

Protein folding investigated by NMR spectroscopy



T. Reid Alderson

Pembroke College
University of Oxford

A thesis submitted for the degree of
Doctor of Philosophy
Michaelmas 2018

Declaration

This dissertation is a summary of research performed in the Department of Chemistry at the University of Oxford and the National Institute of Diabetes and Digestive and Kidney Diseases at the National Institutes of Health between October 2014 and October 2018. The work described herein is wholly my own and not the outcome of a collaboration, except where this has been specifically indicated in the text. The dissertation has not, neither as a whole nor in part, been submitted for a degree, diploma, or any other academic qualification at any other university or academic institution.

T. Reid Alderson

Toronto, Canada, February 2019

Table of contents

CHAPTER 1

Protein folding and molecular chaperones	5
1.1 Proteins: biological workhorses.....	6
1.2 Structures and functions of proteins.....	8
1.3 A spectrum of protein conformational states	12
1.4 Protein misfolding and molecular chaperones.....	15
1.5 Summary	17
1.6 References	19

CHAPTER 2

Biomolecular NMR spectroscopy	24
2.1 NMR spectroscopy: resonating with proteins	25
2.2 Nuclear spin relaxation	28
2.2.1 Spectral density function	28
2.2.2 Spin interactions and transverse and longitudinal relaxation	31
2.3 Chemical exchange and NMR spectroscopy.....	37
2.4 Carr-Purcell-Meiboom-Gill relaxation dispersion	44
2.5 References	75

CHAPTER 3

Structural and dynamical characterization of the HSP27 monomer	79
3.1 Abstract.....	80
3.2 Introduction.....	81
3.2 Results	84
3.2.1 Reduction increases the population of HSP27 monomers and chaperone activity	84
3.2.2 HSP27 monomers are potent chaperones <i>in vitro</i> that readily self-aggregate	89
3.2.3 Dynamics at the HSP27 dimer interface are redox sensitive	93
3.2.4 Partially disordered monomers of cHSP27 characterized by RD NMR.....	101
3.2.5 High pressures and acidic conditions enable direct detection of the cHSP27 monomer	103
3.2.6 Structural and dynamical characterization of the partially disordered cHSP27 monomer	109
3.4 Discussion.....	113
3.5 Materials and Methods	121
3.5.1 Protein expression and purification	121
3.5.2 Site-directed mutagenesis in cHSP27	123
3.5.3 Aggregation of substrate proteins and chaperone activity assays	123
3.5.4 Aggregation of H124K/C137S	125
3.5.5 Native MS	126
3.5.6 Thermal denaturation of cHSP27	126
3.5.7 NMR spectroscopy.....	126
3.5.8 CD spectroscopy and bis-ANS fluorescence of C137S and H124K/C137S.....	135
3.5.9 Random Coil Index	135
3.6 Acknowledgements.....	136
3.7 References	137

CHAPTER 4

Dysregulation of HSP27 by a neuropathy-associated mutation	145
4.1 Abstract	146

4.2 Introduction.....	147
4.3 Results	150
4.3.1 P182L forms very large, soluble oligomers that are chaperone-inactive	150
4.3.2 The P182L mutation significantly lowers binding affinity for the IxI motif.....	152
4.3.3 Mutant and WT IxI/V binding kinetics probed by relaxation dispersion NMR	157
4.3.4 Altered thermodynamic properties of the P182L variant.....	161
4.3.5 Summary and Conclusions	162
4.4 Methods.....	164
4.4.1 Proteins and peptides	164
4.4.2 SEC-MALS, negative-stain EM, and chaperone activity assays	165
4.4.3 NMR spectroscopy.....	166
4.4.4 Two- or three-states	170
4.5 Acknowledgements.....	172
4.6 References	174

CHAPTER 5

Development of pressure-jump hydrogen exchange to study protein folding.....	177
5.1 Abstract	178
5.2 Introduction.....	179
5.3 Results and Discussion	181
5.4 Methods.....	194
5.4.1 Protein expression and purification	194
5.4.2 NMR spectroscopy.....	195
5.4.3 Pressure jump apparatus	195
5.4.4 Pressure-induced changes to the pK _w of water	196
5.4.5 Pressure-induced changes to the pK _a of phosphoric acid.....	197
5.5 Acknowledgements.....	205
5.6 References	206

Chapter 6

Monitoring protein structure by high-pressure, pulsed field gradient diffusion NMR spectroscopy	209
6.1 Abstract	210
6.2 Introduction.....	211
6.3 Results and Discussion	214
6.3.1 Pressure-induced changes to chemical shifts and hydration radii of tracer molecules.....	214
6.3.2 Pressure- and urea-denatured VA2-ubiquitin have similar hydrodynamic properties	223
6.3.3 PFG NMR diffusion experiments indicate minimal compaction of the unfolded state in low denaturant conditions.....	226
6.3.4 Diffusion measurements on equilibria of folded and unfolded protein by 1D PFG NMR	228
6.4 Conclusions.....	232
6.5 Methods.....	234
6.5.1 Protein expression and purification	234
6.5.2 NMR spectroscopy.....	235
6.6 References	239

CHAPTER 7

Measurement of <i>cis</i>-proline in disordered and unfolded proteins	244
7.1 Abstract	245
7.2 Introduction.....	246
7.3 Results	248
7.3.1 Measurement of <i>cis</i> -proline fractions in the IDP α -synuclein	248
7.3.2 Small peptides from α -synuclein overestimate the fraction of <i>cis</i> -proline.....	252
7.3.3 Long-range interactions not involving inter-domain contacts influence <i>cis</i> -proline propensity in α -synuclein.....	254
7.3.4 High-pressure NMR to measure <i>cis</i> -proline fractions in unfolded proteins.....	258
7.3.5 Similar backbone dynamics in the <i>cis</i> - and <i>trans</i> -proline conformers	261
7.4 Conclusions.....	266
7.5 Methods.....	267
7.5.1 Proteins and peptides	267
7.5.2 Preparation of peptide solutions	267
7.5.3 NMR spectroscopy.....	269
7.6 Acknowledgements.....	270
7.7 References	271

CHAPTER 8

Conclusions.....	274
8.1 Concluding remarks.....	275

Acknowledgements

I thank my parents, Mary and Tom, for wholeheartedly supporting and encouraging my education. Without them, and the support of my siblings, Lauren, Henry, and Samantha, none of this would be possible. To my supervisors, Andy Baldwin, Ad Bax, and Justin Benesch: thank you for welcoming me into your research groups and molding me into the scientist that I am today. Thank you for all of your teaching, generosity, enthusiasm, and support over the years, and for thinking about these research projects with me. To my mentor and former supervisor John Markley: thank you for teaching me the foundations of NMR and inspiring me to pursue a career in research. I fondly remember my time at NMRFAM.

Thanks to all of my colleagues in the Baldwin, Bax, and Benesch groups, and those from the neighboring groups of Marius Clore and Carol Robinson. I especially thank my colleague Heidi Gastall for welcoming me into the HSP27 project and for being an enthusiastic, generous mentor. I thank Matteo Degiacomi and Cyril Charlier for their friendship, positive nature, and for passing on their computational wizardry. Thanks to Gogulan Karunanithy for stimulating NMR discussions throughout and for printing my thesis. And finally, thanks to Iva Pritišanac for her support, encouragement, and love – I would not have made it without you.

1

Protein folding and molecular chaperones

1.1 Proteins: biological workhorses

The human genome encodes *ca.* 20,000 different proteins ([Consortium 2001](#), [Kim et al. 2014](#)), nanometer-sized biological molecules that perform diverse cellular roles from metabolizing energy sources to regulating the acidity of blood. The collective actions of proteins and their interactions with nucleic acids at the molecular level enable life at the macroscopic level. Each protein, while unique in composition, is synthesized from the same set of 20 amino acids, with sequential amino acids joined by a peptide bond involving the C-terminal carboxylic acid of amino acid i and the N-terminal amino group of $i + 1$.

Linear chains of amino acids are known as polypeptides or proteins, which typically fold into complex three-dimensional (3D) structures comprised of local secondary structure elements (*e.g.* α -helices ([Fig. 1](#)) and β -strands ([Fig. 2](#)), among other types). The primary sequence of a protein determines its 3D or tertiary structure, and whether it forms higher-order oligomers, or quaternary structure ([Fig. 1](#)). Whereas DNA generally exists as a double helix, proteins can fold into many distinct forms: helices, sheets, barrels, rings, zippers, and polyhedra, for example. Knowledge accumulated from the many thousands of 3D structures has enabled prediction of the structures of some proteins based on their amino acid sequences ([Baker & Sali 2001](#)), and design of proteins ([Kuhlman et al. 2003](#)) with specific functions ([Jiang et al. 2008](#)) or oligomeric architectures ([Bale et al. 2016](#), [Boyken et al. 2016](#)).

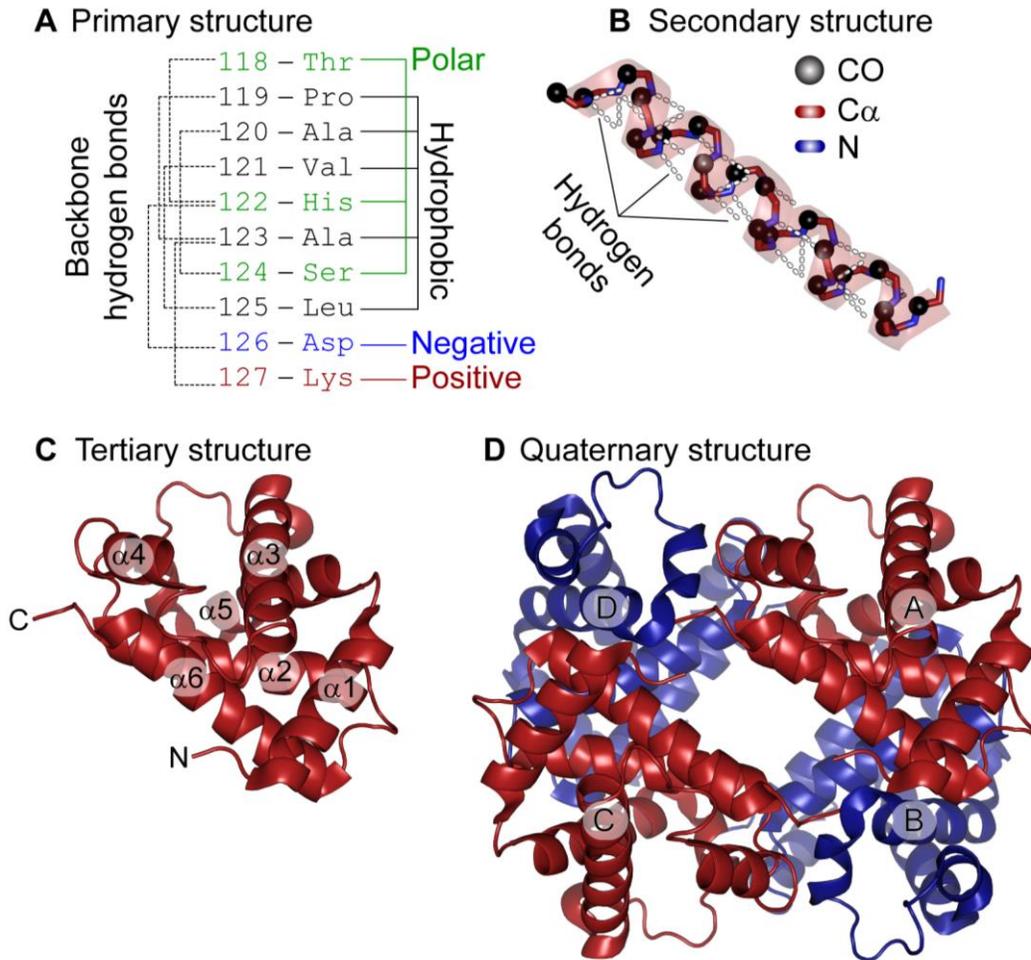


Figure 1. The four levels of protein structure. Human hemoglobin is used as an example (PDB ID: 2hhb) (Fermi et al. 1984). **(A)** Primary structure: the amino acid sequence. This is an abridged sequence of human hemoglobin corresponding to a portion of the $\alpha 6$ helix in panels B and C. The amino acids are colored according to their side-chain groups (*green* for polar, *red* for positively charged, *blue* for negatively charged, and *black* for hydrophobic). The i to $i+4$ backbone hydrogen bonding pattern of an α -helix is indicated with dashed lines on the left. **(B)** Secondary structure: an α -helix. The corresponding α -helix ($\alpha 6$) from panel A is taken from a subunit of hemoglobin (subunit A in panel D). The helix is shown in a transparent cartoon format, with stick representations for backbone N (*blue*) and C α atoms (*red*), and CO atoms shown as *black* spheres. Backbone hydrogen bonds are shown as white dashed lines. **(C)** Tertiary structure: the three-dimensional structure. A subunit of hemoglobin (subunit A) is shown in cartoon format with the six α -helices numerated. The N- and C-termini are indicated. The amino acid sequence from panel A and helix from panel B correspond to helix $\alpha 6$ here. **(D)** Quaternary structure: the arrangement of proteins into a larger assembly. Here, four hemoglobin subunits interact to form a tetramer.

1.2 Structures and functions of proteins

Structural modeling has long played a significant role in understanding the function and assembly of biomolecules (Fersht 2008). For example, the antibiotic penicillin was discovered from a natural source in 1929 by Alexander Fleming (Fleming 1929); by the early 1940's, penicillin could be mass produced, but its chemical synthesis remained intractable, as the chemical formula and structure of penicillin were unknown. The crystal structure of penicillin, solved by Dorothy Crowfoot Hodgkin in 1945 and published in 1949 (Crowfoot et al. 1949), illuminated the precise arrangement of atoms and the chemical formula of this naturally existing antibiotic. Such detailed information enabled the total chemical synthesis of penicillin in the laboratory and the development of antibiotic derivatives, such as cephalosporins.

Atomic models based on experimental data (Franklin & Gosling 1953, Wilkins et al. 1953) were generated during the successful determination of the 3D structure of DNA (Watson & Crick 1953). While the first 3D structures of the proteins myoglobin and hemoglobin (Fig. 1) were respectively determined in 1958 by John Kendrew and coworkers (Kendrew et al. 1958) and 1960 by Max Perutz and colleagues (Perutz et al. 1960), secondary structure elements of proteins had been predicted and already modeled in 1951 by Linus Pauling and colleagues (Pauling & Corey 1951, Pauling et al. 1951). Myoglobin and hemoglobin, the first proteins to have their structures determined, were comprised entirely of α -helices and loops (Fig. 1). (As an aside, Pauling originally named the α -helix an α -spiral; one of his colleagues successfully lobbied for the former (Eisenberg 2003)). Experimental confirmation of a protein containing β -strands, however, did not arrive until 1965 when David Phillips and colleagues determined the structure of

hen egg-white lysozyme (Blake et al. 1965), nearly 15 years after Pauling's original postulation. Elegant crystallographic studies on hen egg-white lysozyme paved the way for delineating the catalytic mechanisms of proteins and enzymes based on structural models and biochemical assays (Johnson & Phillips 1965, Phillips 1970).

Based on the similarity of the original 3D structures from myoglobin (1958) and hemoglobin (1960), it was evident to structural biologists, or molecular biologists at the time, that related proteins would fold into related structures. Nearly 40 years ago, Max Perutz was already speculating on the evolutionary conservation of the 3D structure of hemoglobin (Perutz 1979, 1983), and yet 3D structures were only available for human and horse hemoglobin. Perutz nevertheless surmised that, in order to have a common function, hemoglobins from vertebrates would likely adopt the same or highly similar structures (Perutz 1979, 1983). As Perutz wrote:

My proposal that the tertiary and quaternary structure of the hemoglobins has been conserved throughout evolution from fish to mammals has met with disbelief among biologists and others, but it comes to no surprise to protein crystallographers, who have found that homologous proteins in distant species have closely similar structures. (Perutz 1983)

It is now generally accepted that homologous proteins from closely related, or even evolutionarily distant organisms will likely adopt the same or similar structures. For example, the three proteins shown in Fig. 2 (PDB IDs: 4r3e, 1a58, 1oca) originate from evolutionarily distant species, a bacterium (*Escherichia coli*), a nematode (*Brugia malayi*), and human (*Homo sapiens*), yet exhibit near identical 3D structures: upon alignment to the secondary structural elements of the human protein, the backbone root-mean-square deviations are respectively 0.78 Å and 0.68 Å for *E. coli* and *B. malayi* (Fig. 2). Of the 165 and 177 amino acids in CycA from *E. coli* and *B. malayi*, respectively, 109 and 113 amino

acids have been substituted relative to the sequence of *H. sapiens* Cwc27. Thus, despite conservation levels of 34% and 36%, the *E. coli* and *B. malayi* CycA structures are nearly indistinguishable to the human Cwc27.

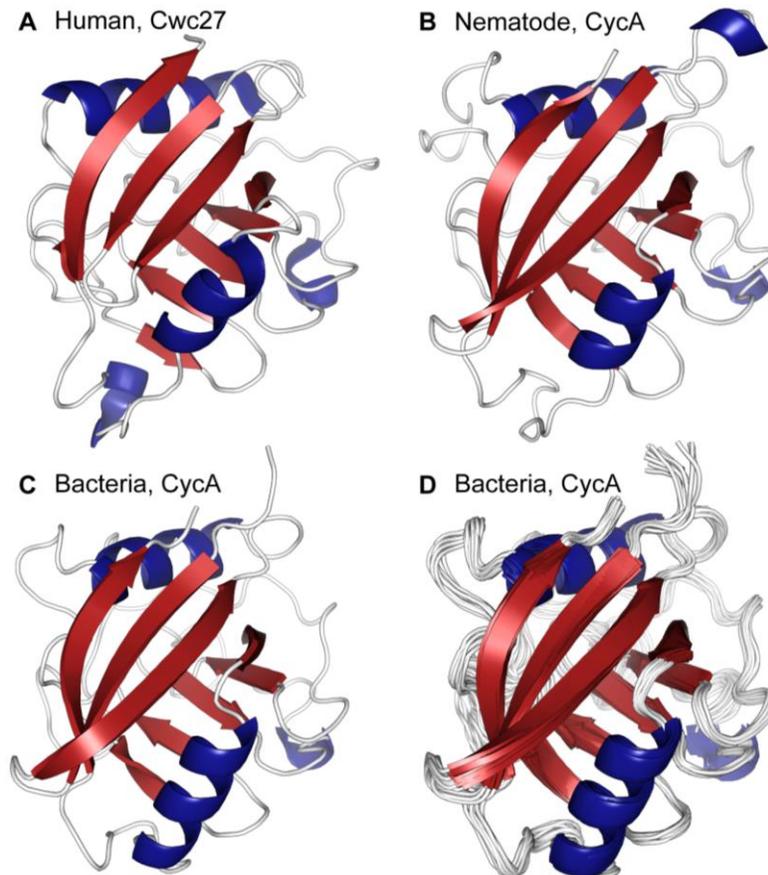


Figure 2. Evolutionary conservation of the 3D structure of proteins. Shown here are the 3D structures of three peptidyl-prolyl isomerases from different organisms. (A) X-ray structure of the protein Cwc27 from *H. sapiens*, (B) X-ray structure of CycA from *B. malayi*, and (C) Lowest energy NMR structure and (D) the 20 deposited NMR structures of CycA from *E. coli*. The β -strands are colored dark red and α -helices dark blue. PDB IDs: 4r3e, 1a58, 1oca. Alignment of the amino acid sequences of the human and bacterial or human and nematode proteins respectively yields 34% and 36% sequence identity. Despite differing in ca. 65% of the amino acids, the three proteins yield overall highly similar structures.

1.3 A spectrum of protein conformational states

Preservation of functionality through the conservation of atomic structures has led to the assumption that the 3D structure of a protein encodes its biological function. Curiously, however, a set of proteins were discovered called intrinsically disordered proteins (IDPs), which lack a fixed 3D structure (Fig. 3). IDPs adopt an ensemble of unfolded or transiently folded, disordered conformations that rapidly interconvert on a pico-nanosecond timescale (Dunker et al. 2002, Dyson & Wright 2005, Wright & Dyson 1999). In the human proteome, ca. 30% of the proteins are anticipated to be IDPs or contain sizeable intrinsically disordered regions (IDRs) (van der Lee et al. 2014) interspersed with folded domains (Ward et al. 2004). Many IDPs were reported in the 1990's and primarily identified through nuclear magnetic resonance (NMR) spectroscopy (Eliezer et al. 1998, Kriwacki et al. 1996, Penkett et al. 1997, Weiss et al. 1990, Zhang et al. 1994). Indeed, NMR spectroscopy has proven to be indispensable for the biophysical characterization of IDPs and IDRs (Dyson & Wright 2001, Jensen et al. 2014, Konrat 2014, Redfield 2004).

Various degrees of secondary structure can exist within an IDP or IDR (Conicella et al. 2016, Marsh et al. 2006) (Fig. 3). The presence of residual secondary structure within a compact, unfolded protein is referred to as a molten globule (Dolgikh et al. 1981, Ohgushi & Wada 1983). Some folded proteins adopt molten globule conformations at low pH (Redfield 2004), with some conformations reported as on-pathway folding intermediates (Dyson & Wright 2017). It is not a prerequisite for IDPs or IDRs to remain always disordered: numerous transcription factors are disordered in their unbound states and fold upon binding to their cognate DNA (Dyson & Wright 2002, Weiss et al. 1990) or other protein partners (Shammas 2017, Sugase et al. 2007). Likewise, a class of proteins known as molecular chaperones (discussed below), which function to prevent the

aggregation of other proteins, contains members that are conditionally disordered ([Bardwell & Jakob 2012](#)). That is, these proteins can exist in a disordered or partially disordered conformation under certain conditions, and fold into a structured state under other conditions. For example, the bacterial protein HdeA, which responds to acid stress, exists as a stable dimer at pH 7 but dissociates and unfolds at acidic pH; the unfolded monomeric form exhibits a significantly higher activity than the folded dimer ([Tapley et al. 2009, 2010](#)). In contrast, the protein 4E-BP2 undergoes a disorder-to-order transition. 4E-BP2 inhibits translation of mRNA and exists *in vitro* as an IDP ([Bah et al. 2015](#)). Upon phosphorylation at multiple sites, however, 4E-BP2 folds into a β -strand-rich structure; the disorder-to-order transition is accompanied by occlusion of a key binding motif, which prevents 4E-BP2 from inhibiting translation ([Bah et al. 2015](#)). Therefore, post-translational modifications of proteins can modulate structural propensities, and provide cells with the ability to regulate the degree of order or disorder through, *e.g.*, kinase and phosphatase activities.

A relatively new class of proteins that lack a fixed structure, the so-called metamorphic proteins, have recently been reported ([Dishman & Volkman 2018, Tuinstra et al. 2008](#)). Metamorphic proteins can adopt more than one stable structure depending on the conditions, even reversibly transitioning between α -helix- and β -strand-rich conformations ([Murzin 2008](#)). Such behavior is reminiscent of the structural transitions that accompany amyloid fibril formation, whereby proteins that are otherwise structured or even natively unfolded undergo large-scale conformational rearrangements and assemble into a β -strand-rich fibril ([Iadanza et al. 2018](#)). While the kinetic barrier for amyloid fibril formation significantly exceeds that associated with metamorphic proteins'

structural transitions, it has been demonstrated that amyloid fibrils are more thermodynamically favorable than native conformations (Baldwin et al. 2011b) under physiological conditions. Even the all- α -helical protein myoglobin can form β -strand-rich amyloid fibrils (Fändrich et al. 2001), as can lysozyme (Booth et al. 1997). In some cases, amyloid fibrils, such as those formed by the prion protein Sup35p, can confer selective advantages to organisms and boost overall fitness (Shorter & Lindquist 2005, True et al. 2004).

Thus, the folding of proteins is not an essential prerequisite for their function and appears to be moldable and context dependent. The degree of structural disorder (Fig. 3) can be “tuned” to respond to cellular conditions and alter the functions of proteins as needed. Such structural plasticity is believed to enable promiscuous interactions with a diverse set of target proteins and nucleic acids, rather than a canonical “lock-and-key” interaction with a specific biomolecule.

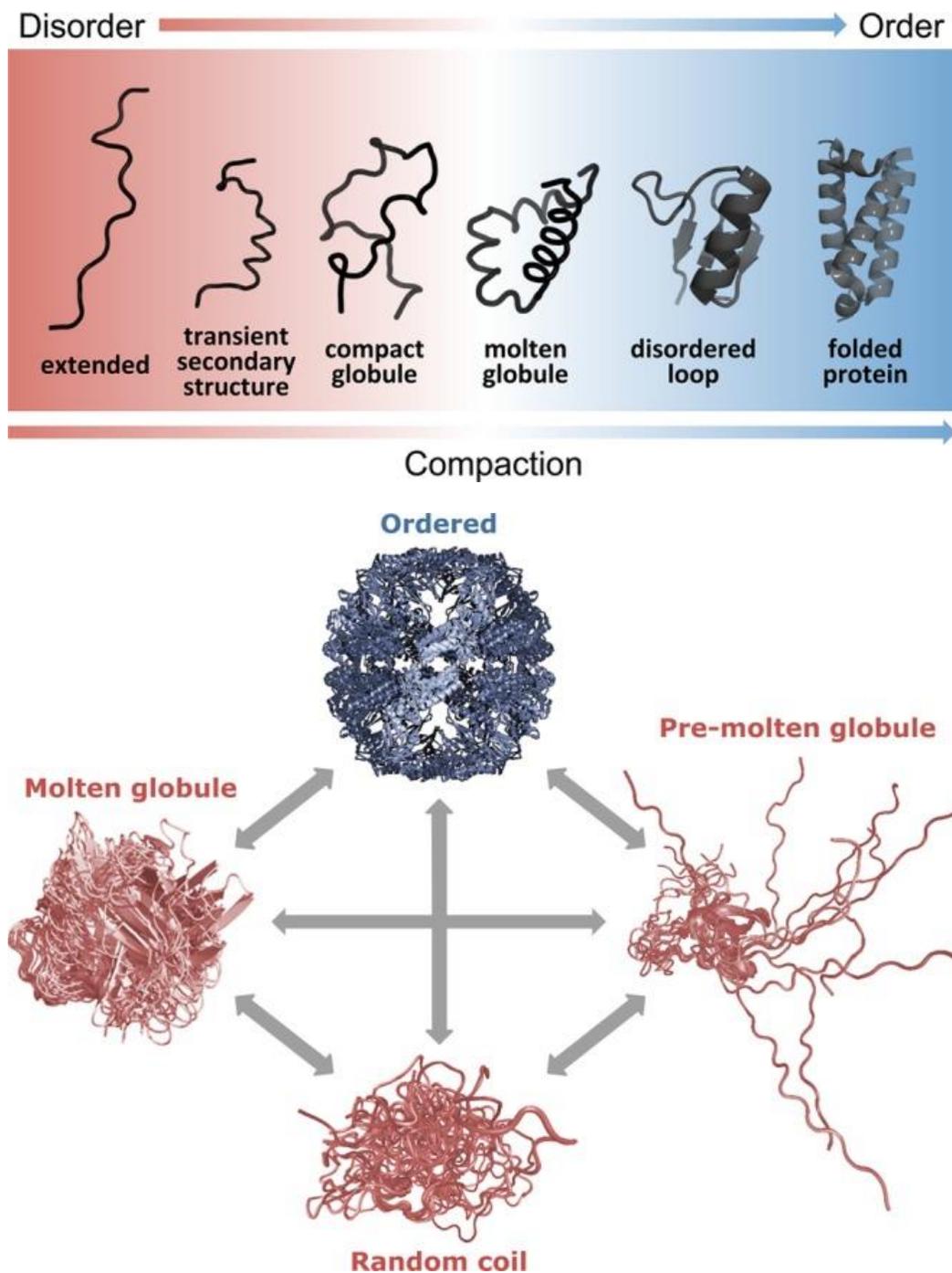


Figure 3. The spectrum of protein conformational states. (*Top*) A schematic of the protein disorder-to-order spectrum, ranging from a random coil (extended) to a fully folded protein. Disordered regions are represented by heavy lines and structured regions by cartoon. (*Bottom*) Structural representations of various states in the spectrum of protein conformational states. The random coil, pre-molten globule, and molten globule depictions are models and not based on experimental restraints. A given protein may interconvert between all four of the depicted forms. Images adapted from [van der Lee et al. 2014](#) with permission from the American Chemical Society.

