaks that show differences in behaviour are annotated. Peaks marked with asterisks between 8.2-8.4 ppm in $^{1}H$ in the pWT spectrum were also present at 0 μM peptide; they likely correspond to minor degradation products of cαBC and should not be considered for this comparison.*

In what follows, I analyse two aspects of the titrations in turn: kinetic and thermodynamic features of the behaviour observed in the above spectra; and spatial clustering of the residues most strongly affected by peptide binding.

### Kinetic and thermodynamic modelling

A range of peak behaviours was observed for different residues over the course of the titration. Representative examples from the pWT titration are shown in **Figure 2.13a**. Most residues showed small peak shifts (group (**i**)). Several residues, mostly those in the β4-β8 groove, underwent severe peak broadening, being lost at 140 μM

peptide (group (**ii**)). A significant fraction of residues demonstrated features of slow exchange, i.e. loss of intensity of the original peak and gain of intensity of a new peak (group (**iii**)). Some residues also showed signs of heterogeneity, e.g. appearance of several peaks and non-linear peak movement (**iv**). The location of peaks in these groups on the c$\alpha$BC structure is shown in **Figure 2.13b**).

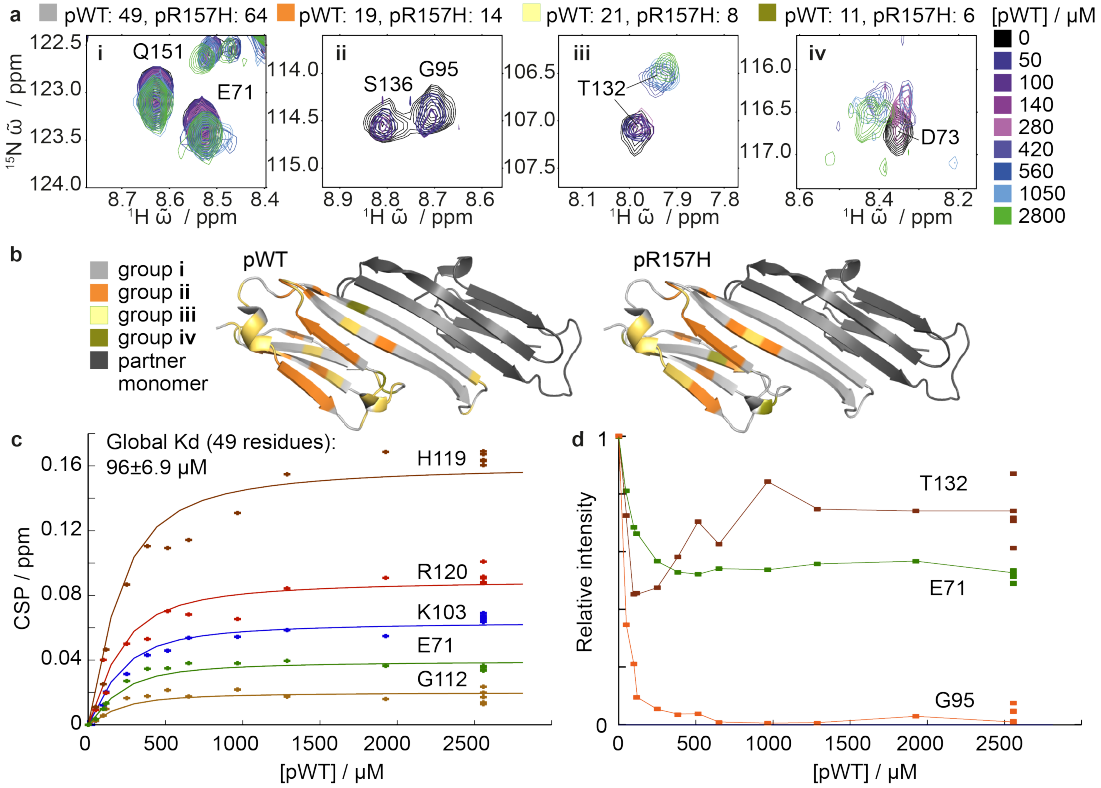



***Figure 2.13: Analysis of peak behaviour in the course of the peptide-c$\alpha$BC titration.*** ***a*** *Examples of types of behaviour in NMR spectra: (**i**) linear peak shifts; (**ii**) dramatic loss of intensity; (**iii**) slow-exchange type behaviour; (**iv**) nonlinear peak movement and appearance of multiple peaks. Peak count in each group is shown above the boxes for both titrations.* ***b*** *Location of peaks in groups above on the structure of c$\alpha$BC dimer.* ***c*** *Chemical shift changes of peaks in group **i** can be fit with assumption of fast exchange (peak position is a population average of free and bound states).* ***d*** *Examples of peak intensity changes. Residues with linear peak shifts (e.g. E71) also show significant changes in intensity, therefore the analysis in **c** is an oversimplification.*

A key question is what structural and dynamic transitions these behaviours are reflecting. The pattern of chemical shift and intensity changes for a given peak is governed by the relative magnitude of the exchange frequency between the different states, e.g. free and peptide-bound c$\alpha$BC, and the chemical shift difference between the states (see **Figure 1.10** in **Section 1.5.1**). Exchange between states can be

global - e.g. the entire molecule is in principle either free or peptide-bound. The chemical shift perturbation is, however, unique to each residue, as it depends on its physical environment. Therefore it is possible that disparate peak behaviours are due to residue-specific variation in chemical shift and underlying chemical exchange is governed by a single set of kinetic and thermodynamic parameters for the different groups.

What exchange states should be considered in this titration? c$\alpha$BC reversibly forms a symmetric dimer, and any monomer can bind a peptide. Previous work (see **Figure 2.2a**, Hilton et al. [2013]) showed that (a) free monomers and dimers can be considered to bind peptide equivalently, (b) binding a second peptide to a dimer is less favourable, and (c) the presence of two peptides affects the dimerisation equilibrium. Thus we have five distinct states of the core domain, $(c\alpha BC)$, $(c\alpha BC)_2$, $(c\alpha BC)(Pep)$, $(c\alpha BC)_2(Pep)$, $(c\alpha BC)_2(Pep)_2$, where $P$ is peptide, and five known reactions are taking place:

$$2(c\alpha BC) \rightleftharpoons (c\alpha BC)_2 \tag{2.1}$$

$$(c\alpha BC) + (Pep) \rightleftharpoons (c\alpha BC)(Pep) \tag{2.2}$$

$$(c\alpha BC)_2 + (Pep) \rightleftharpoons (c\alpha BC)_2(Pep) \tag{2.3}$$

$$(c\alpha BC)_2(Pep) + (Pep) \rightleftharpoons (c\alpha BC)_2(Pep)_2 \tag{2.4}$$

$$2(c\alpha BC)(Pep) \rightleftharpoons (c\alpha BC)_2(Pep)_2 \tag{2.5}$$

Each of these could in principle proceed via intermediates, therefore additional states could be postulated. In the extreme, each residue in the core domain could be considered to behave independently in each of these transitions. Nevertheless, one must start by testing the simplest possible model first and increase complexity to the degree justified by the data at hand, using statistical tools such as the F-test. The physical processes that are discerned in the NMR analysis may not be exactly the same as those detectable by other approaches; different simplifications may be appropriate to different methods and we may discover new features of the system

that are essential to reflect both the newly available and previous data.

In practice, therefore, the first question to ask is whether the titration data can be explained with the most simple, two-state exchange model:

$$(c\alpha BC) + (Pep) \underset{k^-}{\overset{k^+}{\rightleftharpoons}} (c\alpha BC)(Pep) \tag{2.6}$$

Here cαBC is treated as a set of independent monomers which can be either free ($c\alpha BC$) or bound ($c\alpha BC(Pep)$). $k^+$ and $k^-$ are the second-order and first-order rate constants describing the forward and reverse reactions respectively. To begin to explore whether this model applies, we can look at residues that show linear peak shifts (group (**i**), **Figure 2.13a**). For residues in fast exchange, where the sum of the forward and reverse exchange rates between free and bound states is much greater than the difference in chemical shift frequency (in rads$^{-1}$), the position of the peak is a population-weighted average of the two states. Therefore the CSP can be treated as the fraction of the molecules that are in the bound state ($\frac{[(c\alpha BC)(Pep)]}{[(c\alpha BC)]_{total}}$), and the curve describing the variation of CSP with peptide concentration can be fitted to the quadratic equation relating this fraction to $[(Pep)]$, with $K_D = \frac{k^-}{k^+}$ as a fitting parameter. I performed this analysis on peaks in group (**i**), with some examples shown in **Figure 2.13c**. The data conformed to the model well and gave a global $K_D$ for 96±6.9 $\mu$M for pWT and 88±5.7 $\mu$M for pR157H.

This fast exchange model assumes constant intensity of the peaks. In fact, almost all peaks showed significant variation in intensity during the titration. Common peak intensity patterns are shown in **Figure 2.13d**. Most peaks in group (**i**) showed partial loss of intensity, similar to peak E71, sometimes with recovery of intensity at peptide concentrations over 1000 $\mu$M. Therefore the fast exchange assumption in **Figure 2.13c** is not justified, even though the CSP data fits the model well, and more sophisticated modelling is required to explain their behaviour. In addition, the other major groups of residues, (**ii-iv**), with intensity variation represented by G95 and T132 in **Figure 2.13d**, are clearly not in two-state fast

exchange.

The most comprehensive approach to testing kinetic models is to simulate the data using the Bloch-McConnell equations, which describe evolution of magnetisation from first principles with chemical shift, relaxation and chemical exchange incorporated into the calculation (see **Section 1.5.1**, McConnell [1958]). This circumvents the need to assume exchange regimes as they come out naturally from the simulation with appropriate parameters. A software package to perform this type of analysis is under development in our laboratory, with the code written by Prof. Andrew Baldwin. Full analysis of the data with this approach is ongoing and preliminary results are shown here in **Figure 2.14**.

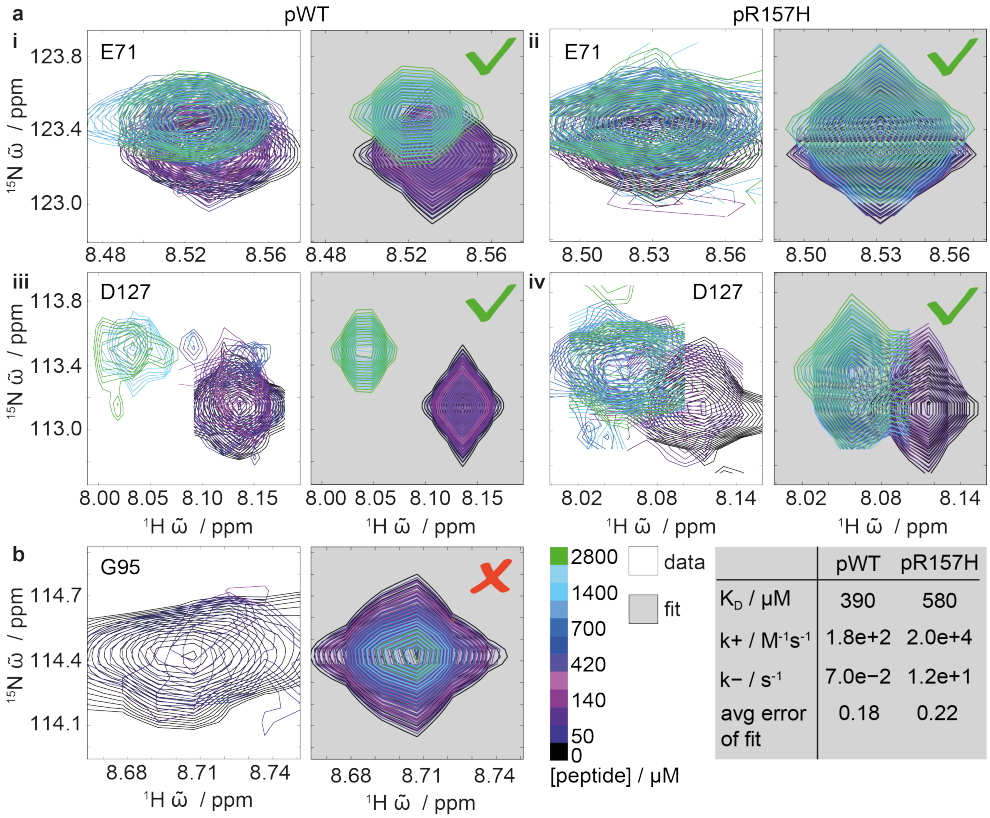



***Figure 2.14: Exchange analysis from peak shape fitting.*** ***a*** *Examples of fast and slow exchange type peaks fit with the same two-state model.* ***b*** *Peaks with dramatic intensity loss, such as the example shown here, do not conform to the model in* ***a****. The preliminary optimised parameters of the models for pWT and pR157H are shown on the bottom right.*

The algorithm performs a simulation of the data with given exchange parameters, in the case of a two-state model $K_D$ and $k^+$, transverse relaxation constants

$R_2$ for both states and chemical shift difference $\Delta\tilde{\omega}$ between the states. It then uses Levenberg-Marquardt least squares optimisation (see **Section 1.5.2**) against the extracted raw peak shape data over the course of the titration to find best-fit values of $K_D$ and $k^+$.

Here we performed the analysis with the initial $\Delta\tilde{\omega}$ set from the 1D analysis of the peaks, of the type shown in **Figure 2.13b**, and $R_2$ of both states fixed at $15\,s^{-1}$, a reasonable estimate for a protein of 10-20 kDa (see e.g. values measured in **Figure 7.7** in **Chapter 7**). The initial values of $K_D$ and $k^+$ were adjusted manually. We performed the fit with $K_D$ and $k^+$ as global parameters on residue groups (**i**) and (**iii**), i.e. fast-exchange-like and slow exchange residues. We found that residues from both groups combined could be well fit to a single two-state exchange model, with examples shown in **Figure 2.14a**. The parameters were different for the pWT and pR157H datasets, with $K_D$ of $390\,\mu$M and $580\,\mu$M and $k^+$ of $1.8\times10^{+2}\,s^{-1}$ and $2.0\times10^{+4}\,s^{-1}$, making $k^-$ $7.0\times10^{-2}\,s^{-1}$ and $1.2\times10^{+1}\,s^{-1}$ respectively. Thus interestingly the forward rate of the transition is over 100-fold faster for the mutant but the reverse rate is nearly 200-fold faster, making the $K_D$ less favourable.

We tested this model on the remaining set of residues that showed dramatic intensity decay (group (**ii**) in **Figure 2.13a**). To account for lack of peak shifts or rising partner peaks, we fixed the $\Delta\tilde{\omega}$ for these residues at an arbitrary large value, 4 ppm. The simulation thus fit only the decay in the intensity of the initial state. We found that the model did not adequately describe the behaviour of these residues: the example fit shown in **Figure 2.14b** demonstrates that this model predicts much more gradual intensity loss of the original peak, with some intensity visible even at the highest concentration of peptide, $2800\,\mu$M, which is clearly not supported by the data. Thus, a single two-state model with two adjustable parameters is likely not an adequate representation of the whole dataset.

The pWT $K_D$ is higher than the range of values previously determined for first

and second peptide binding to the cαBC dimer - 70-105 $\mu$M (Hilton et al. [2013], Delbecq et al. [2012]). The error on the fit of group (**i**) and (**iii**) residues is large and some optimisation of the fits must be carried out, by testing a wider range of initial fitting parameters and adjusting the peak shape. The results presented here nevertheless suggest that cαBC is undergoing at least a three-state transition upon peptide binding, with the groove peaks broadening at low peptide concentrations due to the first peptide binding event and a second conformational change occurring with a higher $K_D$, potentially due to the second peptide binding event:

$$(c\alpha BC)_2 + (Pep) \rightleftharpoons (c\alpha BC)_2(Pep) \tag{2.7}$$

$$(c\alpha BC)_2(Pep) + (Pep) \rightleftharpoons (c\alpha BC)_2(Pep)_2 \tag{2.8}$$

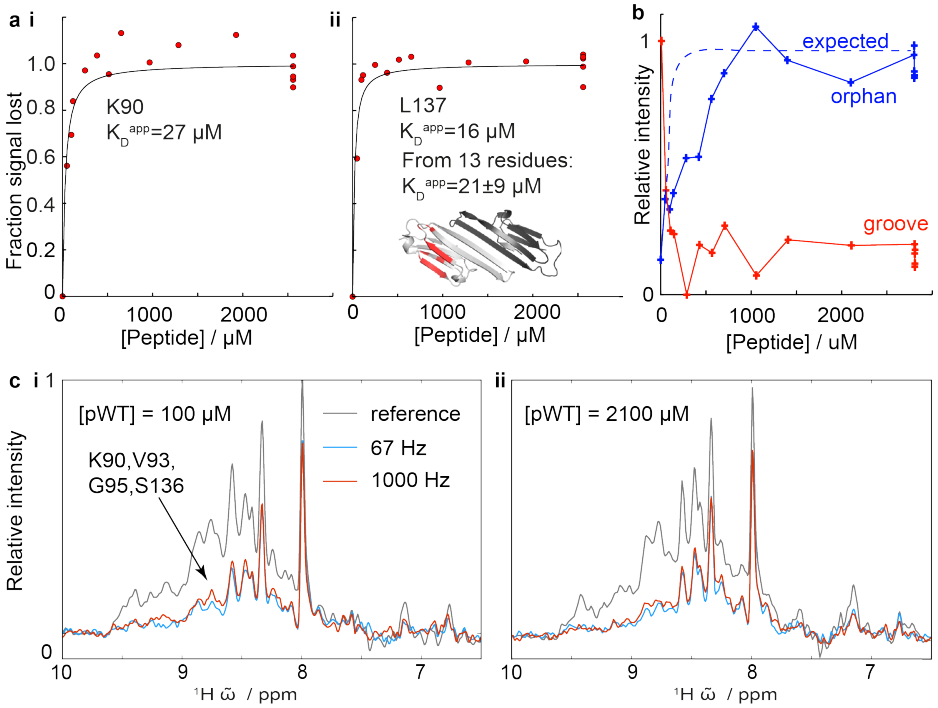



***Figure 2.15: Probing the behaviour of the groove peaks.*** ***a*** *Apparent $K_D$ fits to peak intensity versus peptide concentration assuming slow exchange.* ***b*** *Comparison of the concentration dependence of intensity for a representative orphan peak and groove peak.* ***c*** *1D CPMG experiments at* ***i*** *100 $\mu$M pWT and* ***ii*** *2100 $\mu$M pWT.*

What is the best way to model the group (**ii**) groove peaks? They are clearly

not in fast exchange as they show dramatic intensity loss with increasing peptide concentration. The treatment above followed the hypothesis that they are in slow exchange with a large $\Delta\tilde{\omega}$. In this limit the relative intensity of the original peak corresponds to the population of molecules in the initial state, and the fraction of signal lost is then the fraction $\frac{[(c\alpha BC)(Pep)]}{[(c\alpha BC)]_{total}}$. It can be fit to obtain the $K_D$ of the c$\alpha$BC-Pep interaction, by analogy to the treatment of CSPs in fast exchange (see **Figure 2.13b**). This was possible only for pWT, where data was available at 50 $\mu$M and 100 $\mu$M to define the fit (**Figure 2.15a**). I obtained a $K_D$ of 21$\pm$9 $\mu$M by this method. It is lower than the smallest peptide binding $K_D$ obtained in the MS analysis, which was 70 $\mu$M (Hilton et al. [2013]). Thus it is likely that there is some intermediate exchange contribution to the peaks and the value of 21$\mu$M represents a lower limit on the $K_D$ rather than its true value.

Another consequence of these peaks being in slow exchange is that complementary rising peaks must be observed elsewhere in the spectrum. There were only 5 such peaks versus 19 that were lost in the pWT titration. In addition, the pattern of rise in intensity of the orphan peaks with concentration did not match the loss in intensity of the groove peaks, with an example shown in **Figure 2.15b**. The groove peak intensity is at noise level at 100 $\mu$M peptide, whereas the orphan peak reaches maximum only around 700-1000 $\mu$M peptide. It is thus more likely that the orphan peaks are additional states of group (**iii**) residues (compare e.g. intensity pattern of T132 in **Figure 2.13c**) rather than partner peaks of the residues in group (**ii**).

Support for an intermediate exchange contribution to the group (**ii**) peaks comes from 1D CPMG data recorded in the course of the pWT titration. The data for two concentrations of peptide, 100 $\mu$M and 2100 $\mu$M, is shown in **Figure 2.15c**. In these diagnostic experiments, intermediate exchange is detected as an increase in intensity in the spectrum with 1000 Hz frequency of CPMG pulses relative to the 67 Hz spectrum. This is the case for an area where several groove peaks cluster (indicated by an arrow, see also the 2D spectrum in **Figure 2.9**). The effect is clearer at the lower peptide concentration. This is as expected, since at higher peptide

concentration the exchange rates would be faster, pushing the residues towards the fast exchange regime.

Even in intermediate exchange, the peaks would be expected to reappear at high concentrations of peptide. An explanation for their absence could be the heterogeneity of the bound state, observed directly for some of the other peaks (see e.g. D73 in **Figure 2.13 iv**, where even at the highest peptide concentration there is evidence for at least two overlapped peaks). If these residues are in several environments, the intensity of each peak may be insufficient to observe them. Alternatively these residues could be dynamic on the $\mu$s-ms timescale in the bound state. Intermediate exchange with an alternative state populated to only a few percent could significantly broaden the peaks (Palmer [2014], Anthis and Clore [2015]).

In summary, behaviour of most c$\alpha$BC peaks in pWT and pR157H titrations can be modelled as a two-state transition with the $K_D$ in the 100s of $\mu$M. However, the residues in the groove show a dramatic loss of peak intensity and do not conform to the above model. They are likely to undergo intermediate exchange broadening for at least some peptide concentrations. Thus the NMR data suggests at least two processes with separate kinetic and thermodynamic parameters: a binding event at the groove and a more widespread conformational conversion across the domain, potentially as a result of a second peptide binding to the groove of the partner

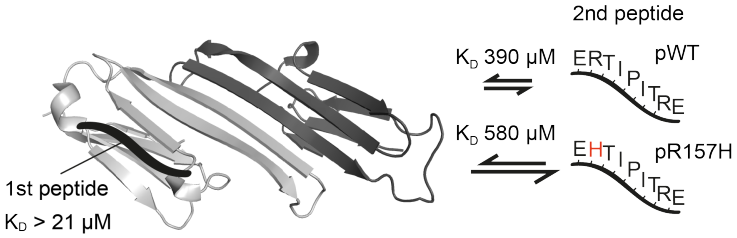



***Figure 2.16: Summary of thermodynamic parameters of peptide binding to c$\alpha$BC.*** *From dramatic broadening of the peaks in the $\beta$4-$\beta$8 groove and the slow-exchange limit $K_D$ fit, the first peptide binding event happens with a $K_D$ greater than 21 $\mu$M. The rest of the peaks show peak and intensity shifts that can be described with a single, higher $K_D$, most likely corresponding to a peptide binding event on the partner monomer. The mutant peptide pR157H has a faster forward rate for this second transition but an even faster reverse rate, making the $K_D$ of this second transition higher than that of pWT.*

monomer (**Figure 2.16**).

### Network analysis

In this section, which is independent from the kinetic analysis above, I aim to discern spatial patterns in the residues most stronly affected by peptide binding, i.e. to define the network of interactions that links the two interfaces in c$\alpha$BC. To this end, I compared the spectra at 0 and 2800 $\mu$M pWT or pR157H (illustrated in black and green in **Figure 2.11** for pWT). I measured the maximum chemical shift perturbations for residues where these were detectable (groups (**i**) and (**iii**) in **Figure 2.13**), and included the peaks that show dramatic intensity loss (group (**ii**)).

The CSPs between the spectra with no peptide and the highest concentration of peptide (2800 $\mu$M) are shown in **Figure 2.17a** and mapped on the structure of c$\alpha$BC in **Figure 2.17b**. I use the crystal structure of free c$\alpha$BC in the APII register (Bagnéris et al. [2009]) rather than the peptide-bound structure in APIII register (Hochberg et al. [2014]), since the predominant register in solution is APII and the binding peptide must exert influence on contacts that are present in the apo protein. As observed previously, the severely broadened residues are located on the strands $\beta$4 and $\beta$8, i.e. in the peptide binding site, but notably two are in the dimer interface: Y122 and I124. Significant CSPs are observed in the loops between $\beta$6+7-$\beta$8 and $\beta$2-$\beta$4, as well as in the $\beta$9 strand, which is located next to the $\beta$8 strand. Notably, little change is observed from residue 103 to 117, the tail end of $\beta$5 and first half of the dimer interface strand $\beta$6+7, with the exception of residue F113. Therefore it is more likely the change is transmitted to this residue across the intra-dimer interface rather than the $\beta$5 and $\beta$6+7 strands in the same chain.

**Figure 2.17c** illustrates the location of the few residues that show notable differences in CSP between the two peptides. The structure used for this figure is the complex between c$\alpha$BC and pWT (Hochberg et al. [2014]). Residues that

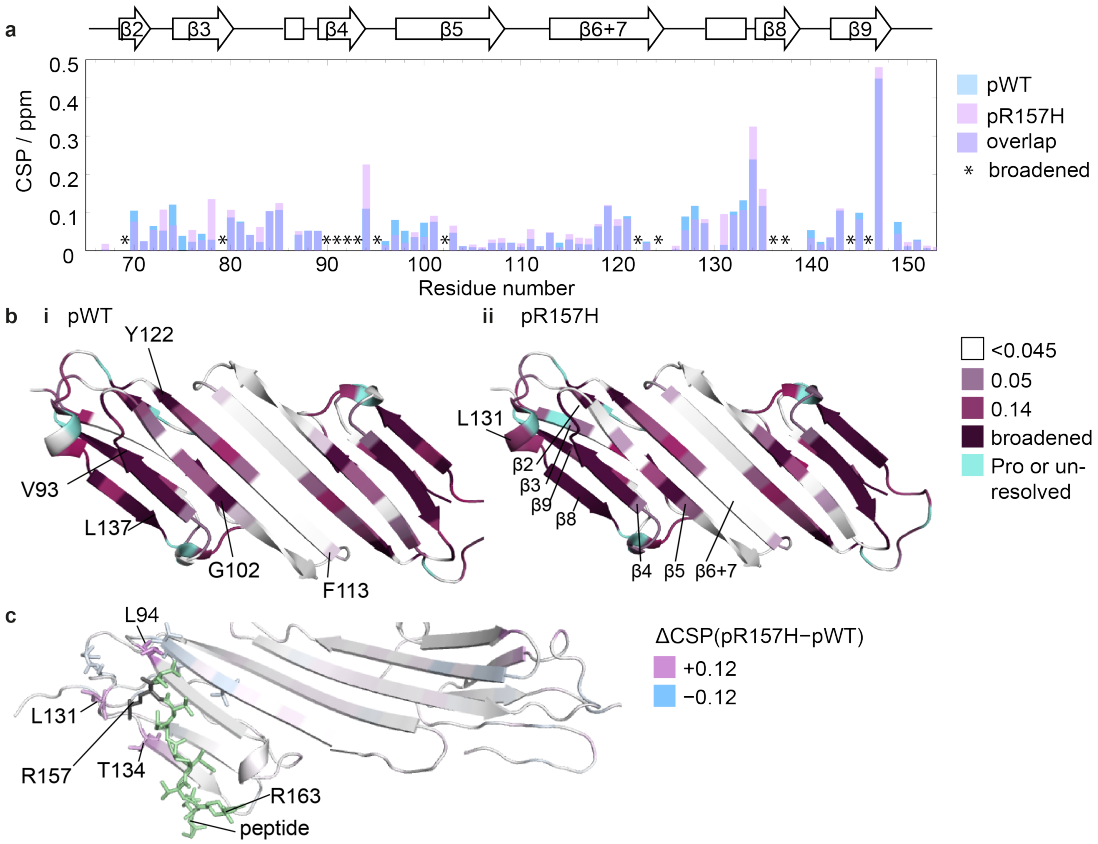



***Figure 2.17: Location of chemical shift perturbations.*** ***a*** *Residue CSPs for titrations of pWT and pR157H, measured between 0 and 2800 μM of peptide. The secondary structure cartoon is shown above the bar chart. Asterisks denote severe peak broadening.* ***b*** *Mapping of the CSPs in* ***a*** *onto cαBC crystal structure in register APII (PDB ID 2WJ7) for* ***i*** *pWT and* ***ii*** *pR157H. Some of the residues with significant CSPs are annotated.* ***c*** *Mapping of relative differences in CSP for the two peptides onto the cαBC crystal structure with the C-terminal wild-type peptide (PDB ID 4M5S). The mutated residue R157 is annotated together with cαBC residues that show the biggest changes (also shown as sticks).*

show greater CSP with pR157H are clustered around one end of the peptide and are in fact the sidechains that surround one of the arginines in the palindromic pWT, ERTIPITRE. No ΔCSP are observed around the other arginine in pWT, R163, also annotated on the structure. Thus the mutant peptide EHTIPITRE binds predominantly in one orientation. Overall, the ΔCSP values are small and the general pattern of CSPs is remarkably similar for pWT and pR157H (see **Figure 2.17a**). Thus both pWT and pR157H affect mostly the same set of residues across the core domain up to the intra-dimer interface.

What are the principal pathways through which communication occurs? To

answer this question, I took into account physical distances between residues as well as the location of maximum CSPs: strongly perturbed residues that are in physical contact should transmit information to each other. Although the experiments I recorded were backbone HSQC, I incorporated sidechains into the analysis of contacts, since it is highly likely they are also affected by peptide binding and predictions about their involvement can be tested using appropriate NMR experiments.

**Figure 2.18a** shows residues that are within 4 Å of each other[1], linked by cylinders coloured by the chemical shift of each residue at each end. For clarity, although distances were calculated between any backbone or sidechain heteroatom, the cylinders in this representation are drawn between C$\alpha$ atoms, and only residues with CSP>0.045 are shown.

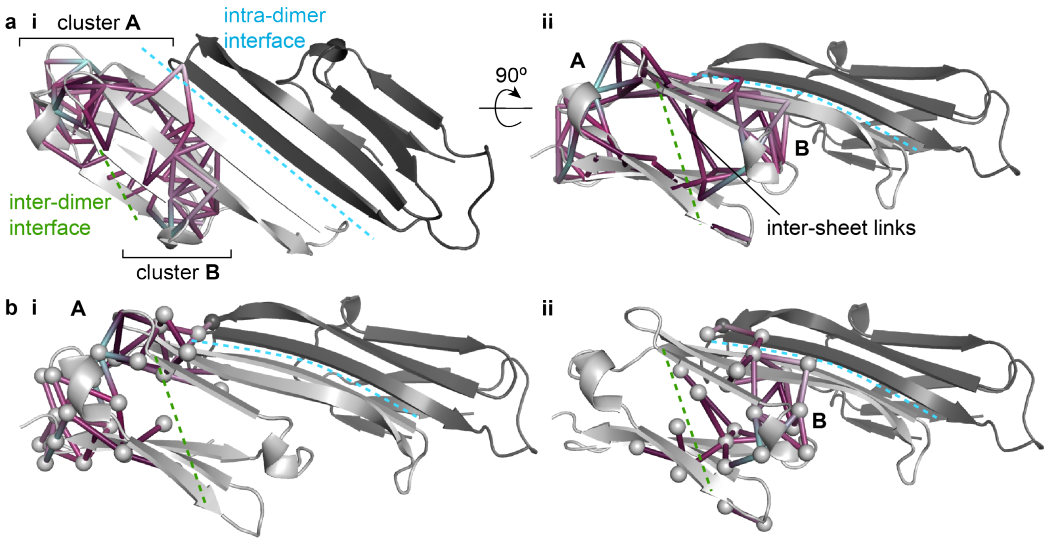



***Figure 2.18: The interface communication network in cαBC.*** ***a*** *The network of residues that are within 4 Å of each other and show significant CSPs, in two orientations. The residues are linked with cylinders coloured by CSP at each end. The colour scale is as in* ***Figure 2.17b****. The intra- and inter-dimer interfaces are shown as dashed lines and two clusters of links are annotated as A and B.* ***b*** *Individual clusters* ***(i)*** *A and* ***(ii)*** *B with the Cβ atoms of each amino acid shown as spheres, to illustrate the orientation of sidechains.*

It is clear from the figure that there is a redundant network of residues with

[1]The distance threshold of 4 Å between heteroatoms was chosen based on covalent and van der Waals radii: the length of a CH bond is 1.14 Å and the van der Waals radii of C and H are 1.68 Å and 1.11 Å respectively (Anslyn and Dougherty [2006]). Therefore the closest theoretical distance between two carbons is about 3.4 Å, and two CH bonds meeting head-on would be at about 4.5 Å of each other. Thus it is reasonable to assume that residues whose heteroatoms are within 4 Å are in immediate contact.

CSPs that links the inter-dimer and intra-dimer interfaces. There are two groups of connections, termed A and B. They cluster around each set of loops on the ends of the $\beta$-sandwich, as seen more clearly in the side view (**Figure 2.18a ii** and **b**). The side view also shows that both the top and bottom sheets of the $\beta$-sandwich are involved in the network and that there are links between the two clusters through the interior of the sandwich. Panel **b** shows the A and B clusters separately with connections drawn between C$\beta$ atoms of each residue, in order to indicate the orientation of the sidechain. The figure demonstrates that residues with sidechains on both the interior and exterior of the sandwich form part of the network.

Interestingly, the CSP network in one monomer converges to the tip of the intra-dimer interface (seen best in **Figure 2.18a i**). Equivalent CSPs occur in the second monomer, impinging on the other tip. Thus the intra-dimer interface is affected from both ends rather than from its centre. Previous native MS work showed that the intra-dimer interface is destabilised by stronger inter-dimer interactions (see **Figure 2.2**, Baldwin et al. [2011d], Hilton et al. [2013]). It is more energetically unfavourable to break the interface from the centre, where the monomers are held in place by neighbouring interactions and the bonds can easily reform. Disruption of edge interactions is less energetically costly, which may be why the dominant intra-dimer interface CSPs are located there. The $\beta$5-$\beta$6+7 loop also shows a range of conformations in X-ray and NMR structures of c$\alpha$BC obtained under different conditions (Clark et al. [2011]), underlining particular flexibility of this region and therefore a potentially easy route into unzipping the interface.

**Figure 2.19a** shows CSP network residues with the maximum number of connections (six in addition to the two sequence neighbours). They are located in loops: I124 and V97 in cluster A and F84 in cluster B. Their mutation should not disrupt the structure of c$\alpha$BC, due to their location in flexible regions, but should significantly perturb the communication network. Therefore their importance for interface communication could be tested by mutagenesis experiments.

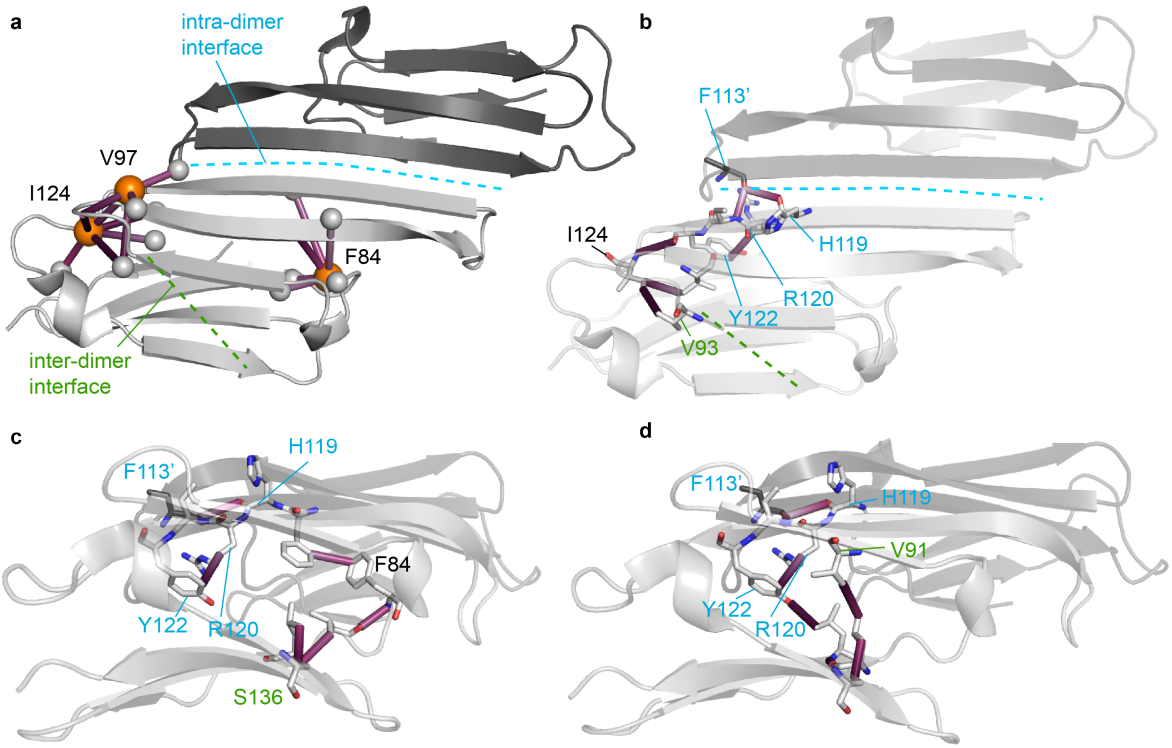



***Figure 2.19: Key network residues.*** ***a*** *Hub residues with the maximum number of connections (primary structure neighbours plus six): F84, V97, I124. The residues are depicted as orange spheres at Cβ locations. Close neighbours are shown as grey spheres at Cβ locations and the connections are coloured by chemical shift.* ***b, c, d*** *Examples of continuous paths between interfaces via residues with significant chemical shift. Residues are shown as sticks and 4 Å connections are shown as cylinders, coloured by chemical shift of the backbone amide at each end as in* ***Figure 2.17****. The intra-dimer interface residues are annotated in cyan: H119, R120, Y122 and F133', the prime symbol indicating this residue is located in the second monomer. Note that the sidechain of F113' was not resolved fully in the crystal structure. The inter-dimer interface residues are annotated in green and the hub residues in black.*

Most residues in the strands β4 and β8 are strongly affected by peptide binding (see **Figure 2.17** and **2.18**). There are thus many starting points for paths that link the inter-dimer and intra-dimer interfaces. **Figure 2.19b-d** shows three examples of interaction chains that start at different points on the inter-dimer interface with details of sidechain and backbone contacts. Two of the examples go via the hub residues I124 or F84, and the last goes mostly through the interior of the sandwich. As already shown above, the pathways converge on the same set of intra-dimer interface residues: H119, R120, Y122 in the same monomer and F113' in the partner monomer (the dash denoting a different potein chain). R120 is known to

be a key residue whose mutation significantly affects structure and function of $\alpha$BC (Bova et al. [1999], Clark et al. [2011], Boncoraglio et al. [2012], Boelens [2014]). **Figure 2.19b-d** shows that the elbow of the R120 sidechain is next to the ring of Y122, thus motion of these residues would affect sidechain packing. These predicted sidechain contacts can be analysed directly with NOESY HSQC or HCCH TOCSY experiments. R120 sidechain points towards the partner monomer and forms a salt bridge with D109'. The $\beta$5-$\beta$6+7 loop residue D109 does not show a significant CSP in the backbone HSQC performed here, but the backbone could be less affected by the perturbation of the salt bridge than the sidechain. R120 also forms an interchain H-bond to F113' via the backbone, and the backbone HSQC data presented here offer direct evidence of this interaction being affected by binding of the peptide.

In summary, the network of interactions affected by the peptides spans the core domain and relies in particular on the loops at both ends of the $\beta$-sandwich. There are many redundant pathways by which information could be transmitted across the core domain, with residues I124, V97 and F84 acting as hubs. The pathways converge on a key interchain contact, R120-F113', at the edge of the intra-dimer interface. The mutations R120G and R157A (and therefore likely R157H) are known to cause an increase in the average size of the oligomeric distribution of $\alpha$BC (Bova et al. [1999], Hilton et al. [2013]). The energetic feedback between the intra-dimer and inter-dimer interfaces suggests that this outcome is achieved through reverse mechanisms: weakening of the intra-dimer interface in $\alpha$BC R120G with compensatory strengthening of the inter-dimer interface, and vice versa for the R157A mutant. The results presented here show how the two interfaces and the residues in question are linked to enable this information transfer.

## 2.4 Conclusions

c$\alpha$BC is very amenable to analysis by solution state NMR. It does not show evidence of significant aggregation for concentrations up to 1 mM and does not suffer from

extensive exchange broadening. Backbone assignment of 80 out of 83 residues was achieved at 15°C in salt-free buffer and assignment of 53 residues was possible at 37°C in phosphate buffer with 100 mM NaCl. Using titration of two peptide variants of C-terminal sequence of $\alpha$BC, wild-type and the mutant R157H, I explored the mechanism of crosstalk between the inter-dimer and intra-dimer interfaces in $\alpha$BC. Taken together with previous native MS results, the data indicate that C-terminus binding in the $\beta$4-$\beta$8 groove destabilises the intra-dimer interface almost 20 Å away through a network of pathways which travel mostly through the inter-strand loops, and converge on the edges of the $\beta$6+7 strands, involving the key disease-relevant residue R120G. The activation of the network proceeds in at least two stages, one of which involves the $\beta$4-$\beta$8 groove residues and the other, with a higher dissociation constant, is more wide-ranging. These results demonstrate the detailed insight that can be obtained from NMR titration experiments and shed light on how elegant information transfer can occur in a protein molecule across a $\beta$-sheet structure, typically seen as rigid, in order to regulate structure and dynamics of a quaternary assembly.

# Chapter 3

# Structure and dynamics of full-length $\alpha$BC

*This Chapter presents insights into the dynamics and assembly of full-length $\alpha$BC oligomers obtained using native mass spectrometry and ion mobility. I demonstrate that subunit exchange is significantly slower if a covalent crosslink is introduced at the intra-dimer interface, which reveals that subunit exchange must proceed with significant involvement of monomers rather than exclusively dimers. Secondly, disulfide crosslinking at the inter-dimer interface shows that up to 50% of subunits form vertices of order three or higher, which is consistent with polyhedral geometry of the oligomers. Finally, I describe the application of a post-ion mobility collision-induced dissociation approach to directly measure collisional cross-sections for individual stoichiometries of $\alpha$BC within the polydisperse ensemble. Remarkably, the measurements show a continuous linear scaling relationship of collisional cross-sections with stoichiometry for both odd and even oligomers.*

## 3.1 Introduction

$\alpha$BC, like most metazoan sHSPs, exists as large, polydisperse and dynamic ensembles of oligomers. While the oligomeric nature of $\alpha$-crystallin was discovered early with electrophoresis, size exclusion chromatography and centrifugation techniques (Resnik [1957], Spector and Katz [1965], Mok and Waley [1967], Spector et al. [1971], Spector [1973]), subunit exchange was detected later (Siezen and Owen [1983], van den Oetelaar et al. [1990]) and quantified using fluorescence (Bova et al. [1997]). Shortly afterwards, native MS provided definitive evidence for polydispersity of $\alpha$BC and enabled detailed characterisation of the oligomeric distribution and measurement of subunit exchange rates on untagged protein (Sobott et al. [2002], Aquilina et al. [2003, 2005], Baldwin et al. [2011d]). While X-ray and NMR structures of the isolated core crystallin domain are available (Bagnéris et al. [2009], Jehle et al. [2010], Laganowsky et al. [2010], Hochberg et al. [2014]), the arrangement of the subunits in oligomers is still a controversial topic. Following on from analysis of cross-talk between interfaces in the core domain of $\alpha$BC in **Chapter 2**, this Chapter presents native MS and IM experiments that probe the architecture and dynamics of subunits within full-length oligomers and bring us closer to high-resolution structures of their individual stoichiometries.

A number of atomic-resolution structures of monodisperse sHSPs are available (**Figure 3.1a**). The 24-meric Hsp16.9 from the archaebacterium *Methanocaldococcus jannaschii* and Hsp14.0 from *Sulfolobus tokodaii* are octahedral with each edge formed by a dimer (Kim et al. [1998], Hanazono et al. [2012]). The N-termini were not clearly visible in either structure but evidence such as the orientation of the first visible residue pointed to their location inside the shell. The wheat dodecameric Hsp16.5 was found to form a double disk of hexamers, with the C-termini linking subunits from both rings into tetramers (van Montfort et al. [2001]). The 32-mer of *Caenorhabditis elegans* sHSP Sip1 forms a hollow globe-shaped assembly, with each hemisphere composed of 8 meridional spokes (Fleckenstein et al. [2015]). At

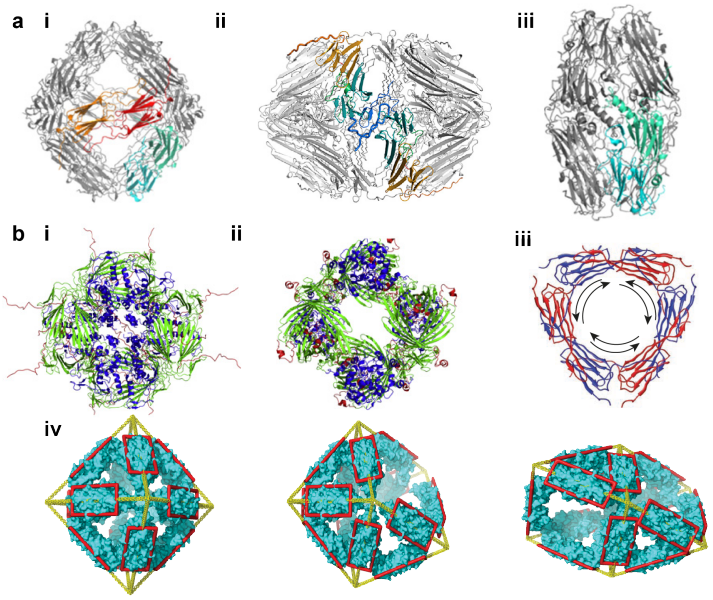



***Figure 3.1: sHSP architecture.*** ***a*** *Structures of large sHSPs.* ***i*** *Methanocaldococcus jannaschii Hsp16.9,* ***ii*** *Caenorhabditis elegans Sip1,* ***iii*** *Schizosaccharomyces pombe Hsp16.0. Adapted from Haslbeck and Vierling [2015] and Fleckenstein et al. [2015].* ***b*** *Structural models of αBC.* ***i*** *NMR, EM and SAXS 24-mer model (Jehle et al. [2011]),* ***ii*** *cryoEM 24-mer model (Braun et al. [2011]). Both of these models are based on a hexameric ring (****iii****) with subunits from neighbouring dimers swapping C-termini in a reciprocal manner (black arrows).* ***iv*** *Coarse-grained models for a 24-, 26- and 28-mer from IMMS data (Baldwin et al. [2011c]).*

the poles, the monomers are linked into a ring by the C-terminus-to-groove interactions, and the hemispheres are connected to each other by analogous but pairwise interactions. The structure of the 16-mer of *Schizosaccharomyces pombe* Hsp16.0 is similar, with four spokes of four subunits extending from one pole to the other (Hanazono et al. [2013]).

The polydisperse αBC is recalcitrant to crystallisation, but according to EM reconstructions, dynamic light scattering (DLS) and IMMS data it forms roughly spherical particles 13-19 nm in diameter with a central cavity (Haley et al. [1998], Bettelheim et al. [1999], Baldwin et al. [2011c]). Solid state NMR data show one set of signals for the core domain of αBC, suggesting that its arrangement in the different oligomers is conformationally equivalent. The flexible N- and C-termini are less well-defined. No clear signals are observed for the N-termini by solid state NMR, suggesting that they are structurally or dynamically heterogeneous (Jehle

et al. [2009, 2011]). The C-termini are known to be in the vicinity of the $\beta$4-$\beta$8-groove on the core domain (Pasta et al. [2004], Laganowsky et al. [2010], Delbecq et al. [2012]) but to remain mostly free in solution at physiological temperatures (Baldwin et al. [2012]). The rate at which they bind the $\beta$4-$\beta$8-groove correlates with the rate of monomer release from the oligomers (Baldwin et al. [2011a]).

Various pseudoatomic models have been proposed for one of the most common stoichiometries of $\alpha$BC, the 24-mer (**Figure 3.1b**). Jehle and colleagues have used the solid state NMR restraints and the core domain structure to fit small-angle X-ray scattering (SAXS) density (Jehle et al. [2010, 2011]). The resulting structure consists of four rings of hexamers, where the N-termini are placed inside the oligomer. Braun and colleagues fit the same core domain structure into cryoEM density at 9.4 Å resolution (Braun et al. [2011]). Their structure differs from that of Jehle et al. mainly in the arrangement of the N-termini, which were placed in clusters outside the hexameric rings to satisfy the EM density.

Both of these approaches used ensemble methods where the SAXS and cryoEM data represent an average of several states, even if an effort is made to classify the EM images. This poses difficulties in the case of a polydisperse protein such as $\alpha$BC. Another route, taken by our laboratories, is to use IMMS, which is capable of resolving charge states of individual stoichiometries. The information it returns is sparse, providing a rotationally averaged collisional cross-section for each stoichiometry together with a measure of its spread, but it can be powerful if used in combination with molecular modelling to discriminate between possible architectures (Ruotolo et al. [2008]). For example, the CCS values of the 11-mer TRAP protein are more consistent with a ring rather than close-packed topology (Ruotolo et al. [2005]). Using computational fitting of IMMS data on $\alpha$BC and assuming a linear relationship between subunit number and CCS, polyhedral models for different stoichiometries of $\alpha$BC were derived (**Figure 3.1b iv**). The proposed structure of the 24-mer is thus an octahedron with each edge formed by a dimer of subunits, similar to the monodisperse sHSPs described above.

Here I use IMMS to test and develop these models. Firstly, IMMS allows explicit consideration of odd stoichiometries, which so far have received little attention. I provide experimental evidence that subunit exchange occurs at least in large part by monomers, supporting the idea that monomers rather than dimers are the basic building block of the ensemble. Secondly, the structural models described above predict different orders of connectivity at the inter-dimer interface: the C-terminus-$\beta$4-$\beta$8 groove interactions should be pairwise in the case of ring structures proposed by Jehle and Braun, and multi-order in polyhedral complexes. I present disulphide crosslinking experiments at this interface that support the latter family of structures. Finally, I extend the novel IM-CID method developed by Shepherd et al. [2015] to obtain experimental CCS values for individual even and odd stoichiometries of $\alpha$BC between 18 and 32 subunits. The pattern is remarkably linear and continuous across all stoichiometries, further supporting polyhedral oligomer architecture with an average vertex degree of 4.5.

## 3.2 Materials and methods

**Mutagenesis** Mutagenesis for the E117C construct was performed using the Quikchange Lightning kit (Agilent) as specified by the manufacturer. The double mutants S135S/I159C and S135C/I161C were produced analogously by Gillian Hilton. The constructs were checked with Sanger sequencing. The mutants were purified as wild-type protein (see below).

**Protein purification** Wild-type full-length $\alpha$BC was produced as described previously (Horwitz et al. [1998], Baldwin et al. [2012]). Briefly, the IPTG-inducible gene was expressed from a pET-28 plasmid in *Escherichia coli* BL21 cells (Agilent), either in LB media for 3 hours at 37°C (unlabelled protein), or in standard M9 media with $^{15}$N ammonium chloride as the sole nitrogen source at 37°C overnight ($^{15}$N-labelled protein). The cells were lysed in 20 mM Tris pH 8.6 with a protease

inhibitor tablet (Roche) and 5 mM $\beta$-mercaptoethanol. The lysate was clarified by centrifugation at 5,000 g for 10 min at 4°C. The protein was purified using anion exchange chromatography (Q XL 16/10 column, GE Healthcare) with buffer A (20 mM Tris pH 8.6, 1 mM EDTA and 0.02% w/v sodium azide) and equivalent buffer B containing 1 M NaCl. The gradient was 0-20% over 100 ml at a flow rate of 5 ml/min. The fractions containing $\alpha$BC were pooled and subjected to size exclusion chromatography on a Superdex 200 26/60 column (GE Healthcare) in 200 mM ammonium acetate pH 6.9, with conditions as specified by the manufacturer. The purity of the fractions was checked using sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE).

**Native MS** Native MS for $\alpha$BC E117C, S135C/I159C, and S135C/I161C mutants was performed on a quadrupole-ToF instrument (Waters) modified for high mass transmission (Kondrat et al. [2015]). The sample was subjected to nano-electrospray from glass capillaries (1.0 mm outer diameter and 0.78 mm inner diameter, Harvard Apparatus) pulled to a sharp tip using a micropipette puller (Sutter Instrument) and gold-coated in-house (Agar Auto sputter coater). The MS parameters were as follows: source pressure $8.7\times10^{-3}$ mbar, capillary 1.8 kV, cone 200 V, extractor 10 V, collision gas Ar, collision gas pressure 0.15 MPa, collision energy either 30 V (low) or 200 V (high). The spectra were recorded with the samples incubated at room temperature.

**Subunit exchange experiments** The isolated proteins at 20 $\mu$M were preincubated for 30 min at 37°C, in 200 mM ammonium acetate pH 6.9. They were then mixed in a 1:1 ratio and returned to 37°C. Aliquots were removed at different timepoints: 2, 5, 10, 15, 20, 25, 30, 35, 40, 50, 60, 130 and 180 min and immediately put on ice. The samples were analysed as above (see section on native MS) with collision voltage at 200 V.

**LC-MS peptide analysis** The trypsin digest and LC-MS analysis was performed according to standard procedures, without reduction and with alkylation of any free disulphides with iodoacetamide (Borner and Fielding [2014]). LC was performed on a Dionex UltiMate 3000 RSLCnano UPLC system (Thermo Fisher) and MS data was acquired on an LTQ Orbitrap XL instrument (Thermo Scientific). The data was analysed using Xcalibur (Thermo Scientific) and the MassMatrix Database Search Engine (Xu et al. [2010]). The fragmentation spectra were compared against the database generated from the appropriate mutant protein sequence with intermolecular disulphide crosslinks allowed. A reversed decoy database was used. The precursor ion tolerance was 3 Da and the product ion tolerance was 0.5 Da. Oxidation of methionine was permitted as a variable modification. Up to 6 missed trypsin cleavages were allowed. Minimum and maximum peptide length was 3 and 40 amino acids respectively. The pp threshold (the negative common logarithm of the likelihood that the match is random) was 5.

**Refolding** For denaturation, urea was added to protein samples in 200 mM ammonium acetate pH 8 (protein concentrations were between 30 and 250 $\mu$M). The final urea concentration was 8 M. The sample was incubated at room temperature for at least 30 min. Refolding was performed by overnight dialysis into 200 mM ammonium acetate. MS analysis was performed on a Synapt G1 instrument as described in the section **Post-IM CID**, except the following parameters: trap voltage 15 V, transfer voltage 15 V for the wild-type and 150 V for the mutants, IM cell wave velocity 150 m/s and IM cell wave height 4 V, transfer wave velocity 248 m/s and transfer wave height 3.0 V. SEC was performed on an Akta Pure chromatography system using a Superdex 200 Increase 5/150 column (GE Healthcare) at 4°C as specified by the manufacturer.

**SEC under denaturing conditions** SEC was performed on an Agilent 1260 Infinity UPLC system with a TSKgel G4000SWxl column in 8 M urea at room

temperature and a flow rate of 0.3 ml/min.

**Drift-tube CID measurement of Hsp16.5** The experiment was performed on a modified Synapt G1 instrument in which the travelling wave mobility cell was replaced with a drift cell (Bush et al. [2010]). The parameters were as follows: source pressure 5.9 mbar, capillary 1.5 kV, cone 30 V, extractor 4 V, collision gas Ar, trap voltage 10 V, trap flow 10 ml/min, trap pressure $1.90\times10^{-1}$mbar, bias 30 V or 45 V, IM gas He, IM flow rate 65 ml/min, IM pressure 8.45 mbar, ToF pressure $3.54\times10^{-6}$ mbar. 11 values of drift voltages between 50 and 200 V were used. The CCS values were determined from the plot of drift time versus reciprocal drift voltage (Kemper et al. [2009]).

**Post-IM CID** Spectra were acquired on a Synapt G1 instrument (Waters). The conditions were as follows, unless stated otherwise: source pressure 5.7 mbar, capillary 1.5 kV, cone 50 V, extractor 3 V, bias 25 V, trap gas Ar, trap flow 10.0 ml/min, trap pressure $7.7\times10^{-2}$ mbar, trap collision energy 8 V, trap wave height 0.2 V, trap wave velocity 300 m/s, IM cell gas $N_2$, IM cell gas flow 22 ml/min, IM cell pressure $5.3\times10^{-3}$ mbar, IM cell wave height 9 V, IM cell wave velocity 250 m/s, static offset 120 V, transfer wave height 1 V, transfer wave velocity 65 m/s; ToF pressure $2.4\times10^{-6}$ mbar. Data was processed with MassLynx (Waters) and Pulsar (Allison et al. [2015]). The analysis pipeline is described in the text. Briefly, assignment of visible product stoichiometries was performed on the data at 200 V in the transfer cell. Arrival time distributions were fitted to a Gaussian for each product peak and m/z-dependent correction was applied (Ruotolo et al. [2008]). The arrival times for different products of the same parent ion were averaged. The selected charge state of each parent stoichiometry was determined with the MSMS selection spectrum at 8 V. The average arrival time was converted to CCS using the appropriate calibration. The calibration was performed using Hsp16.5 from *Methanocaldococcus jannaschii*, glutamate dehydrogenase (GDH), serum amyloid protein (SAP) 10-mer

and pyruvate kinase (PK), using the drift tube CCS measurement for Hsp16.5 and literature CCS values for the other proteins (Bush et al. [2010]). Tandem MS spectra of different charge states of each calibrant were recorded at 200 V in the transfer cell and arrival times of the products were used, by analogy to the processing of $\alpha$BC data. The calibration relationship was $CCS = a + bzt_D^c$, where $z$ is the parent ion charge state and $t_D$ is the m/z-corrected drift time, with $a$, $b$ and $c$ as adjustable parameters.

## 3.3 Results and discussion

### 3.3.1 Crosslinking the intra-dimer interface

To explore the effect of intra-dimer dynamics on the equilibrium distribution and subunit exchange rates of $\alpha$BC, I introduced the E117C mutation in the full-length protein. This mutation was previously made in the core domain (Hochberg et al. [2014]). **Figure 3.2a** shows the location of the mutation on the crystal structure of the E117C c$\alpha$BC dimer and in the oligomers in cartoon form.

The native mass spectra of the mutant and wild-type under low collision energy conditions were similar (**Figure 3.2b**). Thus the mutation did not significantly perturb the oligomeric distribution of $\alpha$BC. Both spectra show the pattern of overlapping peaks characteristic of polydisperse oligomers, centered around 10,000 m/z. **Figure 3.2c** shows non-reducing SDS-PAGE analysis of the sample from **Figure 3.2b**. As expected, mostly covalent dimers were observed. Some monomer was also present. This could be because the environment of the cysteines is such that complete oxidation is not energetically favourable. Some breakage of the disulphide bond is also possible during electrophoresis.

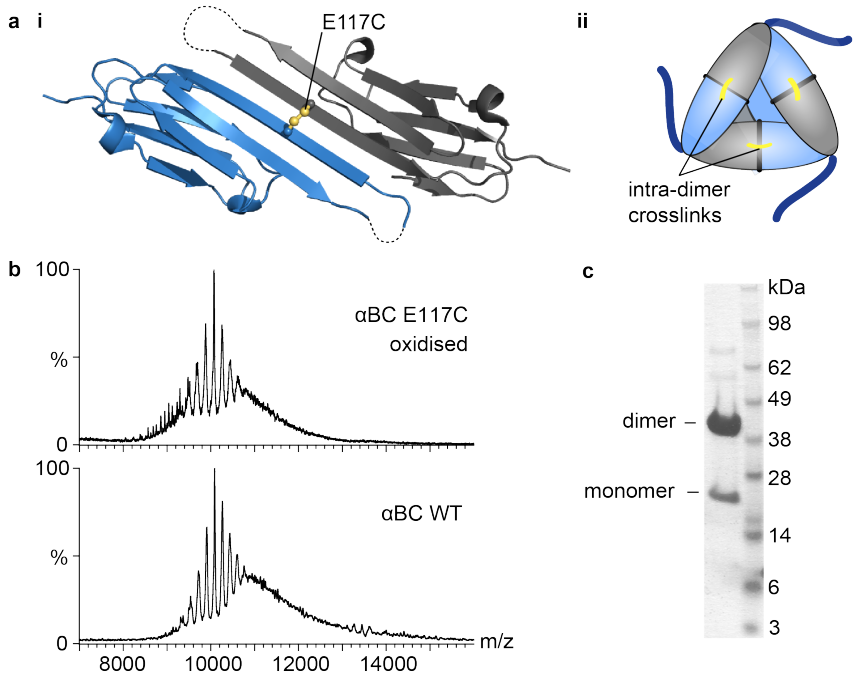



***Figure 3.2: Crosslinking the intra-dimer interface.*** ***a i*** *Illustration of the E117C mutation on the structure of cαBC (PDB ID 4M5T).* ***ii*** *Schematic illustration of the crosslinks in a cartoon of full-length oligomeric αBC.* ***b*** *Native mass spectra of intact oligomers at a collision voltage of 110 V.* ***c*** *Non-reducing SDS-PAGE of oxidised E117C, showing extensive covalent dimerisation.*

### Characterisation of CID spectra

**Figure 3.3** shows the spectrum of the wild-type and mutant proteins at high voltage, at which they undergo CID. As previously described for $\alpha$BC, the dissociation process is asymmetric (Aquilina et al. [2003], Baldwin et al. [2011d]). Highly charged, unfolded monomers are ejected from the oligomers sequentially, leaving stripped oligomers behind. The monomers appear at low m/z and the stripped oligomers, which now carry much less charge, appear at m/z above that of the parent ions. As charge is removed from the stripped oligomers, the peaks become better dispersed and it becomes possible to resolve individual stoichiometries, as analysed below.

The features of the spectra can be understood by considering which dissociation pathways are available to the protein in the absence and presence of crosslinks (**Figure 3.3a i** and **b i**). For wild-type $\alpha$BC, one route is available: successive loss

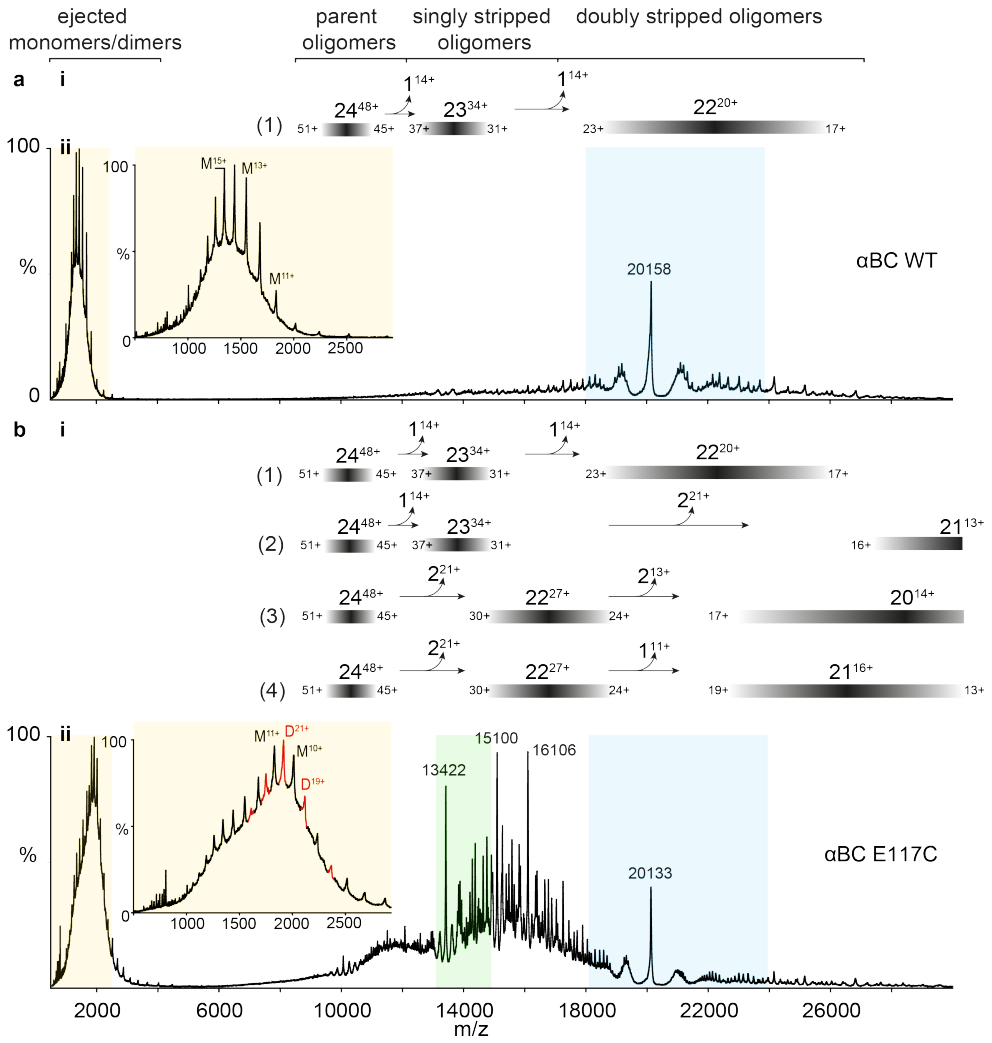



***Figure 3.3: CID of (a) wild-type αBC and (b) αBC E117C.*** *The regions where different species typically appear are annotated above the Figure. (**i**) Possible dissociation pathways illustrated on an example of a 24-mer with 48 charges. The gradient bars indicate the width of the distribution of six charge states. Note that this is one of many stoichiometries that αBC adopts. (**ii**) The CID mass spectra. The insets show the expanded stripped monomer/dimer region (yellow). The areas highlighted in green and blue are shown in **Figure 3.4**. The peaks at 13422, 15100 and 16106 correspond to oligomers with 3 charges for every 2 subunits, 4 charges for every 3 subunits and 5 charges for every 4 subunits respectively. The peaks at m/z 20133 and 20159 correspond to species with as many charges as subunits.*

of monomers. The stripped monomer region for wild-type αBC (yellow in **Figure 3.3**) shows that the average charge state of the ejected monomer is 14+. **Figure 3.3a i** indicates where the corresponding singly and doubly stripped oligomer products would be expected for an example stoichiometry, $24^{48+}$. The dominant outcome is clearly two monomer loss events. A closer look at the doubly stripped region (**Figure 3.4a i**) reveals that there is a preponderance of even stoichiometries,

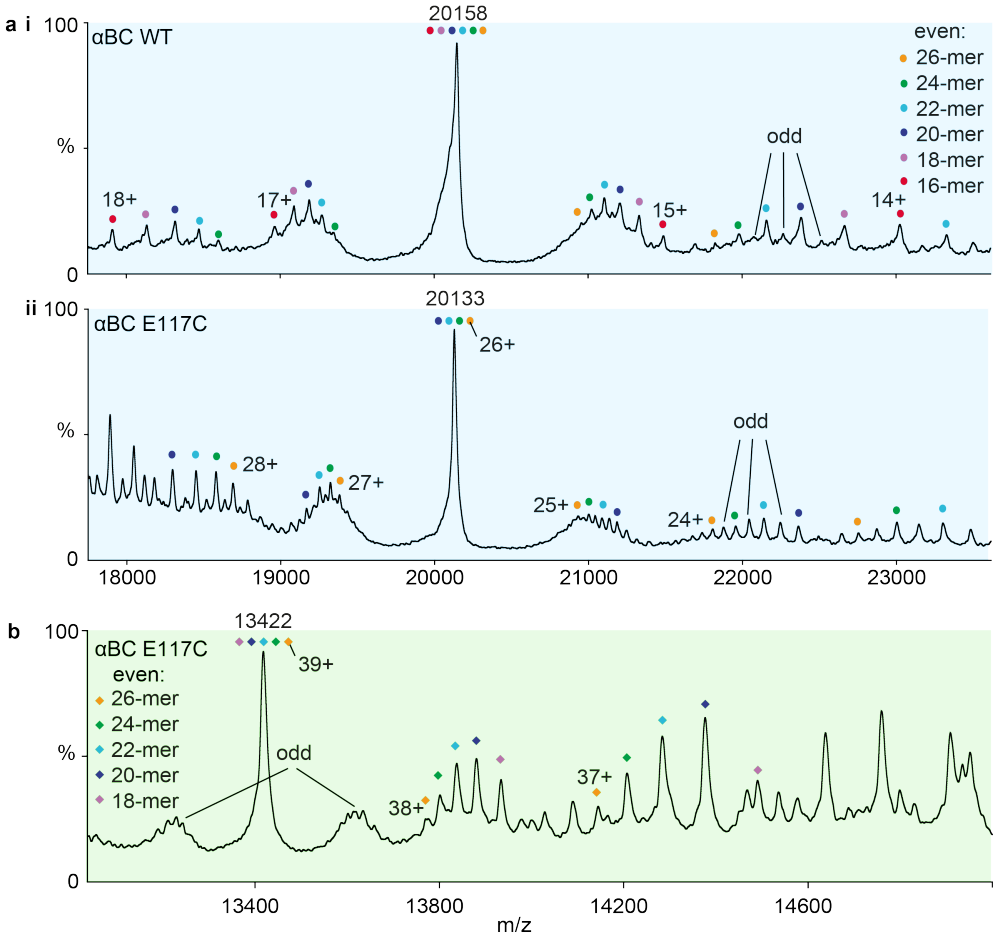



***Figure 3.4: Expanded views of stripped oligomer regions from Figure 3.3.*** ***a*** *The doubly stripped oligomer region for* ***(i)*** *wild-type αBC and* ***(ii)*** *αBC E117C.* ***b*** *Part of the singly stripped oligomer region for αBC E117C. The dominant even stoichiometries and some of their charge states are annotated in all panels. The location of the peaks corresponding to the odd stoichiometries is indicated.*

which has previously been reported for αBC (Aquilina et al. [2004], Baldwin et al. [2011d,c], Hochberg and Benesch [2014]).

The mutant protein oligomers contain predominantly covalent dimers with a small proportion of monomer (see **Figure 3.2c**). As a rough estimate, for a population of 24-mers with 10% uncrosslinked monomer, the average number of uncrosslinked monomers per oligomer would be 2.8. Not all of them might be available for dissociation since some might not be exposed on the surface of the oligomers. In the low m/z region, (yellow, inset in **Figure 3.3b ii**) both monomer and dimer dissociation products are observed, with an average charge state of 11+ and 21+ respectively. The energy required to eject a subunit depends on subunit

size: for a larger species more internal and interfacial interactions need to be broken to eject it. Therefore monomers would be lost preferentially if they are available.

The possibilities are illustrated for a $24^{48+}$ in **Figure 3.3b i**. Although the pathways (1)-(4) show two dissociation events, they could also of course only proceed halfway. Any odd-numbered oligomers necessarily have at least one uncrosslinked monomer and would therefore dissociate by pathways (1) or (2). Even oligomers could dissociate by pathway (1) or (2) if they have at least one uncrosslinked dimer, or by pathways (3) and (4).