1.4 Protein misfolding and molecular chaperones

While typically left to the realm of biophysics, the folding of proteins from unstructured chains into exquisite three-dimensional structures plays a pivotal role in human health. Seemingly disparate diseases such as type II diabetes, cystic fibrosis, ALS, and Alzheimer's share a common molecular origin: the misfolding of proteins ([Knowles et al. 2014](#)). The maintenance of protein homeostasis (proteostasis) comprises an essential biological process ([Kim et al. 2013](#)). *In vivo*, the concentration of macromolecules can reach 300 mg/mL, and newly synthesized polypeptides that are released into highly crowded environments *in vivo* must rapidly fold into their native three-dimensional structures. If protein folding is delayed or has faltered, unwanted associations between exposed hydrophobic amino acids can lead to cytotoxic protein aggregation ([Knowles et al. 2014](#)). Moreover, during the lifetime of an organism, repeated exposures to cellular stresses such as oxidative damage, acidosis, and heat shock disrupt proteostasis and can culminate in cell death ([Kirstein et al. 2015](#), [Walther et al. 2015](#)).

However, cells have armed themselves with an evolutionarily conserved defense mechanism to combat protein misfolding: a class of proteins known as molecular chaperones that can recognize misfolded, aggregation-prone proteins and either correctly refold them or target them for degradation and recycling ([Alderson et al. 2016](#), [Didenko et al. 2012](#), [Kampinga et al. 2015](#), [Mayer & Bukau 2005](#)). Understanding the precise molecular mechanisms that underpin these processes could enable future therapeutics to be designed that help mitigate protein misfolding diseases. A particular class of molecular chaperones known as heat shock proteins (HSPs) upholds proteostasis during basal and stressful conditions by assisting with protein folding, unfolding, re-folding, translocation, assembly, and degradation ([Hartl 2011](#), [Hartl & Hayer-Hartl 2002](#), [Martin & Hartl 1994](#)).

Within the family of HSPs, the small HSPs (sHSPs), named for their monomeric masses (12–43 kDa), function to both refold non-natively folded proteins and keep such substrates competent for subsequent binding to refoldase machinery (e.g. HSP70) (Hochberg & Benesch 2014). These sHSP-mediated processes do not require energy liberated from ATP binding or hydrolysis. In humans, there are 10 sHSPs (HSPB1–HSPB10) (Kampinga et al. 2015), some of which can homo- and hetero-oligomerize into dynamic, polydisperse assemblies ranging from dimers (40-50 kDa) to 50-mers (>1 MDa).

HSP27, the main focus of a significant portion of this thesis, is a human sHSP that assembles into such polydisperse oligomers, which form and continuously exchange via dimeric building blocks (Arrigo 2001). Dimerization occurs through a conserved ~80-residue α -crystallin domain (ACD), which folds into a β -sheet-rich immunoglobulin-like domain (Hochberg et al. 2014). The ACD is flanked by variable N- and C-terminal regions with little sequence homology beyond a conserved IxI/V motif near the C-terminus (Hochberg & Benesch 2014). This IxI/V motif serves to bridge sHSP dimers together and regulate subunit exchange, or quaternary dynamics (Baldwin et al. 2011a,c,d; Hilton et al. 2013). More information about HSP27 will be provided in the sections below.

1.5 Summary

As outlined above, molecular chaperones and protein folding comprise an essential aspect of human health. A thin line exists between correctly folded proteins and cytotoxic aggregates. Understanding the biophysical properties of the folding and misfolding processes, primarily with the aid of solution-state NMR spectroscopy, has been central to this thesis. Various aspects of protein folding, including hydrogen exchange and *cis*-proline formation, are examined alongside the function and dysfunction of a molecular chaperone, human HSP27.

To provide a basis for many of the experimental results that follow, I begin in Chapter 2 by introducing NMR spectroscopy and some of its theoretical underpinnings. Topics that are discussed in detail include nuclear spin relaxation and Carr-Purcell-Meiboom-Gill relaxation dispersion, which are featured heavily in Chapters 3 and 4.

In Chapter 3, I present the order-to-partially disordered transition of the HSP27 α -crystallin domain dimer. I discovered that the HSP27 α -crystallin monomer partially unfolds and exhibits higher chaperone activity than the structured dimeric counterpart. Interestingly, the monomer also has a heightened tendency to self-aggregate, signifying the delicate balance between order and disorder and function and dysfunction.

In Chapter 4, I investigate the molecular basis for the dysregulation of the HSP27 oligomer by a neuropathy-associated hereditary mutation, P182L. Pro182 resides in the highly conserved IxI/V motif of HSP27, and its mutation to Leu is known to be implicated in the onset of a motor neuropathy termed Charcot-Marie-Tooth disease. I demonstrate that the P182L variant of HSP27 forms very large, soluble oligomers *in vitro* that are nearly 30-fold larger than the WT protein. With Carr-Purcell-Meiboom-Gill relaxation dispersion, I probe the binding kinetics of peptides encompassing the WT and P182L IxI/V motifs, and find that the mutant IxI/V binds nearly 10-fold weaker because of an attenuated on-rate.

In Chapter 5, the focus switches to protein folding. I outline novel pressure-jump NMR experiments to monitor hydrogen exchange during protein folding. A pressure-sensitized variant of ubiquitin is used as a model system. My results indicate that the folding of the ubiquitin variant is limited by the formation of hydrogen bonds in a reverse turn. Many amides become protected from solvent exchange after *ca.* 70-80 ms after folding has been initiated, whereas this reverse turn only becomes protected after *ca.* 130

ms. The hydrogen exchange rates in the unfolded protein prior to crossing the transition barrier were also determined.

In Chapter 6, I outline the usage of high-pressure, pulsed field gradient NMR spectroscopy to monitor structural changes in proteins. The same ubiquitin variant from Chapter 5 is employed and its translational diffusion is investigated as a function of pressure and urea concentration. The unfolded state is found to only slightly compact under lowly denaturing conditions, with a decrease in the hydration radius of *ca.* 10% relative to the fully denatured state (*i.e.* at 2500 bar or 8M urea). Pressure-dependent changes to the hydration radii of tracer molecules are discussed.

Finally, in Chapter 7, I present the utility of NMR spectroscopy to characterize *cis*-proline formation in unfolded proteins. Across a total of 15 Xaa-Pro bonds in unfolded proteins or unfolded regions of proteins, the average *cis*-Pro content is *ca.* 6%. This value is considerably lower than what is commonly reported in the literature. I report in this chapter that the usage of small peptides to measure *cis*-Pro content, a common practice, leads to over-estimation of the fraction of *cis*. For unblocked peptides, the origin of the elevated *cis* content is due to electrostatic interactions between oppositely charged termini. For blocked peptides, there appears to be a steric effect that is minimized with increasing length of the peptide.

1.6 References

- Alderson TR, Kim JH, Markley JL. 2016. Dynamical Structures of Hsp70 and Hsp70-Hsp40 Complexes. *Structure*. 24(7): 1014-30.
- Arrigo AP. 2001. HSP27: novel regulator of intracellular redox state. *IUBMB Life*. 52(6):303-07
- Bah A, Vernon RM, Siddiqui Z, Krzeminski M, Muhandiram R, et al. 2015. Folding of an intrinsically disordered protein by phosphorylation as a regulatory switch. *Nature*. 519(7541):106–9
- Baker D, Sali A. 2001. Protein structure prediction and structural genomics. *Science*. 294(5540):93–96
- Baldwin AJ, Hilton GR, Lioe H, Bagn eris C, Benesch JLP, Kay LE. 2011a. Quaternary Dynamics of α B-Crystallin as a Direct Consequence of Localised Tertiary Fluctuations in the C-Terminus. *J. Mol. Biol.* 413(2):310–20
- Baldwin AJ, Knowles TPJ, Tartaglia GG, Fitzpatrick AW, Devlin GL, et al. 2011b. Metastability of Native Proteins and the Phenomenon of Amyloid Formation. *J. Am. Chem. Soc.* 133(36):14160–63
- Baldwin AJ, Lioe H, Hilton GR, Baker LA, Rubinstein JL, et al. 2011c. The polydispersity of α B-crystallin is rationalized by an interconverting polyhedral architecture. *Structure*. 19(12):1855–63
- Baldwin AJ, Lioe H, Robinson C V, Kay LE, Benesch JLP. 2011d. α B-crystallin polydispersity is a consequence of unbiased quaternary dynamics. *J. Mol. Biol.* 413(2):297–309
- Bale JB, Gonen S, Liu Y, Sheffler W, Ellis D, et al. 2016. Accurate design of megadalton-scale two-component icosahedral protein complexes. *Science*. 353(6297):389–94
- Bardwell JCA, Jakob U. 2012. Conditional disorder in chaperone action. *Trends Biochem. Sci.* 37(12):517–25
- Blake CC, Koenig DF, Mair GA, North AC, Phillips DC, Sarma VR. 1965. Structure of hen egg-white lysozyme. A three-dimensional Fourier synthesis at 2 Angstrom resolution. *Nature*. 206(4986):757–61
- Booth DR, Sunde M, Bellotti V, Robinson C V., Hutchinson WL, et al. 1997. Instability, unfolding and aggregation of human lysozyme variants underlying amyloid fibrillogenesis. *Nature*. 385(6619):787–93
- Boyken SE, Chen Z, Groves B, Langan RA, Oberdorfer G, et al. 2016. De novo design of protein homooligomers with modular hydrogen-bond network-mediated specificity. *Science*. 352(6286):680–87
- Conicella AE, Zerze GH, Mittal J, Fawzi NL. 2016. ALS Mutations Disrupt Phase Separation Mediated by α -Helical Structure in the TDP-43 Low-Complexity C-Terminal Domain. *Structure*. 24(9):1537–49
- Consortium IHGS. 2001. Initial sequencing and analysis of the human genome. *Nature*. 409(6822):860–921
- Crowfoot D, Bunn CW, Rogers-Low BW, Turner-Jones A. 1949. X-ray crystallographic investigation of the structure of penicillin. In *Chemistry of Penicillin*, eds. HT Clarke, JR Johnson, R Robinson, pp. 310–67. Princeton University Press
- Didenko T, Duarte AMS, Karag z GE, R diger SGD. 2012. Hsp90 structure and function studied by NMR spectroscopy. *Biochim. Biophys. Acta*. 1823(3):636–47
- Dishman AF, Volkman BF. 2018. Unfolding the Mysteries of Protein Metamorphosis. *ACS Chem. Biol.* 13(6):1438–46
- Dolgikh DA, Gilmanshin RI, Brazhnikov E V, Bychkova VE, Semisotnov G V, et al. 1981. Alpha-Lactalbumin:

compact state with fluctuating tertiary structure? *FEBS Lett.* 136(2):311–15

Dunker AK, Brown CJ, Lawson JD, Iakoucheva LM, Obradović Z. 2002. Intrinsic disorder and protein function. *Biochemistry.* 41(21):6573–82

Dyson HJ, Wright PE. 2001. Nuclear magnetic resonance methods for elucidation of structure and dynamics in disordered states. *Methods Enzymol.* 339:258–70

Dyson HJ, Wright PE. 2002. Coupling of folding and binding for unstructured proteins. *Curr. Opin. Struct. Biol.* 12(1):54–60

Dyson HJ, Wright PE. 2005. Intrinsically unstructured proteins and their functions. *Nat. Rev. Mol. Cell Biol.* 6(3):197–208

Dyson HJ, Wright PE. 2017. How Does Your Protein Fold? Elucidating the Apomyoglobin Folding Pathway. *Acc. Chem. Res.* 50(1):105–11

Eisenberg D. 2003. The discovery of the alpha-helix and beta-sheet, the principal structural features of proteins. *Proc. Natl. Acad. Sci. U. S. A.* 100(20):11207–10

Eliezer D, Yao J, Dyson HJ, Wright PE. 1998. Structural and dynamic characterization of partially folded states of apomyoglobin and implications for protein folding. *Nat. Struct. Biol.* 18(2):148–55

Fändrich M, Fletcher MA, Dobson CM. 2001. Amyloid fibrils from muscle myoglobin. *Nature.* 410(6825):165–66

Fermi G, Perutz MF, Shaanan B, Fourme R. 1984. The crystal structure of human deoxyhaemoglobin at 1.74 Å resolution. *J. Mol. Biol.* 175(2):159–74

Fersht AR. 2008. From the first protein structures to our current knowledge of protein folding: delights and scepticisms. *Nat. Rev. Mol. Cell Biol.* 9(8):650–54

Fleming A. 1929. On the antibacterial action of cultures of a penicillium, with special reference to their use in the isolation of *B. influenzae*. *Br. J. Exp. Pathol.* 10(3):226–36

Franklin RE, Gosling RG. 1953. Molecular Configuration in Sodium Thymonucleate. *Nature.* 171(4356):740–41

Hartl FU. 2011. Chaperone-assisted protein folding: the path to discovery from a personal perspective. *Nat. Med.* 17(10):1206–10

Hartl FU, Hayer-Hartl M. 2002. Molecular chaperones in the cytosol: from nascent chain to folded protein. *Science.* 295(5561):1852–58

Hilton GR, Hochberg GKA, Laganowsky A, McGinnigle SI, Baldwin AJ, Benesch JLP. 2013. C-terminal interactions mediate the quaternary dynamics of B-crystallin. *Philos. Trans. R. Soc. B Biol. Sci.* 368(1617):20110405–20110405

Hochberg GK a., Ecroyd H, Liu C, Cox D, Cascio D, et al. 2014. The structured core domain of B-crystallin can prevent amyloid fibrillation and associated toxicity. *Proc. Natl. Acad. Sci.* 111(16):E1562–70

Hochberg GKA, Benesch JLP. 2014. Dynamical structure of αB-crystallin. *Prog. Biophys. Mol. Biol.* 115(1):11–20

Iadanza MG, Jackson MP, Hewitt EW, Ranson NA, Radford SE. 2018. A new era for understanding amyloid structures and disease. *Nat. Rev. Mol. Cell Biol.*

- Jensen MR, Zweckstetter M, Huang J, Blackledge M. 2014. Exploring Free-Energy Landscapes of Intrinsically Disordered Proteins at Atomic Resolution Using NMR Spectroscopy. *Chem. Rev.* 114(13):6632–60
- Jiang L, Althoff EA, Clemente FR, Doyle L, Rothlisberger D, et al. 2008. De Novo Computational Design of Retro-Aldol Enzymes. *Science.* 319(5868):1387–91
- Johnson LN, Phillips DC. 1965. Structure of some crystalline lysozyme-inhibitor complexes determined by X-ray analysis at 6 Angstrom resolution. *Nature.* 206(4986):761–63
- Kampinga HH, Boer R, Beerstra N. 2015. The Big Book on Small Heat Shock Proteins. , eds. MR Tanguay, EL Hightower, pp. 3–26. Cham: Springer International Publishing
- Kendrew JC, Bodo G, Dintzis HM, Parrish RG, Wyckoff H, Phillips DC. 1958. A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. *Nature.* 181(4610):662–66
- Kim M-S, Pinto SM, Getnet D, Nirujogi RS, Manda SS, et al. 2014. A draft map of the human proteome. *Nature.* 509(7502):575–81
- Kim YE, Hipp MS, Bracher A, Hayer-Hartl M, Hartl FU. 2013. *Molecular Chaperone Functions in Protein Folding and Proteostasis.*, Vol. 82
- Kirstein J, Morito D, Kakihana T, Sugihara M, Minnen A, et al. 2015. Proteotoxic stress and ageing triggers the loss of redox homeostasis across cellular compartments. *EMBO J.* 34(18):2334–49
- Knowles TPJ, Vendruscolo M, Dobson CM. 2014. The amyloid state and its association with protein misfolding diseases. *Nat. Rev. Mol. Cell Biol.* 15(6):384–96
- Konrat R. 2014. NMR contributions to structural dynamics studies of intrinsically disordered proteins. *J. Magn. Reson.* 241:74–85
- Kriwacki RW, Hengst L, Tennant L, Reed SI, Wright PE. 1996. Structural studies of p21Waf1/Cip1/Sdi1 in the free and Cdk2-bound state: conformational disorder mediates binding diversity. *Proc. Natl. Acad. Sci. U. S. A.* 93(21):11504–9
- Kuhlman B, Dantas G, Ireton GC, Varani G, Stoddard BL, Baker D. 2003. Design of a Novel Globular Protein Fold with Atomic-Level Accuracy. *Science.* 302(5649):1364–68
- Marsh JA, Singh VK, Jia Z, Forman-Kay JD. 2006. Sensitivity of secondary structure propensities to sequence differences between alpha- and gamma-synuclein: implications for fibrillation. *Protein Sci.* 15(12):2795–2804
- Martin J, Hartl FU. 1994. Molecular Chaperones in Cellular Protein Folding. *BioEssays.* 16(9):689–92
- Mayer MP, Bukau B. 2005. Hsp70 chaperones: cellular functions and molecular mechanism. *Cell. Mol. Life Sci.* 62(6):670–84
- Murzin AG. 2008. Metamorphic Proteins. *Science.* 320(5884):1725–26
- Ohgushi M, Wada A. 1983. “Molten-globule state”: a compact form of globular proteins with mobile side-chains. *FEBS Lett.* 164(1):21–24
- Pauling L, Corey RB. 1951. The pleated sheet, a new layer configuration of polypeptide chains. *Proc. Natl. Acad. Sci. U. S. A.* 37(5):251–56
- Pauling L, Corey RB, Branson HR. 1951. The structure of proteins; two hydrogen-bonded helical configurations of the polypeptide chain. *Proc. Natl. Acad. Sci. U. S. A.* 37(4):205–11

- Penkett CJ, Redfield C, Dodd I, Hubbard J, McBay DL, et al. 1997. NMR analysis of main-chain conformational preferences in an unfolded fibronectin-binding protein. *J. Mol. Biol.* 274(2):152–59
- Perutz MF. 1979. Molecular adaptation in haemoglobin and thermophile bacteria. *Differentiation.* 13(1):47–50
- Perutz MF. 1983. Species adaptation in a protein molecule. *Mol. Biol. Evol.* 1(1):1–28
- Perutz MF, Rossman MG, Cullis AF, Muirhead H, Will G, North AC. 1960. Structure of haemoglobin: a three-dimensional Fourier synthesis at 5.5-Å resolution, obtained by X-ray analysis. *Nature.* 185(4711):416–22
- Phillips DC. 1970. The development of crystallographic enzymology. *Biochem. Soc. Symp.* 30:11–28
- Redfield C. 2004. Using nuclear magnetic resonance spectroscopy to study molten globule states of proteins. *Methods.* 34(1):121–32
- Shammas SL. 2017. Mechanistic roles of protein disorder within transcription. *Curr. Opin. Struct. Biol.* 42:155–61
- Shorter J, Lindquist S. 2005. Prions as adaptive conduits of memory and inheritance. *Nat. Rev. Genet.* 6(6):435–50
- Sugase K, Dyson HJ, Wright PE. 2007. Mechanism of coupled folding and binding of an intrinsically disordered protein. *Nature.* 447(7147):1021–25
- Tapley TL, Franzmann TM, Chakraborty S, Jakob U, Bardwell JCA. 2010. Protein refolding by pH-triggered chaperone binding and release. *Proc. Natl. Acad. Sci.* 107(3):1071–76
- Tapley TL, Körner JL, Barge MT, Hupfeld J, Schauerte JA, et al. 2009. Structural plasticity of an acid-activated chaperone allows promiscuous substrate binding. *Proc. Natl. Acad. Sci. U. S. A.* 106(14):5557–62
- True HL, Berlin I, Lindquist SL. 2004. Epigenetic regulation of translation reveals hidden genetic variation to produce complex traits. *Nature.* 431(7005):184–87
- Tuinstra RL, Peterson FC, Kutlesa S, Elgin ES, Kron MA, Volkman BF. 2008. Interconversion between two unrelated protein folds in the lymphotactin native state. *Proc. Natl. Acad. Sci.* 105(13):5057–62
- van der Lee R, Buljan M, Lang B, Weatheritt RJ, Daughdrill GW, et al. 2014. Classification of Intrinsically Disordered Regions and Proteins. *Chem. Rev.* 114(13):6589–6631
- Walther DM, Kasturi P, Zheng M, Pinkert S, Vecchi G, et al. 2015. Widespread Proteome Remodeling and Aggregation in Aging *C. elegans*. *Cell.* 161(4):919–32
- Ward JJ, Sodhi JS, McGuffin LJ, Buxton BF, Jones DT. 2004. Prediction and Functional Analysis of Native Disorder in Proteins from the Three Kingdoms of Life. *J. Mol. Biol.* 337(3):635–45
- Watson JD, Crick FHC. 1953. Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature.* 171(4356):737–38
- Weiss MA, Ellenberger T, Wobbe CR, Lee JP, Harrison SC, Struhl K. 1990. Folding transition in the DMA-binding domain of GCN4 on specific binding to DNA. *Nature.* 347(6293):575–78
- Wilkins MHF, Stokes AR, Wilson HR. 1953. Molecular Structure of Nucleic Acids: Molecular Structure of Deoxyribose Nucleic Acids. *Nature.* 171(4356):738–40

Wright PE, Dyson HJ. 1999. Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *J. Mol. Biol.* 293(2):321–31

Zhang O, Kay LE, Olivier JP, Forman-Kay JD. 1994. Backbone ^1H and ^{15}N resonance assignments of the N-terminal SH3 domain of drk in folded and unfolded states using enhanced-sensitivity pulsed field gradient NMR techniques. *J. Biomol. NMR.* 4(6):845–58

2

Biomolecular NMR spectroscopy

2.1 NMR spectroscopy: resonating with proteins

Nuclear magnetic resonance (NMR) spectroscopy is the main experimental method capable of probing the structures and dynamics of molecules at atomic resolution in solution (Cavanagh et al. 2007, Ernst et al. 1991, Wüthrich 1986). While proteins and nucleic acids are typically represented by a single static structure, many studies have demonstrated the importance of describing the motions, plasticity, and structural ensembles of biomolecules (Boehr et al. 2009, Fersht 2013, Karplus & Petsko 1990, McCammon et al. 1977, Wagner & Wuthrich 1978, Wuthrich & Wagner 1975). Numerous powerful NMR experiments have been designed to characterize molecular motions across a large range of timescales, spanning pico-nanoseconds (Jarymowycz & Stone 2006) to micro-milliseconds (Palmer et al. 2001) to seconds-hours and longer (Kleckner & Foster 2011). In the case of spin relaxation experiments that probe rapid ps-ns dynamics, modelling of the relevant spectral density functions at the experimentally accessible frequencies provides insight into the timescale and amplitude of the motions (Clore et al. 1990, Lipari & Szabo 1982). For dynamics on slower timescales such as seconds to hours, the modulation of signal intensities due to the exchange of labile amide protons with solvent protons or deuterons yields insight into local structural propensities and hydrogen bonding patterns (Englander & Mayne 1992).

The theoretical considerations that underlie NMR originate from nuclear spin angular momentum, which itself is quantum mechanical. A full description of the theory of NMR spectroscopy is outside the scope of this work, but interested readers are directed toward more quantitative monographs (Cavanagh et al. 2007, Ernst et al. 1991, Hore et al. 2000, Rule & Hitchens 2006).

Nuclear spin is characterized by two numbers, the spin number (I) and the magnetic quantum number (m), with detection of an NMR signal reliant on a non-zero value of I . Nuclear spin angular momentum is a vector, I , with its magnitude described by:

$$|I| = \sqrt{I \cdot I} = \hbar \sqrt{I(I+1)} \quad (1)$$

where \hbar is Planck's constant divided by 2π . Typically, the projection of I along the z-axis is considered:

$$I_z = \hbar m \quad (2)$$

where $m = [-l, -l + 1, \dots, l - 1, l]$ and thus I_z has $2l + 1$ possible energy values in the presence of a magnetic field. For common biomolecular NMR nuclei such as ^1H , ^{13}C , and ^{15}N , I_z has two values: $m = -1/2$ and $m = +1/2$. Nuclei with odd mass numbers have half-integral spin numbers (e.g. $l = 1/2$ for ^1H , ^{13}C , ^{15}N , ^{19}F , and ^{31}P ; $l = 3/2$ for ^{23}Na), whereas nuclei with even mass numbers and odd or even atomic numbers respectively have integral spin numbers (e.g. $l = 1$ for ^2H , ^{14}N) or spin numbers equal to zero (e.g. $l = 0$ for ^{12}C , ^{32}S). Because of the high natural abundance of ^1H (99.99%) and its $l = 1/2$ value, biomolecules can be studied by NMR spectroscopy without isotopic enrichment. However, to perform heteronuclear NMR on biomolecules in the solution-state, samples are typically enriched in ^{13}C and ^{15}N , which respectively have natural abundances of 1.07 and 0.37%. Natural abundance NMR studies making use of ^{13}C are possible, although isotope enrichment provides a more convenient environment for spin gymnastics.

Table 1. Properties of some nuclei that are used in biomolecular NMR studies, including the gyromagnetic ratio (γ), relative sensitivity to ^1H ($\gamma_x/\gamma_{1\text{H}}$), natural abundance, and spin number. Note that ^3H , commonly referred to as tritium, has a higher gyromagnetic ratio than ^1H , but is radioactive and extremely rare in nature (trace). The isotopes ^2H , ^{13}C , ^{15}N , ^{17}O require biosynthetic enrichment, enabling near 100% incorporation into biomolecules.

Nucleus	γ (rad s ⁻¹ G ⁻¹)	$\gamma_x / \gamma_{1\text{H}}$	Natural abundance (%)	Spin (I)
^1H	26,753	1	99.980	1/2
^2H	4,106	0.153	0.016	1
^3H	28,535	1.067	trace	1/2
^{13}C	6,728	0.251	1.108	1/2
^{15}N	-2,712	0.101	0.370	1/2
^{17}O	-3,628	0.136	0.038	5/2
^{19}F	25,179	0.941	100.000	1/2
^{23}Na	7,081	0.265	100.000	3/2
^{31}P	10,841	0.405	100.000	1/2

In the absence of an applied external magnetic field, the $2I + 1$ values of I_z have identical energies. However, upon application of an external magnetic field, the energies of the $2I + 1$ spin states split into equally spaced, quantized energy levels known as the Zeeman levels:

$$E = -\boldsymbol{\mu} \cdot \mathbf{B} \quad (3)$$

$$E_m = -m\hbar\gamma B_0 \quad (4)$$

where \mathbf{B} is the vector associated with static magnetic field strength along, generally oriented in the z-direction of the laboratory frame (B_0), $\boldsymbol{\mu}$ ($= \gamma I$) is a vector describing the magnetic moment, and γ is the gyromagnetic ratio of the nucleus of interest (Table 1). For example, the two spin states associated with $I = 1/2$ nuclei differ in energy (ΔE) by $\hbar\gamma B_0$. When γ is positive, the lower energy state ($m = 1/2$) that aligns with the static magnetic field is the α state, whereas the higher energy state ($m = -1/2$) is referred to as the β state. Due to the lower energy associated with the α state, a larger number of nuclei preferentially adopt this state, leading to a population difference between the α and β states that scales with the Boltzmann distribution:

$$e^{-\frac{\Delta E}{2k_B T}} = e^{-\frac{\hbar\gamma B_0}{2k_B T}} \quad (5)$$

On the other hand, when γ is negative (e.g. for ^{15}N) the β state is lower in energy. Thus, the population difference between α and β states increases with the gyromagnetic ratio and static magnetic field strength. Such a low population difference renders NMR an insensitive spectroscopic technique; yet, considerable methodological advances in hardware, software, pulse programming, and biosynthetic strategies have established NMR spectroscopy as an indispensable biophysical tool used in laboratories across the world.

2.2 Nuclear spin relaxation

Above, a formalism was outlined that describes the evolution of magnetization during a pulse sequence. At the same time, excited nuclear spins experience relaxation back to equilibrium, and measurement of nuclear spin relaxation rates can provide atomic-level information about biomolecular dynamics across a wide range of timescales. Relaxation of nuclear spins occurs through dipole-dipole (DD) and chemical shift anisotropy (CSA) interactions, which cause fluctuations in the local magnetic field about a given nucleus. In order to induce transitions between spin states, the magnetic field fluctuations must occur at particular frequencies.

2.2.1 Spectral density functions

One can describe the intensity of magnetic field fluctuations at a given frequency, ω , with the spectral density function, $J(\omega)$:

$$J(\omega) = \int_0^{\infty} g(\tau) e^{-i\omega\tau} d\tau \quad (78)$$

where $g(\tau)$ is the normalized autocorrelation function, for isotropic tumbling due to Brownian motion:

$$g(\tau) = \frac{1}{5} e^{-\frac{\tau}{\tau_c}} \quad (79)$$

integrating the above function yields the spectral density function for isotropic tumbling:

$$J(\omega) = \frac{2}{5} \left(\frac{\tau_c}{1 + \omega^2 \tau_c^2} \right) \quad (80)$$

which depends on the rotational correlation time (τ_c), the time required for a molecule to rotate by approximately one radian (Fig. 2).