The high relative intensity of peaks in the region 12,000-18,000 m/z indicates that dissociation of one dimer is the most common event (the first stage of pathways (3) or (4)). Dissociation of a single monomer (the first stage of pathways (1) or (2)) is also common. Both even and odd stoichiometries are observed in this region with a clear dominance of the even-numbered oligomers (**Figure 3.4b**). Interestingly, in the doubly stripped region, 18,000-26,000 m/z, the mutant shows a greater relative abundance of the odd stoichiometries compared to the wild-type protein (**Figure 3.4a ii**). Since the majority of the even oligomers dissociate via a single loss of a dimer and therefore appear in the singly stripped area, the doubly stripped region is effectively depleted of the even stoichiometries, making the odd ones appear relatively more abundant.

### The subunit exchange experiment

With this understanding of the behaviour of the E117C mutant, I proceeded to perform the subunit exchange experiment under both oxidising and reducing conditions, to test whether exchange was occurring predominantly via monomers or dimers. If dimers were the main exchanging unit, then no significant difference in the rate of exchange would be expected under oxidising conditions. If monomers were important players in subunit exchange, the process should be slower in the presence of the crosslink.

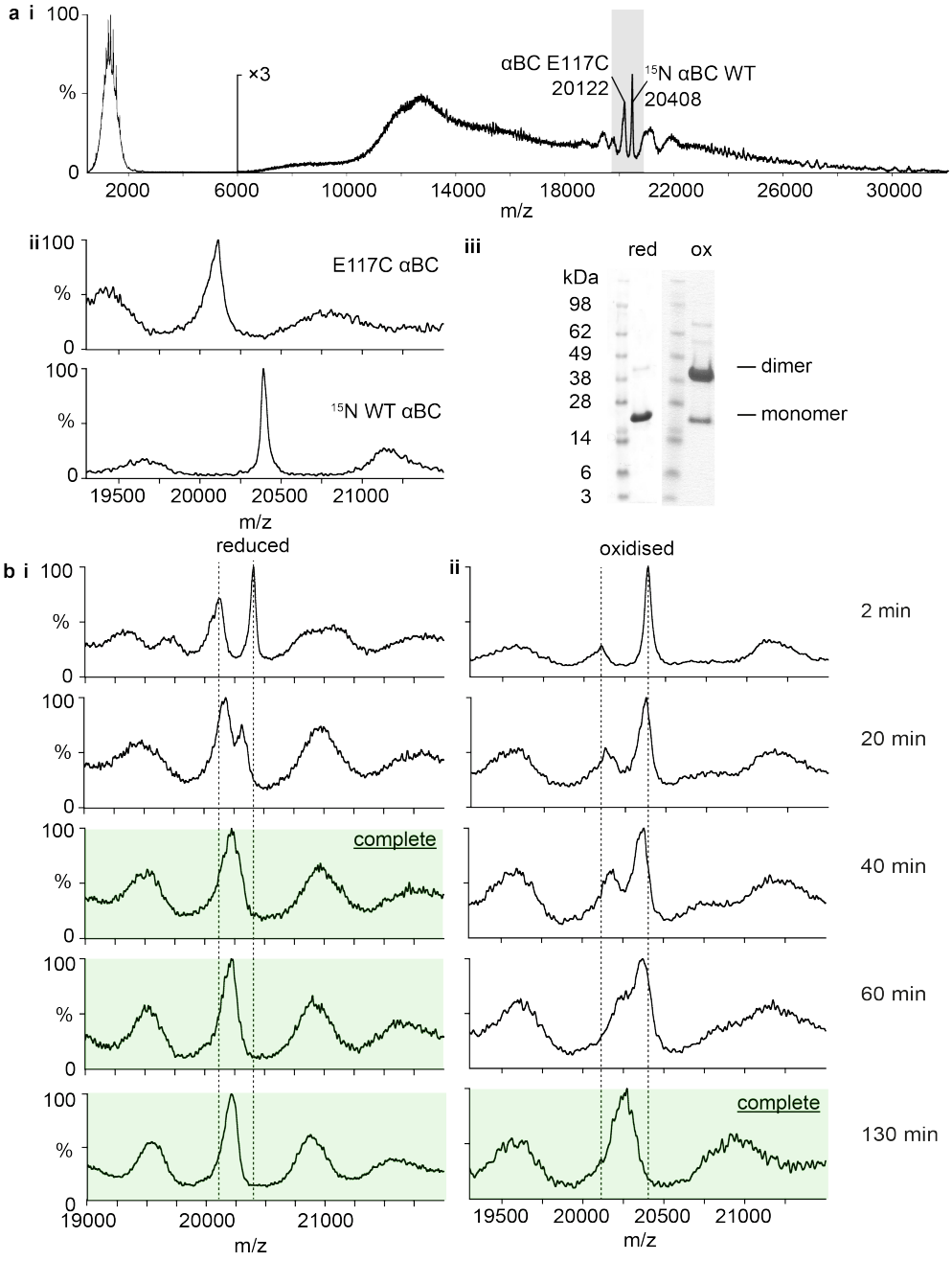



***Figure 3.5: Subunit exchange of αBC slows down in the presence of the intra-dimer crosslink.*** ***a i*** *A CID spectrum of a mixture between non-isotopically labelled αBC E117C and* $^{15}$*N-labelled wild-type αBC. The subunit exchange experiment was carried out by focusing on the mass peaks around 20,000 m/z (grey).* ***ii*** *Control spectra of each protein in isolation to confirm the identity of the mass peaks.* ***iii*** *SDS-PAGE analysis to check whether the E117C αBC samples were indeed predominantly reduced or oxidised.* ***b*** *Example spectra at five timepoints, showing gradual convergence of the two mass peaks in (**i**) reducing and (**ii**) oxidising conditions. Green background indicates spectra in which the individual peaks have converged to the extent that they can no longer be distinguished.*

The experiment is shown in **Figure 3.5**. The native mass spectrometry approach to measuring subunit exchange involves mixing either proteins that naturally differ in mass or isotopically labelled samples, and acquiring a time series of spectra to follow merging of equivalent peaks (Sobott et al. [2002], Baldwin et al. [2011d]). Note that this experiment measures solution rather than gas phase exchange rates: the MS measurement takes milliseconds and provides a snapshot of the exchange reaction that proceeds on the timescale of minutes to hours in solution. In the experiments performed here, $\alpha$BC E117C was not isotopically labelled whereas wild-type $\alpha$BC was labelled with $^{15}$N. A convenient intense and well-resolved peak to follow was the one in the doubly stripped oligomer region corresponding to all oligomers, even and odd, carrying as many charges as subunits (20,158 or 20,133 in **Figure 3.4a**, also highlighted in **Figure 3.5a i**). This peak falls on the m/z corresponding to the molecular mass of the monomer and is therefore called the mass peak. The mass peaks were offset by the expected difference in mass for the mutant and isotopically labelled wild-type protein (**Figure 3.5a**). SDS-PAGE analysis of the oxidised and reduced samples of $\alpha$BC E117C confirmed the dominance of covalent dimers and monomers respectively.

The proteins were mixed and incubated in solution at 37°C. Mass spectra of the mixture were acquired at several time points during incubation (**Figure 3.5b**). The two mass peaks were well resolved 2 minutes after mixing but merged over time as mixed oligomers were formed. In reducing conditions the peaks merged by about 40 minutes. In contrast, under oxidising conditions the two mass peaks could still clearly be distinguished after 40 minutes. A single peak was only observed after 130 minutes, i.e. the oligomers took over 3 times longer to mix in the presence of the E117C crosslink.

Thus introduction of an intra-dimer covalent crosslink clearly slows down subunit exchange. Moreover, the retardation observed here is an underestimate. The reason for this is that although the mass peak contains contributions from many stoichiometries and is therefore a good choice for obtaining an overall measure of

the rate of subunit exchange, we know from the CID analysis in **Figure 3.4** that odd oligomers that have at least one uncrosslinked monomer are overrepresented in the doubly stripped oligomer region. The actual subunit exchange rate is thus expected to be even slower in fully crosslinked oligomers. This means that subunit exchange in wild-type $\alpha$BC cannot take place exclusively by movement of dimers or larger oligomers; rather, a significant contribution comes from movement of individual monomers. The results validate the unbiased Poissonian monomer exchange model previously proposed for $\alpha$BC (Baldwin et al. [2011d]). Thus, although it is often thought that dimers are the basic substructural unit of sHSPs, monomers play an important role for $\alpha$BC, and its odd stoichiometries are exchange-competent members of the oligomeric ensemble in solution.

### 3.3.2 Crosslinking the inter-dimer interface

In addition to the intra-dimer interface explored in the previous section, another set of interactions important in $\alpha$BC oligomers is the binding of the flexible C-termini and $\beta$4-$\beta$8 grooves on a neighbouring monomer. This is termed the inter-dimer interface. The interaction is illustrated in **Figure 3.6a i** on a crystal structure of an $\alpha$BC construct missing the N-terminus (PDB ID 3L1G). The pattern of these interactions depends on the arrangement of the subunits in the oligomer. For example, in a tetrahedral structure the C-terminus-groove interactions might be expected to show a threefold cyclical symmetry. In a disk structure, pairwise swapping of C-termini is predicted (illustrated in **Figure 3.1b iii**).

To investigate the connectivity of subunits in this interface, I studied double cysteine mutants in which disulfide crosslinks could be made between the C-terminus and the groove. The location of the amino acids mutated is shown on the crystal structure in **Figure 3.6a i**. The $\beta$4-$\beta$8 groove mutation was S135C. Mutation of this residue has previously been shown not to result in large structural or functional perturbations of the protein (Delbecq et al. [2012], Mainz et al. [2015]). In the C-

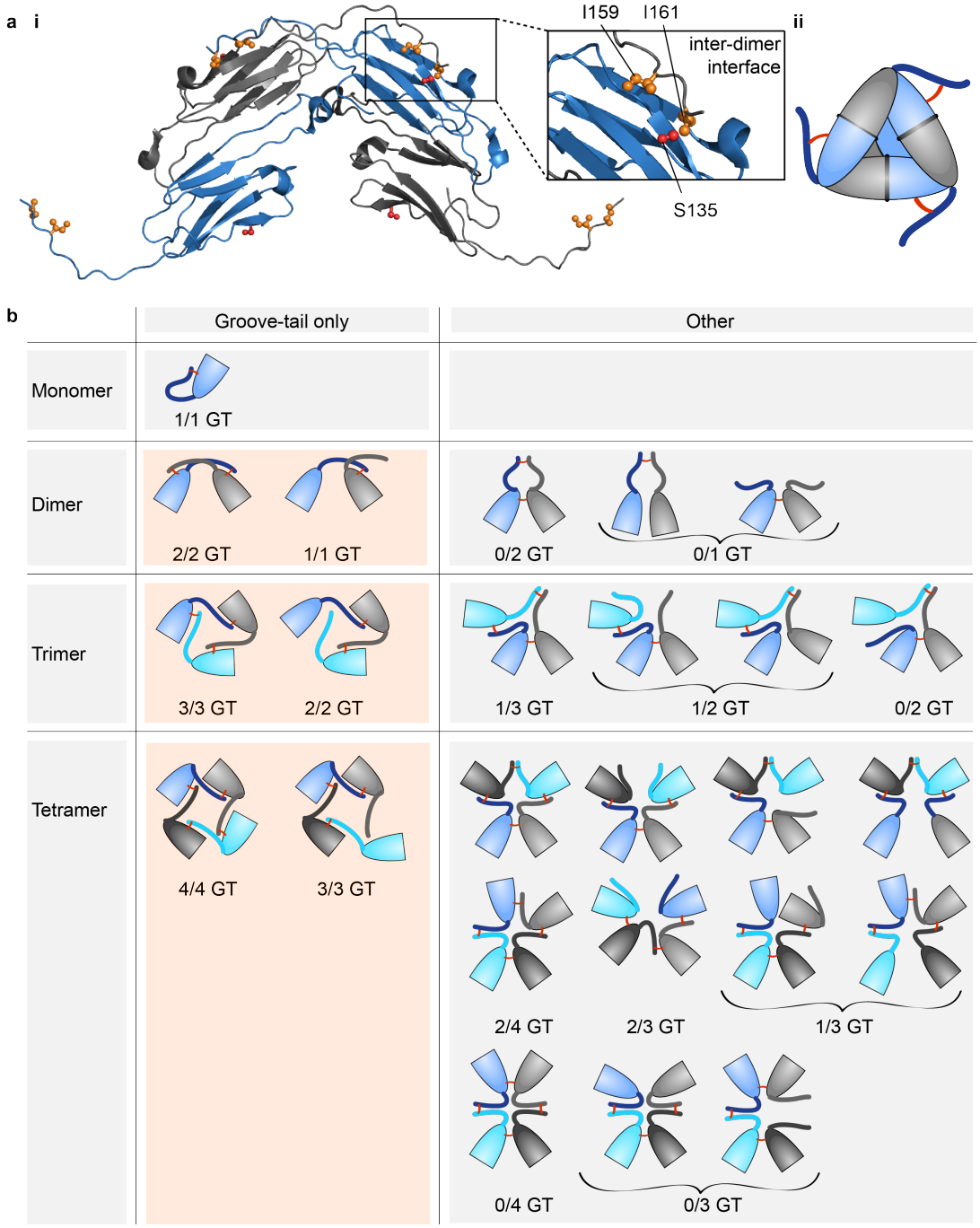



***Figure 3.6: Crosslinking the inter-dimer interface.*** ***a i*** *Location of S135 in the β4-β8 groove, and I159 and I161 on the flexible C-termini. These residues were mutated to cysteines to create disulphide crosslinks at the inter-dimer interface. PDB ID 3L1G.* ***ii*** *Illustration of the crosslinks in a schematic representation of an αBC oligomer.* ***b*** *Table illustrating possible crosslinked species up to tetramer. GT - groove-tail crosslinks (other possibilities are tail-tail and groove-groove). The desirable species with the maximum number of groove-tail crosslinks are shown on a pink background.*

terminus, two alternative mutations were tested: either I159C or I161C, since the tail is flexible in solution and the sequence in this region is palindromic. The crosslinks

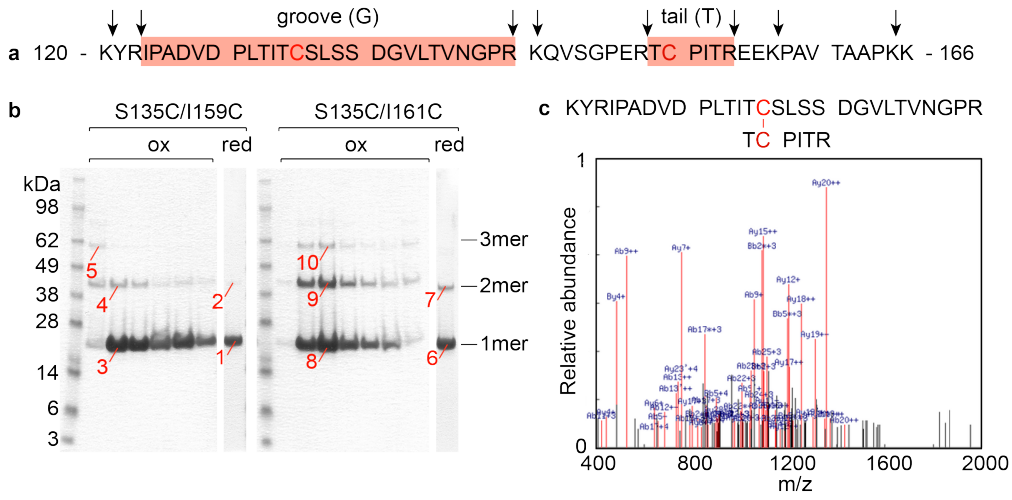
are illustrated diagrammatically in a cartoon oligomer structure in **Figure 3.6a ii**.

I pursued two goals in this project. First, I aimed to characterise which covalently crosslinked species (dimers, trimers, tetramers etc.) would be observed, since this can give clues as to possible geometries in the full set of polydisperse oligomers. Secondly, I aimed to reconstitute subpopulations of the $\alpha$BC ensemble by isolating the croslinked species under denaturing conditions, refolding them and allowing them to reassemble separately (e.g. only trimers). The resulting sets of oligomers should be high symmetry members of the $\alpha$BC distribution. They would be more amenable to structural study due to their reduced polydispersity.

Of course, apart from the desired groove-tail connection, non-specific crosslinks could also be formed. The space of possible crosslinks in species up to tetramer is presented in **Figure 3.6b**. A monomer could crosslink with itself, which would likely make it incompetent for oligomer assembly. As the stoichiometry increases, the variety of species with undesirable groove-groove and tail-tail crosslinks grows. Many of these, such as groove-groove crosslinks, are likely to be very unfavourable sterically, but others, such as tail-tail crosslinks, are quite likely.

To find out which crosslinks were formed predominantly, I carried out a trypsin digest and liquid chromatography mass spectrometry (LC-MS) of different SDS-PAGE bands from reduced and oxidised samples of both S135C/I159C and S135C/I161C double mutants. The results are shown in **Figure 3.7**.

**Figure 3.7d** presents a summary table showing how many peptide spectra corresponding to each type of crosslink were detected in each band. Apart from intrinsic relative abundance of crosslinks, the total number of detection events also depends on the intensity of the excised band. Crosslinks were detected in the reduced bands (1, 2, 6, 7) indicating that either the reduction in solution was incomplete (which would not be detected on SDS-PAGE for the monomers) or there was some scrambling of the disulfides. All oxidised bands contained protein with groove-tail crosslinks, and those for $\alpha$BC S135C/I159C also contained tail-tail crosslinks.




**d**

| | S135C/I159C | GT | TT | GG |
|---|---|---|---|---|
| 1 | monomer (reduced) | 5 | 6 | – |
| 2 | dimer (reduced) | – | – | – |
| 3 | monomer (oxidised) | 21 | 6 | – |
| 4 | dimer (oxidised) | 5 | 1 | – |
| 5 | trimer (oxidised) | 1 | 2 | – |

| | S135C/I161C | GT | TT | GG |
|---|---|---|---|---|
| 6 | monomer (reduced) | 3 | 1 | – |
| 7 | dimer (reduced) | 6 | – | – |
| 8 | monomer (oxidised) | 7 | – | – |
| 9 | dimer (oxidised) | 8 | – | – |
| 10 | trimer (oxidised) | 2 | – | – |

***Figure 3.7: Specificity of the inter-dimer crosslinks.*** ***a*** *Partial sequence of αBC (residues 120-166) with the mutations S135C and I159C, and predicted trypsin cleavage sites (arrows). The peptides containing the cysteines are highlighted in pink. The peptide pattern is the same for the S135C/I161C double mutant.* ***b*** *SDS-PAGE of chromatography fractions of the mutants in the oxidised and reduced conditions. The red arrows indicate which bands were excised for LC-MS analysis.* ***c*** *Example fragmentation spectrum of two crosslinked peptides analysed with the MassMatrix software. The peaks in red correspond to assigned fragments.* ***d*** *Table showing how many fragment spectra could be assigned to each crosslink in each sample (GT: groove-tail, TT: tail-tail, GG: groove-groove).*

The S135C/I159C oxidised monomer had particularly many groove-tail crosslinks, partly because it is a very intense gel band, but also perhaps because this mutation is paticularly favourable for intramolecular crosslinks, since the equivalent band for the S135C/I161C band (8) was equally intense but much fewer crosslinks were detected in it. For αBC S135C/I161C, only groove-tail crosslinks were observed. No groove-groove crosslinks were observed in any samples, probably because formation of this crosslink is highly sterically hindered, and the crosslinked peptide would be too large to ionise well. In summary, αBC S135C/I161C formed more specific crosslinks than the S135C/I159C mutant.

I continued characterisation of both proteins, keeping in mind that, in the

absence of other factors, the S135C/I161C mutant would a better choice due to its specificity. Next, I undertook optimisation of crosslinking conditions in collaboration with a Master's student, Oliver Stach[1].

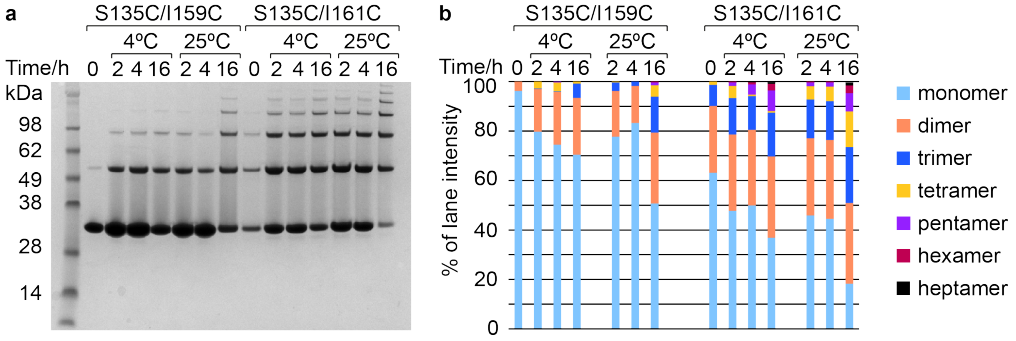



***Figure 3.8: Optimisation of crosslink efficiency.*** ***a*** *SDS-PAGE showing optimisation of crosslinking by dialysis for different temperatures and timepoints.* ***b*** *Densitometry-based quantification of the gel in* ***a****.*

Some of the factors that might enhance formation of disulphides include temperature, pH, which increases nucleophilicity of thiols (Poole [2015]) and time of incubation. A reaction that is too fast might result in formation of predominantly unspecific crosslinks, such as self-crosslinking of a monomer. Thus an array of gentle conditions was tested. We first attempted to oxidise the samples by removing the reducing agent ($\beta$-mercaptoethanol) by size exclusion chromatography (SEC) followed by incubation at three different temperatures (4°C, 25°C, 37°C) and two pH values (7.0 and 8.0) for varying periods of time (data not shown). The S135C/I161C construct formed more crosslinks than S135C/I159C construct, but we did not see any significant variation between the conditions. We then attempted more gradual removal of reducing agent via dialysis at pH 8.0 and two different temperatures (**Figure 3.8**). Overnight incubation at the warmer temperature significantly improved the efficiency of crosslink formation. Thus dialysis at pH 8.0 and at 25°C was chosen as the optimal way to produce crosslinked protein.

We checked that the mutants still formed oligomers under these conditions by recording native mass spectra shown in **Figure 3.9**. The mutants are clearly

[1]Crosslinking optimisation, separation in denaturing conditions and mass spectrometry characterisation of the refolded proteins (data in **Figures 3.8-3.11a**) was performed by Oliver Stach as part of his Master's degree project under my supervision.

oligomeric with a distribution broadly similar to wild-type $\alpha$BC. The S135C/I161C mutant has a tendency to form larger oligomers but the most abundant species are those similar to wild-type. Thus crosslinking under the optimised conditions did not significantly hinder oligomer formation.

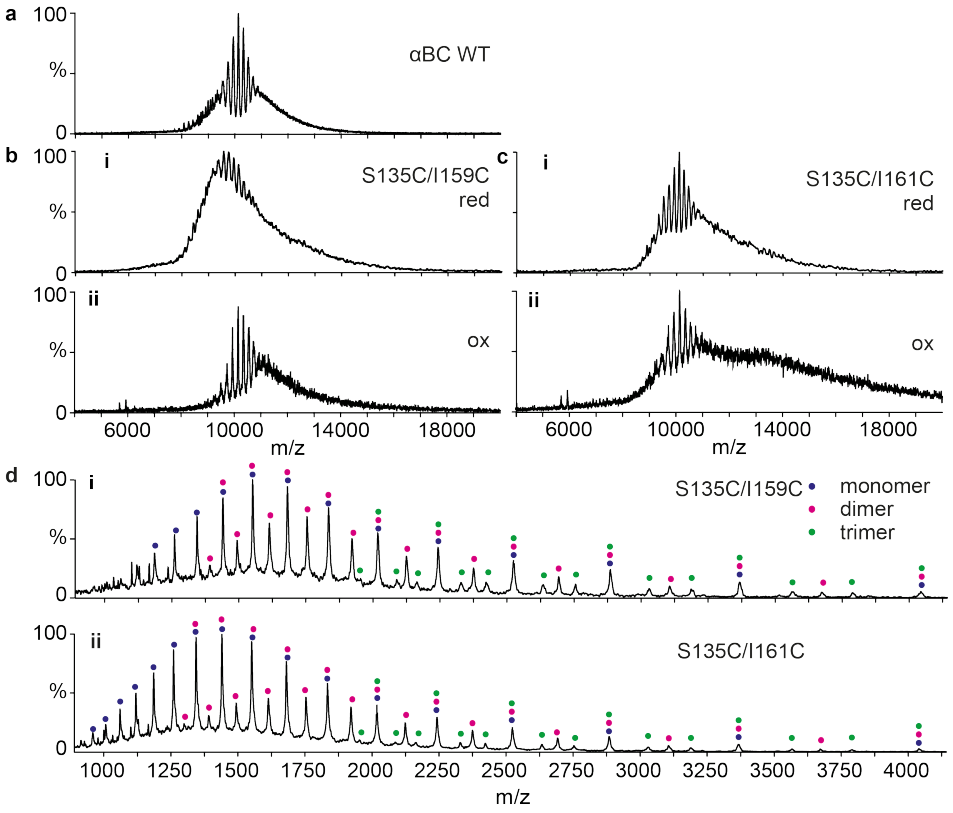



***Figure 3.9: Crosslink formation does not hinder oligomerisation.*** *Native mass spectra of (**a**) wild-type $\alpha$BC, (**b**) $\alpha$BC S135C/I159C and (**c**) $\alpha$BC S135C/I161C. (**i**) Reducing and (**ii**) oxidising conditions. **d** Assigned stripped monomer regions of the CID mass spectra of (**i**) $\alpha$BC S135C/I159C and (**ii**) $\alpha$BC S135C/I161C under oxidising conditions.*

To confirm formation of crosslinks by mass spectrometry, the oxidised mutants were subjected to CID with high collision energy. The stripped oligomer region did not have high enough signal-to-noise ratio to be useful, but the stripped monomer region could be analysed (**Figure 3.9d**). Monomer, dimer and trimer dissociation products were observed for both mutants, confirming existence of covalent crosslinks. Note that generally any smaller species would be dissociated preferentially. Therefore the relative abundance of the higher-order covalent oligomers observed here is an underestimate.

### Separation of crosslinked species

Having optimised the crosslinking conditions and checked that the oligomer distributions of the mutants are similar to wild type, we sought to separate the crosslinked species and reconstitute subpopulations of oligomers from them.

The separation must be performed in denaturant to break the non-covalent interactions holding the native oligomers together. To reconstitute oligomers, the proteins must be refolded from the denaturant. To check that with this procedure it is in principle possible, we first attempted refolding without any separation. Wild-type and mutant $\alpha$BC samples were incubated in 8 M urea for 30 min and dialysed into ammonium acetate pH 8.0 overnight. They were then subjected to native MS analysis and SEC in ammonium acetate. The results of SEC on a Superdex 200 Increase 5/150 column are shown in **Figure 3.10a**. According to GE Healthcare, the manufacturer of the column, a 669 kDa example protein elutes at 1 ml and a 158 kDa example protein elutes at 1.8 ml. From the elution profiles in **Figure 3.10a**, we can therefore conclude that the refolded $\alpha$BC variants must be oligomeric.

To compare the oligomeric distributions before and after refolding, we acquired native mass spectra of the refolded samples prior to any SEC separation (**Figure 3.10b**). A control spectrum of $\alpha$BC not subjected to denaturation and refolding was also recorded. A transfer voltage of 150 V was required to obtain adequate transmission of the refolded proteins, possibly due to residual contamination with urea. Therefore stripped monomers are observed at low m/z. Nevertheless, all samples clearly show the presence of polydisperse oligomers at similar m/z to the control spectrum in **b i**. Panel **c** additionally shows the 2D IMMS spectra of wild-type $\alpha$BC before denaturation and of S135C/I159C after refolding. The drift time distributions of the oligomer region (10-40 ms) are similar in the two spectra (see the summed projections in the rightmost panel). This further confirms that the oligomeric distribution of the refolded mutant is similar to wild-type.

To try and separate the different crosslinked species, we performed SEC at

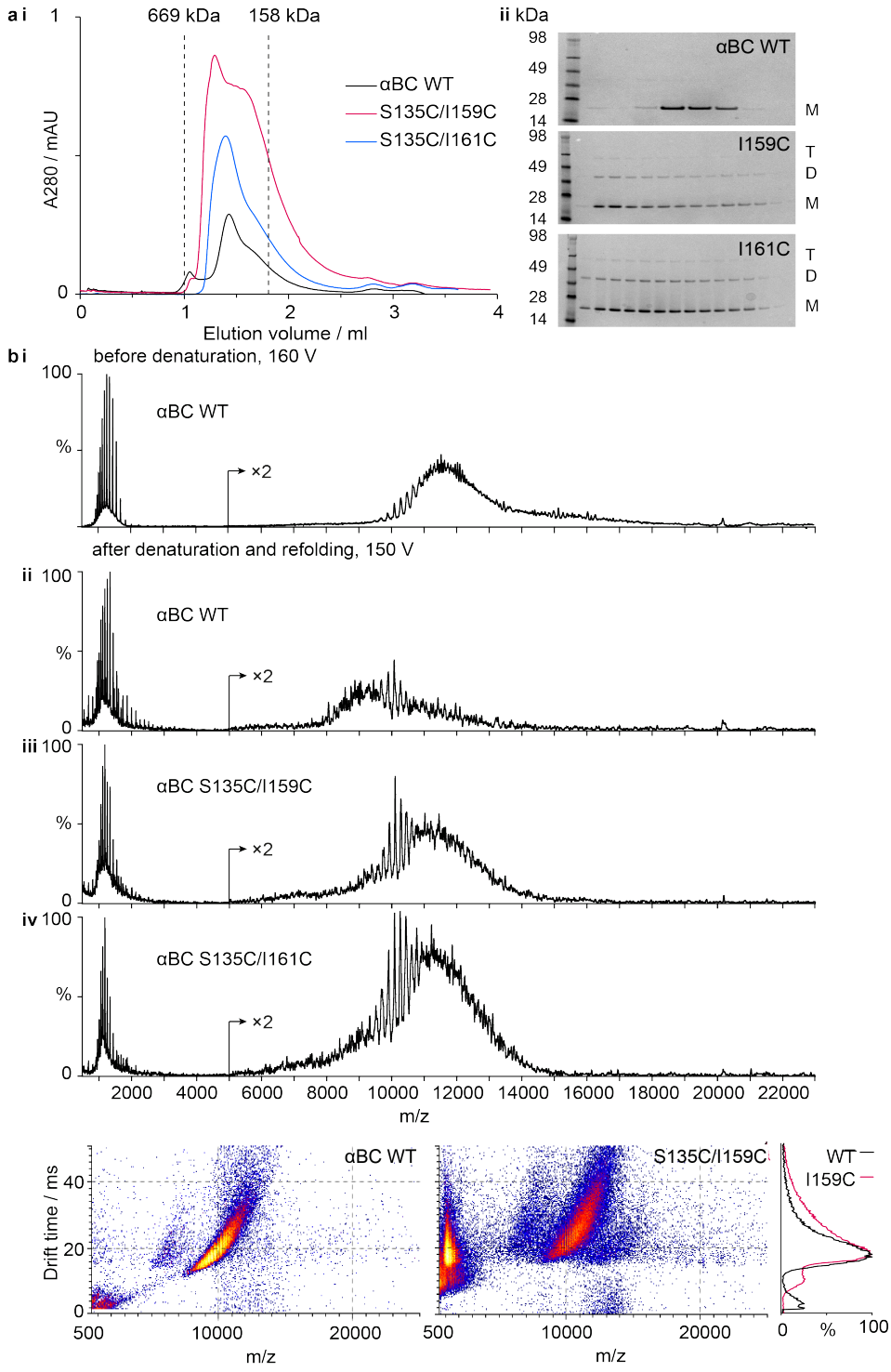



***Figure 3.10: The crosslinked mutants can be denatured and refolded without loss of oligomerisation capacity.*** ***a*** *SEC of wild-type and mutant αBC constructs following denaturation and refolding.* ***i*** *Traces of absorbance at 280 nm.* ***ii*** *SDS-PAGE of the region between 1 and 3 ml.* ***b*** *Mass spectrometry analysis on unseparated samples.* ***i*** *Control spectrum of wild-type αBC before urea treatment (transfer voltage 160 V).* ***ii-iv*** *Spectra of refolded wild-type and mutant αBC (transfer voltage 150 V).* ***c*** *IMMS spectra of wild-type αBC before denaturation (transfer voltage 50 V) and of the refolded S135C/I159C mutant (transfer voltage of 150 V). The drift time projections are shown on the right.*

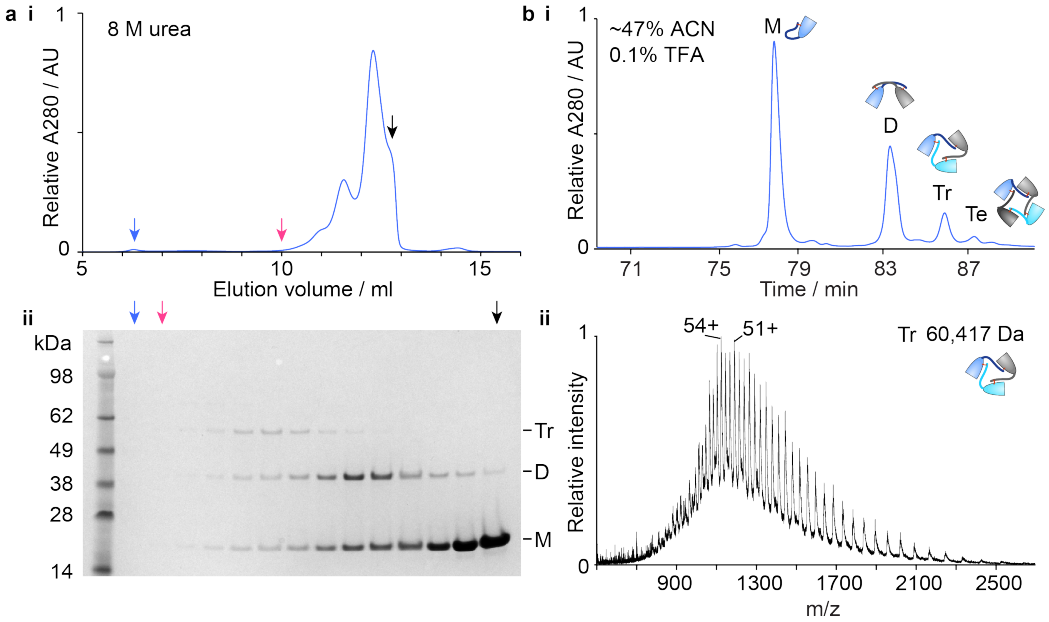



***Figure 3.11: Separation of crosslinked species for αBC S135C/I161C.*** ***a*** *SEC in 8 M urea on the TSKgel G4000SWxl column.* ***i*** *The SEC chromatogram. The arrows refer to the SDS-PAGE fractions in* ***ii****.* ***ii*** *SDS-PAGE of the fractions across the main peak. Monomer (M), dimer (D) and trimer (Tr) bands are indicated.* ***b*** *Reverse phase chromatography on an Agilent C8 column*[2]*.* ***i*** *The chromatogram. The baseline-resolved peaks for monomer, dimer, trimer and tetramer (Te) are annotated. Note that the configuration of crosslinks in the cartoons is illustrative; other possibilities exist for each species (see* ***Figure 3.6b****). The species were identified using SDS-PAGE and MS as in (****ii****).* ***ii*** *The mass spectrum of the pooled trimer fractions. Analogous spectra were recorded to confirm the identity of monomer and dimer.*

room temperature in 8 M urea on the TSKgel G4000SWxl column. The results are shown in **Figure 3.11a**. The SEC profile was complex. SDS-PAGE analysis of the fractions between the red and black arrows showed that it corresponded to overlapping peaks of the trimer, dimer and monomer, although they were not well resolved.

Further optimisation of separation was carried out by India Kelly under the supervision of Shane Chandler. They showed that separation by reverse phase chromatography in acetonitrile results in baseline-resolved peaks for monomer, dimer and trimer (**Figure 3.11b**). This excellent result enables the next stage of the project: refolding the separated fractions alone and in combination and characterising the resulting sets of oligomers of reduced polydispersity.

[2]Data and figures in panel **b** and were kindly provided by India Kelly and Shane Chandler.

In summary, groove-tail crosslinking at the inter-dimer interface is particularly specific for the S135C/I161C mutant. Although the dimer is the most dominant crosslinked species, significant amounts of at least trimer, tetramer and some higher-order oligomers are also observed. This supports polyhedral mixed-degree architecture for $\alpha$BC, rather than ring-based structures (**Figure 3.12**, Jehle et al. [2009], Braun et al. [2011]). The crosslinked protein is assembly-competent in oxidising conditions, even after denaturation and refolding. The experiments on refolding and separation under denaturing conditions reported here show great promise for future work towards reconstituting high symmetry subpopulations of $\alpha$BC oligomers.

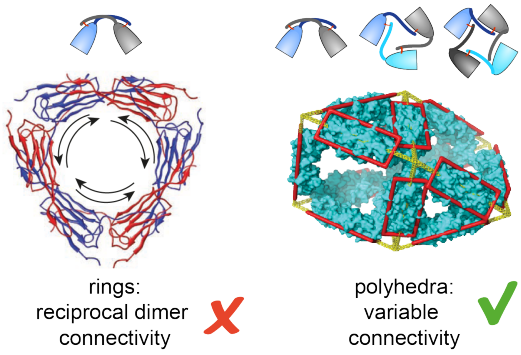



***Figure 3.12: Inter-dimer crosslinking supports polyhedral structures of $\alpha$BC.*** *Ring structures (left) are based on dimers of subunits reciprocally swapping C-termini (black arrows, vertex order 2). Observation of significant amounts of higher-order crosslinks therefore does not support such structures but suggests polyhedral topologies with varied vertex order.*

### 3.3.3 CCS measurements of individual stoichiometries of $\alpha$BC using post-IM CID

The crosslinking results presented in the two previous sections show that $\alpha$BC monomers take part in subunit exchange and that they have variable connectivity at the inter-dimer interface. IMMS can provide aditional structural restraints in the form of collisional cross-sections (CCS), potentially for individual stoichiometries of $\alpha$BC, thus going further than ensemble imaging methods such as EM. The CCS values allow screening of candidate structural models. The data would be particularly interesting for odd oligomers and species where it is difficult to conceive a symmetrical arrangement, such as 19-mers.

CCS measurement is straightforward for monodisperse proteins (Ruotolo et al.

[2008]). It entails measurement of arrival times for peaks corresponding to the species of interest, correcting for the m/z-dependent drift time and for charge and either calculating the CCS from the drift time from first principles (in a drift tube instrument) or using a calibration curve with proteins whose CCS is already known (in a travelling wave instrument). However, knowledge of the mass and charge of the species of interest is a prerequisite. Therefore it is not possible to apply these methods directly to measure CCSs of $\alpha$BC oligomer stoichiometries, because the peaks corresponding to individual m/z states of each are not resolved.

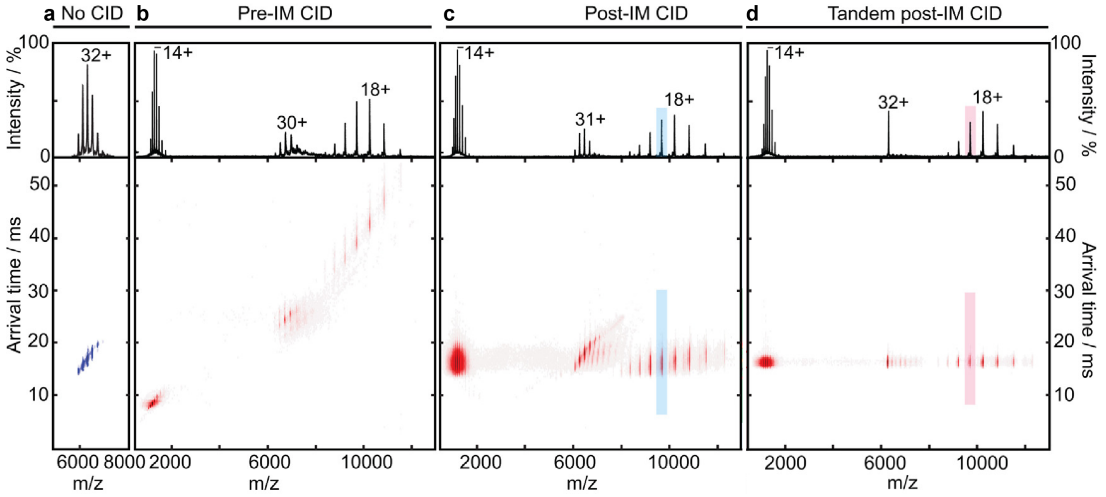



***Figure 3.13: Time-alignment of precursor and products in post-IM CID on wheat Hsp16.9.*** ***a*** *The IMMS spectrum of Hsp16.9 at low collision energy.* ***b*** *CID prior to the mobility measurement (in the trap cell) leads to distortion of arrival times.* ***c*** *CID after the mobility cell preserves the native arrival times. It is not possible to assign CID products unambiguously to charge states of the parent oligomer.* ***d*** *Isolation of a single charge state of the parent oligomers followed by CID shows that the products are time-aligned to the precursor. The products can be directly assigned to the isolated parent charge state. Adapted from Shepherd et al. [2015]).*

It is possible to circumvent this difficulty with a tandem MS-IM-CID approach, recently been developed in our lab (Shepherd et al. [2015], **Figure 3.13**). CID is performed after the mobility separation, so that the products carry the undistorted mobility signature of their parent ions. The stripped oligomer products are much better separated in the m/z dimension due to significant charge loss to the stripped monomers, making it possible to resolve peaks from individual parent stoichiometries. By selecting a single precursor peak or a very narrow region, the mass

and charge of the parent ions can be identified unambiguously. The arrival time distribution for the precursor can be determined from the average of those for its unambiguously assigned products. Thus all the necessary information to determine the CCS is made available.

### Outline of the post-IM CID experiment on $\alpha$BC

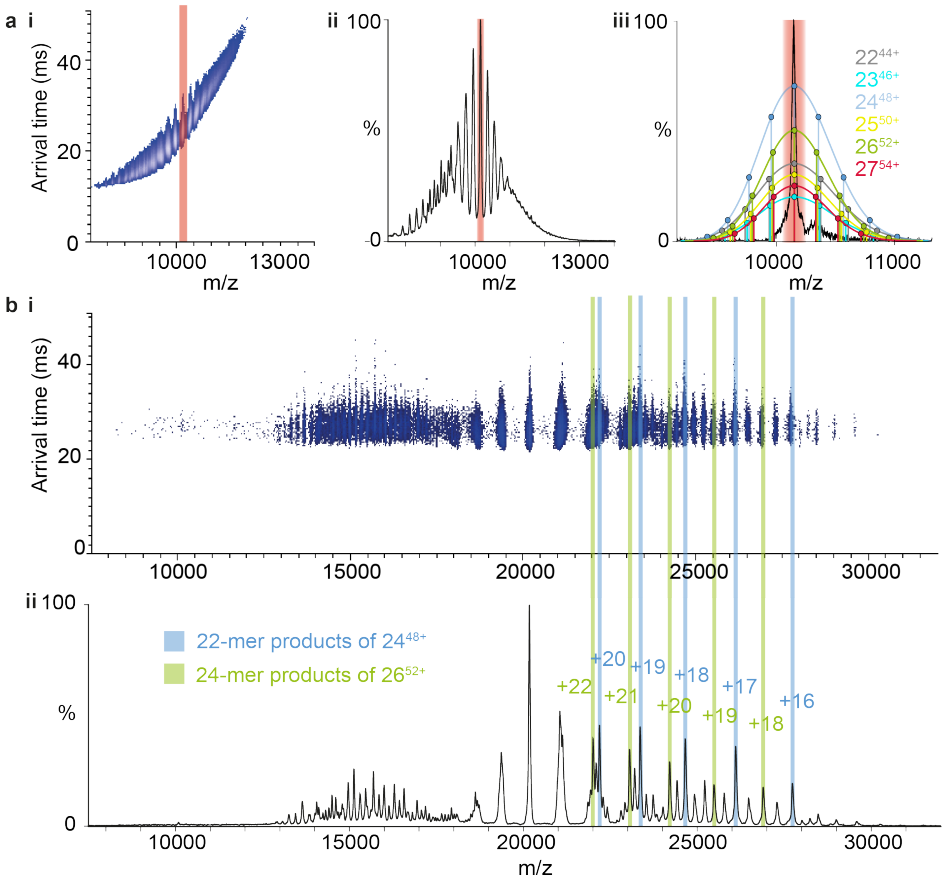



***Figure 3.14: Drift time measurement of individual oligomer charge states of $\alpha$BC by post-IM CID.*** ***a*** *The low collision energy spectra.* ***i*** *2D IMMS spectrum of $\alpha$BC showing an example region to be isolated by the quadrupole (pink).* ***ii*** *The m/z dimension of the spectrum in* ***i****.* ***iii*** *The isolated peak with an illustration of which oligomer charge states contribute to the signal in the selected window.* ***b*** *The high collision energy spectra.* ***i*** *2D IMMS spectrum of the stripped oligomer region.* ***ii*** *The m/z dimension of the spectrum above. Greater peak dispersion in the doubly stripped oligomer area enables assignment of product charge series. Two are annotated here as examples. Their arrival time distributions can be extracted as slices from the 2D spectrum in* ***ii*** *and matched with the parent species from the assignment in* ***a iii****.*

I applied the post-IM CID approach to find CCS values for as many stoi-

chiometries of $\alpha$BC as possible. The experiment is illustrated in **Figure 3.14**. I selected narrow windows of the spectrum using the quadrupole (**Figure 3.14a i-ii**) and recorded their 2D spectra at low and high transfer energy. For each isolation, I analysed which parent stoichiometries with a single charge state fall into this window (**Figure 3.14a iii**). Following assignment of the visible charge series in the high energy product spectrum, I matched the series of products with the correct precursor ions. For example, the charge state series of the 22-mer in the doubly stripped region can only arise from the $24^{48+}$ species. Although the parent ion has one charge state, the product oligomer charge state is variable because the splitting of charge between stripped monomers and oligomers can occur in a number of ways.

Extraction of the arrival time distributions for each product ion thus gives several arrival time distributions for the parent ion. The distributions can be fit with a Gaussian curve. The mean arrival times must be corrected for the m/z-dependent drift time between the IM cell and the ToF pusher, using the m/z of the product, to give true drift times for the parent ion. This correction is on the order of 1-3% of the arrival time. Since the use of a travelling-wave instrument is essential to obtain the required sensitivity in the stripped oligomer region, a calibration with proteins of known CCS must be recorded to convert arrival times into CCS, as described below.

### Variation of drift times with transfer voltage and implications for CCS calibration

Since the transfer cell is located after the IM cell in the instrument (see **Figure 1.21** in **Chapter 1**), variation of the transfer voltage should in principle have no effect on the drift time recorded in the IM cell. Therefore it should even be possible to record $\alpha$BC spectra at 200 V and spectra of calibrants at 8 V in the transfer. This would be beneficial since some of the smaller calibrants undergo extensive scattering and fragmentation at 200 V. To check this assumption of constant drift time, I recorded

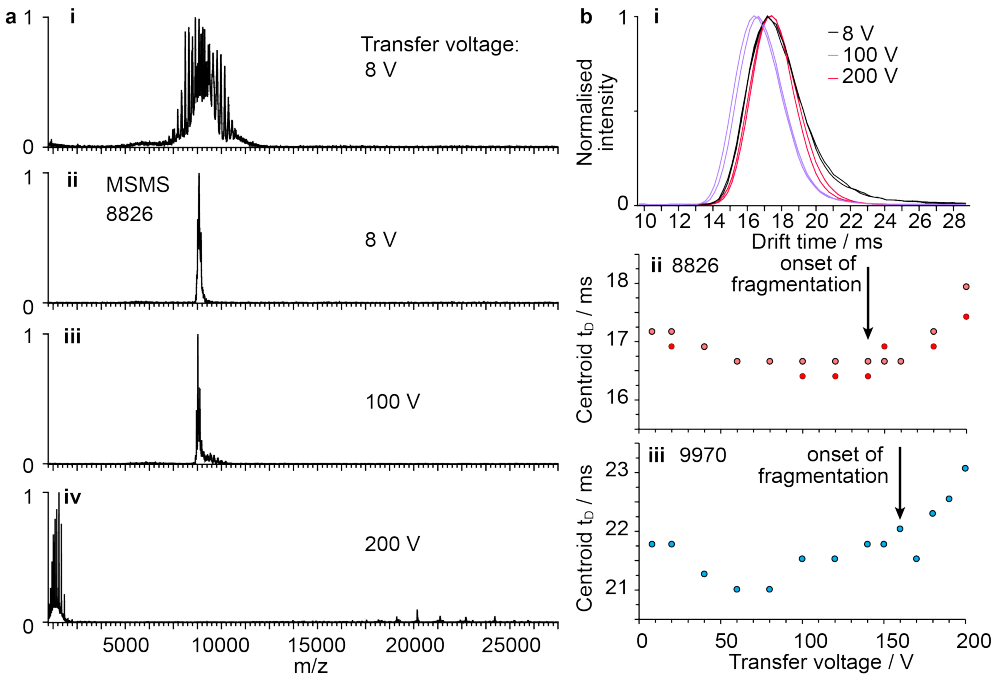



***Figure 3.15: Voltage in the transfer cell affects arrival times.*** ***a*** *Example mass spectra of αBC before any selection (**i**), after selection of a window centered on m/z 8826 at 8 V (**ii**), at 100 V, still prior to the onset of fragmentation (**iii**), and 200 V, when extensive fragmentation occurred (**iv**).* ***b i*** *Arrival time distributions corresponding to the mass spectra in **a**.* ***ii-iii*** *Variation of the centroid of the distributions like those in **i** with transfer voltage for two precursor window selections, centered on 8826 and 9970.*

summed arrival time distibutions of strips of the αBC spectrum at a range of transfer voltages between 5 and 200 V (**Figure 3.15**). Surprisingly, variation of arrival time of up to 2 ms (about 10%) was observed. Arrival time varied significantly even before any fragmentation of the parent ion took place (up to 120-150 V). **Figure 3.15b i** shows that a noticeable shift in the average arrival time of the precursor ion occurred between 8 and 100 V without noticeable effect on its shape. Interestingly, for the selection centred at m/z 8826, at 200 V the distribution was very similar to that at 8 V, but this was not the case for a different selection, centred on m/z 9970, where the arrival time at 200 V was significantly longer (**Figure 3.15b iii**).

**Figure 3.16** shows that this puzzling effect is still observed when the data is analysed in a more detailed way, with application of the m/z-dependent correction to arrival times of individual product peaks. The data was acquired only for voltages where significant intensity for products was seen (150-200 V). There is a clear

increase in the average drift times of the parent ions as a function of transfer voltage in this range, consistent with the overall trend observed in **Figure 3.15**.

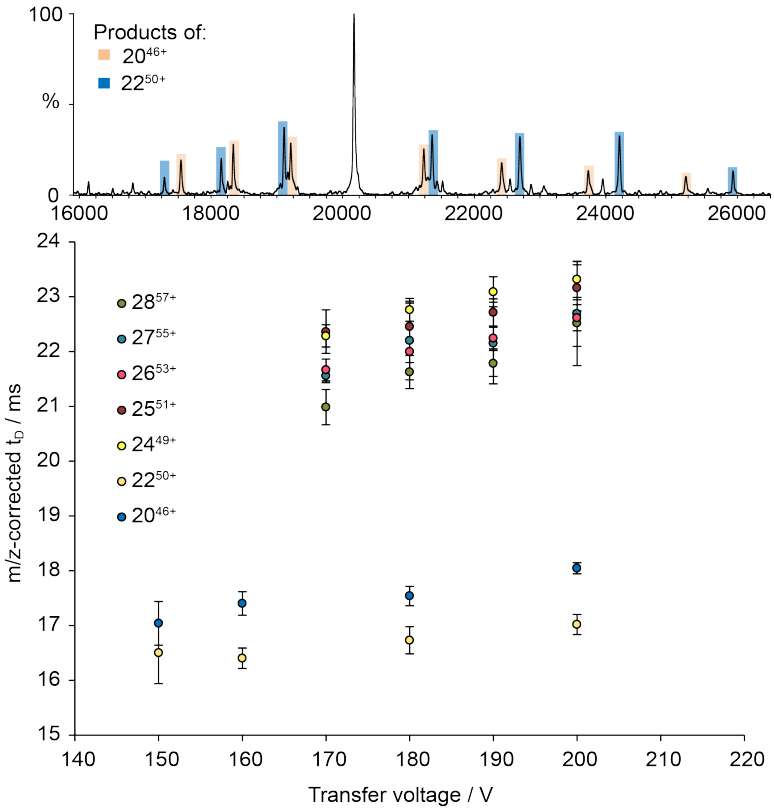



***Figure 3.16: The m/z-dependent correction does not account for the effect of transfer voltage on arrival times.*** ***a*** *Illustration of the stripped oligomer spectrum at a single voltage. In this analysis, arrival time distributions corresponding to individual product peaks were recorded, corrected for m/z-dependent flight time and averaged for a single precursor ion.* ***b*** *Variation of corrected drift times thus obtained with transfer voltage.*

I analysed this effect in more detail on a more amenable protein, *Methanocaldococcus jannaschii* Hsp16.5, which is a monodisperse 24-mer. Since the spectra are much less overlapped and it is possible to isolate a single stoichiometry and charge state, I could analyse the drift times of different types of species: the intact parent ion prior to fragmentation, stripped monomers, singly stripped oligomers and doubly stripped oligomers. The low and high energy spectra with the location of different product groups are illustrated in **Figure 3.17a**.

I ascertained that an analogous pattern of drift time voltage dependence is observed for various charge states of this sHSP. Interestingly, all product types

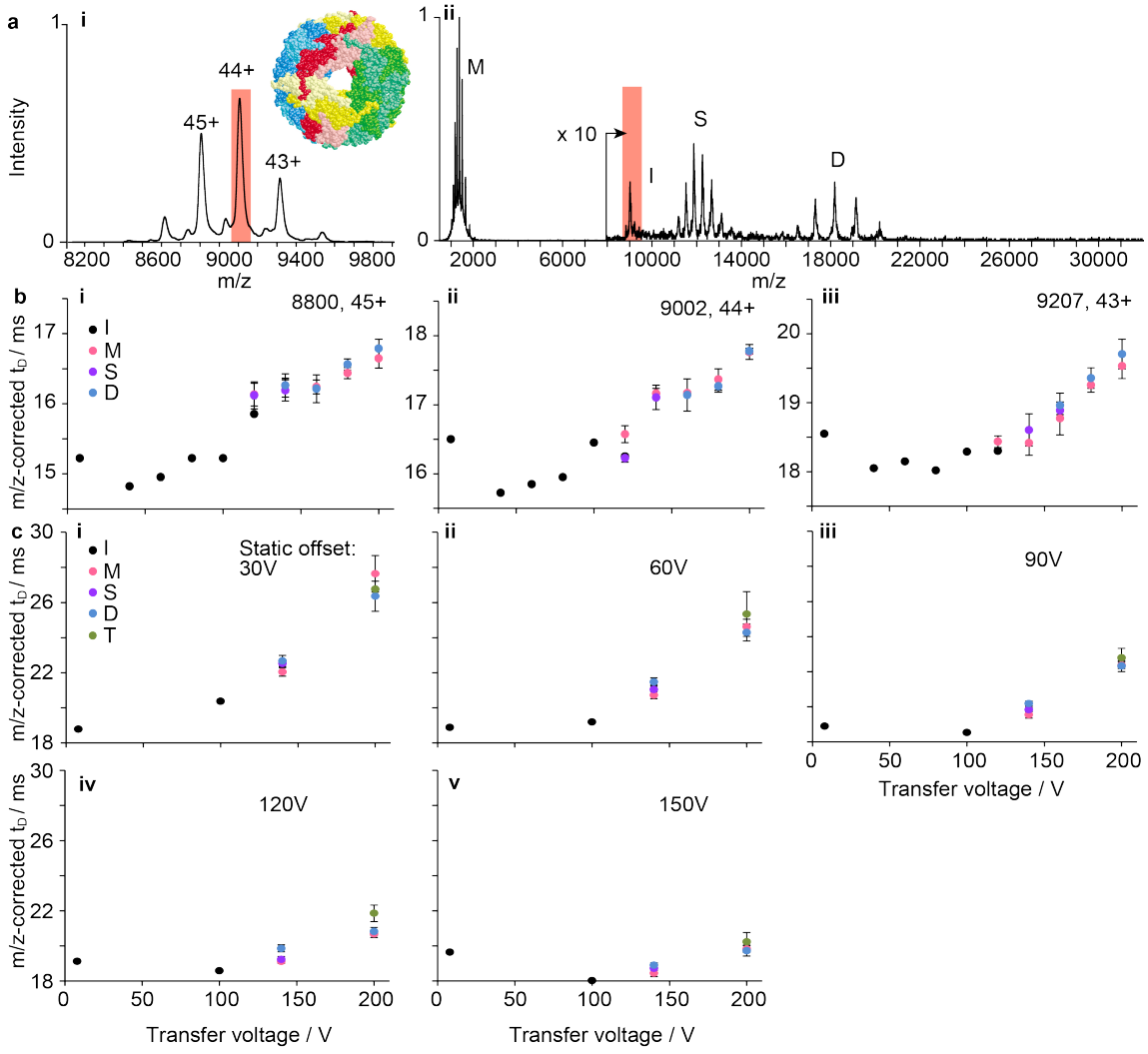



***Figure 3.17: Drift time variation with transfer voltage and static offset for Hsp16.5.*** ***a*** *Mass spectra of Hsp16.5 at (**i**) low and (**ii**) high transfer voltage. Tandem MS selection of a single charge state is indicated in pink.* ***b i-iii*** *Variation of the m/z-corrected drift times for three different charge states, 45+, 44+ and 43+, calculated from different types of species: I - intact parent ion, M - stripped monomer, S - singly stripped oligomer and D - doubly stripped oligomer.* ***c i-v*** *The experiment on the 43+ charge state at a range of static offset voltages. I, M, S, D defined as above, T - triply stripped oligomer.*

behaved similarly, with longer drift times (i.e. a deceleration) at high voltage. This data is in contrast to previous observations on a Synapt G2 instrument (Shepherd et al. [2015]), where firstly much smaller drift time distortions were observed, on the order of a few percent, and secondly different types of products behaved differently: the stripped monomers were accelerated and stripped oligomers decelerated. This could be explained by residual mobility separation of the products in the transfer cell. However, this explanation cannot apply here since there is no m/z dependence

in the behaviour of the different types of product ions.

A possible reason for the distortion is undesired interference between electrodes in the instrument, which could alter the potential the ions are experiencing in the mobility cell. To check for this, I performed the experiment at a range of static offset voltages. The static offset is the reference voltage relative to which all other voltages are set and is calculated relative to the transfer voltage. Changing the static offset thus does not alter potential differences in the instrument, which govern the passage of the ions, but does alter their absolute values. The standard value of static offset used for the above experiments was 120 V.

The data is shown in **Figure 3.17c**. It is clear that there is a large dependence of the distortion on static offset, with the offset at 30 V leading to a very large arrival time increase of almost 40%. In contrast, there is little change at the highest static offset of 150 V. A smaller offset corresponds to larger absolute voltage values. The unexpected dependence of arrival times on absolute values of instrument potentials, with constant potential differences, supports the hypothesis that the distortion is due to voltage crosstalk in the instrument.

One of the most obvious crosstalk effects is a change in travelling wave height or speed, either in the ion mobility cell or in the transfer cell. Travelling waves are a square potential superimposed on a radiofrequency (RF) voltage used to confine the ions. With the assistance of the laboratory manager Todd Mize, we connected a breakout cable assembly to the IM cell and transfer cell voltage terminals, enabling us to monitor the waves on an oscilloscope. The results are shown in **Figure 3.18**. The standard setting of the RF voltage is 350 V. For safety reasons we worked only with RF of up to 20 V in this experiment. We ascertained that there was no difference in the amplitude or periodicity of the waves with transfer voltage, at different values of the static offset or RF voltage.

Thus there was no obvious distortion of travelling wave patterns. Nevertheless, voltage crosstalk cannot be ruled out, since the exact experimental conditions

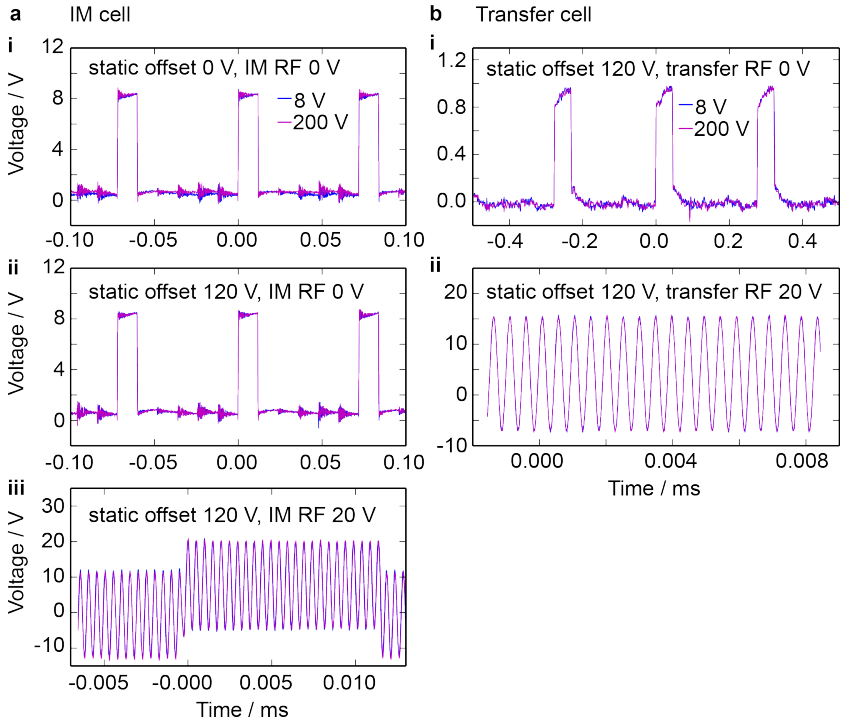



***Figure 3.18: Analysis of travelling wave patterns at different voltage settings.***[3] *Blue and red curves correspond to the data at 8 V and 200 V respectively throughout.* ***a*** *IM cell waves.* ***i-ii*** *Waves at different values of static offset, with the RF voltage turned off.* ***iii*** *Waves with a low RF voltage superimposed.* ***b*** *Transfer cell travelling waves.* ***i*** *Wave pattern without any RF voltage.* ***ii*** *Section of a wave with a low RF voltage on. No difference between wave patterns at 8 V and 200 V in the transfer cell was observed.*

cannot be replicated with the breakout cables due to the dangerously high voltages involved. It is difficult to locate the source of the problem without detailed knowledge of instrument operation and access to relevant boards. Since the arrival time distributions were shifted by up to 10% but not significantly distorted otherwise (see e.g. **Figure 3.15b i**), the easiest way forward was to perform the drift time calibration at the same transfer voltage of 200 V. As any instrumental voltage biases would be the same for the calibrants and the protein of interest, correct CCS values can be obtained in this way.

[3]The experiments shown in this figure were performed with the kind help of Dr Todd Mize.

**CCS calibration at high voltage in the transfer cell**

A number of folded protein mobility calibrants are available (Bush et al. [2010]). However, few of them are of a size comparable with $\alpha$BC. The range is further restricted by the need to record data at 200 V, since smaller calibrants are extensively fragmented and scattered at this high voltage. To add a large calibrant to my dataset, I measured the CCS of Hsp16.5 on a drift tube instrument. First, I ascertained that the conditions were sufficiently gentle not to cause any unfolding or collapse of the protein (see **Figure A.2** in **Appendix A**). I noticed that there was a slight change in arrival time between bias settings of 30 and 45 V. There was no further collapse or expansion at higher bias voltages, thus this change likely did not correspond to excessive activation. Since there was no reason to disregard either dataset, I combined the data at both voltages (**Figure 3.19a**).

The drift tube arrival time $t_A$ can be expressed as follows:

$$t_A = t_D + t_0 \tag{3.1}$$

where $t_D$ is the drift time and $t_0$ is the time spent outside the drift tube. The Mason-Schamp equation relates $t_D$ and CCS directly (Revercomb and Mason [1975], see below). However, it uses $t_D$, not $t_A$. $t_0$ can be determined simply by considering that ion mobility $K$ is defined as $K = \frac{L}{t_D V}$, where $L$ is the length of the drift tube and $V$ is the drift voltage. Therefore we can write:

$$t_A = \frac{L}{KV} + t_0 \tag{3.2}$$

Thus $t_0$ can be determined as the intercept of a plot of $t_A$ versus $\frac{1}{V}$. This plot is shown for all four visible charge states of Hsp16.5 in **Figure 3.19a i**. $t_D$ can then be calculated and inserted into the Mason-Schamp equation, **Equation 3.3**, together with the measured values of $T$ and $P$ and all other constants to obtain the CCS:

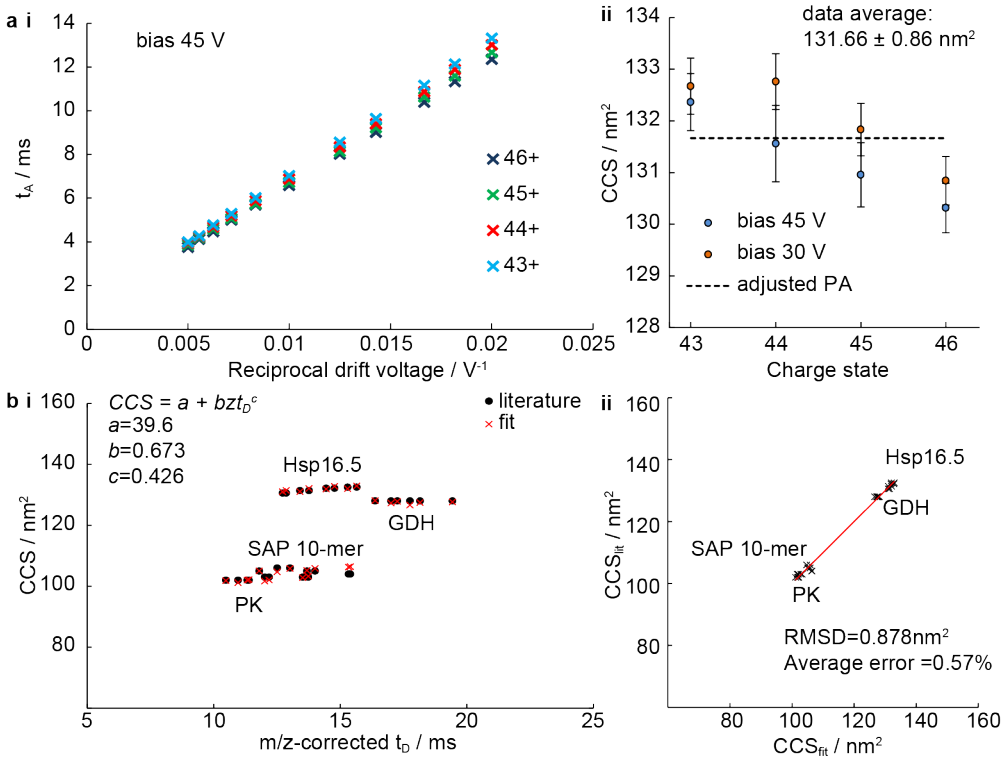



***Figure 3.19: Drift time-CCS calibration.*** ***a*** *Drift tube measurement of the CCS of Hsp16.5.* ***i*** *Plot of arrival time versus the inverse of the drift voltage for four observable charge states of Hsp16.5, at a bias voltage of 45 V. Equivalent data were recorded for the bias voltage of 30 V.* ***ii*** *The measured CCS values for each charge state at two values of bias voltage, 30 V and 45 V. The theoretical projection approximation (PA) prediction based on the crystal structure, adjusted by a factor of 1.14 (Benesch and Ruotolo [2011]) is shown as a dashed line.* ***b*** *Calibration with four proteins: Hsp16.5, glutamate dehydrogenase (GDH), SAP 10-mer (serum amyloid P) and pyruvate kinase (PK). For Hsp16.5, the data obtained in* ***a*** *was used in place of a literature value.* ***i*** *Plot of the best fit calibration function, $CCS = a + bzt_D^c$, with the parameters shown, for two sets of $t_D$ measurements for each protein performed on different days.* ***ii*** *Correlation plot of the literature CCS versus the values obtained with the best-fit calibration function.*

$$CCS = \frac{\sqrt{18\pi}}{16}\frac{ze}{\sqrt{k_bT}}\sqrt{\frac{1}{m_i}+\frac{1}{m_g}}\frac{760}{P}\frac{T}{273.2}\frac{1}{N}\frac{t_DV}{L} \tag{3.3}$$

Here $e$ is the charge on an electron, $k_b$ is Boltzmann constant, $T$ is temperature, $m_i$ is the mass of the ion, $m_g$ is the mass of the buffer gas, here a He atom, $P$ is pressure in the mobility cell, $N$ is the buffer gas number density, $V$ is the drift voltage and $L$ is the length of the cell. The CCS thus calculated for each charge state of Hsp16.5 are shown in **Figure 3.19a ii**. The average CCS obtained for the protein was $131.66 \pm 0.86\,\mathrm{nm}^2$.

A three-parameter calibration was performed to relate the measured values of $t_D$ to CCS (Baldwin et al. [2011c]):

$$CCS = a + bzt_D^c \tag{3.4}$$

where $a$, $b$ and $c$ were adjustable parameters. Calibration was first done with 7 proteins, with multiple charge states for each. I showed that the behaviour of the proteins in standard calibration at 8 V and calibration using product ions at 200 V was not dramatically different, but the two calibrations resulted in different values of the parameters $a$, $b$ and $c$ (see **Figure A.3** in **Appendix A**). To reduce the calibration error, I used the four largest proteins for final calibration at 200 V, which are likely to have similar gas-phase mobility to the large $\alpha$BC. The calibration plots with two sets of data acquired on different days are shown in **Figure 3.19b**. The root mean square deviation for the calibrants was under 1 $nm^2$ and the average error of the calibration was 0.6%.

### Variation of CCS with stoichiometry for $\alpha$BC

The final $\alpha$BC CCS values, obtained using the approach in **Figure 3.14** and calibrated as illustrated in **Figure 3.19**, are presented in **Figure 3.20**. The figure combines data from 446 product ion arrival time distributions covering 15 stoichiometries. The variation of CCS with the number of subunits is remarkably linear. There is a slight undulation of the points around the line of best fit, better visible in the plot of the residuals for each individual measurement (**Figure 3.20b**). Nevertheless the average increase in CCS upon going from an even to an odd stoichiometry is 4.3±0.6 $nm^2$, and from odd to even 4.6±0.5 $nm^2$ - very similar values. Overall, addition of a subunit thus results in an increase of 4.5 $nm^2$.

What are the implications of this trend in terms of assembly of the oligomers? Early models of $\alpha$-crystallin proposed a fluid micelle-like structure that possessed symmetry but no defined arrangement of the subunits (Walsh et al. [1991]). Al-

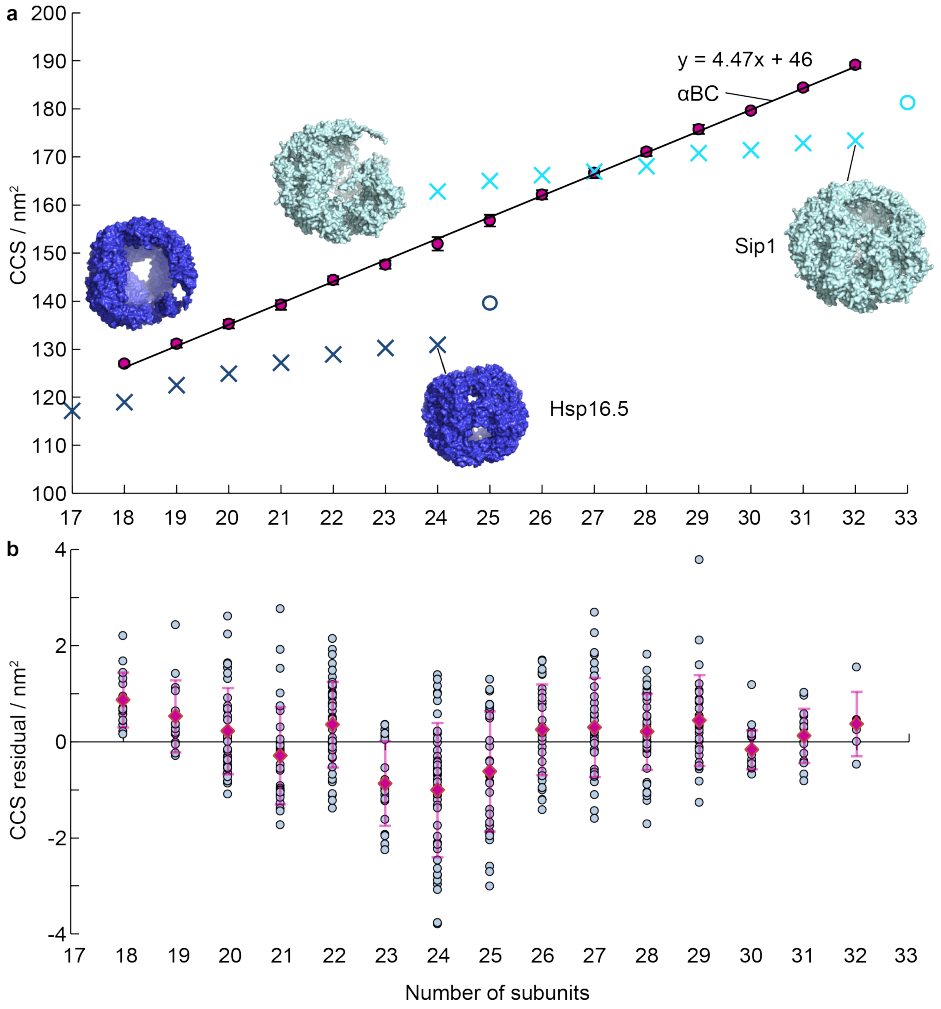



***Figure 3.20: CCS variation in αBC as a function of subunit number. a*** *Plot of CCS versus subunit number. CCS values for αBC are shown as pink circles. Error bars indicate standard deviations of measurements for the various product ions for each stoichiometry. The line of best fit with its equation is shown. Adjusted PA predictions based on crystal structures of Hsp16.5 (PDB ID 1SHS) and Sip1 (PDB ID 4YDZ) are shown as dark blue and cyan crosses respectively. The native stoichiometries are 24 and 32 respectively; predictions for smaller stoichiometries were produced by removing protein chains from the PDB files. Open circles show CCS values for theoretical structures in which a symmetrical monomer was added on the outside of the existing structure.* ***b*** *Residuals of individual measurements from the line of best fit for αBC data. The average and standard deviation of the residuals are indicated by a pink rhombus and bars respectively.*

though knowledge of detailed specific interfaces between subunits in αBC argues against such amorphous structures, they would not be inconsistent with the CCS scaling and uniform environment of αBC deduced from observation of a single set of NMR peaks (Baldwin et al. [2012]).

An alternative is a defined set of modular structures. It is known that the aver-

age projected area of convex objects constitutes a quarter of the surface area (Vouk [1948]). Assuming that $\alpha$BC oligomers are convex, like the known sHSP structures, a constant increase in CCS implies a constant increase in surface area and therefore equivalent orientation of each incoming monomer. This would be consistent with binding via defined interfaces (some of which are the intra- and inter-dimer interfaces explored earlier in the Chapter) and with observation of a single set of peaks for the monomer in solid state NMR (Jehle et al. [2009]). Some flexibility in orientation angles would clearly be necessary to accommodate monomers in increasingly larger structures, which may be made possible by the plasticity of the N- and C-termini.

Also included in **Figure 3.20** are plots of adjusted theoretical projection approximation (PA) CCS for two sHSPs for which crystal structures are available: *M. jannaschii* Hsp16.5 and *C. elegans* Sip1. Their CCS is smaller than the equivalent $\alpha$BC, which is expected since they have a smaller monomer mass. I produced hypothetical CCS predictions for lower stoichiometries of these proteins by removing a subunit at a time and recalculating the CCS. These hypothetical structures are no longer convex, but do nevertheless still show roughly linear scaling with stoichiometry. The average CCS difference per subunit was only 2.0 and 1.3 $nm^2$ respectively, much less than that observed for $\alpha$BC. Thus more dramatic structural rearrangements must occur in the latter. I also made a CCS prediction for addition of an exposed monomer on the outside of the protein shells, thus constructing a 25-mer for Hsp16.5 and a 33-mer for Sip1 (circles in **Figure 3.20a**, see **Figure A.4** in **Appendix A** for images). The rise per subunit was 8-9 $nm^2$, more than enough to account for the rise seen for $\alpha$BC. Thus assembly of $\alpha$BC is more likely to occur by addition of relatively exposed monomers on the outside of the existing structures, rather than by monomers fitting into gaps. A similar linear pattern of CCS versus stoichiometry was also experimentally observed for wheat Hsp16.9 (Shepherd et al. [2015]), with a rise per subunit of about 5.5 $nm^2 monomer^{-1}$. This protein has a monomer mass of about 17 kDa, 3 kDa smaller than $\alpha$BC. Thus incoming monomers must be less exposed in $\alpha$BC than in Hsp16.9.

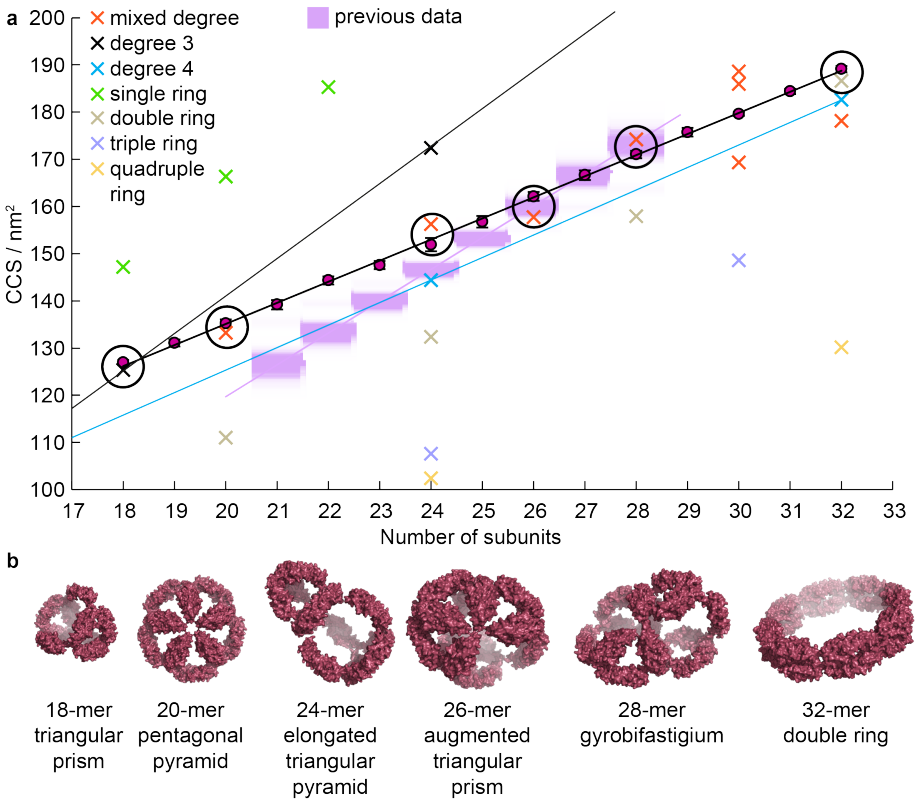



***Figure 3.21: Prediction of candidate oligomer structures for αBC. a*** *Plot of CCS versus subunit number for αBC and various types of theoretical assemblies of core domain dimers, taken from Baldwin et al. [2011c]. The degree of a structure refers to the number of subunits that come together at a vertex. The fitted data from Baldwin et al. [2011c] are superimposed on the figure in purple. Black circles highlight close matches of a prediction to the experimental measurement.* ***b*** *Illustrations of the matching candidate structures highlighted by black circles for six stoichiometries.*

The data can also be compared to previously obtained CCS ranges produced by fitting the 2D IMMS spectrum (**Figure 3.21**, Baldwin et al. [2011c]). Importantly, the experimental measurements show that the assumption of a linear relationship between CCS and stoichiometry is justified. Nevertheless, the previously fitted gradient and intercept are 6.6 $nm^2 monomer^{-1}$ and -11 $nm^2$ whereas the experimental values are 4.5 $nm^2 monomer^{-1}$ and +46 $nm^2$. Thus the fits match the data very well for 26-28-mers, but diverge for the lower stoichiometries. The closest matching candidate structures are still an augmented triangular prism and a gyrobisfastigium for the 26-mer and 28-mer. For the 24-mer, the experimental CCS is closer to an elongated triangular pyramid than to an octahedron. Moreover, new candidate structures can be proposed for the 18-mer, the 20-mer and the 32-mer. The slope

of the experimental CCS line can be compared to the line connecting the degree 3 polyhedra (black) and the line for degree 4 polyhedra (cyan). The experimental CCS values are closer to degree 3 polyhedra at lower stoichiometries but closer to degree 4 for larger stoichiometries, reflecting the need to accommodate more subunits at a vertex for larger oligomers. A higher vertex degree requires shallower connection angles. The structure of the $\alpha$BC molecule with a central domain and mobile tails gives $\alpha$BC the flexibility to accommodate this range of structures.

The judgement on which candidate structures are correct should be based on generation of a larger structural database with a greater range of hypothetical topologies, taking account of the flexible termini, and validating models, for example by mutagenesis. Reconstitution of subpopulations of $\alpha$BC from monomers crosslinked at the inter-dimer interface may enable detailed characterisation of some of the structures and validation of the modelling predictions, and pave the way for structure determination of some of the oligomers using cryoEM. Observation of analogous linear trends in CCS versus stoichiometry in human $\alpha$BC and plant Hsp16.9 suggests that structural plasticity may be a general feature of polydisperse sHSPs. Extension of this work to other sHSPs will show whether this is the case. A further appealing direction for future work is analysis of modified forms of sHSPs, e.g. phosphorylated forms of $\alpha$BC, which might reveal how the modifications regulate the architecture of the ensemble.

## 3.4 Conclusions

In summary, IMMS and inter-dimer crosslinking data presented in this Chapter supports mixed-degree polyhedral architectures of $\alpha$BC. There is a surprisingly smooth linear relationship of CCS versus subunit number, including both odd and even stoichiometries, and $\alpha$BC monomers are important players in exchange of subunits between oligomers. The slope of the CCS line and the crosslinking data indicate the existence of a range of vertex degrees: 2, 3, 4 and higher. The monomers must

therefore be connected at sharper or shallower angles, which is made possible by the plasticity of the $\alpha$BC monomer. The data highlights the power of native MS and IM for structural and dynamic investigation of species in a polydisperse ensemble.

# Chapter 4

# Kinetic analysis of $\alpha$Lac aggregation

*One of the key functions of sHSPs in general, and $\alpha$BC in particular, is inhibition of pathological aggregation. To find out how they achieve this, it is first necessary to understand aggregation of a substrate protein in the absence of chaperone. This is the objective of this Chapter. I present the development of a methodology to look at mechanism of aggregation in quantitative detail, using the model protein $\alpha$Lac as an example. I show that NMR data can be used as a well-defined metric of aggregation, and that it provides sufficient power to distinguish between many kinetic models. I use the methodology to show that $\alpha$Lac aggregation is nucleated and proceeds via at least two parallel pathways.*

## 4.1 Introduction

The molecular mechanism for aggregation is the set of events that lead from soluble protein to the equilibrium ensemble of aggregates, oligomers and residual monomer (or other stoichiometry of origin), together with the kinetic and thermodynamic parameters that govern these reactions. The mechanism depends on the nature of the protein and on the solution conditions. The knowledge of aggregation mechanisms is a prerequisite for understanding how aggregation might cause disease, and how it might be modulated to tackle the disease, for example by addition of molecules that stall aggregation or direct it on a different pathway avoiding the most toxic species.

There are few methods that can deliver comprehensive kinetic and thermodynamic information about the variety of species on the aggregation pathway. Methods based on empirical fitting of light scattering, dynamic light scattering or NMR data have been used to quantify aggregation (e.g. Carver et al. [2002], Stranks et al. [2009], Bumagina et al. [2010]). They are helpful for comparing aggregation propensity under different conditions but do not link the macroscopic changes in the data to the molecular events that underlie them. Native mass spectrometry can give insight into the variety of species that are present especially during early stages of the aggregation process (e.g. the work of Smith et al. [2006] on $\beta$2-microglobulin or Young et al. [2015] on a screening method for small-molecule amyloid inhibitors). Nevertheless, it has limited buffer compatibility and converting spectrum intensity into absolute concentrations is challenging.

A state-of-the-art approach is to apply chemical kinetics models to data that report on one or more species in the aggregating mixture. For example, Cremades et al. [2012] used single-molecule fluorescence combined with chemical kinetics to track populations of $\alpha$-synuclein oligomers and derive parameters of their interconversion. A recently developed, more comprehensive method, applicable to amyloid aggregates, involves recording ThT fluorescence traces for different concentrations of the protein, which report on the total concentration of aggregates, and fitting these

curves with kinetic mechanisms (Meisl et al. [2016]). For example, in this way it was possible to show that A$\beta$42 aggregates according to a mechanism that includes surface-catalysed secondary nucleation, but not fragmentation, in the solution conditions under study, and ascribe rate constants for primary nucleation, elongation and secondary nucleation (Cohen et al. [2013]).

In this Chapter, I describe a new chemical kinetics approach to derive aggregation mechanisms. It is based on following the disappearance of monomeric species during aggregation by NMR and using numerical simulations of trial mechanisms to fit the data at a range of protein concentrations. Unlike the approach based on ThT data, it is not restricted to amyloid aggregates and does not rely on the use of analytical solutions to the problem, allowing a wider diversity of aggregation models to be tested. To enable the development of a robust methodology, I focused on the readily available model protein bovine $\alpha$-lactalbumin ($\alpha$Lac). It is a known *in vitro* substrate for $\alpha$BC, and **Chapter 5** deals with the extension of the approach to enable the study of chaperone action of $\alpha$BC on $\alpha$Lac.

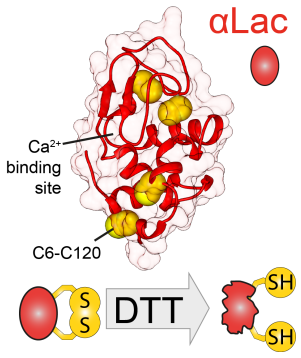



***Figure 4.1: The structure of apo $\alpha$Lac*** *(PDB ID 1F6R). The disulphide bridges are shown as yellow spheres. The Cys6-Cys120 bond, which is broken first in aggregation, and the location of $Ca^{2+}$ ion binding site in holo $\alpha$Lac is annotated. Bottom panel: breakage of the disulphide bonds by a reducing agent such as DTT causes misfolding of $\alpha$Lac and makes it aggregation-prone.*

$\alpha$Lac is a 14.2 kDa globular protein found in bovine milk. It is the regulatory component of lactose synthetase, the enzyme that produces lactose from glucose and UDP-galactose (Fitzgerald et al. [1970]). $\alpha$Lac has a mixed $\alpha$-helix and $\beta$-sheet structure and is related to the lysozyme family (**Figure 4.1**). It binds one $Ca^{2+}$ ion per protein molecule which can be stripped without substantial changes to protein structure (Chrysina et al. [2000]). $\alpha$Lac contains 8 cysteines which form four disulphide bonds. Reduction of these bonds leads to aggregation and precipitation of the protein. The timescale of breakage of each disulphide has been characterised

(Carver et al. [2002]). The first critical bond to be broken is the Cys6-Cys120 bond. This allows formation of a kinetic molten globule state, i.e. a state in which the native secondary structure is still present but there is no well-defined tertiary structure (Ewbank and Creighton [1991], Lindner et al. [2001]). The remaining three disulfides are broken at a significantly slower rate. The transient nature of the kinetic molten globule has made it difficult to characterise, but a lot of work has been done on equilibrium molten globules which $\alpha$Lac forms in a variety of conditions: at a moderate concentration of a denaturant such as urea, at low pH (the so-called A-state), at neutral pH in the absence of $Ca^{2+}$ and salt (reviewed in Kuwajima [1996]), and if some or all of the cysteines are mutated to alanine. In these molten globules, the $\alpha$-helical domain is more structured whereas the $\beta$-sheet domain is largely unfolded (Wu et al. [1995], Schulman et al. [1997]).

Formation of large aggregates following the conformational transition to the molten globule state has been studied by light scattering, DLS, SAXS, ThT fluorescence and NMR (Horwitz et al. [1998], Lindner et al. [2001], Carver et al. [2002], Kulig and Ecroyd [2012], Hochberg et al. [2014]). Bumagina et al. [2010] proposed formation of "start aggregates" of 80-100 molecules of $\alpha$Lac which associate to form larger oligomers, based on DLS data. This contrasts with the observation of a continuous size increase of $\alpha$Lac oligomers by Bettelheim et al. [1999] and Regini et al. [2010] by DLS and SAXS. Carver et al. [2002] followed the changes in the apo $\alpha$Lac spectrum over time following addition of DTT. They observed a change in the appearance of the spectrum within seconds, consistent with misfolding, and a decay in the intensity of the spectrum over time as $\alpha$Lac was incorporated into large aggregates. They quantified the apparent exponential rate of intensity loss for a resolved tyrosine peak with a rate constant of $1.50(\pm 0.04)\times 10^{-3}\,\mathrm{s}^{-1}$ at 37°C (140 $\mu$M), but did not relate it to the association and dissociation reactions the protein might be involved in. Stranks et al. [2009] proposed a mathematical model of light scattering traces based on empirical scaling laws and geometric assumptions. Although it is unclear whether the nonlinear dependence of light scattering signal on aggregate size

and geometry was taken into account, the authors report that light scattering data of $\alpha$Lac aggregating amorphously are consistent with a nucleated pathway where more than one monomer is involved in nucleus formation.

In summary, while the equilibrium molten globule state of $\alpha$Lac and the empirical features of its aggregation have been well studied, a comprehensive mechanism of $\alpha$Lac aggregation has not yet been reported. Here, I use NMR in combination with chemical kinetics simulations to obtain a quantitative mechanistic description of $\alpha$Lac aggregation and derive the kinetic and thermodynamic parameters of its interconversion between the different states. I show that key processes are monomer misfolding and formation of the dimeric nucleus, and that aggregation proceeds via at least two pathways which differ in the kinetics and thermodynamics of nucleation.

## 4.2 Materials and methods

**Materials** Full-length human $\alpha$BC (UniprotKB ID P02511) and the core domain construct (residues 68-153) were prepared as described in **Sections 3.2** and **2.2** respectively. Bovine $\alpha$Lac was purchased from Sigma Aldrich (L6010). All reagents were purchased from Sigma Aldrich unless stated otherwise. Carbon-coated electron microscopy grids were purchased from Electron Microscopy Sciences (catalogue number CF300-CU-50).

**Aggregation reactions** Unless stated otherwise, light scattering assays were performed at 37°C in 50 mM sodium phosphate pH 7.0, 100 mM NaCl and 2mM EDTA. For NMR-based assays, 5% $D_2O$ was also added. Aggregation was initiated by addition of 20 mM DTT.

**Light scattering and ThT fluorescence assays** The assays were performed on a FLUOstar Omega spectrophotometer (BMG LabTech). Light scattering was measured by changes in apparent absorbance at 340 nm. For ThT fluorescence,

excitation was at 440 nm and emission was measured at 485 nm. Data was recorded every 60 s.

**FTIR spectroscopy measurements** Experiments were performed on a Varian FTS7000 spectrometer equipped with a DuraSamplIR Diamond attennuated total reflectance (ATR) element. 2 mM $\alpha$Lac samples were allowed to aggregate before centrifugation at 10,000g for 5 min. The pelleted aggregates were transferred onto the ATR element. For the no-aggregation control, the spectrum for lyophilised $\alpha$Lac powder was measured. Solvent signal was subtracted and the data was normalised to the highest point.

**EM** Transmission electron microscopy was performed on a Tecnai 12 (FEI) transmission electron microscope operated at 120 kV. Samples at 5 $\mu$M in 5$\mu$l were loaded on glow-discharged carbon grids, washed with water and stained with 2% uranyl acetate for 10 s. Images were acquired with a Gatan OneView CMOS camera.

**NMR-based aggregation assays** NMR experiments were performed on a narrow-bore Varian 14.1 T NMR spectrometer with a 5-mm z-axis gradient triple resonance room temperature probe for solution samples. The following parameters were typically used for $^1$H watergate experiments: acquisition time 100 ms, number of points 1798, 32 scans per free induction decay (FID), recycle delay 1 s. The duration of the $^1$H watergate experiments was 40 s. Spectra were analyzed using NMRPipe (Delaglio et al. [1995]). Processing included digital water suppression where necessary, exponential apodization at 10 Hz, phase correction and linear baseline correction. The range 0.77-2.73 ppm was taken for integration (no buffer components present). Data for five $\alpha$Lac concentrations was acquired as follows: 100 $\mu$M - 4 repeats, 200 $\mu$M - 4 repeats, 400 $\mu$M - 2 repeats, 800 $\mu$M - 1 repeat, 2000 $\mu$M - 2 repeats.

**NMR diffusion experiments** $^{1}$H, $^{15}$N-edited $^{1}$H and $^{13}$C-edited $^{1}$H pulsed field gradient diffusion experiments were carried out over 11 gradient field strengths (see 2.2) and a diffusion delay of 100 ms or 200 ms, with 1600 ($^{1}$H), 576 ($^{15}$N$^{1}$H) or 574 ($^{13}$C$^{1}$H) points, an acquisition time of 200 ms ($^{1}$H) or 64 ms ($^{15}$N$^{1}$H,$^{13}$C$^{1}$H) and 88 ($^{1}$H), 80 ($^{15}$N$^{1}$H) or 64 ($^{13}$C$^{1}$H) transients per FID. The gradient duration was 2 ms ($^{1}$H), or 0.5 ms ($^{15}$N$^{1}$H, $^{13}$C$^{1}$H). Exponential decay in signal at each ppm point in the spectrum was analysed independently and the diffusion coefficient and its error was determined as the mean and standard deviation of a Gaussian distribution fitted to these data (see example analysis in **Figure 2.6**).

**Chemical kinetics simulations and numerical fitting** Briefly, complete kinetic models were constructed to follow the change in concentration of individual oligomeric species (up to 4,000-mers were considered), parameterised by a number of rate constants. These were integrated numerically and used to simulate the experimental data. The observable was defined as the total concentration of monomeric protein, with a relative intensity factor for the misfolded species which was treated as a fitting parameter. The data was averaged across repeats and downsampled to 17 points per $\alpha$Lac concentration, while preserving the shape of each curve. Three out of the five concentrations (100 $\mu$M, 400 $\mu$M and 2000 $\mu$M) were included in the model selection process, with the two remaining concentrations (200 $\mu$M, 800 $\mu$M) set aside to be fit only with the final model as a validation step. Rate constants were optimised using an algorithm that combined Levenberg-Marquardt minimisation (Press et al. [2007]) and a basin hopping routine (Olson et al. [2012]). The spectral intensity factor for the second misfolded form of $\alpha$Lac was also included in the fit as a free parameter. The $\chi^2$ values of the fits to the experimental NMR time-courses were compared using the F-test ($p<0.05$). The errors in the parameters were estimated by performing the fit for each of the five concentrations independently.

## 4.3 Results and discussion

### 4.3.1 Characterisation of αLac aggregation by light scattering

I first sought to look at the aggregation behaviour of αLac by light scattering, an established method. **Figure 4.2a** shows light scattering traces produced by αLac aggregating over time at a range of concentrations. The increase in light scattering was sigmoidal and occurred on the timescale of minutes to hours in phosphate buffer at 37°C. The length of the lag period before the sigmoid rise and the final level of signal were dependent on total protein concentration. The curves were most reproducible at higher concentrations of αLac. At 50 μM or in the absence of reducing agent no scattering was observed. The increase in scattering could be completely suppressed by the addition of either αBC or cαBC in a 1:1 ratio, at 200 μM each (**Figure 4.2b**). This is in line with previously published data (Carver et al. [2002], Bumagina et al. [2010], Hochberg et al. [2014]).

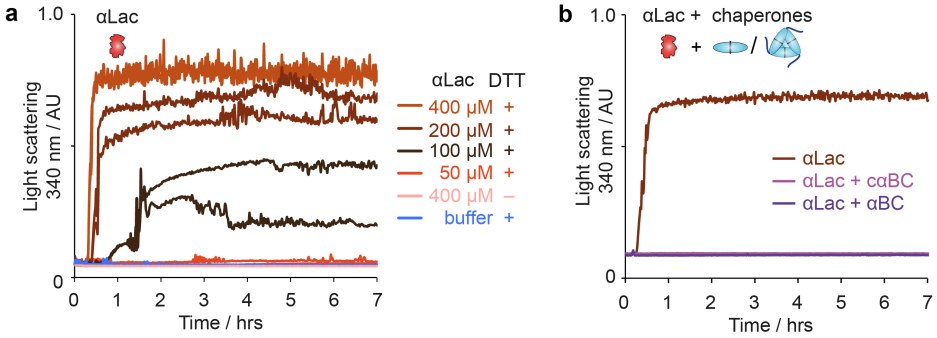



***Figure 4.2: αLac light scattering assays in phosphate buffer.*** ***a*** *αLac assay at total protein concentrations between 50-400 μM. No DTT and buffer-only controls are also included.* ***b*** *Light scattering traces in the absence and presence of either full-lenght αBC or the isolated αBC core domain (cαBC). Each protein was at 200 μM.*

Interpreting the light scattering curves to describe mechanism is challenging. Light scattering has a complex, non-linear dependence on the size and shape of aggregates and can be analysed quantitatively only in ideal systems, such as spherical particles (Hall et al. [2016]). Therefore, I limited myself to exploring in a semi-

quantitative way to what extent the kinetics of aggregation could be manipulated by solution conditions. I tested the influence of several factors described below.

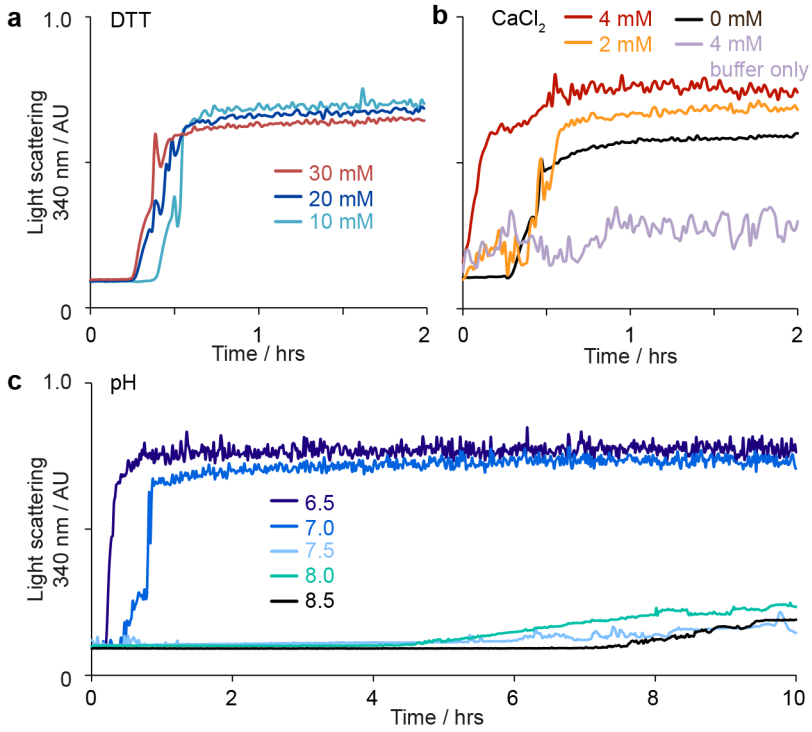



***Figure 4.3: Effect of solution conditions on αLac light scattering assay.*** ***a*** *DTT concentration;* ***b*** *addition of calcium chloride;* ***c*** *pH of the phosphate buffer.*

**DTT** Since I was interested in intrinsic aggregation properties of $\alpha$Lac, it was preferable to avoid conditions in which the reducing agent was rate-limiting. I therefore compared traces as a function of DTT concentration (**Figure 4.3a**). I found that DTT had a minor effect on the kinetics in this concentration range. The lag time was about 12 minutes longer at 10 mM DTT. There was little difference between the traces at 20 and 30 mM DTT, suggesting that DTT was already close to saturating at 20 mM.

**$Ca^{2+}$ ions** $Ca^{2+}$ is known to stabilise $\alpha$Lac and reduce the rate of reduction of its disulfide bonds (Carver et al. [2002]). Nevertheless, in the presence of $CaCl_2$ (in imidazole buffer), the protein is known to precipitate faster than without $CaCl_2$ (in phosphate buffer), possibly due to the dominance of different partially reduced intermediates with greater exposed hydrophobic area and therefore higher aggrega-

tion propensity (Carver et al. [2002]). I tested the feasibility of using $CaCl_2$ as an additive to manipulate the rate of aggregation (**Figure 4.3b**). Addition of $CaCl_2$ in phosphate buffer did lead to shorter lag times and higher plateau signal levels, thus suggesting that $Ca^{2+}$ ions indeed enhance aggregation, but the control experiment revealed significant precipitation of $CaCl_2$ itself (purple curve in **Figure 4.3b**). Thus poor solubility of $CaCl_2$ in phosphate buffer means it is of limited use for manipulating the rate of aggregation.

**Effect of pH** DTT becomes more reactive as a reducing agent as the pH is increased from 6.5 towards its thiol pKa of 9.2, since it is its deprotonated form that acts as a nucleophile (Whitesides et al. [1977], Lukesh et al. [2012]). One could therefore expect that $\alpha$Lac aggregation would speed up at higher pH. Changes in pH did indeed have a dramatic effect on aggregation, but with more acidic solution conditions strongly enhancing aggregation and more basic conditions slowing it down to beyond practical (**Figure 4.3c**). The observation that aggregation in fact slows down at higher pH could have several explanations. Firstly, in line with the effect of $Ca^{2+}$ ions, which impede full reduction of disulfide bonds, at lower pH DTT has less reducing power, so the partially reduced species, which could be more aggregation-prone, could become more prevalent. Alternatively, there could be pH-dependent changes in the conformation of the protein that affect the accessibility of disulfide bonds or the protein's surface properties, and therefore aggregation propensity.

**Ammonium acetate** To check whether aggregation assays could be performed in ammonium acetate, a buffer compatible with native MS, I carried out experiments shown in **Figure 4.4**. Panel **a** shows two comparisons between sodium phosphate and ammonium acetate buffer conditions at the same $\alpha$Lac concentration. In ammonium acetate, the lag phase before the increase in light scattering was shorter, but the final level of signal was lower, suggesting that aggregation is faster but the

steady-state distribution of oligomers is shifted to smaller sizes. Thus the kinetics and characteristics of the aggregates are different in the two buffers. To slow down the reaction, I used only 1 mM DTT in the following experiments. **Figure 4.4b** demonstrates that, unlike in sodium phosphate buffer, neither $\alpha$BC nor c$\alpha$BC have a significant effect on $\alpha$Lac aggregation in ammonium acetate. This suggests that they do not act as effective chaperones in these solution conditions. The result is consistent with previous observations that $\alpha$BC is less effective as a chaperone for species that aggregate fast (Lindner et al. [2001], Carver et al. [2002]), providing support to the idea of kinetic competition between aggregation and chaperone binding (Lindner et al. [2001], Jacobs et al. [2016]). Since we were interested in developing methodology that would allow the study of chaperone function of $\alpha$BC, sodium phosphate buffer was used as the solution of choice for the remainder of the experiments described in this thesis.

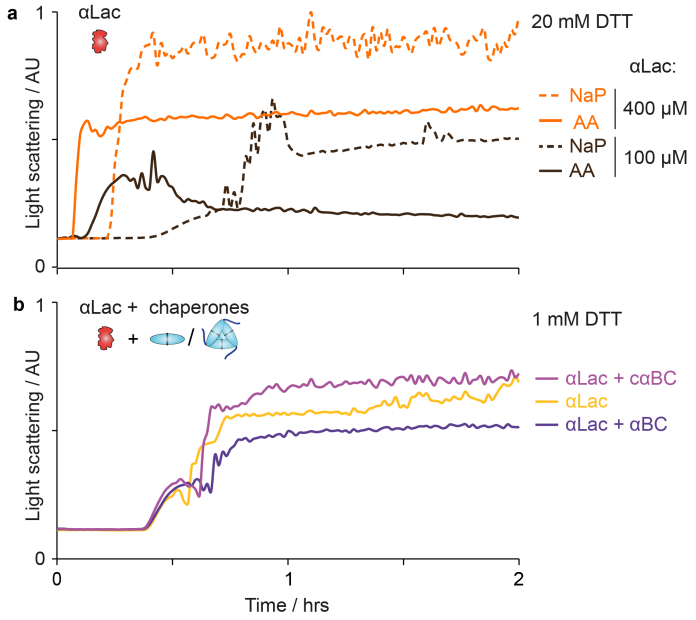



***Figure 4.4: $\alpha$Lac light scattering assays in ammonium acetate.*** ***a*** *Comparisons of reactions in sodium phosphate and ammonium acetate at the same $\alpha$Lac concentrations: either 100 $\mu$M or 400 $\mu$M.* ***b*** *Reaction at 100 $\mu$M $\alpha$Lac with or without $\alpha$BC or c$\alpha$BC in a 1:1 ratio to $\alpha$Lac.*

In summary, I could follow $\alpha$Lac aggregation by light scattering at 340 nm, obtaining results in agreement with previous studies. The reaction was sensitive to solution conditions such as pH and the nature of the salts in the buffer. It is clear that the effect of particular solution conditions on the kinetics of aggregation of $\alpha$Lac is not easily predictable. This is not surprising as many factors are at play: reactivity of the reducing agent, accessibility and reactivity of disulfide bonds,

conformation of any intermediates and their stability in the presence of ligands, acids, bases or salts.

### 4.3.2 $\alpha$Lac shows features of amyloid aggregation

In terms of the morphology of the aggregates, $\alpha$Lac is known to form both amorphous and amyloid aggregates depending on the conditions used. $\alpha$Lac saturated with $Zn^{2+}$ forms amorphous aggregates (Goers et al. [2002]). Conditions favouring amyloid formation include reduction and carboxymethylation (Kulig and Ecroyd [2012]), low pH and low salt concentration (Goers et al. [2002], Ghahghaei et al. [2010]). Amyloid aggregates are the form associated with many human diseases such as Alzheimer's, Parkinson's and type II diabetes (Eisenberg and Jucker [2012], Knowles et al. [2014], Eisele et al. [2015]). I therefore sought to determine the behaviour of the system under the conditions employed here. I used three different methods as set out below.

**ThT fluorescence** ThT is an aromatic small molecule which fluoresces at 482 nm when it intercalates between $\beta$-sheets in amyloid fibrils (Xue et al. [2017]). To check whether $\alpha$Lac aggregates bind ThT, I performed an assay in which light scattering and ThT fluorescence was measured in parallel for three $\alpha$Lac concentrations and controls (**Figure 4.5a**). There was a clear concentration-dependent increase in ThT fluorescence in $\alpha$Lac samples reduced with DTT, and no increase in buffer or no-DTT controls. The initial decrease in fluorescence has been observed previously and linked to the dependence of ThT fluorescence on temperature, as the samples warmed up to 37°C (Ecroyd et al. [2008], Hudson et al. [2009]). Thus, formation of $\alpha$Lac aggregates results in ThT fluorescence, as would be expected from amyloid aggregates.

**Fourier transform infra-red spectroscopy** FTIR spectroscopy is a method that reports on secondary structure of proteins. It is particularly useful for aggregates as it can be performed on solid samples. It is used extensively for character-

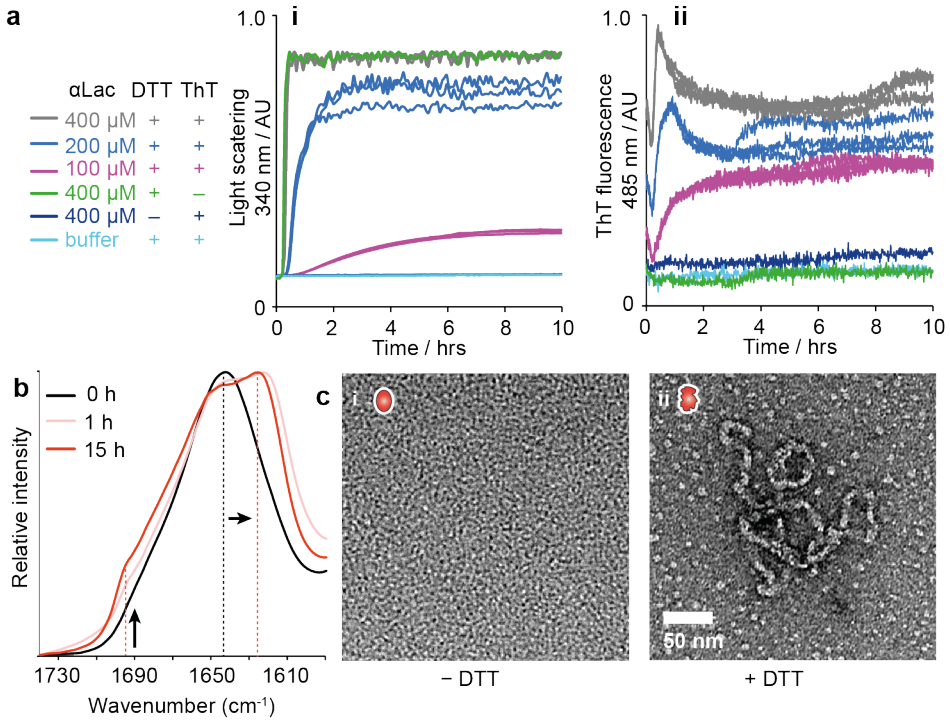



***Figure 4.5: αLac shows features of amyloid aggregation.*** ***a*** *Parallel light scattering (**i**) and ThT fluorescence measurements (**ii**) at a range of αLac concentrations. Increases in both light scattering and ThT fluorescence are observed at higher αLac concentrations.* ***b*** *FTIR spectroscopy of soluble αLac (black) and its aggregates at early (pink) and late (red) timepoints. The emergence of a peak around 1680 cm$^{-1}$ and the shift of the main peak towards lower wavenumbers are characteristic of increased β-sheet content in the aggregates.* ***c*** *Negative stain EM of soluble αLac (**i**) and 12 hours after the addition of a reducing agent (**ii**). Large fibrillar particles are observed, consistent with formation of amyloid.*

ising protein aggregates (Sarroukh et al. [2013]). I obtained FTIR spectra of αLac in three states: lyophilised powder prior to reduction with DTT, protein pellet one hour after reduction and 15 hours after reduction (**Figure 4.5b**). The amide-II band of the spectrum, between 1730 and 1640 cm$^{-1}$, shows features characteristic of the appearance of β-sheets following reduction: the shift of the main peak to lower wavenumbers and the appearance of a peak around 1690 cm$^{-1}$. This data is therefore also consistent with formation of amyloid aggregates by αLac upon reduction with DTT.

**Negative stain EM** Protein aggregates can be observed using negative stain electron microscopy. After careful optimisation of the grid preparation protocol (see

**Figure A.5** in the **Appendix**), $\alpha$Lac samples before and after aggregation were compared. A sample of $\alpha$Lac before the addition of DTT showed no visible features, as expected since $\alpha$Lac monomers are too small to be observed by EM (**Figure 4.5c i**). In contrast, upon addition of DTT two morphologies of aggregate were observed (**Figure 4.5c ii**): smaller spherical aggregates with an average radius of about 4.7± 1.0 nm and longer, flexible string-like aggregates of thickness 9.2±2.9 nm. They are similar to the amyloid aggregates previously documented for apolipoprotein C-II (Hatters et al. [2000, 2003]). No significant changes in qualitative features were observed at different $\alpha$Lac concentrations or samples taken at different timepoints of aggregation. Quantitative treatment, except a simple estimate of average particle size, was not attempted with the EM data as this was difficult to do comprehensively.

Taken together, the ThT fluorescence, FTIR spectra and negative stain electron microscopy data strongly suggest that at least a significant fraction of $\alpha$Lac aggregates have amyloid-like properties. Nevertheless it cannot be excluded that some fraction of non-amyloid, amorphous aggregates are also formed.

### 4.3.3 Following aggregation of $\alpha$Lac monomer by NMR

NMR is sensitive to the size of the particles under investigation - small species tumble quickly, leading to slower relaxation and sharp NMR signals, whereas signal broadens for large species, essentially making them invisible. Real-time NMR has previously been used to follow aggregation of proteins (e.g. Carver et al. [2002], Svane et al. [2008], Mishra et al. [2009]). It was therefore interesting to see how aggregation would affect NMR spectra of $\alpha$Lac. **Figure 4.6** shows a comparison between a light scattering experiment and an NMR time series experiment at 200 $\mu$M $\alpha$Lac.

In the absence of reductant, apo ($Ca^{2+}$-free) $\alpha$Lac showed a rich 1D $^1$H NMR spectrum typical of a well-folded protein (**Figure 4.6b i**). On addition of DTT, the spectrum changed within the 3-minute dead time required to manipulate the

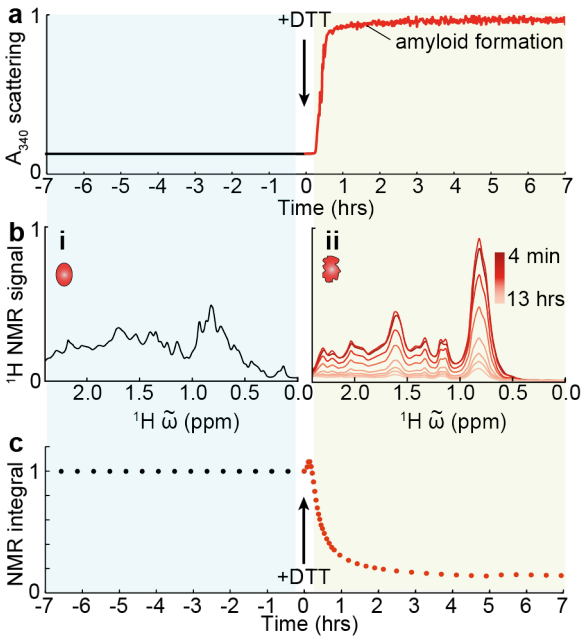



***Figure 4.6: Following αLac aggregation by NMR.*** ***a*** *A light scattering experiment at 200 μM αLac.* ***b i*** *0.0-2.5 ppm region of the 1D $^1H$ NMR spectrum of αLac prior to the addition of DTT.* ***ii*** *Time series of 1D $^1H$ spectra of αLac after the addition of DTT.* ***c*** *A plot of the integral under the spectra in* ***b*** *over time for 200 μM αLac.*

sample, with both position and intensity of the peaks affected (**Figure 4.6b ii**). The spectrum generally became less resolved, consistent with a conformational change to a molten globule observed and studied previously (Rajaraman et al. [1998], Carver et al. [2002], Rösner and Redfield [2009]). To follow aggregation over time, I recorded a series of 40-second $^1H$ experiments over 13 hours, some of which are shown in **Figure 4.6b ii**. There was little further change to the shape of the spectrum, but the overall intensity decreased over time, consistent with formation of large oligomers which are not directly visible by NMR. I quantified this change by calculating the integral of the spectrum where no peaks from buffer components were present (0.77-2.73 ppm) and plotting this integral over time relative to the integral before addition of DTT (downsampled curve shown in **Figure 4.6c**). Quantified in this way, there was a small initial increase in signal in the first few minutes due to the conformational rearrangement followed by decrease in signal by 80-90% as aggregation proceeded. The most significant decrease in NMR signal occurred at the same time as the rise in light scattering signal (compare **Figures 4.6a** and **c**).

To understand more about the nature of the observable in the NMR experiment, I recorded NMR diffusion experiments, measuring the diffusion coefficient of visible αLac species before and after aggregation, as well as several controls (**Figure**

**4.7**).

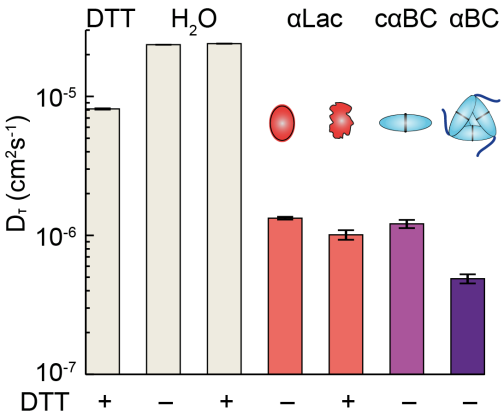



***Figure 4.7: Probing the NMR observable with diffusion measurements.*** *Diffusion coefficients of (left to right) DTT, water in a mixture of αLac and cαBC to rule out any differences in convection upon aggregation, αLac before addition of DTT, αLac visible in the NMR spectrum at 12 hours after the addition of DTT, cαBC and full-length αBC.*

The diffusion coefficients of DTT and water were measured as references to give a sense of scale for the protein measurements. In addition, I checked that the diffusion coefficient of water did not vary before and after aggregation in the presence of significant concentrations of proteins (200 $\mu$M each of $\alpha$Lac and c$\alpha$BC) to rule out any significant change in viscosity and convection upon formation of large particles.

The diffusion coefficient of $\alpha$Lac visible in the NMR spectrum decreased from $1.33\times10^{-6}$ before aggregation to $1.01\times10^{-6}$ cm$^2$s$^{-1}$ after aggregation, i.e. by about 24%. It was much closer to the coefficient of folded c$\alpha$BC, which exists in an equilibrium between 10 kDa monomer and 20 kDa dimer under these conditions, than to the coefficient of full-length oligomeric $\alpha$BC which has an average size of about 600 kDa. Using the Stokes-Einstein equation ($D = \frac{k_B T}{6\pi\eta R_H}$, where $k_B$ is the Boltzmann constant, $T$ is temperature, $\eta$ is solvent viscosity and $R_H$ is hydrodynamic radius), the decrease in $D$ of 24% corresponds to a 1.3-fold increase in hydrodynamic radius. Wilkins et al. [1999] describe empirical relationships for estimating $R_H$ for folded and unfolded proteins of given size. For $\alpha$Lac, the fold increase in $R_H$ upon complete unfolding is predicted to be about 1.8. Our data is therefore consistent with observation of partially unfolded monomers. Of note, the expected fold increase in $R_H$ for formation of a folded dimer is 1.2 and that for unfolded dimer is 2.7. Reduction of disulphides means that folded dimers cannot form. A small contribution to the signal from misfolded oligomers such as dimers or trimers is possible, but is unlikely to be significant since in that case a greater decrease in

the observed diffusion coefficient would be expected. In what follows, therefore, I interpreted the 1D watergate experiments as reporting on $\alpha$Lac monomers.

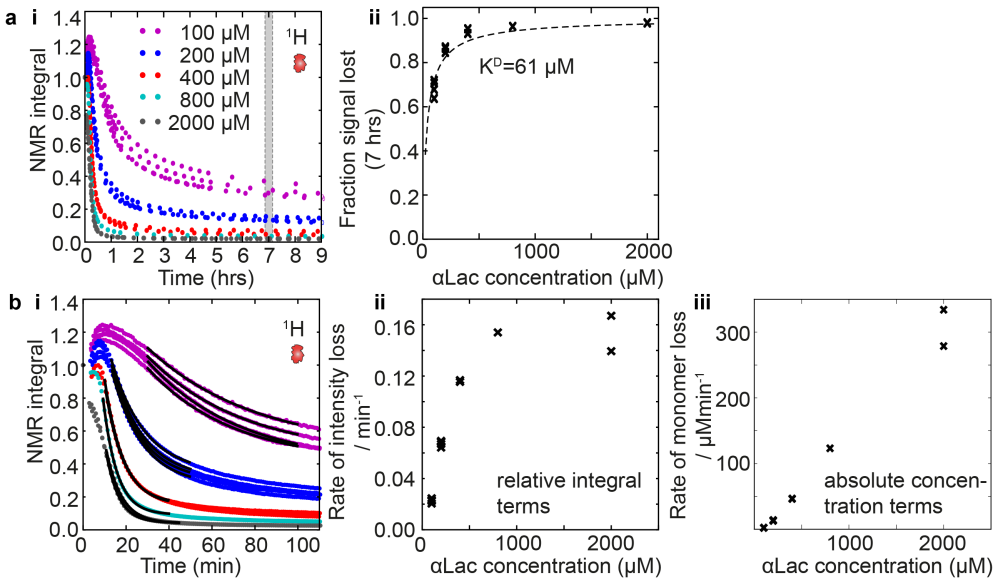



***Figure 4.8: The $\alpha$Lac aggregation dataset with basic analysis. a i*** *Traces for five concentrations with repeats, recorded for at least 9 hours. By 7 hours the traces have mostly reached the plateau phase. The grey strip indicates the region taken for detemination of the equilibrium fraction of signal lost, plotted versus $\alpha$Lac concentration in **ii**. **b** Extended view of the same traces between 0-100 min to illustrate the differences in initial decay of signal. Black lines indicate exponential fits in manually chosen regions. The rate constants for these fits are plotted versus $\alpha$Lac concentration in (**ii**). Panel (**iii**) shows the same rates without the normalisation by initial concentration, to give the approximate initial rate of monomer loss in $\mu Mmin^{-1}$.*

**Figure 4.8** shows the NMR aggregation assay at five concentrations of $\alpha$Lac. The data was much smoother and more reproducible than light scattering traces (**Figure 4.2**). As the data was normalised to the integral before the addition of DTT, there was less apparent variation between repeats in the high-concentration traces. Some concentration-dependent patterns were clear in the data: the decay in signal was more rapid at higher $\alpha$Lac concentrations, masking the initial increase in signal above 400 $\mu$M $\alpha$Lac, and the steady-state proportion of signal remaining was also concentration-dependent. I determined the steady-state signal levels at 7 hours (**Figure 4.8a i**) and the initial rates of decay, fitted to an exponential in manually chosen regions (**Figure 4.8b i**). The hyperbolic shape of the concentration dependence of residual monomer (**Figure 4.8a ii**) is indicative of the steps in $\alpha$Lac

aggregation being reversible (else all signal would be expected to be lost at long times). Treating the residual NMR signal as the fraction of monomeric $\alpha$Lac as discussed above, and the fraction signal lost as the summed concentration of all other species, the plot in **Figure 4.8a ii** can be fit using Oosawa's linear polymerisation model, in which all polymerisation steps occur with the same forward and reverse rates (Oosawa and Kasai [1962]). The model fit the data well, yielding a $K_D$ of $61 \pm 5\,\mu$M.

The relative initial decay rates also increase with concentration and eventually level off (**Figure 4.8b ii**). However, normalisation to the zero point masks the true dependence of monomer loss on protein concentration in this plot. **Figure 4.8b iii** shows the unnormalised initial monomer loss rates in real concentration terms (the rates in (**ii**) multiplied by total protein concentration). From this it is clear that the initial rates increase with total concentration throughout. This means that the rate-limiting step is not zeroth order - it could be an association or a unimolecular conversion. The linear polymerisation model (Oosawa and Kasai [1962]) did not fit the initial rates well (shown in **Figure 4.9a ii**). This suggests that linear polymerisation is an oversimplification for this system, and that the information content in the full traces would allow discrimination between more sophisticated models.

In summary, aggregation of $\alpha$Lac can be followed by NMR in real time. The data is easily quantifiable by the overall intensity integral of the spectrum; it is smooth and reproducible. Importantly, there is only a small decrease in the diffusion coefficient of visible $\alpha$Lac signal that can be explained by misfolding of the monomer. Therefore the contribution of large species to the spectrum is minor and in the simplest approximation the experiment can be taken to report on the residual monomeric $\alpha$Lac. Such clear definition of the observable paves the way for comparison with mechanistic models of aggregation. Although some simple analysis can be performed by looking at patterns such as the steady state signal levels, a more sophisticated approach is clearly needed to fully exploit the information in the shape of the NMR aggregation traces.

### 4.3.4 Chemical kinetics modelling of the NMR data

#### Overview

I sought to develop a strategy that interprets the complete 1D NMR dataset by comparison to simulations of different chemical kinetics models for aggregation. As discussed above, the data is taken to report on changes in the total concentration of misfolded monomer. We implemented an approach in which the changes in concentrations of each species (monomer, dimer, trimer and so on) were explicitly followed. The interconversion between them was parameterised by rate constants. An aggregation model was thus a set of reactions describing possible interconversions and the rate constants that govern them. **Figures 4.9a i** and **b i** illustrate such reaction schematics for two example models: the simplest model (Oosawa's linear polymerisation, Oosawa and Kasai [1962]) and the two-pathway nucleated model. For each model, the expected signal for the NMR experiment could be calculated and compared to the data. The rate constants were optimised numerically to give the best fit

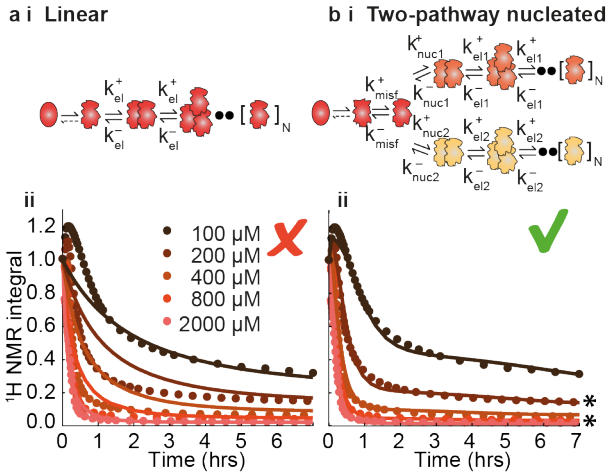



***Figure 4.9: Fitting models to the NMR data.*** ***a-b i*** *Schematics of the linear polymerisation model (**a**) and the two-pathway nucleated model (**b**).* ***a-b ii*** *Fits of the models above to the NMR aggregation traces. The linear model fits poorly and can therefore be discarded. The two-pathway nucleated model fits well. Asterisks indicate data that was excluded from the model selection process and was fit only by the final model as a validation step (200 μM and 800 μM). The ability of the model to fit these excluded data is additional evidence that the it is a good representation of the aggregation process.*

to the aggregation data for a given model. Simple models such as linear polymerisation produced poor fits (**Figure 4.9a ii**). Models of increasing complexity were built, and the improvement in fit was tested for statistical significance using the F-test (see **Section 1.5.2**), taking into account the increase in the number of fitting parameters. A total of 17 models were tested. The simplest model that reproduced the features of the NMR data was the two-pathway nucleated model, which was chosen as final in this analysis (**Figure 4.9b ii**).

Below I describe how an aggregation mechanism can be cast into a set of equations and fit to the NMR data. I consider key variables controlling optimisation of the rate constants, and present the model selection strategy in more detail.

### Setup of an aggregation model

**The observable** The diffusion experiments described above justify the definition of the NMR signal as the fraction of monomeric species, relative to the total protein concentration of the solution prior to addition of DTT:

$$S_{data} = S_{sim} = \frac{[A]_{mon}}{[A]_{total}} \tag{4.1}$$

In the implementation of any aggregation model, the predicted value of $S$ was evaluated at every timepoint and compared against the experimental data to calculate the least square $\chi^2$ value for that model. The least square $\chi^2$ was defined as follows, where the sum is over all datapoints:

$$\chi^2_{LS} = \sum (S_{data} - S_{sim})^2 \tag{4.2}$$

**Example rate equations** Let us consider the simplest model, linear polymerisation. It postulates that aggregation proceeds by monomer addition with forward and backward rates $k^+_{el}$ and $k^-_{el}$ which are the same for all steps. Thus it proceeds

as follows:

$$2A \underset{k_{el}^-}{\overset{k_{el}^+}{\rightleftharpoons}} A_2 \tag{4.3}$$

$$A + A_2 \underset{k_{el}^-}{\overset{k_{el}^+}{\rightleftharpoons}} A_3 \tag{4.4}$$

$$A + A_3 \underset{k_{el}^-}{\overset{k_{el}^+}{\rightleftharpoons}} A_4 \tag{4.5}$$

$$\ldots \underset{k_{el}^-}{\overset{k_{el}^+}{\rightleftharpoons}} \ldots \tag{4.6}$$

Each reaction can be considered in turn and the changes in concentration of each species involved can be represented as differential equations. The following rate equations were used for simulating this linear model:

$$\left.\begin{aligned} d[A] &= -k_{el}^+[A]^2 dt \\ d[A_2] &= +k_{el}^+[A]^2 dt \end{aligned}\right\} \quad \text{equation 4.3 forward}$$

$$\left.\begin{aligned} d[A] &= +k_{el}^-[A_2] dt \\ d[A_2] &= -k_{el}^-[A_2] dt \end{aligned}\right\} \quad \text{equation 4.3 reverse}$$

$$\left.\begin{aligned} d[A] &= -k_{el}^+[A][A_2] dt \\ d[A_2] &= -k_{el}^+[A][A_2] dt \\ d[A_3] &= +k_{el}^+[A][A_2] dt \end{aligned}\right\} \quad \text{equation 4.4 forward}$$

$$\left.\begin{aligned} d[A] &= +k_{el}^-[A_3] dt \\ d[A_2] &= +k_{el}^-[A_3] dt \\ d[A_3] &= -k_{el}^-[A_3] dt \end{aligned}\right\} \quad \text{equation 4.4 reverse}$$

$$\left.\begin{aligned} \ldots &= \ldots \\ \ldots &= \ldots \end{aligned}\right\} \quad \text{equation 4.5 forward}$$

Here $[A]$ is the concentration of $\alpha$Lac monomer, $[A_2]$ is the concentration of the dimer and so on, and $dt$ is the time step of the simulation. Similar equations were set up for species up to the maximum number considered (at least 2,000).

In later models, a conversion step to an aggregation-competent monomer, was

explored, with its associated rate equations. If $A^*$ is the first monomer form and $A$ is the second, aggregation-competent form, the conversion can be written as:

$$A^* \underset{k^-_{misf}}{\overset{k^+_{misf}}{\rightleftharpoons}} A \tag{4.7}$$

$$\left.\begin{aligned} d[A^*] &= -k^+_{misf}[A^*]dt \\ d[A] &= +k^+_{misf}[A^*]dt \end{aligned}\right\} \quad \text{equation 4.7 forward}$$

$$\left.\begin{aligned} d[A^*] &= +k^-_{misf} \times [A] \times dt \\ d[A] &= -k^-_{misf} \times [A] \times dt \end{aligned}\right\} \quad \text{equation 4.7 reverse}$$

In the models were the misfolding step was included, the observable was defined as follows:

$$S_{sim} = \frac{[A^*] + \phi_1[A]}{[A]_{total}} \tag{4.8}$$

with $\phi_1$ included as an additional fitting parameter. The subsequent oligomers were composed of $A$ and not $A^*$.

Using the rate equations, the state of the system at time $t+dt$ can be calculated from its state at time $t$. The starting conditions, i.e. the state at $t = 0$, are known ([A] equal to total protein concentration). Thus, the timecourse of the concentration changes in each species can be evaluated, the signal $S$ can be calculated at any time using the definition above and compared to the experimental data.

**Range of allowed stoichiometries** I explicitly calculated concentrations of species up to 4,000-mers in single-pathway models, up to 2,000-mers in two-pathway models and up to 3,000-mers in the three-pathway model. Plots of mass concentration versus oligomer number were checked to make sure that this stoichiometry space is sufficient and that there are no edge effects such as a buildup of species of maximum stoichiometry.

**Parameter optimisation for each model** Numerical optimisation was used to find the best values of rate constants for a given kinetic model. The cost function was $\chi^2_{LS}$ between the data points ($S_{data}$) and the simulated signal $S_{sim}$ from the model (**Equation 4.2**). The protocol used for optimising the parameters was the Levenberg-Marquardt algorithm (LMA) combined with a basin hopping routine. The principles of these methods are described in **Section 1.5.2**. Here I consider the choice of some key parameters for the optimisation process.

*Starting conditions* The parameter landscape in such a multidimensional problem is likely to be rugged. Incorporating the basin hopping steps helps avoid the problem of being confined to local minima. However, especially with increasing number of parameters, the choice of starting conditions is very important to guide the algorithm to a relevant region of the parameter space. The starting conditions for the simplest model (values of $k^+_{el}$, $k^-_{el}$ and $\phi_1$) were chosen by hand to give simulated curves that roughly resembled the data. For subsequent models, the output parameters from the simpler models were used as starting conditions in the more complex models.

*Levenberg-Marquardt algorithm (LMA) convergence criteria* Important parameters in LMA are the criteria for completion of an optimisation run. To judge convergence, values of $\chi^2$ in successive iterations were compared. When the difference was less than a defined threshold, the run was considered to have converged to a minimum. This value ($1\times10^{-4}$) was chosen so that the variation between successive runs was much smaller than the variation between experimental repeats. Typically, fewer than 10 iterations were required to achieve the convergence threshold in cases where this was possible. Secondly, if the model is incorrect and the parameter landscape is very shallow, it may be very difficult or impossible to reach the convergence threshold. For such cases, the algorithm was set up to perform a maximum of 20 iterations.

*Basin hopping step size* The extent of random variation of each parameter

between successive LMA runs controls the sampling scope of the parameter space. Since we were interested in finding a model in which the parameters are of the correct order of magnitude, rather than precisely defined values, we needed to perform broad searches of the parameter space. Therefore scaled lognormal parameter sampling was chosen over linear Gaussian sampling. Here 'heat' is the standard deviation of the log-space Gaussian centred on the current parameter value.

I tested heat values between 0.1-5.0 on the same model and starting conditions (**Figure 4.10**). I found that low and high values led to the algorithm performing many iterations. At low heat, the step was not large enough to come out of the initial parameter well and many short calculations were performed dropping to the same minimum. At high values of heat, time was spent testing many unrealistic combinations of very high or very low trial rate constants. At intermediate values of heat the number of iterations was lower. Runs with values up to 3.0 resulted in a similar minimum $\chi^2$ whereas the run with heat equal to 5.0 did not. For heat values 0.1-3.0, the logarithmic standard deviation between the output parameters was between 0.08-1.41, i.e. the optimised parameters were mostly of the same order of magnitude and the algorithm could be said to converge to the same minimum. I therefore selected the value of heat of 1.0, which achieved a low $\chi^2$ in the minimum number of iterations.

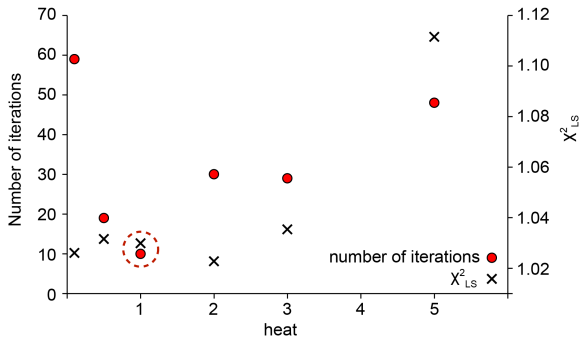



***Figure 4.10: Selection of the optimal step scale (heat) for basin hopping.*** *'Heat' is defined as the standard deviation of the log-space Gaussian centered on the current parameter value. The value that could achieve a low $\chi^2$ in the smallest number of iterations (1.0) was chosen (red dashed circle).*

*Basin hopping convergence criteria* The basin hopping protocol was set to stop after performing 20 successive iterations without any improvement in $\chi^2$. There was no limit on the number of simulated annealing steps if improvement in $\chi^2$ was achieved.

**Validation by excluding a subset of the data** Data were recorded for five concentrations of $\alpha$Lac. However, model testing was performed on only three out of the five datasets (the lowest, middle and highest concentrations, $100, 400$ and $2000\,\mu$M). The best model was fit to the remaining two curves (200 and $800\,\mu$M). This provides an additional test of the appropriateness of the model: if it indeed describes the aggregation process accurately, it should be able to fit the excluded data with similar parameters to the data it was selected on.

### The model selection strategy

Two principles were at the heart of the model selection process. Firstly, the appropriate model should explain the features of the data, such as the initial increase in spectral intensity at low concentrations of $\alpha$Lac, the rates of signal decay and equilibrium levels of signal, for all curves. Secondly, the model should contain the lowest number of parameters needed to achieve this, i.e. it should be as simple as possible. I therefore started with very simple models and gradually increased the complexity until a good fit could be found. The list of all models in the order attempted is presented in **Figure 4.11**. The strategy is illustrated graphically in **Figure 4.12**, where the lowest $\chi^2_{LS}$ following LMA and basin hopping optimisation is plotted in **Figure 4.12**, together with the number of rate constants, a measure of model complexity. As model complexity increased, the fits improved. Models with more adjustable parameters will, however, inherently provide better fits, leading to overfitting. To judge whether the improvement in fit was significant enough to justify the increase in complexity, the F-test was performed between appropriate pairs of models (orange bars in **Figures 4.11** and **4.12**). Where the number of parameters was identical, a lower $\chi^2$ was a sufficient condition for the model to be taken forward.

Initially, linear and nucleated polymerization models were tested (models 1, 2). These were both in poor agreement with the data, as neither could properly

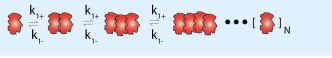
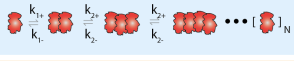
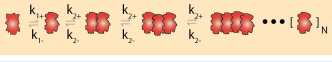
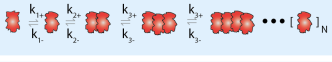
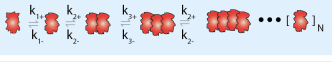
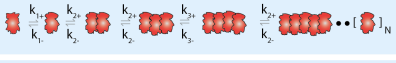
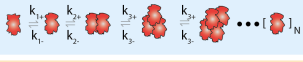
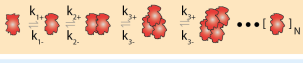
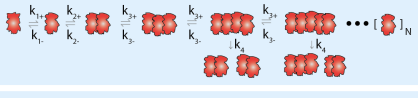
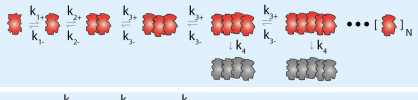
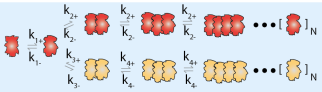
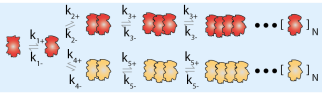
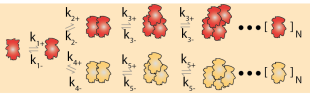
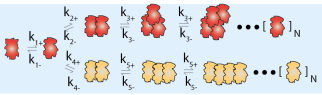
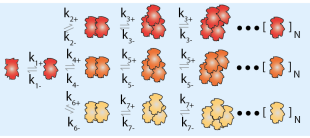
| | Model | | Rate constants | Scaling of $k_+$ | Scaling of $k_-$ | $\chi^2_{LS}$ | F-test p-level |
|---|---|---|---|---|---|---|---|
| 1 | linear |  | 2 | 1 | 1 | 1.371 | |
| 2 | nucleated linear |  | 4 | 1 | 1 | 1.241 | |
| 3 | linear, initial misfolding |  | 4 | 1 | 1 | 0.3211 | vs model 1: 4.1E-17 PASS |
| 4 | nucleated linear, initial misfolding, nucleus=dimer |  | 6 | 1 | 1 | 0.1828 | |
| 5 | nucleated linear, initial misfolding, nucleus=trimer |  | 6 | 1 | 1 | 0.2911 | |
| 6 | nucleated linear, initial misfolding, nucleus=tetramer |  | 6 | 1 | 1 | 0.3718 | |
| 7 | nucleated linear, initial misfolding, nucleus=10mer |  | 6 | 1 | 1 | 0.4717 | |
| 8 | nucleated linear, initial misfolding, nucleus=20mer |  | 6 | 1 | 1 | 0.8267 | |
| 9 | nucleated Poisson-type, initial misfolding, nucleus=dimer |  | 6 | 1 | $i^{2/3}$ | 0.1841 | |
| 10 | nucleated spherical, initial misfolding, nucleus=dimer |  | 6 | $i^{2/3}$ | $i^{2/3}$ | 0.1721 | vs model 3: 1.7E-07 PASS |
| 11 | nucleated linear, initial misfolding, nucleus=dimer fragmentation |  | 7 | 1 | 1 | 0.1980 | |
| 12 | nucleated linear, initial misfolding, nucleus=dimer irreversible side aggregation |  | 7 | 1 | 1 | 0.1777 | |
| 13 | two-pathway (linear + nucleated linear) initial misfolding |  | 8 | 1<br>1 | 1<br>1 | 0.1660 | |
| 14 | two-pathway (2 x nucleated linear) initial misfolding |  | 10 | 1<br>1 | 1<br>1 | 0.1384 | |
| 15 | two-pathway (2 x nucleated spherical) initial misfolding |  | 10 | $i^{2/3}$<br>$i^{2/3}$ | $i^{2/3}$<br>$i^{2/3}$ | 0.1343 | vs model 10: 0.020 PASS |
| 16 | two-pathway (nucleated spherical+nucleated linear) initial misfolding |  | 10 | $i^{2/3}$<br>1 | $i^{2/3}$<br>1 | 0.1555 | |
| 17 | three-pathway (3 x nucleated spherical) initial misfolding |  | 14 | $i^{2/3}$<br>$i^{2/3}$<br>$i^{2/3}$ | $i^{2/3}$<br>$i^{2/3}$<br>$i^{2/3}$ | 0.1342 | vs model 15: 1.0 FAIL |

***Figure 4.11: List of models tested for $\alpha$Lac aggregation data in the absence of chaperone.*** *Models that resulted in key improvements are highlighted in orange. The number of parameters in each model was the number of rate constants plus one for the intensity ratio of the misfolded monomer ($\phi_1$). The scaling columns refer to adjustment of the elongation rates depending on aggregate geometry: no scaling, i.e. 1, for fibrils and a scaling of $i^{2/3}$ for spherical aggregates. The scaling for the Poisson model (9) is described in Baldwin et al. [2011b]. $\chi^2_{LS}$ is defined in* ***Equation 4.2****. The threshold p-value to pass the F-test was 0.05.*

account for the increase in signal intensity observed at early times at the lower concentrations. The inclusion of an initial monomer misfolding step (models 3, 4) resulted in a substantial decrease in $\chi^2_{LS}$, as this step does produce an increase in signal at early times due to appearance of a species with a higher NMR signal intensity on the timescale of minutes. The nucleated model (model 4) was found to better reproduce the concentration dependence in the kinetic data compared to the non-nucleated one. To test the identity of the nucleus, a series of models was trialled, varying the oligomeric size of the nucleus from 2 to 20 (models 4-8). The model with a dimeric nucleus was statistically justified. Next, I tested how the oligomer growth rates scaled with number of $\alpha$Lac molecules in the aggregate, which is determined by aggregate geometry (models 4, 9-10). The difference in $\chi^2_{LS}$ favoured spherical growth (rate scaling with $i^{2/3}$ for an aggregate of stoichiometry $i$), although the magnitude of the difference with a linear growth model was small, suggesting that linear and spherical growth cannot be clearly distinguished here statistically. I also investigated including fragmentation of oligomers (model 11) and off-pathway irreversible aggregation (model 12), neither of which improved the fit.