The autocorrelation function depends on the rotational diffusion of the molecule and the presence of any internal motions that are faster than the rotational correlation time (τ_c). Knowledge of the rotational diffusion behavior (e.g. isotropic, axially symmetric, or anisotropic) must be determined in order to accurately interpret relaxation data.

Motional model	Autocorrelation function	Spectral density function
Isotropic	$g(\tau) = \frac{1}{5} e^{-\frac{\tau}{\tau_c}}$	$\frac{2}{5} \left(\frac{\tau_c}{1 + \omega^2 \tau_c^2} \right)$
Isotropic, fast internal motions	$g(\tau) = \frac{1}{5} e^{-\frac{\tau}{\tau}}$	$\frac{2}{5} \left(\frac{\tau_c}{1 + \omega^2 \tau^2} \right)$
Isotropic, fast/spatially restricted internal motions	$g(\tau) = \frac{1}{5} \left(S^2 e^{-\frac{\tau}{\tau_m}} + (1 - S^2) e^{-\frac{\tau}{\tau}} \right)$	$\frac{2}{5} \left(\frac{S^2 \tau_c}{1 + \omega^2 \tau_c^2} + \frac{[1 - S^2] \tau}{1 + \omega^2 \tau^2} \right)$
Axially symmetric	$g(\tau) = \frac{1}{5} \left(A_1 e^{-\frac{\tau}{\tau_1}} + A_2 e^{-\frac{\tau}{\tau_2}} + A_3 e^{-\frac{\tau}{\tau_3}} \right)$	$\frac{2}{5} \sum_{m=1}^3 A_m \left(\frac{\tau_j}{1 + \omega^2 \tau_j^2} \right)$
Anisotropic	$g(\tau) = \frac{1}{5} \sum_{m=1}^5 A_m e^{-\frac{\tau}{\tau_m}}$	$\frac{2}{5} \sum_{m=1}^5 A_m \left(\frac{\tau_j}{1 + \omega^2 \tau_j^2} \right)$

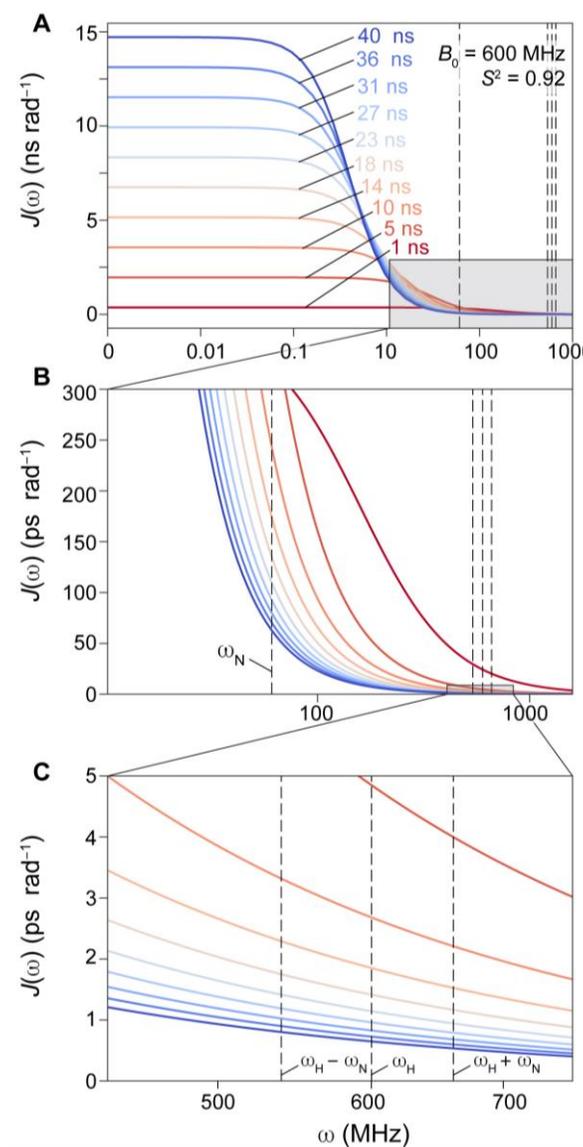


Figure 2. Spectral density functions simulated for various correlation times. (A, B, C) Isotropic tumbling was assumed and the $J(\omega)$ values for a N-H vector were simulated as a function of increasing correlation time, ranging from 1 ns (red) to 40 ns (blue) for a fixed S^2 of 0.92. As can be seen, slowly tumbling molecules have higher amplitude motions at lower frequencies (A), whereas faster tumbling molecules have higher amplitude motions at high frequencies (B, C). The frequencies for ω_N , ω_H , ω_{H-N} , and ω_{H+N} are indicated assuming $B_0 = 14.1$ T.

2.2.2 Spin interactions and transverse and longitudinal relaxation

Nuclear spin relaxation rate constants for the S spin in an I–S spin system, where I = ^1H and S = ^{15}N in this example, are described by the following set of equations (Cavanagh et al. 2007):

$$d = \left(\frac{\mu_0 \hbar \gamma_N \gamma_H}{4 r_{H,N}^3} \right)^2 \quad (81)$$

$$c = \frac{\omega_N^2 \Delta\sigma^2}{3} \quad (82)$$

$$R_1 = d [J(\omega_H - \omega_N) + 3J(\omega_N) + 6J(\omega_H + \omega_N)] + cJ(\omega_N) \quad (83)$$

$$R_2 = \frac{d}{2} [4J(0) + J(\omega_H - \omega_N) + 3J(\omega_N) + 6J(\omega_H) + 6J(\omega_H + \omega_N)] + \frac{c}{6} [4J(0) + 3J(\omega_N)] \quad (83)$$

$$NOE = 1 + d \frac{\gamma_H}{\gamma_N} [(6J(\omega_H + \omega_N) - J(\omega_H - \omega_N))] T_1 \quad (84)$$

$$J(0) = \frac{2}{5} (S^2\tau_c + [1 - S^2]\tau_e) \quad (85)$$

$$J(\omega) = \frac{2}{5} \left(\frac{S^2\tau_c}{1 + \omega^2\tau_c^2} + \frac{[1 - S^2]\tau_e}{1 + \omega^2\tau_e^2} \right) \quad (86)$$

with d denoting the dipolar interaction constant that contains the terms: μ_0 , the permittivity of free space; \hbar , the reduced Planck's constant; γ_H and γ_N , the respective gyromagnetic ratios of ^1H and ^{15}N ; and $r_{H,N}$, the internuclear H–N bond length. The constant for CSA, c , includes ω_N , the Larmor frequency (rad s^{-1}) of ^{15}N , and $\Delta\sigma$, the CSA of the ^{15}N spin. R_1 and R_2 respectively stand for the longitudinal relaxation rate constant (*i.e.* T_1^{-1}) and transverse relaxation rate constant (*i.e.* T_2^{-1}), and NOE for the steady-state value of the nuclear Overhauser enhancement. S^2 , τ_c , and τ_e are the order parameter, rotational correlation time, and fast internal correlation time.

Transverse and longitudinal relaxation manifest via differently weighted, linear contributions of spectral density functions at discrete frequencies, with R_1 depending on $J(\omega_H - \omega_N)$, $J(\omega_H + \omega_N)$, and $J(\omega_N)$, whereas R_2 depends on the same spectral density functions supplemented by $J(\omega_H)$ and $J(0)$. The steady-state NOE depends on R_1 and spectral density functions at high frequencies, $J(\omega_H - \omega_N)$ and $J(\omega_H + \omega_N)$. The spectral density functions, of course, depend on the rotational correlation time and rotational diffusion of the molecule. Therefore, residues within a protein that exhibit different motional behavior can be identified based on their specific relaxation rates. For example, ^{15}N relaxation studies identified a flexible linker within the protein calmodulin that was originally thought to be a stable α -helix (Barbato et al. 1992).

It is apparent from the above equations that R_1 , R_2 , and NOE values change with the static magnetic field strength, encoded both by changes to the Larmor frequencies of I (^1H) and S (^{15}N) spins and due to the CSA term. Additionally, because R_2 depends on $J(0)$, which increases with the τ_C of a macromolecule, the measurement of R_1 and R_2 rate constants can provide insight into the rotational correlation time of the molecule (Kay et al. 1989) (Fig. 3). The dependence of the R_2/R_1 ratio on the B_0 strength is shown below (Fig. 4).

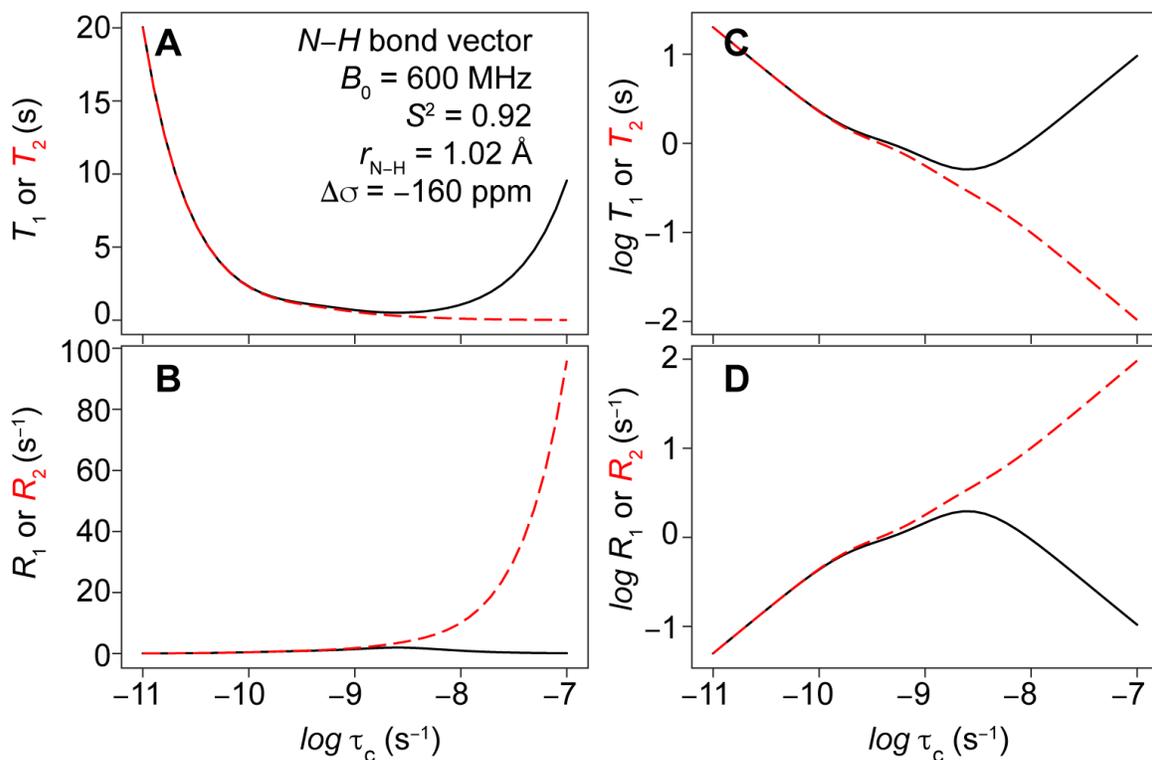


Figure 3. Dependence of ^{15}N transverse and longitudinal relaxation rates on the correlation time. Relaxation rates or times are shown for an N-H bond vector at a static magnetic field strength of 14.1 T (600 MHz), with a fixed S^2 of 0.92, a bond length of 1.02 Å, and a ^{15}N CSA of -160 ppm. **(A)** The calculated T_1 (black) or T_2 (red, dashed) times as a function of increasing correlation time. **(B)** The same as A, but the rates are plotted. **(C, D)** The same as A and B, except for values on the y-axes have been computed as the logarithm.

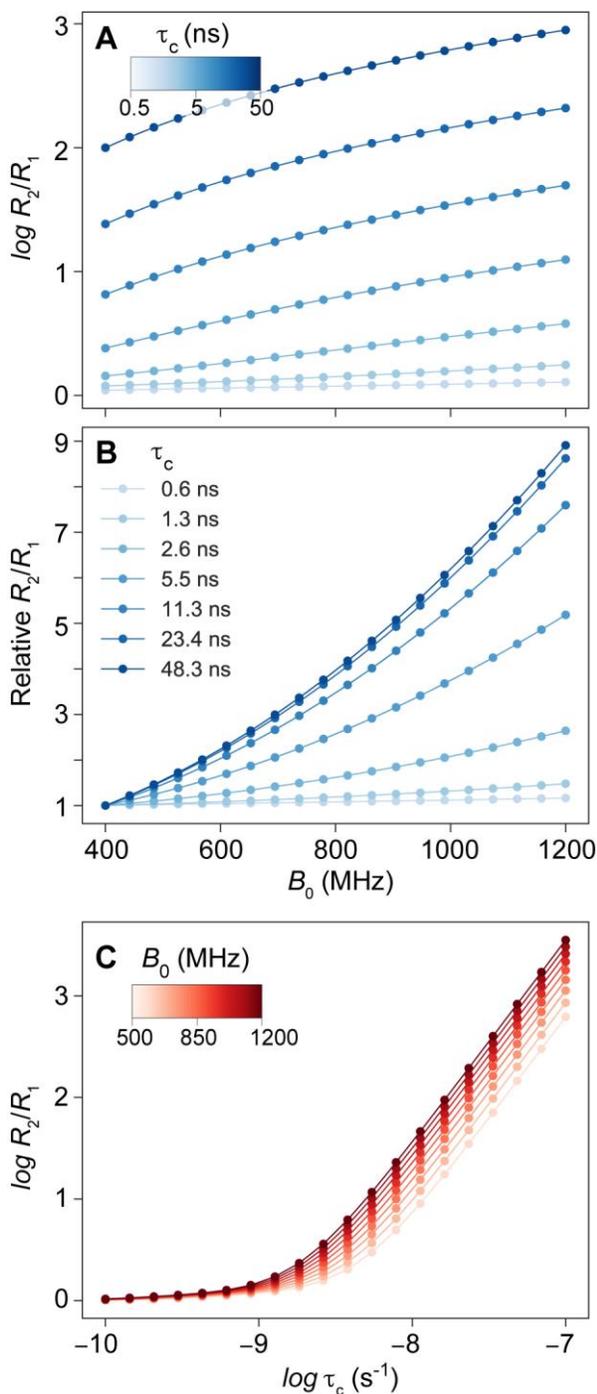


Figure 4. ^{15}N R_2/R_1 ratios are sensitive to the correlation time and magnetic field strength. (A) The logarithm of ^{15}N R_2/R_1 ratios is computed as a function of the static magnetic field strength (x-axis) and the rotational correlation time (color gradient). As can be seen, the ratio depends on the B_0 field and rotational correlation time. (B) For each rotational correlation time, the data in panel A were normalized to equal 1 at the lowest B_0 . The growth in ^{15}N R_2/R_1 ratios with increasing magnetic field strength differs for differing rotational correlation times. (C) The same data in panel B, but shown as a function of increasing correlation time (x-axis) for various B_0 fields (color gradient).

Because NMR signals decay with an exponential factor $e^{-R_2 t}$, linewidths become broader as the transverse relaxation time gets shorter, or as R_2 gets larger. The τ_c roughly scales with the molecular mass of a biomolecule, and thus, among other factors, an estimate of τ_c can be useful in determining the feasibility of an NMR project. An estimate of τ_c can be obtained from Stoke's

Law, which relates the isotropic τ_c to the solution viscosity (η), the effective hydrodynamic radius (R_h), and the temperature (T):

$$\tau_c \approx \frac{4\pi \eta R_h^3}{3 k_B T} \quad (87)$$

This expression assumes that the particle of interest is spherical and undergoes isotropic rotational diffusion. For many proteins, neither of these assumptions will be true, as evidenced by the experimentally observed anisotropic behavior of the small, model protein ubiquitin (76-residue, 8.4 kDa protein) (Tjandra et al. 1995). Nevertheless, using Stoke's Law to estimate τ_c can provide useful preliminary insight. The R_h of a folded protein can be estimated based on its molecular mass (M), the average density of a protein (ρ , 1.37 g cm⁻³), Avogadro's number (6.022×10^{23} mol⁻¹), and the depth of the hydration shell (r_w , ranging from one-half to one water molecule deep, 1.6 to 3.2 Å) (Cavanagh et al. 2007). Alternatively, an empirical relationship has been derived to obtain the expected R_h based on the number of amino acids in a protein (AA) (Wilkins et al. 1999). Both of these equations provide similar R_h values, e.g. for the 76-residue protein, the equations yield $R_h = 16.7$ Å (empirical) and $R_h = 15.8$ Å (volume-based), with the experimentally measured R_h of ubiquitin at 298 K equal to 15.7 Å (e.g. see Chapter 6).

$$R_h \approx 4.75 AA^{0.29} \quad (88)$$

$$R_h \approx \frac{3M}{\sqrt[3]{4\pi\rho N_A}} + r_w \quad (89)$$

As can be noted from the above equations, the rotational correlation time is inversely related to temperature; that is, τ_c decreases at higher temperatures and increases at lower temperatures. Importantly, the viscosity of the solution, which also depends strongly on temperature, additionally impacts τ_c and therefore the relaxation rates (Fig. 5). As such, it is sometimes favorable to collect NMR data at elevated temperatures where viscosity is lower.

However, in some cases, e.g. when using backbone N-H-based experiments, elevated amide exchange rates under such conditions can be prohibitive.

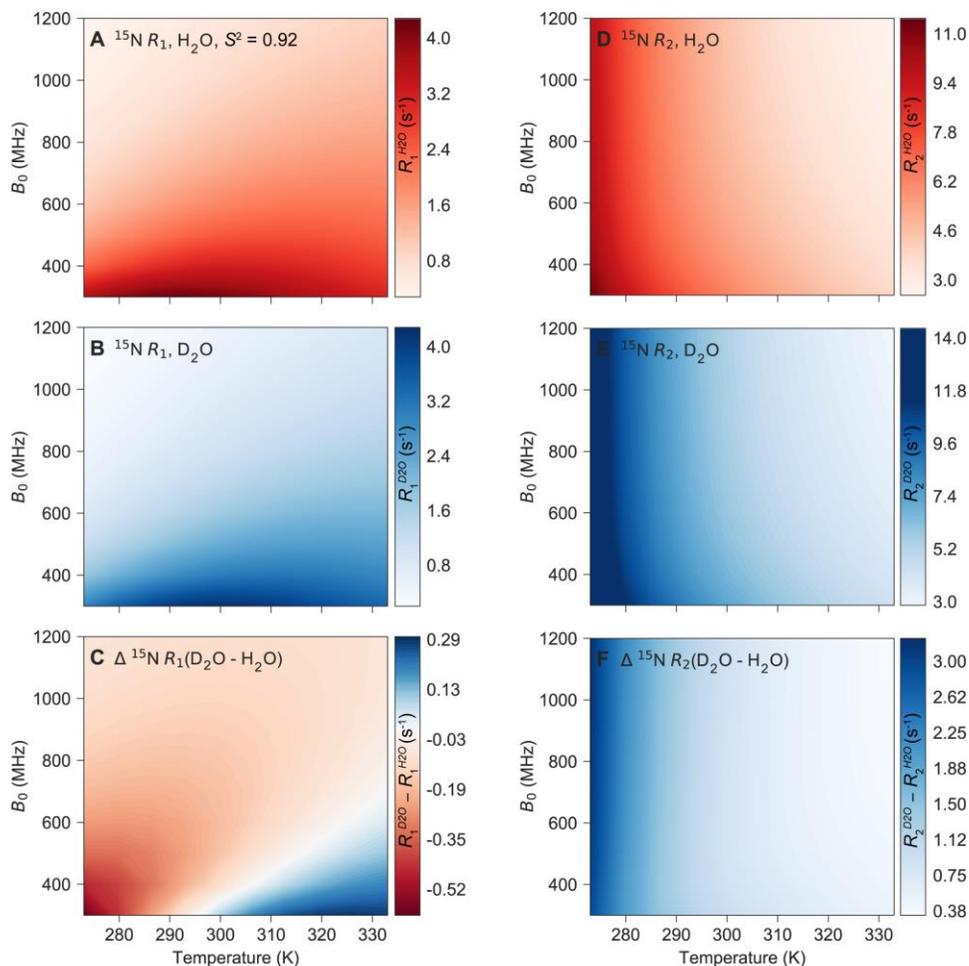


Figure 5. Calculation of $^{15}\text{N } R_1$ and R_2 rates as a function of temperature and B_0 in different solvents. Using the equations outlined above, $^{15}\text{N } R_1$ and R_2 rates can be calculated for a given rotational correlation time, S^2 , B_0 strength, N-H bond length, and ^{15}N CSA value. Given the ability to predict the rotational correlation time from Stoke's Law or empirically, and assuming the change in a solvent's viscosity with temperature is known, one can compare the relaxation behavior of a molecule in various solvents as a function of temperature. (A) $^{15}\text{N } R_1$ rates in H_2O as a function of temperature (x-axis) and B_0 strength (y-axis). The R_1 values increase at lower B_0 fields and higher temperatures. The known temperature dependence of the viscosity of water was used (Huber et al. 2009) (B) The same calculation as A, but in $^2\text{H}_2\text{O}$, *i.e.* D_2O . The viscosity of D_2O differs from H_2O , as does the temperature dependence. (C) The difference in R_1 rates ($\text{D}_2\text{O} - \text{H}_2\text{O}$). (D-F) The same as A-C, but for $^{15}\text{N } R_2$. A solvent-exposed N-H group would not be observable in D_2O , whereas a buried N-H group could be observable. This comparison is mainly for illustrative purposes.

2.3 Chemical exchange and NMR spectroscopy

In solution, molecules dynamically interconvert between two (or more) conformations, including aromatic ring flips in small molecules and proteins ([Wuthrich & Wagner 1975](#)), the puckering of nucleic acids ([Al-Hashimi 2013](#)), or conformational changes during enzymatic catalysis ([Lisi & Loria 2017](#)). The free energy differences between such conformations are typically on the order of a few kcal mol⁻¹ at room temperature, leading to a highly-skewed population difference between the lower energy (major, or ground) and higher energy (minor, or conformationally excited) state ([Baldwin & Kay 2009a](#), [Korzhnev & Kay 2008a](#)).

The inherent dynamics and low populations of such systems often render minor states “invisible” to crystallographic methods and undetectable by most biophysical methods. For example, to enable detection of a minor protein state by intrinsic tryptophan fluorescence, the indole ring must exist in different chemical environments in each state: in proteins with one tryptophan, however, a subtle structural rearrangement therefore might remain invisible. NMR spectroscopy, which provides a spectroscopic probe at nearly every moiety, is highly sensitive to dynamic processes across a wide-range of timescales, and has proven indispensable for the characterization of biomolecular chemical exchange ([Palmer 2014](#)). Importantly, chemical exchange-based NMR methods do not necessitate *direct* observation of resonances from the lowly populated conformation; rather, *indirect* characterization of the minor state is enabled through encoding or transferring of the magnetization properties from the minor state to the resonances of the visible, major state ([Palmer et al. 2001](#), [Vallurupalli et al. 2017](#)). This enables detection of short-lived, minor conformations that are populated to as little as 0.5%, ranging from atomic-level differences in protein folding intermediates ([Neudecker et al. 2009](#), [Zhuravleva & Korzhnev 2017](#)) to the kinetics of exchange between amyloid- β monomers and proto-fibrils ([Fawzi et al. 2011](#)). These and other applications of chemical exchange-based NMR spectroscopy in proteins and nucleic acids have led to unprecedented atomic-level insight into transiently

populated, minor biomolecular conformations (Bothe et al. 2011, Bouvignies et al. 2011, Hansen et al. 2008b, Korzhnev et al. 2004, 2010; Palmer 2015, Vallurupalli et al. 2008).

The utilization of modern, chemical exchange-based NMR experiments does not require detailed knowledge of the underlying theory (Bain 2003). However, an understanding of the theoretical considerations provides more insight into the outcomes and potential limitations of the experiments. As such, some theoretical considerations involving chemical exchange and its manifestation in NMR spectroscopy are outlined below.

For a system undergoing two-state chemical exchange, $A \rightleftharpoons B$, where the forward arrow denotes the rate constant (in s^{-1}) k_{AB} and the backward arrow k_{BA} (in s^{-1}), the linewidths, intensities, and chemical shifts of both resonances can be significantly impacted (Palmer et al. 2001). As will be discussed below, biomolecular NMR typically examines skewed population differences, whereby the population of state B (p_B) is far less than that of state A (p_A), *i.e.* $p_B \ll p_A$ or $p_B \ll 1$, with $p_A + p_B = 1$. For the discussion below, a case of equal populations is considered in which $p_A = p_B = 0.5$. This is mainly for illustrative purposes and is not particularly common for populations of proteins and nucleic acids.

The extent of NMR signal modulation caused by chemical exchange depends on the exchange regime, defined by $\theta = \frac{k_{ex}}{|\Delta\omega|}$, where $|\Delta\omega|$ stands for the absolute value of the difference in chemical shifts for states A and B (in $rad\ s^{-1}$) and k_{ex} for the interconversion rate, or the sum of k_{AB} and k_{BA} (s^{-1}). A parameter termed α has been established (Millet et al. 2000), which varies from 0 – 2 (Fig. 6) and also characterizes the exchange regime:

$$\alpha = \frac{2 \left(\frac{k_{ex}}{|\Delta\omega|} \right)^2}{1 + \left(\frac{k_{ex}}{|\Delta\omega|} \right)^2} \quad (90)$$

$0 < \alpha < 1$ = slow exchange

$\alpha \sim 1$ = intermediate exchange

$1 < \alpha < 2$ = fast exchange

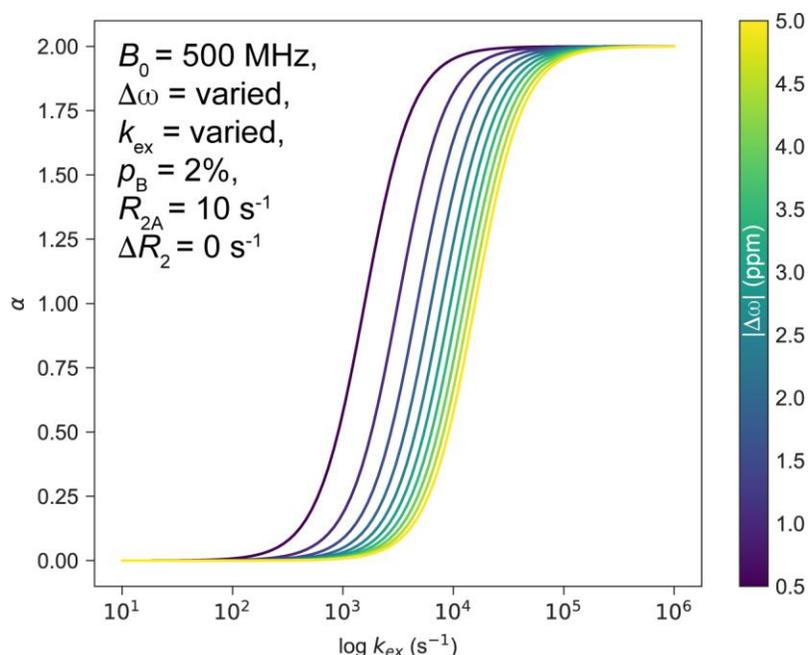


Figure 6. The exchange regime as defined by the parameter α . Simulations were performed using the parameters in the upper left-hand corner as a function of varying $|\Delta\omega|$ value (colors). The values of k_{ex} were systematically varied to alter the exchange regime. CPMG RD experiments are only sensitive to k_{ex} values in the range of *ca.* 10^2 to $10^{3.8}$ s^{-1} , and thus intermediate exchange is accessible to $|\Delta\omega|$ values that are < 2.5 ppm (1H) under these conditions. Collecting data at higher B_0 values will shift the curves to the right, as increasing B_0 effectively lowers the value of α .

In slow exchange, when $k_{ex} \ll |\Delta\omega|$ and $\theta \ll 1$, two separate resonances are observed for states A and B, with chemical shifts at ω_A and ω_B , respectively. Under such conditions, the intensities of each chemical shift report on the respective populations of the two states (Fig. 7 A). Longitudinal exchange spectroscopy can be employed to investigate the rate constants associated with slow exchange behavior. Experiments such as zz-exchange (exchange spectroscopy, EXSY) (Farrow et al. 1994) and real-time NMR provide insight into reactions with k_{ex} in the *ca.* 0.5 – 10 s^{-1} range (Kleckner & Foster 2011), including slow protein folding reactions and interconversion between other structural forms. Chemical exchange saturation transfer (CEST) (Vallurupalli et al. 2012, 2017) or analysis of lineshapes can also be used to characterize exchange in the *ca.* 10 – 100 s^{-1} regime, *e.g.* by monitoring changes to resonances during a titration with ligand. Slow exchange contributions to the linewidth of an observed resonance, R_{ex} , scale

independently of the magnetic field strength (Palmer et al. 2001), and vary with the rate k_{AB} ($= p_B k_{ex}$). The expression for R_{ex} in the slow exchange regime is:

$$R_{ex} = p_B k_{ex} \quad \text{as} \quad \frac{k_{ex}}{|\Delta\omega|} \rightarrow 0 \quad (91)$$

As k_{ex} further increases to fast-intermediate ($k_{ex} > |\Delta\omega|$) and fast exchange ($k_{ex} \gg |\Delta\omega|$) (Fig. 7 G–I), the resonance regains intensity and approaches the chemical shift of the population-weighted averaged ($p_A\omega_A + p_B\omega_B$), where p_X reflects the population of state X. When exchange is sufficiently fast, the intensity of the population-weighted chemical shift will be equivalent to that of the sum of the initial conditions (Fig. 7 I). For a ^1H spin undergoing two-state chemical exchange with a $|\Delta\omega|$ of 2.5 ppm at 500 MHz (7853 rad s^{-1}), k_{ex} must exceed ca. 7850 s^{-1} before α exceeds 1. When $k_{ex} > |\Delta\omega|$, in the intermediate-fast to fast exchange regime, experiments such as paramagnetic relaxation enhancement (PRE) and $R1\rho$ relaxation dispersion offer the ability to characterize exchange reactions in the 10,000–50,000 ($R1\rho$) (Palmer & Massi 2006) and 100,000 s^{-1} (PRE) range (Cloue et al. 2007). R_{ex} also scales quadratically with the magnetic field strength under fast exchange conditions (Palmer et al. 2001), providing a useful diagnostic of the exchange regime.

$$R_{ex} = \frac{p_A p_B |\Delta\omega|^2}{k_{ex}} \quad \text{as} \quad \frac{k_{ex}}{|\Delta\omega|} \rightarrow \infty \quad (92)$$

Between these two scenarios exists the intermediate exchange regime, in which $k_{ex} \sim |\Delta\omega|$. Under such conditions, the intensities no longer quantitatively reflect the population of each state (Fig. 7) and severe signal attenuation occurs due to line broadening from chemical exchange. At $k_{ex} \sim |\Delta\omega|$, the two signals coalesce into one, significantly broadened resonance. Depending on the nucleus of interest, the static magnetic field strength, and the difference in chemical shifts between the two conformations, intermediate exchange typically falls in the μs – ms regime ($k_{ex}^{-1} = \tau_{ex}$), a range where many interesting biological dynamics take place (Korzhnev & Kay 2008b, Lisi & Loria 2017). Processes such as allosteric changes, loop fluctuations, protein

folding/misfolding, and enzymatic catalysis can occur on the μs – ms timescale, and thus it proves critical to understand the structural, kinetic, and thermodynamic aspects of these molecular fluctuations. As previously mentioned, $R_{1\rho}$ relaxation dispersion is sensitive to exchange events in the range of $10,000$ – $50,000\text{ s}^{-1}$ (Palmer & Massi 2006), which may encompass intermediate exchange for nuclei with large $|\Delta\omega|$ values and higher gyromagnetic ratios (e.g. ^1H with $|\Delta\omega| = 2$ ppm at 900 MHz, 11309 rad s^{-1}). For nuclei with lower gyromagnetic ratios and $|\Delta\omega|$ values, $R_{1\rho}$ relaxation dispersion offers a route to characterize intermediate-fast/fast exchange events, as originally outlined for ^{15}N amides in the model protein ubiquitin.

Carr-Purcell-Meiboom-Gill relaxation dispersion (CPMG RD) (Baldwin 2014, Carr & Purcell 1954, Carver & Richards 1972, Hansen et al. 2008a, Loria et al. 1999, Meiboom & Gill 1958), in contrast, is sensitive to k_{ex} values in the ca. 200 – 2000 s^{-1} window, with additional information from the exchange-induced shift (Vallurupalli et al. 2011) (*vide infra*) and differences in chemical shifts of HSQC and HMQC spectra (Skrynnikov et al. 2002) increasing the upper range to ca. 6000 s^{-1} . Therefore, CPMG RD provides an experimental method to probe dynamics in the millisecond regime; spectroscopically, this enables characterization of intermediate exchange for nuclei of lower gyromagnetic ratios and smaller $|\Delta\omega|$ values (e.g. ^{15}N with $|\Delta\omega| = 2$ ppm at 900 MHz, 1146 rad s^{-1}). CPMG RD has blossomed into a powerful NMR method to describe the structures of lowly populated biomolecular conformations present at as little as 0.5% (Baldwin & Kay 2009b, Hansen et al. 2008c, Korzhnev & Kay 2008b, Vallurupalli et al. 2008), and the kinetics and thermodynamics that govern their chemical exchange with the observable ground state. CPMG RD was widely employed in this thesis, and therefore its theoretical underpinnings will be discussed below.

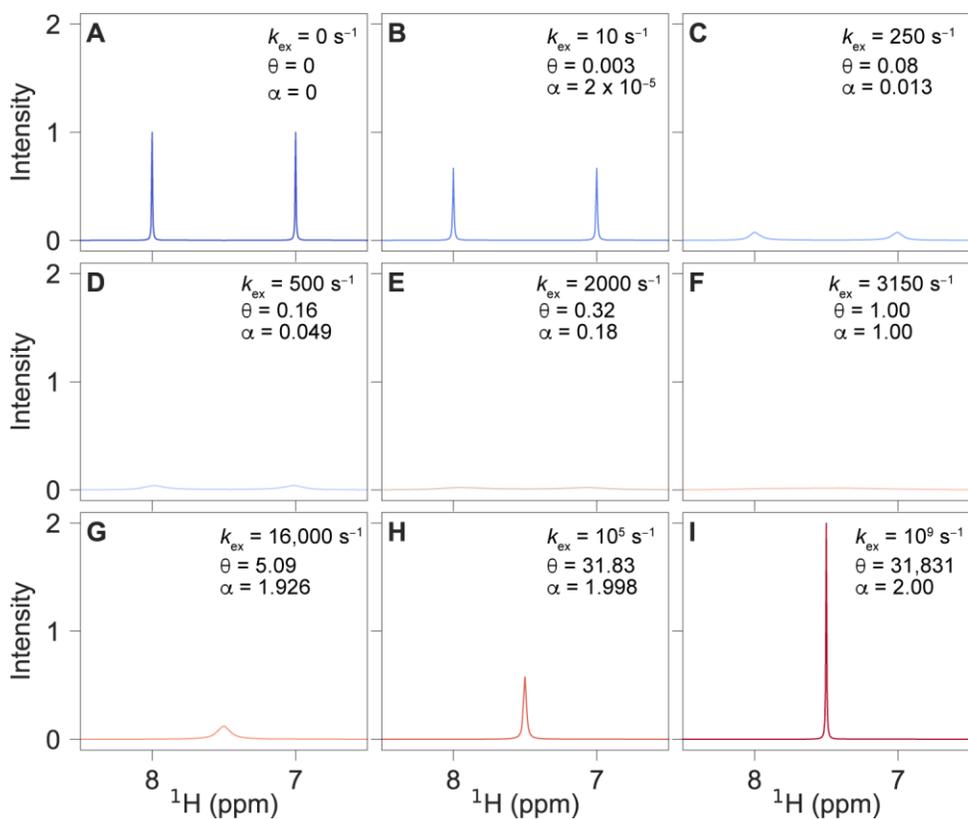


Figure 7. Simulated 1D ¹H NMR spectra for a system undergoing two-state chemical exchange. The parameters for the simulation are: $p_A = p_B = 50\%$, $B_0 = 500$ MHz, $|\Delta\omega| = 1$ ppm (3141.6 rad s^{-1}) in ¹H, $R_{2A} = R_{2B} = 10$ s^{-1} . The rate of interconversion between the two states, k_{ex} , is equal to 0 in panel **A**, but is increased in panels **B** – **I**. The exact k_{ex} values are listed in the upper-right corner of each plot. The estimate of the exchange regime, θ and α , are included. See text for their definitions. The semi-transparent black spectrum shown in panels (B-I) is the original spectrum in the absence of exchange (A). Unequal populations are shown in [Fig. 8](#) and [Fig. 9](#).

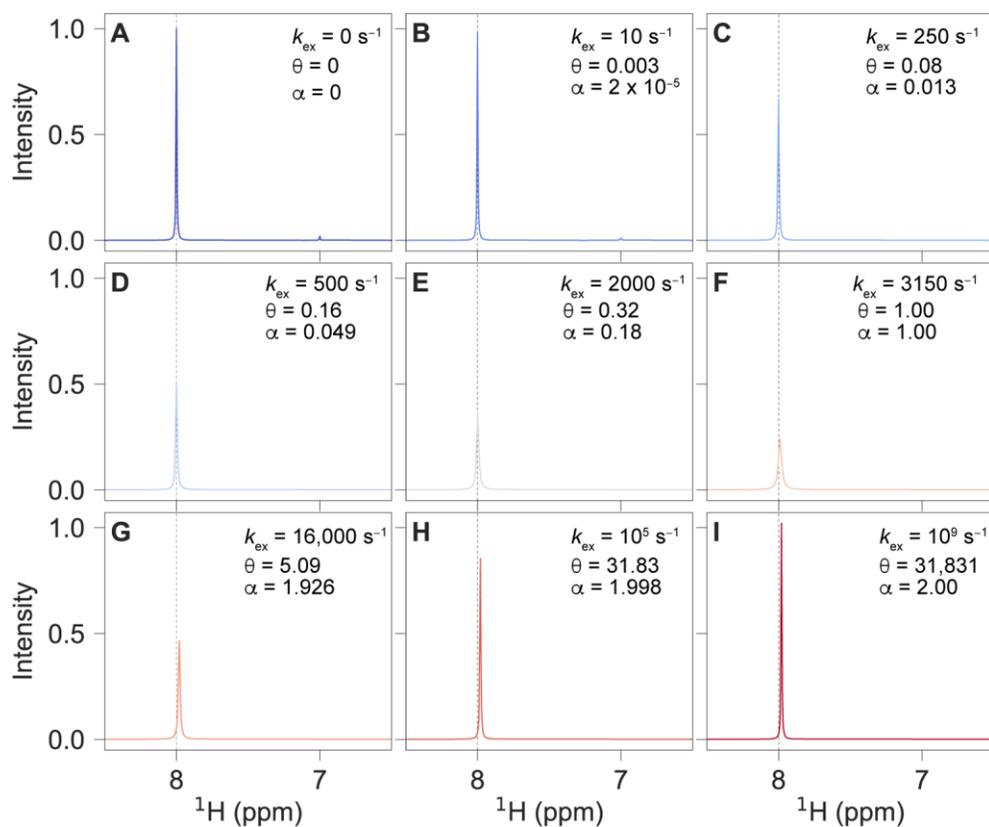


Figure 8. Simulated 1D NMR spectra as in Fig. 7, except that $p_A = 98\%$ and $p_B = 2\%$. The dashed line at 8 ppm indicates the original position of ω_A , with deviations at higher k_{ex} values caused by the exchange-induced shift (see text). Note that the lone observable resonance in panels C-G is significantly broader than the first and last panels. This is caused by exchange-induced broadening (see text). Zoomed-in and overlaid spectra are shown in Fig. 9.

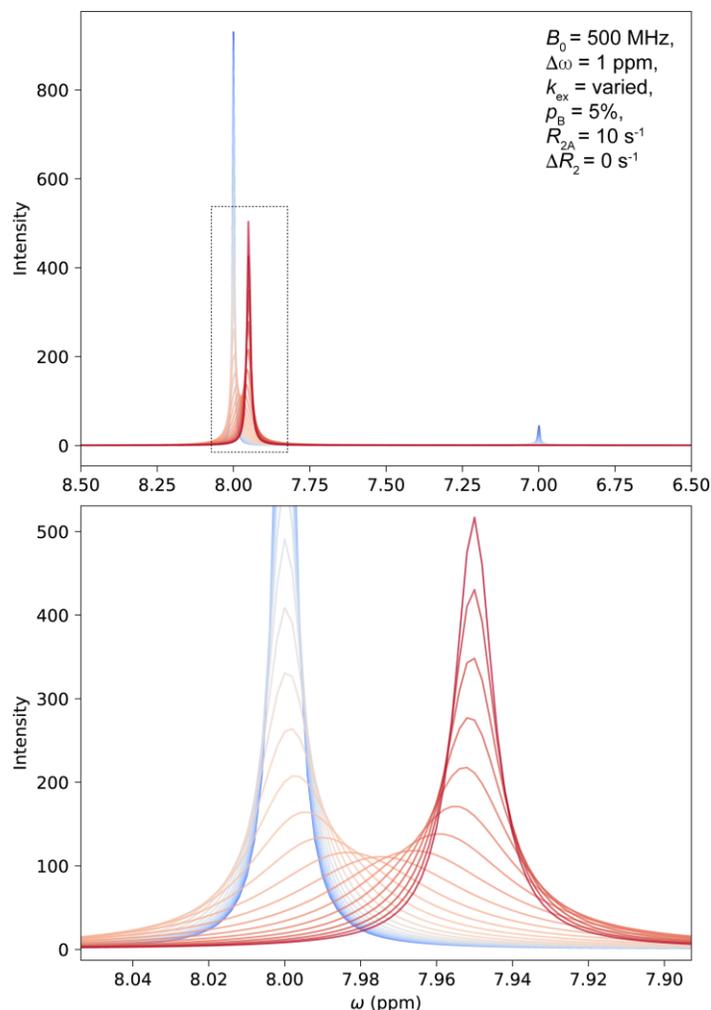


Figure 9. Simulated 1D NMR spectra as in Fig. 8, except with all spectra overlaid and k_{ex} reaching only to $20,000 \text{ s}^{-1}$. In addition, $p_A = 95\%$ and $p_B = 5\%$. The bottom panel shows a zoomed-in region of the top panel (boxed), revealing the exchanged-induced shift and broadening of the major resonance. In the fast exchange regime at higher k_{ex} values (red), the population-averaged resonance appears at 7.95 ppm ($p_A\omega_A + p_B\omega_B = 7.95$).

2.4 Carr-Purcell-Meiboom-Gill relaxation dispersion

CPMG RD experiments utilize a variable number of spin echoes during a fixed, constant-time relaxation delay (T_{CPMG}) to monitor the effective transverse relaxation rate constant $R_{2,\text{eff}}$ as a function of the number of 180° refocusing pulses (Hansen et al. 2008a, Loria et al. 1999, Palmer et al. 2001). CPMG RD data can be analytically modeled with the Carver-Richards equation (Carver & Richards 1972, Palmer et al. 2001):

$$R_{2,\text{eff}} = \frac{1}{2} \left(R_{2,A} + R_{2,B} + k_{\text{ex}} - \left(\frac{1}{2\tau} \right) \cosh^{-1} (D_+ \cosh(\eta_+) - D_- \cos(\eta_-)) \right) \quad (93)$$

$$D_{\pm} = \frac{1}{2} \left(\pm 1 + \frac{\Psi + 2\Delta\omega^2}{\sqrt{\Psi^2 + \zeta^2}} \right)$$

$$\eta_{\pm} = \frac{2\tau}{\sqrt{2}} \sqrt{\pm\Psi + \sqrt{\Psi^2 + \zeta^2}}$$

$$\Psi = (R_{2,A} - R_{2,B} - p_A k_{\text{ex}} + p_B k_{\text{ex}})^2 - \Delta\omega^2 + 4p_A p_B k_{\text{ex}}^2$$

$$\zeta = 2\Delta\omega + (R_{2,A} - R_{2,B} - p_A k_{\text{ex}} + p_B k_{\text{ex}})$$

where $R_{2,\text{eff}}$ refers to the effective transverse relaxation rate constant, $R_{2,A}$ to the intrinsic R_2 value of state A (also noted as $R_{2,A}^0$ in the literature), $R_{2,B}$ to the intrinsic R_2 value of state B (also noted as $R_{2,B}^0$ in the literature), p_A to the population of state A, p_B to the population of state B ($p_B = 1 - p_A$), k_{ex} to the rate of interconversion between states A and B ($k_{\text{ex}} = k_{AB} + k_{BA}$), $\Delta\omega$ to the difference in chemical shifts between states A and B (in rad s^{-1}), and τ to the time between the centers of successive 180° refocusing pulses. Analytical expressions for three-site (Grey et al. 2003) and N -site chemical exchange (Koss et al. 2018) have also been reported, but will not be discussed here.

As has been demonstrated previously (Baldwin 2014), and also below, modeling CPMG RD data using the approximate Carver-Richards equation introduces errors into the analysis. These errors can significantly exceed the experimental error in the experiment (Baldwin 2014), and it is therefore not recommended to fit CPMG RD data to the Carver-Richards equation. Nonetheless, despite the author's opinion that it is difficult to correctly implement the Carver-Richards equation (*i.e.* by avoiding typos due to the many signed quantities), many publications report fitting to this equation (Narayanan et al. 2018, Nunes et al. 2018). Indeed, many softwares that have been developed to fit CPMG RD data rely on the Carver-Richards equation (Bieri & Gooley 2011, Kleckner & Foster 2012, Mazur et al. 2013, Sugase et al. 2013). A better solution

for fitting CPMG RD data is the usage of a Bloch-McConnell exchange matrix, as described in previous publications (Hansen et al. 2008c, Vallurupalli et al. 2008). The full matrix amounts to a square matrix of size 31-by-31, and captures all necessary spin physics, including the impact of off-resonance effects. However, the large matrix size can lead to slow calculation times due to the need for many iterations of matrix exponentiation.

Here, only the simplest 2-by-2 Bloch-McConnell matrix is considered, which enables calculation of the evolution of the transverse magnetization during the constant-time period in a CPMG RD experiment (Palmer et al. 2001). This analysis utilizes the fact that observed magnetization at the end of the CPMG RD experiment corresponds to $\hat{M}^+ = \hat{M}_x + i\hat{M}_y$ and $\hat{M}^- = \hat{M}_x - i\hat{M}_y$. For two species, A and B, the Bloch equation that describes the free precession of transverse magnetization is:

$$\frac{d}{dt} \begin{bmatrix} \hat{M}_A^+ \\ \hat{M}_B^+ \end{bmatrix} = \begin{bmatrix} -R_{2,A} + i\omega_A & 0 \\ 0 & -R_{2,B} + i\omega_B \end{bmatrix} \begin{bmatrix} \hat{M}_A^+ \\ \hat{M}_B^+ \end{bmatrix} \quad (94)$$

where ω_A and ω_B are the chemical shifts of states A and B, respectively, and \hat{M}_A^+ and \hat{M}_B^+ are the transverse magnetization components, as described above, from states A and B, respectively. Calculation of a spectrum requires solving the equation with a matrix exponential:

$$\hat{M}^+(t) = \hat{M}_A^+(0)e^{-(R_{2,A}+i\omega_A)t} + \hat{M}_B^+(0)e^{-(R_{2,B}+i\omega_B)t} \quad (95)$$

While iterating over the total acquisition time, $t = np \times dt$, where np is the number of complex points and dt is the dwell time or time between points, the above solution can be determined at each iteration and stored in a data array. Following a Fourier transform, the resultant spectrum will contain resonances centered at ω_A and ω_B with linewidths related to $R_{2,A}$ and $R_{2,B}$ and intensities governed by the initial states, $\hat{M}_A^+(0)$ and $\hat{M}_B^+(0)$. To incorporate chemical exchange between the states A and B, the Bloch equations are modified to the Bloch-McConnell equation:

$$\frac{d}{dt} \begin{bmatrix} \widehat{M}_A^+ \\ \widehat{M}_B^+ \end{bmatrix} = \begin{bmatrix} -R_{2,A} - k_{AB} + i\omega_A & k_{BA} \\ k_{AB} & -R_{2,B} - k_{BA} + i\omega_B \end{bmatrix} \begin{bmatrix} \widehat{M}_A^+ \\ \widehat{M}_B^+ \end{bmatrix} \quad (96)$$

If the initial state of transverse magnetization is a matrix termed $\mathbf{M}(0)$ and the first matrix on the right-hand side is termed $i\omega - \mathbf{R} + \mathbf{k}$, then the solution to the Bloch-McConnell equations is obtained by:

$$\mathbf{M}(t) = e^{(i\omega - \mathbf{R} + \mathbf{k})t} \mathbf{M} \quad (97)$$

While the solutions are typically obtained with a computer, they can be explicitly determined in this case (Palmer et al. 2001):

$$\begin{bmatrix} \widehat{M}_A^+ \\ \widehat{M}_B^+ \end{bmatrix} = \begin{bmatrix} a_{11}(t) & a_{12}(t) \\ a_{21}(t) & a_{22}(t) \end{bmatrix} \begin{bmatrix} \widehat{M}_A^+(0) \\ \widehat{M}_B^+(0) \end{bmatrix} \quad (98)$$

in which:

$$a_{11}(t) = \frac{1}{2} \left[\left(1 - \frac{i\Delta\omega - \Delta R_2 + k_{AB} - k_{BA}}{(\lambda_+ - \lambda_-)} \right) e^{-t\lambda_-} + \left(1 + \frac{i\Delta\omega - \Delta R_2 + k_{AB} - k_{BA}}{(\lambda_+ - \lambda_-)} \right) e^{-t\lambda_+} \right]$$

$$a_{12}(t) = \frac{k_{BA}}{(\lambda_+ - \lambda_-)} (e^{-t\lambda_-} - e^{-t\lambda_+})$$

$$a_{21}(t) = \frac{k_{AB}}{(\lambda_+ - \lambda_-)} (e^{-t\lambda_-} - e^{-t\lambda_+})$$

$$a_{22}(t) = \frac{1}{2} \left[\left(1 + \frac{i\Delta\omega - \Delta R_2 + k_{AB} - k_{BA}}{(\lambda_+ - \lambda_-)} \right) e^{-t\lambda_-} + \left(1 - \frac{i\Delta\omega - \Delta R_2 + k_{AB} - k_{BA}}{(\lambda_+ - \lambda_-)} \right) e^{-t\lambda_+} \right]$$

$$\lambda_{\pm} = \frac{1}{2} \left[(-i\omega_A - i\omega_B + R_{2,A} + R_{2,B} + k_{ex}) \pm \sqrt{[i\Delta\omega - \Delta R_2 + k_{AB} - k_{BA}]^2 + 4k_{AB}k_{BA}} \right]$$

As above, the spectrum is obtained by iterating over time and solving the Bloch-McConnell equations for each time step. Saving the output in a data array and taking the Fourier transformation will yield a spectrum; however, the intensities for the chemical shifts of state A and state B, ω_A and ω_B , are impacted by exchange and no longer reflect the exact populations of either state. This is demonstrated in Fig. 10, where two-site exchange is simulated for $k_{ex} = 0 \text{ s}^{-1}$ as a function of population ($p_A = 1 - p_B$), and the intensities of each peak reflect the inputted population

values. In contrast, when $k_{\text{ex}} = 1000 \text{ s}^{-1}$, and thus the system is in intermediate exchange ($\frac{k_{\text{ex}}}{|\Delta\omega|} = 0.32$), the intensity of the major (ground) state resonance is severely attenuated under conditions when $p_{\text{B}} > 0\%$. In addition, both the linewidths and the chemical shifts contain contributions from exchange that lead to additional line broadening (R_{ex}) and modulation of the chemical shift (δ_{ex}) (Vallurupalli et al. 2011):

$$R_{\text{ex}} = k_{AB} \frac{\rho(1 + \rho) + \xi^2}{(1 + \rho)^2 + \xi^2} \quad (99a) \quad \delta_{\text{ex}} = k_{AB} \frac{\xi}{(1 + \rho)^2 + \xi^2} \quad (99b)$$

where $\xi = \frac{\Delta\omega}{k_{\text{ex}}}$ and $\rho = \frac{\Delta R_2}{k_{\text{ex}}}$ in the equations above. As can be seen in Fig. 10, δ_{ex} varies from a few hundredths of a ppm toward the population-weighted average, $p_{A\omega_A} + p_{B\omega_B}$, which occurs when in the fast exchange limit ($k_{\text{ex}} \gg |\Delta\omega|$). The δ_{ex} value can be used as a fitting parameter to extend the timescale in which CPMG RD is sensitive (Vallurupalli et al. 2011). For example, fitting δ_{ex} values alongside $R_{2,\text{eff}}$ (and also the difference in ground state ω values obtained from HMQC and HSQC experiments) extended the timescale of accuracy from ca. 2000 s^{-1} to near 6000 s^{-1} , and also enabled extraction of accurate $|\Delta\omega|$ values (Vallurupalli et al. 2011).

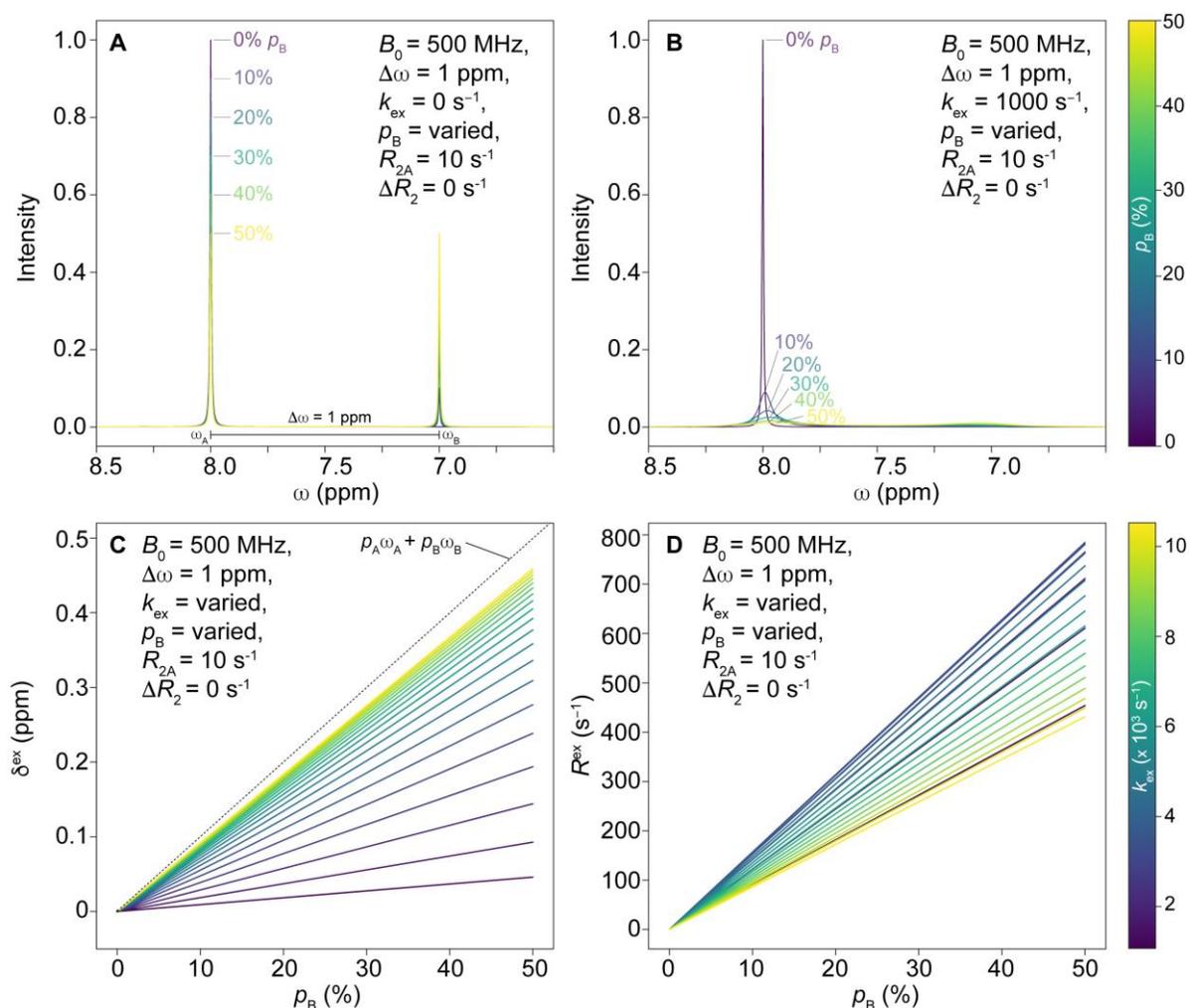


Figure 10. 1D NMR spectra simulated as a function of the population of a second state. (A) 1D NMR spectra simulated in the absence of exchange ($k_{\text{ex}} = 0 \text{ s}^{-1}$) as a function of the second state (colors). The parameters are indicated. **(B)** The same simulation as A except for k_{ex} was set to 1000 s^{-1} . **(C)** The exchange-induced shift plotted as a function of k_{ex} (colors) and population of the minor state (x-axis). **(D)** The same as C, except for the exchange-induced broadening.