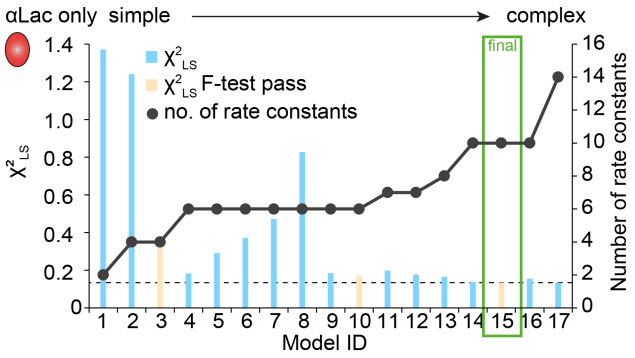



***Figure 4.12: The model selection process for $\alpha$Lac.*** *Model numbers correspond to those in* ***Figure 4.11****. $\chi^2_{LS}$ and the number of rate constants is shown for each model. Orange bars correspond to models that pass the F-test relative to a simpler model and therefore incorporate an essential molecular process. The final model is the two-pathway nucleated model, with the fit shown in* ***Figure 4.9****.*

All models up until here failed to account for what appeared to be a second, slow kinetic decrease in the data, and so a second aggregation pathway was

introduced. Various permutations of nucleated, linear and spherical pathway combinations were tested (models 13-16). Overall, the model containing two spherical nucleated pathways (model 15) was found to the optimal model. An even more complex model with three parallel pathways was tested (model 17), but the reduction in $\chi^2_{LS}$ was not statistically significant given the greater number of fitting parameters, according to an F-test. In summary, an initial monomer misfolding step, a nucleation step and inclusion of two parallel aggregation pathways provided notable improvements in the quality of the fit (**Figures 4.9** and **4.13**). The rate equations describing the final model are listed in **Section A.1** in **Appendix A**.

In sum, I tested a total of 17 models of $\alpha$Lac aggregation, optimising the free parameters for each, to determine the simplest best-fit mechanism and the values of the corresponding parameters. I identified a model characterised by 10 rates that best explains our data in the entire concentration range. Further increase in model complexity did not afford a statistically justified improvement.

### 4.3.5 An aggregation mechanism for $\alpha$Lac

Our final model for the aggregation of $\alpha$Lac includes the formation of an on-pathway aggregation-prone monomeric species, and two parallel nucleated pathways (**Figure 4.13a**). Each of these features can be justified in terms of their visible effects on the data. Without the monomer conversion to an aggregation-prone form, we could not accommodate the increase in signal at early times; without nucleation we could not properly account for the concentration dependence of the early decrease in signal intensity; and a single pathway was not sufficient for describing the bi-phasic decay observed at longer times (**Figure 4.13a ii**). Thus the model allows us to explain, on a molecular level, the features that are observable in the data.

**Figure 4.13b** shows parameters from fits of each concentration individually. The values of each parameter are mostly of the same order of magnitude across concentrations. Two of those were not used in the model selection process (200 $\mu$M and

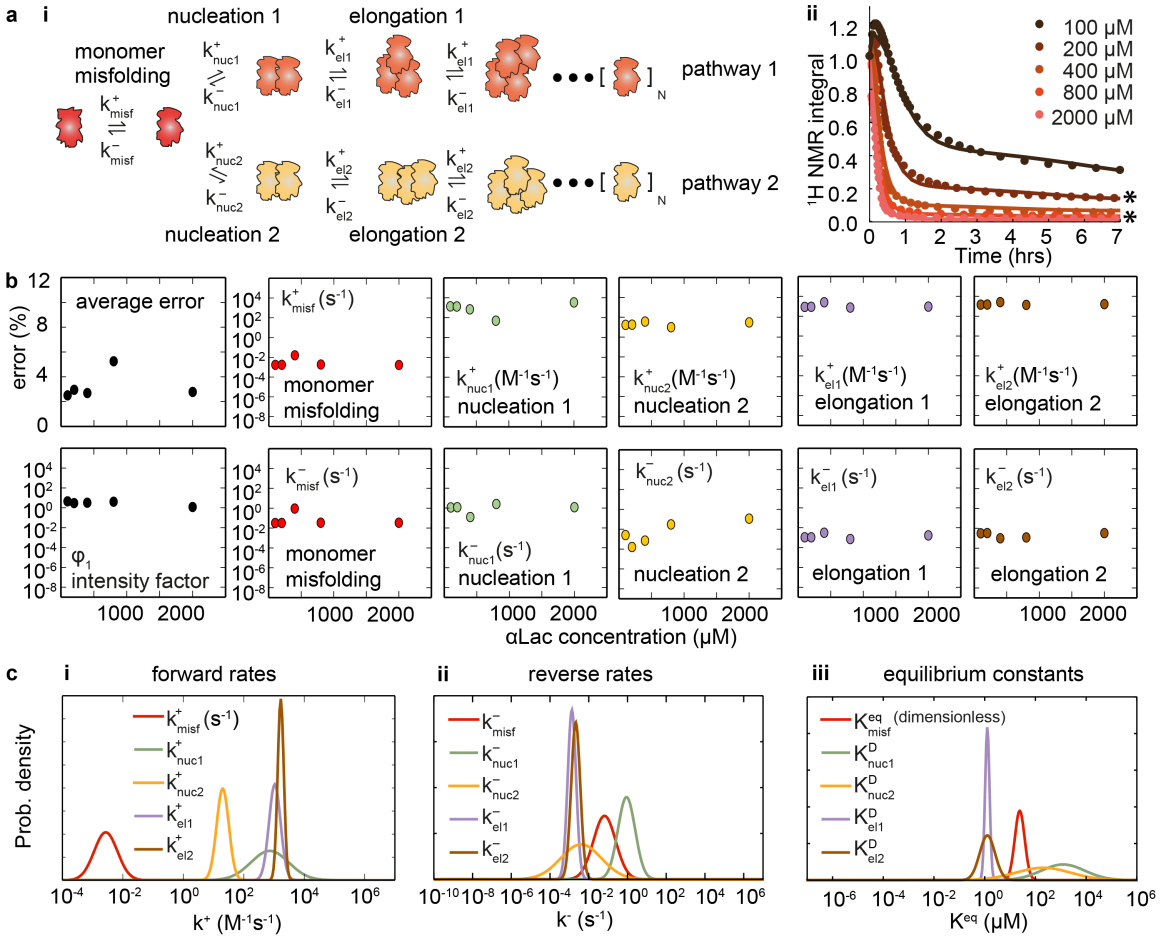



***Figure 4.13: Detailed overview of the final model and its parameters for αLac.*** ***a i*** *Annotated model schematic.* ***ii*** *Model fit to the data, reproduced from* ***Figure 4.9b*** *for convenience. Asterisks indicate the curves that were not included in model selection and whose fits therefore constitute validation of the model.* ***b*** *The average fitting error and values of parameters obtained in individual fits of each of the curves in* ***a ii.*** ***c*** *Distribution representations of the forward (**i**) and reverse rate constants (**ii**) and dissociation constants (**iii**) for each process. The mean and standard deviation for each Gaussian correspond to the distribution of the five fits in* ***b****.*

800 $\mu$M $\alpha$Lac). The observation that they are also fit well increases confidence in the results, since a good model should be able to explain data across the concentration range with the same parameters.

The need for an additional misfolding step links to the presence of several disulphide bonds in the protein. Based on iodoacetamide trapping and MS analysis, Carver et al. [2002] report that at 23°C the first two disulphide bonds are nearly fully reduced after 5 min but the remaining two bonds are broken much more slowly. At 37°C this process is likely to occur faster. Thus the conversion step we are

observing on the timescale of minutes could correspond to breakage of one of the last disulphide bonds. This could be confirmed with matching kinetics of reduction in an equivalent iodoacetamide trapping experiment under the conditions used here.

**Figure 4.13c** shows that the steps with the highest $K_D$ values are the two nucleation events. It is interesting to consider the ways in which nucleation 1 and nucleation 2 are rate-limiting. For nucleation 1 (green curves), the rate of nucleus formation is relatively high, similar to the elongation rate constants, but the rate of nucleus breakdown ($k^-_{nuc1}$) is also very high. So the nucleus can form fairly easily but then quickly decays back to monomer. In contrast, for nucleation 2 (yellow), $k^-_{nuc2}$ is lower, but the rate of nucleus formation, $k^+_{nuc2}$, is significantly slower than the other nucleation and elongation rates. Therefore for nucleation 2, formation of the nucleus in the first place is difficult. The rates of elongation for the two pathways are rather similar. Therefore the key distinction between the pathways is on the level of the nuclei.

With the knowledge of the underlying mechanism and the rate constants governing aggregation, we can look at the aggregation process as a function of time and total concentration in a comprehensive way. **Figure 4.14a** is a kinetic phase diagram showing the changes in the concentration of monomer $\alpha$Lac. A threshold contour at which aggregation sets in can be defined in time and concentration space (black line in **Figure 4.14a**, corresponding to 20% loss of monomer). At concentrations of $\alpha$Lac below 20-30 $\mu$M little, if any, aggregation is predicted for up to 100 hours. In accord with this prediction, we observed no aggregation over a timescale of hours at 10 $\mu$M $\alpha$Lac (**Figure 4.14c**).

The behaviour of the two predicted pathways can also be compared, shown in **Figure 4.14b**. Looking their kinetic and thermodynamic parameters in **Figure 4.13c**, they have similar elongation $K_D$ values but pathway 2 has a slightly lower nucleation $K_D$. Thus although both pathways operate in similar regions of aggregation space, pathway 2 is dominant. The electron micrographs of $\alpha$Lac shown in **Figure**

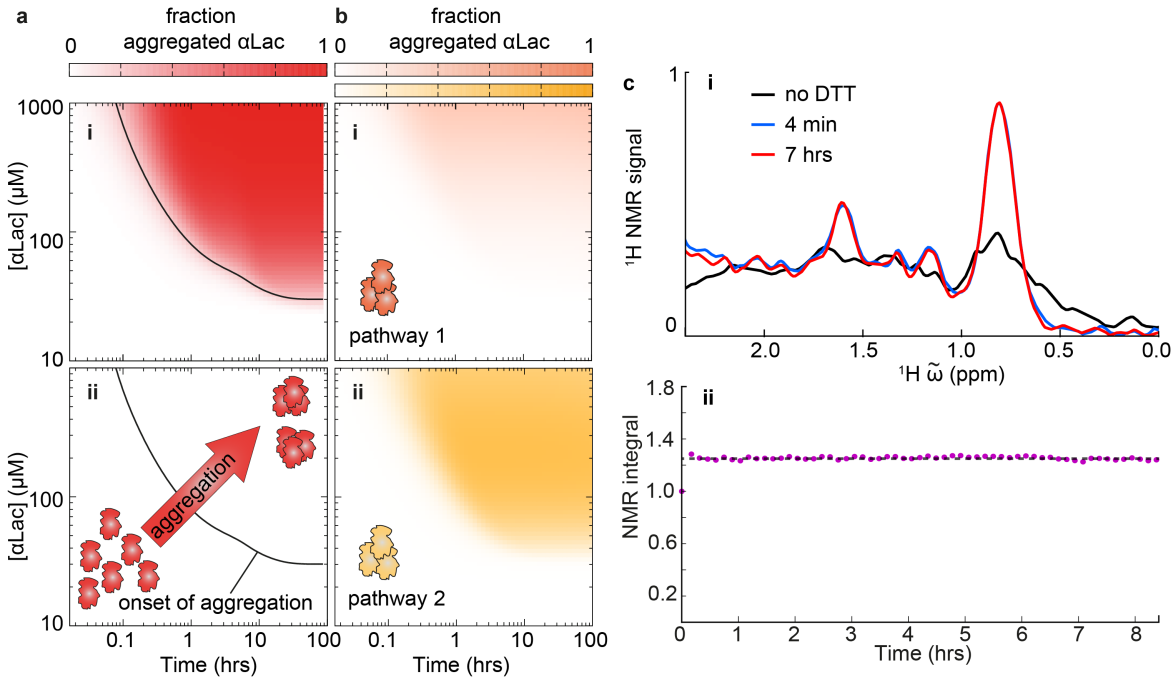



***Figure 4.14: Analysis of αLac aggregation mechanism.*** ***a i*** *A surface in time and total αLac concentration space showing the fraction of αLac that is in non-monomeric form. The black contour line shows the threshold for aggregation (20% monomer loss).* ***ii*** *A cartoon illustration of the meaning of the different regions in* ***i****.* ***b*** *Equivalent surfaces showing the partition of αLac aggregates into pathway 1 and pathway 2, with numbering as in* ***Figure 4.13****. The majority of the protein aggregates via pathway 2.* ***c*** *1D watergate experiments at 10 μM αLac, a concentration at which the surface in a predicts no significant aggregation on the timescale of hours.* ***i*** *A region of the spectrum before, immediately after and 7 hours after the addition of DTT, showing that the conformational change occurs but there is no subsequent loss of intensity.* ***ii*** *The spectral integral time course. A dashed straight line is added at y=1.32 to confirm that there is no signal loss.*

**4.5** and also **Figure 6.4** (larger images) show two morphologies of aggregate, which is consistent with there being two aggregation pathways. The concept of parallel pathways is also consistent with the reports that αLac forms different types of aggregates depending on solution conditions (Goers et al. [2002], Ghahghaei et al. [2010], Kulig and Ecroyd [2012]). Presence of particular salts or the solution pH might make one pathway dominant over others.

By fitting chemical kinetic models to our NMR data, we have therefore been able to obtain an in-depth, quantitative understanding of the aggregation mechanism, identifying and characterising key types of species, and providing the predictive power as a function of time and protein concentrations. The approach rests

on being able to clearly define the observable in the experiment, so that the output from the chemical kinetics simulation can be compared to the data. It also relies on the information content of the NMR spectral intensity curves. While the discrimination potential has its limits (e.g. the $\chi^2_{LS}$ difference between the spherical vs linear geometries of the aggregates, models 14 and 15, is small and does not clearly rule out one or the other), the rich shape of the data did allow us to disprove 16 models before selecting the best-fit mechanism. In addition, the NMR data does not assume amyloid or amorphous aggregation, and does not rely on analytical solutions of mechanisms. Instead, it relies on numerical simulation of any given set of rate equations, so it is limited only by the computational resources available and allows the testing of a broad range of models. The $^1$H NMR experiments are label-free and simple to record. The approach is therefore widely applicable to other aggregating systems.

## 4.4 Conclusions

Studying protein aggregation is a challenge - it is a non-equilibrium process and involves hundreds of species that constantly interconvert. Real-time 1D NMR in combination with chemical kinetics and statistical principles has enabled me to tackle such a system to derive a representation of $\alpha$Lac aggregation in terms of rate equations and determine the associated rate constants for each. I showed that $\alpha$Lac aggregation involves a conformational change of the monomer, rate-limiting nucleation and elongation along at least two parallel pathways. A mechanistic understanding of $\alpha$Lac aggregation paves the way for investigating how it might be slowed down by chaperones, which is the subject of the next Chapter.

# Chapter 5

# Mechanism of $\alpha$BC chaperone action

*In this Chapter I look at the mechanism by which $\alpha$BC inhibits protein aggregation. I extend the NMR and chemical kinetics methodology developed in the previous chapter to mixtures of the aggregating substrate $\alpha$Lac and one of two chaperones: the simpler core domain of $\alpha$BC (c$\alpha$BC) and large and polydisperse full-length $\alpha$BC. After extensive model testing, I derive mechanisms by which $\alpha$BC and its isolated core domain suppress aggregation of $\alpha$Lac. The mechanisms show that the chaperones delay the aggregation of $\alpha$Lac monomer in two ways: by disassembling the dimeric $\alpha$Lac nuclei and by competitive binding of larger aggregates. The full-length chaperone is more effective at rescuing monomeric $\alpha$Lac from aggregation and more selective in the species it binds compared to the isolated core domain.*

## 5.1 Introduction

$\alpha$BC is well-known to act as a molecular chaperone for various aggregating substrates *in vivo* and *in vitro* (Horwitz [1992], Boelens [2014], Treweek et al. [2015]). Various theories have been put forward as to how it might inhibit aggregation, both in terms of the stage of the aggregation process in which it acts and in terms of where the binding sites for the substrates might be located. **Chapters 6** and **7** describe experiments aimed at detecting complexes and discerning the location of substrate binding sites on the chaperone, whereas this Chapter is concerned with the former question: of the variety of species formed by the substrate during aggregation, which ones are targeted by $\alpha$BC and what kinetic and thermodynamic parameters govern these interactions?

The isolated core domain of $\alpha$BC, c$\alpha$BC, has been reported to have potent chaperone activity in light scattering assays, equivalent to that of full-length $\alpha$BC in the cases of $\alpha$Lac and A$\beta_{1-42}$, and to reduce A$\beta_{1-42}$-associated toxicity to cultured cells (Hochberg et al. [2014]). This suggests that the C- and N-termini are dispensable for chaperone function *in vitro*. However, light scattering assays are a rather coarse measure of chaperone function. Therefore I investigate the differences between the core domain and the full-length chaperone in more detail here.

I focus on the action of $\alpha$BC and c$\alpha$BC on the model substrate $\alpha$Lac, whose mechanism of aggregation in the absence of chaperone is the subject of the previous Chapter. Aggregation of $\alpha$Lac is triggered by addition of a reducing agent such as DTT, which does not affect either c$\alpha$BC or $\alpha$BC as these proteins have no cysteines in their sequence. The key elements of the aggregation mechanism of $\alpha$Lac are conformational change of the monomer prior to aggregation, formation of a dimeric nucleus, and growth of the aggregates (see **Figure 4.13**). At least two sets of nuclei and aggregates with different kinetic and thermodynamic properties were required to explain the NMR data describing its aggregation. In principle $\alpha$BC could act on any of these stages. As it does not consume ATP, a likely mode of its action is

kinetic competition with the substrate (Lindner et al. [2001], Jacobs et al. [2016]). In support of this hypothesis, $\alpha$BC is known to be a less effective chaperone in systems that aggregate quickly (e.g. in the presence of aggregation-accelerating salts such as sodium sulfate or with holo rather than apo $\alpha$Lac, Lindner et al. [2001], Carver et al. [2002]).

Is $\alpha$BC likely to bind the monomer, the nucleus or larger aggregates? Carver and colleagues have looked at the NMR spectrum of $\alpha$Lac following addition of a reducing agent. This spectrum corresponds to the molten globule state of $\alpha$Lac that precedes aggregation (Lindner et al. [1997, 2001], Carver et al. [2002]). They demonstrated that $\alpha$BC slows down the rate at which the intensity of this spectrum is lost over time. The authors argue that $\alpha$BC stabilises the molten globule monomer by weak and transient interactions. The group of McHaourab and colleagues have described two-mode binding of a set of T4 lysozyme mutants to $\alpha$A-crystallin and $\alpha$B-crystallin variants (Mchaourab et al. [2002], Koteiche and McHaourab [2003], Sathish et al. [2003]). On the other hand, no interaction of $\alpha$BC was observed between monomeric $\alpha$-synuclein (2D NMR, Waudby et al. [2010]) or A$\beta$ (sedimentation velocity, Raman et al. [2005]).

Hatters et al. [2001] investigated the mechanism of how $\alpha$-crystallin (a mixture of $\alpha$A- and $\alpha$B-crystallin) might inhibit amyloid formation by the human plasma protein apolipoprotein CII. They could not detect any interaction with the unfolded monomer or accumulation of oligomers of intermediate size. They therefore concluded that $\alpha$BC most likely interacts transiently with rate-limiting aggregation nuclei, promoting their dissociation into monomers. They did not consider the interaction of $\alpha$-crystallin with mature aggregates.

However, $\alpha$BC was shown to bind amyloid aggregates of $\alpha$-synuclein (Waudby et al. [2010]), as well as A$\beta$ fibrils with a $K_D$ of about 2 $\mu$M (Raman et al. [2005], Shammas et al. [2011]). In both of these cases $\alpha$BC was found to inhibit fibril elongation. Rekas et al. [2004] suggest that $\alpha$BC might redirect $\alpha$-synuclein from

amyloid to amorphous aggregation, based on negative stain EM imaging. $\alpha$BC has also been observed to be part of large complexes with $\alpha$Lac (Carver et al. [2002], Kulig and Ecroyd [2012], Smirnova et al. [2013]), which suggests that it acts at the stage of mature aggregates.

Thus $\alpha$BC has been proposed to act at all stages of aggregation for different substrates: the soluble monomer, oligomers including aggregation nuclei or larger aggregates, through either transient or more long-lived interactions. The interpretation of the data is challenging because the substrate and the chaperone act in a system of interconnected equilibria, so that chaperone binding to a given species also affects the upstream and downstream steps. This Chapter builds on the approach presented in **Chapter 4** to tackle such complex mixtures and allow integrated analysis of chaperone action within the network of coupled equilibria. I demonstrate that both $\alpha$BC and c$\alpha$BC intervene at several stages of the aggregation process, and that the full-length chaperone is more potent and selective compared to the core domain. The analysis reveals a key previously unreported activity of both chaperones - the ability to break down the dimeric $\alpha$Lac nuclei, thus shifting the equilibrium of the system towards monomers.

## 5.2 Materials and methods

Please see **Section 4.2** for details on materials, aggregation reactions and NMR diffusion experiments, **Section 3.2** for $\alpha$BC purification and **Section 2.2** on c$\alpha$BC purification.

**1D NMR-based aggregation assays** The $^{1}$H experiments were performed as described in Section 4.2. For the heteronucleus-edited experiments, the following parameters were used: $^{15}$N$^{1}$H 1D HSQC: acquisition time 67 ms, number of points 607, 72 transients per FID, recycle delay 1 s. $^{13}$C$^{1}$H 1D constant-time HSQC: number of points 576, acquisition time 64 ms, 64 transients per FID, recycle delay 1 s,

transmitter offset in $^{13}$C dimension 35 ppm. The duration of the experiments was 81 s for $^{15}$N$^{1}$H HSQC and 73 s for $^{13}$C$^{1}$H HSQC. Spectra were analyzed using NMRPipe (Delaglio et al. [1995]) and Sparky (Lee et al. [2015]). The following ranges were taken for integration (no buffer components present): $^{15}$N$^{1}$H spectra: 6.5-9.5 ppm, $^{13}$C$^{1}$H spectra: 0.84-2.33 ppm.

**Chemical kinetics simulations and numerical fitting** This was carried out as described in **Section 4.2**. One dataset consisted of one $^{1}$H timecourse and either a $^{15}$N$^{1}$H (for c$\alpha$BC) or a $^{13}$C$^{1}$H (for $\alpha$BC) timecourse, with both fit together. Each curve was downsampled to 17 points, preserving its overall shape. The $\alpha$Lac aggregation model and parameters were fixed to those derived in **Chapter 4**. The kinetic models were extended by addition of interactions with the chaperone, parameterised by forward and reverse rate constants. The spectral intensity factor of the chaperone relative to misfolded $\alpha$Lac was included as a fitting parameter. Model selection was performed on two datasets (100 $\mu$M and 2000 $\mu$M $\alpha$Lac) for c$\alpha$BC and on three datasets (100 $\mu$M, 400 $\mu$M and 2000 $\mu$M $\alpha$Lac) for $\alpha$BC. The remaining datasets (200 $\mu$M, 400 $\mu$M and 800 $\mu$M $\alpha$Lac for c$\alpha$BC and 200 $\mu$M and 800 $\mu$M $\alpha$Lac for $\alpha$BC) were set aside to be fit only by the final models as a validation step.

## 5.3 Results and discussion

### 5.3.1 Chaperone action followed over time by NMR

Real-time 1D NMR provides a convenient way to follow aggregation - the signal from small species such as monomers decays over time as they become incorporated into large aggregates, which are not visible in the spectra. The shape of the time traces gives insight into the kinetic and thermodynamic relationships between the different species involved, as I explore in **Chapter 4**. Here I take this approach further, using isotopic labelling to follow different proteins in a mixture in a non-invasive

way. For studying $\alpha$Lac in the presence of chaperones, I used isotopically unlabelled $\alpha$Lac and differentially labelled chaperones: c$\alpha$BC was $^{15}$N-labelled and $\alpha$BC was $^{13}$C-labelled. Full-length, oligomeric $\alpha$BC is too large to be fully visible by solution NMR; only the flexible C-termini can be seen (Carver et al. [1992], Ghahghaei et al. [2009], Baldwin et al. [2012]). $^{13}$C labelling allowed me to obtain a higher signal-to-noise ratio in the spectra. For the much smaller c$\alpha$BC, standard $^{15}$N labelling was used. I recorded two types of experiment on the mixtures as interleaved series: $^{1}$H watergate experiments reporting on the sum of signal from $\alpha$Lac and chaperone, and heteronucleus-edited HSQC experiments reporting only on the chaperone within the mixture. **Figure 5.1** illustrates this for control spectra in the absence of DTT.

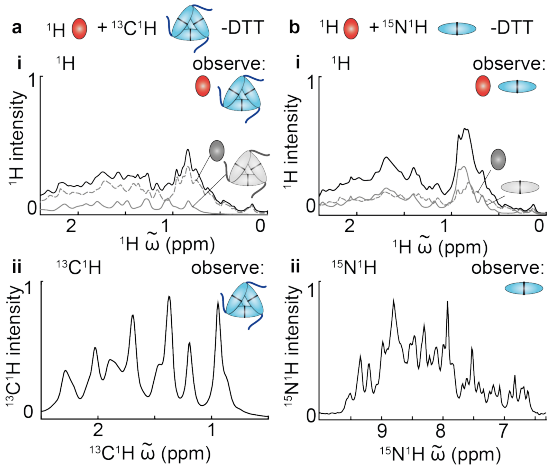



***Figure 5.1: Using isotope labelling to follow different species within a mixture.*** ***a*** *Mixture of $\alpha$Lac with $\alpha$BC,* ***b*** *$\alpha$Lac with c$\alpha$BC, both without DTT. (**i**) $^{1}$H spectra reporting on both species; (**ii**) isotope-edited $^{1}$H spectra reporting selectively on the chaperone. The $^{1}$H spectra are a sum of the spectra of the individual components (gray traces and cartoons), indicating that the proteins do not interact in the absence of aggregation.*

The real-time NMR experiments are illustrated in **Figure 5.2**. In the absence of DTT, the $^{1}$H NMR spectra of these mixtures constituted a sum of the spectra of $\alpha$Lac and the chaperone alone (**Figure 5.1**) and did not vary over time, i.e. no interaction between the proteins could be detected (**Figure 5.2a i** and **d i**, blue panels). After the addition of DTT, the $^{1}$H spectra showed an immediate change in shape followed by a global decay in intensity (**Figure 5.2a ii** and **d ii**). In both sets of heteronucleus-edited experiments, no change in the shape of the spectra was observed upon addition of DTT, but there was a time-dependent intensity decrease (**Figure 5.2b** and **e**). Preservation of the shape of the spectrum suggested we were observing residual unbound chaperone, and that any change in the shape of $^{1}$H spectra was due to a change in conformation of $\alpha$Lac only.

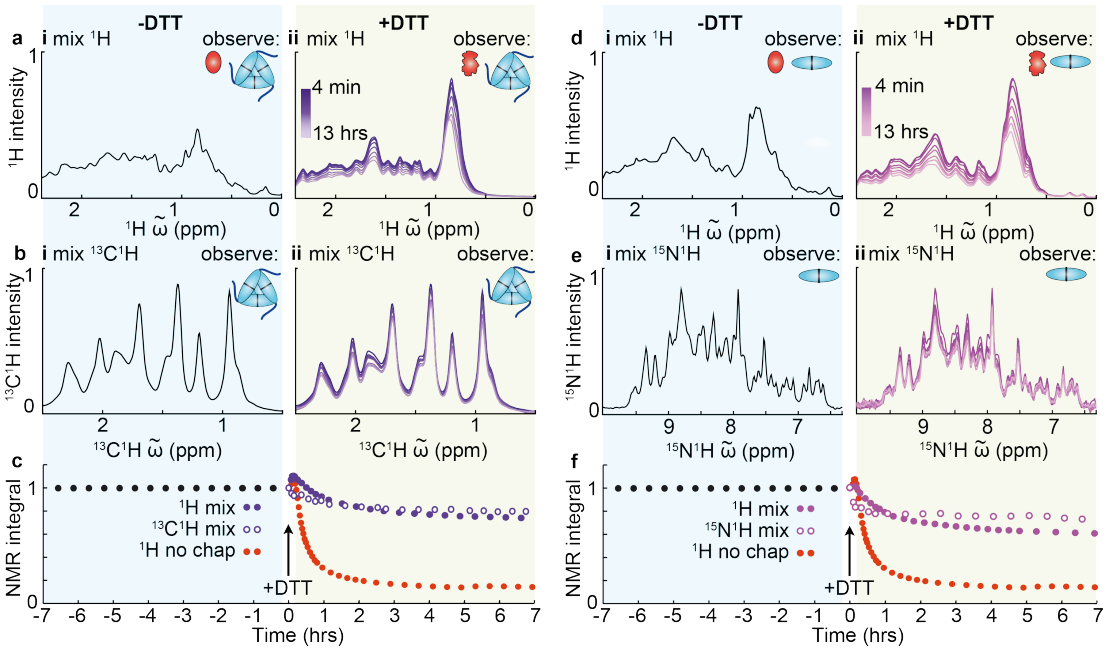



***Figure 5.2:*** *Following chaperone action by NMR.* ***a-c*** *αBC,* ***d-f*** *cαBC. Data obtained before addition of DTT is always shown in panels (**i**) and data after addition of DTT is shown in panels (**ii**).* ***a, c*** *$^{1}H$ spectra reporting on the sum of the signals from αLac and the chaperone.* ***b, d*** *$^{15}N^{1}H$ or $^{13}C^{1}H$ spectra selectively reporting on the chaperone in the mixture.* ***c, f*** *Quantification of the area under the respective spectra above, shown as time traces. The equivalent curve for αLac in the absence of chaperone is also included for comparison (see **Figure 4.6**). All proteins were at 200 μM.*

I quantified the data by integrating the spectra and plotting them over time relative to the respective spectra before addition of DTT (**Figure 5.2c** and **f**). The traces from the $^{1}H$ and heteronucleus-edited experiments in both **Figure 5.2c** and **f** do not overlay, i.e. each type of experiment carries distinct information about the kinetics and thermodynamics of the reaction. There is clearly much less overall loss of signal compared to the reaction in the absence of chaperone (red in **Figure 5.2 c** and **f**).

I repeated the experiments at constant chaperone concentration of 200 μM and five different αLac concentrations between 100-2000 μM (**Figure 5.3**). These can be compared with the set of $^{1}H$ traces for αLac in the absence of chaperone (**Figure 4.8** in **Chapter 4**). From the shape of the traces in **Figure 5.3** it can be seen that the behaviour of full-length αBC is distinct from that of cαBC (compare panels **a ii** and **b ii**, or **d ii** and **e ii**).

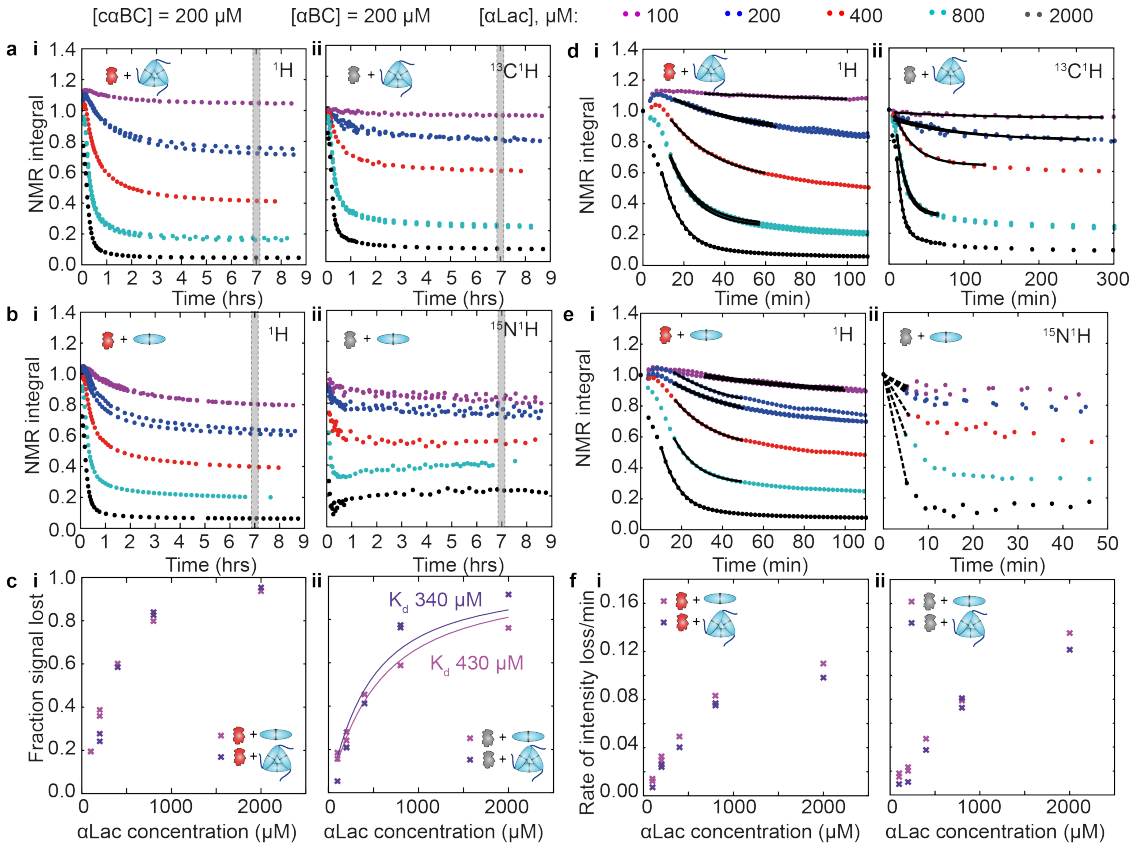



***Figure 5.3: The full set of NMR time traces for mixtures of αLac with chaperones.*** *Panels (**i**) always show $^{1}H$ data and panels (**ii**) show $^{15}N^{1}H$ or/and $^{13}C^{1}H$ data.* ***a*** *Mixtures with αBC.* ***b*** *Mixtures with cαBC. The grey strips indicate the region taken for determination of equilibrium level of signal, plotted versus αLac concentration in panel* ***c****.* ***d-e*** *Extended view of the first 50-300 minutes of aggregation to illustrate exponential fitting of the initial rate of signal loss in manually defined regions, for αBC and cαBC respectively. In the case of the $^{15}N^{1}H$ experiment for cαBC, the decay was mostly within the dead time of the experiment and the rate was therefore linearly approximated between time 0 and the first measured timepoint (panel* ***e ii****).* ***f*** *Plots of the fitted initial signal decay rate versus αLac concentration.*

**Figures 5.3c** and **f** present some basic analysis of the shape of the traces. The equilibrium levels of signal, taken at 7 hours, are plotted versus αLac concentration in panel **c**, with increasing fractions of signal lost at higher αLac concentration. The hyperbolic shape of the relationship is indicative of the system behaving as a set of reversible, concentration-dependent equilibria. If aggregation and/or chaperone binding was irreversible, all traces would be expected to decay to zero. In the heteronuclear experiments, a simplified model can be applied to extract an apparent dissociation constant for the chaperone-substrate interaction. Assuming that the reaction is $A + B \rightleftharpoons AB$, where $A$ is αLac and $B$ is chaperone, and that the fraction

of signal lost corresponds to the bound chaperone, the fit to the data gave a $K_D$ of 430 $\mu$M for c$\alpha$BC and 340 $\mu$M for $\alpha$BC. The theoretical curve fits the data for $\alpha$BC quite poorly, suggesting that this model is not a good approximation for the underlying set of reactions. On the other hand, the curve for c$\alpha$BC fits the data well, suggesting that there is perhaps one dominant interaction in the system. **Figure 5.3f** shows the variation of initial rates of signal loss with $\alpha$Lac concentration. The rates increase with concentration, as would be expected if the rate-limiting step is not zeroth order in concentration of both $\alpha$Lac and chaperone.

In order to understand more about the species that each experiment reported on, I recorded $^1$H and heteronucleus-edited $^1$H diffusion experiments. The results are shown in **Figure 5.4**.

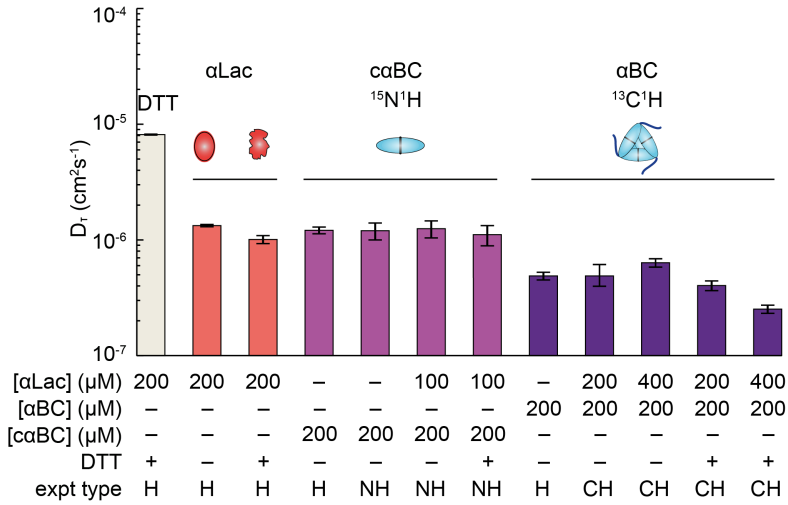



***Figure 5.4: Diffusion data chaperones in mixtures with $\alpha$Lac.*** *Diffusion coefficients for DTT and $\alpha$Lac with and without DTT are included for comparison.*

For c$\alpha$BC, no significant difference in diffusion was observed regardless of the presence of $\alpha$Lac or DTT. This is consistent with the absence of significant changes in the shape of the 1D HSQC spectra and confirms that it is the residual unbound chaperone that is observed in these experiments. For $\alpha$BC, a decrease in diffusion coefficient was observed at higher concentrations of aggregating $\alpha$Lac. However, similarly to c$\alpha$BC, the shape of the 1D HSQC spectra stayed the same. The data can be explained by considering the oligomeric nature of full-length $\alpha$BC. The C-termini in $\alpha$BC are seen by virtue of their mobility. If an $\alpha$Lac molecule

is bound to or in the vicinity of a C-terminal tail, the latter would be expected to lose mobility and therefore become invisible in both the HSQC and the diffusion experiment. However, the remaining free tails on the same oligomer would show slower translational diffusion as they are now part of a larger particle, without any changes to their HSQC spectra. Thus the HSQC and diffusion data are consistent with the observation of residual unbound C-terminal tails on $\alpha$BC.

To summarise, 1D NMR can be used to follow aggregation and chaperone action in real time. Aggregation is slowed down in the presence of c$\alpha$BC or $\alpha$BC. The two chaperone constructs act differently since they give rise to traces of different shape. The 1D watergate, HSQC and diffusion data is consistent with two observables in the experiments on mixtures of $\alpha$Lac and chaperone: the sum of residual monomeric $\alpha$Lac and residual unbound chaperone ($^1$H experiments) and the residual unbound chaperone alone (heteronucleus-edited experiments). These definitions enable us to simulate the data based on kinetic models, as described in the following section.

### 5.3.2 Chemical kinetics modelling of the NMR data in the presence of chaperone

#### Overview

In order to interpret the 1D NMR data in terms of mechanism, I built models in the form of rate equations and compared their predictions to the NMR traces. **Chapter 4** presents the initial phase of this analysis, which entailed constructing an appropriate model for $\alpha$Lac aggregation in the absence of chaperone. The model with its parameters is depicted in **Figure 4.13**. The rate equations comprising this model were included throughout all work on chaperones with the parameters fixed to their best-fit values from the analysis in **Chapter 4**. This is equivalent to assuming that the chaperone operates by kinetic competition and does not change the

structure of free $\alpha$Lac, which is the simplest assumption one could make about an ATP-independent chaperone. New rate equations were added to describe binding and release of various $\alpha$Lac species by the chaperone. The rate constants for these chaperone binding reactions were treated as adjustable parameters and were optimised for each model using the Levenberg-Marquardt algorithm with basin hopping (the same methodology used in **Chapter 4**). Models were devised and statistically tested according to a strategy that aimed to find the minimal necessary and sufficient set of steps required to explain the experimental data.

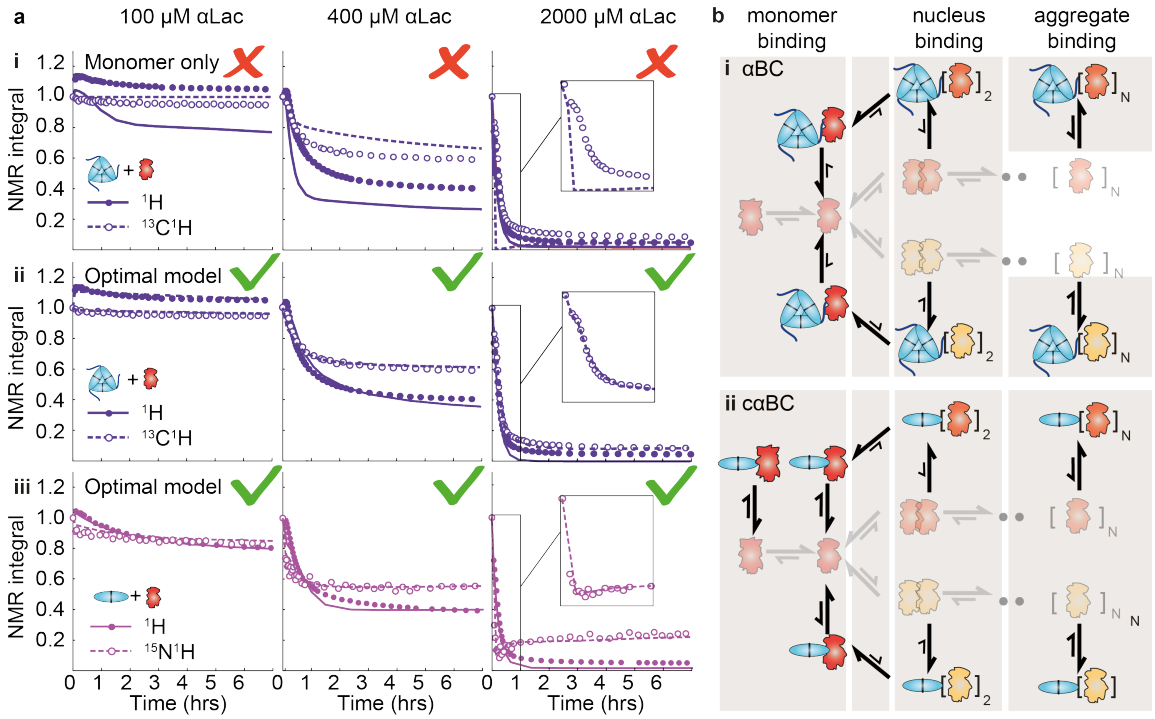



***Figure 5.5: Elucidation of chaperone mechanism from NMR data.*** ***a*** *Spectral integral time courses for three concentrations of $\alpha$Lac in the absence and presence of either $\alpha$BC or c$\alpha$BC at 200 $\mu$M, with fits by different models.* ***i*** *Poor fit produced by the monomer only binding model (lines) to $\alpha$BC data (points).* ***ii*** *Fit of the optimal model for full-length $\alpha$BC.* ***iii*** *Fit of the optimal model for c$\alpha$BC.* ***b*** *Mechanistic framework for the best fit models in* ***a*** *(**i** - c$\alpha$BC, **ii** - $\alpha$BC). The brown boxes indicate distinct sets of chaperone binding rates. The relative size of the arrows indicates the relative magnitude of the $K_D$ for each step.*

**Figure 5.5** illustrates the approach. It shows the poor fit produced by an initial simple model, incorporating chaperone binding only to monomeric $\alpha$Lac, and the much improved fit obtained for the models selected as final, in which the

chaperone acts at several points in $\alpha$Lac aggregation. Below I explain the setup of the chaperone binding models, using the monomer-only binding case as an example, and discuss the model selection strategy. For more details of the minimisation algorithm the reader is referred to **Section 4.3.4** in the preceding Chapter.

### Connecting the data and the simulation

The data on mixtures of $\alpha$Lac and the chaperones was fit in sets of two curves for each $\alpha$Lac concentration: the $^1$H time trace, reporting on the sum of the signal from the two proteins, and the corresponding heteronucleus-edited trace, reporting only on the chaperone within the mixture. Both were recorded concurrently on the same sample. The NMR spectra were normalised to those prior to addition of reducing agent. Thus, if $A^*$ and $A$ are the two misfolded monomeric forms of $\alpha$Lac through which aggregation proceeds (see **Section 4.3.5** and **Section A.1** in **Appendix A**), and $C$ is free chaperone, the NMR data reports on relative concentration of these species, weighted by appropriate spectral intensity factors:

$$S_{data1} = \frac{[A^*] + \phi_1[A] + \phi_2[C]}{[A]_{total} + \phi_2[C]_{total}} \tag{5.1}$$

$$S_{data2} = \frac{[C]}{[C]_{total}} \tag{5.2}$$

$S_{data1}$ corresponds to the observable of the $^1$H experiment and $S_{data2}$ to that of the heteronucleus-edited experiment. $\phi_1$ is the spectral intensity ratio between $A$ and $A^*$, and $\phi_2$ is the equivalent ratio between $C$ and $A$. $\phi_1$ was a constant known from the analysis in **Chapter 4**, and $\phi_2$ was included as a fitting parameter here. An aggregation model allows us to calculate the values of $[A^*]$, $[A]$ and $[C]$ at any time point; thus we can compute the predicted values $S_{sim1}$ and $S_{sim2}$. To obtain a measure of the goodness of fit of the model, we calculate the least-squares $\chi^2$ over all timepoints:

$$\chi^2_{LS} = \sum (S_{data1} - S_{sim1})^2 + \sum (S_{data2} - S_{sim2})^2 \tag{5.3}$$

**Setup of a chaperone binding model**

We thus require a set of rate equations that would allow us to calculate the concentrations of the monomeric species at any given time. A subset of those would involve interconversion of various $\alpha$Lac species without the involvement of chaperone (the $\alpha$Lac-only model). The best-fit model from **Chapter 4** was used, with parameter values treated as constants (see **Figure 4.13** and **Section A.1** in **Appendix A**).

Here, in addition to this basic $\alpha$Lac aggregation model, another set of differential equations was included describing chaperone binding to the various $\alpha$Lac species. For example, a model that involves chaperone binding all $\alpha$Lac monomer species was represented by the following reactions:

$$A^* + C \underset{k^-_{mon}}{\overset{k^+_{mon}}{\rightleftharpoons}} A^*C \tag{5.4}$$

$$A + C \underset{k^-_{mon}}{\overset{k^+_{mon}}{\rightleftharpoons}} AC \tag{5.5}$$

The rate equations for changes in concentrations of involved species were then as follows:

$$\left.\begin{aligned} d[A^*] &= -k^+_{mon}[A^*][C]dt \\ d[C] &= -k^+_{mon}[A^*][C]dt \\ d[A^*C] &= +k^+_{mon}[A^*][C]dt \end{aligned}\right\} \quad \text{equation 5.4 forward}$$

$$\left.\begin{aligned} d[A^*] &= +k^-_{mon}[A^*C]dt \\ d[C] &= +k^-_{mon}[A^*C]dt \\ d[A^*C] &= -k^-_{mon}[A^*C]dt \end{aligned}\right\} \quad \text{equation 5.4 reverse}$$

$$\left.\begin{aligned} d[A] &= -k^+_{mon}[A][C]dt \\ d[C] &= -k^+_{mon}[A][C]dt \\ d[AC] &= +k^+_{mon}[A][C]dt \end{aligned}\right\} \quad \text{equation 5.5 forward}$$

$$\left.\begin{aligned} d[A] &= +k^{-}_{mon}[AC]dt \\ d[C] &= +k^{-}_{mon}[AC]dt \\ d[AC] &= -k^{-}_{mon}[AC]dt \end{aligned}\right\} \quad \text{equation 5.5 reverse}$$

No self-association of chaperone was assumed in the models I tested. For full-length polydisperse $\alpha$BC, the C-termini visible in the NMR experiment were assumed to be independent of each other. Thus at the start of the reaction $[C] = [C]_{total} = 200\mu M$ and $[A^*] = [A]_{total}$. Using the rate equations, we can calculate $[A]$, $[A^*]$ and $[C]$ over time for given values of $k^{+}_{mon}$ and $k^{-}_{mon}$. We can therefore compute the values of $S_{sim1}$ and $S_{sim2}$ and compare these with the experimental data using **Equation 5.3**.

The fit of this monomer binding model for full-length $\alpha$BC is shown in **Figure 5.5a i**. Evidently, the fit to the data is poor at all three concentrations of $\alpha$Lac, therefore this cannot be an adequate representation of $\alpha$BC action and chaperone binding to other species should be considered. Through the model selection process described in the next section, I derived the necessary and sufficient sets of interactions for both $\alpha$BC and c$\alpha$BC, which are shown in **Figure 5.5b** and **5.13**. The full set of differential equations for both optimal models is given in **Sections A.2** for $\alpha$BC and **A.3** for c$\alpha$BC in **Appendix A**.

**Model selection strategy** A great diversity of possible chaperone mechanisms exists. Therefore an objective strategy is required to navigate the model space. The approach I adopted is explained in **Figure 5.6**. I used a combination of forward selection and reductive elimination to find the simplest set of parameters that would explain the data well. The statistical significance of improvements in fit was judged using the F-test (**Section 1.5.2**).

**Figure 5.7** illustrates the path taken through model space for each chaperone. The details of models tested in the forward selection stage are listed in **Figure 5.9** at the end of this section. The reductive elimination models for full-length $\alpha$BC are

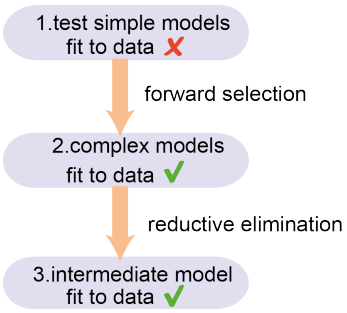



***Figure 5.6: The model selection strategy for chaperone mechanisms.*** *Simple models such as monomer binding produced poor fits to the data. In forward selection, I increased model complexity to find all steps necessary to explain the data and therefore achieve a good fit. In reductive elimination, I removed elements of the best-fitting complex model to produce the simplest sufficient set of steps that preserved the goodness of fit to the data.*

shown in **Figure 5.10** and those for cαBC are shown in **Figure 5.11**.

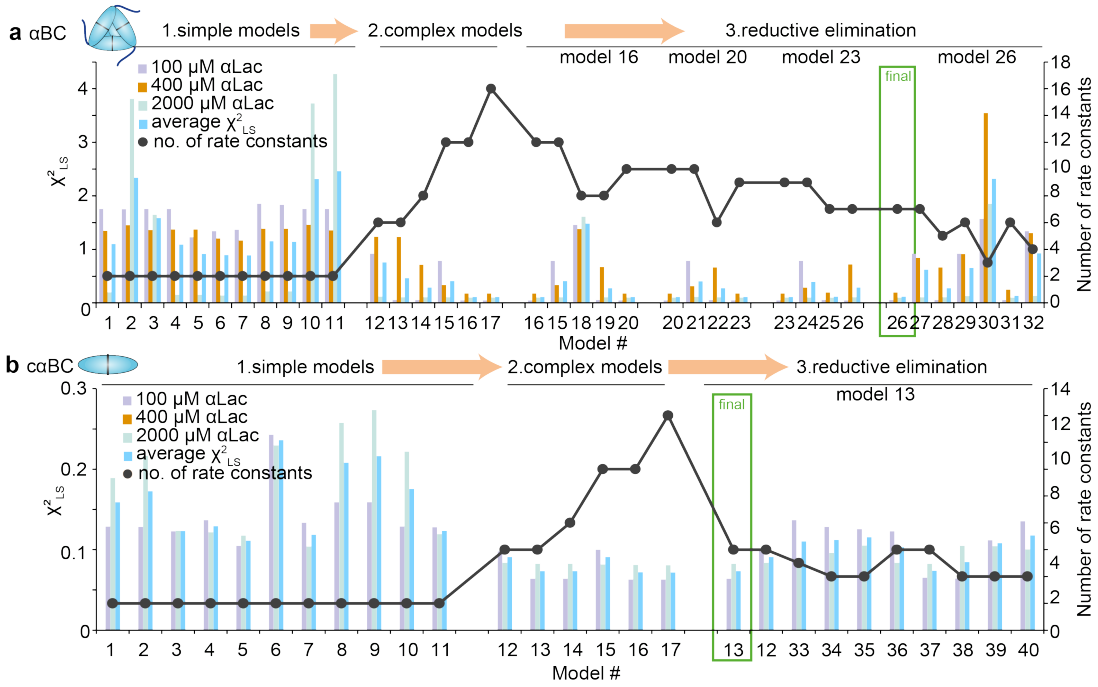



***Figure 5.7: Overview of modelling for data with chaperones.*** *Model ID numbers refer to* ***Figures 5.9-5.11****.* ***a*** *Mixtures of αLac and αBC. Model selection was performed on three datasets: 100 μM, 400 μM and 2000 μM αLac.* ***b*** *Mixtures of αLac and cαBC. Model selection was performed on two datasets: 100 μM and 2000 μM αLac. The graphs show the least squares $\chi^2$ for each dataset and the average of all datasets for each model, together with the number of rate constants in each model. The final model, for which all its reductive elimination sub-models resulted in a significantly higher average $\chi^2_{LS}$, is highlighted by a green box for each chaperone.*

I started with very simple models with either equal binding to all species or single-species binding (such as monomers only or nuclei only). None of these produced good fits. I then tested models incorporating combinations of these steps (forward selection). Increasingly comprehensive models improved in their fit to the data.

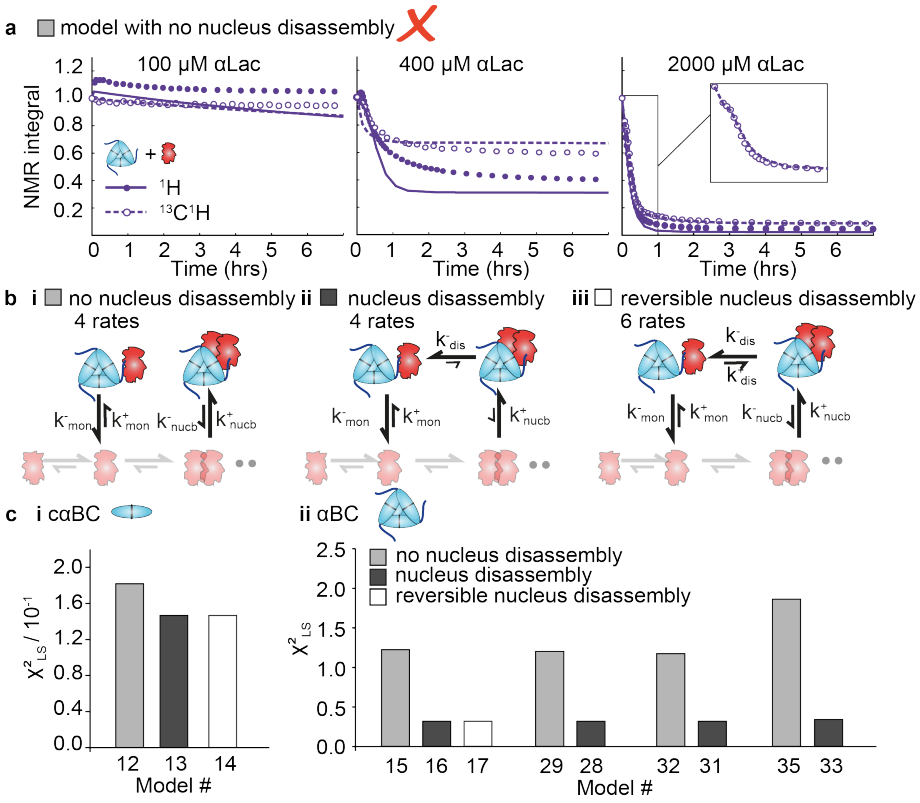



***Figure 5.8: The need for the nucleus disassembly step to fit the NMR aggregation data.*** ***c*** *The poor fit of a model without nucleus disassembly (αBC model 15).* ***b*** *Different models for the interaction of the chaperone with the nucleus, illustrated with αBC: (**i**) simple reversible binding without nucleus disassembly, (**ii**) nucleus binding and disassembly where chaperone-bound dimeric nucleus goes to chaperone-bound monomers, and (**iii**) reversible nucleus disassembly where chaperone-bound monomers can also come together to form a chaperone-bound nucleus.* ***c*** *$\chi^2_{LS}$ values for the models depicted in* ***b*** *for (**i**) cαBC and (**ii**) αBC. Model numbers correspond to those in* ***Figures 5.9-5.11****. For both chaperones, the model with nucleus disassembly conferred a significant improvement in $\chi^2_{LS}$ over the simple binding model (black vs gray), whereas the model with reverse nucleus disassembly did not result in further improvement in fit (white).*

However, at low concentrations of αLac the decay of $^{1}$H signal in the data was slower than that predicted (for examples see **Figure 5.5a i** and also **Figure 5.8a**). In other words, there was more αLac monomer than expected. How could the chaperone lead to the release of additional monomeric αLac? A possibility is that the binding energy of the chaperone could influence the conformation of bound αLac in such a way as to make its dissociation more favourable. I tested incorporation of this effect for the smallest αLac oligomers, the dimeric nuclei, by including conversion of chaperone-bound nuclei to chaperone-bound monomers (**Figure 5.8b**

**ii**). I also tested the companion model in which the nucleus disassembly step was reversible, i.e. the chaperone-bound monomers could associate to form chaperone-bound dimers (**Figure 5.8b iii**). The improvement in $\chi^2_{LS}$ afforded by inclusion of nucleus disassembly is shown in **Figure 5.8c**. The improvement was especially notable for full-length $\alpha$BC. In contrast, there was no further improvement in fit for the fully reversible model, i.e. the rate of association of chaperone-bound monomers was negligible (**Figure 5.8b iii** and **c**). Therefore the intermediate complexity model in **Figure 5.8b ii** was included in subsequent analysis.

The same set of models was tested for both chaperones in forward selection, but different models were justified to be taken forward for reductive elimination (see **Figure 5.9**): for $\alpha$BC it was justified to allow the chaperone to bind species in the two pathways with different sets of rate constants (model 16), whereas for c$\alpha$BC, no distinction was made between the two parallel pathways of $\alpha$Lac aggregation (model 13). Having found models that fit the data well, of the standard shown in **Figure 5.5a ii-iii**, I sought to simplify them to remove any unnecessary steps. I removed steps one by one, or set common rates for distinct processes and looked at the $\chi^2_{LS}$ of the resulting sub-models (see groups of models under reductive elimination in **Figure 5.7** and in **Figures 5.10** and **5.11**). For example, for model 16 in **Figure 5.10**, which includes independent binding to monomers, nuclei and aggregates and nucleus disassembly, some examples of submodels are (a) no nucleus disassembly, (b) no monomer binding or (c) no distinction between types of aggregate. If one of the sub-models preserved the goodness of fit of its parent as judged by the F-test, I took it forward and repeated the reductive elimination process. Note that the condition for a "pass" in an F-test depends on the type of analysis. The confidence interval $P$ gives the probability that the complex model can result in a fit of given quality by chance. Therefore in forward selection (**Figure 5.9**) a more complex model can be adopted if $P$ <0.05, whereas in reductive elimination the complex model in question must be abandoned for a simpler one if $P$ >0.05 (**Figures 5.10-5.11**). For models comprising an identical number of parameters, a lower $\chi^2_{LS}$ is a sufficient condition to

accept a model (denoted as "equal dof" in **Figures 5.9-5.11**). The model in which none of the steps could be removed without jeopardising the fit was selected as final (model 26 for $\alpha$BC and model 13 for c$\alpha$BC). The need for nucleus disassembly was tested at each stage of reductive elimination for both chaperones and was found to be indispensable throughout.

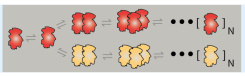
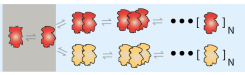
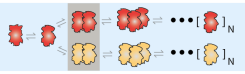
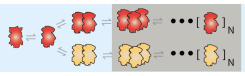
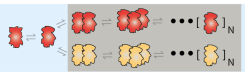
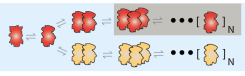
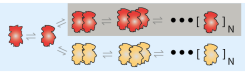
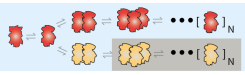
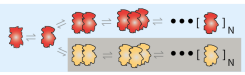
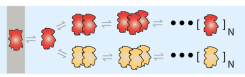
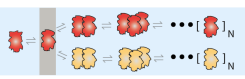
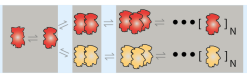
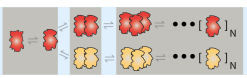
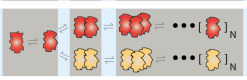
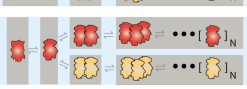
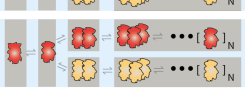
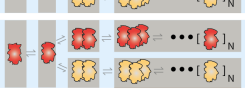
| | Model description | | Nucleus disassembly? | Rate constants | cαBC $\chi^2_{LS}$ sum | cαBC F-test p-level | αBC $\chi^2_{LS}$ sum | αBC F-test p-level |
|---|---|---|---|---|---|---|---|---|
| **Simple models: one set of association and dissociation rates** | | | | | | | | |
| 1 | monomer, nucleus, aggregate binding, all with same rates |  | n | 2 | 3.178E-01 | | 3.295E-00 | |
| 2 | binding both monomers with same rates |  | n | 2 | 3.452E-01 | | 7.012E-00 | |
| 3 | nucleus binding with same rates |  | n | 2 | 2.464E-01 | | 4.757E-00 | |
| 4 | aggregate binding all with same rates |  | n | 2 | 2.585E-01 | | 3.273E-00 | |
| 5 | aggregate, nucleus binding all with same rates |  | n | 2 | 2.224E-01 | | 2.749E-00 | |
| 6 | aggregate pathway 1 binding only, all with same rates |  | n | 2 | 4.723E-01 | | 2.674E-00 | |
| 7 | pathway 1 aggregate and nucleus binding, all with same rates |  | n | 2 | 2.375E-01 | | 2.671E-00 | |
| 8 | pathway 2 aggregate binding only, all with same rates |  | n | 2 | 4.165E-01 | | 3.452E-00 | |
| 9 | pathway 2 aggregate and nucleus binding, all with same rates |  | n | 2 | 4.325E-01 | | 3.431E-00 | |
| 10 | first monomer binding |  | n | 2 | 3.509E-01 | | 6.939E-00 | |
| 11 | second monomer binding |  | n | 2 | 2.470E-01 | | 7.384E-00 | |
| **Complex multi-rate models** | | | | | | | | |
| 12 | monomers, nucleus and aggregate binding distinct rates |  | n | 6 | 1.818E-01 | vs model 1: 1.5E-09 PASS | 2.265E-00 | vs model 1: 5.2E-08 PASS |
| 13 | as 12, nucleus disassembly |  | y | 6 | 1.467E-01 | vs model 12: equal dof PASS | 1.389E-00 | vs model 12: equal dof PASS |
| 14 | as 12, reversible nucleus disassembly |  | y | 8 | 1.467E-01 | vs model 13 1.0 FAIL | 8.700E-03 | vs model 13: 9.3E-13 PASS |
| 15 | as 12, monomer types and two pathways have distinct rates |  | n | 12 | 1.817E-01 | vs model 12: 1.0 FAIL | 1.225E-02 | vs model 14: higher $\chi^2$ FAIL |
| 16 | as 15, nucleus disassembly |  | y | 12 | 1.440E-01 | vs model 13: 1.0 FAIL | 3.206E-01 | vs model 14: 3.3E-23 PASS |
| 17 | as 15, reversible nucleus disassembly |  | y | 16 | 1.436E-01 | vs model 13: 1.0 FAIL | 3.206E-01 | vs model 16: 1.0 FAIL |

***Figure 5.9: Catalogue of models used for forward selection.*** *In model schematics, chaperone binding is represented by brown boxes over the corresponding αLac species. The $\chi^2_{LS}$ value given is a sum of $\chi^2_{LS}$ for all αLac concentrations tested. Orange boxes highlight models that passed an F-test against a simpler counterpart. Models 16 and 13 were justified to be taken forward for reductive elimination for αBC and cαBC respectively. 'Equal dof' means that the number of parameters was the same in the models in question and the F-test could not be performed; the $\chi^2_{LS}$ values were compared directly in those cases.*

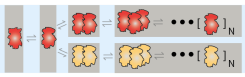
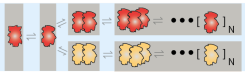
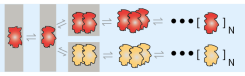
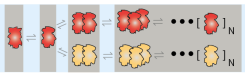
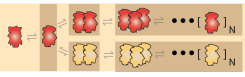
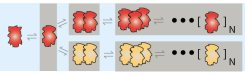
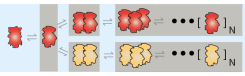
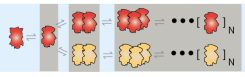
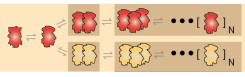
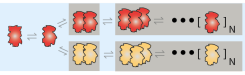
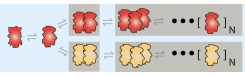
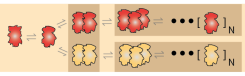
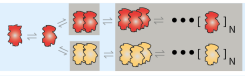
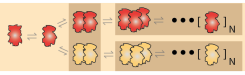
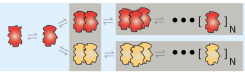
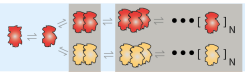
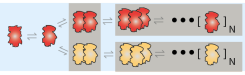
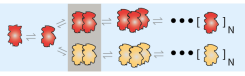
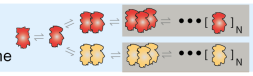
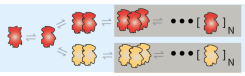
| Step removed | Model description | Nucleus disassembly? | Rate constants | $\chi^2_{LS}$ sum | F-test p-level |
|---|---|---|---|---|---|
| **REDUCTIVE ELIMINATION FOR MODEL 16** | | | | | |
| 16 carried forward |  | y | 12 | 3.206E-01 | |
| 15 as 16, no nucleus disassembly |  | n | 12 | 1.225E-00 | vs model 16: equal dof FAIL |
| 18 as 16, no aggregate binding |  | y | 8 | 4.443E-00 | vs model 16: 4.0E-69 FAIL |
| 19 as 16, no distinction between pathways |  | y | 8 | 8.219E-01 | vs model 16: 2.9E-18 FAIL |
| 20 as 16, no monomer 1 binding |  | y | 10 | 3.207E-01 | vs model 16: 1.0 PASS |
| **REDUCTIVE ELIMINATION FOR MODEL 20** | | | | | |
| 20 carried forward |  | y | 10 | 3.207E-01 | |
| 21 as 20, no nucleus disassembly |  | n | 10 | 1.204E-00 | vs model 20: equal dof FAIL |
| 22 as 20, no distinction between pathways |  | y | 6 | 8.236E-01 | vs model 20: 7.7E-23 FAIL |
| 23 as 20, no direct monomer binding |  | y | 9 | 3.212E-01 | vs model 20: 0.97 PASS |
| **REDUCTIVE ELIMINATION FOR MODEL 23** | | | | | |
| 23 carried forward |  | y | 9 | 3.212E-01 | |
| 24 as 23, no nucleus disassembly |  | n | 9 | 1.175E-00 | vs model 23: equal dof FAIL |
| 25 as 23, no distinction between nuclei from pathway 1 and 2 |  | y | 7 | 3.428E-01 | vs model 23: 0.16 PASS |
| 26 as 23, no distinction between aggregates from pathway 1 and 2 |  | y | 7 | 8.697E-01 | vs model 23: 3.3E-28 FAIL |
| **REDUCTIVE ELIMINATION FOR MODEL 26** | | | | | |
| 26 carried forward |  | y | 7 | 3.428E-01 | |
| 27 as 26, no nucleus dissembly |  | n | 7 | 1.862E-00 | vs model 26: same dof FAIL |
| 28 as 26, no distinction between aggregates from pathway 1 and 2 |  | y | 5 | 8.235E-01 | vs model 26: 1.03E-25 FAIL |
| 29 as 26, no monomer dissociation following nucleus disassembly |  | y | 6 | 1.956E-00 | vs model 26: 2.3E-55 FAIL |
| 30 as 26, no aggregate binding |  | y | 3 | 6.961E-00 | vs model 26: 1.3E-89 FAIL |
| 31 as 26, monomer dissociation and nucleus disassembly rates the same |  | y | 6 | 3.992E-01 | vs model 26: 5.2E-05 FAIL |
| 32 as 26, no nucleus binding or nucleus disassembly |  | n | 4 | 2.774E-00 | vs model 26: 3.7E-62 FAIL |

***Figure 5.10: Catalogue of reductive elimination models for full-length αBC.*** *Model 16 was the result of forward selection for αBC. Several rounds of simplification were performed. For model 26 all the submodels resulted in a significant worsening of the fit. Therefore model 26 is both sufficient and necessary to explain the NMR data for mixtures of αLac and αBC.*

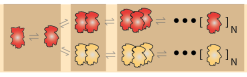
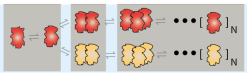
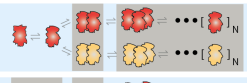
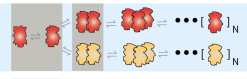
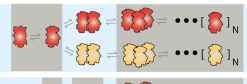
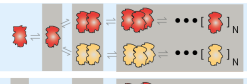
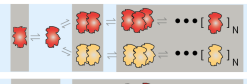
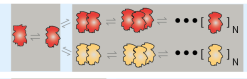
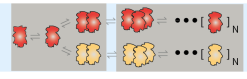
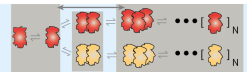
| Step removed | Model description | Nucleus disassembly? | Rate constants | $\chi^2_{LS}$ sum | F-test p-level |
|---|---|---|---|---|---|
| **REDUCTIVE ELIMINATION FOR MODEL 13** | | | | | |
| 13 carried forward |  | y | 6 | 1.467E-01 | |
| 12 as 13, no nucleus disassembly |  | n | 6 | 1.818E-01 | vs model 13: equal dof FAIL |
| 33 as 13, no monomer binding |  | y | 5 | 2.204E-01 | vs model 13: 1.4E-09 FAIL |
| 34 as 13, no aggregate binding |  | y | 4 | 2.244E-01 | vs model 13: 1.1E-08 FAIL |
| 35 as 13, no nucleus binding and disassembly |  | n | 4 | 2.309E-01 | vs model 13: 2.7E-09 FAIL |
| 36 as 13, no monomer 1 binding |  | y | 6 | 2.065E-01 | vs model 13: equal dof FAIL |
| 37 as 13, no monomer 2 binding |  | y | 6 | 1.478E-01 | vs model 13: equal dof FAIL |
| 38 as 13, nucleus and aggregate binding with same rates |  | y | 4 | 1.691E-01 | vs model 13: 6.3E-03 FAIL |
| 39 as 13, nucleus and monomer binding with same rates |  | y | 4 | 2.161E-01 | vs model 13: 6.6E-08 FAIL |
| 40 as 13, monomer and aggregate binding with same rates |  | y | 4 | 2.358E-01 | vs model 13: 9.8E-10 FAIL |

***Figure 5.11: Catalogue of reductive elimination models for the core domain chaperone, cαBC.*** *Model 13 was the result of forward selection for cαBC. None of the sub-models presented here preserved the goodness of fit, as judged by the F-test. Therefore the parent model 13 is both sufficient and necessary to explain the NMR data for mixtures of αLac and cαBC. 'Equal dof' means that the number of parameters was the same in the models in question and the F-test could not be performed; the $\chi^2_{LS}$ values were compared directly in those cases.*

### 5.3.3 Chaperone mechanisms for αBC and cαBC

The final models of chaperone action for αBC and cαBC are presented in **Figure 5.12** and **Figure 5.13**. **Figure 5.12** shows the fits of the models to the data. The fits were performed to pairs of $^1$H and heteronucleus-edited curves for each αLac concentration individually. Asterisks mark the data that was not used in model selection but was included here as a validation step. For both αBC and cαBC, the fit is very good, including these additional data, which increases confidence in the models.