Practical aspects of CPMG RD data collection

Below, some practical considerations about CPMG RD experiments on ^{15}N -labeled proteins will be discussed. Many excellent chapters have been written on the topic already, and the reader is directed there for more information ([Gopalan et al. 2018](#), [Ishima 2014](#), [Sauerwein & Hansen 2015](#)).

Assuming there is evidence of conformational exchange on the ms timescale that warrants further investigation, CPMG RD experiments can often be performed at early stages in a protein

NMR project: resonance assignments are helpful for structural interpretation of the data, but not required. Thus, CPMG RD data collection can proceed once conditions have been established in which the protein remains soluble or stable inside the NMR spectrometer at a relatively high concentration for a reasonable period of time. A reasonable experimental time to obtain 20 x 2D spectra as a function of ν is *ca.* 10 hours (4 scans, inter-scan delay of 3 s, maximum $^1\text{H}/^{15}\text{N}$ acquisition times of 64/19 ms, T_{CPMG} of 40 ms, 35* points in ^{15}N). Lower signal-to-noise samples will of course require more experimental time.

For data analysis purposes (*vide infra*), generally at least two CPMG RD datasets are collected at different static magnetic field strengths (Kovrigin et al. 2006), and thus the sample must remain in as close to identical conditions for the duration of the experiments. Even subtle changes to the population of the minor state or the interconversion rate can yield significant effects on the measured dispersions. For example, if the system is a non-covalent dimer (major) in exchange with a monomer (minor state), sample precipitation that occurred between CPMG RD datasets collected at the first and second B_0 fields will lower the effective total protein concentration, and thereby increase the population of the monomeric form. Moreover, the interconversion rate for a bimolecular exchange reaction (here, monomer-dimer) depends on the concentration of the monomer (Palmer et al. 2001), and thus the k_{ex} of the system will also change upon sample precipitation. Finally, the temperature of the sample comprises an important variable, as both k_{ex} (the interconversion rate) and p_{B} (the population of the minor state) can be significantly impacted by changes in temperature. The temperature inside an NMR spectrometer is controlled via the computer, but can be manually calibrated using a standard (e.g. d_4 -methanol). Thus, caution at multiple levels is required to ensure high levels of sample integrity.

For a CPMG RD dataset of a ^{15}N -labeled protein, typically 20 x 2D-planes are acquired in which a variable number of spin echoes are applied in a constant-time relaxation delay (T_{CPMG}). For small-to-medium proteins, T_{CPMG} ranges from about 30 to 100 ms, depending on the signal attenuation due to transverse relaxation in the presence and absence of T_{CPMG} . The 180° pulses

during the spin echoes are applied as a function of varying τ time, in which the more commonly reported value of τ is ν , equivalent to $\frac{1}{2\delta}$ where $\delta = 2\tau + 2p_N$, and $2p_N$ refers to the duration of the 180° pulse (Hansen et al. 2008a). Generally, excluding recently developed high-power CPMG RD experiments (Reddy et al. 2018), ν values range from 50 to 1000 Hz for ^{15}N CPMG RD experiments, meaning the time between successive 180° refocusing pulses ranges from 10 ms ($\nu = 50$ Hz) to 0.5 ms ($\nu = 1000$ Hz).

Evidence of conformational exchange on the millisecond timescale in CPMG RD experiments manifests through an increase in the intensity of a given resonance as a function of the number of spin echoes, or an increasing ν value. Typically, a residue is considered to be undergoing exchange on the millisecond timescale if $R_{2,\text{eff}}(\nu=\infty) - R_{2,\text{eff}}(\nu=0 \text{ Hz}) > 2 \text{ s}^{-1}$, or some other cut-off value. A more rigorous test involves a statistical comparison (F -test) of a fit to the Bloch-McConnell equation, with versus without a non-zero k_{ex} parameter. In order to generate the effective transverse relaxation rate constant ($R_{2,\text{eff}}$) values from the raw data, a reference 2D plane is recorded in the absence of T_{CPMG} and used to calculate:

$$R_{2,\text{eff}} = \frac{-1}{T_{\text{CPMG}}} \ln\left(\frac{I}{I_0}\right) \quad (100)$$

where I_0 is the transverse magnetization in the absence of the constant time period and I is the transverse magnetization present after the CPMG pulse train. Thus, higher values of $R_{2,\text{eff}}$ indicate faster transverse relaxation due to conformational exchange, which manifests as lower peak intensities. As the number of spin echoes is increased, the effect of conformational exchange on the $R_{2,\text{eff}}$ value is diminished and the intensity of an exchanging peak increases. The precise dependence of $R_{2,\text{eff}}$ on ν varies, to differing extents, with B_0 , p_B , k_{ex} , R_{2A} , ΔR_2 , and $|\Delta\omega|$, and therefore CPMG RD data provide detailed information about the structures ($|\Delta\omega|$), kinetics ($k_{AB} = p_B k_{\text{ex}}$; $k_{BA} = p_A k_{\text{ex}}$), and thermodynamics (p_B) of the exchanging system.

In the author's experience, predicting how various parameters impact CPMG RD can be sometimes unituitive, e.g. it is difficult to know *a priori* how the off-rate will impact the observed dispersion in a monomer-dimer equilibrium. Thus, in order to visualize how various parameters and chemical exchange schemes impact CPMG RD data (Fig. 11), a Python software program termed Chemical Exchange Simulator And Regression Instrument for Cpmg Analysis (*CESARICA*) was written by the author. The acronym was inspired by the life and music of the late Oliver Dragojević. The software is capable of simulating and fitting CPMG RD data, analyzing the χ^2 error surface, and more. All of the figures henceforth demonstrate some of the features of *CESARICA*. A publication that outlines its capability in full will be detailed in full elsewhere.

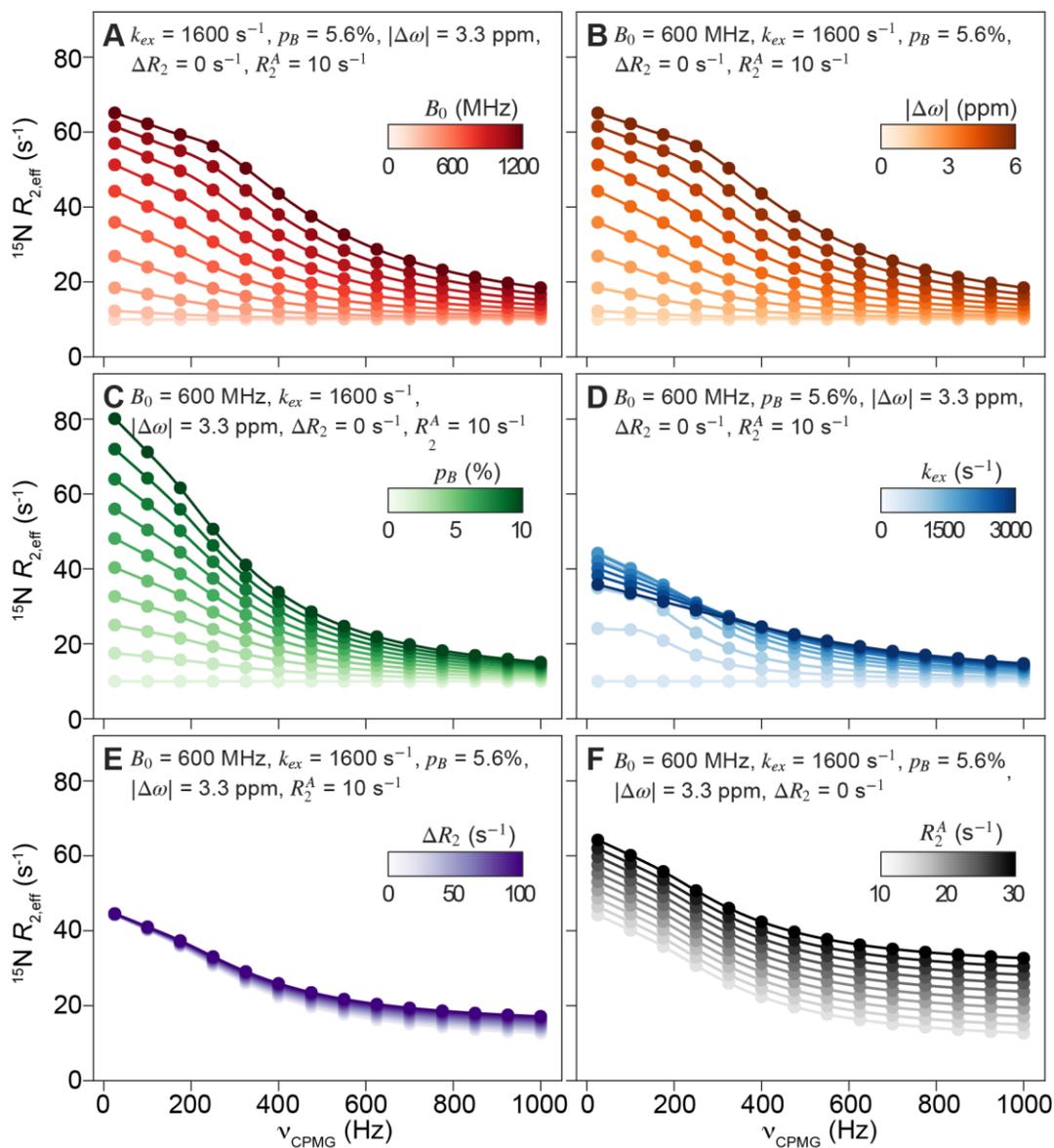


Figure 11. Example CPMG relaxation dispersion curves. The dispersions shown here are computed as a function of (A) the static magnetic field strength, (B) the difference in chemical shifts, (C) the population of the minor state, (D) the interconversion rate, (E) the difference in intrinsic transverse relaxation rates, and (F) the intrinsic relaxation rate. In all cases, the non-variable parameters are identical between panels (e.g. $|\Delta\omega| = 3.3$ ppm for panels A and C–F) to enable ease of comparison. Note that increasing the B_0 field (A) for a fixed k_{ex}/p_B combination has the same impact as increasing $|\Delta\omega|$ (B).

The interconversion rate, k_{ex} , and thus CPMG dispersions depend on the reaction order (Palmer et al. 2001). The following dependence of k_{ex} on the reaction rates for a model of two-state exchange apply, where state A is a monomer that is in exchange with a form A_n , where n describes the number of units, *i.e.* $n=2$ for a dimer, $n=4$ for a tetramer, etc:

$$k_{ex} = k'_{AB} + k_{BA} \quad (101)$$

$$k'_{AB} = n k_{AB}[A]^{n-1} \quad (102)$$

In the case of a simple protein-ligand interaction ($A + L$ yielding AL), $k'_{AB} = k_{AB}[L]$. In the figure below, a bimolecular exchange reaction is considered in which the major (ground or observable) state corresponds to a dimer (A_2) and the minor (excited or invisible) state is a monomer ($2A$). Under such conditions:

$$k_{ex} = 2k_{AB}[A] + k_{BA} \quad (103)$$

and therefore the dispersion will depend on the total monomer concentration ($[A]$, in units of M), the association- or on-rate (k_{on} , in units of $M^{-1} s^{-1}$) and the dissociation- or off-rate (k_{off} , in units of s^{-1}). The latter two values are related to the dissociation constant (K_d) by:

$$K_d = \frac{k_{off}}{k_{on}} \quad (104)$$

Moreover, for a non-covalent dimeric protein the population of the free monomer will depend on the K_d and the total monomer concentration. For a dimer with a K_d in the low micromolar range that is present at NMR quantities (*i.e.* hundreds of micromolar or near millimolar), the population of the free monomer will be rather low, *e.g.* for a $1 \mu M$ K_d and a total protein (monomer) concentration of 1 mM , there would exist 2.21% free monomer or $22.1 \mu M$. If the total protein concentration was lowered to 0.1 mM , there would exist 6.83% monomer or $68.1 \mu M$. Since the population of the free monomer comprises p_B in this scenario, it has a significant impact on the overall profile of the dispersion. Given the K_d and the total protein (monomer) concentration (M_T), the amount of free monomer (p_B) can be calculated as follows:

$$p_B = \frac{-K_d + \sqrt{K_d^2 + 8M_T K_d}}{4M_T} \quad (105)$$

The figure below demonstrates how, for a non-covalent dimer, CPMG RD data can vary widely as a function of the total protein concentration (Fig. 12A), the K_d (Fig. 12B), the dissociation rate (Fig. 12C), or the association rate (Fig. 12D). For instance, when the K_d is $1 \mu M$ and the association and dissociation rates are $10^8 M^{-1} s^{-1}$ and $100 s^{-1}$, respectively, an experimentally

realistic situation that may exist when the dimer is formed through electrostatically-mediated interactions, the R_{ex} values of the dispersions are largest when the total monomer concentration (M_T) is relatively low (e.g. 100 μM) and approach 50 s^{-1} . This is of course because the population of the monomeric state is increased as M_T is decreased. As can be seen in panel A, as M_T is increased to 1 mM and beyond, the R_{ex} values become progressively smaller and smaller, eventually becoming flat above 10 mM. Likewise, for a fixed M_T of 0.3 mM (Fig. 12B) with variable K_d values ranging from 10 nM to 100 μM (10^{-8} to 10^{-4} M), the R_{ex} values on the dispersions get larger as the K_d approaches M_T . The $R_{2,\text{eff}}$ values in panel B above ca. 60-70 s^{-1} would likely be unmeasurable due to severe exchange broadening, but they are included for illustrative purposes. Similar observations are made as the k_{off} and k_{on} values are varied for a fixed K_d . At low concentrations of free monomer, [A], k_{ex} depends to a large extent on k_{on} – e.g. for $K_d = 1 \mu\text{M}$ and $M_T = 1 \text{ mM}$, where $p_B = 2.21\%$ and therefore $[A] = 22.1 \mu\text{M}$, typical values of $2k_{\text{on}}[A]$ could be 0.442 s^{-1} for $k_{\text{on}} = 1 \times 10^4 \text{ M}^{-1} \text{ s}^{-1}$, 4.42 s^{-1} for $k_{\text{on}} = 1 \times 10^5 \text{ M}^{-1} \text{ s}^{-1}$, or 44.2 s^{-1} for $k_{\text{on}} = 1 \times 10^6 \text{ M}^{-1} \text{ s}^{-1}$. Under such scenarios, k_{off} values contribute negligibly to k_{ex} , ranging from 0.01 s^{-1} to 1 s^{-1} as k_{on} increases from $1 \times 10^4 \text{ M}^{-1} \text{ s}^{-1}$ to $1 \times 10^6 \text{ M}^{-1} \text{ s}^{-1}$. Only when k_{off} reaches 100 s^{-1} or higher does it appreciably contribute to k_{ex} , and therefore the observed dispersion (Fig. 12C). Similarly, for a K_d of $1 \mu\text{M}$ and M_T of 0.3 mM, k_{on} must exceed $1 \times 10^6 \text{ M}^{-1} \text{ s}^{-1}$ before significant dispersions are observed. The full variation of k_{ex} for an exchanging dimeric system as a function of M_T , K_d , k_{off} , and k_{on} can be found in Fig. 12E and Fig. 12F.

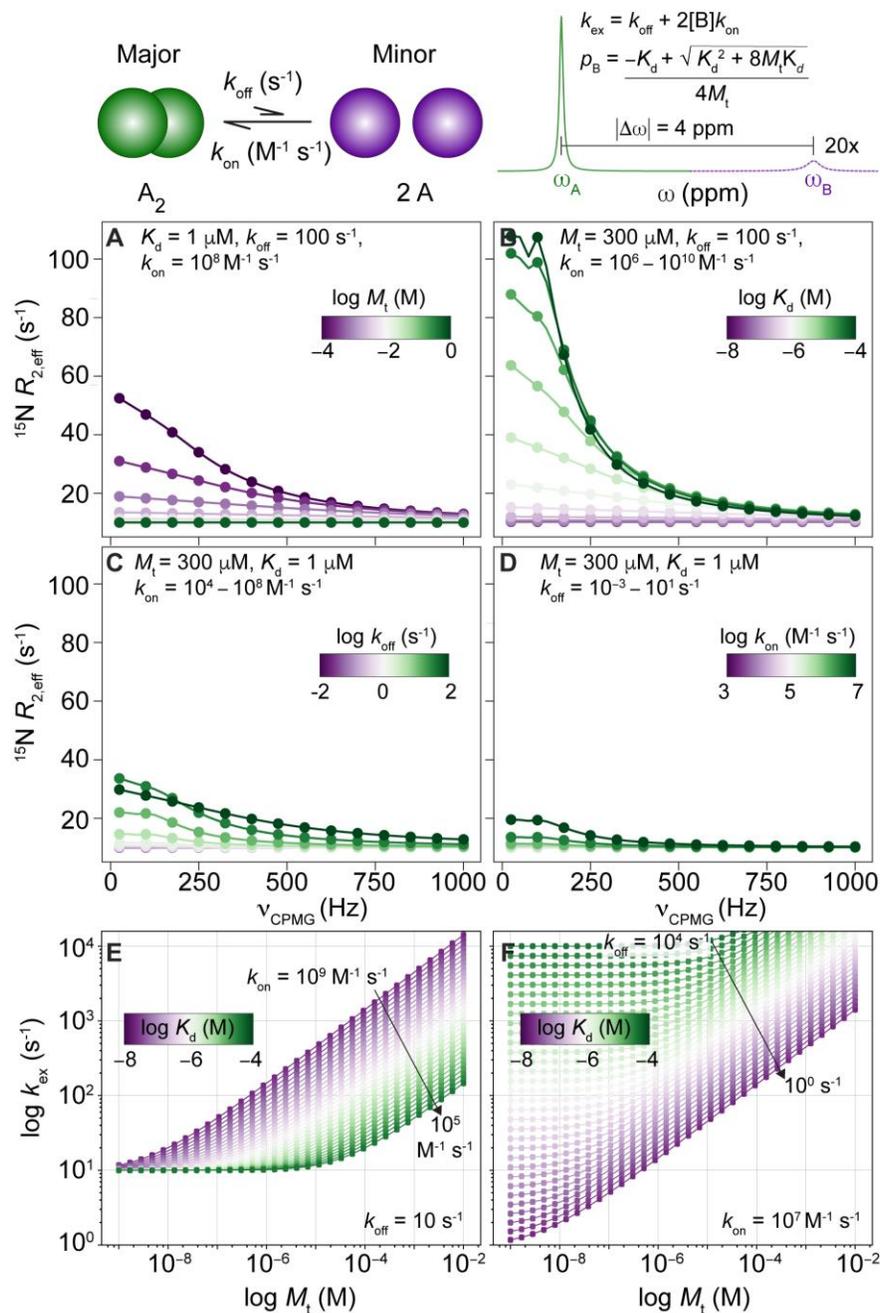


Figure 12. Monomer-dimer association and dissociation rates significantly impact CPMG RD measurements. Example CPMG relaxation dispersion curves for a dimer-monomer equilibrium computed as a function of (A) the total monomer concentration, (B) the dimer dissociation constant, and (C, D) the dissociation and association rates. A schematic of the dimer-monomer equilibrium is depicted above panels A and B. Dissociation of the dimer is first-order (s^{-1}), whereas association of monomers is second-order ($\text{M}^{-1} \text{s}^{-1}$). The expressions for k_{ex} and ρ_B can be calculated based on the known rates and ρ_B depending on the dissociation constant and total monomer concentration. An example 1D NMR spectrum for a system undergoing two-site chemical exchange is shown with the following parameters: $k_{ex} = 500 \text{ s}^{-1}$, $\rho_B = 1\%$ (magnified by 20x for illustration), $R_2^A = 10 \text{ s}^{-1}$, $\Delta R_2 = 0 \text{ s}^{-1}$, $B_0 = 500 \text{ MHz}$. All parameters besides k_{ex} and ρ_B are the same in panels A-D. (E) and (F) depict the impact of M_t , K_d , k_{off} , and k_{on} on the observed k_{ex} . In (E) k_{off} is fixed and in (F) k_{on} is fixed.

Likewise, for a protein-ligand reaction in which the major state is the free (^{15}N -labeled) protein, P, and the minor state is the bound protein-ligand complex (PL), CPMG RD can also be simulated as a function of the K_d and various reaction rates (Fig. 13). In this situation, k_{ex} and p_B are described by (note that equation 105 also applies):

$$k_{ex} = k_{on}[L] + k_{off} \quad (106)$$

$$p_B = \frac{[L]}{[L] + K_d} \quad (107)$$

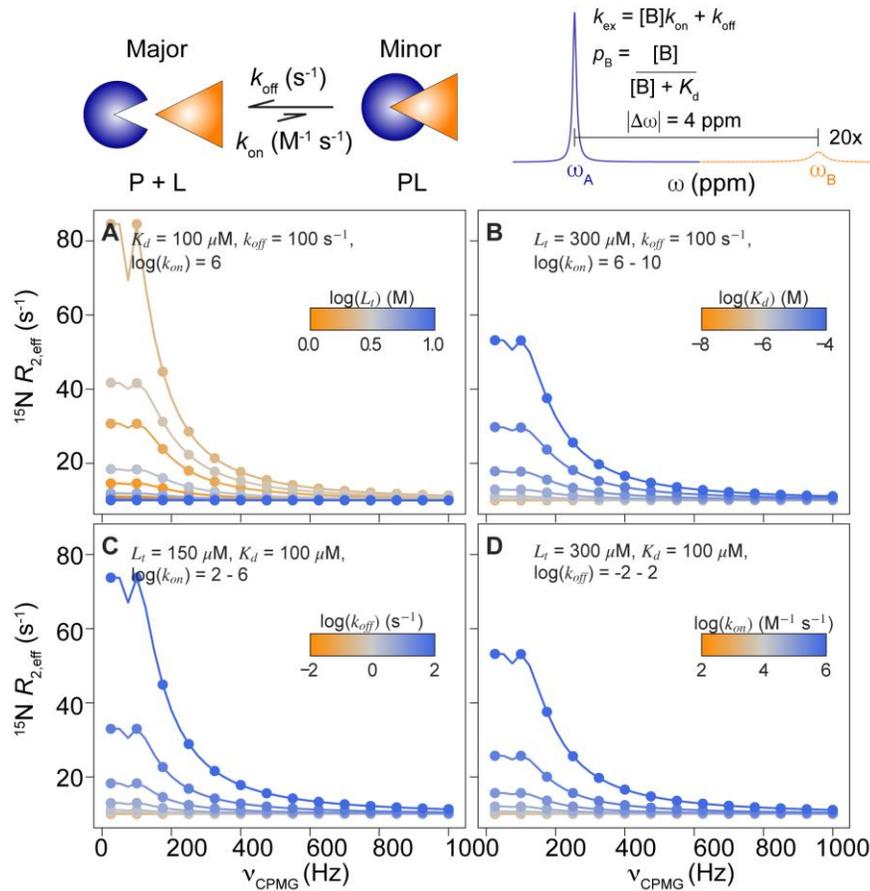


Figure 13. Protein-ligand binding kinetics significantly impact CPMG RD measurements. Example CPMG relaxation dispersion curves for a protein-ligand binding event computed as a function of (A) the total ligand concentration, (B) the protein-ligand dissociation constant, and (C, D) the dissociation and association rates. A schematic of the protein-ligand equilibrium is depicted above panels A and B. Dissociation of the ligand is first-order (s⁻¹), whereas association is second-order (M⁻¹ s⁻¹). As above, k_{ex} and p_B can be computed explicitly. An example 1D NMR spectrum for a system undergoing two-site chemical exchange is shown with the following parameters: $k_{ex} = 500 \text{ s}^{-1}$, $p_B = 1\%$ (magnified by 20x for illustration), $R_2^A = 10 \text{ s}^{-1}$, $\Delta R_2 = 0 \text{ s}^{-1}$, $B_0 = 500 \text{ MHz}$. All parameters besides k_{ex} and p_B are the same in panels A-D.

Fitting CPMG RD data

In order to fit the resultant $R_{2,\text{eff}}$ vs. ν curves, the Bloch-McConnell equations can be used as a target function. However, rather than computing the entire spectrum as above, simply the evolution of transverse magnetization during the constant time CPMG period can be calculated using the 2-by-2 Bloch-McConnell equations as follows (Libich et al. 2015):

$$M(t) = (AA^*A^*A)^n M(0) \quad (108)$$

$$A = e^{-\begin{bmatrix} -R_{2,A} - p_B k_{ex} & (1-p_B)k_{ex} \\ p_B k_{ex} & -R_{2,B} - i\Delta\omega - p_A k_{ex} \end{bmatrix} \tau} \quad A^* = e^{-\begin{bmatrix} -R_{2,A} - p_B k_{ex} & (1-p_B)k_{ex} \\ p_B k_{ex} & -R_{2,B} + i\Delta\omega - p_A k_{ex} \end{bmatrix} \tau} \quad (109)$$

where $M(t)$ is the transverse magnetization, A is the propagator, n is the number of cycles in the CPMG pulse train (constant time duration divided by 2τ), and everything else as described above. The $R_{2,\text{eff}}$ value is then calculated as above, with $I_0 \propto$ the initial transverse magnetization (*i.e.* p_B), and so I reflects the magnetization after a given number of spin echoes. This route enables relatively rapid calculation of $R_{2,\text{eff}}$ values for a given set of parameters, and can be implemented with modern, non-linear optimization methods. Minimization of a target function can be implemented by minimizing the chi-square (χ^2) value:

$$\chi^2 = \sum_{i=1}^N \left(\frac{F^i_{\text{calc}} - F^i_{\text{exp}}}{\sigma^i_{\text{exp}}} \right)^2 \quad (110)$$

$$\chi^2_{\text{red}} = \frac{\chi^2}{\text{DF}} \quad (111)$$

where F^i_{calc} and F^i_{exp} are the calculated and experimental data points for data point i , and σ^i_{exp} represents the experimental uncertainty of data point i . The sum is over all data points. The reduced χ^2 (χ^2_{red}) is simply the χ^2 divided by the degrees of freedom (DF).

For a CPMG RD dataset acquired at a single static magnetic field strength, the fitting parameters include R_{2A} , R_{2B} , k_{ex} , p_B , and $\Delta\omega$. As mentioned above, CPMG RD experiments are only sensitive to the absolute value of the difference in chemical shifts, $|\Delta\omega|$, and so the sign of $\Delta\omega$ is not obtained from fitting. Instead, the sign can be determined through recording HSQC spectra

at multiple static magnetic field strengths and monitoring the exchange-induced shift, or, if a chemical shift difference between major and minor states also exists for the ^1H nucleus, by recording HSQC and HMQC spectra and monitoring the difference in position of the major state resonance (Skrynnikov et al. 2002).

Typically, unless otherwise known from inspection or by statistical analysis, residues that show evidence of μs - ms motions (*i.e.* dispersions, whereby $R_{2,\text{eff}}(\nu=\infty) - R_{2,\text{eff}}(\nu=0 \text{ Hz}) > 2 \text{ s}^{-1}$) are assumed to undergo global exchange, *i.e.* the k_{ex} and ρ_{B} values are global fitting parameters. In addition, CPMG RD data are generally recorded at multiple, static magnetic field strengths, so as to introduce a magnetic field-dependence to the $|\Delta\omega|$ parameter. Therefore, for CPMG RD data recorded at B static magnetic field strengths on a protein with M dispersions, there will be $k_{\text{ex}} + \rho_{\text{B}} + M$ values of $|\Delta\omega|$ + MB values of $R_{2\text{A}}$ + MB values of $R_{2\text{B}} = 2 + M + 2BM$ fitting parameters. It has previously been shown that the difference in transverse relaxation rate constants, $\Delta R_2 = R_{2\text{A}} - R_{2\text{B}}$, if reasonably small (*i.e.* same order of magnitude), does not significantly impact the fitted results, and so it can be ignored in some cases, reducing the number of fitted parameters to $2 + M + BM$. Thus, for a protein with 20 dispersions and data recorded at two static magnetic field strengths, there will be 102 or 82 fitted parameters, depending on if $R_{2\text{B}}$ values are fitted. Even for $2 + M + 2BM$ fitted parameters, this value is far lower than the number of data points, which for n values of ν reaches BMn , equaling 800 in this example ($\nu = n = 20$). Therefore, for data collected at B static magnetic fields for M resonances over n values of ν :

$$\chi^2 = \sum_{k=1}^n \sum_{j=1}^B \sum_{i=1}^M \left(\frac{R_{2,\text{eff calc}}^{i,j,k} - R_{2,\text{eff}}^{i,j,k}}{\sigma_{i,j,k}^{R_{2,\text{eff}}}} \right)^2 \quad (112)$$

and therefore, the difference between the calculated $R_{2,\text{eff}}$ value and the experimentally measured value is summed over each ν value ($k = 1, 2, \dots, n$) at each B_0 field ($j = 1, 2, \dots, B$) for each residue number ($i = 1, 2, \dots, M$), *i.e.* resonance. A fit to both the Bloch-McConnell equation and the Carver-Richards equation is shown below for a single residue undergoing two-state exchange (Fig. 14).

CPMG RD data were generated at three static magnetic field strengths using the Bloch-McConnell equations, with a small amount of noise added afterward to introduce deviations from ideality.

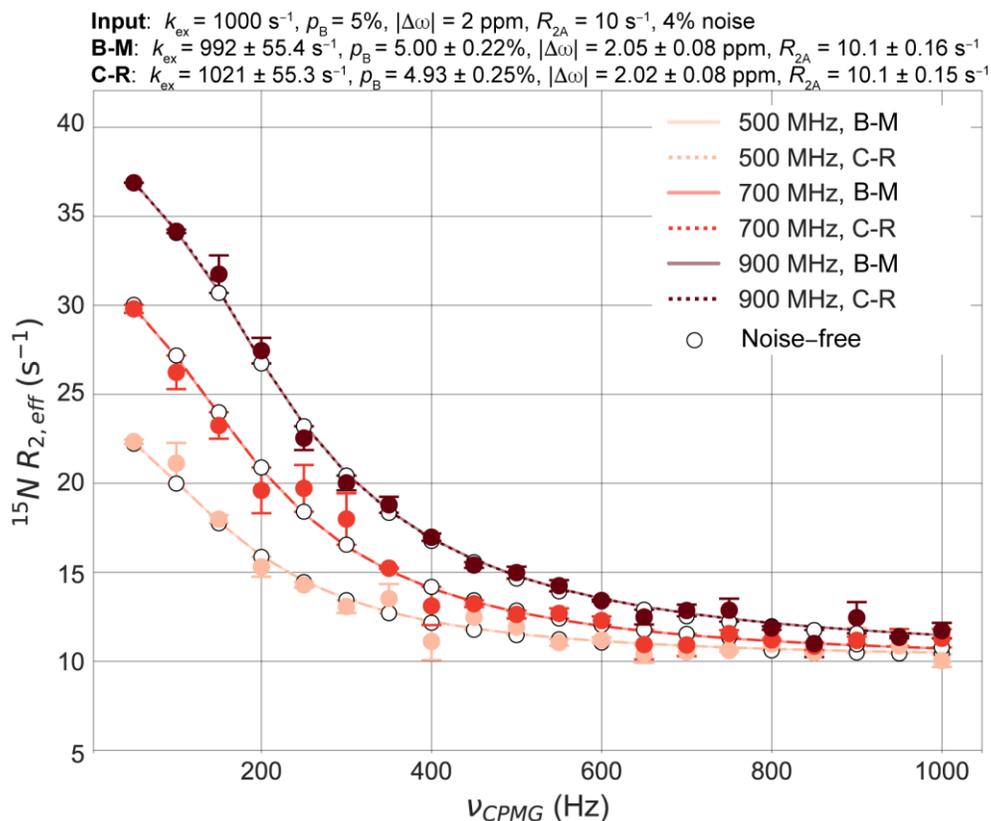


Figure 14 Fitting synthetic CPMG RD data. The solid circles represent $R_{2,\text{eff}}$ values generated from numerically evaluating the Bloch-McConnell equations (B-M) for the input parameters indicated. To introduce noise into the data points, a normal distribution was generated and centered at zero with a width of the 4% of data point of interest. A value was then randomly selected from this distribution and added to the synthetic data point. The error bars thus reflect the deviation from the ideal $R_{2,\text{eff}}$ values generated in the absence of noise. $R_{2,\text{eff}}$ values were generated at three static magnetic field strengths (500, 700, and 900 MHz) and the data were fit globally (*i.e.* a shared k_{ex} and p_{B}) with local fitted values of $|\Delta\omega|$ and R_{2A} . Data were either fit to the B-M equation (solid line) or the Carver-Richards equation (dotted line, C-R). In both cases, a good fit is obtained that replicates the inputted values. The χ^2_{red} values are 1.076 and 1.079 for the B-M and C-R equations, respectively.

Comparison of the Carver-Richards and Bloch-McConnell equations

Intriguingly, although the Carver-Richards equation is capable of yielding seemingly “good” fits to CPMG RD data (Fig. 14) with similar χ^2 values as the Bloch-McConnell equation, a closer inspection indicates that $R_{2,\text{eff}}$ values generated from Carver-Richards equation actually deviate

from those synthesized with the Bloch-McConnell equation. This becomes especially evident at low levels of added noise (Fig. 15), and therefore the addition of higher levels of noise slightly obviates this effect. A previous publication on this topic has explored the deviation in more mathematical detail (Baldwin 2014), and derived a new analytical expression. Here, the deviation between Carver-Richards and the Bloch-McConnell equation is examined across a wider range of parameter space and also in the context of fitting. Since the Carver-Richards equation is simply an analytical approximation of the Bloch-McConnell equation, all synthetic CPMG RD were therefore generated by numerically evolving the Bloch-McConnell equation, as described above.

For a given set of input parameters at two static magnetic field strengths, the resultant $R_{2,\text{eff}}$ values as determined by numerically evolving the Bloch-McConnell equations are shown in Fig. 15A as colored circles. For the same set of input parameters, the $R_{2,\text{eff}}$ values as calculated by the Carver-Richards equation are shown as empty circles. The subtle, but significant deviation at low ν values manifests as errors in the resultant fit to the Bloch-McConnell equation (Fig. 15B). By iteratively fitting the data at 500 and 800 MHz in the presence of a very small amount of random noise (normal distribution centered at 0 with a width of 0.1% of $R_{2,\text{eff}}$) that changes in each iteration and plotting the outputted k_{ex} and ρ_{B} values, it becomes clear that fitting to the Carver-Richards equation introduces errors to k_{ex} and ρ_{B} that significantly exceed the noise level. The input k_{ex} and ρ_{B} were 1000 s^{-1} and 2.5%, respectively, but the mean values obtained from fitting the data to the Carver-Richards equation were respectively 1028 s^{-1} and 2.45% (Fig. 15B). These deviations from the inputted values represent errors of 2.8% and 2% for k_{ex} and ρ_{B} , respectively, whereas the noise was inputted at a level of 0.1% of the $R_{2,\text{eff}}$ values. For comparison, iteratively fitting to the Bloch-McConnell equation in the presence of 0.1% noise yields a mean k_{ex} of 1000 s^{-1} and ρ_{B} of 2.5% (Fig. 15B), indicating the overall accuracy and precision.

This scenario represents one of many potential combinations of k_{ex} and ρ_{B} , and I thus varied systematically these parameters to assess the errors introduced by fitting to the Carver-Richards equation. This is of interest as, under certain exchange regimes, the differences between

Bloch-McConnell and Carver-Richards may become exacerbated or minimized. Therefore, a large array of synthetic CPMG RD data were generated over 1024 combinations of k_{ex} and p_{B} at two static magnetic field strengths (500 and 800 MHz) in the presence of 0.1% noise, as above. The grid of k_{ex} and p_{B} comprised a 32 x 32 matrix with k_{ex} values ranging from 300 to 3000 and p_{B} values spanning 0.5 to 10%. The $|\Delta\omega|$ and R_{2A} values were kept the same as before, namely 4 ppm and 10 s^{-1} , respectively. For each combination of k_{ex} and p_{B} , the data were fit 200 times to either the Carver-Richards equation or the Bloch-McConnell equation, and the mean value of k_{ex} , p_{B} , and $|\Delta\omega|$ was saved in an array. As above, the noise, which was randomly added, varied in each iteration of the 200 fits. A contour plot in Fig. 15 of the resultant values:

$$\Delta k_{\text{ex}}(\%) = 100 \frac{|k_{\text{ex}}^{\text{calc}} - k_{\text{ex}}^{\text{fit}}|}{k_{\text{ex}}^{\text{calc}}}$$

$$\Delta p_{\text{B}}(\%) = 100 \frac{|p_{\text{B}}^{\text{calc}} - p_{\text{B}}^{\text{fit}}|}{p_{\text{B}}^{\text{calc}}}$$

$$\Delta |\Delta\omega|(\%) = 100 \frac{|\Delta\omega^{\text{calc}} - \Delta\omega^{\text{fit}}|}{\Delta\omega^{\text{calc}}}$$

thus indicates the percent errors in the fitted values of k_{ex} , p_{B} , and $|\Delta\omega|$ relative to the input values.

The contour plots derived from fitting to the Bloch-McConnell equations (Fig. 15C) indicate overall high levels of accuracy, with maximum errors in the low single digit values, particularly for conditions when $p_{\text{B}} < 1\%$. These errors still exceed the noise threshold, likely due to the small dispersions that are present under such conditions, but pale in comparison to errors from fitting to the Carver-Richards equation. As can be seen in Fig. 15D, fitting to the Carver-Richards equation over this grid of CPMG RD data introduces very large errors, especially under conditions of intermediate/intermediate-slow exchange where errors reach and exceed 15%. The errors introduced into $|\Delta\omega|$ by fitting to the Carver-Richards equation are relatively smaller, but still reach up to 5%. Therefore, it is not recommended to fit CPMG RD data to the Carver-Richards equation,

especially if the system is in the slow/intermediate-slow exchange regime. For low signal-to-noise data, the errors introduced by using the Carver-Richards equation may be masked by the experimental uncertainty; as such, my usage of high signal-to-noise, synthetic data herein provided a convenient way to assess errors due to the choice of the target function.

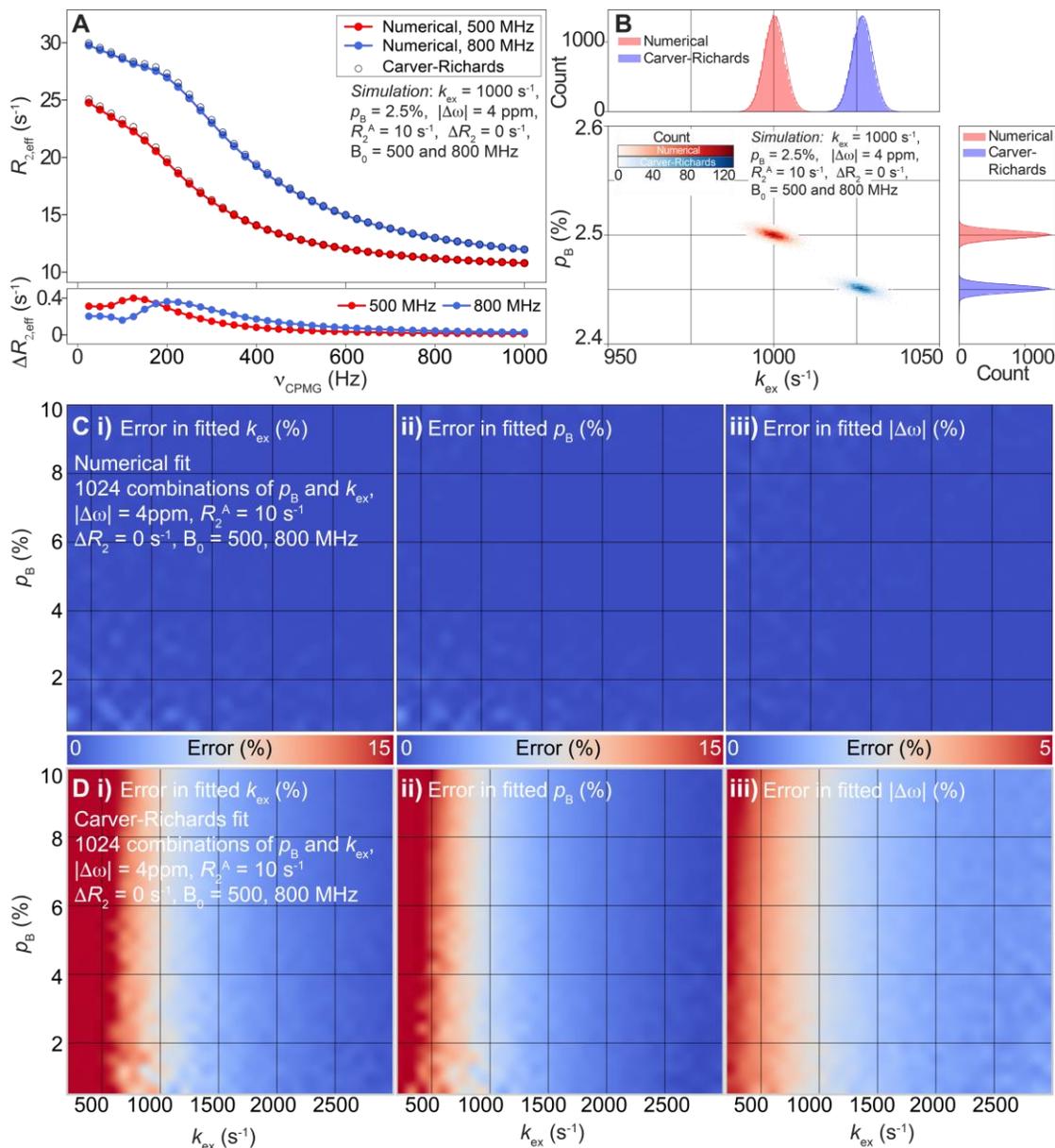


Figure 15. Errors introduced by fitting CPMG RD data to the Carver-Richards equation. (A) Simulated CPMG RD traces at two static magnetic field strengths, with the data generated using either numerical

simulation (filled circles) or the Carver-Richards equation (empty circles). In both cases, the input parameters were identical. The bottom inset depicts the difference in $R_{2,\text{eff}}$ values, i.e. Carver-Richards minus numerical, and a static magnetic field-dependence is evident. **(B)** Synthetic CPMG RD data generated numerically at two static magnetic field strengths were fit to either the Carver-Richards equation (blue) or the 2-by-2 Bloch-McConnell matrix (red). A small amount of normally distributed noise (mean 0.1% of $R_{2,\text{eff}}$) was added to the synthetic data. The resultant 2D histogram depicts the outcome of many fits to the data, with 1D histograms plotted on the k_{ex} and p_{B} axes. As can be seen for the given parameters, fitting to the Carver-Richards equation introduces errors in the fitted k_{ex} and p_{B} values that exceed the input noise. **(C, D)** CPMG RD data with a fixed were simulated numerically at two static magnetic field strengths over a vast range of k_{ex} and p_{B} values (1024 combinations), whereas all other parameters were held constant. The synthetic data were then fit either to the 2x2 Bloch-McConnell matrix (C) or the Carver-Richards equation (D). Errors in the resultant values of k_{ex} , p_{B} , and $|\Delta\omega|$ are respectively shown in panels *i*, *ii*, and *iii*. Fitting to the Carver-Richards equation can introduce significant errors in k_{ex} and p_{B} , in excess of 15%, whereas the error in $|\Delta\omega|$ reaches only 5%.

Fitting CPMG RD data obtained at a single magnetic field strength

Having established that, for high accuracy fits, CPMG RD data should not be fit using the Carver-Richards equation, I next investigated how much data are required to obtain a “good” fit. As mentioned above, CPMG RD data are typically collected at more than one static magnetic field strength, such that a magnetic field-dependence is introduced to the $|\Delta\omega|$ term. Thus, while only a single $|\Delta\omega|$ is actually fit among the multiple datasets, the dependence of $|\Delta\omega|$ on the static magnetic field strength provides an independent parameter.

Previously, the importance of having datasets from multiple static magnetic fields was investigated by Loria and coworkers ([Kovrigin et al. 2006](#)) and Kay and coworkers ([Neudecker et al. 2006](#)). The latter publication noted the essential nature of multiple datasets when fitting to models of chemical exchange involving three states ([Neudecker et al. 2006](#)). The former publication demonstrated that significant correlation between fitted parameters exists when only using data from one magnetic field strength, leading to inaccuracy in the fitted parameters ([Kovrigin et al. 2006](#)). However, Loria and coworkers mainly investigated the impact on a single resonance under a single condition (*i.e.* one exchange regime), and only probed global fitting of three resonances that varied from slow-to-intermediate exchange ([Kovrigin et al. 2006](#)). In a

typical protein NMR study, however, 20 or more resonances that encompass all exchange regimes may be jointly fit to a single k_{ex} and p_{B} combination, thus providing a large amount of data to fit.

My rationale for doing so is thus: depending on the nature of the specific NMR project that utilizes CPMG RD, perhaps only qualitative values of k_{ex} , p_{B} , or $|\Delta\omega|$ are necessary in order to move forward or obtain insightful results. Moreover, for samples that are unstable and do not remain soluble for collection of CPMG RD data at more than one magnetic field strength, perhaps collection of data is only feasible at a single field. Finally, for protein-ligand studies or other semi-high-throughput investigations, collecting CPMG RD data at a single field (and cutting down the number of ν values or using NUS) would diminish experimental acquisition times, and could lead to the screening of more samples in a fixed period of time. Further gains could be acquired from using 1D CPMG RD experiments and qualitatively fitting the data. As such, I sought to investigate in more detail the potential pitfalls and outcomes of fitting CPMG RD data obtained at a single static magnetic field strength.

First, I generated synthetic CPMG RD data for a single resonance using the Bloch-McConnell equations at a single static magnetic field strength (Fig. 16) for k_{ex} of 1000 s^{-1} , p_{B} of 2.5%, ^{15}N $|\Delta\omega|$ of 8 ppm, at B_0 of 500 MHz ($\alpha = 0.27$). Noise was added to these data, as described above, and fit using both the Carver-Richards equation and the Bloch-McConnell equations. This fit was repeated three times while the noise varied between datasets, and indeed as found previously, a large spread of fitted values was obtained. In all cases, the $|\Delta\omega|$ appears relatively well described (Hansen et al. 2008), but the k_{ex} and p_{B} values can fluctuate considerably. Such variation in the fitted parameters originates from the added noise, as can be seen in the bottom panel of Fig. 16.

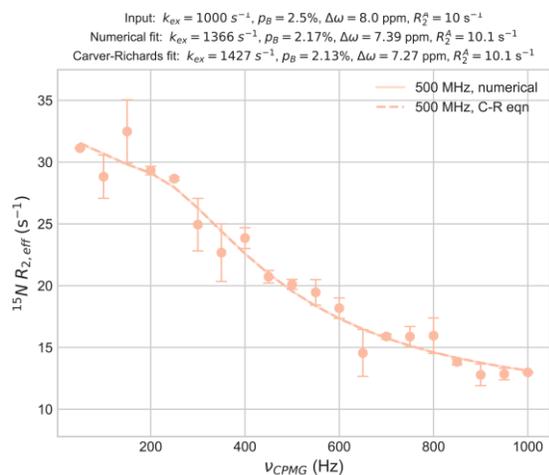
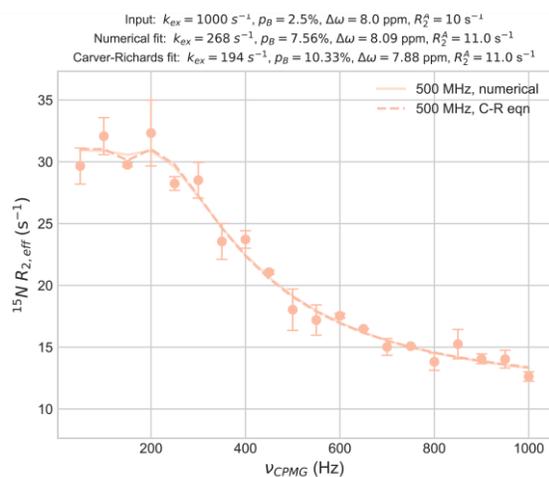
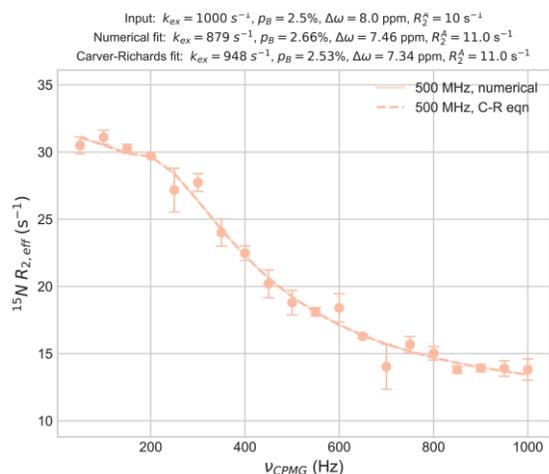


Figure 16. Fitting CPMG RD data at a single static magnetic field strength. Synthetic CPMG RD was generated at 500 MHz for a ^{15}N nucleus undergoing two-state chemical exchange. Random noise (5%) was added to the input data three times, and the data were fit three separate times to either the Bloch-McConnell or Carver-Richards equations. The resultant values of the fits are listed atop each panel. Note the large variation in fitted parameters.



Thus, to assess the variation in the fitted parameters as a function of the added noise, I conducted many fits for a dispersion with k_{ex} of 1000 s^{-1} , p_{B} of 2%, $^{15}\text{N } |\Delta\omega|$ of 2 ppm, at B_0 of 1200 MHz ($\alpha = 0.6$); N.B. that this magnetic field strength currently does not exist, but will be available in the near future. Noise was added to the synthetic data at 1% (Fig. 17A), 2.5% (Fig. 17B), 5% (Fig. 17C), or 10% of $R_{2,\text{eff}}$ (Fig. 17D), and each fit was performed 10,000 times, with the noise randomly varying each iteration. The 2D histograms in Fig. 17 demonstrate how increasing the noise level increases the uncertainty in the fitted parameters, as expected. However, it is interesting to note that, even at noise levels of 5% and 10%, the mean values of k_{ex} and p_{B} remain close to the input parameters. Thus, for CPMG RD data collected at a single field, conducting many fits and taking the mean value appears to yield relatively correct outputs (Fig. 17) For example, at 5% noise, the mean k_{ex} and p_{B} values are 1057 s^{-1} and 5.4%, respectively, which correspond to errors of 5.7% and 8%. At 10% noise, the mean values of k_{ex} and p_{B} do not change considerably, but the width of the distributions have significantly increased. The uncertainty in the fitted parameters increases rapidly when using a single field to fit CPMG RD data.

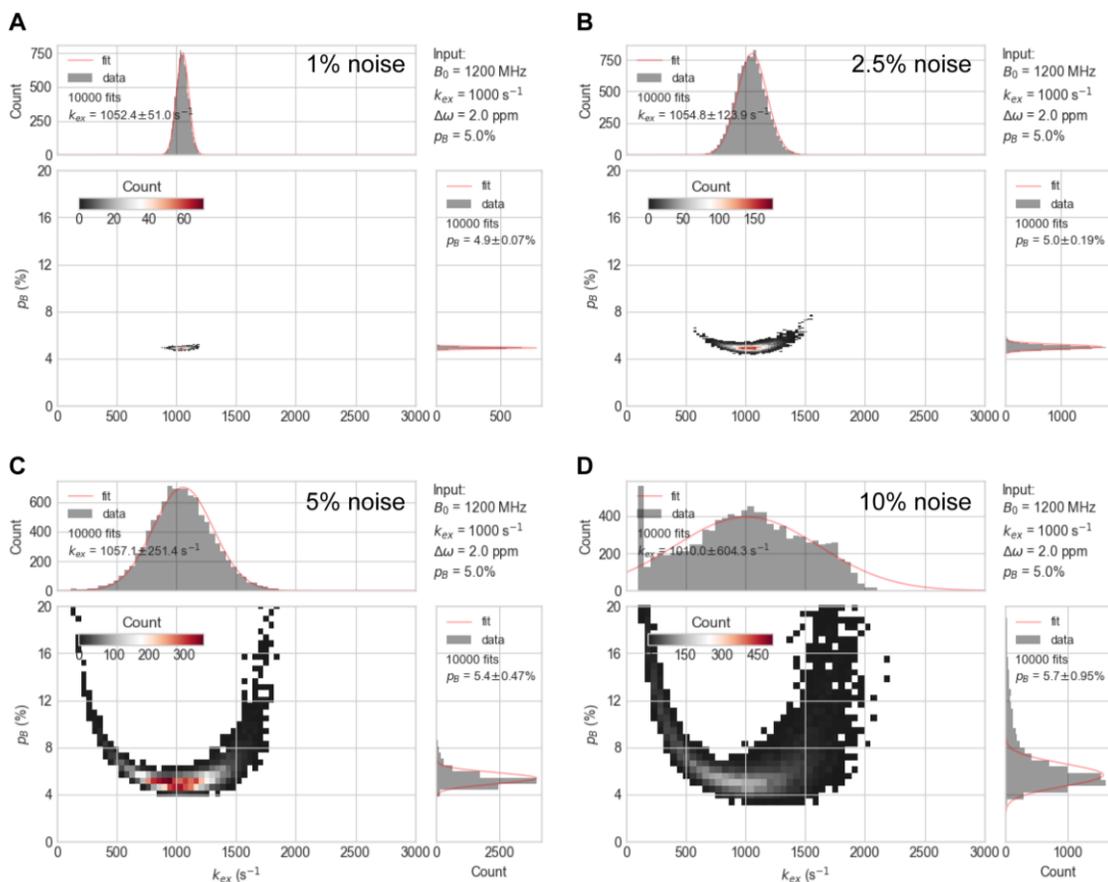


Figure 17. Distributions of fitted k_{ex} and p_{B} values obtained from CPMG RD data at a single magnetic field strength. Synthetic CPMG RD data were generated at a B_0 of 1200 MHz (1.2 GHz, the highest spectrometer frequency currently available at the time of writing) for a k_{ex} of 1000 s^{-1} , p_{B} of 5%, and ^{15}N $|\Delta\omega|$ of 2 ppm. In all panels, random noise was added to each data point, with the noise obtained from a normal distribution centered at zero with a width proportional to the value of $R_{2,\text{eff}}$ at a given data point. The added noise was equivalent to 1% (A), 2.5% (B), 5% (C), or 10% (D) of the $R_{2,\text{eff}}$ values. The 2D histograms depict the distributions of fitted k_{ex} (x-axis) and p_{B} (y-axis) values as obtained from 10,000 fits, where the random noise varied from fit-to-fit. The red lines in the 1D histograms represent fits to a Gaussian.

As all of the above fits of CPMG RD data at a single field were for a lone dispersion, I generated a synthetic set of CPMG RD data at a single magnetic field strength (500 MHz) for a group of three resonances that experience a global k_{ex} and p_{B} of 1000 s^{-1} and 5% (Fig. 18). The $|\Delta\omega|$ values were set to 2, 4, and 8 ppm ($\alpha = 1.42, 0.76,$ and 0.27) to ensure that a range of exchange regimes are included. These resonances were globally fit together to a model of two-state chemical exchange with the Bloch-McConnell equations. To my surprise, the fitted values of

k_{ex} , p_{B} , and $|\Delta\omega|$ all fell very close to the input values with small errors on the fit. The mean values contained errors that ranged from 0.8% (p_{B}) to 2.75% ($|\Delta\omega|$ for 8 ppm) (Fig. 18, bottom). Globally fitting a group of dispersions significantly reduces the variation in fitted parameters.