**Figure 5.13a** shows schematics of the steps included in the models for each chaperone. In terms of αLac species they target, full-length αBC binds the nucleus

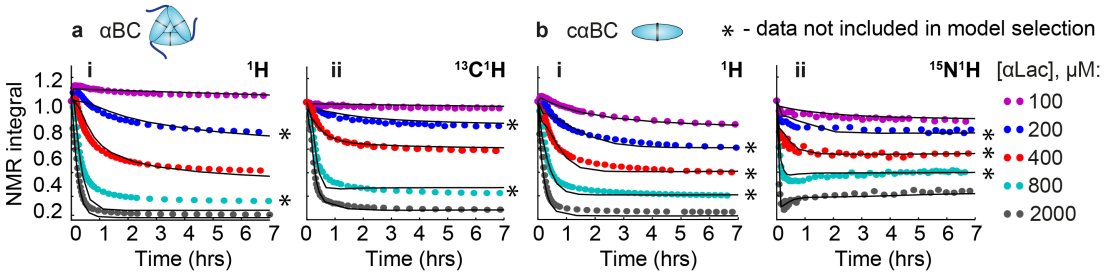



***Figure 5.12: The fits of the optimal models to the full datasets.*** *(**i**) αBC, (**ii**) cαBC. The chaperones were at 200 μM and αLac concentrations are indicated in the legend. The curves marked with asterisks were excluded from model selection described in the previous section and only fitted with the final models as a validation step.*

and the two types of aggregate at distinct rates, and is able to disassemble the nuclei and release the resulting monomers. In contrast, the core domain binds all species - the monomers, the nuclei and larger aggregates, without distinguishing between the two pathways. Thus αBC is more selective than its isolated core domain.

**Figure 5.13b** shows the values of the optimised parameters for fits of individual datasets at each concentration of αLac. Most parameters are within the same order of magnitude, although a few vary over several orders of magnitude. There are several possible reasons for the inconsistensies. Firstly, models of higher complexity may need to be tested. Secondly, the cost function surface could be very rugged so that the optimisation algorithm struggled to find the best minimum. Finally, there could be correlations between parameters. These could be revealed by a bootstrapping analysis. Ideally the model should be optimised simultaneously across several concentrations (a global fit). This was not feasible with the current speed of calculation, but with further improvement of the algorithm (for example by using Markov models rather than sequential integration) this may be possible, allowing more stringent model selection.

The average parameters from all five concentrations are presented as distributions in **Figure 5.13c** and **d**. It is interesting to compare the rates of chaperone binding to various species to the rates of αLac interconversions in the absence of chaperone, determined in **Chapter 4** and included here in **Figure 5.13e**. For ex-

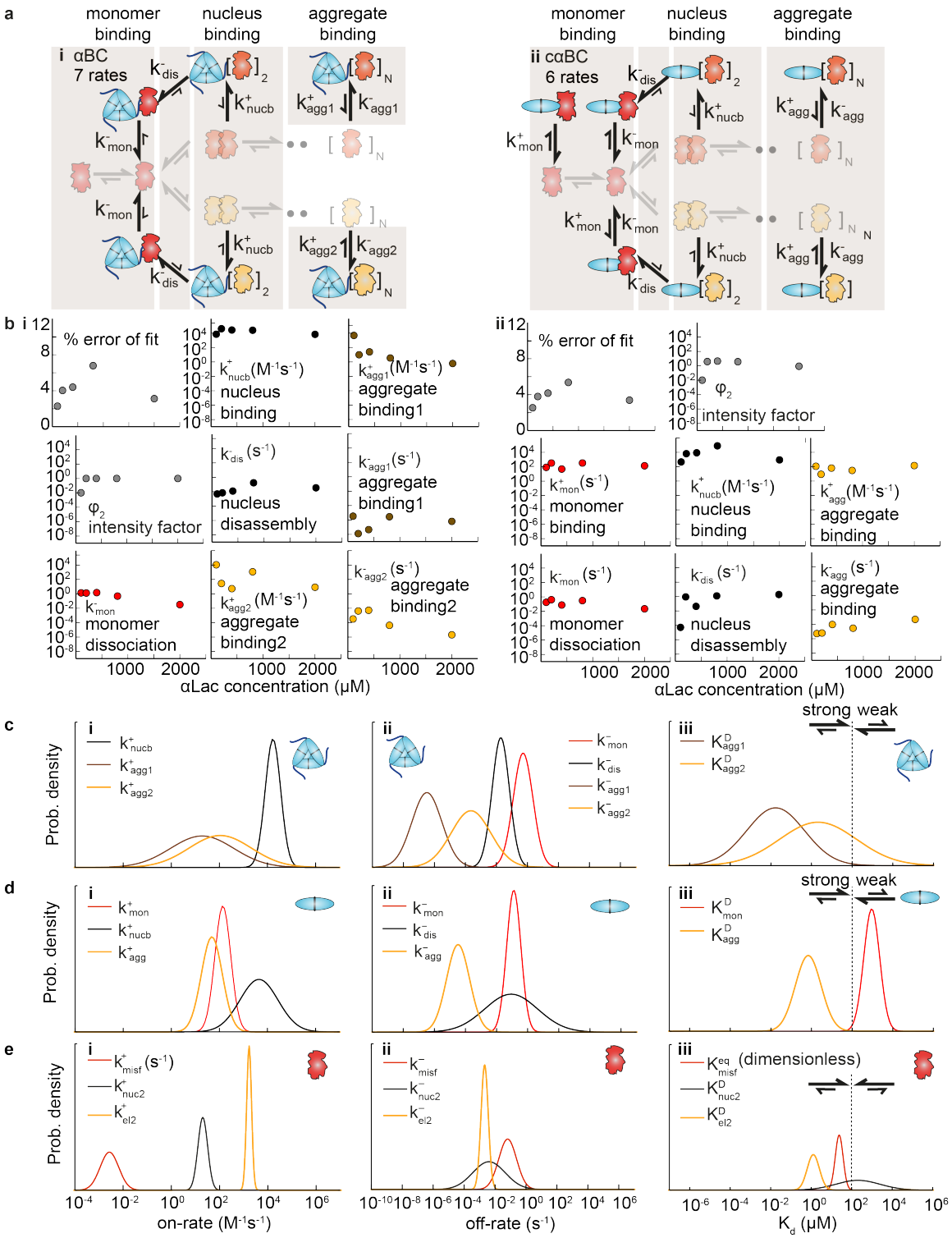



***Figure 5.13: Analysis of chaperone action models.*** ***a*** *Model schematics for* ***(i)*** *αBC and* ***(ii)*** *cαBC. The relative size of the arrows indicates the magnitude of the $K^D$ according to the threshold shown in* ***c/d iii****. The smallest arrows refer to processes that are so slow that the inclusion of a rate constant was not needed.* ***b*** *Parameter values for the fit at each αLac concentration.* ***(i)*** *αBC and* ***(ii)*** *cαBC.* ***c*** *The average optimised rate and dissociation constants for the αBC model in* ***a i****. The parameters are represented as lognormal distributions with the average and standard deviation of the five values in* ***b****.* ***d*** *The average parameters for cαBC.* ***e*** *The average parameters for the major pathway of αLac aggregation from* ***Figure 4.13****, included for comparison.*

ample, $\alpha$BC binds $\alpha$Lac nuclei faster than the nuclei elongate to form aggregates (black curves in **c i** and **d i** vs orange in **e i**). The rates of nucleus disassembly and monomer release by the full-length chaperone (black and red curves in **c ii**) are also faster than the rate of spontaneous nucleus breakdown (black curve in **e ii**). The rates of nucleus disassembly by $\alpha$BC and c$\alpha$BC are comparable, but nucleus binding is faster for $\alpha$BC. Therefore $\alpha$BC provides an alternative pathway from $\alpha$Lac nucleus to $\alpha$Lac monomer with a lower activation energy, with the full-length chaperone being more efficient than the isolated core domain.

Considering aggregate binding, both the on- and off-rates for this are slower for $\alpha$BC and c$\alpha$BC compared to forward and backward elongation rates of $\alpha$Lac, with the $K_D$ values roughly similar (yellow curves in **Figure 5.13c-e**). Thus the chaperone can compete with $\alpha$Lac for association with aggregates. This would be expected to retard aggregation by reducing the effective concentration of $\alpha$Lac aggregate sites available for elongation. Taking a tetramer as an example, we can represent the rate of change of its concentration due to forward elongation as $\frac{d[A_4]}{dt} = -k_{el}^{+}[A_4][A]$. If some of the tetramers are bound by the chaperone, their concentration will be a fraction of that in the absence of chaperone: $\frac{d[A_4]}{dt} = k_{el}^{+}c[A_4][A]$, where $c < 1$. This fraction can be incorporated into the elongation rate constant, thus giving an effective rate that is slower than in the absence of chaperone: $\frac{d[A_4]}{dt} = -ck_{el}^{+}[A_4][A] = -k_{app}^{+}[A_4][A]$, where $k_{app}^{+} < k_{el}^{+}$. The action of the chaperone can therefore be seen as increasing the activation energy barrier for aggregate growth.

The free $\alpha$Lac monomer is right at the beginning of the aggregation pathway. In the cell, this form would be most amenable to ATP-dependent chaperones capable of refolding substrates back to the native form. It is therefore interesting to consider how the presence of the chaperones affects the amount of this species. This can be calculated using the aggregation simulation and the optimised rate constants. **Figure 5.14a-c** shows the fraction of $\alpha$Lac monomer in time and concentration space. With increasing concentration of chaperone, the red aggregation region, in which the

free monomer is depleted, is pushed back to longer times and higher $\alpha$Lac concentrations, i.e. $\alpha$Lac is protected from aggregation in a wider time and concentration space. However, at sufficiently long times and high substrate concentrations the chaperone can no longer counter monomer loss. In addition, at high concentrations of c$\alpha$BC (**Figure 5.14c ii**), there is generally less free $\alpha$Lac monomer (the surface is pink) because of the tendency of c$\alpha$BC to bind it directly. Interesting insight is obtained for equivalent surfaces for unbound chaperone (**Figure 5.14d-e**). In the regions of greater suppression of $\alpha$Lac aggregation (in the vicinity of the black isoline), the chaperone is mostly unbound. At long times and high $\alpha$Lac concentrations the chaperone is fully bound and effectively overwhelmed.

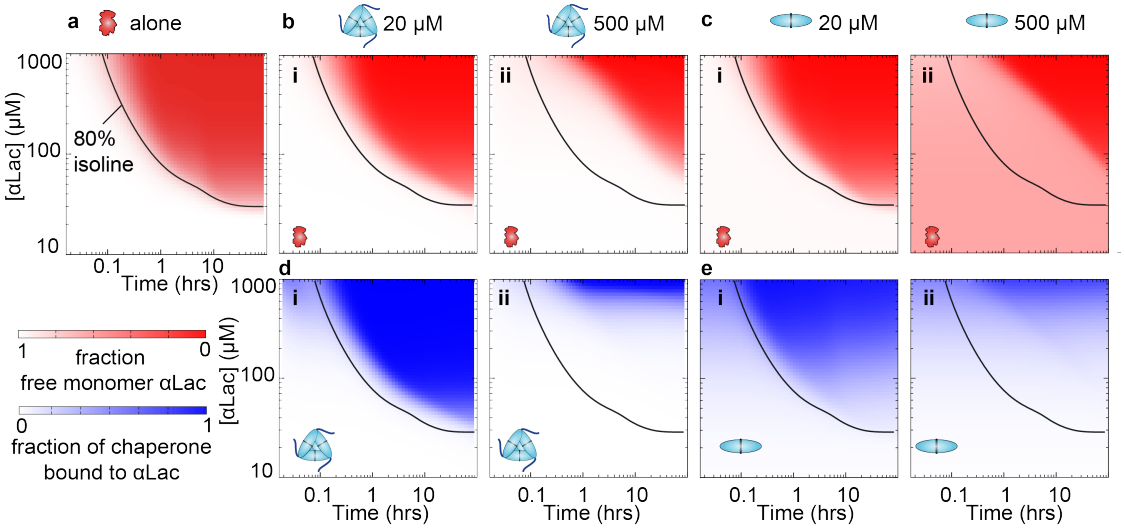



***Figure 5.14: Kinetic phase diagrams for free monomer $\alpha$Lac and free chaperone. a-c*** *Surfaces showing the fraction of free monomer $\alpha$Lac as a function of time and total $\alpha$Lac concentration, in the absence of chaperone (**a**), with $\alpha$BC (**b**) and with c$\alpha$BC (**c**). (**i**) 20 $\mu$M chaperone, (**ii**) 500 $\mu$M chaperone. The black line shows the onset of aggregation in the absence of chaperone (see **Figure 4.14**).* ***d-e*** *Surfaces showing fraction of chaperone bound to any $\alpha$Lac species at different total chaperone concentrations as above.*

The effect of the chaperone on $\alpha$Lac monomer can be seen more clearly by normalising the surfaces. **Figure 5.15** shows the plots from **Figure 5.14b** and **c** divided by the surface in **a**. These normalised surfaces make it clear where there are changes in monomer concentration in the presence of chaperone. The plots reveal that there are regions on the edge of the aggregation zone where the presence of chaperone leads to an increase in the fraction of monomeric $\alpha$Lac. I termed these

areas regions of protection from aggregation, and the fold increase in monomer concentration the rescue factor. The regions of protection are broader at higher concentration of chaperone, and the maximum rescue factor is higher (it is under 2 in the presence of 20 $\mu$M of either chaperone and up to 10 in the presence of 500 $\mu$M chaperone. The contour plots in **Figure 5.15c** show how the boundaries of the protection region evolve with increasing chaperone concentration. To the bottom left of this zone (short times and low total $\alpha$Lac concentration) $\alpha$Lac does not aggregate, and to the top right it aggregates so quickly that the chaperone is overwhelmed. It is also clear that at equivalent concentration $\alpha$BC is more effective at preventing monomeric $\alpha$Lac from aggregation than the isolated core domain as its protection region is larger.

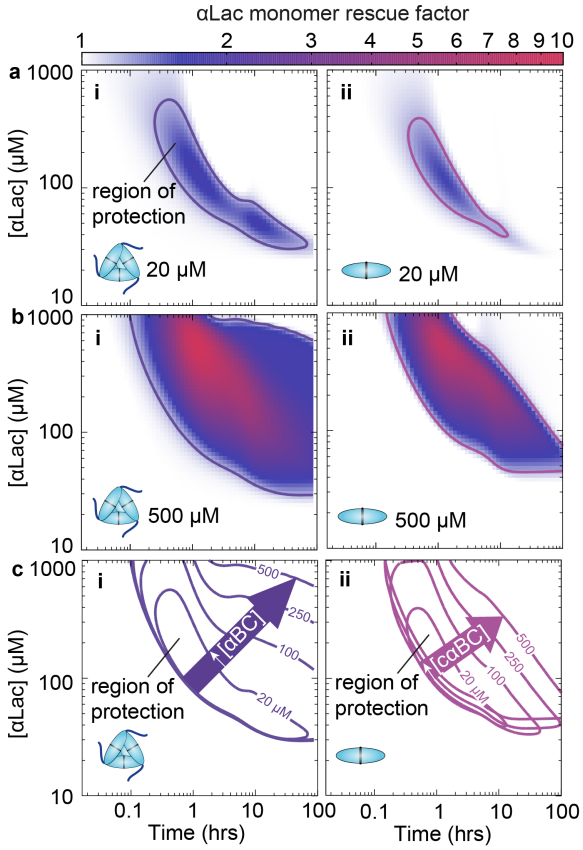



***Figure 5.15:*** ***$\alpha$BC protects $\alpha$Lac monomer from aggregation in a limited time and concentration window.*** *Panels **(i)** always show $\alpha$BC data and panels **(ii)** show c$\alpha$BC data.* ***a*** *Surfaces showing factor increase in the concentration of free $\alpha$Lac monomer in the presence of 20 $\mu$M chaperone as a function of time and total $\alpha$Lac concentration, relative to the monomer concentration in the absence of chaperone. A contour line is drawn at the value of 1.2.* ***b*** *The equivalent surfaces for 500 $\mu$M chaperone.* ***c*** *Contour plots of the region of protection for **(i)** $\alpha$BC and **(ii)** c$\alpha$BC at a range of chaperone concentrations.*

We can analyse how the chaperones achieve their protective effect by looking at the importance of the different steps their models include. **Figure 5.16** shows how the region of protection is affected if different processes in the models are turned off. Turning off the ability for nucleus disassembly has a strong effect on

the shape of the surface, i.e. it is significant in the chaperone function of both cαBC and αBC. Removing this function is equivalent to abrogating all interaction with nuclei. Hence the main protective action of αBC and cαBC with regard to nuclei is not the capacity to bind them but the capacity to disassemble them into monomers. Aggregate binding also plays a role. For cαBC, direct monomer binding is not significant as removing this step does not affect the protection region (**Figure 5.16b v**).

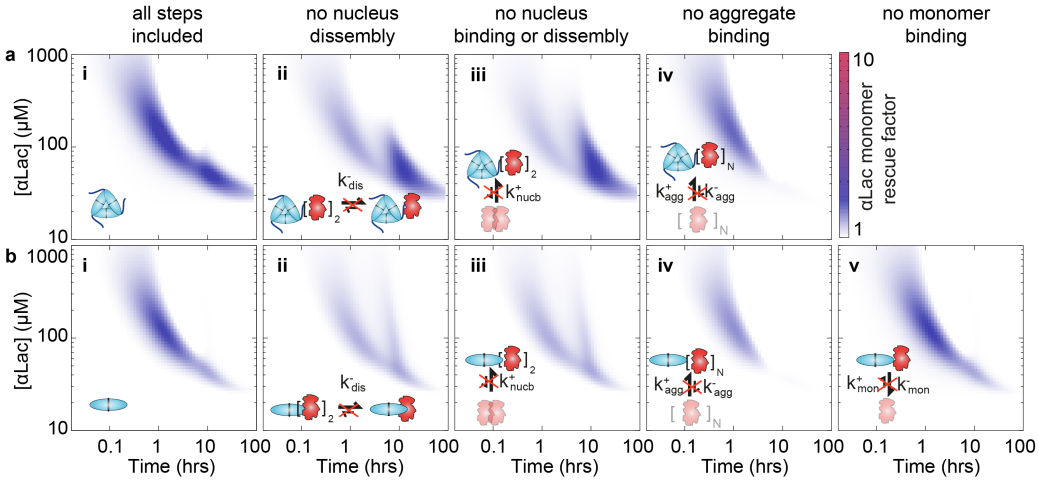



***Figure 5.16: The contributions of different chaperone interactions to protection of αLac monomer from aggregation.*** *The plots show the fold increase in free αLac monomer (αLac monomer rescue factor) in the presence of 20 μM αBC (**a**) or cαBC (**b**) **i** Extent of monomer αLac protection as predicted by full models derived here. **ii-v** The effect of removing different types of interaction as annotated. A significant difference from **i** indicates that this type of interaction is particularly important for preserving αLac monomer. Note that compensation by other types of interactions can occur in **ii-iv**, resulting in some regions being more intense than in i or different shape of the protection region.*

In summary, αBC is able to delay the loss of αLac monomer in a finite time and concentration window, i.e. it changes the kinetics but not the thermodynamics of aggregation. This is an expected mode of action for a chaperone that does not consume ATP. It acts in two main ways: by providing an alternative pathway from nuclei to monomers through nucleus disassembly and increasing the kinetic barrier to elongation through competitive binding. This is summarised in **Figure 5.17**. The isolated core domain acts by broadly the same mechanism as the full-length protein but is less selective, for example binding αLac monomer directly with a low

affinity, which does not contribute significantly to its protection, and effective in a smaller area of time and concentration space compared to its full-length counterpart.

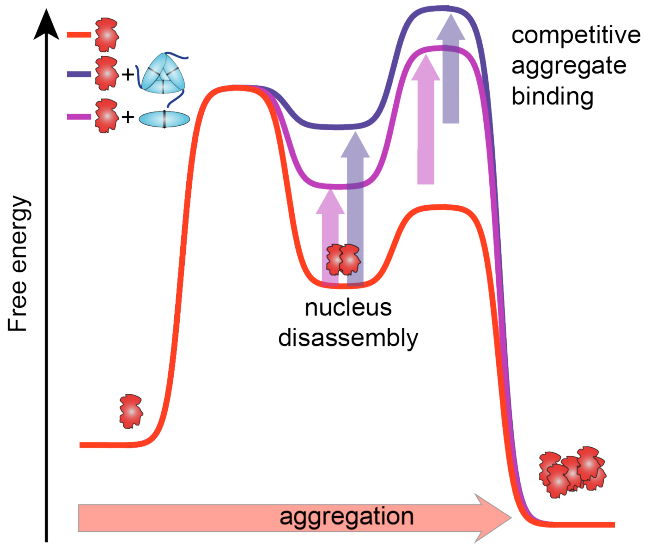



***Figure 5.17: Summary of the mechanism of action of αBC.*** *The chaperone drives the system towards monomers through two main kinetic effects: nucleus disassembly, which provides a more accessible path from nuclei to monomers, and competitive binding of aggregates.*

While aggregate binding has been observed for αBC by a number of groups, nucleus disassembly is a surprising new feature in the mechanism of αBC and the sHSP family. Nucleation is the rate-limiting step of aggregation, therefore targeting this step is an especially effective strategy. While binding to nuclei might be expected, the ability to break them down is remarkable for an ATP-independent chaperone. Nevertheless, other examples of ATP-independent disaggregases have previously been reported (e.g. in the chloroplast recognition particle, Jaru-Ampornpan et al. [2013]). In the case of αBC, the binding energy of the chaperone to the nucleus might be sufficient to disrupt the interactions holding the αLac monomers together. Direct study of this interaction is a great challenge as it is difficult to isolate free or chaperone-bound nuclei. A microfluidics approach has recently been shown to allow quantitative study of single nucleation events (Knowles et al. [2011]), and could be a fruitful strategy for analysing the disassembly function of αBC.

### 5.3.4 Limitations of the method

While the NMR and chemical kinetics approach is very powerful for studying heterogeneous protein mixtures, it is important to consider the assumptions made in

the analysis and the intrinsic limitations of the method.

Firstly, as with any chemical kinetics approach, a mechanism can clearly be shown to be incorrect in given conditions but cannot be definitively proven, even if it agrees with the experimental data. However, in such cases the principle of parsimony should be applied, i.e. the simplest mechanism that fits the evidence should be selected. In addition, predictions based on the accepted mechanism can be put forward and tested, as I did here by excluding some data to be fitted only with the final model, and additional experimental evidence can be used to help decide between alternatives.

Secondly, the nature of the numerical integration of rate equations is linear, i.e. each timepoint depends on the previous one. Thus, while the code can be parallelised for optimisation of each parameter, the parallelisation of individual runs is difficult and the calculations are computationally demanding. Thus, for example, global model selection on chaperone datasets for multiple $\alpha$Lac concentrations was not computationally feasible. The protein aggregation problem is analogous to small molecule polymerisation, for which Monte Carlo-based simulations have been developed (Chaffey-Millar et al. [2007], Van Steenberge et al. [2014]). The applicability of these more efficient algorithms could be explored and could potentially enable the use of more advanced metaheuristics for better exploration of the multidimensional parameter space, for example particle swarm optimisation methods.

A more efficient calculation algorithm would also enable more detailed representation of the interconversions between the various species. For example, in the current schemes direct interconversion between chaperone-bound oligomers and self-association of chaperone were not accounted for. While the ability to fit the data in the absence of these processes suggests that they are perhaps less critical, it could also mean that the method has limited distinguishing power for variables that are far removed from the observable, such as oligomer interconversions that do not involve the directly detected monomer. Whether or not this is the case would become clear

from comparing the simulated data with and without these interactions.

Along the same line, in this analysis, full-length $\alpha$BC has not been modelled as polydisperse oligomers but rather as a collection of independent C-termini. This is an oversimplification, for example because in a large oligomer not all C-termini can be expected to be exposed on the surface. Also, the substrate can bind the N-termini or the body of the core domain without significant effect on the C-termini, which would underestimate the effectiveness of the chaperone. Parallel analysis of the isolated core domain has shown that it acts by broadly the same mechanism as the full-length molecule but is less selective and less efficient. To assess the relative effect of the C-termini, the core domain and the N-termini systematically, another $\alpha$BC construct could be explored in which only the N-terminus has been truncated. This construct has been produced before and is soluble (Laganowsky et al. [2010]).

## 5.4 Conclusions

In summary, I have derived quantitative kinetic mechanisms for chaperone action by $\alpha$BC and its core domain using NMR and chemical kinetics. I proposed a new feature in the mechanism of $\alpha$BC - the ability to disassemble chaperone-bound nuclei to monomers, and have shown that nucleus disassembly and competitive aggregate binding by $\alpha$BC serve to delay aggregation of $\alpha$Lac monomer in a finite time and concentration window. Full-length $\alpha$BC is more selective in the species it binds and more efficient than its isolated core domain. The methodology presented here is widely applicable to other complex macromolecular mixtures. With isotopic labelling, quantitative analysis of macromolecular interactions by NMR and chemical kinetics can be performed in the presence of other cellular factors, in cell lysates or even inside cells.

# Chapter 6

# Characterisation of chaperone-substrate complexes

*In this Chapter I present the experiments aimed at detecting and characterising the species present in mixtures of full-length αBC and cαBC chaperones with αLac by a variety of biophysical methods. The results indicate formation of large aggregates containing both the chaperone and the aggregating substrate. I show that complexes with cαBC are more labile whereas those with αBC can be isolated by SEC. I measure their average drift time distribution by ion mobility mass spectrometry, and show that the density of the complexes is similar to that of free αBC.*

## 6.1 Introduction

One of the key questions in the study of sHSP chaperones, including the human sHSP $\alpha$BC, is how they interact with their aggregating clients. This is the third in a series of four Chapters in which I investigate different aspects of the interaction between $\alpha$BC and the model client $\alpha$Lac. The chemical kinetics analysis presented in **Chapters 4** and **5** established that there are two main chaperone functions of $\alpha$BC: disassembling the dimeric $\alpha$Lac nuclei to monomers, and competitive binding of $\alpha$Lac aggregates. Here I aim to detect and isolate these predicted complexes between $\alpha$BC and $\alpha$Lac and to characterise their size, shape and stoichiometry.

This task requires acknowledgement of the heterogeneity and dynamic nature of aggregating mixtures, and application of a variety of methods with awareness of their limitations and complementarity. In particular, those methods that rely on mass transport, such as SEC and analytical ultracentrifugation (AUC), must factor in subunit exchange rates relative to the separation timescale (Cann and Goad [1970], Stevens [1989]). Notably, if the rate of subunit exchange is comparable to or faster than the rate of separation, the observed size distribution can be skewed. For example, an apparent decrease in average mass of over 25% was obtained for an $\alpha$BC mutant when SEC was performed at 27°C compared to 4°C (Hochberg and Benesch [2014]). The difference was due to an increase in subunit exchange rate and not due a real shift in the size distribution. Thus care must be taken when interpreting shifts in mass transport profiles for dynamic systems.

Coming back to the biological context of this Chapter, an attractive model for the function of oligomeric sHSPs is that the oligomers represent a storage form of the active dimeric chaperone unit (van Montfort et al. [2002], Haslbeck and Vierling [2015], Strauch and Haslbeck [2016]). This hypothesis is based on the study of the yeast sHSP Hsp26, which has been shown to dissociate at heat shock temperatures with a concomitant enhancement of chaperone function (Haslbeck et al. [1999]). Further proposed evidence for this mode of action is that some chaperone-active sHSPs

exist exclusively as dimers (Basha et al. [2013], Bepperling et al. [2012]) and that when $\alpha$BC is phosphorylated, it exhibits a smaller average size and improved ability to prevent aggregation of the substrate p53 (Peschek et al. [2013]). Nevertheless in the latter case significant dissociation into dimers was not demonstrated.

An alternative hypothesis is that disassembly of oligomers is not required for chaperone action. Heat-induced dissociation and activation is not a general feature of the sHSPs and does not occur for $\alpha$BC (Haslbeck et al. [1999], Reddy et al. [2000]). Glutaraldehyde crosslinking of $\alpha$-crystallin, while preserving the average size of the oligomers, did not impair its ability to inhibit aggregation of the substrate $\beta$-crystallin (Augusteyn [2004]), suggesting that neither transition to smaller oligomers nor subunit exchange *per se* is necessary for chaperone function. Similarly, there was no correlation between subunit exchange rates and chaperone potency in different constructs of $\alpha$A-crystallin (Aquilina et al. [2005]).

Moreover, small chaperone-substrate complexes, such as chaperone dimers bound to one or two molecules of the aggregating substrate, have rarely, if at all, been observed. SEC profiles typically show that complexes are larger than the chaperone itself, often eluting in the void volume of the available columns, which indicates that their molecular mass is on the scale of MDa (Rao et al. [1993], Ehrnsperger et al. [1997], Lee et al. [1997], Kulig and Ecroyd [2012]). This is in agreement with electron microscopy analysis of the chaperone-substrate mixtures. With amorphously aggregating substrates the predominant species observed are globular particles with the diameter on the order of tens of nanometres, corresponding to megadaltons in mass (Ehrnsperger et al. [1997], Lee et al. [1997], Haslbeck et al. [1999], Rajagopal et al. [2015]). With amyloid-forming substrates, elongated assemblies (Ecroyd et al. [2007]) or fibrils with associated immunogold-labelled chaperone (Waudby et al. [2010], Shammas et al. [2011]) have been observed by EM. While formation of large complexes does not preclude initial dissociation of the oligomeric chaperone, there is also no need to invoke the latter to explain these data.

Native mass spectrometry is especially powerful for analysing the composition of polydisperse proteins and could help reconcile the above views. In a state-of-the-art study, Stengel et al. [2010] used native MS to look at pea Hsp18.1. They showed that the oligomeric distribution of this protein is sensitive to temperature, with the monodisperse 12-mer undergoing a transition to a larger polydisperse set of oligomers, as well as significant amounts of free monomer and dimer, at 43°C. Upon addition of the client luciferase, large and polydisperse complexes were formed. The authors were able to resolve and assign the most common stoichiometries, with the most abundant being $[\text{Hsp18.1}]_{24}[\text{Luc}]_2$. Thus although quaternary reorganisation of this sHSP occurs, it is the large oligomers that bind the client rather than the free monomers and dimers.

Below I describe the sizes and shapes of particles formed in mixtures of $\alpha$Lac and either c$\alpha$BC or $\alpha$BC. I show that large complexes with $\alpha$Lac can be detected for both chaperones by native MS. Measurement of the average IM arrival time distribution for the complexes with $\alpha$BC allows me to propose that although there is an increase in particle size upon $\alpha$Lac binding, there is no significant change in particle density.

## 6.2 Materials and methods

Please see **Section 4.2** for details on aggregation reactions and negative stain electron microscopy, **Section 3.2** for $\alpha$BC purification and **Section 2.2** on c$\alpha$BC purification. All reagents were from Sigma unless stated otherwise.

**Dynamic light scattering** Measurements were made on the Viscotek 802 instrument using a quartz glass cuvette (Hellma) in aggregation buffer (50 mM sodium phosphate pH 7.0, 100 mM NaCl, 2 mM EDTA) at 37°C. The data were an average of at least 10 experiments of 10 s each. Data was processed using the software OmniSize 3.0. Mass distributions were derived using the in-built regularisation routine

which determines the least polydisperse distribution that gives minimum residuals to the correlation data.

**Pellet-supernatant analysis** Samples were centrifuged in a benchtop Biofuge Fresco (Heraeus) for 10 min at 13,000 g and 4°C. After aspirating the supernatant, the pellet was dissolved in lithium dodecyl sulfate SDS-PAGE loading buffer (Life Technologies). Equivalent amounts of both supernatant and pellet material were taken for SDS-PAGE.

**Ultracentrifugation** Centrifugation was performed sequentially, with the supernatant collected and centrifuged at increasingly higher speeds. The experiments were performed using an SW 60 Ti rotor in an Optima L-90K ultracentrifuge (Beckman Coulter) at 4°C, for 30 min at 20,000 g, 80,000 g and 160,000 g and for 4 hours at 320,000 g. After aspirating the supernatant, the pellet was dissolved in 6 M guanidinium hydrochloride at room temperature. Equivalent amounts of both supernatant and pellet material were taken for SDS-PAGE.

**Sucrose gradients** Sucrose solutions were prepared in aggregation buffer (50 mM sodium phosphate pH 7.0, 100 mM NaCl, 2 mM EDTA). Continuous gradients (10-30% w/v) were prepared using a Hoefer gradient mixer in Ultra-Clear 11x60 mm tubes (Beckman Coulter), total volume 4 ml. 100 $\mu$l samples were added to the top of the gradient. The tubes were centrifuged in an SW 60 Ti rotor in an Optima L-90K ultracentrifuge (Beckman Coulter) at 483,750 g for 1 hour and 20 minutes at 4°C. 150 $\mu$l fractions were taken with a pipette from the top of each tube.

**Size exclusion chromatography** Analysis was performed on an Akta Pure instrument (GE Healthcare) equipped with a Superose 6 10/300 column (GE Healthcare) at 4°C, equilibrated with various buffers as stated in the text. The flow rate was 0.5 ml/min. The injection volume of the sample was 200 $\mu$l.

**Ion mobility mass spectrometry** Where stated, buffer exchange was performed using Micro BioSpin columns (BioRad) as specified by the manufacturer. Samples were ionised by nanoelectrospray from gold-coated glass capillaries made in-house. For mixtures with c$\alpha$BC, spectra were acquired on a quadrupole-ToF instrument (Waters) modified for high mass transmission (Kondrat et al. [2015]) with the following parameters: source pressure $7\times10^{-4}$ mbar (labelled low) or $3\times10^{-3}$ mbar (labelled high), capillary 1.5 kV, cone 40 V, collision gas pressure 0.15 MPa, collision gas Ar, collision energy 30 V (labelled low) or 200 V (labelled high), ToF pressure $1\times10^{-6}$ mbar. For mixtures with $\alpha$BC, spectra were acquired on a Synapt G1 instrument (Waters). The conditions were as follows, unless stated otherwise: source pressure 5.9 mbar, capillary 1.5 kV, cone 50 V, extractor 3 V, trap gas Ar, trap flow 7.0 ml/min, trap pressure $3.5\times10^{-2}$ mbar, trap collision energy 8 V, trap wave height 0.5 V, trap wave velocity 300 m/s, IM cell gas $N_2$, IM cell gas flow 22 ml/min, IM cell pressure $3.1\times10^{-3}$ mbar, IM cell wave height 8 V, IM cell wave velocity 200 m/s; transfer wave height 0.2 m/s, transfer wave velocity 248 m/s, ToF pressure $1.7\times10^{-6}$ mbar.

## 6.3 Results and discussion

### 6.3.1 Overall sizes and shapes of particles

I combined two methods to look at the sizes and shapes of aggregates in mixtures of $\alpha$Lac with either c$\alpha$BC or $\alpha$BC: dynamic light scattering (DLS) and negative stain electron microscopy (EM). They are both low-resolution, rapid, equilibrium methods that involve no prior separation by mass transport. DLS gives size distributions of all species present in solution, although its dynamic range may be limited: if the scattering from large species is very strong, the signal from smaller species may not be detected. For EM, particles must be over about 200 kDa in mass to be readily detectable and unlike DLS this method allows visualisation of aggregate shape.

Analysis of many particles is required to obtain distributions of sizes from EM. Here I used it as a qualitative method to give an overall estimate of the shapes and sizes of the particles present.

**Dynamic light scattering** In DLS, intensity of laser light scattered by a solution at a given angle is measured at a number of time points (**Figure 6.1**). The correlation of this intensity over time is used to derive a size distribution. For a monodisperse sample, the decay of the correlation function is a single exponential. For a sample with particles of several sizes, algorithms such as non-negative least squares or Tikhonov regularisation can be used to find a size distribution with minimum polydispersity required to minimise residuals to the autocorrelation curve (Provencher [1982], Scotti et al. [2015]).

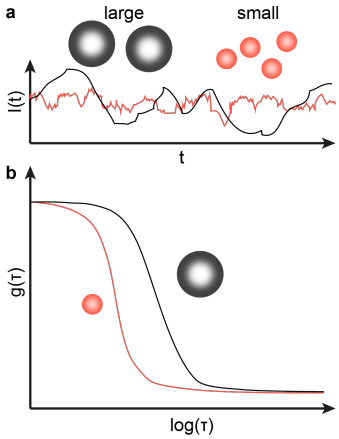



***Figure 6.1: The principle of the DLS experiment.*** *Scattered laser light intensity is measured at one or more angles over time t (**a**). Fluctuations due to diffusion of the particles are observed. Smaller particles diffuse faster and their fluctuations contain more high-frequency components. The fluctuations can be quantified with an autocorrelation function $g(\tau)$ (**b**), which shows the probability of correlation between particle positions at time t and $t+\tau$.*

For spherical particles in the Rayleigh limit, i.e. if the particle size is significantly smaller than the wavelength of scattered light, scattering intensity is proportional to $R_h^6$, where $R_h$ is the hydrodynamic radius of the particles. The intensity distribution is therefore heavily biased to large species. An example intensity distribution is shown in blue in **Figure 6.2** for a mixture of aggregates formed from 100 $\mu$M $\alpha$Lac with 200 $\mu$M c$\alpha$BC. There are two populations of particles, with the larger set of particles scattering much more light. Assuming the particles are indeed, in the simplest case, spherical, scaling the data by $\frac{1}{R_h^3}$ gives a quantity that is proportional to $\frac{R_h^6}{R_h^3} = R_h^3$, i.e. to the volume of the particles, and therefore, assuming constant density, to their mass. **Figure 6.2** shows the mass distribution of the same

sample of aggregates in black. It is equivalent to protein monomer concentration in particles of given $R_h$. Unlike in the intensity distribution, it is clear that there is much more material in the population of particles with smaller $R_h$ than in the population with the larger $R_h$.

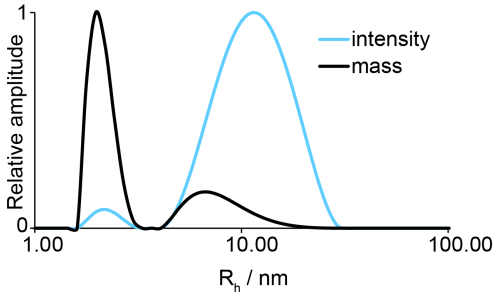



***Figure 6.2: Examples of the intensity and mass distributions**. The distributions are shown for a sample of aggregates formed after 22 hours in the presence of 100 μM αLac and 200 μM cαBC.*

The data below is presented in the more intuitive form of mass distributions. I recorded DLS data for each component alone and for mixtures with and without DTT at a range of timepoints. The data is shown in **Figure 6.3**. Panel **a** shows the mass distributions for the individual protein controls. αLac and cαBC appear as peaks at an $R_h$ of about 2 nm, corresponding to monomer (αLac) and/or dimer (for cαBC) whereas the distribution of full-length, oligomeric αBC is centered around 8 nm. These are consistent with expected sizes and previous measurements (e.g. Haley et al. [1998], Bumagina et al. [2010]).

**Figure 6.3b** shows the distribution obtained from a mixture of αLac and cαBC at 200 μM each incubated for 22 hours in the presence of DTT, together with the relevant controls. Two distinct populations are observed for an aggregation reaction in the presence of DTT (purple curve): around an $R_h$ of 2 nm and 10-20 nm. The peak at smaller $R_h$ is consistent with αLac monomers and/or unbound cαBC, as is evident from the overlapping controls. It is unusually narrow, which is likely to be an artifact of the regularisation routine in this particular experiment, since the narrow peak width was not reproduced in other experiments in this series (see below). The existence of unbound chaperone is also in agreement with the previous NMR analysis: HSQC and diffusion data on an equimolar mixture of αLac and cαBC showed that there is a significant amount of unbound chaperone (**Figures**

**5.3-5.4** in **Chapter 5**). Conversely, residual free $\alpha$Lac is expected to contribute little under these conditions - **Figure 5.14** shows that the predicted amount of residual $\alpha$Lac monomer is very low at this time even at a higher concentration of chaperone (500 $\mu$M).

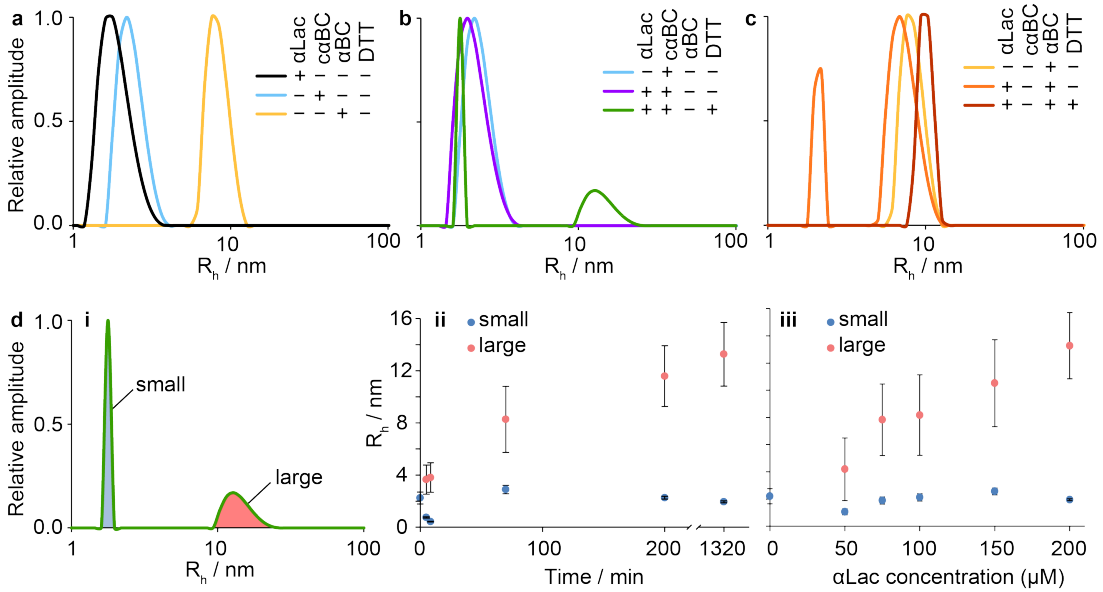



***Figure 6.3: DLS measurements on aggregating samples of $\alpha$Lac with and without $\alpha$BC and c$\alpha$BC.*** *All proteins were at 200 $\mu$M.* ***a*** *Single protein controls.* ***b*** *$\alpha$Lac with c$\alpha$BC.* ***c*** *$\alpha$Lac with $\alpha$BC.* ***d i*** *Illustration of the two populations of species with small and large $R_h$ in the mixture of $\alpha$Lac and c$\alpha$BC.* ***ii, iii*** *Variation of the mean size of the large and small species with time at 200 $\mu$M protein each (**ii**) and $\alpha$Lac concentration (**iii**) after 22 hours of aggregation. The error bars are the standard deviations of the data underlying the mass distribution peaks.*

The equivalent measurements for the mixture of $\alpha$Lac with full-length $\alpha$BC are shown in **Figure 6.3c**. For the mixture without DTT, two populations corresponding to those seen for the isolated components are observed. After prolonged incubation with DTT, however, the peak at low $R_h$, corresponding to any free $\alpha$Lac or free $\alpha$BC monomer or dimer, is not seen, and the larger population shifts to higher $R_h$. While the signal from small particles could be suppressed due to the limited dynamic range measurable with DLS, this is unlikely since the aggregates are not much larger than the $\alpha$BC oligomers observed in the control sample without DTT, and in this latter sample the signal from monomeric $\alpha$Lac is very strong (orange curve in **Figure 6.3**). On the contrary, the lack of any small species is consistent with the prediction that the amount of free $\alpha$Lac is negligible under these

conditions. This data also suggests that there is no significant amount of any free $\alpha$BC monomers being released from the oligomers, or formation of small complexes with $\alpha$Lac. The sizes of the oligomers formed are in agreement with the EM data described below.

**Figure 6.3d** shows an investigation of the behaviour of the distributions for $\alpha$Lac and the core domain chaperone c$\alpha$BC as a function of time and $\alpha$Lac concentration. The population of the larger more polydisperse oligomers (red) grows with both time and $\alpha$Lac concentration, but the population of small species, corresponding mostly to free c$\alpha$BC as argued above, remains at constant $R_h$ (blue). Thus, interestingly, the aggregates grow to form a distinct distribution of sizes which is not continuous with monomer except at the shortest times and lowest substrate concentrations.

**Electron microscopy** To assess the shape of the aggregates and confirm their average size estimated by DLS, I subjected the samples to negative stain electron microscopy analysis. In order to obtain useful data, it was necessary to optimise the grid preparation protocol to prevent artifacts from salt crystals, which can be mistaken for amorphous aggregates (see **Figure A.5** in the Appendix).

The electron micrographs of control samples are shown in **Figure 6.4a**. Soluble $\alpha$Lac and c$\alpha$BC are too small to be observed by EM, but full-length $\alpha$BC appears as roughly spherical particles of radius 8.7±1.7 nm, similar to what has been observed previously (Haley et al. [1998], Jehle et al. [2011]) and in agreement with the above DLS data.

**Figure 6.4b** shows images of aggregates with and without chaperone. In the absence of chaperone, $\alpha$Lac forms visible aggregates of two types (panels **b i** and **c**). The first type are large and flexible filaments, some of which look circular. Their thickness is 9.2±2.9 nm. They are more flexible than rods most frequently observed for amyloid fibrils (see e.g. **Figure 1.3** in **Chapter 1**) but very similar to

previously observed amyloid aggregates of human apolipoprotein CII (Hatters et al. [2000, 2003]). The second type are smaller roughly spherical particles with a radius of 4.7± 1.0 nm.

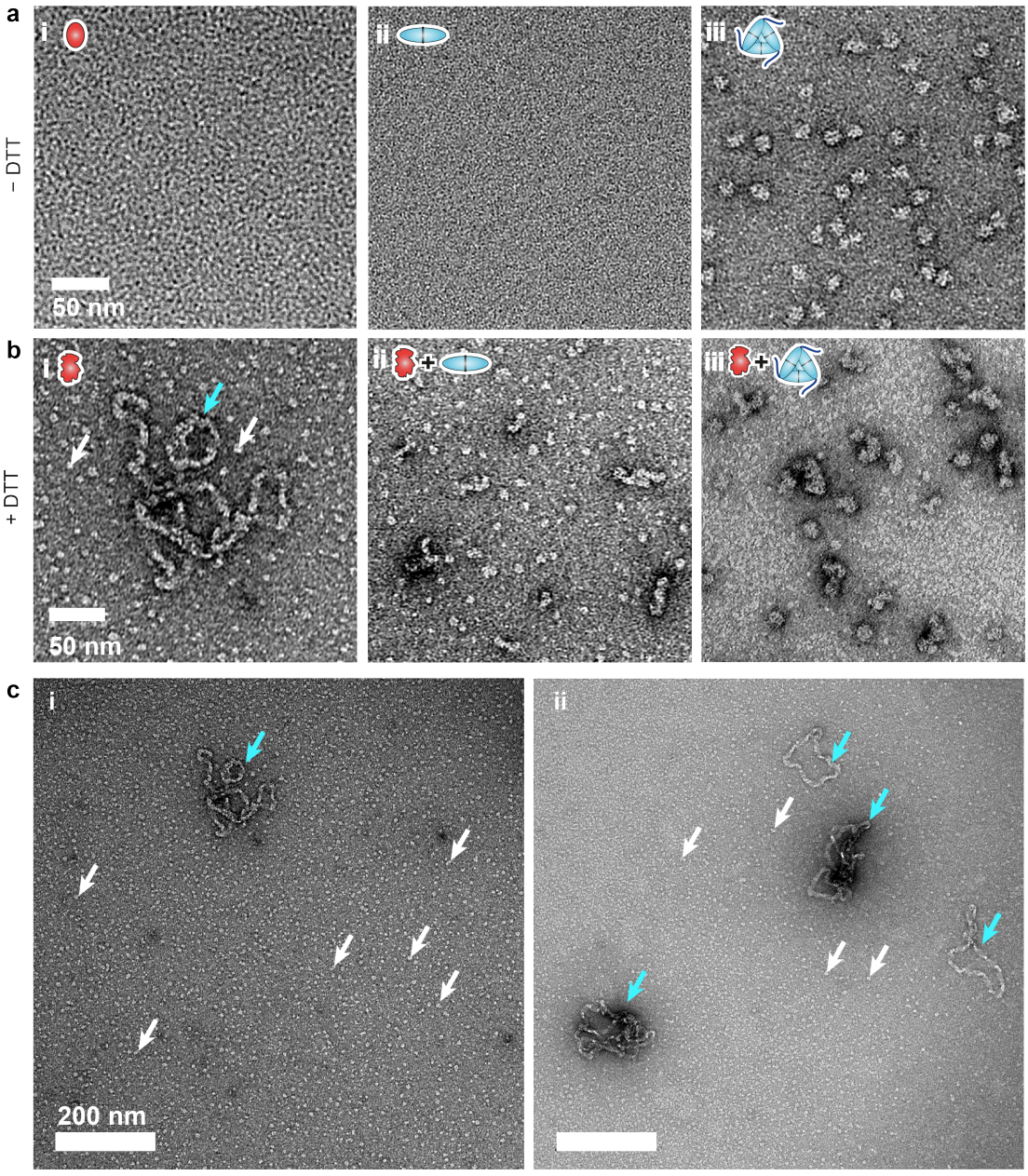



***Figure 6.4: Representative negative stain electron microscopy imaging of aggregates and chaperones.*** ***a*** *Control samples in the absence of DTT.* ***(i)*** *αLac,* ***(ii)*** *cαBC,* ***(iii)*** *αBC.* ***b*** *Samples 22 hours after the addition of DTT.* ***(i)*** *αLac alone,* ***(ii)*** *αLac and cαBC,* ***(iii)*** *αLac and αBC.* ***c*** *Lower magnification images of the sample in* ***b*** ***(ii)*** *to illustrate two morphologies of aggregate. Long filamentous aggregates are indicated by cyan arrows and small spherical aggregates by white arrows.*

In the presence of chaperones the aggregate morphology is different. With

c$\alpha$BC (**Figure 6.4b ii**), the long filaments are no longer observed but shorter elongated aggregates are seen, of thickness 7.2±3.0 nm. There are also smaller aggregates of radius 4.4±2.1 nm, rather similar to the small aggregates observed for $\alpha$Lac alone. With $\alpha$BC (**Figure 6.4b iii**), globular particles somewhat larger than $\alpha$BC alone are present (average radius 11.4±3.5 nm). Thus the assumption of sphericity for the analysis of the DLS data, while not strictly true, is not unreasonable for the mixtures with $\alpha$BC. Accordingly, the average aggregate sizes measured by the two methods agree well.

In summary, DLS and EM data together show that particles with radii on the order of several nanometres are formed in mixtures of $\alpha$Lac with either c$\alpha$BC or $\alpha$BC. Formation of long filamentous $\alpha$Lac aggregates is suppressed. The morphology of particles is different with the two chaperones: with c$\alpha$BC, they are often elongated and irregular; with $\alpha$BC, they are globular and larger than the chaperone alone. The size of these large aggregates increases with aggregation time and with total $\alpha$Lac concentration. With c$\alpha$BC, DLS clearly shows a second distinct population of small species, most likely free c$\alpha$BC, at all concentrations and times tested. In the case of mixtures with $\alpha$BC, no significant amount of small particles was detected with DLS.

### 6.3.2 Detection of complexes with low-resolution mass transport methods

Bulk methods like DLS and EM do not provide information on which protein is present in which type of particle. I therefore implemented several separation methods, aiming to observe correlations in behaviour of the substrate and the chaperone which would indicate formation of complexes between them.

**Benchtop centrifugation** During $\alpha$Lac aggregation with no or little chaperone, a visible white precipitate formed, corresponding to formation of very large aggregates.

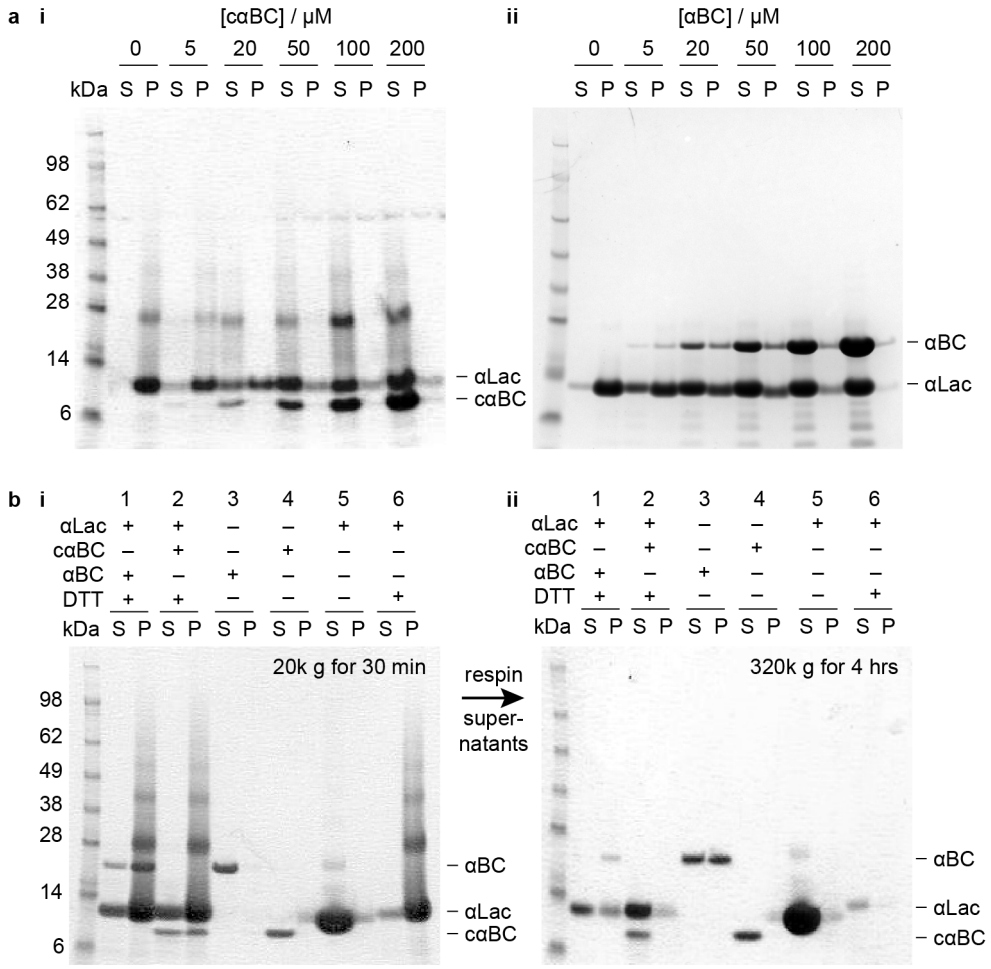



***Figure 6.5: Centrifugation analysis of steady-state aggregates of $\alpha$Lac in the presence of chaperones.*** *S - supernatant, P - pellet.* ***a*** *Pellet-supernatant analysis at low centrifugation speed (13,000 g for 10 minutes) in the presence of increasing concentrations of (**i**)c$\alpha$BC or (**ii**) $\alpha$BC. $\alpha$Lac was at 200 $\mu$M throughout.* ***b*** *Ultracentrifugation of $\alpha$Lac-chaperone mixtures and controls.* ***i*** *Centrifugation of the whole reaction mixture at 20,000 g for 30 minutes.* ***ii*** *Centrifugation of the supernatants from* ***i*** *at 320,000 g for 4 hours.*

To observe whether the chaperone formed part of these large aggregates, I performed a simple pellet-supernatant analysis. I incubated mixtures of 200 $\mu$M aggregating $\alpha$Lac with either c$\alpha$BC or $\alpha$BC at a range of concentrations for 12 hours at 37°C, and then subjected the samples to centrifugation at 13,000 g for 10 min at 4°C. $\alpha$BC, which has a size distribution between 200-800 kDa, sediments at accelerations over 70,000 g (e.g. Hatters et al. [2001]), so at 13,000 g only particles significantly larger than 800 kDa will form a pellet. SDS-PAGE was used to analyse the composition of the supernatant and the pellet. **Figure 6.5a** shows the material found in either fraction at each chaperone concentration. In the absence of chaperone, most of

$\alpha$Lac was found in the pellet. As the concentration of chaperone was increased, more $\alpha$Lac partitioned to the supernatant, consistent with the reduction in light scattering caused by the chaperone. At 100 and 200 $\mu$M chaperone no visible pellet was observed. c$\alpha$BC was seen only in the supernatant at any ratio of $\alpha$Lac to c$\alpha$BC (**Figure 6.5a i**). $\alpha$BC also partititioned mostly into the supernatant but there was also a clear band for it in the pellet lanes (**Figure 6.5a ii**). These results could imply that, unlike $\alpha$BC, c$\alpha$BC does not interact with megadalton-scale aggregates of $\alpha$Lac, or that the lifetime of this interaction is much shorter than the centrifugation transport time, as explained in the introduction to this Chapter. In contrast, the interactions of full-length $\alpha$BC with large $\alpha$Lac aggregates appear more long-lived. I discuss this reasoning further in the section on SEC below.

**Ultracentrifugation** To investigate the sizes of aggregates and complexes further, I performed more powerful ultracentrifugation of the mixtures and the proteins individually as controls. I used 2 mM $\alpha$Lac and 100 $\mu$M $\alpha$Lac for these experiments in order to saturate the chaperone with the substrate. **Figure 6.5b i** shows the partitioning into the supernatant and the pellet after centrifugation at 20,000 g for 30 minutes. Under these conditions, the soluble protein controls were found in the supernatant (lane pairs 3-5) while aggregated $\alpha$Lac was mostly in the pellet (lane pair 6). For the mixtures of $\alpha$Lac and chaperone, more $\alpha$Lac was found in the supernatant compared to $\alpha$Lac-only control (lanes 1 and 2 S compared to 6 S). As for the chaperones, they were found in both the supernatant and the pellet fraction, including c$\alpha$BC. Thus both form part of aggregates larger than 800 kDa, the maximum size $\alpha$BC alone.

I centrifuged the supernatant fractions of these samples sequentially at even higher speeds, 80,000 g for 30 minutes, 160,000 g for 30 minutes and 320,000 g for 4 hours. Only at the highest acceleration were changes in the distribution of the proteins observed (**Figure 6.5b ii**). Notably, under these conditions, the full-length $\alpha$BC oligomers started partitioning into the pellet (lane pair 3). Their mass range

is 200-800 kDa. We thus have a rough standard against which to compare the behaviour of the mixtures. The smaller soluble c$\alpha$BC and $\alpha$Lac were still found only in the supernatant (lane pairs 4, 5 and 6). (Note that no material was found in the pellet from the aggregated $\alpha$Lac control in lane 6 P as large aggregates were removed by the first centrifugation step in **b i**.) For the substrate-chaperone mixtures, lane pair 1 shows that all remaining $\alpha$BC with some $\alpha$Lac was found in the pellet, although most of $\alpha$Lac was still in the supernatant. c$\alpha$BC, on the other hand, remained in the supernatant with most of the $\alpha$Lac (lane pair 2). This suggests that $\alpha$BC forms part of aggregates that are several hundred kDa in mass, whereas c$\alpha$BC does not engage in a long-lived interaction with such aggregates. The presence of c$\alpha$BC in the supernatant at even the highest speeds is consistent with the DLS and NMR data which indicate significant amounts of unbound c$\alpha$BC in the aggregating mixtures (**Figures 6.3**, **5.3-5.4**). The high intensity of the $\alpha$Lac bands in the supernatants relative to the pellet also suggests that there is a significant population of $\alpha$Lac aggregates smaller than $\alpha$BC oligomers, especially for mixtures with c$\alpha$BC, which is corroborated by EM (**Figure 6.4**).

**Sucrose gradients** The supernatant-pellet experiments described above are binary, i.e. they categorise the components into large and small based on an arbitrary threshold which is difficult to define. To obtain more information on the distribution of the sizes of the complexes, I performed continuous sucrose gradient experiments. Sucrose gradients are used for e.g. virus capsid purification (Abraham and Colonno [1984], Marjomaki et al. [2014]). Given the large size of aggregates, this method in which the separating power is tunable by adjusting the density of the gradients has advantages over simpler SEC when a very limited arsenal of columns is available. The samples are also recoverable and following removal of the sucrose can be subjected to e.g. MS analysis.

For these experiments, mixtures of 200 $\mu$M chaperone with 100 $\mu$M $\alpha$Lac were allowed to aggregate overnight. This ratio was chosen to avoid very large complexes

which would not be amenable to MS. I also prepared control samples of each component individually, as well as a mixture of soluble protein standards: thyroglobulin (660 kDa), $\beta$-amylase (200 kDa) and lysozyme (14 kDa). The results of the experiment for the 10-30% sucrose gradient are shown in **Figure 6.6**. Panel **a** shows that this gradient range gave limited separation of the proteins. The three standards and the control samples are found in neighbouring fractions in **Figure 6.6a**. There is also considerable spread of the monodisperse proteins (**a i-iii**)over several fractions, comparable to the polydisperse full-length $\alpha$BC (**a iv**) which is observed over 6 fractions (fraction volume was 150 $\mu$l). Panels **b-c** shows the analysis of $\alpha$Lac-chaperone mixtures with and without DTT. The $\alpha$Lac bands are more intense one or two fractions further into the gradient in the samples with DTT, consistent with formation of larger species, but there is little change in the distributions of the chaperones.

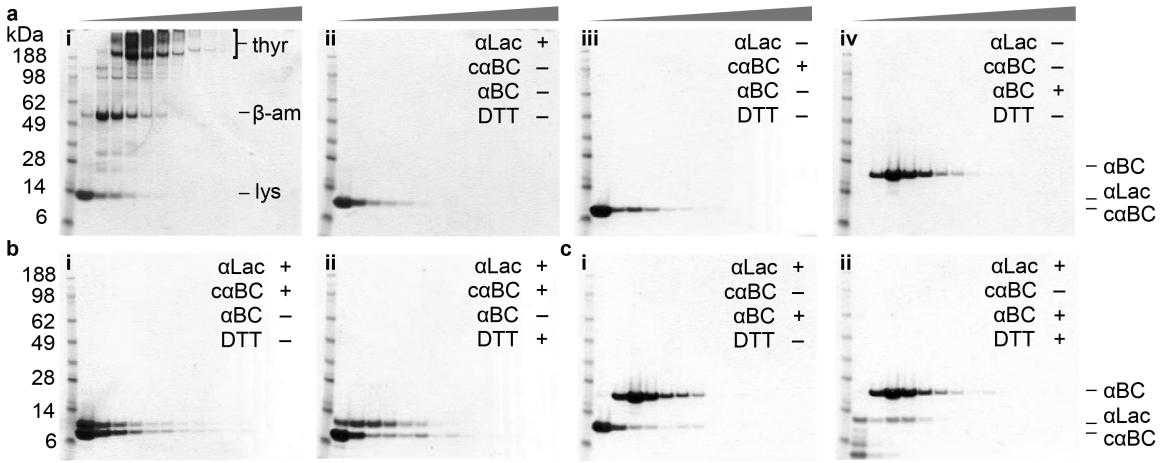



***Figure 6.6: Sucrose gradient analysis of chaperone-substrate interaction.*** *The gradient strength is from 10% to 30% sucrose from left to right in each gel (triangle).* ***a*** *Control samples.* ***i*** *Standards: thyr - thyroglobulin, $\beta$-am - $\beta$-amylase, lys - lysozyme.* ***ii-iv*** *Soluble $\alpha$Lac, c$\alpha$BC and $\alpha$BC centrifuged individually.* ***b-c*** *Mixtures of $\alpha$Lac with c$\alpha$BC or $\alpha$BC respectively, in the absence (**i**) and presence (**ii**) of DTT.*

Thus the experiment did not resolve complexes of $\alpha$Lac with either chaperone. Significant further optimisation of the procedure is clearly needed which could include preparing longer, shallower gradients. Longer gradients could also help narrow the peaks if their broadness is due to mixing of the sample during fractionation. In any case, the additional purification steps needed to remove sucrose would make the

yield of any complexes isolated on the gradient low. For these reasons, I did not pursue further optimisation of this laborious approach and focused on SEC with a Superose 6 column (fractionation range 1 MDa-1 kDa) to isolate complexes between $\alpha$Lac and $\alpha$BC chaperones.

**Size exclusion chromatography** I performed SEC on mixtures of $\alpha$Lac with $\alpha$BC and c$\alpha$BC after 12 hours of aggregation, together with single protein controls. The reaction was carried out at 37°C and SEC was done at 4°C in the same buffer to slow down dissociation kinetics and prevent further aggregation. The results are shown in **Figure 6.7**. As seen in panel **a**, $\alpha$Lac in the absence of DTT elutes between 18-19 ml. Full-length $\alpha$BC elutes between 12-15 ml. For the two mixtures of $\alpha$BC with $\alpha$Lac at different ratios, two features were observed. The late-eluting peak (17-20 ml) contained only $\alpha$Lac as confirmed by SDS-PAGE. The early-eluting peak, with a substantial fraction in the void (around 8 ml) for the 1:2 $\alpha$BC:$\alpha$Lac ratio, contained both $\alpha$BC and $\alpha$Lac. Thus $\alpha$BC forms part of aggregates with $\alpha$Lac that are larger than $\alpha$BC oligomers alone. The absence of $\alpha$BC at elution volumes longer than those characteristic for its intact oligomers suggests that dissociation of $\alpha$BC to bind $\alpha$Lac monomer does not occur, although dissociation and reassociation into the larger aggregates cannot be ruled out.

**Figure 6.7b** shows the equivalent data for c$\alpha$BC. c$\alpha$BC elutes around 17-18 ml, close to monomeric $\alpha$Lac, as expected from their similar molecular mass (14 kDa for $\alpha$Lac and 20 kDa for c$\alpha$BC dimer). At 1:0.5 and 1:2 ratios of chaperone to substrate, a range of larger species is formed as seen from broad peaks at elution volume 8-16 ml. However, SDS-PAGE showed that these fractions contained predominantly $\alpha$Lac. c$\alpha$BC eluted as a second peak at its control elution volume, together with some $\alpha$Lac. Thus no significant extent of complex formation between $\alpha$Lac and c$\alpha$BC was detected with SEC. Additional experiments showed that this behaviour was independent of the time that the mixture was allowed to aggregate (between 3 and 12 hours).

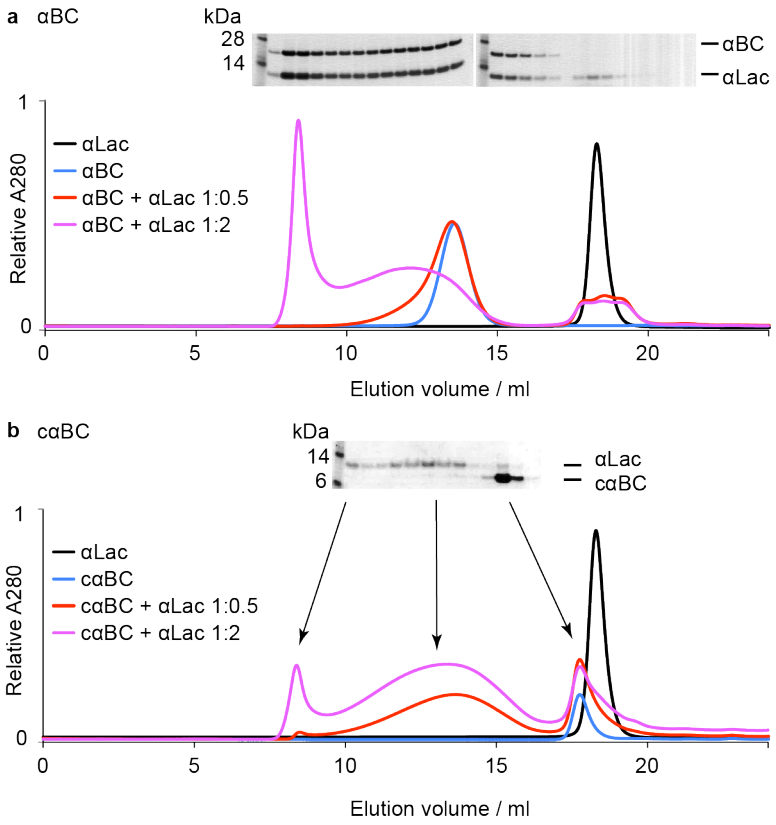



***Figure 6.7: Size exclusion traces in aggregation buffer for both chaperones.*** *Chaperone concentration 200 μM. αLac concentration in the control sample 100 μM. In b, the αLac trace has been scaled to 50% height.*

Thus the existence of complexes with $\alpha$Lac could be demonstrated for full-length $\alpha$BC but not for the isolated core domain c$\alpha$BC. Similarly to centrifugation, this observation does not exclude complex formation between $\alpha$Lac and c$\alpha$BC. Competing interactions with the stationary phase of the column cannot be ruled out, although they are unlikely. A more likely culprit could be the tradeoff between the rate of mass transfer and exchange rates of c$\alpha$BC with the aggregates. Looking at the rates of chaperone-aggregate interactions determined by chemical kinetics analysis in **Chapter 5** (**Figure 5.13c-d**), the $K_D$ for interactions with the major type of $\alpha$Lac aggregates (yellow curves) is similar for c$\alpha$BC and $\alpha$BC, but the on- and off-rates are an order of magnitude slower for the core domain. The effect of kinetic control on mass transport is complex and best assessed using a simulation. Stevens [1989] has carried out simulations of a reversibly binding system separated by SEC, at a constant $K_D$ and a range of rate constant magnitudes. His results show that

a system that was initially at equilibrium tends to appear as distinct components upon separation at faster on- and off-rates, which does not align with the behaviour observed here. Thus interference between kinetics of the interaction and mass transport also does not explain why a significant amount of complexes of $\alpha$Lac and c$\alpha$BC is not observed by SEC.

**Figure 5.14** from mechanistic analysis of the interactions shows that in fact in this $\alpha$Lac and chaperone concentration range and time most of the chaperone is unbound. Therefore complexes with c$\alpha$BC could be difficult to detect simply because their population is small. Complexes with full-length $\alpha$BC are similarly low in abundance. However, they can be detected more readily due to oligomeric nature of $\alpha$BC: a single interaction between an $\alpha$Lac molecule in an aggregate and an $\alpha$BC monomer is potentially enough to "pull down" an entire $\alpha$BC oligomer. Thus information on thermodynamic partitioning of the chaperone derived from the 1D NMR analysis provides an explanation of these puzzling SEC results.

### 6.3.3 Characterisation of complexes with native MS

Native MS is the state-of-the-art technique for looking at polydisperse mixtures. It is a fast experiment, with the time between ionisation and detection on the timescale of milliseconds. Thus MS has greater potential to capture short-lived interactions. Its resolution is sufficient to potentially assign individual stoichiometries of complexes (Stengel et al. [2010]) and in combination with ion mobility it can provide structural restraints (Shepherd et al. [2015], see also the work on $\alpha$BC in **Chapter 3**). Therefore I attempted native MS analysis of mixtures of $\alpha$Lac and $\alpha$BC, without prior separation by SEC, and those of $\alpha$Lac and $\alpha$BC with SEC.

#### c$\alpha$BC and $\alpha$Lac

For the study of complexes between these two proteins by native MS, I performed aggregation reactions in phosphate buffer, and exchanged the samples into ammonium

acetate just before MS analysis by either dialysis or a microcentrifuge desalting column. To begin with, I analysed the spectra of control samples of the single proteins and of the mixture without DTT. This data is shown in **Figure 6.8**. Small and large species require different instrumental conditions for optimal transmission; the source pressure is particularly important with higher values being more favourable for larger species due to increased collisional cooling and focusing (Chernushevich and Thomson [2004], Hernandez and Robinson [2007], Landreh et al. [2015], see also **Section 1.5.3**). Therefore I recorded data for all samples under both low (0.7 $\mu$bar) and high (3 $\mu$bar) source pressure.

The spectra in the absence of DTT are shown in **Figure 6.8**. $\alpha$Lac was mostly monomeric with traces of dimer and higher order oligomers visible at higher source pressure (note the much lower signal-to-noise ratio in the latter spectrum, indicating the low abundance of these species). c$\alpha$BC showed both monomer and dimer peaks and minor amounts of oligomers up to octamer at higher source pressure. To confirm that these were not non-specific artifacts of the electrospray process (Benesch et al. [2007]), I checked that the relative intensity of the peaks did not change with solution concentration of the protein (20-60 $\mu$M range), which was the case. The spectra of the mixture of $\alpha$Lac and c$\alpha$BC could be explained by combination of the above single-component spectra.

For aggregate mixtures incubated in the presence of DTT, the buffer exchange into ammonium acetate proved problematic. With both microcentrifuge chromatography columns (Biospin) and dialysis, SDS-PAGE analysis showed that $\alpha$Lac precipitated and/or was retained on the column or in the cassette (**Figure 6.9a**). I acquired spectra of the remaining material following Biospin buffer exchange at several timepoints after aggregation (**Figure 6.9b-d**).

**Figure 6.9b** shows the spectrum at low source pressure, which did not change with time. Peaks corresponding to c$\alpha$BC and $\alpha$Lac were observed, but no peaks corresponding to the two species bound to each other were detectable. Panel **c**

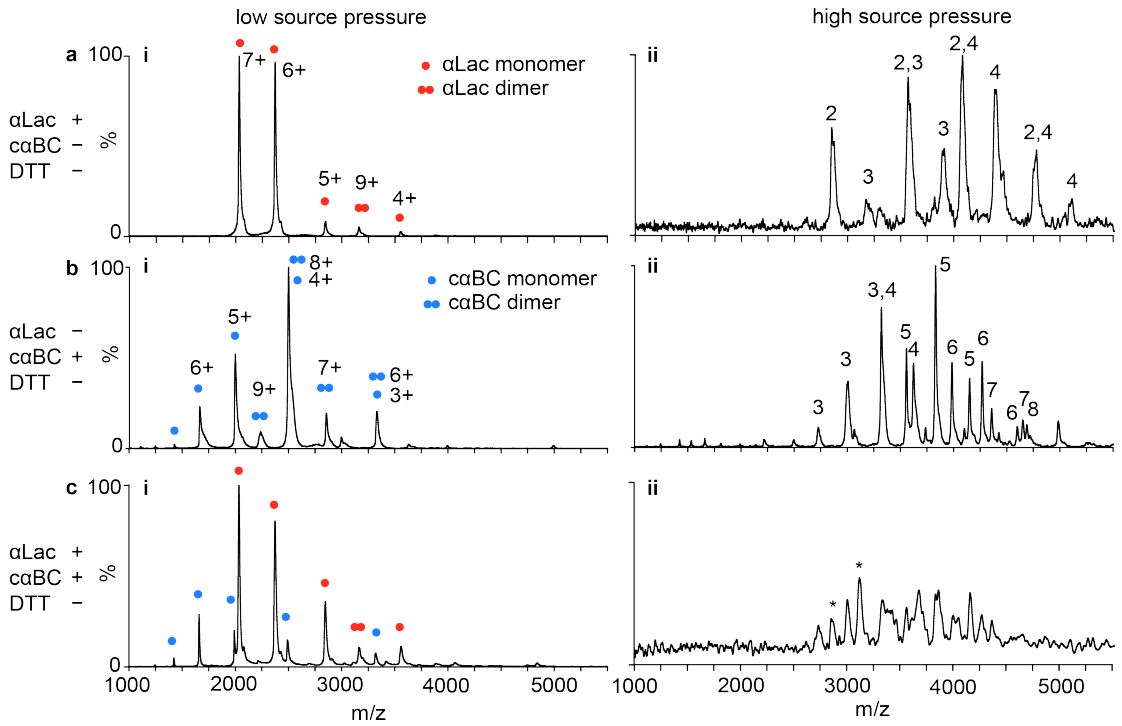



***Figure 6.8: Native mass spectra of control samples.*** *(**a**) αLac alone, (**b**) cαBC alone and (**c**) both proteins together in the absence of DTT at (**i**) low (0.7 μbar) and (**ii**) high (3 μbar) source pressure. Stoichiometry and charge state assignment is indicated for each peak in (**i**) whereas only stoichiometry is indicated in (**ii**) as a number for clarity. The peaks in (**c ii**) that cannot be matched to either control spectrum above are marked with an asterisk.*

shows the spectra at high source pressure. The pattern of overlapping peaks is characteristic of polydisperse oligomers and is different to the control without DTT (**Figure 6.8c ii**). Moreover, the distribution shifted to higher m/z over time, indicating formation of larger aggregates. This is consistent with previously obtained DLS data (**Figure 6.3**). I sought to show that cαBC was part of these aggregates. Tandem MS analysis - the standard way to accomplish this - was not possible due to low signal-to-noise ratio of the parent spectrum. Therefore I performed CID without isolation of a parent population, but setting the transmission profile such that only species above 5000 m/z were transmitted. The low m/z region of the resulting product spectrum is shown in **Figure 6.9d**. Both cαBC and αLac are observed as dissociation products, with higher charge states than in the low collision energy spectrum in panel **b**.

Thus native MS analysis shows that large oligomers consisting of both αLac

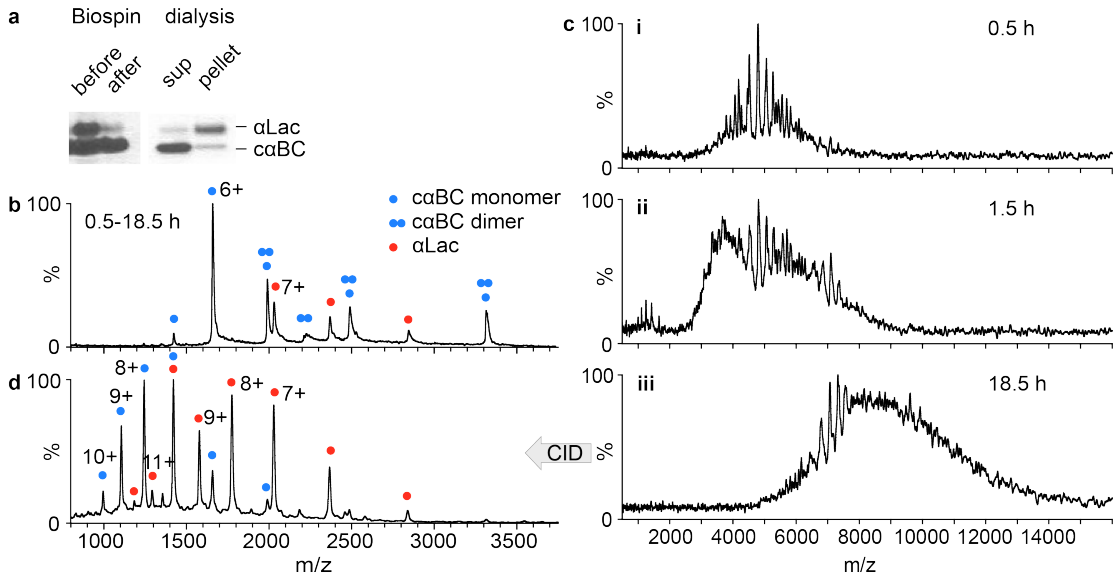



***Figure 6.9: Native MS of aggregating mixtures of αLac and cαBC. a*** *Gels to show loss of material through buffer exchange into ammonium acetate.* ***b*** *A low source pressure spectrum, which did not change significantly between 0.5 hrs and 18.5 hrs after the addition of DTT.* ***c*** *Spectra of the mixture at high source pressure at* ***(i)*** *0.5 hrs,* ***(ii)*** *1.5 hrs and* ***(iii)*** *18.5 hrs.* ***d*** *The low m/z region of following collision-induced dissociation of the oligomers in* ***c iii****.*

and cαBC are formed on aggregation, and that some free αLac monomer and cαBC persist in solution throughout. The aggregates become larger over time. The data is consistent with the shape and behaviour of the DLS distributions. Moreover, CID confirmed that the oligomers contained both proteins. Nevertheless, significant loss of material upon buffer exchange and the low signal-to-noise ratio of the oligomer spectra meant that further analysis and any assignment of stoichiometries was not practical for the mixtures with cαBC.

### αBC and αLac

This work was performed in collaboration with Dr Frances Kondrat[1]. Since it was possible to isolate complexes between αLac and full-length αBC using SEC in phosphate buffer (**Figure 6.7** above), we focused on analysing these samples rather than the unseparated mixtures. I allowed αLac to aggregate in the presence of αBC in

[1]The data in **Figures 6.10-6.12** and **6.14** was obtained in the course of joint work with Dr Frances Kondrat, to whom I am grateful for her engagement and insight.

phosphate buffer and carried out SEC of the mixture into ammonium acetate, thus performing the required buffer exchange and separating the fractions containing the complexes from those containing only $\alpha$BC or only $\alpha$Lac.

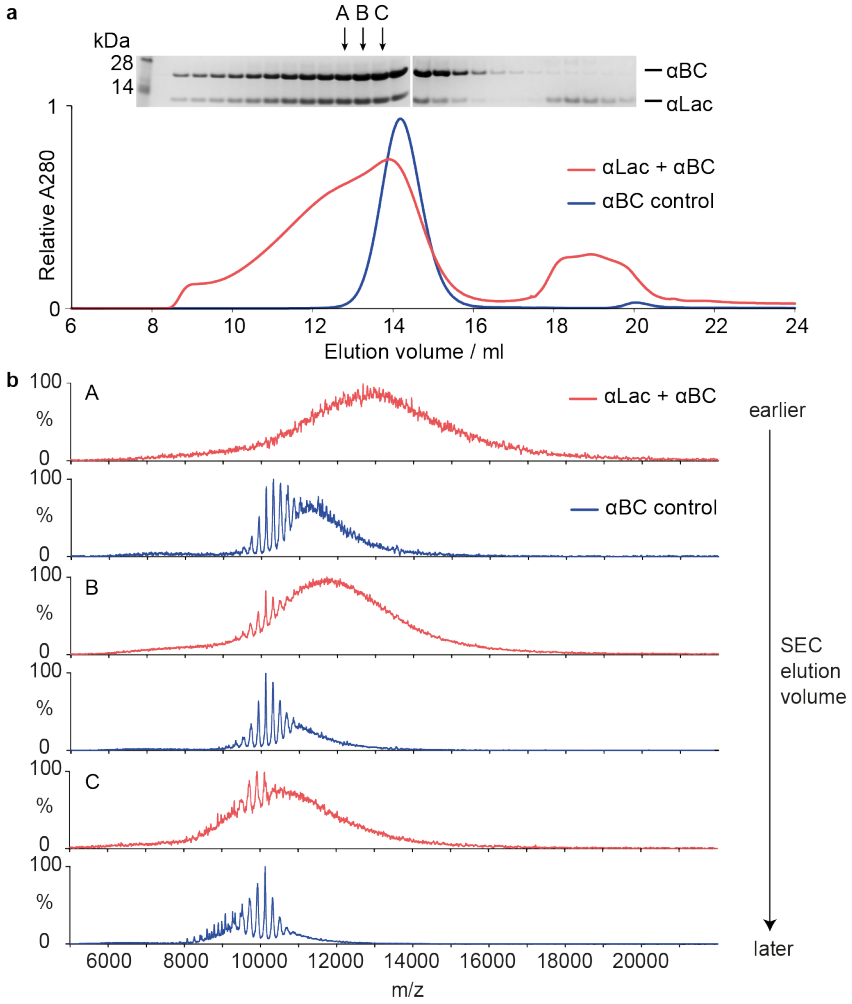



***Figure 6.10: Analysis of the SEC profile of a mixture of $\alpha$Lac and $\alpha$BC.*** ***a*** *Chromatography traces for the mixture of the two proteins (red) and $\alpha$BC only (blue). SDS-PAGE analysis of the mixture sample is shown above. A, B and C indicate the fractions taken for native MS analysis.* ***b*** *Native mass spectra of the fractions A, B and C from panel* ***a*** *(red), with the corresponding fractions from the $\alpha$BC-only control run (blue).*

An example chromatography trace is shown in **Figure 6.10a** for a mixture of 200 $\mu$M each of $\alpha$Lac and $\alpha$BC. SDS-PAGE confirmed that the behaviour of the mixture in ammonium acetate is similar to SEC in phosphate buffer (**Figure 6.7a**), with the broad peak at early elution volumes containing both proteins. We analysed fractions just upstream of the $\alpha$BC control peak, since they were expected to be enriched in $\alpha$Lac compared to those that eluted concurrently with the control peak,

and would be more amenable to CID analysis than very large complexes eluting earlier. They were analysed together with their counterparts from the $\alpha$BC-only control. The spectra of the fractions A, B and C are shown in **Figure 6.10b**. The spectra show broad peaks corresponding to many overlapping charge distributions of large and polydisperse oligomers. There is a clear decrease in average m/z of the aggregates at later elution times, and the spectra are broader and less resolved than those of $\alpha$BC, indicating greater polydispersity of the aggregates in the mixture.

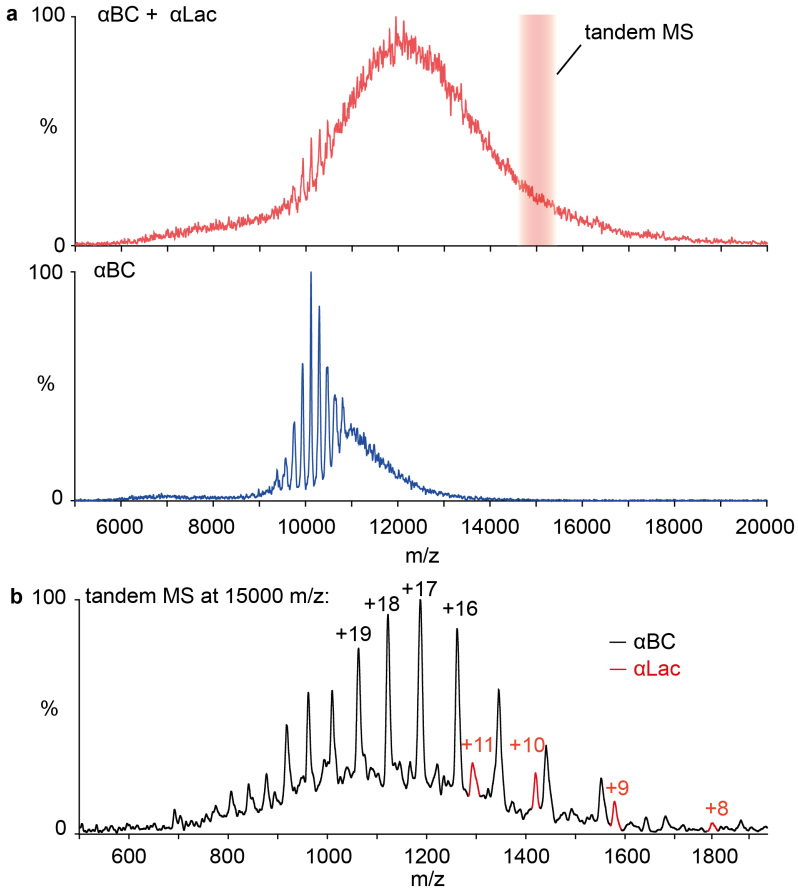



***Figure 6.11: Identifying components of large aggregates.*** ***a*** *Parent spectrum of the mixture at low collision voltage, indicating the region that was isolated for CID. The counterpart control $\alpha$BC spectrum has no signal in this region (blue).* ***b*** *The low m/z region of the product CID spectrum, with charge state assignment for $\alpha$BC and $\alpha$Lac.*

To confirm that $\alpha$BC is present in the large complexes together with $\alpha$Lac, we performed tandem MS. We chose a SEC fraction in which significant signal was observed in a high m/z area in the mixture spectrum but not in the $\alpha$BC-alone spectrum (**Figure 6.11a**). Any $\alpha$BC products from this region in the mixture must

therefore come from aggregates larger than free $\alpha$BC oligomers and the contribution from free $\alpha$BC can be reduced. The low m/z region of the product spectrum (**Figure 6.11c**) shows that both proteins were detected as dissociation products. The peaks of $\alpha$BC are much more intense than those of $\alpha$Lac. Since an $\alpha$BC monomer is larger than $\alpha$Lac (20 and 14 kDa respectively), in principle $\alpha$BC should be harder to dissociate than $\alpha$Lac. The reverse difference in intensity suggests that the stoichiometry of the complexes is biased towards $\alpha$BC, so that the probability of an $\alpha$BC dissociation event is higher than that of $\alpha$Lac (and possibly also that $\alpha$BC is on the outside of the complexes, making it easier to dissociate). Indeed, SDS-PAGE analysis across the large aggregate peak (**Figure 6.10a**) suggests that the complexes contain at least twice the amount of $\alpha$BC relative to $\alpha$Lac.

To obtain an estimate of the sizes of these complexes, we performed native MS in combination with ion mobility (IM). **Figure 6.12** shows the resulting two-dimensional IMMS spectra of one of the SEC fractions, with the m/z dimension as the x-axis and the ion mobility arrival time as the y-axis. Panel **a** shows the intact oligomers in low collision energy conditions. The $\alpha$BC control appears as the expected broad band of peaks, with species at higher m/z also appearing at longer arrival times due to their greater size and concomitantly higher charge. The mixed aggregate spectrum in panel **a ii** has a similar but more extended distribution. Its continuity and similarity to the control spectrum of $\alpha$BC alone suggests that the geometry of the $\alpha$Lac aggregates and complexes is not dramatically different from $\alpha$BC oligomers.

To interrogate the composition of the complexes, we subjected them to CID in the transfer cell. This cell is located after the ion mobility cell (**Figure 1.21**), therefore the dissociation products preserve the mobility signature of their parent ions (see also work in **Chapter 3**). **Figure 6.12b** shows the results of this experiment. For many protein complexes, dissociation is asymmetric, with highly charged unfolded monomers and corresponding charge-reduced oligomers as products (Benesch et al. [2007]). Each dissociation event results in further charge reduction,

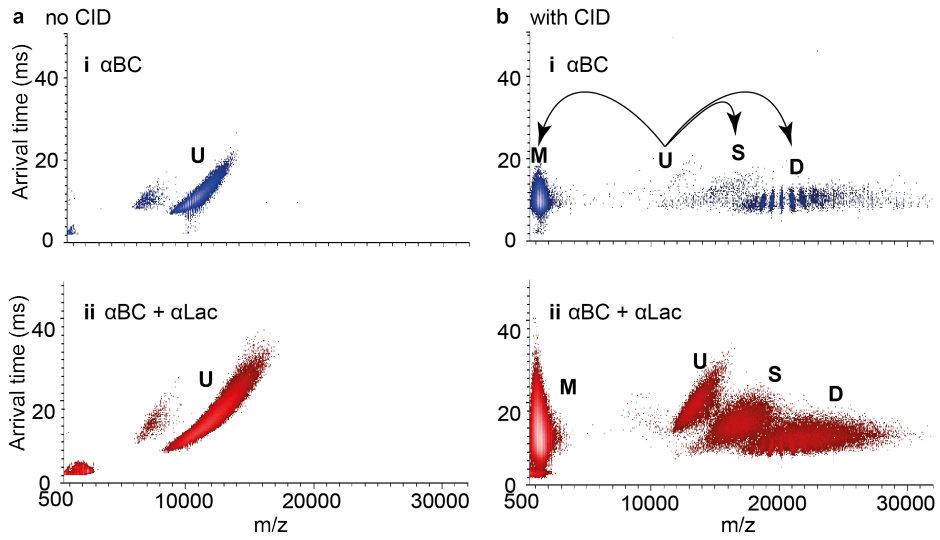



***Figure 6.12: Overview of the ion mobility analysis.*** ***a*** *IMMS spectra of equivalent SEC fractions at low transfer collision energy, i.e. in the absence of collision-induced dissociation (CID).* ***(i)*** *αBC alone,* ***(ii)*** *mixture of αBC and αLac.* ***b*** *Spectra of the same fractions at high transfer collision energy, leading to asymmetric CID.* ***(i)*** *αBC alone,* ***(ii)*** *mixture of αBC and αLac. U - undissociated oligomers, M - stripped monomers, S - singly stripped oligomers, D - doubly stripped oligomers.*

shifting the product oligomer population to higher m/z. αBC readily undergoes two dissociation events at 200 V in the transfer cell, and its predominant products are doubly stripped oligomers. For the mixture with αLac, where the size range is wider, this collision energy is not sufficient to fully dissociate the largest oligomers (marked U). For medium-sized oligomers, the energy is enough to lead to one dissociation (S), and the smallest oligomers that appear in the same range as the αBC control undergo two dissociation events (D).

**Figure 6.13** shows the analysis of the stripped monomer region of the CID spectra. These monomers could arise from an oligomer of any size that has undergone at least one dissociation event. Their arrival times therefore give a measure of the overall arrival time distribution of all oligomers in the spectrum. I extracted the slices of the 2D spectrum in **Figure 6.12b**, obtaining the arrival time distributions of each charge state of either protein monomer (**Figure A.6** in **Appendix A**), and combined them to obtain the average arrival time distribution of αBC in the absence of αLac, in the presence of αLac and of αLac in the mixture (**Figure 6.13**). The average arrival time of αBC increased in the mixture and the arrival time of αLac

correlated with that of $\alpha$BC. This further confirms that the proteins form complexes and that $\alpha$BC interacts with the whole population of $\alpha$Lac aggregates, as opposed to, for example, only the smaller aggregates.

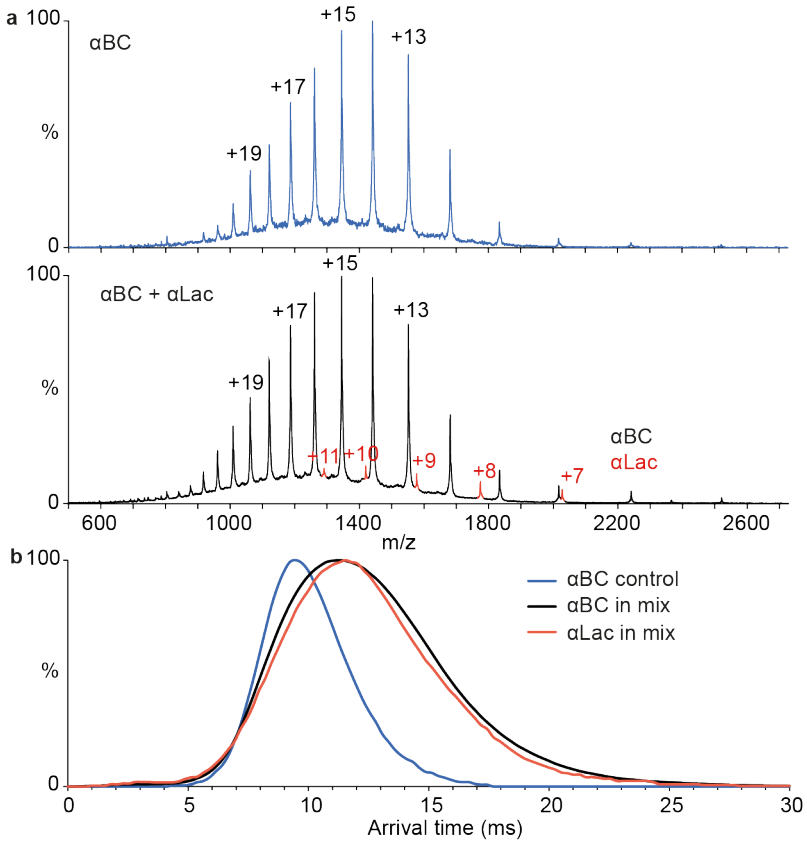



***Figure 6.13: Analysis of monomer CID product drift times.*** ***a*** *Expanded view of the stripped monomer (M) region of* ***Figure 6.12b*** *in the m/z dimension for the equivalent SEC fractions of $\alpha$BC alone and $\alpha$BC-$\alpha$Lac mixture samples.* ***b*** *Arrival time distributions of each protein averaged across all observed charge states.*

The monomer region gives no information on the stoichiometries of complexes, since monomers that come from oligomers of different size or charge cannot be distinguished. This information can in principle be obtained from the stripped oligomer region. Following charge reduction through dissociation of a highly charged monomer, peaks in this area become more widely spaced with charge series corresponding to individual stoichiometries becoming resolved. In **Chapter 3**, I exploited this to obtain collisional cross section (CCS) values for different oligomer stoichiometries of $\alpha$BC.

**Figure 6.14** shows the stripped oligomer area for an $\alpha$BC-only control and

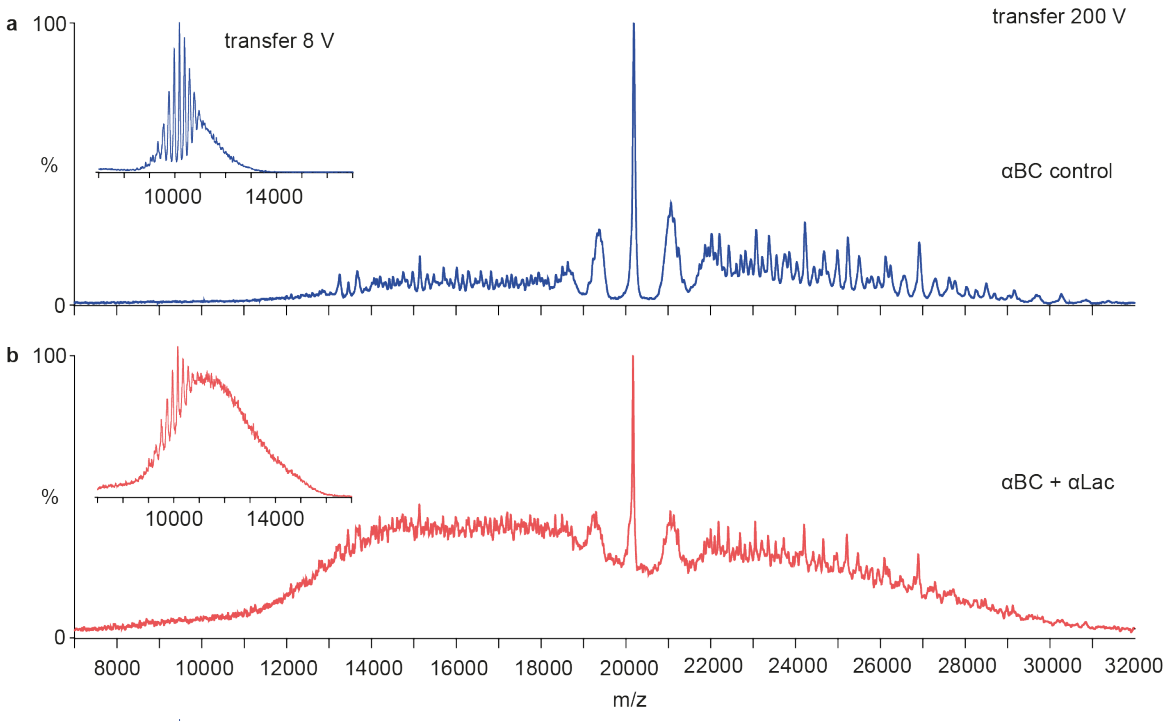



***Figure 6.14: The stripped oligomer regions of the CID spectra in Figure 6.13.*** *Insets show the undissociated parent spectra.*

the corresponding mixture spectrum, corresponding to the U, S and D areas in **Figure 6.12b**. Several overlapping series of peaks are visible, with the prominent peak around 20,000 m/z arising from all $\alpha$BC oligomers in which the charge is double the number of subunits. For $\alpha$BC, spectra of this type are assigned in **Chapter 3**. In **Figure 6.14b**, any peaks that are different from those in the control would correspond to species containing $\alpha$Lac. However, no such peaks are clearly observed. The baseline in this spectrum is high, especially at the lower m/z of around 14,000. These observations suggest that the larger oligomers containing $\alpha$Lac have not undergone sufficient dissociation to resolve them. The mobility spectrum in **Figure 6.12b ii** confirms that it is indeed the larger oligomers that remain unfragmented.

Since the collision energy was already at maximum, I attempted using a different gas, sulfur hexafluoride ($SF_6$), to enhance dissociation of the larger oligomers. $SF_6$ is larger than argon atoms and therefore results in more energetic collisions. Fragmentation was indeed improved (**Figure 6.15**). With argon as the dissocia-

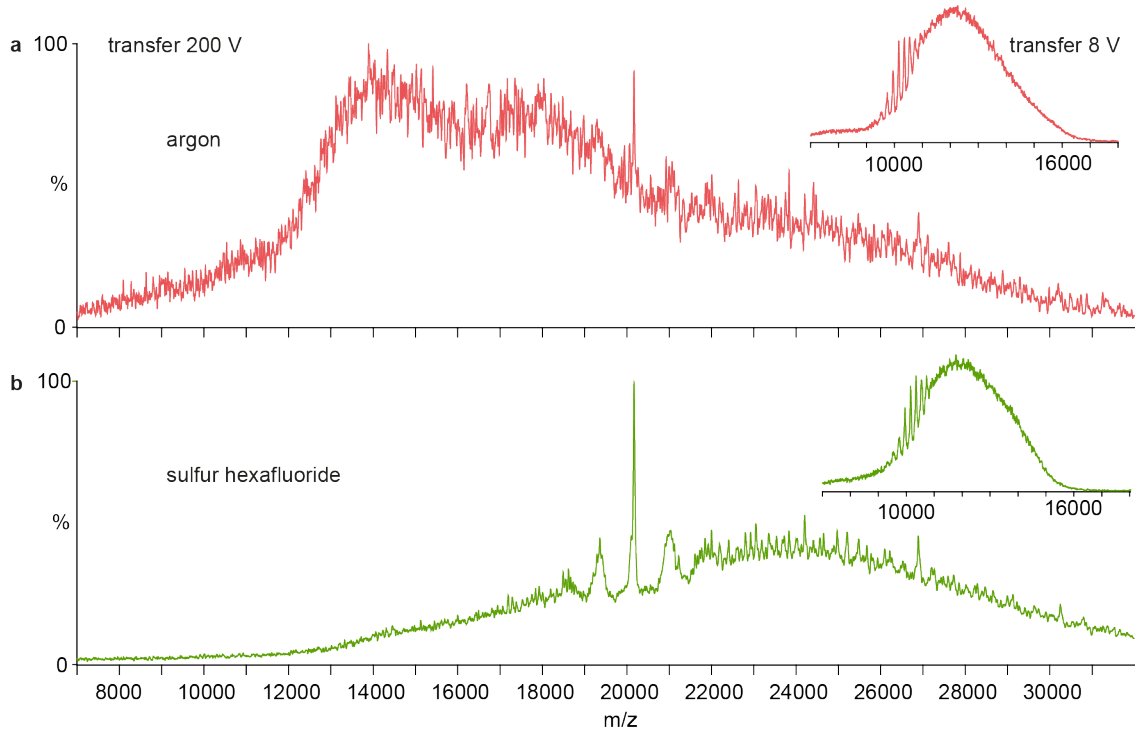



***Figure 6.15: Dissociation efficiency as a function of collision gas identity.*** ***a*** *The stripped oligomer region of a SEC fraction of a mixture of αBC and αLac, with argon as collision gas.* ***b*** *The same spectrum with $SF_6$ as collision gas. The insets show the parent spectra prior to CID.*

tion gas, the most prominent band is that corresponding to undissociated oligomers around 14,000 m/z, followed by the singly stripped oligomers around 17,000 m/z, With $SF_6$, these areas are low in intensity and the most prominent area is that of the doubly stripped oligomers around 24,000 m/z. Thus the use of $SF_6$ is promising for resolving complexes containing αLac. Further adjustment of other transmission conditions is required to optimise sensitivity with this gas.

Even with improved fragmentation efficiency, the stripped oligomer spectra still show rather poor resolution, making identification of peaks corresponding to specific stoichiometries difficult. To make the spectra more amenable to analysis, tandem MS should be performed, where only a narrow region of the parent spectrum is dissociated. In addition, generation of smaller complexes, e.g. by collection earlier in the aggregation reaction or later in the elution profile, should be explored. A third avenue for further work is the use of Orbitrap instruments with native MS capability which can provide much greater resolution than time-of-flight instruments

(Rose et al. [2012], Gault et al. [2016]). Nevertheless, due to the timescale mismatch between IM pulses and ion trap acquisition time a platform that combines the two for analysis of large protein complexes is still in development (see Ibrahim et al. [2016] for such analysis on hydrocarbons and peptides).

### Density of complexes

With optimisation of resolution in the stripped oligomer region, stoichiometries and sizes of the complexes will be defined in detail. Nevertheless, known scaling laws can be used to obtain insight into the properties of the complexes from the average arrival time measurements in the stripped monomer region. Arrival time is linked to particle volume via the CCS calibration, and the average charge depends on particle mass (Fernandez De La Mora [2000], Stengel et al. [2010]). If ratios of these properties are taken, many of the constants cancel, enabling a robust estimation of the density of the complexes relative to unbound $\alpha$BC oligomers.

We know the m/z range for $\alpha$BC and the complexes, and also their arrival time distributions. The simplest travelling wave CCS calibration relationship is $CCS = bzt_D^c$ (Ruotolo et al. [2008]), where $z$ and $t_D$ are protein charge and drift time and $b$ and $c$ are the calibration constants. The value of $c$ depends on the wave conditions but varies little, usually being between 0.3-0.5. The CCS ratio of two species with different average charge states and arrival times, here free $\alpha$BC oligomers and $\alpha$BC-$\alpha$Lac complexes, is therefore $\frac{CCS_1}{CCS_2} = \frac{z_1}{z_2}(\frac{t_{D1}}{t_{D2}})^c$. Here the m/z-dependent correction to the arrival time to convert it into a drift time has been omitted, since it is on the order of a few percent. The average charge of a particle can be approximated by the relationship $z_{avg} = pM^q$, with p=0.047 and p=0.5 (Stengel et al. [2010]). Using this relationship, the ratios of masses and average charges of the two types of species can be derived in terms of the observed ratio in m/z: $\frac{M_1}{M_2} = (\frac{M_1/z_1}{M_2/z_2})^{1/(1-q)}$ and $\frac{z_1}{z_2} = (\frac{M_1/z_1}{M_2/z_2})^{q/(1-q)}$.

Using the approximate values of $m/z = 10,000$ and $t_D = 9\,\text{ms}$ for $\alpha$BC and

$m/z$ = 13,000 and $t_D$ = 12 ms for the complexes (**Figures 6.12** and **6.13**), and a value of $c$ of 0.4, we obtain a mass ratio $\frac{M_{\alpha BC}}{M_{complexes}} = 0.59$ and an approximate CCS ratio $\frac{CCS_{\alpha BC}}{CCS_{complexes}} = 0.69$. Since for convex particles CCS is proportional to surface area (Vouk [1948]), particle volume scales as $CCS^{2/3}$. The ratio of volumes is therefore $\frac{V_{\alpha BC}}{V_{complexes}} = 0.56$. The ratio of masses and volumes gives a change in density on complex formation of 1.04. Thus the density of complexes is the same as the density of free $\alpha$BC oligomers. This result is robust: the density ratio is 1.0±0.1 if c is varied between 0.3-0.5, m/z between 9,000-13,000 for $\alpha$BC and between 11,000-17,000 for the complexes and $t_D$ between 6-15 ms for $\alpha$BC and 7-24 ms for the complexes.

The constant density rules out encapsulation of significant amounts of $\alpha$Lac within the $\alpha$BC scaffold. Incorporation of a single $\alpha$Lac inside the average $\alpha$BC oligomer, e.g. a 28-mer, with a change in mass but constant particle volume, would lead to a 3 % increase in density. With the approximate 2 $\alpha$BC : 1 $\alpha$Lac stoichiometry of interaction from SDS-PAGE, a density increase of 35 % would be expected, i.e. more than estimated here. Thus significant encapsulation is inconsistent both with the observed increase in size of the complexes relative to $\alpha$BC and with the finding that density of the assemblies remains constant. The interior of $\alpha$BC oligomers shows small, if any, gaps in electron density and is thought to accommodate close-packed N-termini (Baldwin et al. [2011c], Jehle et al. [2011]). The absence of density changes in the complexes therefore also rules out structures with large central cavities. In summary, the DLS, EM, SEC and IMMS data on the $\alpha$BC-$\alpha$Lac complexes supports formation of larger, globular, polydisperse oligomers with packing density similar to that of apo $\alpha$BC.

## 6.4 Conclusions

In this Chapter, I determined sizes and shapes of species in aggregating mixtures of $\alpha$Lac and $\alpha$BC or c$\alpha$BC chaperones by a variety of biophysical methods: DLS,

negative stain EM, centrifugation techniques, SEC, native MS and ion mobility. The techniques yield constistent and complementary results, and are in agreement with the chemical kinetics analysis in **Chapter 5**. I demonstrated that aggregating mixtures of $\alpha$Lac and c$\alpha$BC contain two types of species: the free chaperone, even at long aggregation times, and larger aggregates containing both proteins. The latter are spherical or elongated. For mixtures with $\alpha$BC, globular aggregates of size similar to and larger than free $\alpha$BC were detected, also containing both $\alpha$Lac and $\alpha$BC. Intact $\alpha$BC oligomers were also present in the mixture. However, no direct evidence for $\alpha$BC action via dissociation of dimers was obtained - DLS, EM and SEC in combination did not detect any complexes smaller than unbound oligomeric $\alpha$BC. Using IMMS, I demonstrated that the arrival time distributions of $\alpha$Lac and $\alpha$BC within the large complexes were unimodal, smooth and correlated, i.e. the chaperone-substrate complexes exist as a continuous polydisperse distribution of hetero-oligomers, larger than $\alpha$BC but with similar packing density.

# Chapter 7

# Residue-resolution analysis of chaperone function

*The aim of this Chapter is to identify elements of the amino acid sequence of $\alpha$BC that are important for its chaperone function towards the aggregating substrate $\alpha$Lac. In Chapters 4-6 I showed that $\alpha$BC acts by reversible binding to a variety of species on the aggregation pathway, and that a significant fraction of the complexes reaches several hundred kilodaltons in mass. I therefore employed NMR and MS experiments that probe for structural and dynamic changes in the chaperone at a range of timescales, with an awareness that the complexes it forms are large and heterogeneous. I demonstrate that the substrate has a dual impact on the regions controlling quaternary dynamics of $\alpha$BC: via the C-terminus and the core domain, but not through competitive binding in the $\beta$4-$\beta$8 groove.*

## 7.1 Introduction

This Chapter is the last in the series of four investigating the interaction of the sHSP αBC with the model aggregating substrate αLac. **Chapters 4** and **5** explore which types of aggregating species are bound by the chaperone and derive the kinetic and thermodynamic parameters of these interactions. **Chapter 6** deals with detection and characterisation of the predicted complexes between the two proteins. This Chapter concerns itself with the identification of primary structure regions of the chaperone that are either directly engaged in, or indirectly affected by, the interaction with αLac in these complexes.

αBC, like many sHSPs, is a promiscuous chaperone, both for many model aggregating proteins and many *in vivo* clients (Horwitz [1992], Kumar and Reddy [2009], Arrigo [2013], Fu [2014], Haslbeck and Vierling [2015]). This remarkable property is evidently useful for their function in maintaining proteostasis. At the same time it raises many interesting questions. Is there a defined binding site that exploits a common property of all substrate proteins, or does the chaperone engage different binding sites for different substrates? Moreover, in **Chapter 5** I showed that αBC can interact with different types of aggregating species, including monomers, nuclei and large aggregates. What binding determinants does the chaperone use for different forms of the same protein?

A reasonable hypothesis is that the chaperone recognises hydrophobic sites that become exposed upon misfolding (van Montfort et al. [2002], Haslbeck and Vierling [2015]). In accord with this, a correlation between binding hydrophobic probes such as bis(8-anilinonaphthalene-1-sulfonate) (bis-ANS) and chaperone activity has been observed for several sHSPs, in particular at higher temperatures (Lee et al. [1997], Yang et al. [1999], Haslbeck et al. [1999]). The correlation is not universal, for instance it holds for αA-crystallin but not αB-crystallin (Reddy et al. [2000]). Further evidence for hydrophobic interactions includes identification of the relatively hydrophobic N-terminus of sHSPs as a potential substrate binding site by

crosslinking, peptide mapping and deletion mutant studies (Ahrman et al. [2007], Aquilina and Watt [2007], Ghosh et al. [2006]). NMR studies have pointed to the N-terminus of $\alpha$BC as the binding site for lysozyme and the hydrophobic $\beta$4-$\beta$8 groove as the binding site for A$\beta_{1-40}$ (Mainz et al. [2015]).

Interestingly, sHSPs are commonly thought to interact with substrates that have lost a significant fraction of the native structure but are not completely unfolded, i.e. forms that have exposed some but not all hydrophobic sites. For example, the fluorescence properties of several substrates in the presence of $\alpha$-crystallin were much closer to the folded rather than the denatured control (Das et al. [1996]). Using hydrogen exchange mass spectrometry (HX-MS), malate dehydrogenase was shown to retain the solvent inaccessibility of its core in the presence of two plant sHSPs (Cheng et al. [2008]), and more recently similar results were obtained with two yeast sHSPs (Ungelenk et al. [2016]). The work of McHaourab and colleagues has shown that $\alpha$A- and $\alpha$B-crystallin bind monomeric, folded mutants of T4 lysozyme with an affinity that inversely correlates with their stability, and that two modes of binding can be observed for one mutant depending on solution conditions (Mchaourab et al. [2002], Koteiche and McHaourab [2003]). The authors hypothesised that the chaperone recognises the small population of unfolded state with exposed hydrophobic surfaces.

In addition to hydrophobicity, it is becoming apparent that other types of interactions also play a role for different substrates (Fu [2014], Treweek et al. [2015]). For example, reduced and carboxymethylated $\alpha$Lac is an unfolded polypeptide that has little exposed hydrophobicity and yet $\alpha$BC can still curtail its aggregation (Kulig and Ecroyd [2012]). The isolated core domain of $\alpha$BC shows significant chaperone activity (Feil et al. [2001], Hochberg et al. [2014] and the work in **Chapter 5** of this thesis), so the N-terminus is not essential for substrate binding. Even peptides corresponding to very short stretches of the core domains of $\alpha$A- and $\alpha$B-crystallin, containing both hydrophobic and charged residues, have been reported to be chaperone-active (Sharma et al. [2000], Bhattacharyya et al. [2006]). Deletion

of the polar C-terminus of $\alpha$BC abrogated its chaperone activity towards several substrates (Ghosh et al. [2006]), pointing to its direct or indirect importance. A plethora of mutations in $\alpha$BC are known to cause cataracts (e.g. R11H, R56W, R120G, D140N) and myopathies *in vivo* (e.g. R120G, Q151X, R157H) (Reddy and Reddy [2015]). It is not always clear, however, whether this is due to direct disruption of interactions with *in vivo* partners or an indirect effect through altered quaternary structure and dynamics.

In summary, the varied results obtained so far suggest that, for sHSPs, the location and type of interactions depend on the identity and form of the substrate, although hydrophobicity can play a strong role. In what follows, I investigate the interaction of $\alpha$Lac with the largest hydrophobic groove on $\alpha$BC with the help of mutagenesis. I go on to employ a range of 2D NMR experiments to study the effects of aggregating $\alpha$Lac on the C-terminus and the core domain, and use HX-MS to probe the whole protein including the N-terminus. Key properties of the system that must be remembered for interpretation of the results are quaternary dynamics of $\alpha$BC itself, reversible binding of the chaperone to the aggregates, their heterogeneity and large size.

## 7.2 Materials and methods

Please see **Section 4.2** for details on light scattering aggregation assays, **Section 3.2** for $\alpha$BC purification and **Section 2.2** on c$\alpha$BC purification. All reagents were from Sigma unless stated otherwise.

**Mutagenesis** Mutagenesis was performed using the Quikchange kit (Agilent) as specified by the manufacturer. The primers used were 5'-atcag atgac aggga ctgag taatg gtgag agggt ctaca tcagc tg-3' and 5'-cagct gatgt agacc ctctc accat tactc agtcc ctgtc atctg at-3' (S135Q) and 5'-atcag atgac aggga ccaag taatg gtgag agggt ctaca t-3' and 5'-atgta gaccc tctca ccatt acttg gtccc tgtca tctga t-3' (S135W). The constructs

were checked with Sanger sequencing. The mutants were purified as wild-type protein (see **Section 4.2**).

**Peptide binding experiments** Experiments were performed on Synapt G1 (Waters) in ToF mode. The samples were made up from concentrated stocks so that the final concentrations were 8 $\mu$M for c$\alpha$BC and between 4 and 80 $\mu$M for the peptide. The parameters were: source pressure 3.2 mbar, capillary 1.1 kV, cone 20 V, extractor 4 V, bias 4 V, trap 5 V, transfer 5 V. Wave parameters: trap velocity 300 $\mathrm{ms}^{-1}$, height 0.5 V, transfer velocity 248 $\mathrm{ms}^{-1}$, height 0.2 V.

**Hydrogen exchange mass spectrometry** Experiments were performed on the Waters HDX-MS system, including a Synapt G2 Si mass spectrometer coupled to an Agilent Acquity UPLC system and an HX-MS sample manipulation robot (Waters). Aggregation reactions were prepared at 200 $\mu$M of each protein as described in **Section 4.2**. Following an overnight incubation at 37°C, the samples were filtered and kept on ice. They were diluted tenfold and deuterated aggregation buffer was added to initiate proton-deuteron exchange. Separate samples were incubated at 4°C for 10 s, 2, 5, 30, 70 and 140 min. Each timepoint was acquired in triplicate. Quench buffer (aggregation buffer at pH 2.0) was added to stop the reaction. The samples were injected into the UPLC system in water, which included online pepsin digestion (Waters Enzymate BEH pepsin column) followed by peptide trapping (Waters C18 BEH column). The trapping time was 3 min at the flow rate of 200 $\mathrm{mlmin}^{-1}$. The peptides were eluted and separated with a water-acetonitrile gradient of 3-35% at a flow rate of 40 $\mu\mathrm{lmin}^{-1}$ over 7 min. MS data acquisition was performed in $\mathrm{MS}^e$ mode (Levin et al. [2011]) with leucine enkephalin as lock mass, with parameters as specified by the manufacturer. The low energy and elevated energy thresholds were 250 and 100 counts respectively. Data was processed using ProteinLynx Global Server and DynamX software (Waters). For peptide identification, a minimum of 3 fragment matches were required and peptide mass tolerance was 2 ppm. 47 peptides

were identified for $\alpha$BC with full sequence coverage.

**$^{15}$N$^{1}$H CPMG-RD** Relaxation dispersion experiments were recorded on a 600 MHz Varian instrument at 37°C. The number of points was 1150 and 100, acquisition time was 64 ms and 24 ms respectively. The constant time relaxation period was 40 ms. 7 values of CPMG pulse frequencies were applied in randomised order: 50, 150, 250, 750, 500, 1000 Hz. The highest frequency was acquired twice to derive a measure of uncertainty. A 2.5 s recycle delay was applied between successive experiments. Data was processed using NMRPipe (Delaglio et al. [1995]) and FuDa (Hansen et al. [2007], http://www.biochem.ucl.ac.uk/hansen/fuda/). Processing included solvent suppression, a Gauss-Lorentz window function, zero-filling twice, and linear prediction in the indirect dimension. The initial linewidth and radius of each peak was adjusted manually.

**$^{13}$C$^{1}$H CPMG-RD** Experiments were acquired as in the previous section with the pulse sequences adapted for $^{1}$3C. The number of points in the direct and indirect dimensions was 1154 and 100, acquisition time was 64 ms and 21.4 ms respectively. 15 values of CPMG pulse frequencies were applied in randomised order: 50, 100, 150, 200, 250, 300, 350, 450, 500, 600, 650, 750, 850, 900, 1000 Hz. Processing was as for the $^{15}$N$^{1}$H CPMG experiments.

**CEST** Experiments were recorded on a 600 MHz Varian instrument at 37°C. For the experiment on c$\alpha$BC and $\alpha$Lac, the proteins were at 1000 $\mu$M and 400 $\mu$M respectively and direct intensity of the water peak was measured. The strength of the selective pulse was 300 Hz and it was applied every 0.26 ppm. For the experiment on full-length $\alpha$BC and $\alpha$Lac, the proteins were at 200 $\mu$M and 100 $\mu$M respectively and the ratio of water peak intensity in successive pairs of spectra with and without the selective pulse was measured. The strength of the selective pulse was 100 Hz and it was applied every 0.49 ppm. The number of points was 19231, acquisition

time was 2 s and 1 transient per FID was acquired in both experiments.

## 7.3 Results and discussion

### 7.3.1 The $\beta$4-$\beta$8 groove of $\alpha$BC is not involved in the interaction with the aggregating substrate

The $\beta$4-$\beta$8 groove on $\alpha$B-crystallin core domain serves as an interaction site for the C-terminus in the full-length molecule, as has been observed by crystallography (Laganowsky et al. [2010]), NMR (Baldwin et al. [2011a], Delbecq et al. [2012]) and native MS (Hilton et al. [2013]). This groove is the largest hydrophobic patch on the molecule, as illustrated in **Figure 7.1**, and has been implicated in binding aggregating substrates (Mainz et al. [2015]). I set out to test this hypothesis for the case of $\alpha$Lac. To this end, I produced two constructs of c$\alpha$BC with a point mutation in the groove, S135Q and S135W (**Figure 7.1a**). Glutamine was chosen as a relatively large hydrophilic residue known to inhibit C-terminal peptide binding (Delbecq et al. [2012]) and tryptophan as a large sidechain that would make the groove less accessible. The substitutions were checked using the software Chimera to make sure that there were rotamers of the mutant sidechains that did not produce significant steric clashes with the neighbouring residues, to avoid excessive structural perturbation of the protein.

The mutant proteins were stable and gave native mass spectra similar to wild type, with the expected difference in mass. To demonstrate that the groove was indeed blocked in the mutants, I performed C-terminal peptide binding experiments using native MS. The results are shown in **Figure 7.1b-d**. c$\alpha$BC appears as two overlapping charge series, for the monomer and the dimer. In the presence of increasing concentrations of the C-terminal peptide ERTIPITRE, peaks corresponding to c$\alpha$BC monomer with one peptide, c$\alpha$BC dimer with one peptide and c$\alpha$BC dimer with two peptides appeared and grew in intensity compared to the peaks corre-

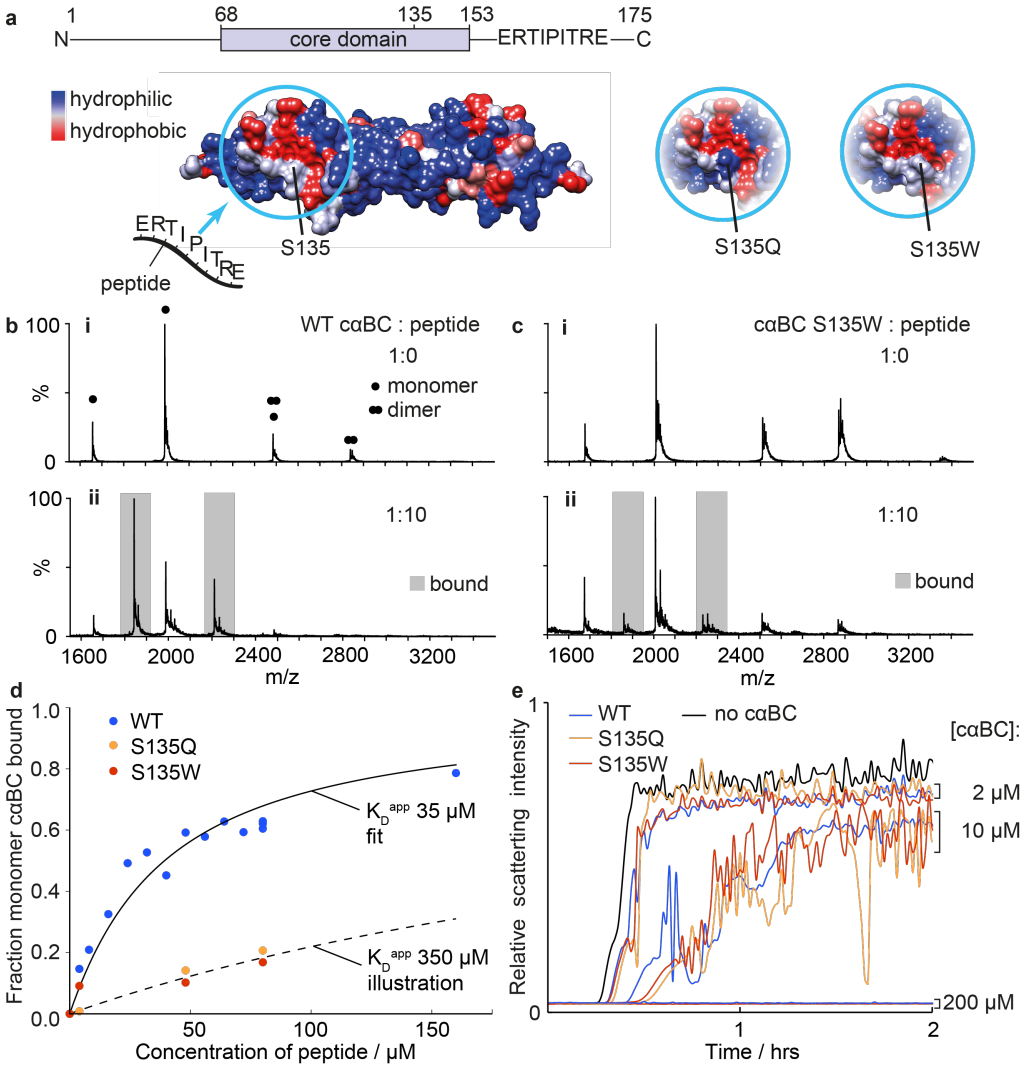



***Figure 7.1:*** ***αLac does not bind in the β4-β8 groove by analogy to the C-terminus of αBC.*** ***a*** *Illustration of the β4-β8 groove on the structure of cαBC coloured by hydrophobicity, the mutations S135Q and S135W and the C-terminal region which interacts with the groove.* ***b,c*** *Example mass spectra of wild-type cαBC and cαBC S135W respectively, without the peptide (**i**) and with a 10-fold excess of peptide (**ii**). The peaks corresponding to some of the bound forms are highlighted in grey.* ***d*** *Estimation of the binding affinity from the fraction of bound monomer cαBC. The solid black line shows a fit to a two-component equilibrium with the apparent $K_D$, and the dashed line illustrates the analogous curve for a tenfold higher $K_D$.* ***e*** *Light scattering aggregation assays at 200 μM αLac and variable chaperone concentration for wild-type cαBC and the two mutants.*

sponding to free c$\alpha$BC monomer and dimer. The most intense peaks corresponding to bound c$\alpha$BC are highlighted in grey in **Figure 7.1b** and **c**. The data was analysed using UniDec software to extract the area under each species. Since this is a complex equilibrium involving six different species (c$\alpha$BC monomer and dimer, singly bound monomer, singly bound dimer, fully bound dimer and free peptide, see **Section 2.3.4**), a multiparameter fitting must be performed to extract binding constants between them. The signal-to-noise ratio of the data was insufficient to perform such fits, especially in the case of the weakly binding mutants. Therefore, to judge the relative ability of these proteins to bind the C-terminal peptide, I compared the fraction of peptide-bound monomer relative to the total observed monomer signal, assuming a simplified equilibrium $C + P \rightleftharpoons CP$, which enabled me to determine the apparent monomer binding $K_D^{app}$. The data is shown in **Figure 7.1d** for all three proteins. I derived an expression for the bound fraction in terms of $K_D^{app}$ and performed a least squares fit to estimate $K_D^{app}$ for the wild-type protein, obtaining a value of 35 $\mu$M. Previously reported true $K_D$ for the interaction between the apo wild-type core domain and peptide is 70 $\mu$M (Hilton et al. [2013]), so the simple approximation used here resulted in a value of the same order of magnitude. Equivalent fitting was not performed for the mutants due to weak binding and concomitant low signal-to-noise ratio, but for illustration the expected curve for a $K_D$ of 350 $\mu$M is shown in **Figure 7.1d**, which comes close to the data points. It is clear that peptide binding is significantly impaired in the mutants, i.e. the $\beta$4-$\beta$8 groove is obstructed.

To relate this to the chaperone function of $\alpha$BC, I performed light scattering aggregation assays with c$\alpha$BC and both mutants, at a constant $\alpha$Lac concentration of 200 $\mu$M and a range of chaperone concentrations. Wild-type and S135Q constructs have a small extinction coefficient and since accurate concentration measurement is critical for a fair comparison of aggregation curves between the chaperones, c$\alpha$BC concentration was measured with a more accurate Bradford assay instead of absorbance at 280 nm. Each reaction was performed in triplicate. **Figure 7.1d** shows

representative traces for the control and three concentrations of chaperone. While some of the curves show a lot of scatter, there is no evidence for significant difference in the chaperone function between the wild type c$\alpha$BC and either of the two mutants.

Thus while binding to the hydrophobic groove of c$\alpha$BC is clearly impaired, the chaperone function, as measured by light scattering, remains unaffected. While $\alpha$Lac could still bind to hydrophobic areas around S135, it does not bind in the groove in the same mode as the C-terminus of $\alpha$BC. This is consistent with the absence of IXI motifs in the sequence of $\alpha$Lac.

### 7.3.2 Greater exposure of the N-terminus implicated in chaperone function by HX-MS

HX-MS reports on solvent accessibility of exchangeable protons in the protein of interest, and therefore potentially on location of macromolecular interfaces. It is performed by mixing the sample with deuterated buffer, thus initiating exchange of protons to deuterons, incubating the sample for a defined period of time and quenching the sample to stop any further exchange before analysis by mass spectrometry. Digestion of the protein into peptides by pepsin following exchange allows localisation of any differences to short sequence regions. Greater solvent accessibility of a sequence stretch will result in more extensive exchange and therefore higher mass of the fragment. In principle, two kinetic limits of exchange behaviour exist (**Figure 7.2**). If the rate of release of a binding partner or conformational change is faster than the intrinsic rate of amide proton exchange (EX2 regime), then the latter is rate-limiting and the envelope of the peptide spectrum shifts to higher m/z in a linear fashion. If the rate of the conformational change or binding is significantly slower than the rate of amide proton exchange (EX1 regime), then as soon as a different state is accessed, the exposed amides exchange quickly. Therefore in the case of two-state exchange a bimodal distribution is observed with the relative

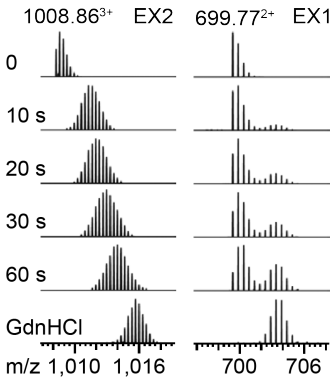



***Figure 7.2: Exchange limits in an HX-MS experiment.*** *Variation in m/z of two example peptides with duration of the exchange period (from protein ESCRT-III, modified from Yang et al. [2015]). In the EX2 regime, a linear shift of the unimodal peptide envelope to higher m/z occurs. In the EX1 regime, a second more exchanged population appears and increases in intensity. The fully exchanged control denatured with guanidinium hydrochloride (GdnHCl) is also shown.*

populations changing over time.

I compared hydrogen exchange profiles for $\alpha$BC in the presence and absence of $\alpha$Lac to pinpoint areas of the chaperone that could be affected either by direct binding of $\alpha$Lac or by any changes in the dynamics or structure of the chaperone.

The aggregation reaction was performed overnight at 37°C as described in **Section 4.2**, with $\alpha$BC and $\alpha$Lac at 200 $\mu$M each. I then monitored the extent of exchange of amide protons in $\alpha$BC after increasingly long periods of incubation with deuterated buffer, with 6 timepoints between 10 s and 140 min. **Figure 7.3** shows the fractional uptake of deuterium, i.e. the number of exchanged sites relative to the total number of exchangeable amides and sidechains in each peptide, across the $\alpha$BC sequence.

I obtained full coverage of $\alpha$BC with 47 peptic peptides (mapped as points across the sequence in **Figure 7.3a**). In general, deuterium uptake increased with time for each peptide, as expected. There was significant variation in the extent of maximum uptake across the protein, with over 50% in some parts of the N-terminus and no significant uptake at all over this time period in a region of the core domain. In general, the termini were much more exchangeable than the core domain. This is as expected for the exposed and flexible C-termini. The N-termini are thought to be on the inside of the $\alpha$BC oligomers (e.g. Delbecq and Klevit [2013]), and observation of their significant deuterium uptake means that if this is the case, subunit exchange must occur to allow them to be solvent-accessible.

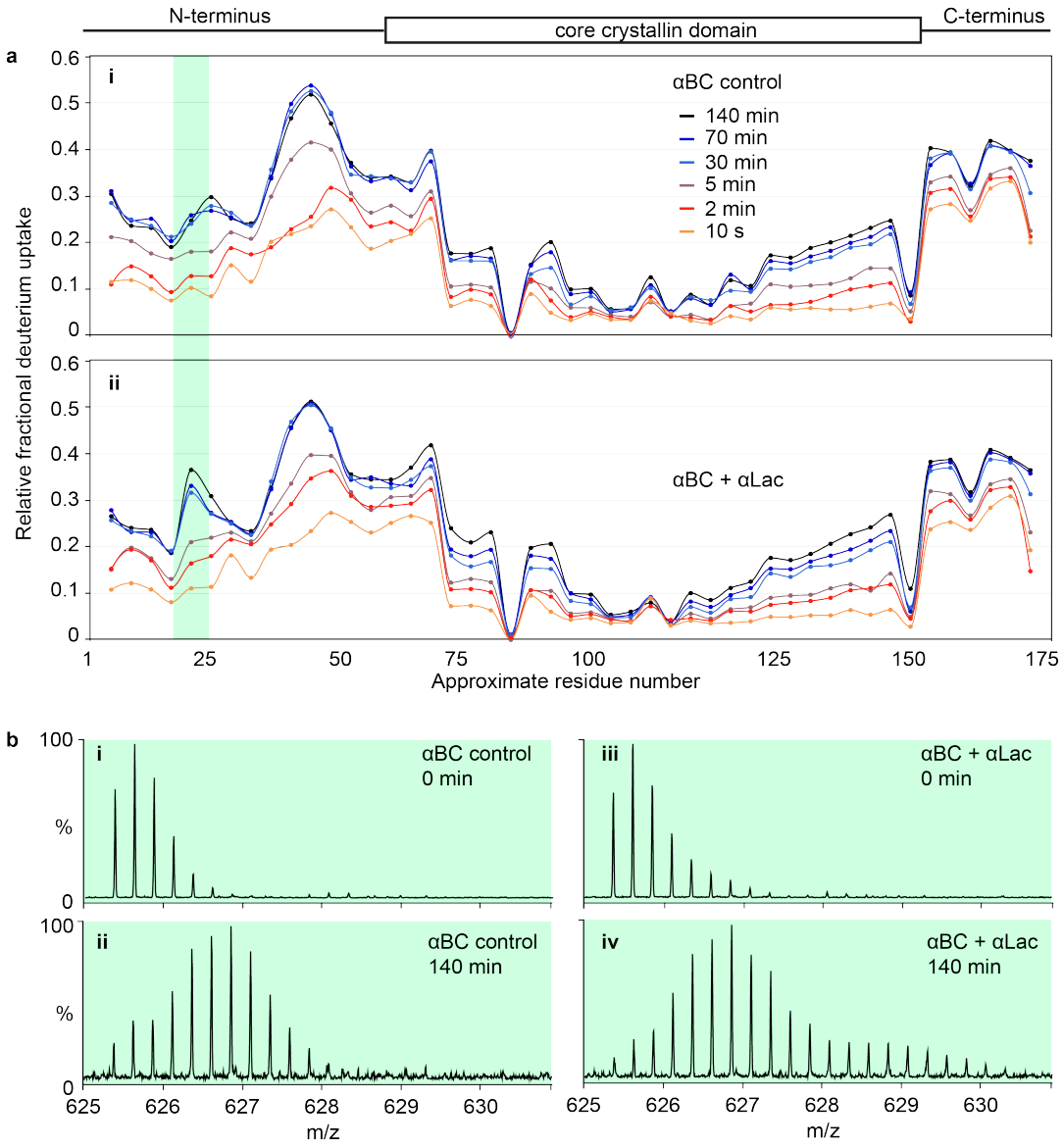



***Figure 7.3: Effect of aggregating substrate on hydrogen exchange of αBC. a*** *Profiles of relative fractional uptake of deuterium over time for 47 identified peptides for both the αBC control (**i**) and for the αBC-αLac mixture (**ii**). Each vertical set of points corresponds to a peptide. The region of most significant change, corresponding to the peptide in (**b**), is highlighted in green. **b** Spectra for the peptide 4-AIHHPWIRRPFFPFHSPSRL-23, charge state 4+, at time 0 and at the longest timepoint for αBC (**i-ii**) and the mixture (**iii-iv**), showing that a second minor population of peptide with greater extent of exchange is observed in the mixture but not in the control.*

Surprisingly little difference was observed as a function of substrate (**i** vs **ii** **Figure 7.3a**). The region of most significant difference was a peptide close to the beginning of the N-terminus, where, surprisingly, more deuterium uptake was seen in the presence of substrate. Example mass spectra of the corresponding peptide

are shown in **Figure 7.3b**. While the control spectra at t=0 min are similar, and at long times the envelope shifts to higher m/z in both cases, in the presence of $\alpha$Lac a minor second population of peptide with higher uptake is observed (**Figure 7.3b iv**). The appearance of this second population was also seen for two other charge states of this peptide.

Therefore the data suggests that the presence of the substrate leads to deprotection of a region, with appearance of a second more exposed population of peptide. This deprotection could either be static or dynamic. An example of the former would be formation of complexes with $\alpha$Lac in which the $\alpha$BC N-termini are on the outside. This hypothesis could be tested, for example, by limited proteolysis of the complexes. An alternative explanation for greater exposure of the N-terminus could be enhanced subunit exchange rate. Dissociation into smaller chaperone-active oligomers has been reported for some sHSPs but has not been seen for $\alpha$BC (see discussion in **Section 6.1**). However, it is known that $\alpha$BC can exhibit different rates of subunit exchange in different conditions even without significant change in average oligomer size (Baldwin et al. [2011d]). To test the hypothesis that subunit exchange rate is faster in the presence of $\alpha$Lac, direct measurement of subunit exchange rates could be performed using isotopically labelled samples of $\alpha$BC, as described in **Section 3.3.1**.

Why is so little difference in deuterium uptake seen for $\alpha$BC with and without substrate? Firstly, from chemical kinetics analysis in **Chapter 5** it is known that a significant proportion of $\alpha$BC is unbound under these conditions (e.g. 90% of the chaperone signal is observed by NMR at long times for this concentration ratio in **Figure 5.12**). The difference in the exchange profiles from the relatively small fraction of $\alpha$BC in complexes may not be sufficient to be observed in this experiment and therefore a higher ratio of $\alpha$Lac to $\alpha$BC should be tested. Another possibility is that denaturation of the complexes upon quenching led to formation of aggregates that are less accessible to pepsin. This could be checked e.g. by performing DLS on the samples before and after quenching. If this is the case, adding less substrate to

produce smaller complexes (as seen by SEC in **Chapter 6**) could help improve the efficiency of digestion, as could manipulation of flow rate and trapping time for this step.

Cheng et al. [2008] performed the analogous experiment with the plant sHSPs PsHsp18.1 and TaHsp16.9 and also could not detect any changes in HX-MS profiles upon addition of the substrates malate dehydrogenase (MDH) and luciferase. The authors proposed that the interaction is dynamic, with the off-rate much faster than than detectable by deuterium exchange. To assess the validity of this explanation for the case of $\alpha$BC and $\alpha$Lac, we can compare the predicted substrate release rates from chemical kinetics analysis in **Chapter 5** with estimated intrinsic amide exchange rates for equivalent conditions. The group of Englander and co-workers have established a comprehensive dataset on intrinsic amide exchange rates, which are affected by pH, temperature, salt concentration and primary structure context (Bai et al. [1993]). At 20°C, pH 7 and 100 mM NaCl, most residues exchange their amides at rates between $1\times10^{-3}$ and $1\,s^{-1}$. At 37°C, they are expected to be faster. As shown in **Figure 5.13**, the predicted rates of release of $\alpha$Lac aggregates, the most abundant species at the end of the aggregation reaction, are on the order of $1\times10^{-3}$ to $1\times10^{-7}\,s^{-1}$ at 37°C, i.e. some orders of magnitude slower than the intrinsic amide exchange rates. The chemical kinetics analysis thus suggests that transient interactions are an unlikely explanation for why the substrate has little effect on the amide exchange rates of the chaperone. In fact, observation of additional EX1-type exchange for the N-terminus (**Figure 7.2** and **7.3b**) is consistent with the relatively slow rates of aggregate binding derived by chemical kinetics.