I thus tested whether the high accuracy afforded by this fit was due to the multiple exchange regimes present or due to the information content of three dispersions. Therefore a second global fit was performed on synthetic data that had been generated using the same values of k_{ex} and p_{B} , but with $|\Delta\omega|$ set to 4, 4, and 4 ppm ($\alpha = 0.76$) (Fig. 18, top). The fits output from this dataset were remarkably worse, with errors in $|\Delta\omega|$ reaching up to 25%, k_{ex} to 67%, and p_{B} to 71%. Thus, the quality of a fit of CPMG RD data obtained at a single static magnetic field strength not only depends on the number of dispersions, but also on the exchange regimes present (Fig. 18). This is particularly important because residues in a protein undergoing chemical exchange will have a variety of $|\Delta\omega|$ values and therefore a range of exchange regimes.

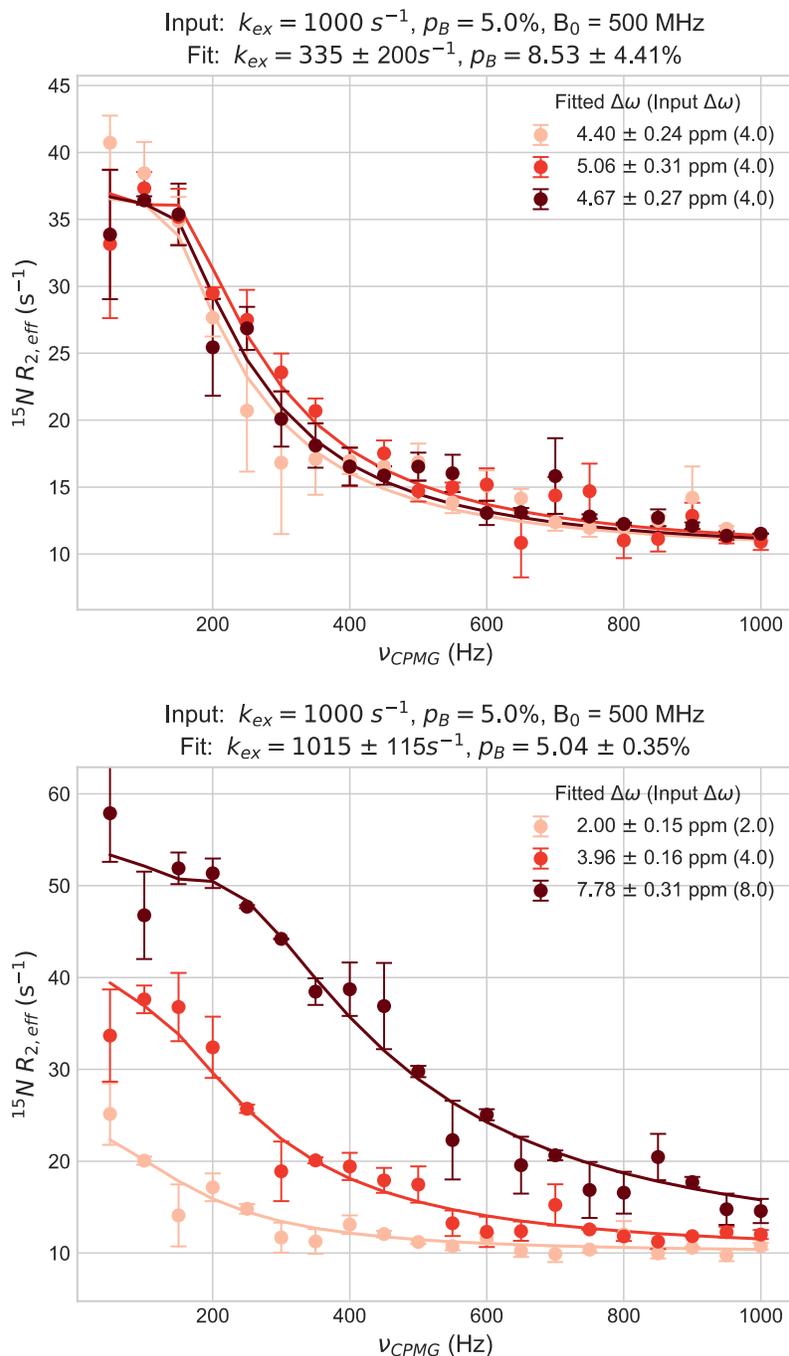


Figure 18. Globally fitting CPMG RD data obtained at a single static magnetic field strength. (Top) Synthetic CPMG RD data were generated at 500 MHz for three resonances with ^{15}N $|\Delta\omega|$ values of 4 ppm and global k_{ex} and p_B values of 1000 s^{-1} and 5%, respectively. The exchange regime is ca. slow-intermediate/intermediate ($\alpha = 0.76$). R_{2A} was set to 10 s^{-1} and 10% noise was added as described in the text. The three resonances were globally fit to the Bloch-McConnell equations with the fitted values of k_{ex} , p_B , and $|\Delta\omega|$ indicated on the plot. **(Bottom)** The same k_{ex} , p_B , and R_{2A} values were used to generate synthetic CPMG RD data for three resonances with ^{15}N $|\Delta\omega|$ values of 2, 4, and 8 ppm, which span fast/fast-intermediate, slow-intermediate/intermediate, and slow exchange regimes ($\alpha = 1.42, 0.76$, and 0.27). The fitted parameters are indicated on the plot, as above.

Increasing the number of fitted dispersions from $n = 3$ to $n = 8$ further improved the accuracy of the fitted parameters (Fig. 19). The input k_{ex} and p_B values were kept at 1000 s^{-1} and 5%, respectively, with B_0 equal to 500 MHz, as before. Instead, however, the ^{15}N $|\Delta\omega|$ values were varied from 1 to 8 ppm in 1 ppm increments, spanning all exchange regimes. Performing a global fit with a shared k_{ex} and p_B yields mean values of 1005 s^{-1} and 4.86%, which deviate from the input values by 0.5% and 2.8%, respectively. In addition, the fitted $|\Delta\omega|$ values exhibit high accuracy and precision, with errors relative to the input values ranging from 1% (1 ppm) to 3.3% (7 ppm) and fitted standard deviations ranging from 12% (1 ppm) to 1.8% (8 ppm) (Fig. 19).

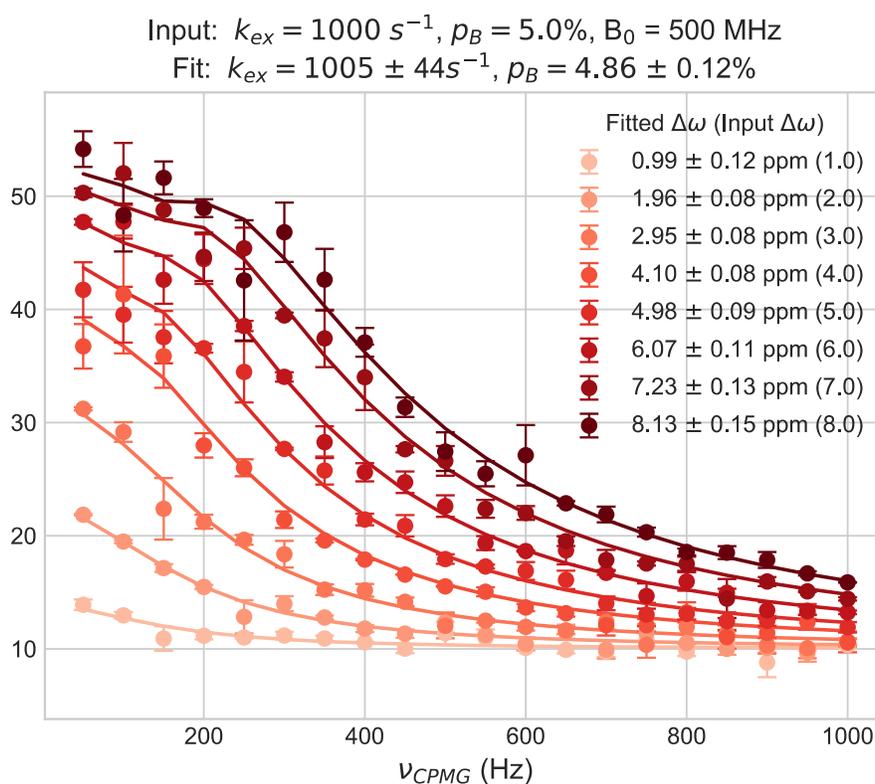


Figure 19. Globally fitting CPMG RD data obtained at a single static magnetic field strength. Synthetic CPMG RD data were generated at 500 MHz for eight resonances with ^{15}N $|\Delta\omega|$ values of 1, 2, 3, 4, 5, 6, 7, and 8 ppm and global k_{ex} and p_B values of 1000 s^{-1} and 5%, respectively. The exchange regime varies from fast (1 ppm, $\alpha = 1.82$) to intermediate (3 ppm, $\alpha = 1.05$) to slow (8 ppm, $\alpha = 0.26$). The data were fitted to a global k_{ex} and p_B , with the best-fit values indicated atop the figure. The standard deviation of the fit is determined from the variance-covariance matrix. Values for fitted $|\Delta\omega|$ values are indicated on the right, with input values indicated in parentheses.

Finally, as mentioned above, in a typical protein NMR study, as many as 20 or more dispersions from a protein may be fit globally to determine k_{ex} and p_{B} . I therefore generated synthetic CPMG RD datasets at one field strength as a function of increasing number of fitted dispersions and examined the distribution of k_{ex} and p_{B} values. As can be seen in Fig 20., increasing the number of globally fit dispersions from 2 to 10 to 20 and then 40 has the impact of considerably narrowing the distribution of fitted k_{ex} and p_{B} values. All datasets contained $|\Delta\omega|$ values of 1 and 3 ppm, and datasets with n dispersions (more than two dispersions) also contained $n-2$ randomly sampled values between 1 and 3 ppm. The amount of added noise in all cases was 10% of the $R_{2,\text{eff}}$ values, as described above, and the input k_{ex} and p_{B} were 1000 s^{-1} and 5%. The dispersions were simulated at a magnetic field strength of 500 MHz (Fig. 20).

In conclusion, CPMG RD provides highly insightful information about chemically exchanging systems, including details about the structural changes and kinetics and thermodynamics that accompany them. The precise dependence of dispersion profiles on the input parameters can, at times, be unintuitive. Thus, a software named *CESARICA* has been written to enable routine simulation of CPMG RD data across a wide range of parameter space and for a range of exchanging systems. At present, these systems include unimolecular exchange ($A \rightleftharpoons B$) and bimolecular exchange, encompassing both monomer-dimer exchange ($2A \rightleftharpoons A_2$ or $A_2 \rightleftharpoons 2A$) and a protein-ligand binding event ($P + L \rightleftharpoons PL$ or $PL \rightleftharpoons P + L$), where the term written first represents the major (observable) conformation. Future additions to the software will include exchange between higher-order oligomers (e.g. dimer-tetramer) and three-state exchange (e.g. linear or branched models), and a module to input thermodynamical parameters such as ΔH and ΔS for the forward and backward reactions, so as to model the temperature-dependence of CPMG RD data. Finally, *CESARICA* also contains modules to fit CPMG RD data to either the Carver-Richards or the Bloch-McConnell equations at one or more static magnetic field strengths. For now, the 2-by-2 Bloch-McConnell equation has been implemented, but future efforts will

include implementing the full 32-by-32 matrix that takes into account off-resonance effects, and other spin physics. The figures were not included here, but CESARICA can also conduct a grid search over a wide-range of k_{ex} and p_{B} parameters and map the χ^2 surface to analyze the true errors. A bootstrapping procedure will also be included in future developments.

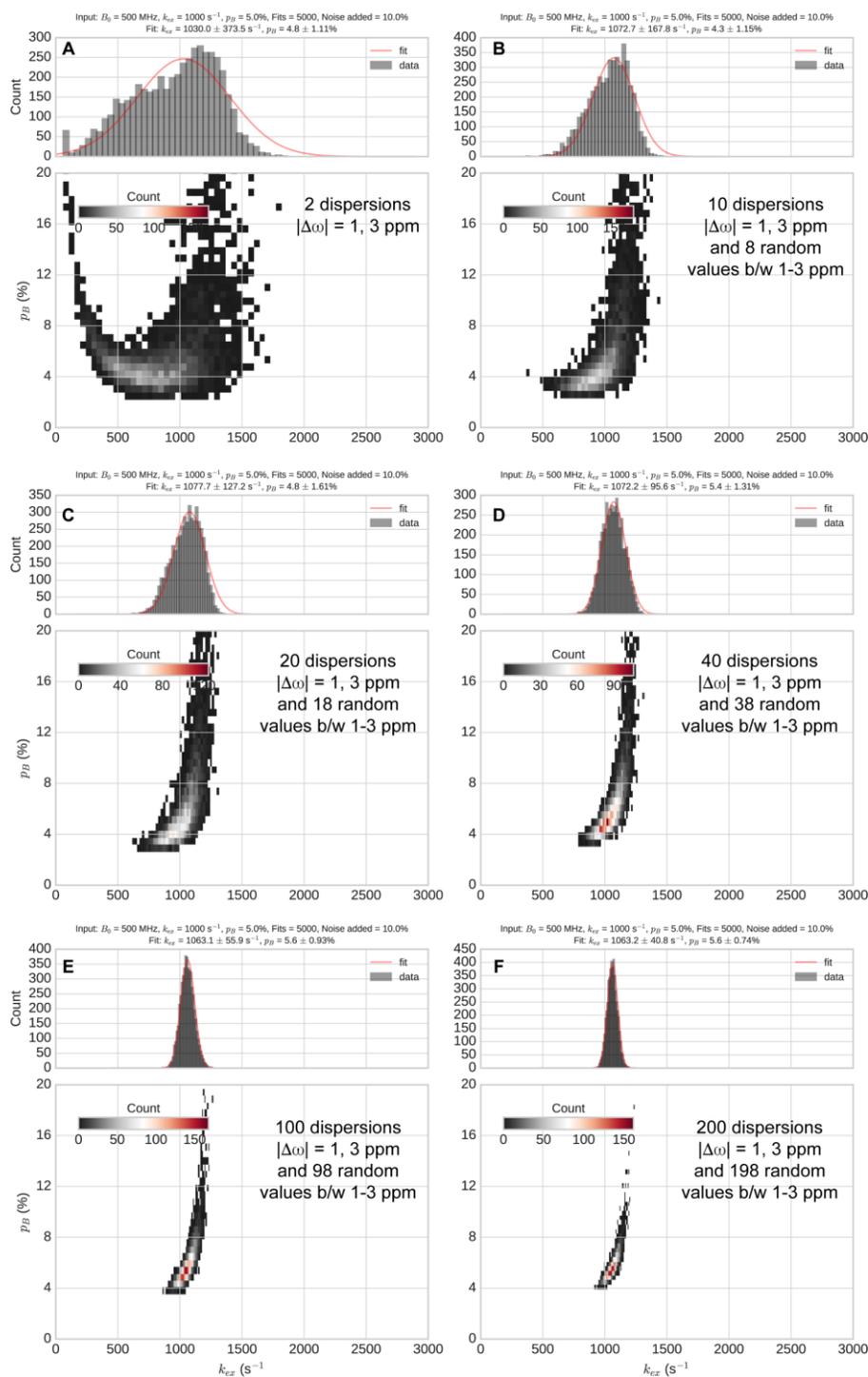


Figure 20. Synthetic CPMG RD data were generated at 500 MHz for a range of $|\Delta\omega|$ values between 1 and 3 ppm. For > 2 residues globally fit, a random value between 1 and 3 ppm was selected, but 1 and 3 ppm were always included. As the number of globally fit residues is increased, the distribution of fitted parameters becomes more narrow. 2D histograms are shown for (A) 2, (B) 10, (C) 20, (D) 40, (E) 100, and (F) 200 residues. The latter two are likely not practical, but were investigated for illustrative purposes.

2.5 References

- Al-Hashimi HM. 2013. NMR studies of nucleic acid dynamics. *J. Magn. Reson.* 237:191
- Allard P, Helgstrand M, Härd T. 1998. The Complete Homogeneous Master Equation for a Heteronuclear Two-Spin System in the Basis of Cartesian Product Operators. *J. Magn. Reson.* 134(1):7–16
- Bain AD. 2003. Chemical exchange in NMR. *Prog. Nucl. Magn. Reson. Spectrosc.* 43(3–4):63–103
- Baldwin AJ. 2014. An exact solution for $R_{2,eff}$ in CPMG experiments in the case of two site chemical exchange. *J. Magn. Reson.* 244:114–24
- Baldwin AJ, Kay LE. 2009a. NMR spectroscopy brings invisible protein states into focus. *Nat. Chem. Biol.* 5(11):808–14
- Baldwin AJ, Kay LE. 2009b. NMR spectroscopy brings invisible protein states into focus. *Nat. Chem. Biol.* 5(11):808–14
- Barbato G, Ikura M, Kay LE, Pastor RW, Bax A. 1992. Backbone dynamics of calmodulin studied by nitrogen-15 relaxation using inverse detected two-dimensional NMR spectroscopy: the central helix is flexible. *Biochemistry.* 31(23):5269–78
- Bieri M, Gooley PR. 2011. Automated NMR relaxation dispersion data analysis using NESSY. *BMC Bioinformatics.* 12(1):421
- Boehr DD, Nussinov R, Wright PE. 2009. The role of dynamic conformational ensembles in biomolecular recognition. *Nat. Chem. Biol.* 5(11):789–96
- Bothe JR, Nikolova EN, Eichhorn CD, Chugh J, Hansen AL, Al-Hashimi HM. 2011. Characterizing RNA dynamics at atomic resolution using solution-state NMR spectroscopy. *Nat. Methods.* 8(11):919–31
- Bouvignies G, Vallurupalli P, Hansen DF, Correia BE, Lange O, et al. 2011. Solution structure of a minor and transiently formed state of a T4 lysozyme mutant. *Nature.* 477(7362):111–14
- Carr HY, Purcell EM. 1954. Effects of Diffusion on Free Precession in Nuclear Magnetic Resonance Experiments. *Phys. Rev.* 94(3):630–38
- Carver J., Richards R. 1972. A general two-site solution for the chemical exchange produced dependence of T_2 upon the carr-Purcell pulse separation. *J. Magn. Reson.* 6(1):89–105
- Cavanagh J, Fairbrother WJ, Palmer AG, Rance M, Skelton NJ. 2007. *Protein NMR Spectroscopy : Principles and Practice.* Academic Press
- Clore GM, Szabo A, Bax A, Kay LE, Driscoll PC, Gronenborn AM. 1990. Deviations from the simple two-parameter model-free approach to the interpretation of nitrogen-15 nuclear magnetic relaxation of proteins. *J. Am. Chem. Soc.* 112(12):4989–91
- Clore GM, Tang C, Iwahara J. 2007. Elucidating transient macromolecular interactions using paramagnetic relaxation enhancement. *Curr. Opin. Struct. Biol.* 17(5):603–16
- Englander SW, Mayne L. 1992. Protein Folding Studied Using Hydrogen-Exchange Labeling and Two-Dimensional NMR. *Annu. Rev. Biophys. Biomol. Struct.* 21(1):243–65

Ernst RR, Bodenhausen G, Wokaun A. 1991. *Principles of Nuclear Magnetic Resonance in One and Two Dimensions*. Clarendon Press

Farrow NA, Zhang O, Forman-Kay JD, Kay LE. 1994. A heteronuclear correlation experiment for simultaneous determination of ^{15}N longitudinal decay and chemical exchange rates of systems in slow equilibrium. *J. Biomol. NMR*. 4(5):727–34

Fawzi NL, Ying J, Ghirlando R, Torchia DA, Clore GM. 2011. Atomic-resolution dynamics on the surface of amyloid- β protofibrils probed by solution NMR. *Nature*. 480(7376):268–72

Fersht AR. 2013. Profile of Martin Karplus, Michael Levitt, and Arieh Warshel, 2013 Nobel Laureates in Chemistry. *Proc. Natl. Acad. Sci. U. S. A.* 110(49):19656

Gopalan AB, Hansen DF, Vallurupalli P. 2018. CPMG Experiments for Protein Minor Conformer Structure Determination. , pp. 223–42

Grey MJ, Wang C, Palmer AG. 2003. Disulfide Bond Isomerization in Basic Pancreatic Trypsin Inhibitor: Multisite Chemical Exchange Quantified by CPMG Relaxation Dispersion and Chemical Shift Modeling. *J. Am. Chem. Soc.* 125(47):14324–35

Griffiths DJ. 2005. *Introduction to Quantum Mechanics*

Hansen DF, Vallurupalli P, Kay LE. 2008a. An Improved ^{15}N Relaxation Dispersion Experiment for the Measurement of Millisecond Time-Scale Dynamics in Proteins. *J. Phys. Chem. B*. 112(19):5898–5904

Hansen DF, Vallurupalli P, Lundström P, Neudecker P, Kay LE. 2008b. Probing Chemical Shifts of Invisible States of Proteins with Relaxation Dispersion NMR Spectroscopy: How Well Can We Do? *J. Am. Chem. Soc.* 130(8):2667–75

Hore PJ, Jones JA, Wimperis S. 2000. *NMR, the Toolkit*. Oxford University Press

Huber ML, Perkins RA, Laesecke A, Friend DG, Sengers J V, et al. 2009. New International Formulation for the Viscosity of H_2O

Ishima R. 2014. CPMG Relaxation Dispersion. , pp. 29–49

Jarymowycz VA, Stone MJ. 2006. Fast time scale dynamics of protein backbones: NMR relaxation methods, applications, and functional consequences. *Chem. Rev.* 106(5):1624–71

Karplus M, Petsko GA. 1990. Molecular dynamics simulations in biology. *Nature*. 347(6294):631–39

Kay LE, Torchia DA, Bax A. 1989. Backbone dynamics of proteins as studied by ^{15}N inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease. *Biochemistry*. 28(23):8972–79

Kleckner IR, Foster MP. 2011. An introduction to NMR-based approaches for measuring protein dynamics. *Biochim. Biophys. Acta - Proteins Proteomics*. 1814(8):942–68

Kleckner IR, Foster MP. 2012. GUARDD: user-friendly MATLAB software for rigorous analysis of CPMG RD NMR data. *J. Biomol. NMR*. 52(1):11–22

Korzhnev DM, Kay LE. 2008a. Probing Invisible, Low-Populated States of Protein Molecules by Relaxation Dispersion NMR Spectroscopy: An Application to Protein Folding. *Acc. Chem. Res.* 41(3):442–51

Korzhnev DM, Kay LE. 2008b. Probing invisible, low-populated States of protein molecules by relaxation

- dispersion NMR spectroscopy: an application to protein folding. *Acc. Chem. Res.* 41(3):442–51
- Korzhnev DM, Religa TL, Banachewicz W, Fersht AR, Kay LE. 2010. A transient and low-populated protein-folding intermediate at atomic resolution. *Science.* 329(5997):1312–16
- Korzhnev DM, Salvatella X, Vendruscolo M, Di Nardo AA, Davidson AR, et al. 2004. Low-populated folding intermediates of Fyn SH3 characterized by relaxation dispersion NMR. *Nature.* 430(6999):586–90
- Koss H, Rance M, Palmer AG. 2018. General expressions for Carr-Purcell-Meiboom-Gill relaxation dispersion for N-site chemical exchange. *Biochemistry.* acs.biochem.8b00370
- Kovrigin EL, Kempf JG, Grey MJ, Loria JP. 2006. Faithful estimation of dynamics parameters from CPMG relaxation dispersion measurements. *J. Magn. Reson.* 180(1):93–104
- Libich DS, Tugarinov V, Clore GM. 2015. Intrinsic unfoldase/foldase activity of the chaperonin GroEL directly demonstrated using multinuclear relaxation-based NMR. *Proc. Natl. Acad. Sci.* 112(29):8817–23
- Lipari G, Szabo A. 1982. Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 2. Analysis of experimental results. *J. Am. Chem. Soc.* 104(1):4559–70
- Lisi GP, Loria JP. 2017. Allostery in enzyme catalysis. *Curr. Opin. Struct. Biol.* 47:123–30
- Loria JP, Rance M, Palmer AG. 1999. A Relaxation-Compensated Carr-Purcell-Meiboom-Gill Sequence for Characterizing Chemical Exchange by NMR Spectroscopy. *J Am Chem Soc*
- Mazur A, Hammesfahr B, Griesinger C, Lee D, Kollmar M. 2013. ShereKhan--calculating exchange parameters in relaxation dispersion data from CPMG experiments. *Bioinformatics.* 29(14):1819–20
- McCammon JA, Gelin BR, Karplus M. 1977. Dynamics of folded proteins. *Nature.* 267(5612):585–90
- Meiboom S, Gill D. 1958. Modified Spin-Echo Method for Measuring Nuclear Relaxation Times. *Rev. Sci. Instrum.* 29(8):688–91
- Millet O, Loria JP, Kroenke CD, Pons M, Palmer AG. 2000. The Static Magnetic Field Dependence of Chemical Exchange Linebroadening Defines the NMR Chemical Shift Time Scale. *J. Am. Chem. Soc.* 122:2867–77
- Narayanan C, Bernard DN, Bafna K, Gagné D, Agarwal PK, Doucet N. 2018. Ligand-Induced Variations in Structural and Dynamical Properties Within an Enzyme Superfamily. *Front. Mol. Biosci.* 5:54
- Neudecker P, Korzhnev DM, Kay LE. 2006. Assessment of the Effects of Increased Relaxation Dispersion Data on the Extraction of 3-site Exchange Parameters Characterizing the Unfolding of an SH3 Domain. *J. Biomol. NMR.* 34(3):129–35
- Neudecker P, Lundström P, Kay LE. 2009. Relaxation Dispersion NMR Spectroscopy as a Tool for Detailed Studies of Protein Folding. *Biophys. J.* 96(6):2045
- Nunes AM, Minetti CASA, Remeta DP, Baum J. 2018. Magnesium Activates Microsecond Dynamics to Regulate Integrin-Collagen Recognition. *Structure.* 26(8):1080–1090.e5
- Palmer AG. 2014. Chemical exchange in biomacromolecules: past, present, and future. *J. Magn. Reson.* 241:3–17
- Palmer AG. 2015. Enzyme Dynamics from NMR Spectroscopy. *Acc. Chem. Res.* 48(2):457–65

- Palmer AG, Kroenke CD, Loria JP. 2001. Nuclear magnetic resonance methods for quantifying microsecond-to-millisecond motions in biological macromolecules. *Methods Enzymol.* 339(1997):204–38
- Palmer AG, Massi F. 2006. Characterization of the Dynamics of Biomacromolecules Using Rotating-Frame Spin Relaxation NMR Spectroscopy. *Chem. Rev.* 106(5):1700–1719
- Reddy JG, Pratihari S, Ban D, Frischkorn S, Becker S, et al. 2018. Simultaneous determination of fast and slow dynamics in molecules using extreme CPMG relaxation dispersion experiments. *J. Biomol. NMR.* 70(1):1–9
- Rule GS, Hitchens TK. 2006. *Fundamentals of Protein NMR Spectroscopy.* Springer
- Sauerwein AC, Hansen DF. 2015. Relaxation Dispersion NMR Spectroscopy. , pp. 75–132. Springer US
- Skrynnikov NR, Dahlquist FW, Kay LE. 2002. Reconstructing NMR Spectra of “Invisible” Excited Protein States Using HSQC and HMQC Experiments. *J. Am. Chem. Soc.* 124(41):12352–60
- Sugase K, Konuma T, Lansing JC, Wright PE. 2013. Fast and accurate fitting of relaxation dispersion data using the flexible software package GLOVE. *J. Biomol. NMR.* 56(3):275
- Tjandra N, Feller SE, Pastor RW, Bax A. 1995. Rotational diffusion anisotropy of human ubiquitin from ¹⁵N NMR relaxation. *J. Am. Chem. Soc.* 117(50):12562–66
- Vallurupalli P, Bouvignies G, Kay LE. 2011. Increasing the Exchange Time-Scale That Can Be Probed by CPMG Relaxation Dispersion NMR. *J. Phys. Chem. B.* 115(49):14891–900
- Vallurupalli P, Bouvignies G, Kay LE. 2012. Studying “Invisible” Excited Protein States in Slow Exchange with a Major State Conformation. *J. Am. Chem. Soc.* 134(19):8148–61
- Vallurupalli P, Hansen DF, Kay LE. 2008. Structures of invisible, excited protein states by relaxation dispersion NMR spectroscopy. *Proc. Natl. Acad. Sci.* 105(33):11766–71
- Vallurupalli P, Sekhar A, Yuwen T, Kay LE. 2017. Probing conformational dynamics in biomolecules via chemical exchange saturation transfer: a primer. *J. Biomol. NMR.* 67(4):243–71
- Wagner G, Wuthrich K. 1978. Dynamic model of globular protein conformations based on NMR studies in solution. - PubMed - NCBI. *Nature.* 275(5677):247–48
- Wilkins DK, Grimshaw SB, Receveur V, Dobson CM, Jones J a., Smith LJ. 1999. Hydrodynamic radii of native and denatured proteins measured by pulse field gradient NMR techniques. *Biochemistry.* 38(50):16424–31
- Wuthrich K, Wagner G. 1975. NMR investigations of the dynamics of the aromatic amino acid residues in the basic pancreatic trypsin inhibitor. *FEBS Lett.* 50(2):265–68
- Wüthrich K. 1986. *NMR of Proteins and Nucleic Acids*
- Zhuravleva A, Korzhnev DM. 2017. Protein folding by NMR. *Prog. Nucl. Magn. Reson. Spectrosc.* 100:52–77

3

Structural and dynamical characterization of the HSP27 monomer

3.1 Abstract

The small heat-shock protein HSP27 is a redox-sensitive molecular chaperone that is expressed throughout the human body. Here, I describe redox-induced changes to the structure, dynamics, and function of HSP27 and its conserved α -crystallin domain (ACD), and provide the first structural characterization of a small heat-shock protein monomer using nuclear magnetic resonance (NMR) spectroscopy. While HSP27 assembles into oligomers, I show that the transiently populated monomers that are released upon reduction are highly active chaperones *in vitro*, but are kinetically unstable and susceptible to uncontrolled self-aggregation. By using relaxation dispersion and high-pressure NMR spectroscopy, I observe that the pair of β -strands that mediate dimerization partially unfold in the monomer. Strikingly, I note that numerous HSP27 mutations associated with inherited neuropathies cluster to this dynamic region. The high degree of sequence conservation in the ACD amongst mammalian sHSPs suggests that the exposed, disordered interface within the context of oligomers or free in solution may be a general, functional feature of these molecular chaperones.