In summary, the HX-MS experiment suggests that the N-terminus becomes more solvent-exposed in the presence of aggregating substrate. Thus the N-termini of $\alpha$BC are on the outside of the complexes with $\alpha$Lac, or the subunit exchange rate of $\alpha$BC is accelerated by the substrate. Both hypotheses may also be concurrently true. They can be tested by direct subunit exchange measurements and limited proteolysis experiments to probe the architecture of the complexes.

### 7.3.3 Peak position and intensity variation in 2D HSQC spectra

**Chapter 5** presents detailed analysis of one-dimensional HSQC spectra of $\alpha$BC and the isolated core domain c$\alpha$BC in the presence of aggregating $\alpha$Lac. These spectra showed a uniform decrease in intensity as a function of substrate concentration without significant changes in shape. From these observations, NMR diffusion experiments and DLS data, it was concluded that it was residual free chaperone that was being observed in these experiments without significant contribution from $\alpha$Lac-bound species.

The unbound chaperone is in dynamic equilibrium with $\alpha$Lac, i.e. it is undergoing chemical exchange between free and bound states, of which there may be more than one. In the bound forms, at least some chemical shifts must be distinct from those of the free state. Detection of these effects is made difficult by the heterogeneity and large size of the interaction partners. **Figure 1.10** in **Chapter 1** illustrates the simulated effects on the spectrum for exchange in different regimes (slow, intermediate and fast) for the case where there is a significant difference in transverse relaxation rate $R_2$ between the exchanging states, as would be the case if one binding partner is much larger than the other. In slow exchange, loss of signal without line broadening would be expected. Intermediate exchange would lead to broadening, and in fast exchange a chemical shift change together with broadening would be seen. The regime for each residue depends on the binding rates as well as on the difference in chemical shifts for that residue ($\Delta\tilde{\omega}$) in the free and bound state.

These residue-specific effects may be subtle and difficult to detect in 1D experiments. Therefore, I recorded the equivalent 2D correlation NMR experiments on c$\alpha$BC and $\alpha$BC in the presence and absence of $\alpha$Lac, examined in detail below. The spectra were acquired after the steady state had been reached, i.e. at about 12 hours after the start of the aggregation reaction.

**c$\alpha$BC**

**$^{15}$N$^{1}$H-HSQC backbone spectra** **Figure 7.4** shows $^{15}$N$^{1}$H HSQC spectra of c$\alpha$BC on its own (black), with $\alpha$Lac in the absence of DTT (cyan) and with $\alpha$Lac and DTT, after 13 hours of aggregation (red). c$\alpha$BC was at 1000 $\mu$M and $\alpha$Lac was at 400 $\mu$M in this experiment.

The spectra of c$\alpha$BC alone, and with $\alpha$Lac in the absence of reductant, were virtually identical. In contrast, small but reproducible peak shifts were observed in the presence of aggregated substrate (i.e. only in the mixture with DTT, marked with arrows in **Figure 7.4a**). These peak shifts were also seen in the equivalent spectra for other recorded concentration ratios of the mixture where the chaperone was in excess (200 $\mu$M c$\alpha$BC and 100 $\mu$M or 200 $\mu$M $\alpha$Lac). I also tested an extreme concentration ratio of 2000 $\mu$M vs. 200 $\mu$M c$\alpha$BC, where extensive loss of intensity of peaks across the spectrum precluded observation of chemical shift changes. The CSPs from **Figure 7.4a** are mapped on the structure of c$\alpha$BC dimer in **Figure 7.4c i**. The residues with the greatest backbone CSP are dispersed around the protein. Some of them are in the $\beta$4-$\beta$8 groove and others in the intra-dimer interface. Thus no well-defined substrate binding patch is apparent from this analysis.

I also measured the intensity of each peak in the spectrum of the chaperone alone and in the steady state phase of aggregation (black and red in **Figure 7.4a**). A correlation plot of these intensities is shown in **Figure 7.4b**. If a peak's intensity were unaffected upon aggregation of the substrate, it would fall on the $y = x$ line. All points fall below this line, consistent with the general decay of the intensity of the chaperone spectrum already observed in the 1D experiments in **Chapter 5**. Around 50% overall decrease in intensity is observed at this concentration ratio. Some peaks that fall noticeably above and below the line are marked on the correlation plot. The deviations of peak intensity from the line of best fit are shown on the structure of c$\alpha$BC in **Figure 7.4c ii**. The residues with disproportionately low intensity cluster around the beginning of strand $\beta$9 and the neighbouring loop between strands $\beta$3

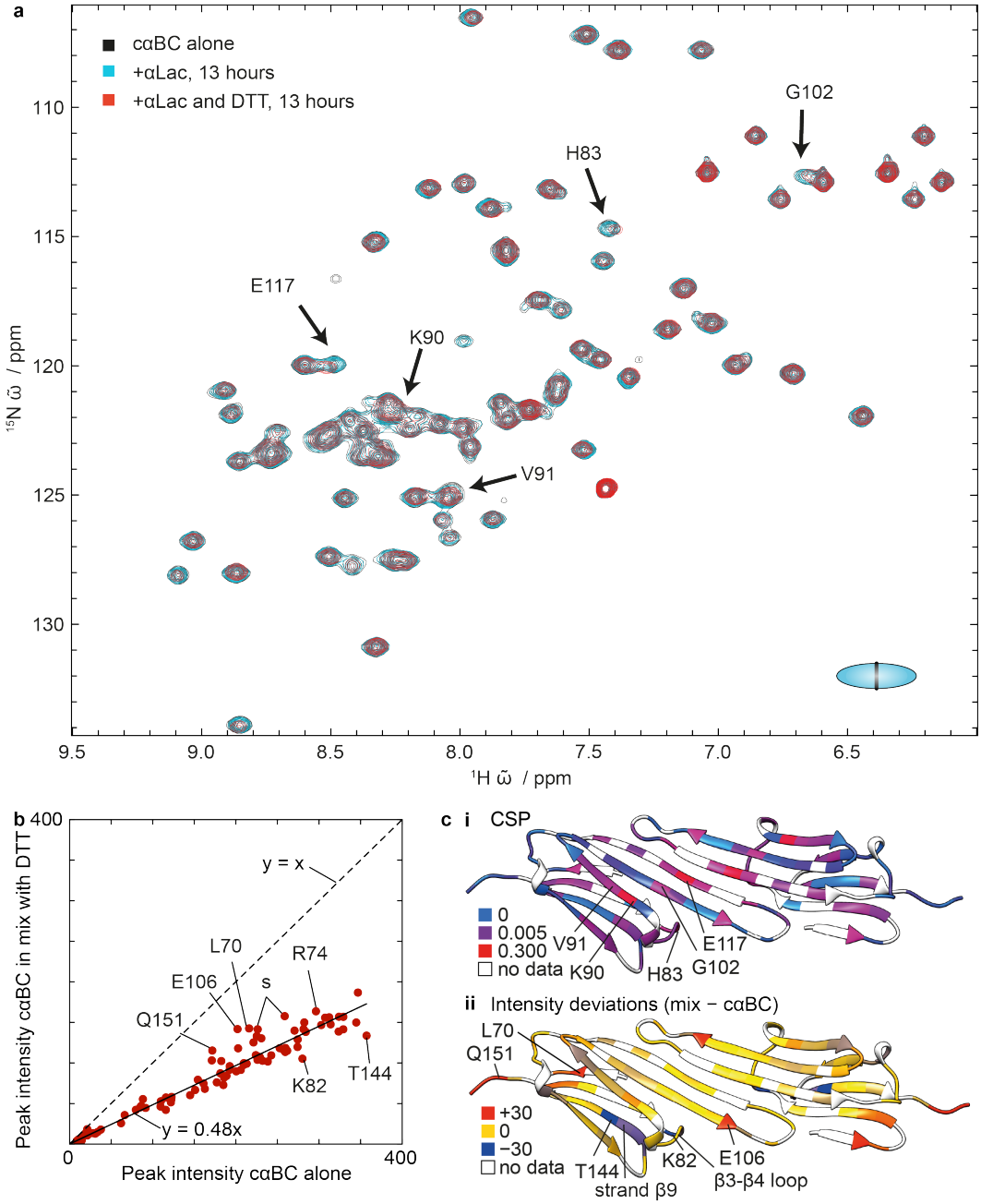



***Figure 7.4: Analysis of 2D $^{15}N$ spectra of cαBC with aggregating substrate.*** ***a*** *$^{15}N$-HSQC overlay for cαBC alone and in a mixture with either αLac in the absence of reductant or with αLac and DTT after 13 hours of aggregation. cαBC is at a total concentration of 1000 μM and αLac is at a total concentration of 400 μM.* ***b*** *Correlation plot of peak intensities in the two spectra above. s - sidechains.* ***c*** *The peak shifts (**i**) and intensity deviations (**ii**) displayed on the structure of the cαBC dimer.*

and $\beta$4. This additional peak broadening could result from intermediate exchange with the substrate (e.g. suspected for K82 and T144).

The residues with disproportionately high intensity localise to the ends of $\beta$-strands and the C-terminus. Unlike broadening, relative sharpening of the peaks is difficult to explain (see the **Introduction** to this Chapter). A possible explanation is that especially mobile residues that are unaffected by substrate are directly visible in the spectra, thus the loss of signal is less than expected. For examples, the very intense peak for the C-terminal residue S153 is not shown in **Figure 7.4b** as it lies far away from the rest of the points. It lies far above the line of best fit (signal-to-noise ratio 1520 units for c$\alpha$BC alone and 1270 units in the presence of aggregating substrate, thus a fraction of 0.83). This peak is visible even in the presence of a 10-fold excess of $\alpha$Lac. 2D $^{15}$N-edited diffusion could be performed to check whether these peaks have a significant translational contribution from a larger species. Another possibility could be that existing intermediate-timescale exchange is quenched by substrate binding. However, as seen from the discussion on CPMG, significant intermediate exchange could be detected only on the dimer interface of c$\alpha$BC, and the disproportionately sharp residues in question are not on the dimer interface.

There are therefore widespread CSPs in c$\alpha$BC induced by substrate binding, but the intensity deviations are more localised and suggest that the area around the start of strand $\beta$9 and the $\beta$3-$\beta$4 loop could be a substrate binding site. The observation that this is distinct from the $\beta$4-$\beta$8 groove is also consistent with the finding of **Section 7.3.1** that the latter region is not important for chaperone function.

**Isoleucine, leucine, valine (ILV) spectra** Deuteration combined with methyl labelling extends the molecular mass range observable by NMR. There are several reasons for this. Firstly, deuteration hinders one of the dominant relaxation mechanisms in proteins - proton dipolar coupling. Secondly, each methyl carbon has three

protons attached and hence gives more intense signal than e.g. an amide group with one proton. Thirdly, the methyl group has favourable relaxation properties, and it is possible to devise experiments that exploit interference between different relaxation mechanisms to find conditions where they cancel out (termed transverse relaxation-optimised spectroscopy (TROSY), Ollerenshaw et al. [2003]). Although most of the methyl groups are often on the inside of the protein due to their hydrophobicity, any exposed sidechains may also be more sensitive to the binding of the aggregating client than the backbone.

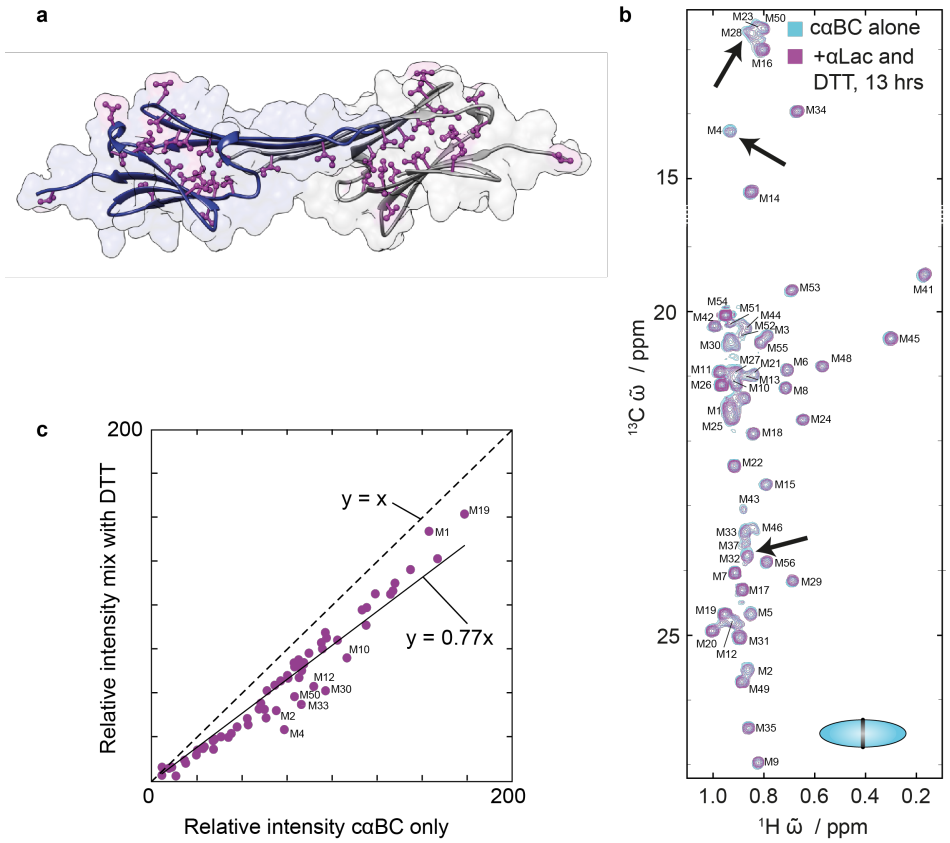



***Figure 7.5: Analysis of methyl spectra of cαBC in the presence of αLac.*** ***a*** *Location of the isoleucine, leucine and valine sidechains in the cαBC dimer structure.* ***b*** $^{13}C^{1}H$*-HSQC overlay for cαBC alone and in a mixture with αLac after 13 hours of aggregation. cαBC was at a total concentration of 200 μM and αLac was at a total concentration of 100 μM.* ***c*** *Correlation plot of peak intensities in the two spectra above.*

To make use of these advantages, I prepared a deuterated isoleucine, leucine, valine (ILV) methyl-labelled sample of cαBC and recorded methyl TROSY spectra in the absence and presence of aggregating αLac, at concentrations of 200 μM and

100 $\mu$M respectively. The results are shown in **Figure 7.5**, with the overlaid spectra in panel **b** and the intensity correlation plot in panel **c**. Surprisingly, the spectra showed little difference beyond the expected general decrease in intensity, which was comparable to the $^{15}$N-HSQC spectra at equivalent protein concentrations (see **Figure 5.3**). Some minor peak shifts were observed, marked with arrows in **Figure 7.5b**. There was also more variation in intensity around the line of best fit (**Figure 7.5c**), with a few peaks falling noticeably below the line. Those sidechains could potentially be affected by the chemical exchange from binding of the substrate. No assignment of the methyl sidechains was performed for this pilot experiment, therefore any changes cannot yet be mapped on the structure of c$\alpha$BC. Assignment experiments such as the HCCH TOCSY in conjunction with the already assigned HNCA could be used for this size of protein. Repeats of the experiment are required to decide whether the observed small CSPs and intensity deviations constitute a significant and reproducible pattern. The experiment should also be performed at a range of $\alpha$Lac concentrations to find conditions where the effects of any chemical exchange are most pronounced.

### $\alpha$BC

Full-length $\alpha$BC is itself large and oligomeric, with an average molecular weight of about 600 kDa, thus further complicating the analysis of interactions with the substrate. Only the flexible C-terminus is visible by solution-state NMR. I recorded 2D correlation experiments on uniformly $^{15}$N$^{1}$H-labelled and $^{13}$C$^{1}$H-labelled samples of full-length $\alpha$BC to assess any residue-specific effects of aggregating $\alpha$Lac on the C-terminal tails. The results are shown in **Figure 7.6**.

**$^{15}$N$^{1}$H-HSQC backbone spectra** Only a few peaks were visible in backbone $^{15}$N-HSQC (**Figure 7.6a**). In the presence of aggregating substrate, no significant peak shifts were observed, and the decrease in peak intensity was uniform. This suggests that the peaks observed here are in the slow exchange regime with respect

to free and $\alpha$Lac-bound forms of the chaperone.

**$^{13}$C$^{1}$H-HSQC spectra** The spectrum of the $^{13}$C-labelled sample showed more peaks compared to $^{15}$N-labelled one. In the range of frequencies recorded, the $C_\alpha$ carbons and the sidechain carbons of the C-terminus are visible but not the carbonyls, whose carbon chemical shifts are in the 160-200 ppm range. Note that decoupling was turned off in this experiment so some of the peaks appear as doublets. $\alpha$BC was at 200 $\mu$M and $\alpha$Lac was at 100 $\mu$M in this sample. No notable changes in peak position were observed **Figure 7.6b i**. However, there were selective changes in intensity of individual peaks (panel **b ii**). The scatter around the line of best fit is comparable to that of methyl-labelled c$\alpha$BC in **Figure 7.5**.

**Figure 7.6b iii** demonstrates the selective changes in intensity on 1D spectra of $\alpha$BC in the presence of more $\alpha$Lac - 800 $\mu$M, fourfold excess over chaperone. There is a clear overall decrease in intensity of the spectrum in the presence of aggregating substrate as observed previously. If this low-intensity spectrum is scaled to the highest peak in the reference spectrum, it becomes clear that the pattern of intensity is distinct: while some peaks overlay with the reference, others are clearly above or below, i.e. the decrease in intensity is not uniform across all peaks. Because of the lack of assignment of backbone and sidechain carbons of $\alpha$BC, these selective changes in intensity cannot yet be mapped on the structure. They could be investigated with exchange experiments as a function of $\alpha$Lac concentration.

### 7.3.4 Intermediate exchange analysis with $\alpha$BC: CPMG-RD

The 2D HSQC spectra of both chaperones have revealed some residue-specific peak and intensity changes against the backdrop of uniform intensity loss in the presence of aggregating substrate. Thus while chaperone resonances are predominantly in the slow exchange regime, there are some indications of intermediate or fast exchange for some peaks. I therefore recorded 2D Carr-Purcell-Meiboom-Gill relaxation dis-

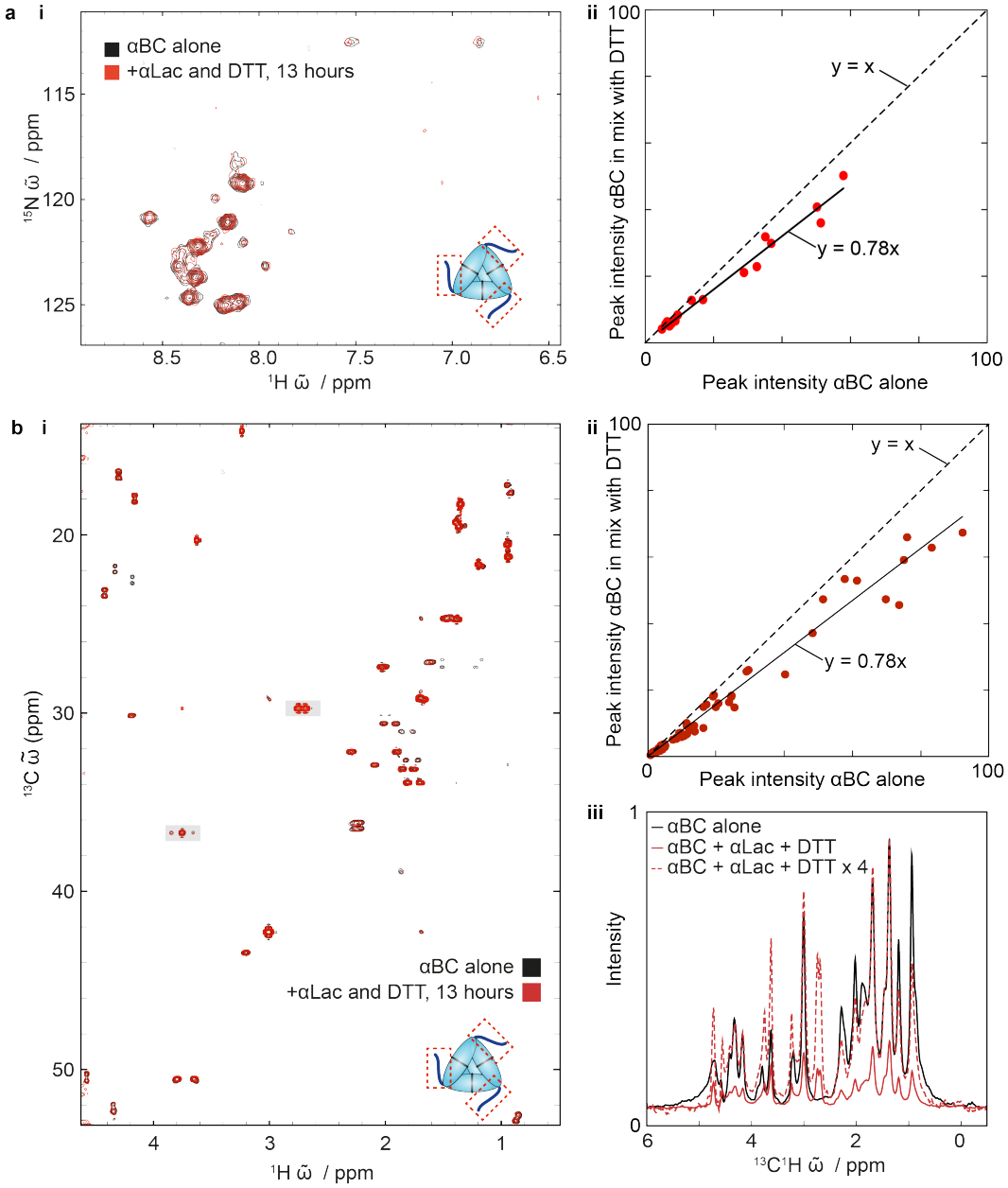



***Figure 7.6: Analysis of $^{15}N$ and $^{13}C$ spectra of full-length αBC in the presence of aggregating substrate.*** ***a i*** *$^{15}$N-HSQC overlay for αBC alone (black) and in a mixture with αLac after 13 hours of aggregation (red). αBC was at 200 μM and αLac was at 100 μM.* ***ii*** *Correlation plot of peak intensities in the two spectra in* ***a****.* ***b i*** *Equivalent overlay of $^{13}$C-HSQC spectra. Grey boxes mark peaks coming from DTT.* ***ii*** *Correlation plot of peak intensities in the $^{13}$C-HSQC spectra in* ***i****.* ***iii*** *Demonstration of differential effects on peak intensity of αBC in the presence of αLac. Black - reference spectrum of αBC at 200 μM. Red, solid line - spectrum of αBC in the presence of αLac 12 hours after aggregation. Red, dashed line - the previous spectrum scaled fourfold to normalise it to the highest peak in the reference spectrum.*

persion (CPMG-RD) experiments to explore any residue-specific changes in motions on the ps-ms timescale in c$\alpha$BC upon addition of aggregating substrate. I recorded $^{15}$N-edited CPMG-RD experiments on 1000 $\mu$M $^{15}$N-labelled c$\alpha$BC in the absence and presence of 400 $\mu$M $\alpha$Lac, and $^{13}$C-edited CPMG-RD experiments on 200 $\mu$M $^{15}$N- and ILV methyl-labelled c$\alpha$BC in the absence and presence of 100 $\mu$M $\alpha$Lac. The principle of the experiment is illustrated in **Figure 7.7a**, with sample data from the $^{13}$C-edited experiment (peak number M5). In these trial experiments, acquired at a single temperature and field strength, I estimated the values of $R_2$ and $R_{ex}$ from the plot of $R_2^{eff}$ versus pulse frequency as shown in **Figure 7.7a i**.

**$^{15}$N-CPMG-RD on c$\alpha$BC** **Figure 7.7b** shows the estimated values of $R_2$ and $R_{ex}$ for the c$\alpha$BC backbone. The spectrum with location of assigned peaks is shown in **Chapter 2** (**Figure 2.10**). The data from 12 unassigned c$\alpha$BC peaks labelled U1-12 is also shown. The error bars are calculated from two repeats of the measurement at the highest pulse frequency.

There was a general increase in the $R_2$ values across the protein in the presence of $\alpha$Lac. This is characteristic of lifetime broadening observed in other aggregating systems (Fawzi et al. [2010], Bodner et al. [2009], Libich et al. [2013]) where interaction with a large species is involved. This finding should be confirmed by dedicated experiments for measuring $R_2$ more reliably than estimated here.

The $R_{ex}$ values were generally small, except on the dimer interface. Rajagopal et al. [2015] have also reported relaxation dispersions in the dimer interface of c$\alpha$BC and concluded that these are due to monomer-dimer exchange. They analysed the process at pH 7.5 and 22°C. The $R_{ex}$ values they observed were e.g. $10\,s^{-1}$ for E117 and $18\,s^{-1}$ for R120, in good agreement with the values seen here. The $K_D$ for the reaction was 36 $\mu$M and the rate constants were $k_{md} = 1.1 \times 10^7\,\mathrm{M^{-1}s^{-1}}$ (dimerisation) and $k_{dm} = 400\,s^{-1}$ (dimer dissociation).

The overall pattern of $R_{ex}$ values is similar in the presence of $\alpha$Lac, but there

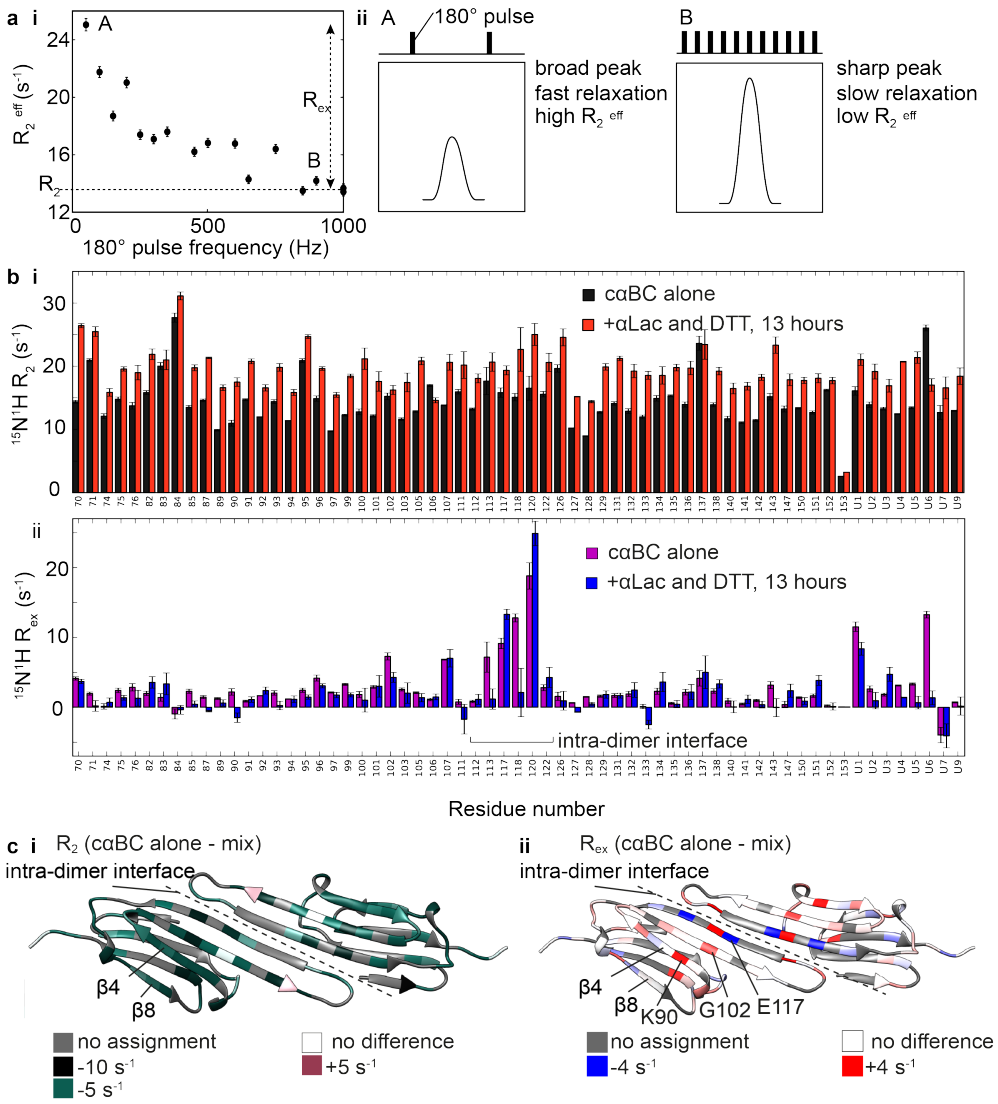



***Figure 7.7:* $^{15}$*N CPMG-RD of cαBC with and without the aggregating substrate.*** ***a*** *The principle of the CPMG-RD experiment with* ***(i)*** *example data from peak M5 in the* $^{13}$*C-CPMG experiment and* ***(ii)*** *illustration of the effect of increasing the frequency of the 180° pulses in the pulse sequence.* ***b i*** $R_2$ *values for each peak attributable to cαBC in the absence and presence of αLac.* ***ii*** *The corresponding* $R_{ex}$ *values. The unassigned peaks are labelled U1-12.* ***c*** *The differences between the control and mixture samples (control-mixture) mapped onto the crystal structure of the cαBC dimer for* ***(i)*** $R_2$ *and* ***(ii)*** $R_{ex}$*.*

are some noticeable differences in $R_{ex}$ of individual peaks. For example for E117, R120 and K121 $R_{ex}$ increases in the presence of $\alpha$Lac but for F118 there is a dramatic decrease. $R_{ex}$ depends on many parameters of the exchange process, including the populations of the ground and excited state $p_a$ and $p_b$, the exchange timescale $k_{ex}$, the $R_2$ of the ground and excited state and the chemical shift difference between them, $\Delta\tilde{\omega}$ (Palmer et al. [2001], Baldwin [2014], Palmer [2014]). For example, in the case of

two-state exchange with no difference in $R_2$ between the ground and excited state the $R_{ex}$ can be approximated as $R_{ex} = \frac{p_a p_b k_{ex}}{1 + (k_{ex}/\Delta\tilde{\omega})^2}$ (Palmer [2014]). If the binding of the substrate modified any of the parameters referring to the overall process, i.e. $p_a$, $p_b$ or $k_{ex}$, a similar pattern in the change of $R_{ex}$ should be expected for the affected residues. However, $\Delta\tilde{\omega}$ and $R_2$ are specific for each residue. Thus depending on how the binding of the substrate affects the conformation of the residues on the dimer interface, their $R_{ex}$ could either increase or decrease.

**Figure 7.7c** shows the differences in $R_2$ and $R_{ex}$ between the samples of the chaperone alone and in the presence of aggregating $\alpha$Lac mapped on the structure of the c$\alpha$BC dimer. The $R_2$ plot illustrates the overall increase of the estimated $R_2$ across the protein, especially towards the $\beta$4-$\beta$8 groove. **Figure 7.7c ii** shows that the residues whose $R_{ex}$ is affected by the presence of substrate form a network that links the dimer interface with the $\beta$4-$\beta$8 groove. There is some correlation between residues where the largest chemical shifts and differences in $R_{ex}$ are observed, for example, K90, G102 and E117 show both a chemical shift perturbation and an $R_{ex}$ change.

I also recorded the $^{15}$N-edited CPMG-RD experiment for the 200 $\mu$M deuterated, $^{15}$N- and ILV methyl-labelled c$\alpha$BC in the absence and presence of 100 $\mu$M $\alpha$Lac (see **Figure A.7** in the **Appendix**). The overall pattern of $R_2$ and $R_{ex}$ variation across the protein was similar, but the increase in $R_2$ in the presence of $\alpha$Lac was not clearly observed. The signal-to-noise ratio in this sample was lower, and since no repeat measurements were made, the error is difficult to estimate. The different concentrations used also affect the timescale of the exchange.

**$^{13}$C-CPMG-RD** To analyse any intermediate exchange behaviour of the sidechain methyls, I recorded a $^{13}$C-edited CPMG experiment for the ILV methyl-labelled sample of c$\alpha$BC at 1000 $\mu$M in the absence and presence of 400 $\mu$M $\alpha$Lac. The data is shown in **Figure 7.8**. The peak numbering corresponds to the labels in **Figure 7.5**; please note it is arbitrary as there is no assignment of the c$\alpha$BC methyl sidechains.

**Figure 7.8c** shows an example fit and CPMG plot for peak M4 to provide an illustration of the quality of the data.

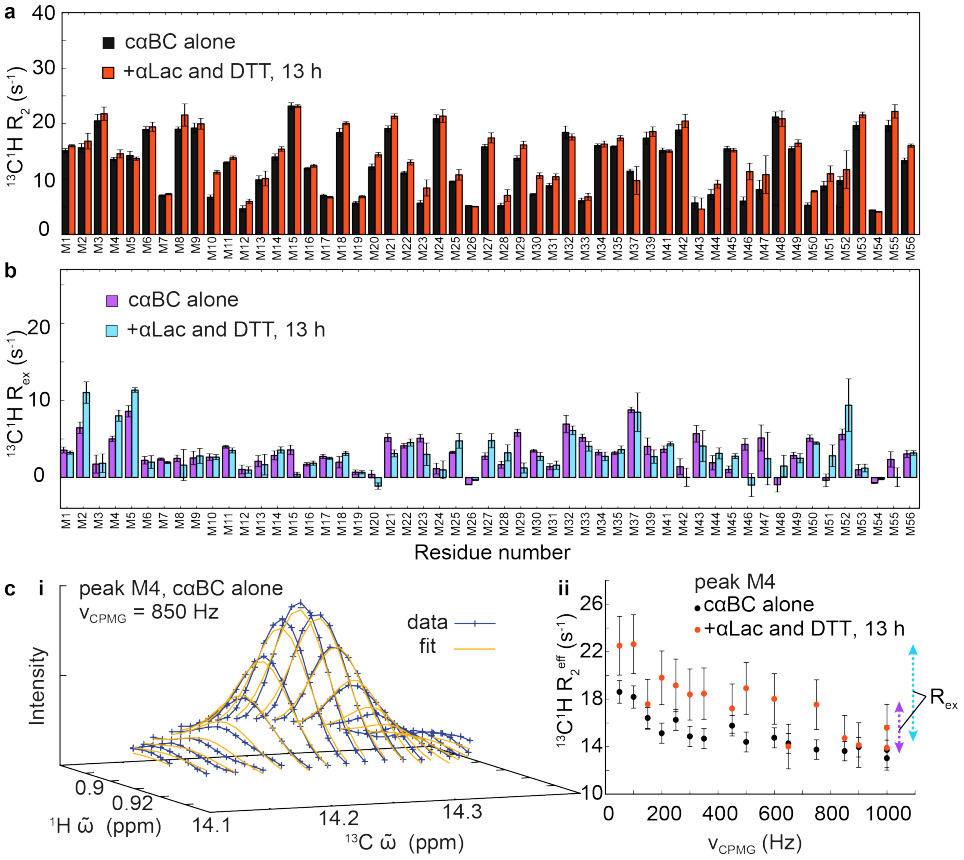



***Figure 7.8:*** ***$^{13}$C CPMG-RD on cαBC in the absence and presence of aggregating substrate.*** ***a*** *$^{13}C^1H$ $R_2$ values for each cαBC methyl peak in the absence and presence of αLac. The numbering is arbitrary since no assignments are available for the methyl spectrum of cαBC.* ***b*** *The corresponding $^{13}C^1H$ $R_{ex}$ values.* ***c*** ***i*** *Example peak fit for peak M4 to illustrate the quality of the fit.* ***ii*** *Example plot of effective $R_2$ for peak M4 as a function of the CPMG frequency from which the $R_2$ and $R_{ex}$ values in panels* ***a*** *and* ***b*** *were derived.*

The lack of assignment and therefore random ordering of peaks makes it difficult to see any sequential correlations. However, similarly to the $^{15}$N experiment, $R_2$ was generally slightly higher in the presence of aggregating substrate than in its absence. The $R_{ex}$ values were generally low with small differences upon addition of αLac. A few residues such as M2, M4, M5 and M52 had higher than average values of $R_{ex}$ and could correspond to the intra-dimer interface.

In summary for the CPMG-RD data, an increase in the effective $R_2$ is observed across cαBC in the presence of aggregating substrate, confirming that the chaperone

is in exchange with large aggregates. Some significant changes in $R_{ex}$ are observed upon addition of $\alpha$Lac particularly in the intra-dimer interface, which could reflect either a shift in the dimerisation equilibrium of c$\alpha$BC or substrate binding in this region.

### 7.3.5 Slow exchange analysis: DEST

The chemical exchange saturation transfer (CEST) experiment and its variant, DEST, is suitable for analysis of slow exchange (Anthis and Clore [2015], Fawzi et al. [2010]). DEST uses the same pulse sequence but is applied in cases of exchange with large species and differs in the observed effects and analysis of the data. These experiments rely on the ability to saturate resonances with a selective pulse, and on transfer of saturation from one state to another via chemical exchange. An array of selective pulses is applied across the spectrum one at a time (**Figure 7.9a**), while the intensity of a ground state peak is monitored. When the pulse excites the ground state directly, significant loss of intensity of this peak is observed (**Figure 7.9a ii**, direct saturation). If an exchange state with a much larger $R_2$ (the 'dark' state) exists, its resonances are very broad and can be saturated at chemical shifts away from the ground state peak. The intensity profile therefore shows a broader trough (**Figure 7.9a ii**). If a family of datasets is acquired at multiple concentrations and multiple pulse frequencies, and $R_2$ of the ground state is measured, the intensity profiles (**Figure 7.9a ii**) can be analysed using Bloch-McConnell equations to derive the kinetic and thermodynamic parameters of the exchange (Anthis and Clore [2015]). Here, I carried out preliminary experiments to show that DEST is a useful approach to study this aggregating system.

The CEST/DEST pulse sequence is under development in our laboratory by Gogulan Karunanithy, and in its isotope-edited form will be used to monitor specific peaks in $\alpha$BC spectra. Here, data from experiments with initial versions of the pulse sequence is presented: rather than looking at protein resonances, I monitored

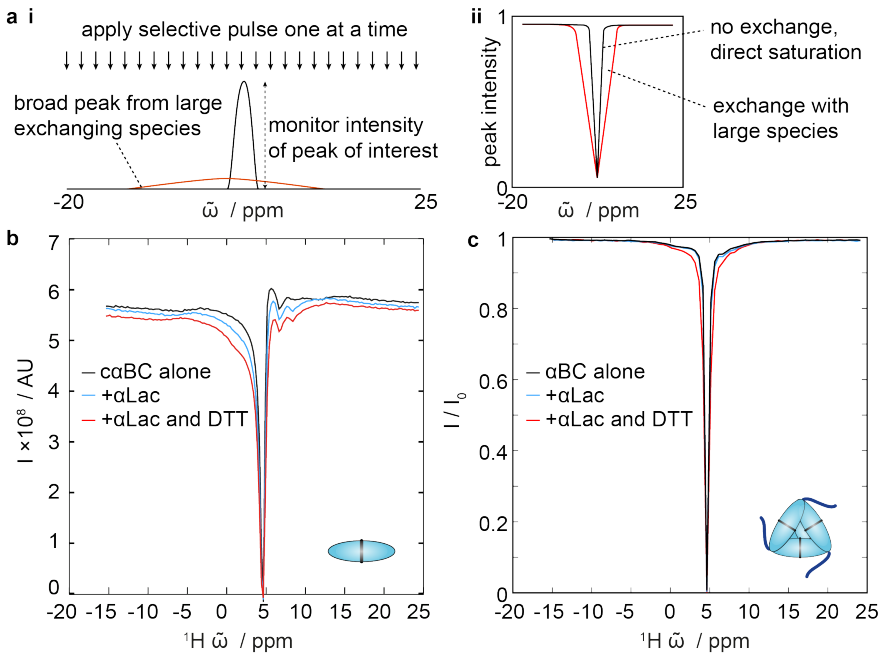



***Figure 7.9: DEST on water as a means of monitoring aggregation.*** ***a*** *Explanation of the DEST experiment.* ***b*** *DEST experiment with cαBC on its own and in the presence of non-aggregating and aggregating αLac, in which absolute intensity of the water resonance was monitored. The concentration of the chaperone was 1000 μM and that of αLac was 400 μM.* ***c*** *A DEST experiment with full-length αBC with and without aggregating substrate and DTT, in which the ratio of intensities with and without the selective pulse was monitored. The concentration of the chaperone was 200 μM and that of αLac was 100 μM.*

intensity of the water signal as the selective pulse was scanned between -15 and +24 ppm. The aim was to investigate whether exchange of protons in water shows differences in the presence of large aggregates.

The experiment with cαBC (**Figure 7.9a**) recorded direct intensity of the water peak. The instrument suffered from periodic fluctuations in signal due to incorrectly controlled humidity of the room. The experiment with the full-length αBC used an improved version of the pulse sequence which records pairs of experiments with the pulse on and off, allowing normalisation of the pulse-on experiment to the immediately preceding pulse-off one. This reduced any artefactual systematic variation in intensity of the spectra as a function of chemical shift. In both experiments, a noticeable difference in trough width is observed in the presence of the aggregating substrate (red curves in **Figure 7.9b** and **c**). Although the relative change is small, the water peak is extremely intense, therefore intensity changes of a few percent

can be reliably measured. Thus the DEST effect observed here is consistent with formation of large particles corresponding to $\alpha$Lac aggregates and complexes with the chaperone, as shown by multiple methods in **Chapter 6**.

The experiments described here focused on exchange of water. The same experiments can be performed on the chaperone, since chemical kinetics analysis in **Chapter 5** predicts slow exchange of the chaperone with $\alpha$Lac aggregates. Thus 2D isotope-edited DEST experiments monitoring residue-specific chaperone resonances promise to be of great use for further study of residue-specific elements of chaperone function of $\alpha$BC.

## 7.4 Conclusions

**Figure 7.10** shows a summary of the results of this Chapter. It maps residues and sequence regions that were affected by the presence of aggregating $\alpha$Lac in HX-MS, 2D HSQC and CPMG experiments, and shows the position of S135 which was shown not to be important for c$\alpha$BC chaperone function by mutagenesis. $\alpha$Lac clearly binds to or near the C-termini and the core domain, since their NMR signals are attennuated in the presence of $\alpha$Lac. However, $\alpha$Lac does not bind in the hydrophobic $\beta$4-$\beta$8 groove by analogy to the C-terminus of $\alpha$BC. Instead, the residues affected by $\alpha$Lac cluster in the $\beta$9 strand, the $\beta$3-$\beta$4 loop and the start of the $\beta$4 strand, and, interestingly, in the intra-dimer interface. Thus $\alpha$Lac could either bind directly in the intra-dimer interface, or on the outside of the $\beta$-sandwich. It is also possible that multiple $\alpha$Lac molecules bind to the various residues detected here. Binding in the intra-dimer interface would attach special significance to the involvement of monomers in subunit exchange of $\alpha$BC, demonstrated in **Chapter 3**, since the intra-dimer interface is exposed in these units.

Mainz et al. [2015] have shown that the $\alpha$BC N-terminus is important for binding lysozyme, a protein that is structurally related to $\alpha$Lac. Nevertheless, the

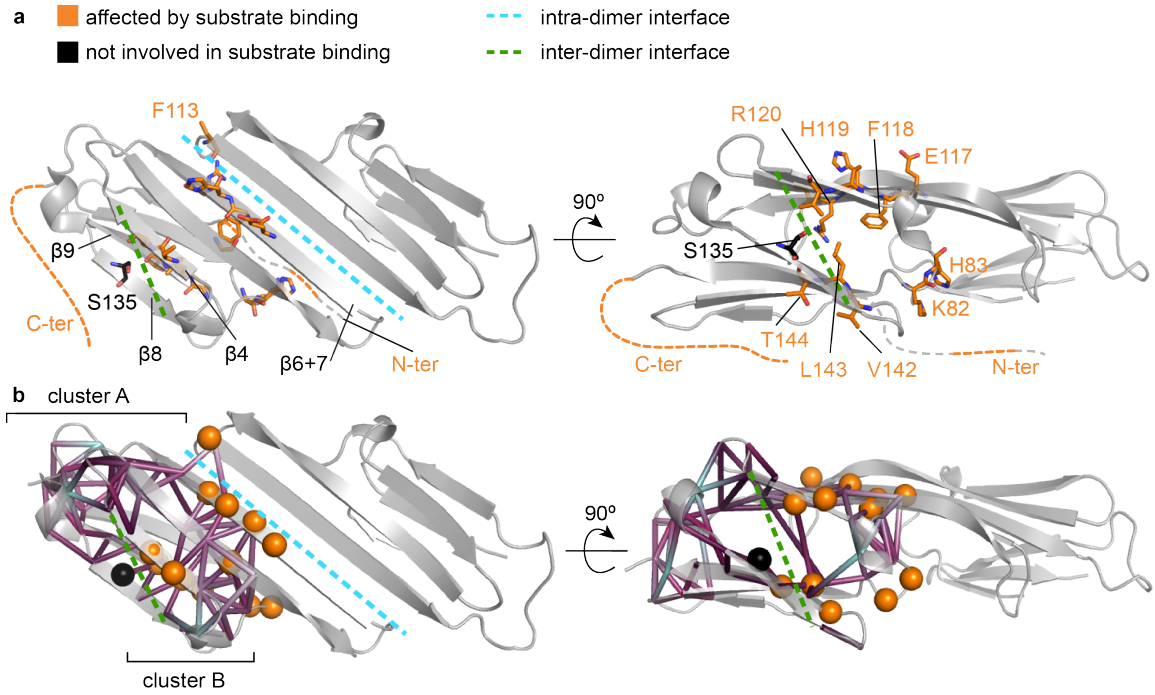



***Figure 7.10: Understanding chaperone function of αBC at residue-level.*** ***a*** *Summary of residues and regions of αBC that were either found to be implicated in chaperone function of αBC (orange) or known not to be important for chaperone function (black, S135) from the experiments performed in this Chapter. Residues are shown as sticks on the X-ray structure of cαBC (PDB ID 2W7J), and the termini are added schematically.* ***b*** *Comparison of the core domain residues identified in* ***a****, rendered as spheres at Cα positions, with the interface communication network derived in* ***Chapter 2****, shown as cylinders. The cylinders are coloured as in* ***Figures 2.17*** *and* ***2.18****. Interaction with aggregating substrate perturbs one of the two connection clusters of the network (cluster B).*

sequence identity between hen lysozyme and bovine αLac is 38%, and the dominant roles for the C-terminus and core domain for αLac may be reflecting the recognition of different primary structure elements for the two proteins, in line with known promiscuity and versatility of αBC.

Panel **b** of **Figure 7.10** shows an overlay of the residues with the interface communication network elucidated in **Chapter 2**. The set of core domain residues affected by the presence of aggregating αLac falls within one of the clusters of the network, cluster B, and impinges on the same inter-monomer contact, R120-F113'. Thus although αLac does not mimic the C-terminus of αBC, it still engages a subset of the same communication pathways which control the energetics of the quaternary assembly. Observation of greater solvent accessibility of the N-terminus

also supports quaternary reorganisation in the presence of aggregating substrate.

In summary, the C-terminus and a cluster of residues linking one edge of the $\beta$-sandwich and the intra-dimer interface are implicated in chaperone function of $\alpha$BC towards $\alpha$Lac. The identified residues have varied chemical identity, supporting substrate-tailored rather than generic hydrophobic or charge-based binding. Engagement of the C-termini and of one leg of the interface communication network of $\alpha$BC constitutes a dual mechanism for substrate-stimulated quaternary reorganisation of the oligomeric ensemble.

# Chapter 8

# Conclusions

Protein aggregation is a hallmark of many debilitating diseases, including Alzheimer's and Parkinson's. Molecular chaperones such as $\alpha$BC are part of the innate defences of cells and organisms against aggregation. In the work described in this thesis, I used a combination of cutting-edge methods to tackle the heterogeneous and dynamic nature of aggregating systems, exploring the structural and functional dynamics of the polydisperse molecular chaperone $\alpha$B-crystallin *in vitro* with and without the model aggregating substrate $\alpha$Lac. The key results are summarised in **Figure 8.1**.

As part of the investigation of the the remarkable quaternary plasticity of $\alpha$BC (top panel of **Figure 8.1**), I determined experimental CCS values of individual odd and even stoichiometries within the ensemble, from the 18-mer to the 32-mer, going beyond previous ensemble average measurements and demonstrating the power of IMMS to interrogate the full range of such assemblies. The CCS values can be used as restraints or means of validation in structural modelling. The CCS scaling and crosslinking results rule out previously published ensemble average structures, and point to mixed-degree vertices and polyhedral geometry of the oligomers (**Figure 8.1a**). The ability to control the polydispersity of $\alpha$BC by cysteine crosslinking potentially paves the way for high-resolution structure determination of the high-

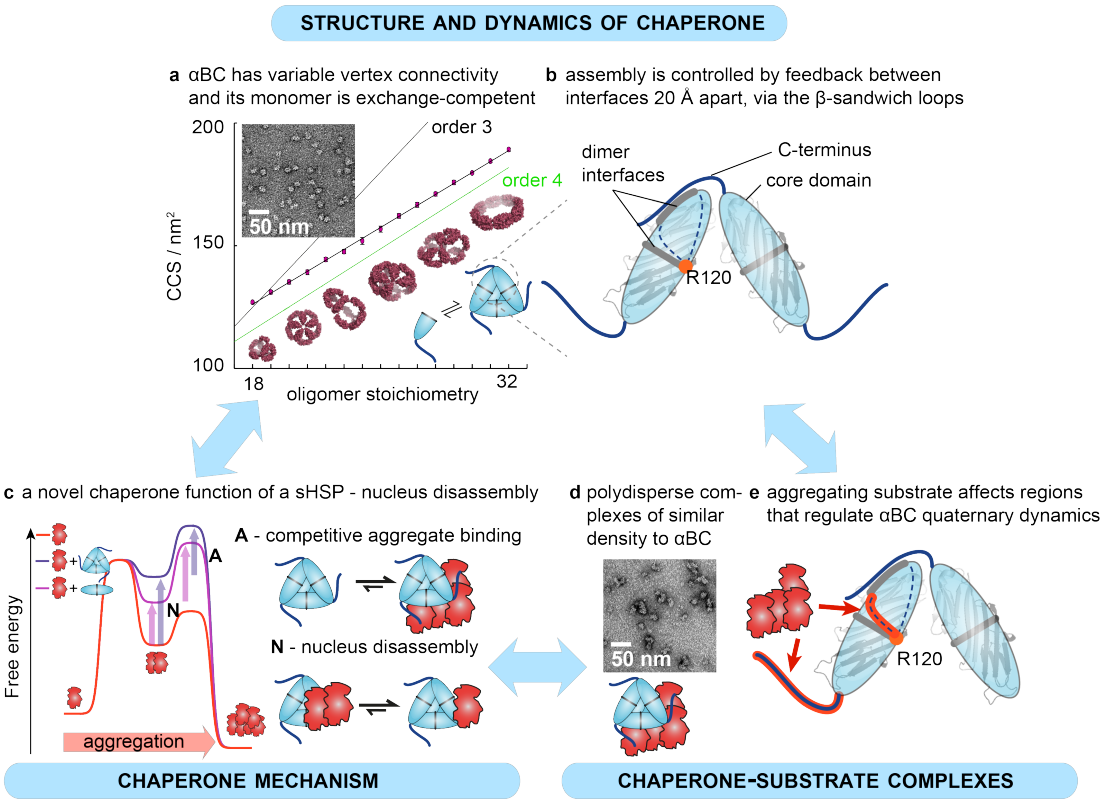



***Figure 8.1: Summary of thesis findings.*** *The work in this thesis can be grouped into three interconnected themes: structure and dynamics of the chaperone αBC in isolation, its chaperone function and the properties of its complexes with the aggregating substrate αLac. I determined CCS values of individual stoichiometries of αBC (**a**) and uncovered the pathways of crosstalk between distinct monomer interfaces within the oligomers (**b**). I developed a new methodology to derive kinetic and thermodynamic descriptions of chaperone action for αBC and its isolated core domain cαBC (**c**). Going on to investigate the predicted complexes of the chaperone and αLac both at coarse (**d**) and residue-specific (**e**) levels, I showed that the substrate impinges on the regulatory elements of αBC identified in the top panel. The results show the tremendous potential of combining NMR, computational analysis and native MS for investigating heterogeneous and dynamic assemblies of large proteins.*

symmetry stoichiometries. Moreover, the variation in vertex order implies a range of angles at which the assembly units come together, highlighting the plasticity of the αBC monomer with its central domain and flexible tails.

The network of communication between two different monomer interfaces, elucidated by NMR, constitutes a complementary mechanistic insight into oligomer assembly (**Figure 8.1b**). The network has many redundant pathways, mainly through the loops at both sides of the β-sandwich (dashed lines in **Figure 8.1b**), but it re-

markably converges on the edges of the intra-dimer interface, including a key residue whose mutation results in cataracts, myopathy and neuropathy, R120. Even more strikingly, the presence of aggregating substrate $\alpha$Lac leads to perturbation of one leg of the network (**Figure 8.1e**), including the residue R120. $\alpha$Lac also binds to or near the C-terminus, which forms part of the inter-dimer interface in $\alpha$BC. Thus the substrate has a dual effect on the structural elements regulating quaternary dynamics of $\alpha$BC. Application of IMMS to the complexes of $\alpha$BC with the substrate, using the methods described in **Chapter 3**, and other methods such as solid-state NMR, will reveal the resulting perturbations of the quaternary ensemble of $\alpha$BC and the composition and architecture of the complexes between $\alpha$BC and the aggregating substrate. Furthermore, together with the observation of a range of connectivities in $\alpha$BC oligomers, and therefore exposure of the C-terminus and core domain regions, these results suggest that different stoichiometries might vary in their ability to chaperone the substrate.

In order to study aggregation and chaperone action in a comprehensive and quantitative way, I developed a new framework based on NMR and chemical kinetics (**Figure 8.1c**). Applying the framework to $\alpha$Lac, I found that $\alpha$BC interacts with multiple $\alpha$Lac species on the aggregation pathway. It slows down aggregation of $\alpha$Lac monomers through two main mechanisms: by kinetic competition with large aggregates and, surprisingly, by disassembly of the aggregation nuclei. The latter is a previously unknown function of sHSPs, difficult to detect by other methods due to the low abundance of the unstable nuclei. It is nevertheless a very effective chaperone action since nucleation is a rate-limiting early step in aggregation and at least in the case of $\alpha$Lac yields monomeric protein, which in the cell would be accessible to ATP-dependent refolding chaperones.

The work described here enables the investigation of $\alpha$BC with disease-relevant substrates, such as $\alpha$-synuclein and A$\beta$, as well as any other heterogeneous and dynamic system. In particular, with the use of isotope labelling the NMR and chemical kinetics analysis can be performed in the presence of other proteins, for example

downstream chaperones, in cell lysates or even inside cells. It is very powerful for informed study of aggregation inhibitors, both macromolecules and small-molecule drugs, and can be combined with other techniques, such as the ThT-based chemical kinetics method, to obtain increasingly robust information on these systems. With constant advances in resolution and buffer compatibility, native MS also has exciting potential in analysis of *in vivo*-like, heterogeneous mixtures. Finally, in cases where direct interpretation of experimental data is difficult, comparison of quantitative simulations to data is a powerful approach to tackle complexity. By applying these and other methods in combination and validating findings *in vivo*, we can lay the foundations for eliminating the formidable neurodegenerative diseases, discern fundamental principles of protein self-assembly and learn more about the fascinating diversity of mechanisms that organisms have evolved to maintain proteostasis.

# References

G. Abraham and R. J. Colonno. Many rhinovirus serotypes share the same cellular receptor. *Journal of Virology*, 51(2):340–5, 1984.

E. Ahrman, W. Lambert, J. A. Aquilina, C. V. Robinson, and C. S. Emanuelsson. Chemical cross-linking of the chloroplast localized small heat-shock protein, Hsp21, and the model substrate citrate synthase. *Protein Science*, 16(7):1464–1478, 2007.

T. M. Allison, E. Reading, I. Liko, A. J. Baldwin, A. Laganowsky, and C. V. Robinson. Quantifying the stabilizing effects of protein-ligand interactions in the gas phase. *Nature Communications*, 6:8551, 2015.

E. V. Anslyn and D. A. Dougherty. *Modern Physical Organic Chemistry*. University Science Books, Sausalito, California, 2006.

N. J. Anthis and G. M. Clore. Visualizing transient dark states by NMR spectroscopy. *Quarterly Reviews of Biophysics*, 1:35–116, 2015.

J. A. Aquilina and S. J. Watt. The N-terminal domain of αB-crystallin is protected from proteolysis by bound substrate. *Biochemical and Biophysical Research Communications*, 353(4):1115–1120, 2007.

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. *Proceedings of the National Academy of Sciences of the United States of America*, 100(19):10611–6, 2003.

J. A. Aquilina, J. L. Benesch, L. L. Ding, O. Yaron, J. Horwitz, and C. V. Robinson. Phosphorylation of $\alpha$B-crystallin alters chaperone function through loss of dimeric substructure. *Journal of Biological Chemistry*, 279(27):28675–28680, 2004.

J. A. Aquilina, J. L. P. Benesch, L. D. Lin, O. Yaron, J. Horwitz, and C. V. Robinson. Subunit exchange of polydisperse proteins: Mass spectrometry reveals consequences of $\alpha$A-crystallin truncation. *Journal of Biological Chemistry*, 280(15): 14485–14491, 2005.

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 $\alpha$B-crystallin and HSP27. *The Journal of Biological Chemistry*, 288(19):13602–9, 2013.

A. P. Arrigo. Human small heat shock proteins: Protein interactomes of homo- and hetero-oligomeric complexes: An update. *FEBS Letters*, 587(13):1959–1969, 2013.

R. C. Augusteyn. Dissociation is not required for $\alpha$-crystallin's chaperone function. *Experimental Eye Research*, 79:781–784, 2004.

S. Baglioni, F. Casamenti, M. Bucciantini, L. M. Luheshi, N. Taddei, F. Chiti, C. M. Dobson, and M. Stefani. Prefibrillar Amyloid Aggregates Could Be Generic Toxins in Higher Organisms. *Journal of Neuroscience*, 26(31):8160–8167, 2006.

C. Bagnéris, O. A. Bateman, C. E. Naylor, N. Cronin, W. C. Boelens, N. H. Keep, and C. Slingsby. Crystal Structures of $\alpha$-Crystallin Domain Dimers of $\alpha$B-Crystallin and Hsp20. *Journal of Molecular Biology*, 392(5):1242–1252, 2009.

Y. Bai, J. Milne, L. Mayne, and W. Englander. Primary Structure Effects on Peptide Group Hydrogen Exchange. *Proteins*, 17(1):75–86, 1993.

A. J. Baldwin. An exact solution for R2,eff in CPMG experiments in the case of two site chemical exchange. *Journal of Magnetic Resonance*, 244:114–124, 2014.

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. *Journal of Molecular Biology*, 413(2): 310–20, 2011a.

A. J. Baldwin, T. P. J. Knowles, G. G. Tartaglia, A. W. Fitzpatrick, G. L. Devlin, S. Shammas, C. A. Waudby, M. F. Mossuto, S. Meehan, S. L. Gras, J. Christodoulou, S. J. Anthony-Cahill, P. D. Barker, M. Vendruscolo, and C. M. Dobson. Metastability of Native Proteins and the Phenomenon of Amyloid Formation. *Journal of the American Chemical Society*, 133:14160–3, 2011b.

A. J. Baldwin, H. Lioe, G. R. Hilton, L. A. Baker, J. L. Rubinstein, L. E. Kay, and J. L. P. Benesch. The polydispersity of αB-crystallin is rationalized by an interconverting polyhedral architecture. *Structure*, 19(12):1855–63, 2011c.

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. *Journal of Molecular Biology*, 413(2):297–309, 2011d.

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. *Journal of the American Chemical Society*, 2012.

S. Banerjee and S. Mazumdar. Electrospray Ionization Mass Spectrometry: A Technique to Access the Information beyond the Molecular Weight of the Analyte. *International Journal of Analytical Chemistry*, 2012:1–40, 2012.

E. Basha, K. L. Friedrich, and E. Vierling. The N-terminal arm of small heat shock proteins is important for both chaperone activity and substrate specificity. *Journal of Biological Chemistry*, 281(52):39943–39952, 2006.

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. *Journal of Molecular Biology*, 425(10):1683–1696, 2013.

N. J. Baxter and M. P. Williamson. Temperature dependence of 1H chemical shifts in proteins. *Journal of Biomolecular NMR*, 9(4):359–369, 1997.

J. L. P. Benesch and B. T. Ruotolo. Mass spectrometry: Come of age for structural and dynamical biology. *Current Opinion in Structural Biology*, 21(5):641–649, 2011.

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. *Chemical Reviews*, 107(8):3544–67, 2007.

M. D. Benson, R. A. Smith, G. Hung, B. Kluve-Beckerman, A. D. Showalter, K. W. Sloop, and B. P. Monia. Suppression of choroid plexus transthyretin levels by antisense oligonucleotide treatment. *Amyloid*, 17(2):43–49, 2010.

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. *Proceedings of the National Academy of Sciences*, 109(50):20407–20412, 2012.

J. L. Berk, O. B. Suhr, L. Obici, Y. Sekijima, S. R. Zeldenrust, T. Yamashita, M. A. Heneghan, P. D. Gorevic, W. J. Litchy, J. F. Wiesman, E. Nordh, M. Corato, A. Lozza, A. Cortese, J. Robinson-Papp, T. Colton, D. V. Rybin, A. B. Bisbee, Y. Ando, S.-I. Ikeda, D. C. Seldin, G. Merlini, M. Skinner, J. W. Kelly, and P. J. Dyck. Repurposing Diflunisal for Familial Amyloid Polyneuropathy. *Journal of the American Medical Association*, 310(24):2658–67, 2013.

K. Bessonov and G. Harauz. Molecular dynamics investigation of myelin basic protein stability on lipid membranes. *Studies by Undergraduate Researchers at Guelph*, 4(1), 2010.

F. A. Bettelheim, R. Ansari, Q. F. Cheng, and J. S. Zigler. The mode of chaperoning of dithiothreitol-denatured $\alpha$-lactalbumin by $\alpha$-crystallin. *Biochemical and Biophysical Research Communications*, 261(2):292–7, 1999.

J. Bhattacharyya, E. G. Padmanabha Udupa, J. Wang, and K. K. Sharma. Mini-$\alpha$B-crystallin: a functional element of $\alpha$B-crystallin with chaperone-like activity. *Biochemistry*, 45(9):3069–3076, 2006.

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. *Progress in Biophysics and Molecular Biology*, 86(3):407–485, 2004.

C. R. Bodner, C. M. Dobson, and A. Bax. Multiple Tight Phospholipid-Binding Modes of $\alpha$-Synuclein Revealed by Solution NMR Spectroscopy. *Journal of Molecular Biology*, 390(4):775–90, 2009.

D. Boehr, D. McElheny, H. Dyson, and P. Wright. The dynamic energy landscape of dihydrofolate reductase catalysis. *Science*, 313(5793):1638–42, 2006.

W. C. Boelens. Cell biological roles of $\alpha$B-crystallin. *Progress in Biophysics and Molecular Biology*, 115(1):3–10, 2014.

A. Boncoraglio, M. Minoia, and S. Carra. The family of mammalian small heat shock proteins (HSPBs): Implications in protein deposit diseases and motor neuropathies. *International Journal of Biochemistry and Cell Biology*, 44(10):1657–1669, 2012.

G. H. H. Borner and A. B. Fielding. Using in-gel digestion and an orbitrap mass spectrometer to analyze the proteome of clathrin-coated vesicles. *Cold Spring Harbor Protocols*, pages 1188–1191, 2014.

M. Bourrelle-Langlois, G. Morrow, S. Finet, and R. M. Tanguay. In vitro structural and functional characterization of the small heat shock proteins (sHSP) of the cyanophage S-ShM2 and its host, Synechococcus sp. WH7803. *PLoS ONE*, 11(9): 1–21, 2016.

M. P. Bova, L. L. Ding, J. Horwitz, and B. K. K. Fung. Subunit exchange of $\alpha$A-crystallin. *Journal of Biological Chemistry*, 272(47):29511–29517, 1997.

M. P. Bova, O. Yaron, Q. Huang, L. Ding, D. A. Haley, P. L. Stewart, and J. Horwitz. Mutation R120G in B-crystallin, which is linked to a desmin-related myopathy, results in an irregular structure and defective chaperone-like function. *Proceedings of the National Academy of Sciences*, 96(11):6137–6142, 1999.

M. P. Bova, Q. Huang, L. Ding, and J. Horwitz. Subunit exchange, conformational stability, and chaperone-like function of the small heat shock protein 16.5 from Methanococcus jannaschii. *Journal of Biological Chemistry*, 277(41):38468–38475, 2002.

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 $\alpha$B-crystallin elucidated by a triple hybrid approach. *Proceedings of the National Academy of Sciences*, 108(51):20491–20496, 2011.

J. S. Brodbelt. Ion Activation Methods for Peptides and Proteins. *Analytical Chemistry*, 88(1):30–51, 2016.

Z. M. Bumagina, B. Y. Gurvits, N. V. Artemova, K. O. Muranov, I. K. Yudin, and B. I. Kurganov. Mechanism of suppression of dithiothreitol-induced aggregation of bovine $\alpha$-lactalbumin by $\alpha$-crystallin. *Biophysical Chemistry*, 146(2-3):108–117, 2010.

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. *Analytical Chemistry*, 82(22):9557–65, 2010.

S. Campioni, B. Mannini, M. Zampagni, A. Pensalfini, C. Parrini, E. Evangelisti, A. Relini, M. Stefani, C. M. Dobson, C. Cecchi, and F. Chiti. A causative link

between the structure of aberrant protein oligomers and their toxicity. *Nature Chemical Biology*, 6(2):140–147, 2010.

J. R. Cann and W. B. Goad. *Interacting macromolecules: the Theory and Practice of Their Electrophoresis, Chromatography, and Ultracentrifugation.* Academic Press, New York and London, 1970.

H. Y. Carr and E. M. Purcell. Effects of diffusion on free precession in nuclear magnetic resonance experiments. *Physical Review*, 94(3):630–8, 1954.

J. A. Carver, J. A. Aquilina, R. J. W. Truscott, and G. B. Ralston. Identification by 1H NMR spectroscopy of flexible C-terminal extensions in bovine lens $\alpha$-crystallin. *October*, 3(2):143–145, 1992.

J. A. Carver, R. A. Lindner, C. Lyon, D. Canet, H. Hernandez, C. M. Dobson, and C. Redfield. The interaction of the molecular chaperone $\alpha$-crystallin with unfolding $\alpha$-lactalbumin: A structural and kinetic spectroscopic study. *Journal of Molecular Biology*, 318(3):815–827, 2002.

H. Chaffey-Millar, D. Stewart, M. M. T. Chakravarty, G. Keller, and C. Barner-Kowollik. A parallelised high performance Monte Carlo simulation approach for complex polymerisation kinetics. *Macromolecular Theory and Simulations*, 16(6): 575–592, 2007.

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. *Journal of Biological Chemistry*, 283(39):26634–26642, 2008.

I. V. Chernushevich and B. A. Thomson. Collisional Cooling of Large Ions in Electrospray Mass Spectrometry. *Analytical Chemistry*, 76:1754–1760, 2004.

F. Chiti, P. Webster, N. Taddei, A. Clark, M. Stefani, G. Ramponi, and C. M. Dobson. Designing conditions for in vitro formation of amyloid protofilaments

and fibrils. *Proceedings of the National Academy of Sciences*, 96(7):3590–3594, 1999.

E. D. Chrysina, K. Brew, and K. R. Acharya. Crystal structures of apo- and holo-bovine $\alpha$-lactalbumin at 2.2-A resolution reveal an effect of calcium on inter-lobe interactions. *Journal of Biological Chemistry*, 275(47):37021–9, 2000.

T. Cierpicki, I. Zhukov, R. A. Byrd, and J. Otlewski. Hydrogen bonds in human ubiquitin reflected in temperature coefficients of amide protons. *Journal of Magnetic Resonance*, 157(2):178–80, 2002.

A. R. Clark, C. E. Naylor, C. Bagnéris, N. H. Keep, and C. Slingsby. Crystal structure of R120G disease mutant of human $\alpha$B-crystallin domain dimer shows closure of a groove. *Journal of Molecular Biology*, 408(1):118–34, apr 2011.

T. Coelho, L. F. Maia, A. Martins da Silva, M. Waddington Cruz, P. Plante-Bordeneuve, V. Lozeron, O. B. Suhr, J. M. Campistol, I. M. Conceicao, H. H. Schmidt, P. Trigo, J. W. Kelly, R. Labaudiniere, J. Chan, J. Packman, A. Wilson, and D. R. Grogan. Tafamidis for transthyretin familial amyloid polyneuropathy: a randomized, controlled trial. *Neurology*, 79(8):785–792, 2012.

T. Coelho, D. Adams, A. Silva, P. Lozeron, P. N. Hawkins, T. Mant, J. Perez, J. Chiesa, S. Warrington, E. Tranter, M. Munisamy, R. Falzone, J. Harrop, J. Cehelsky, B. R. Bettencourt, M. Geissler, J. S. Butler, A. Sehgal, R. E. Meyers, Q. Chen, T. Borland, R. M. Hutabarat, V. A. Clausen, R. Alvarez, K. Fitzgerald, C. Gamba-Vitalo, S. V. Nochur, A. K. Vaishnaw, D. W. Sah, J. A. Gollob, and O. B. Suhr. Safety and Efficacy of RNAi Therapy for Transthyretin Amyloidosis. *New England Journal of Medicine*, 369(9):819–29, 2013.

S. I. A. Cohen, S. Linse, L. M. Luheshi, E. Hellstrand, D. A. White, L. Rajah, D. E. Otzen, M. Vendruscolo, C. M. Dobson, and T. P. J. Knowles. Proliferation of amyloid-$\beta$42 aggregates occurs through a secondary nucleation mechanism.

*Proceedings of the National Academy of Sciences of the United States of America*, 110(24):9758–63, 2013.

N. Cremades, S. I. A. Cohen, E. Deas, A. Y. Abramov, A. Y. Chen, A. Orte, M. Sandal, R. W. Clarke, P. Dunne, F. A. Aprile, C. W. Bertoncini, N. W. Wood, T. P. J. Knowles, C. M. Dobson, and D. Klenerman. Direct observation of the interconversion of normal and toxic forms of $\alpha$-synuclein. *Cell*, 149(5):1048–1059, 2012.

E. Cuevas, D. Zaldívar, M. Pérez-Cisneros, H. Sossa, and V. Osuna. Block matching algorithm for motion estimation based on Artificial Bee Colony (ABC). *Applied Soft Computing*, 13(6):3047–3059, 2013.

K. P. Das, J. M. Petrash, and W. K. Surewicz. Conformational Properties of Substrate Proteins Bound to a Molecular Chaperone $\alpha$-Crystallin. *Communication*, 271(18):10449–10452, 1996.

E. De Hoffmann and V. Stroobant. *Mass spectrometry: principles and applications*. Wiley, 3rd edition, 2012.

W. W. de Jong, G. J. Caspers, and J. A. Leunissen. Genealogy of the $\alpha$-crystallin–small heat-shock protein superfamily. *International Journal of Biological Macromolecules*, 22:151–162, 1998.

A. Dehner and H. Kessler. Diffusion NMR spectroscopy: Folding and aggregation of domains in p53. *ChemBioChem*, 6(9):1550–1565, 2005.

F. Delaglio, S. Grzesiek, G. W. Vuister, G. Zhu, J. Pfeifer, and A. Bax. NMRPipe - a Multidimensional Spectral Processing System Based on Unix Pipes. *Journal of Biomolecular NMR*, 6(3):277–293, 1995.

S. P. Delbecq and R. E. Klevit. One size does not fit all: the oligomeric states of $\alpha$B-crystallin. *FEBS letters*, 587(8):1073–80, 2013.

S. P. Delbecq, S. Jehle, and R. Klevit. Binding determinants of the small heat shock protein, $\alpha$B-crystallin: recognition of the 'IxI' motif. *EMBO Journal*, 31 (24):4587–94, 2012.

W. H. DePas and M. R. Chapman. Microbial manipulation of the amyloid fold. *Research in Microbiology*, 163(9-10):592–606, 2012.

M. Dole, L. L. Mack, R. L. Hines, R. C. Mobley, L. D. Ferguson, and M. B. Alice. Molecular Beams of Macroions. *The Journal of Chemical Physics*, 49(5):2240–2249, 1968.

H. Ecroyd, S. Meehan, J. Horwitz, J. A. Aquilina, J. L. P. Benesch, C. V. Robinson, C. E. Macphee, and J. A. Carver. Mimicking phosphorylation of $\alpha$B-crystallin affects its chaperone activity. *The Biochemical Journal*, 401(1):129–41, 2007.

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 $\alpha$-casein. *Journal of Biological Chemistry*, 283(14):9012–9022, 2008.

M. Ehrnsperger, S. Gräber, M. Gaestel, and J. Buchner. Binding of non-native protein to Hsp25 during heat shock creates a reservoir of folding intermediates for reactivation. *EMBO Journal*, 16(2):221–229, 1997.

Y. S. Eisele, C. Monteiro, C. Fearns, S. E. Encalada, R. L. Wiseman, E. T. Powers, and J. W. Kelly. Targeting protein aggregation for the treatment of degenerative diseases. *Nature reviews. Drug discovery*, 14(11):759–80, 2015.

D. Eisenberg and M. Jucker. The amyloid state of proteins in human diseases. *Cell*, 148(6):1188–1203, 2012.

J. J. Ewbank and T. E. Creighton. The molten globule protein conformation probed by disulphide bonds. *Nature*, 350:518–520, 1991.

P. P. Fagerholm, B. T. Philipson, and B. Lindstroem. Normal human lens - the distribution of protein. *Experimental Eye Research*, 33(6):615–620, 1981.

N. L. Fawzi, J. Ying, D. A. Torchia, and G. M. Clore. Kinetics of amyloid beta monomer-to-oligomer exchange by NMR relaxation. *Journal of the American Chemical Society*, 132(29):9948–51, 2010.

I. K. Feil, M. Malfois, J. Hendle, H. Van der Zandt, and D. I. Svergun. A Novel Quaternary Structure of the Dimeric $\alpha$-Crystallin Domain with Chaperone-like Activity. *Journal of Biological Chemistry*, 276(15):12024–12029, 2001.

J. B. Fenn, M. Mann, C. K. A. I. Meng, S. F. Wong, and C. M. Whitehouse. Electrospray Ionization for Mass Spectrometry of Large Biomolecules. *Science*, 246(4926):64–71, 1989.

L. S. Fenn and J. A. McLean. Biomolecular structural separations by ion mobility-mass spectrometry. *Analytical and Bioanalytical Chemistry*, 391(3):905–909, 2008.

J. Fernandez De La Mora. Electrospray ionization of large multiply charged species proceeds via Dole's charged residue mechanism. *Analytica Chimica Acta*, 406(1): 93–104, 2000.

D. K. Fitzgerald, B. Colvin, R. Mawal, and K. E. Ebner. Enzymic Assay for Galactosyl Transferase Activity of Lactose Synthetase and $\alpha$-Lactalbumin in Purified and Crude Systems. *Analytical Biochemistry*, 36:43–61, 1970.

T. Fleckenstein, A. Kastenmueller, M. L. Stein, C. Peters, M. Daake, M. Krause, D. Weinfurtner, M. Haslbeck, S. Weinkauf, M. Groll, and J. Buchner. The Chaperone Activity of the Developmental Small Heat Shock Protein Sip1 Is Regulated by pH-Dependent Conformational Changes. *Molecular Cell*, 58(6):1067–1078, 2015.

X. Fu. Chaperone function and mechanism of small heat-shock proteins. *Biochimica et Biophysica Acta - Sin*, 46(5):347–356, 2014.

J. Gault, J. A. C. Donlan, I. Liko, J. T. S. Hopper, K. Gupta, N. G. Housden, W. B. Struwe, M. T. Marty, T. Mize, C. Bechara, Y. Zhu, B. Wu, C. Kleanthous, M. Belov, E. Damoc, A. Makarov, and C. V. Robinson. High-resolution mass spectrometry of small molecules bound to membrane proteins. *Nature Methods*, 13(4):333–336, 2016.

A. Ghahghaei, A. Rekas, J. A. Carver, and R. C. Augusteyn. Structure/function studies of dogfish $\alpha$-crystallin, comparison with bovine $\alpha$-crystallin. *Molecular Vision*, 15:2411–2420, 2009.

A. Ghahghaei, A. Divsalar, and N. Faridi. The effects of molecular crowding on the amyloid fibril formation of $\alpha$-lactalbumin and the chaperone action of $\alpha$-casein. *Protein Journal*, 29(4):257–264, 2010.

J. G. Ghosh, A. K. Shenoy, and J. I. Clark. N- and C-terminal motifs in human $\alpha$B-crystallin play an important role in the recognition, selection, and solubilization of substrates. *Biochemistry*, 45(46):13847–13854, 2006.

K. Giles, S. D. Pringle, K. R. Worthington, D. Little, J. L. Wildgoose, and R. H. Bateman. Applications of a travelling wave-based radio-frequency-only stacked ring ion guide. *Rapid Communications in Mass Spectrometry*, 18(20):2401–2414, 2004.

F. Glover and M. Laguna. *Tabu Search*. Kluwer Academic Publishers, Norwell, MA, 1997.

J. Goers, S. E. Permyakov, E. A. Permyakov, V. N. Uversky, and A. L. Fink. Conformational prerequisites for $\alpha$-lactalbumin fibrillation. *Biochemistry*, 41(41):12546–51, 2002.

D. E. Goldberg. *Genetic Algorithms in Search, Optimization and Machine Learning*. Addison-Wesley Longman Publishing Co., Boston, MA, 1989.

D. Goodsell. Molecule of the Month: DNA. *DOI: 10.2210/rcsb_pdb/mom_2001_11. RCSB PDB*, 2001.

D. Goodsell. Molecule of the Month: Tobacco Mosaic Virus. *DOI: 10.2210/rcsb_pdb/mom_2009_1. RCSB PDB*, 2009.

J. Grove and M. Marsh. The cell biology of receptor-mediated virus entry. *Journal of Cell Biology*, 195(7):1071–1082, 2011.

D. A. Haley, J. Horwitz, and P. L. Stewart. The Small Heat-Shock Protein, αB-Crystallin, has a Variable Quaternary Structure. *Journal of Molecular Biology*, 277:27–35, 1998.

D. Hall, R. Zhao, I. Dehlsen, N. Bloomfield, S. R. Williams, F. Arisaka, Y. Goto, and J. A. Carver. Protein aggregate turbidity: Simulation of turbidity profiles for mixed-aggregation reactions. *Analytical Biochemistry*, 498:78–94, 2016.

P. Hammarstrom. Trans-Suppression of Misfolding in an Amyloid Disease. *Science*, 293(5539):2459–2462, 2001.

Y. Hanazono, K. Takeda, M. Yohda, and K. Miki. Structural studies on the oligomeric transition of a small heat shock protein, StHsp14.0. *Journal of Molecular Biology*, 422(1):100–108, 2012.

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*, 21(2):220–228, 2013.

D. F. Hansen, D. Yang, H. Feng, Z. Zhou, S. Wiesner, Y. Bai, and L. E. Kay. An exchange-free measure of 15N transverse relaxation: An NMR spectroscopy application to the study of a folding intermediate with pervasive chemical exchange. *Journal of the American Chemical Society*, 129(37):11468–11479, 2007.

D. F. Hansen, P. Vallurupalli, P. Lundstroem, P. Neudecker, and L. E. Kay. Probing chemical shifts of invisible states of proteins with relaxation dispersion NMR spectroscopy: How well can we do? *Journal of the American Chemical Society*, 130(8):2667–2675, 2008.

F. U. Hartl, A. Bracher, and M. Hayer-Hartl. Molecular chaperones in protein folding and proteostasis. *Nature*, 475(7356):324–332, 2011.

M. Haslbeck and E. Vierling. A first line of stress defense: Small heat shock proteins and their function in protein homeostasis. *Journal of Molecular Biology*, 427(7): 1537–1548, 2015.

M. Haslbeck, S. Walke, T. Stromer, M. Ehrnsperger, H. E. White, S. Chen, H. R. Saibil, and J. Buchner. Hsp26: a temperature-regulated chaperone. *EMBO Journal*, 18(23):6744–51, 1999.

D. M. Hatters, C. E. MacPhee, L. J. Lawrence, W. H. Sawyer, and G. J. Hewlett. Human apolipoprotein C-II forms twisted amyloid ribbons and closed loops. *Biochemistry*, 39(28):8276–8283, 2000.

D. M. Hatters, R. A. Lindner, J. A. Carver, and G. J. Howlett. The Molecular Chaperone, $\alpha$-Crystallin, Inhibits Amyloid Formation by Apolipoprotein C-II. *Journal of Biological Chemistry*, 276(36):33755–61, 2001.

D. M. Hatters, C. A. MacRaild, R. Daniels, W. S. Gosal, N. H. Thomson, J. A. Jones, J. J. Davis, C. E. MacPhee, C. M. Dobson, and G. J. Howlett. The circularization of amyloid fibrils formed by apolipoprotein C-II. *Biophysical Journal*, 85(6):3979–90, 2003.

K. Henzler-Wildman and D. Kern. Dynamic personalities of proteins. *Nature*, 450 (7172):964–972, 2007.

H. Hernandez and C. V. Robinson. Determining the stoichiometry and interactions of macromolecular assemblies from mass spectrometry. *Nature Protocols*, 2(3): 1479–89, 2007.

J. Heuser. Three-dimensional visualization of coated vesicle formation in fibroblasts. *Journal of Cell Biology*, 84(3):560–83, 1980.

G. R. Hilton, G. K. A. Hochberg, A. Laganowsky, S. I. McGinnigle, A. J. Baldwin, and J. L. P. Benesch. C-terminal interactions mediate the quaternary dynamics of αB-crystallin. *Philosophical transactions of the Royal Society B*, 368:20110405, 2013.

K. M. Hines, J. C. May, J. A. McLean, and L. Xu. Evaluation of Collision Cross Section Calibrants for Structural Analysis of Lipids by Traveling Wave Ion Mobility-Mass Spectrometry. *Analytical Chemistry*, 88(14):7329–7336, 2016.

G. K. A. Hochberg and J. L. P. Benesch. Dynamical structure of αB-crystallin. *Progress in Biophysics and Molecular Biology*, 115(1):11–20, 2014.

G. K. A. 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. *Proceedings of the National Academy of Sciences*, 111(16):E1562–70, 2014.

C. J. Hogan, J. A. Carroll, H. W. Rohrs, P. Biswas, and M. L. Gross. Combined charged residue-field emission model of macromolecular electrospray ionization. *Analytical Chemistry*, 81(1):369–377, 2009.

J. Hong, Q. Jing, and L. Yao. The protein amide 1HN chemical shift temperature coefficient reflects thermal expansion of the N-H···O=C hydrogen bond. *Journal of Biomolecular NMR*, 55(1):71–78, 2013.

J. Horwitz. α-crystallin can function as a molecular chaperone. *Proceedings of the National Academy of Sciences of the United States of America*, 89(21):10449–53, 1992.

J. Horwitz. α-crystallin. *Experimental Eye Research*, 76(2):145–153, 2003.

J. Horwitz. α-crystallin: The quest for a homogeneous quaternary structure. *Experimental Eye Research*, 88(2):190–194, 2009.

J. Horwitz, Q.-L. Huang, L. Ding, and M. P. Bova. Lens $\alpha$-crystallin: Chaperone-like properties. *Methods in Enzymology: Molecular Chaperones*, 290:365–383, 1998.

S. A. Hudson, H. Ecroyd, T. W. Kee, and J. A. Carver. The thioflavin T fluorescence assay for amyloid fibril detection can be biased by the presence of exogenous compounds. *FEBS Journal*, 276(20):5960–5972, 2009.

R. L. Hull, G. T. Westermark, P. Westermark, and S. E. Kahn. Islet amyloid: A critical entity in the pathogenesis of type 2 diabetes. *Journal of Clinical Endocrinology and Metabolism*, 89(8):3629–3643, 2004.

Y. M. Ibrahim, S. V. B. Garimella, S. A. Prost, R. Wojcik, R. V. Norheim, E. S. Baker, I. Rusyn, and R. D. Smith. Development of an Ion Mobility Spectrometry-Orbitrap Mass Spectrometry Platform. *Analytical Chemistry*, 88:12152–60, 2016.

J. V. Iribarne and B. A. Thomson. On the evaporation of small ions from charged droplets. *The Journal of Chemical Physics*, 64(6):2287–2294, 1976.

T. Iwaki, A. Kume-Iwaki, and E. Goldman. Cellular Distribution of aB-Crystallin in Non-Lenticular Tissues. *The Journal of Histochemistry and Cytochemistry*, 38 (1):31–39, 1990.

W. M. Jacobs, T. P. Knowles, and D. Frenkel. Oligomers of Heat-Shock Proteins: Structures That Don't Imply Function. *PLoS Computational Biology*, 12(2):1–14, 2016.

J. T. Jarrett and P. T. Lansbury. Amyloid fibril formation requires a chemically discriminating nucleation event: studies of an amyloidogenic sequence from the bacterial protein OsmB. *Biochemistry*, 31(49):12345–12352, 1992.

P. Jaru-Ampornpan, F. C. Liang, A. Nisthal, T. X. Nguyen, P. Wang, K. Shen, S. L. Mayo, and S. O. Shan. Mechanism of an ATP-independent Protein Disaggregase: II. Distinct molecular interactions drive multiple steps during aggregate disassembly. *Journal of Biological Chemistry*, 288(19):13431–13445, 2013.

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. *Journal of Molecular Biology*, 385(5):1481–97, 2009.

S. Jehle, P. Rajagopal, B. Bardiaux, S. Markovic, R. Kühne, 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. *Nature Structural & Molecular Biology*, 17(9):1037–42, 2010.

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. *Proceedings of the National Academy of Sciences of the United States of America*, 108(16): 6409–14, 2011.

W. Jiao, M. Qian, P. Li, L. Zhao, and Z. Chang. The essential role of the flexible termini in the temperature- responsiveness of the oligomeric state and chaperone-like activity for the polydisperse small heat shock protein IbpB from Escherichia coli. *Journal of Molecular Biology*, 347(4):871–884, 2005.

G. Kaklamanos, E. Aprea, and G. Theodoridis. Mass Spectrometry: Principles and Instrumentation. In *Encyclopedia of Food and Health*, pages 661–668. Elsevier, 2016.

G. Kappé, E. Franck, P. Verschuure, W. C. Boelens, J. A. M. Leunissen, and W. W. de Jong. The human genome encodes 10 α-crystallin-related small heat shock proteins: HspB1-10. *Cell Stress & Chaperones*, 8(1):53–61, 2003.

P. R. Kemper, N. F. Dupuis, and M. T. Bowers. A new, higher resolution, ion mobility mass spectrometer. *International Journal of Mass Spectrometry*, 287: 46–57, 2009.

J. Kennedy and R. Eberhart. Particle Swarm Optimization. *Proceedings of the IEEE International Conference on Neural Networks*, pages 1942–1945, 1995.

S.-H. Kim, K. K. Kim, and R. Kim. Crystal structure of a small heat-shock protein. *Nature*, 394(6693):595–599, 1998.

S. Kirkpatrick and J. R. Gelatt. Optimization by simulated annealing. *Science*, 220 (4598):671–80, 1983.

I. R. Kleckner, P. Gollnick, and M. P. Foster. Mechanisms of allosteric gene regulation by NMR quantification of microsecond-millisecond protein dynamics. *Journal of Molecular Biology*, 415(2):372–381, 2012.

R. Klemenz, E. Fröhli, R. H. Steiger, R. Schäfer, and A. Aoyama. $\alpha$B-crystallin is a small heat shock protein. *Proceedings of the National Academy of Sciences of the United States of America*, 88(9):3652–56, 1991.

J. Klucken, Y. Shin, E. Masliah, B. T. Hyman, and P. J. McLean. Hsp70 reduces alpha-synuclein aggregation and toxicity. *Journal of Biological Chemistry*, 279 (24):25497–25502, 2004.

T. P. J. Knowles, C. A. Waudby, G. L. Devlin, S. I. A. Cohen, A. Aguzzi, M. Vendruscolo, E. M. Terentjev, M. E. Welland, and C. M. Dobson. An analytical solution to the kinetics of breakable filament assembly. *Science*, 326(5959):1533–37, 2009.

T. P. J. Knowles, D. A. White, A. R. Abate, J. J. Agresti, S. I. A. Cohen, R. A. Sperling, E. J. De Genst, C. M. Dobson, and D. A. Weitz. Observation of spatial propagation of amyloid assembly from single nuclei. *Proceedings of the National Academy of Sciences*, 108(36):14746–14751, 2011.

T. P. J. Knowles, M. Vendruscolo, and C. M. Dobson. The amyloid state and its association with protein misfolding diseases. *Nature Reviews. Molecular Cell Biology*, 15(6):384–96, 2014.

F. D. Kondrat, W. B. Struwe, and J. L. P. Benesch. Native mass spectrometry: towards high-throughput structural proteomics. *Methods in Molecular Biology*, 1261:349–71, 2015.

H. A. Koteiche and H. S. McHaourab. Mechanism of chaperone function in small heat-shock proteins. Phosphorylation-induced activation of two-mode binding in $\alpha$B-crystallin. *Journal of Biological Chemistry*, 278(12):10361–7, 2003.

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 Journal*, 24(10):3633–42, 2010.

M. Kulig and H. Ecroyd. The small heat-shock protein $\alpha$B-crystallin uses different mechanisms of chaperone action to prevent the amorphous versus fibrillar aggregation of $\alpha$-lactalbumin. *Biochemical Journal*, 448(3):343–52, 2012.

P. A. Kumar and G. B. Reddy. Modulation of $\alpha$-crystallin chaperone activity: A target to prevent or delay cataract? *IUBMB Life*, 61(5):485–495, 2009.

K. Kuwajima. The molten globule state of $\alpha$-lactalbumin. *FASEB Journal*, 10(1): 102–9, 1996.

A. Laganowsky, J. L. P. Benesch, M. Landau, L. Ding, M. R. Sawaya, D. Cascio, Q. Huang, C. V. Robinson, J. Horwitz, and D. Eisenberg. Crystal structures of truncated $\alpha$A and $\alpha$B crystallins reveal structural mechanisms of polydispersity important for eye lens function. *Protein Science*, 19(5):1031–43, 2010.

M. Lakadamyali, M. J. Rust, and X. Zhuang. Endocytosis of influenza viruses. *Microbes and Infection*, 6(10):929–36, 2004.

M. Landreh, I. Liko, P. Uzdavinys, M. Coincon, J. T. S. Hopper, D. Drew, and C. V. Robinson. Controlling release, unfolding and dissociation of membrane protein complexes in the gas phase through collisional cooling. *Chem. Commun.*, 51:15582–15584, 2015.

G. J. Lee and E. Vierling. A small heat shock protein cooperates with heat shock protein 70 systems to reactivate a heat-denatured protein. *Plant Physiology*, 122 (1):189–98, 2000.

G. J. Lee, A. M. Roseman, H. R. Saibil, and E. Vierling. A small heat shock protein stably binds heat denatured model substrates and can maintain a substrate in a folding competent state. *EMBO Journal*, 16(3):659–71, 1997.

W. Lee, M. Tonelli, and J. L. Markley. NMRFAM-SPARKY: Enhanced software for biomolecular NMR spectroscopy. *Bioinformatics*, 31(8):1325–1327, 2015.

A. C. Leney and A. J. R. Heck. Native Mass Spectrometry: What is in the Name? *Journal of the American Society for Mass Spectrometry*, 28(1):5–13, 2017.

Y. Levin, E. Hradetzky, and S. Bahn. Quantification of proteins using data-independent analysis (MS E) in simple and complex samples: A systematic evaluation. *Proteomics*, 11(16):3273–3287, 2011.

D. S. Libich, N. L. Fawzi, J. Ying, and G. M. Clore. Probing the transient dark state of substrate binding to GroEL by relaxation-based solution NMR. *Proceedings of the National Academy of Sciences*, 110(28):11361–11366, 2013.

R. A. Lindner, A. Kapur, and J. A. Carver. The interaction of the molecular chaperone, alpha-crystallin, with molten globule states of bovine alpha-lactalbumin. *The Journal of Biological Chemistry*, 272(44):27722–9, 1997.

R. A. Lindner, T. M. Treweek, and J. A. Carver. Aggregation To Stabilize a Monomeric Molten-Globule Form of $\alpha$-Lactalbumin. *Biochemical Journal*, 87: 79–87, 2001.

N. Lorenzen, S. B. Nielsen, Y. Yoshimura, B. S. Vad, C. B. Andersen, C. Betzer, J. D. Kaspersen, G. Christiansen, J. S. Pedersen, P. H. Jensen, F. A. A. Mulder, and D. E. Otzen. How epigallocatechin gallate can inhibit $\alpha$-synuclein oligomer toxicity in vitro. *Journal of Biological Chemistry*, 289(31):21299–21310, 2014.

P. Lössl, M. van de Waterbeemd, and A. J. Heck. The diverse and expanding role of mass spectrometry in structural and molecular biology. *EMBO Journal*, 35(24): 2634–2657, 2016.

J. Lowe, H. McDermott, I. Pike, I. Spendlove, M. Landon, and R. J. Mayer. $\alpha$B crystallin expression in non-lenticular tissues and selective presence in ubiquitinated inclusion bodies in human disease. *Journal of Pathology*, 166(1):61–68, 1992.

J. C. Lukesh, M. J. Palte, and R. T. Raines. A potent, versatile disulfide-reducing agent from aspartic acid. *Journal of the American Chemical Society*, 134(9): 4057–4059, 2012.

K. Lundmark, G. Westermark, S. Nystroem, C. Murphy, A. Solomon, and P. Westermark. Transmissibility of systemic amyloidosis by a prion-like mechanism. *Proceedings of the National Academy of Sciences*, 99(10):6979–84, 2002.

L. L. Mack, P. Kralik, A. Rheude, and M. Dole. Molecular Beams of Macroions. II. *Journal of Chemical Physics*, 52(10):4977–86, 1970.

A. Mainz, J. Peschek, M. Stavropoulou, K. C. Back, B. Bardiaux, S. Asami, E. Prade, C. Peters, S. Weinkauf, J. Buchner, and B. Reif. The chaperone $\alpha$B-crystallin uses different interfaces to capture an amorphous and an amyloid client. *Nature Structural & Molecular Biology*, 22(11), 2015.

J. Malmström, M. Beck, A. Schmidt, V. Lange, E. W. Deutsch, and R. Aebersold. Proteome-wide cellular protein concentrations of the human pathogen Leptospira interrogans. *Nature*, 460(7256):762–765, 2010.

M. A. Mansfield and J. O. E. L. Key. Synthesis of the Low Molecular Weight Heat Shock Proteins in Plants. *Plant Physiology*, 84:1007–1017, 1987.

J. Marcoux, P. P. Mangione, R. Porcari, M. T. Degiacomi, G. Verona, G. W. Taylor, S. Giorgetti, S. Raimondi, S. Sanglier-Cianferani, J. L. Benesch, C. Cecconi, M. M. Naqvi, J. D. Gillmore, P. N. Hawkins, M. Stoppini, C. V. Robinson, M. B.

Pepys, and V. Bellotti. A novel mechano-enzymatic cleavage mechanism underlies transthyretin amyloidogenesis. *EMBO Molecular Medicine*, 7(10):1337–1349, 2015.

V. Marjomaki, T. Lahtinen, M. Martikainen, J. Koivisto, S. Malola, K. Salorinne, M. Pettersson, and H. Hakkinen. Site-specific targeting of enterovirus capsid by functionalized monodisperse gold nanoclusters. *Proceedings of the National Academy of Sciences*, 111(4):1277–1281, 2014.

H. M. McConnell. Reaction Rates by Nuclear Magnetic Resonance. *The Journal of Chemical Physics*, 28(3):430–431, 1958.

H. S. Mchaourab, E. K. Dodson, and H. A. Koteiche. Mechanism of chaperone function in small heat shock proteins: Two-mode binding of the excited states of T4 lysozyme mutants by $\alpha$A-crystallin. *Journal of Biological Chemistry*, 277(43): 40557–66, 2002.

A. Mehlkopf, D. Korbee, and T. Tiggelman. Sources of tI Noise in Two-Dimensional. *Journal of Magnetic Resonance*, 58:315–323, 1984.

S. Meiboom and D. Gill. Modified spin-echo method for measuring nuclear relaxation times. *Rev. Sci. Instrum.*, 29(8):688–691, 1958.

G. Meisl, X. Yang, E. Hellstrand, B. Frohm, J. B. Kirkegaard, S. I. A. Cohen, C. M. Dobson, S. Linse, and T. P. J. Knowles. Differences in nucleation behavior underlie the contrasting aggregation kinetics of the A$\beta$40 and A$\beta$42 peptides. *Proceedings of the National Academy of Sciences of the United States of America*, 111(26): 9384–9, 2014.

G. Meisl, J. B. Kirkegaard, P. Arosio, T. C. T. Michaels, M. Vendruscolo, C. M. Dobson, S. Linse, and T. P. J. Knowles. Molecular mechanisms of protein aggregation from global fitting of kinetic models. *Nature Protocols*, 11(2):252–272, 2016.

C. J. Merrifield and M. Kaksonen. Endocytic Accessory Factors and Regulation of Clathrin-Mediated Endocytosis. *Cold Spring Harbor Perspectives in Biology*, pages 1–16, 2014.

R. Mishra, M. Geyer, and R. Winter. NMR spectroscopic investigation of early events in IAPP amyloid fibril formation. *ChemBioChem*, 10(11):1769–1772, 2009.

C. Mok and S. G. Waley. Structural Studies on Lens Proteins. *Biochemical Journal*, 108:128–34, 1967.

K. L. Moreau and J. A. King. Protein misfolding and aggregation in cataract disease and prospects for prevention. *Trends in Molecular Medicine*, 18(5):273–282, 2012.

H. N. Motlagh, J. O. Wrabl, J. Li, and V. J. Hilser. The ensemble nature of allostery. *Nature*, 508(7496):331–339, 2014.

P. Narayan, A. Orte, R. W. Clarke, B. Bolognesi, S. Hook, K. A. Ganzinger, S. Meehan, M. R. Wilson, C. M. Dobson, and D. Klenerman. The extracellular chaperone clusterin sequesters oligomeric forms of the amyloid-beta 1-40 peptide. *Nature Structural & Molecular Biology*, 19(1):79–83, 2012.

S. Nguyen and J. B. Fenn. Gas-phase ions of solute species from charged droplets of solutions. *Proceedings of the National Academy of Sciences of the United States of America*, 104(4):1111–1117, 2007.

Y. Nishiyama, H. Taguchi, M. Hara, S. A. Planque, Y. Mitsuda, and S. Paul. Metal-dependent amyloid $\alpha$-degrading catalytic antibody construct. *Journal of Biotechnology*, 180:17–22, 2014.

J. E. Ollerenshaw, V. Tugarinov, and L. E. Kay. Methyl TROSY: explanation and experimental verification. *Magnetic Resonance in Chemistry*, 41(10):843–852, 2003.

B. Olson, I. Hashmi, K. Molloy, and A. Shehu. Basin Hopping as a General and

Versatile Optimization Framework for the Characterization of Biological Macromolecules. *Advances in Artificial Intelligence*, 2012:1–19, 2012.

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*, 144(1):67–78, 2011.

F. Oosawa and M. Kasai. A theory of linear and helical aggregations of macromolecules. *Journal of Molecular Biology*, 4(1):10–21, 1962.

K. Pagel, S.-J. Hyung, B. T. Ruotolo, and C. V. Robinson. Alternate dissociation pathways identified in charge-reduced protein complex ions. *Analytical Chemistry*, 82(12):5363–72, 2010.

A. G. Palmer. Chemical exchange in biomacromolecules: Past, present, and future. *Journal of Magnetic Resonance*, 241(1):3–17, 2014.

A. G. Palmer, C. D. Kroenke, and J. P. Loria. Nuclear magnetic resonance methods for quantifying microsecond-to-millisecond motions in biological macromolecules. *Methods in Enzymology*, 339(1997):204–238, 2001.

S. Y. Pasta, B. Raman, T. Ramakrishna, and C. M. Rao. The IXI/V motif in the C-terminal extension of alpha-crystallins: alternative interactions and oligomeric assemblies. *Molecular vision*, 10(May):655–62, 2004.

J. Peschek, N. Braun, J. Rohrberg, K. C. Back, T. Kriehuber, A. Kastenmüller, S. Weinkauf, and J. Buchner. Regulated structural transitions unleash the chaperone activity of $\alpha$B-crystallin. *Proceedings of the National Academy of Sciences*, 108(51):20491–6, 2013.

L. B. Poole. The basics of thiols and cysteines in redox biology and chemistry. *Free Radical Biology and Medicine*, 80:148–157, 2015.

W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical Recipes 3rd Edition: The Art of Scientific Computing*. Cambridge University Press, New York, 3 edition, 2007.

S. W. Provencher. Contin: A General Purpose Constrained Regularization Program for Inverting Noisy Linear Algebraic and Integral Equations. *Computer Physics Communications*, 27:229–242, 1982.

H. Puchtler, F. Sweat, and M. Levine. On the Binding of Congo Red By Amyloid. *Journal of Histochemistry & Cytochemistry*, 10(3):355–364, 1962.

P. Rajagopal, E. Tse, A. J. Borst, S. P. Delbecq, L. Shi, D. R. Southworth, and R. E. Klevit. A conserved histidine modulates HSPB5 structure to trigger chaperone activity in response to stress-related acidosis. *eLife*, 4:1–21, 2015.

K. Rajaraman, B. Raman, T. Ramakrishna, and C. Rao. The chaperone-like $\alpha$-crystallin forms a complex only with the aggregation-prone molten globule state of $\alpha$-lactalbumin. *Biochemical and Biophysical Research Communications*, 249 (3):917–921, 1998.

B. Raman, T. Ban, M. Sakai, S. Y. Pasta, T. Ramakrishna, H. Naiki, Y. Goto, and C. M. Rao. AlphaB-crystallin, a small heat-shock protein, prevents the amyloid fibril growth of an amyloid beta-peptide and beta2-microglobulin. *Biochemical Journal*, 392(Pt 3):573–581, 2005.

M. Ramirez-Alvarado, J. S. Merkel, and L. Regan. A systematic exploration of the influence of the protein stability on amyloid fibril formation in vitro. *Proceedings of the National Academy of Sciences*, 97(16):8979–8984, 2000.

P. V. Rao, J. Horwitz, and J. S. Zigler. $\alpha$-crystallin, a molecular chaperone, forms a stable complex with carbonic anhydrase upon heat denaturation. *Biochemical and Biophysical Research Communications*, 190(3):786–793, 1993.

G. B. Reddy, K. P. Das, J. M. Petrash, and W. K. Surewicz. Temperature-dependent

Chaperone Activity and Structural Properties of Human αA- and αB-crystallins. *Journal of Biological Chemistry*, 275(7):4565–70, 2000.

V. S. Reddy and G. B. Reddy. Emerging role for αB-crystallin as a therapeutic agent: pros and cons. *Current Molecular Medicine*, 15(1):47–61, 2015.

J. W. Regini, H. Ecroyd, S. Meehan, K. Bremmell, M. J. Clarke, D. Lammie, T. Wess, and J. A. Carver. The interaction of unfolding α-lactalbumin and malate dehydrogenase with the molecular chaperone αB-crystallin: a light and X-ray scattering investigation. *Molecular Vision*, 16:2446–56, 2010.

A. Rekas, C. G. Adda, A. J. 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. *Journal of Molecular Biology*, 340(5):1167–83, 2004.

T. L. Religa, R. Sprangers, and L. E. Kay. Dynamic Regulation of Archaeal Proteasome Gate Opening As Studied by TROSY NMR. *Science*, 328(5974):98–102, 2010.

K. Renkawek, C. E. Voorter, G. J. Bosman, F. P. van Workum, and W. W. de Jong. Expression of αB-crystallin in Alzheimer's disease. *Acta Neuropathologica*, 87(2): 155–60, 1994.

R. A. Resnik. Lens proteins I. α-crystallin of calf lens. *American Journal of Ophthalmology*, 44(5):357–62, 1957.

H. E. Revercomb and E. A. Mason. Theory of Plasma Chromatography Gaseous Electrophoresis - Review. *Anal Chem*, 47(7):970–983, 1975.

R. J. Rose, E. Damoc, E. Denisov, A. Makarov, and A. J. R. Heck. High-sensitivity Orbitrap mass analysis of intact macromolecular assemblies. *Nature Methods*, 9 (11):2–6, 2012.

H. I. Rösner and C. Redfield. The Human $\alpha$-Lactalbumin Molten Globule: Comparison of Structural Preferences at pH 2 and pH 7. *Journal of Molecular Biology*, 394(2):351–62, 2009.

C. A. Ross. Polyglutamine Pathogenesis : Minireview. Emergence of Unifying Mechanisms for Huntington's Disease and Related Disorders. *Neuron*, 35:819–22, 2002.

B. T. Ruotolo, K. Giles, I. Campuzano, A. M. Sandercock, R. H. Bateman, and C. V. Robinson. Evidence for Macromolecular Protein Rings in the Absence of Bulk Water. *Science*, 310(5754):1658–1661, 2005.

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. *Nature protocols*, 3(7):1139–1152, 2008.

R. Sarroukh, E. Goormaghtigh, J. M. Ruysschaert, and V. Raussens. ATR-FTIR: A "rejuvenated" tool to investigate amyloid proteins. *Biochimica et Biophysica Acta*, 1828(10):2328–38, 2013.

H. A. Sathish, R. A. Stein, G. Yang, and H. S. McHaourab. Mechanism of chaperone function in small heat-shock proteins. Fluorescence studies of the conformations of T4 lysozyme bound to $\alpha$B-crystallin. *Journal of Biological Chemistry*, 278(45): 44214–21, 2003.

M. R. Sawaya, S. Sambashivan, R. Nelson, M. I. Ivanova, S. A. Sievers, M. I. Apostol, M. J. Thompson, M. Balbirnie, J. J. W. Wiltzius, H. T. McFarlane, A. Ø. Madsen, C. Riekel, and D. Eisenberg. Atomic structures of amyloid cross-$\beta$ spines reveal varied steric zippers. *Nature*, 447(7143):453–457, 2007.

B. A. Schulman, P. S. Kim, C. M. Dobson, and C. Redfield. A residue-specific NMR view of the non-cooperative unfolding of a molten globule. *Nature Structural Biology*, 4(8):686–689, 1997.

A. Scotti, W. Liu, J. S. Hyatt, E. S. Herman, H. S. Choi, J. W. Kim, L. A. Lyon, U. Gasser, and A. Fernandez-Nieves. The CONTIN algorithm and its application

to determine the size distribution of microgel suspensions. *The Journal of chemical physics*, 142(23):234905, 2015.

T. R. Serio, A. G. Cashikar, A. S. Kowal, G. J. Sawicki, J. J. Moslehi, L. Serpell, M. F. Arnsdorf, and S. L. Lindquist. Nucleated Conformational Conversion and the Replication of Conformational Information by a Prion Determinant. *Science*, 289(5483):1317–1321, 2000.

S. L. Shammas, C. A. Waudby, S. Wang, A. K. Buell, T. P. J. Knowles, H. Ecroyd, M. E. Welland, J. A. Carver, C. M. Dobson, and S. Meehan. Binding of the molecular chaperone $\alpha$B-crystallin to A$\beta$ amyloid fibrils inhibits fibril elongation. *Biophysical Journal*, 101(7):1681–9, 2011.

K. K. Sharma, R. S. Kumar, G. S. Kumar, and P. T. Quinn. Synthesis and characterization of a peptide identified as a functional element in $\alpha$A-crystallin. *Journal of Biological Chemistry*, 275(6):3767–71, 2000.

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. *International Journal of Mass Spectrometry*, 377:663–671, 2015.

M. Siddique, S. Gernhard, P. Von Koskull-Döring, E. Vierling, and K. D. Scharf. The plant sHSP superfamily: Five new members in Arabidopsis thaliana with unexpected properties. *Cell Stress and Chaperones*, 13(2):183–197, 2008.

R. J. Siezen and E. A. Owen. Physicochemical characterization of high-molecular-weight $\alpha$-crystallin subpopulations from the calf lens nucleus. *Biochimica et Biophysica Acta*, 749(3):227–37, 1983.

P. Sikorski, E. D. T. Atkins, and L. C. Serpell. Structure and texture of fibrous crystals formed by Alzheimer's A$\beta$(11-25) peptide fragment. *Structure*, 11(8): 915–926, 2003.

D. Sinnaeve. The Stejskal–Tanner Equation Generalized for Any Gradient Shape—An Overview of Most Pulse Sequences Measuring Free Diffusion. *Concepts in Magnetic Resonance*, 40A(2):39–65, 2012.

E. Smirnova, N. Chebotareva, and B. Gurvits. Transient transformation of oligomeric structure of $\alpha$-crystallin during its chaperone action. *International Journal of Biological Macromolecules*, 55:62–8, 2013.

A. M. Smith, T. R. Jahn, A. E. Ashcroft, and S. E. Radford. Direct Observation of Oligomeric Species formed in the Early Stages of Amyloid Fibril Formation using Electrospray Ionisation Mass Spectrometry. *Journal of Molecular Biology*, 364 (1):9–19, 2006.

F. Sobott, J. L. P. Benesch, E. Vierling, and C. V. Robinson. Subunit Exchange of Multimeric Protein Complexes. *Journal of Biological Chemistry*, 277(41):38921–9, 2002.

A. Spector. The Aging of $\alpha$-Crystallin: A Review. *Experimental Eye Research*, 16: 115–21, 1973.

A. Spector and E. Katz. The Deaggregation of Bovine Lens $\alpha$-Crystallin. *Journal of Biological Chemistry*, 240(5):1979–85, 1965.

A. Spector, L. K. Li, R. C. Augusteyn, A. Schneider, and T. Freund. $\alpha$-Crystallin: The isolation and characterization of distinct macromolecular fractions. *Biochemical Journal*, 124:337–43, 1971.

M. Stefani and C. M. Dobson. Protein aggregation and aggregate toxicity: New insights into protein folding, misfolding diseases and biological evolution. *Journal of Molecular Medicine*, 81(11):678–699, 2003.

E. O. Stejskal and J. E. Tanner. Spin diffusion measurements: spin echoes in the presence of a time-dependent field gradient. *The Journal of Chemical Physics*, 42 (1):288–292, 1965.

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. *Proceedings of the National Academy of Sciences*, 107(5):2007–12, 2010.

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. *Biophysical Journal*, 55(6):1155–1167, 1989.

D. Stock, A. G. Leslie, and J. E. Walker. Molecular architecture of the rotary motor in ATP synthase. *Science*, 286(5445):1700–5, 1999.

S. D. Stranks, H. Ecroyd, S. Van Sluyter, E. J. Waters, J. A. Carver, and L. Von Smekal. Model for amorphous aggregation processes. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, 80(5), 2009.

A. Strauch and M. Haslbeck. The function of small heat-shock proteins and their implication in proteostasis. *Essays In Biochemistry*, 60(2):163–172, 2016.

S. Studer and F. Narberhaus. Chaperone activity and homo- and hetero-oligomer formation of bacterial small heat shock proteins. *Journal of Biological Chemistry*, 275(47):37212–8, 2000.

A. S. P. Svane, K. Jahn, T. Deva, A. Malmendal, D. E. Otzen, J. Dittmer, and N. C. Nielsen. Early stages of amyloid fibril formation studied by liquid-state NMR: the peptide hormone glucagon. *Biophysical Journal*, 95(1):366–77, 2008.

K. Tanaka, H. Waki, Y. Ido, S. Akita, Y. Yoshida, T. Yoshida, and T. Matsuo. Protein and polymer analyses up to m/z 100 000 by laser ionization time-of-flight mass spectrometry. *Rapid Communications in Mass Spectrometry*, 2(8):151–153, 1988.

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*, 2(10):e1046, 2007.

T. M. Treweek, S. Meehan, H. Ecroyd, and J. A. Carver. Small heat-shock proteins: Important players in regulating cellular proteostasis. *Cellular and Molecular Life Sciences*, 72(3):429–451, 2015.

S. Ungelenk, F. Moayed, C.-T. Ho, T. Grousl, A. Scharf, A. Mashaghi, S. Tans, M. P. Mayer, A. Mogk, and B. Bukau. Small heat shock proteins sequester misfolding proteins in near-native conformation for cellular protection and efficient refolding. *Nature Communications*, 7:13673, 2016.

V. Uversky and D. Eliezer. Biophysics of Parkinsons Disease: Structure and Aggregation of α-Synuclein. *Current Protein & Peptide Science*, 10(5):483–99, 2009.

P. J. van den Oetelaar, P. F. van Someren, H. J. Hoenders, J. A. Thomson, R. J. Siezen, P. J. van den Oetelaar, and R. J. Siezen. A Dynamic Quaternary Structure of Bovine α-Crystallin As Indicated from Intermolecular Exchange of Subunits. *Biochemistry*, 29(14):3488–3493, 1990.

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. *Nature Structural Biology*, 8(12):1025–30, 2001.

R. L. van Montfort, C. Slingsby, and E. Vierling. Structure and function of the small heat shock protein/α-crystallin family of molecular chaperones. *Advances in Protein Chemistry*, 59:105–56, 2002.

P. H. M. Van Steenberge, D. R. D'hooge, M. F. Reyniers, and G. B. Marin. Improved kinetic Monte Carlo simulation of chemical composition-chain length distributions in polymerization processes. *Chemical Engineering Science*, 110:185–199, 2014.

P. S. Vassar and C. F. A. Culling. Fluorescent stains, with special reference to amyloid and connective tissues. *AMA Archives of Pathology*, 68:487–98, 1959.

V. Vouk. Projected Area of Convex Bodies. *Nature*, (4116):439–40, 1948.

D. M. Walsh, I. Klyubin, J. V. Fadeeva, W. K. Cullen, R. Anwyl, M. S. Wolfe, M. J. Rowan, and D. J. Selkoe. Naturally secreted oligomers of amyloid beta protein potently inhibit hippocampal long-term potentiation in vivo. *Nature*, 416(6880): 535–539, 2002.

M. T. Walsh, A. C. Sen, and B. Chakrabarti. Micellar subunit assembly in a three-layer model of oligomeric $\alpha$-crystallin. *Journal of Biological Chemistry*, 266(30): 20079–20084, 1991.

X. Wang, R. Wang, C. Ma, X. Shi, Z. Liu, Z. Wang, Q. Sun, J. Cao, and S. Xu. Massive expansion and differential evolution of small heat shock proteins with wheat (Triticum aestivum L.) polyploidization. *Scientific Reports*, 7(1):2581, 2017.

C. A. Waudby, T. P. J. 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 $\alpha$B-crystallin with mature $\alpha$-synuclein amyloid fibrils inhibits their elongation. *Biophysical Journal*, 98(5):843–851, 2010.

G. M. Whitesides, J. E. Lilburn, and R. P. Szajewski. Rates of thiol-disulfide interchange reactions between mono- and dithiols and Ellman's Reagent. *Journal of Organic Chemistry*, 42(2):332–8, 1977.

R. B. Wickner, H. K. Edskes, F. Shewmaker, and T. Nakayashiki. Prions of fungi: inherited structures and biological roles. *Nature Reviews Microbiology*, 5(8):611–618, 2007.

D. K. Wilkins, S. B. Grimshaw, V. Receveur, C. M. Dobson, J. A. Jones, and L. J. Smith. Hydrodynamic radii of native and denatured proteins measured by pulse field gradient NMR techniques. *Biochemistry*, 38(50):16424–16431, 1999.

B. Winner, R. Jappelli, S. K. Maji, P. A. Desplats, L. Boyer, S. Aigner, C. Hetzer, T. Loher, M. Vilar, S. Campioni, C. Tzitzilonis, A. Soragni, S. Jessberger,

H. Mira, A. Consiglio, E. Pham, E. Masliah, F. H. Gage, and R. Riek. In vivo demonstration that alpha-synuclein oligomers are toxic. *Proceedings of the National Academy of Sciences*, 108(10):4194–4199, 2011.

P. Wolfe. Convergence conditions for ascent methods. *SIAM Review*, 11(2):226–235, 1969.

L. C. Wu, Z. Y. Peng, and P. S. Kim. Bipartite structure of the $\alpha$-lactalbumin molten globule. *Nature Structural Biology*, 2(11):983–9, 1995.

H. Xu, P. H. Hsu, L. Zhang, M. D. Tsai, and M. A. Freitas. Database search algorithm for identification of intact cross-links in proteins and peptides using tandem mass spectrometry. *Journal of Proteome Research*, 9(7):3384–3393, 2010.

C. Xue, T. Y. Lin, D. Chang, and Z. Guo. Thioflavin T as an amyloid dye: fibril quantification, optimal concentration and effect on aggregation. *Royal Society Open Science*, 4(1), 2017.

T. Yamazaki, W. Lee, C. H. Arrowsmith, D. R. Muhandiram, and L. E. Kay. A Suite of Triple Resonance NMR Experiments for the Backbone Assignment of 15N, 13C, 2H Labeled Proteins with High Sensitivity. *Journal of the American Chemical Society*, 116:11655–11666, 1994.

R. Yan and R. Vassar. Targeting the $\beta$ secretase BACE1 for Alzheimer's disease therapy. *The Lancet Neurology*, 13(3):319–329, 2014.

B. Yang, G. Stjepanovic, Q. Shen, A. Martin, and J. H. Hurley. Vps4 disassembles an ESCRT-III filament by global unfolding and processive translocation. *Nature Structural & Molecular Biology*, 22(6):492–498, 2015.

H. Yang, S. Huang, H. Dai, Y. Gong, and C. Zheng. The Mycobacterium tuberculosis small heat shock protein Hsp16 . 3 exposes hydrophobic surfaces at mild conditions: Conformational flexibility and molecular chaperone activity. *Protein Science*, 8:174–9, 1999.

L. M. Young, J. C. Saunders, R. A. Mahood, C. H. Revill, R. J. Foster, L.-H. Tu, D. P. Raleigh, S. E. Radford, and A. E. Ashcroft. Screening and classifying small-molecule inhibitors of amyloid formation using ion mobility spectrometry-mass spectrometry. *Nature chemistry*, 7(1):73–81, 2015.

Y. Q. Zhang and K. D. Sarge. Celastrol inhibits polyglutamine aggregation and toxicity though induction of the heat shock response. *Journal of Molecular Medicine*, 85(12):1421–1428, 2007.

# Appendix A

# Supplementary figures and equations

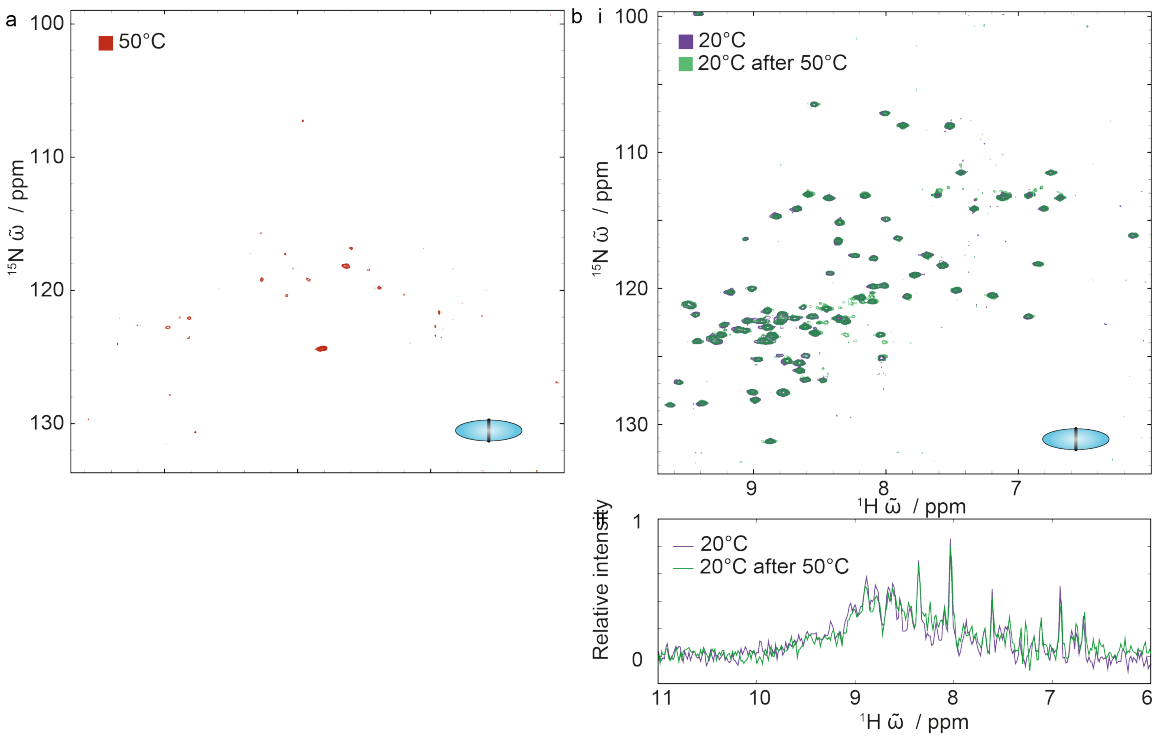



***Figure A.1:*** *$^{15}N^{1}H$ HSQC spectrum of cαBC at 50°C.* ***a*** *$^{15}N^{1}H$ HSQC spectrum of cαBC at 50°C, showing loss of most peaks.* ***b i*** *Spectra at 20°C before and after the 50°C treatment.* ***ii*** *The proton projections of the spectra in* ***i****. Recoverable intensity of peaks shows that the change occurring at 50°C must be reversible.*

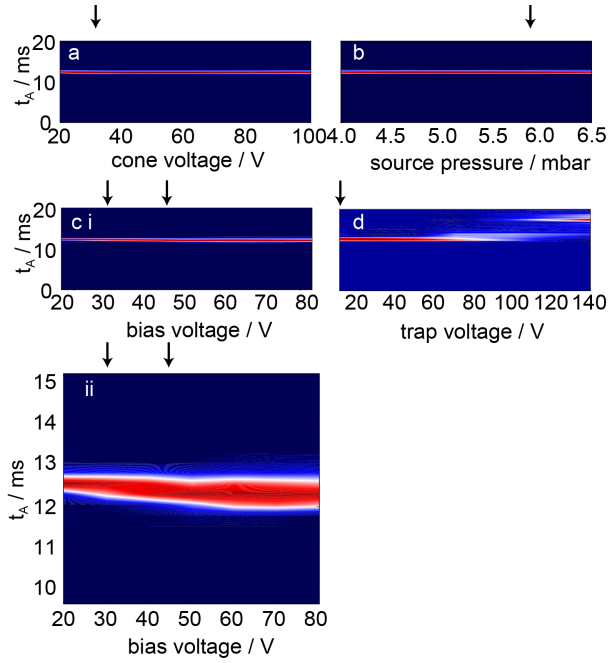



***Figure A.2: Activation checks for Hsp16.5.*** *Arrival time ($t_A$) plots as a function of various relevant voltages on the drift tube instrument prior to CCS measurement of Hsp16.5, shown in* ***Figure 3.19****. The arrows indicate the values subsequently used in the CCS measurement experiment. A minor change of arrival time was observed with low values of bias. This does not correspond to activation since there are no further changes in $t_A$ even for very high bias voltages. Therefore CCS was measured at two representative values of bias voltage (30 V and 45 V) and their average was taken.*

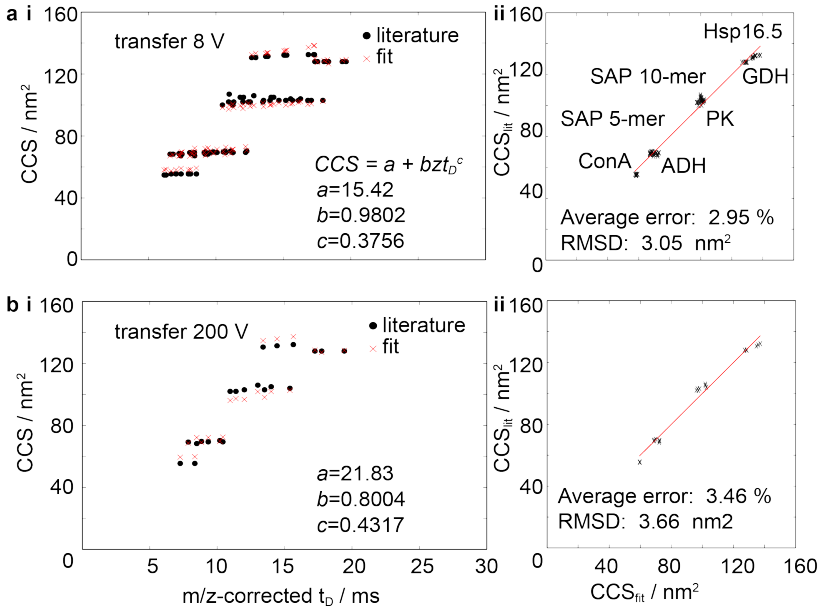



***Figure A.3: CCS calibration at 8 V and 200 V.*** *Comparison of CCS calibrant measurements at 8 V (**a**, standard methods (Ruotolo et al. [2008])) and at 200 V using post-IM CID (**b**). Panels (**i**) show the literature and best-fit value of CCS as a function of $t_D$, the corrected drift time, for each charge state and calibrant recorded. The fitted values were calculated using the relationship $CCS = a + bzt_D^c$, where z is charge and a, b, c are calibration parameters. Panels (**ii**) show the correlation between the literature and fitted CCS values. The quality of the correlation is comparable at the two voltages but the calibration parameters vary.*

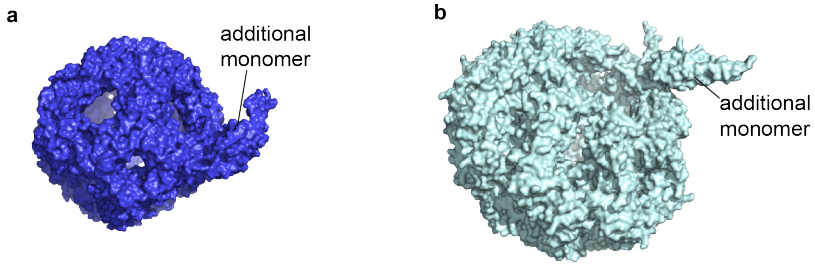



***Figure A.4: Adding exposed monomers to Hsp16.5 and Sip1.*** *Images of structures with an additional unpaired monomer added using crystal packing symmetry. **a** Hsp16.5, PDB ID 1SHS, **b** Sip1, PDB ID 4YDZ.*

## A.1 Kinetic formulation of the best-fit model for $\alpha$Lac only

The following differential equations describe the optimal model describing aggregation of $\alpha$Lac. Here $A*$ denotes the first misfolded species and $A$ denotes the second misfolded species.

Monomer misfolding forward:

$$\frac{d[A*]}{dt} = -k^{+}_{misf}[A*] \tag{A.1}$$

$$\frac{d[A]}{dt} = +k^{+}_{misf}[A*] \tag{A.2}$$

Monomer misfolding reverse:

$$\frac{d[A*]}{dt} = +k^{-}_{misf}[A] \tag{A.3}$$

$$\frac{d[A]}{dt} = -k^{-}_{misf}[A] \tag{A.4}$$

Nucleation 1 forward:

$$\frac{d[A]}{dt} = -2k^{+}_{nuc1}[A]^2 \tag{A.5}$$

$$\frac{d[A_2]}{dt} = +k^{+}_{nuc1}[A]^2 \tag{A.6}$$

Nucleation 1 reverse:

$$\frac{d[A]}{dt} = +2k^{-}_{nuc1}[A_2] \tag{A.7}$$

$$\frac{d[A_2]}{dt} = -k^{-}_{nuc1}[A_2] \tag{A.8}$$

Elongation 1 forward ($2 > i > 2,000$):

$$\frac{d[A]}{dt} = -i^{2/3}k_{el1}^{+}[A][A_{i-1}] \tag{A.9}$$

$$\frac{d[A_{i-1}]}{dt} = -i^{2/3}k_{el1}^{+}[A][A_{i-1}] \tag{A.10}$$

$$\frac{d[A_i]}{dt} = +i^{2/3}k_{el1}^{+}[A][A_{i-1}] \tag{A.11}$$

Elongation 1 reverse:

$$\frac{d[A]}{dt} = +i^{2/3}k_{el1}^{-}[A_i] \tag{A.12}$$

$$\frac{d[A_{i-1}]}{dt} = +i^{2/3}k_{el1}^{-}[A_i] \tag{A.13}$$

$$\frac{d[A_i]}{dt} = -i^{2/3}k_{el1}^{-}[A_i] \tag{A.14}$$

The factor $i^{2/3}$ in the elongation reactions reflects spherical aggregate geometry (rates are proportional to aggregate surface area, which scales with mass as $1^{2/3}$). Analogous nucleation and elongation equations were included to describe the second parallel aggregation pathway starting from $A$ with the corresponding rates $k_{nuc2}^{+/-}$ and $k_{el1}^{+/-}$, with the corresponding oligomer concentrations (denoted as $A_2'$, $A_3'$ etc.) stored as separate matrix entries.

## A.2 Kinetic formulation of the best-fit model for $\alpha$Lac with $\alpha$BC

The additional differential equations describing the interaction of $\alpha$BC with $\alpha$Lac in the optimal model, number 26 in **Figure 5.7**, were as follows. $C$ denotes chaperone and $\alpha$Lac notation is as in the section above. Note that direct monomer binding was not required in this model but chaperone-bound monomers can still be generated by nucleus disassembly.

Monomer release:

$$\frac{d[A]}{dt} = +k^{-}_{mon}[CA] \tag{A.15}$$
$$\frac{d[C]}{dt} = +k^{-}_{mon}[CA] \tag{A.16}$$
$$\frac{d[CA]}{dt} = -k^{-}_{mon}[CA] \tag{A.17}$$

Nucleus 1 binding:

$$\frac{d[A_2]}{dt} = -i^{2/3}k^{+}_{nucb}[C][A_2] \tag{A.18}$$
$$\frac{d[C]}{dt} = -2k^{+}_{nucb}[C][A_2] \tag{A.19}$$
$$\frac{d[C_2A_2]}{dt} = +k^{+}_{nucb}[C][A_2] \tag{A.20}$$

Nucleus 1 disassembly:

$$\frac{d[C_2A_2]}{dt} = -k^{-}_{dis}[C_2A_2] \tag{A.21}$$
$$\frac{d[CA]}{dt} = +2k^{-}_{dis}[C_2A_2] \tag{A.22}$$

Nucleus 2 binding and disassembly was treated analogously with the identical rate constants, $k^{+}_{nucb}$ and $k^{-}_{dis}$. Aggregate 1 binding:

$$\frac{d[A_i]}{dt} = -k^{+}_{agg1}[C][A_i] \tag{A.23}$$
$$\frac{d[C]}{dt} = -ik^{+}_{agg1}[C][A_i] \tag{A.24}$$
$$\frac{d[C_iA_i]}{dt} = +k^{+}_{agg1}[C][A_i] \tag{A.25}$$

Aggregate 1 release:

$$\frac{d[A_i]}{dt} = +k^-_{agg1}[C_iA_i] \tag{A.26}$$

$$\frac{d[C]}{dt} = +ik^-_{agg1}[C_iA_i] \tag{A.27}$$

$$\frac{d[C_iA_i]}{dt} = -k^-_{agg1}[C_iA_i] \tag{A.28}$$

Aggregate 2 binding and release were treated analogously with distinct rate constants, $k^+_{agg2}$ and $k^-_{agg2}$.

## A.3 Kinetic formulation of the best-fit model for $\alpha$Lac with c$\alpha$BC

The optimal model for the interaction of c$\alpha$BC with $\alpha$Lac, model number 13 in **Figure 5.7**, was implemented by analogy to the model for full-length $\alpha$BC in the previous section, with the following differences. Explicit binding of both misfolded monomer forms was permitted:

Monomer $A^*$ binding:

$$\frac{d[A^*]}{dt} = -k^+_{mon}[C][A^*] \tag{A.29}$$

$$\frac{d[C]}{dt} = -k^+_{mon}[C][A^*] \tag{A.30}$$

$$\frac{d[CA^*]}{dt} = +k^+_{mon}[C][A^*] \tag{A.31}$$

$$\tag{A.32}$$

Monomer $A^*$ release:

$$\frac{d[A^*]}{dt} = +k^-_{mon}[CA^*] \tag{A.33}$$

$$\frac{d[C]}{dt} = +k^-_{mon}[CA^*] \tag{A.34}$$

$$\frac{d[CA^*]}{dt} = -k^-_{mon}[CA^*] \tag{A.35}$$

Monomer $A$ was treated analogously with the same rate constants, $k^+_{mon}$ and $k^-_{mon}$.

Nucleus disassembly and aggregate binding was implemented as above, except both sets of $\alpha$Lac aggregates were bound and released by c$\alpha$BC with the same rates, $k^+_{agg}$ and $k^-_{agg}$.

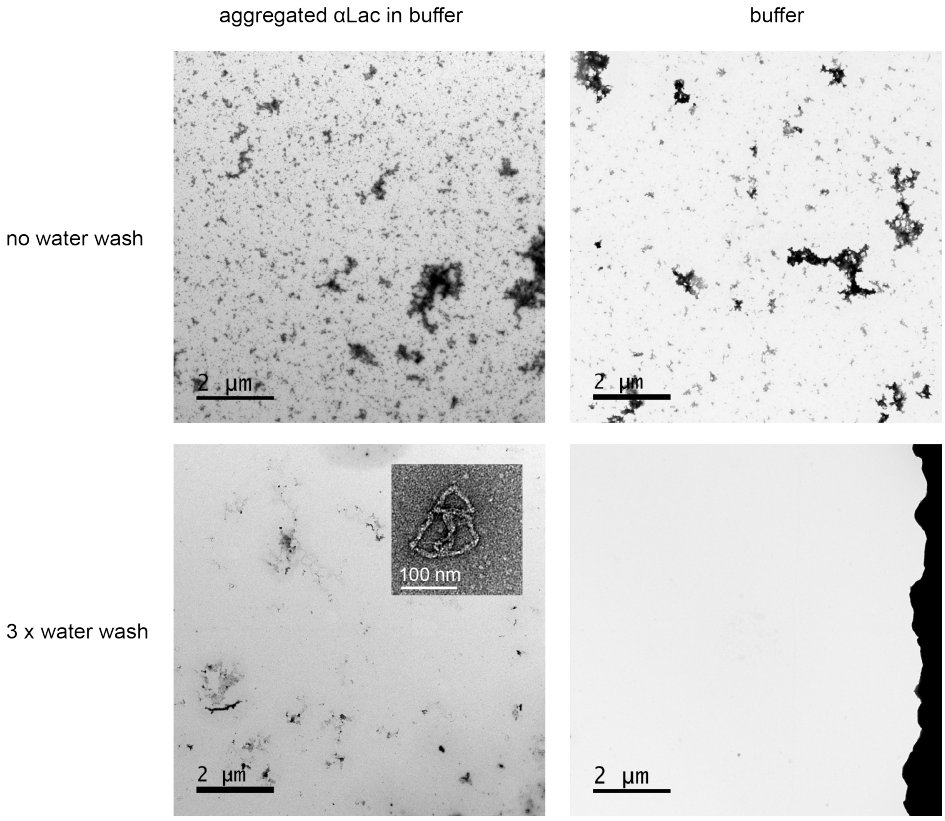



***Figure A.5: Comparison of images with and without a water wash in the EM grid preparation protocol.*** *The standard protocol for preparation of grid involves glow-discharging the grid, application of the sample, wicking off the buffer, application of uranyl acetate stain, wicking off the stain and drying the grid. I found that this protocol resulted in observation of features that could easily be confused with amorphous aggregates but in fact also appeared in soluble protein controls. I tested whether this effect depended on the identity of the protein sample, freshness of the buffer, repeated filtering of the buffer, or was in fact an artifact of the grids. I determined that the aggregate-like structures came from the buffer, i.e. they were most likely to be salt crystals, and could be easily removed with a water wash step in the grid preparation protocol. All images shown in the main text of this thesis were subsequently performed with this optimised protocol. This case illustrates the great care that must be taken when working with poorly defined aggregating samples as misleading results can easily be obtained.*

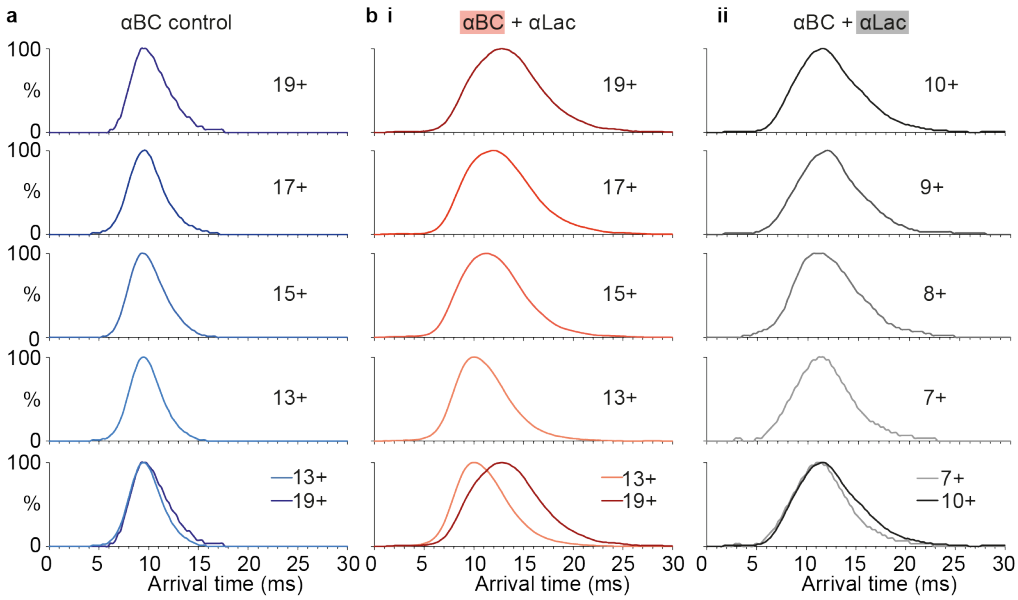



***Figure A.6: Detailed analysis of monomer CID products in $\alpha$BC-$\alpha$Lac mixtures.*** *Arrival time distributions for individual charge states of monomer dissociation products in (**a**) $\alpha$BC alone and (**b**) a mixture of $\alpha$BC and glsaLac, extracted from the 2D spectra in Figure 6.12. There is some variation in the distributions with charge state, notably in the case of $\alpha$BC in the mixture sample, shown in the bottom panels of Figure A.6. This could indicate that the higher charge state products predominantly come from complexes between $\alpha$BC and $\alpha$Lac, which are larger and therefore carry more charge. Each unfolded monomer carries away about half of the total charge on dissociation.*

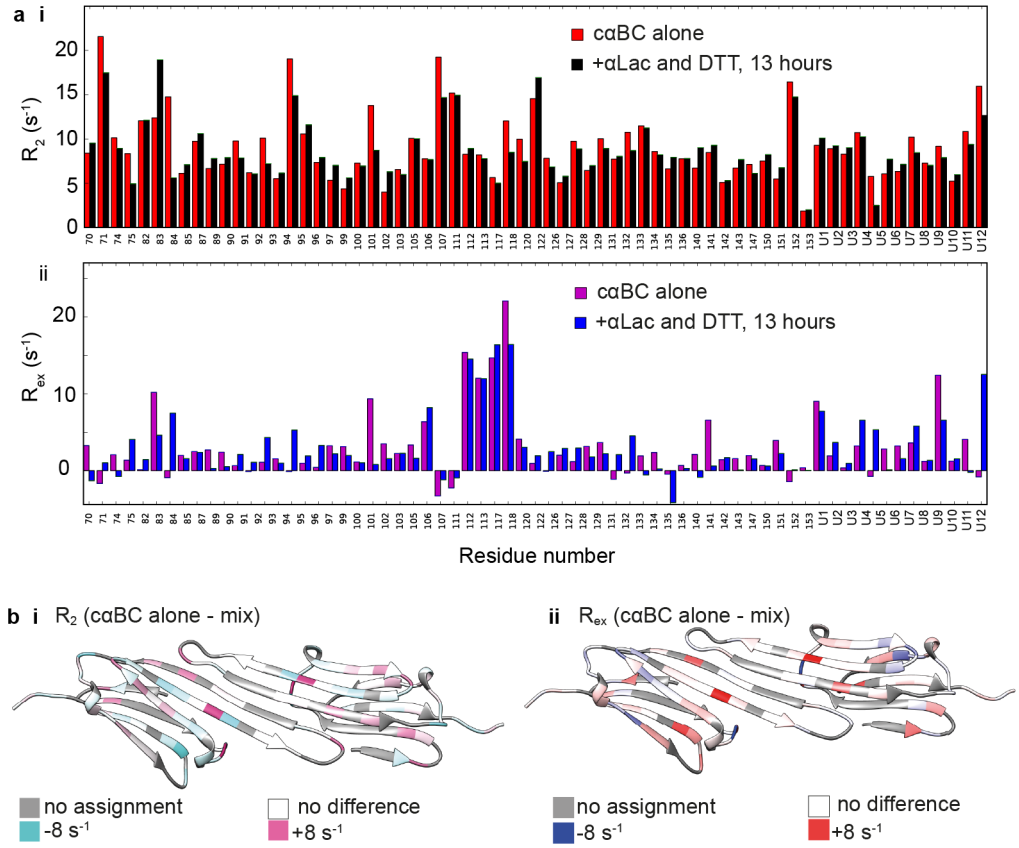



***Figure A.7:*** ***$^{15}$N-edited CPMG-RD experiments on $^{15}$N- and ILV-labelled sample of cαBC with and without αLac.*** ***a*** *$R_2$ and $R_{ex}$ values for 200 μM $^{15}$N- and ILV-labelled cαBC alone and in the presence of 100 μM aggregating αLac. The unassigned peaks are labelled U1-12.* ***b*** *The data in* ***a*** *mapped on the crystal structure of the cαBC dimer.*