3.2 Introduction

Small heat-shock proteins (sHSPs) are a class of molecular chaperones present in all kingdoms of life and exhibit diverse functionality, from modulating protein aggregation to maintaining cytoskeletal integrity and regulating apoptosis ([Kampinga et al. 2015](#)). The most abundant sHSP in humans ([Vos et al. 2009](#)), HSP27 (or HSPB1), is systemically expressed under basal conditions and upregulated by oxidative stress ([Yu et al. 2008](#)), during aging ([Weindruch et al. 2000](#)), and in cancers ([Ciocca & Calderwood 2005](#)) and protein deposition diseases ([Outeiro et al. 2006](#)). Numerous mutations in HSP27 have been linked to different neuropathies, including distal hereditary motor neuropathy (dHMN) and Charcot-Marie-Tooth (CMT) disease ([Evgrafov et al. 2004](#), [Houlden et al. 2008](#)), the most commonly inherited neuromuscular disorder. These maladies are themselves linked to oxidative stress ([Lin & Beal 2006](#), [Niemann et al. 2014](#)), and recent studies have indicated that the reducing environment of the cytosol progressively transitions to an oxidizing environment over the lifetime of an organism ([Kirstein et al. 2015](#), [Walther et al. 2015](#)).

HSP27 is directly sensitive to the intracellular redox state via its lone cysteine residue (C137), which controls dimerization by forming an intermolecular disulfide bond *in vivo* even under the reducing conditions of the cytosol ([Arrigo et al. 2005](#)). This cysteine is highly conserved in HSP27 orthologues but not found in other mammalian sHSPs ([Diaz-Latoud et al. 2005](#)), implying that it plays an important functional role. Accordingly, the presence of this disulfide bond impacts on the activity of HSP27 *in vitro* ([Chalova et al. 2014a](#), [Pasupuleti et al. 2010](#), [Zavialov et al. 1998](#)) and on the resistance of cells to oxidative stress ([Arrigo 2001](#), [Arrigo et al. 2005](#), [Diaz-Latoud et al. 2005](#), [Ye et al. 2016](#)). Like other mammalian sHSPs, HSP27 assembles to form a wide range of oligomers ([Haslbeck & Vierling 2015](#)) whose constituent monomers and dimers freely exchange between oligomers ([Baldwin et al. 2011](#), [Hochberg & Benesch 2014](#)). The chaperone activities of many sHSPs have been characterized *in vitro* ([Mymrikov et al. 2016](#)), but the active sHSP species remains unclear, with large oligomers ([Bepperling et al. 2012](#), [Franzmann et al. 2005](#)), small

oligomers (Fleckenstein et al. 2015, Stengel et al. 2010), and dimers (Van Montfort et al. 2001) all implicated. Intriguingly, variants of HSP27 that have an increased tendency to form free monomers display hyperactivity both *in vitro* and *in vivo* (Almeida-Souza et al. 2010, 2011).

Although functionally relevant, no sHSP monomer has yet been characterized at atomic resolution, as they are typically present at low abundance in equilibrium with higher-order oligomers. Obtaining high-resolution structural information on HSP27 is challenging, as it assembles into a polydisperse ensemble of inter-converting oligomers ranging from approximately 12 to 36 subunits (Aquilina et al. 2013, Jovcevski et al. 2015, Rogalla et al. 1999) of average molecular mass of ca. 500 kDa. Removal of the C-terminal region (CTR) and N-terminal domain (NTD) leaves a conserved ~80-residue, 'α-crystallin' domain (ACD) that does not assemble beyond a dimer (Fig. 1a). The subunits in the dimer adopt an immunoglobulin-like fold, and assemble through the formation of an extended β-sheet upon pairwise association of their β6+7 strands (Baranova et al. 2011, Hochberg et al. 2014, Mchaourab et al. 1997, Rajagopal et al. 2015a). Under oxidizing conditions, the dimer interface in HSP27 is reinforced by an intermolecular disulfide bond involving C137 from adjacent subunits centered on a two-fold axis (Baranova et al. 2011, Hochberg et al. 2014, Rajagopal et al. 2015a). Based on evidence from the closely related human sHSP paralog, αB-crystallin (HSPB5) (Jehle et al. 2009), the ACD is likely structurally similar in the context of the full-length oligomeric protein and in its isolated dimeric form. The excised ACD of both αB-crystallin and HSP27 can display chaperone activity *in vitro* (Cox et al. 2016, Hochberg et al. 2014) suggesting that important aspects of sHSP function are encoded within this domain.

Here, I have employed NMR and native mass spectrometry to interrogate the impact of redox-induced changes to the structural features of HSP27, its excised ACD (cHSP27), and mutants (C137S and H124K/C137S) that affect its ability to dimerize. Against a range of client proteins including Parkinson's disease related α-synuclein (αS), citrate synthase (CS), malate

dehydrogenase (MDH), α -lactalbumin (α Lac), insulin, and glyceraldehyde-3-phosphate dehydrogenase (GAPDH), I find that HSP27 is a more active chaperone under conditions that favor the release of free monomers. Under all conditions tested, the monomeric form of the ACD is a more effective chaperone than its dimer counterpart, and against α Lac the monomeric form almost entirely recapitulates the activity of the full-length chaperone. I demonstrate that neither the oligomeric distribution of HSP27, the conformations and 'fast' (ps/ns) dynamics of the disordered CTR, the structure of the cHSP27 dimer, or its fast dynamics vary appreciably with formation of the disulfide bond. Instead, reduction of the disulfide bond leads to monomerization, whereby monomeric cHSP27 exists in a partially unfolded conformation. Taken together, I conclude that an altered structure of the monomeric form is responsible for the redox dependent chaperone activity.

To interrogate the structure of the transiently populated monomers, I have used a combination of Carr-Purcell-Meiboom-Gill (CPMG) relaxation dispersion (RD) and high-pressure solution-state nuclear magnetic resonance (NMR) spectroscopy methods. My data reveal that monomeric cHSP27 becomes highly dynamic and disordered in the region that previously constituted the dimer interface. While I find the cHSP27 monomer to be highly chaperone-active *in vitro*, I demonstrate that increasing the abundance of the monomer results in a heightened tendency for uncontrolled self-aggregation. The importance of the unstructured region in this delicate balance between function and malfunction can be linked by mutations in HSP27 that are associated with hereditary neuropathies, which mainly cluster to the disordered region of the monomer.

3.3 Results

3.3.1 Reduction increases the population of HSP27 monomers and chaperone activity

I first examined full-length HSP27 to analyze redox-dependent changes to its oligomeric distribution. Native mass spectra of reduced and oxidized HSP27 were highly similar, with overlapping signals in the 5,000–15,000 m/z region (Fig. 1a), consistent with previous data (Aquilina et al. 2013, Jovcevski et al. 2015). This reveals that HSP27 assembles into large, polydisperse oligomers with similar oligomeric distributions under both conditions. I also observed monomeric and dimeric HSP27 in the spectra of both oxidized and reduced forms, with a significant increase in the population of free monomer upon reduction (Fig. 1a).

To ascertain whether the presence of monomers impacts on chaperone function, I initially used the model substrate citrate synthase (CS) (Buchner et al. 1998, Mymrikov et al. 2016) to probe the activity of HSP27 *in vitro*. The addition of 0.5 μ M reduced HSP27 significantly suppressed aggregation, a result that contrasts with oxidized HSP27, which appeared to co-aggregate with CS and enhance aggregation (Fig. 1b). Other sHSPs have been found to co-aggregate with substrates (Bova et al. 1999, Treweek et al. 2005, Ungelenk et al. 2016). The oligomerization of HSP27 is concentration-dependent (Cox et al. 2016, Jovcevski et al. 2015), and at ca. 2 μ M HSP27, 50% of the populated stoichiometries are expected to be dimers when oxidized (Jovcevski et al. 2015). At 0.5 μ M, I expect the majority of HSP27 to be present either as dimers (oxidized) or a mixture of non-covalent dimers and monomers (reduced, Fig. 1a) (Jovcevski et al. 2015). My chaperone activity data suggest that monomerization regulates the chaperone activity of HSP27, rendering it more effective at suppressing aggregation *in vitro*.

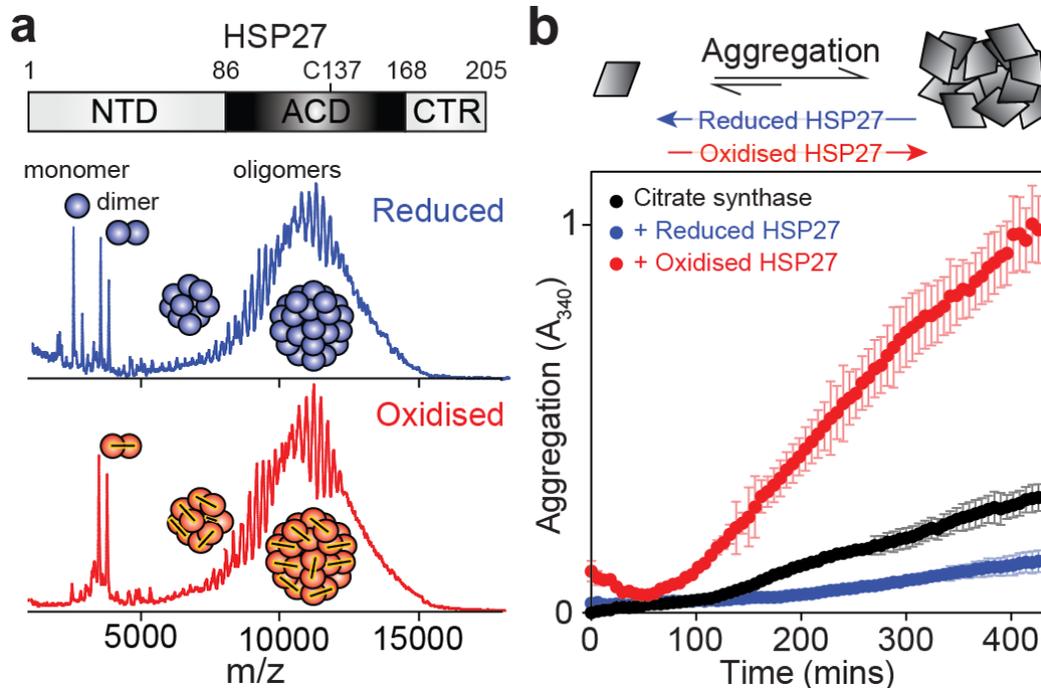


Fig. 1: Reduction of HSP27 releases highly active monomers from the polydisperse oligomers. (a) Domain architecture of the human molecular chaperone HSP27, which forms polydisperse oligomers that reach >500 kDa. Native mass spectra collected at 25 μ M total monomer concentration for both oxidized (*red*) and reduced (*blue*, +250 μ M dithiothreitol, DTT) HSP27 reveal the formation of polydisperse oligomers. More monomers are present in the reduced sample, although the oligomeric distributions are highly similar. (b) The chaperone activity of HSP27 was assayed by monitoring the increase in light scattering at 340 nm of 10 μ M of a thermo-sensitive substrate (CS) in the presence or absence of 0.5 μ M reduced (*blue*, 5 mM 2-mercaptoethanol, BME) or oxidized (*red*) HSP27. At this concentration, dimeric HSP27 comprises more than half of all populated stoichiometries.

To test the generality of this result, I examined the ability of HSP27 to suppress aggregation for a range of aggregating proteins including the Parkinson's disease-related protein α S and thermo-sensitive clients MDH and GAPDH (Fig. 2). Aggregation curves of MDH and GAPDH were independent of the addition of DTT, whereas amyloid formation by α S was approximately 2.5-fold faster (Fig. 3). Normalization of the reduced and oxidized α S curves led to a near perfect overlay, suggesting that the mechanism of aggregation was accelerated, but not altered by redox changes, thus allowing us to qualitatively compare chaperone activity. Strikingly, while the activity of HSP27 depends on the specific aggregating protein under study, HSP27 is a more effective chaperone under conditions that favor the release of free monomers.

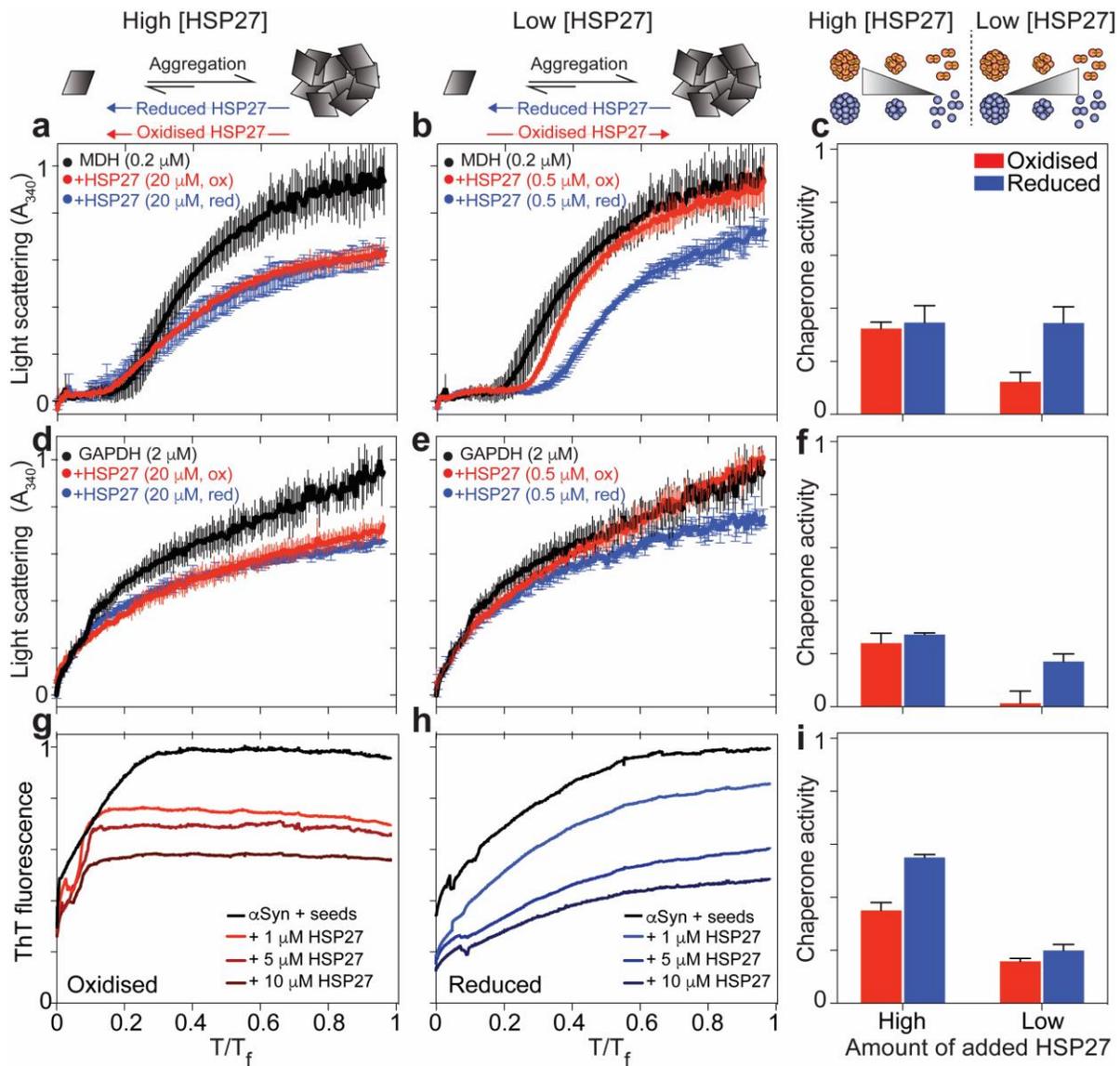


Fig. 2: HSP27 monomers are potent chaperones *in vitro*. The aggregation of MDH and GAPDH (*black*) was monitored by light scattering at 340 nm, and seeded amyloid fibril formation by α S was monitored by ThT fluorescence, each in the presence (reduced) and absence (oxidized) of 5 mM BME. The average \pm one SD of three replicates is shown in each case. The y axes were normalized for comparison, and the x axes were scaled by a factor T_f to normalize time. The T_f values used for each substrate were 5 (a, b), 5 (d, e), and 170 (g, h). In the case of α S, fibrillation when reduced was faster by a factor of two under oxidizing conditions. Experiments were repeated with either high ($\geq 10 \mu\text{M}$) or low ($\leq 1 \mu\text{M}$) HSP27. Chaperone activity is quantified as one minus the ratio of average signal over the time course with and without chaperone (Methods). The activity of HSP27 is client dependent, although in each case HSP27 showed enhanced activity under reducing conditions that favor release of free monomers.

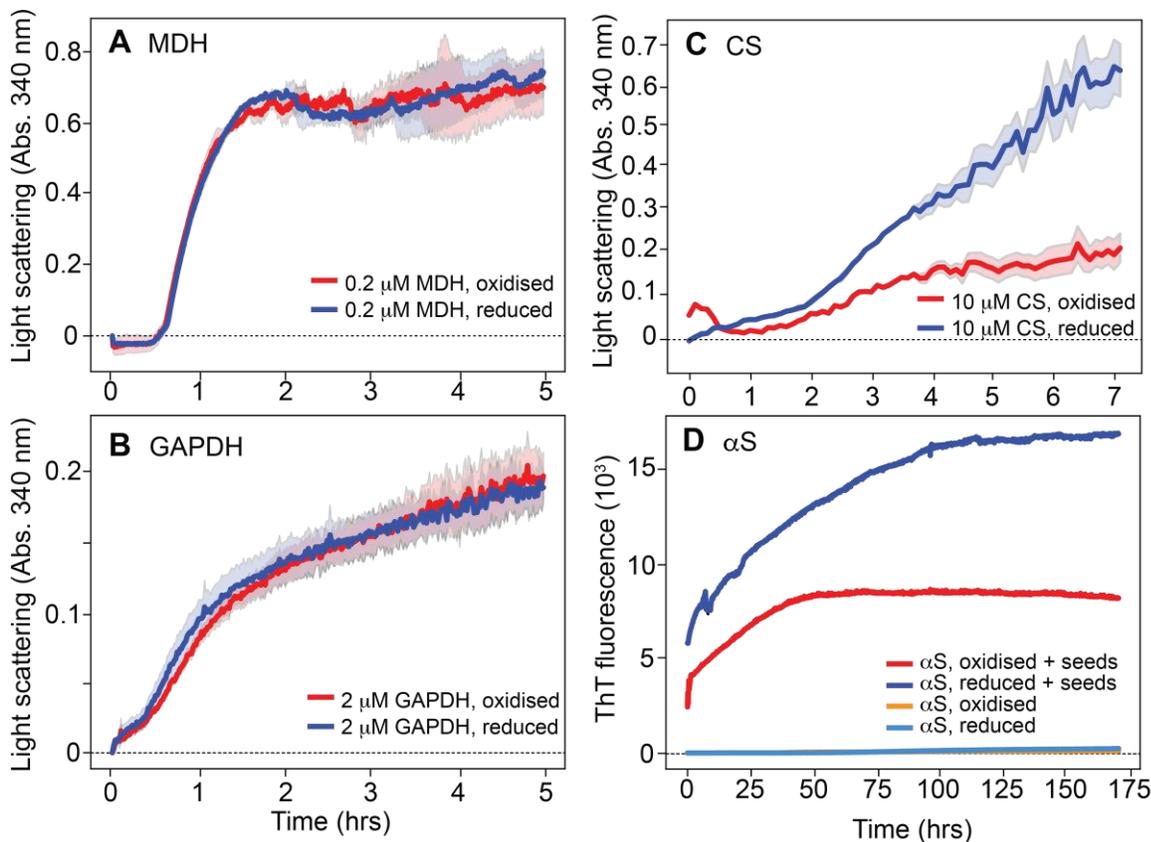


Fig. 3: Effect of oxidation on the aggregation of CS, GAPDH, and MDH and fibrillation of α S. Amorphous aggregation monitored by light scattering at 340 nm versus time for (A) malate dehydrogenase (MDH), (B) glyceraldehyde-3-phosphate dehydrogenase (GAPDH), and (C) citrate synthase (CS). The concentrations of the substrates are indicated in the lower-right corner. *Red* traces are data recorded in the absence of reducing agent and *blue* traces correspond to data obtained in the presence of 5 mM reducing agent (DTT or BME). The buffers contained either 30 mM sodium phosphate, 3 mM EDTA at pH 7 (A, B) or 30 mM sodium phosphate, 3 mM EDTA, and 100 mM NaCl at pH 7 (C). Data were obtained at 40 $^{\circ}$ C (A, B) or 37 $^{\circ}$ C (C). (D) Fibril formation by α -synuclein (α S) over time monitored by ThT fluorescence at 485 nm following excitation at 440 nm. The concentration of α S was 200 μ M and the buffer was phosphate buffered saline (PBS) at pH 7.4 with data obtained at 37 $^{\circ}$ C. In the traces indicated with “+ seeds”, a small aliquot of pre-fibrillized seeds (200 μ M initial concentration) was added. The samples contained either 5 mM DTT or no reducing agent. The addition of DTT increased the total fibril amount and diminished the initial rate of fibril formation, perhaps by altering solution conditions, e.g. ionic strength. For all data shown here, the solid lines represent the average of $n = 3$ (A, B, D) or $n = 2$ (C) with error bars representing the standard deviation from such measurements.

To assess if the dissociation of the dimers, rather than modulation of the oligomers, is the major consequence of reduction, I used solution-state NMR to examine the CTR, which can mediate the assembly of sHSPs (Basha et al. 2012, Hilton et al. 2013). As HSP27 oligomers have an average mass of ca. 500 kDa, only the disordered CTR from 15 N-labeled HSP27 can be

observed in a 2D ^1H - ^{15}N heteronuclear single quantum coherence (HSQC) NMR spectrum (Cabrita et al. 2016, Christodoulou et al. 2004, Deckert et al. 2016). To quantitatively probe the local dynamics in this region, I recorded NMR spin relaxation experiments that characterize motions on the ps-ns timescale (Fig. 4). No significant differences in the conformation or fast backbone motions were detected between oxidized and reduced forms of HSP27. My combined native MS and NMR data on the polydisperse ensemble populated by HSP27 demonstrate that the primary impact of reduction is the release of free monomers.

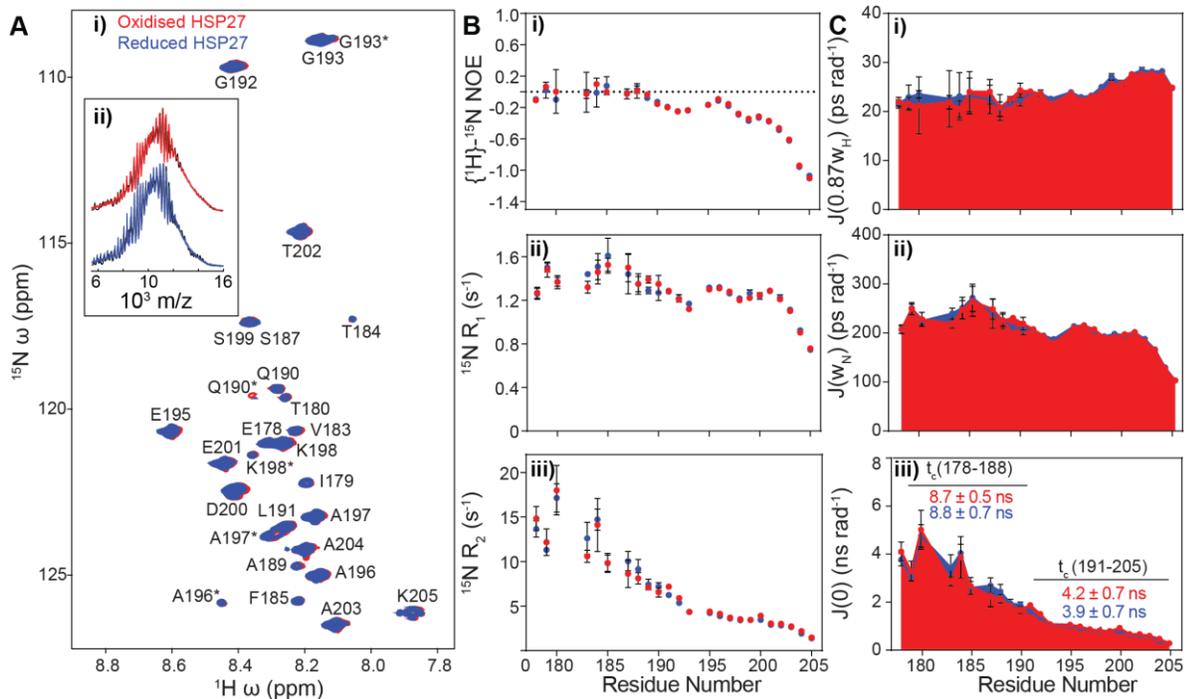


Fig. 4: Analysis of NMR and mass spectra from full-length, oligomeric HSP27. (Ai) 2D ^1H - ^{15}N HSQC spectrum of oxidized (red) HSP27 at 2 mM monomer concentration in 30 mM NaH_2PO_4 , 2 mM EDTA, 2 mM NaN_3 , 100 mM NaCl , pH 7.0 at 25 $^\circ\text{C}$. Observed resonances were assigned to residues E178-K205. Residues with asterisks arise from a minor state that has been ascribed to formation of a *cis*-G193-P194 peptide bond. The addition of 5 mM BME (blue) had no discernible effect on either the chemical shifts of the resonances, nor their intensities. (ii) The native electrospray mass spectra of HSP27 in oxidized and reduced forms (Fig. 1) can be simulated (black) to estimate the underlying distribution of oligomers. The spectra here can be well explained by assuming the oligomeric distribution follows a Gaussian distribution with means and standard deviations for (reduced/oxidized) of 23 ± 3 / 27 ± 3 and 7 ± 2 / 8 ± 2 . (B) ^{15}N spin relaxation data collected on a 14.1 T NMR spectrometer at 25 $^\circ\text{C}$. (i) $\{^1\text{H}\}$ - ^{15}N NOE, (ii) ^{15}N R_1 , and (iii) ^{15}N R_2 measurements were acquired and reported here as a function of residue number and redox state. (C) Reduced spectral density functions were calculated from the ^{15}N relaxation data, yielding three frequencies of the spectral density function (i) $J(0)$, (ii) $J(\omega_N)$, and (iii) $J(0.87\omega_H)$. No significant differences were observed in the spectral density functions between the two different redox states. It is likely that a combination of reduced motional freedom and chemical exchange result in the disappearance of residues prior to E178.

3.3.2 HSP27 monomers are potent chaperones in vitro that readily self-aggregate

Given the potent chaperone activity of the HSP27 monomer, I sought to characterize the monomer:dimer equilibrium in more detail. To isolate this equilibrium from higher order oligomer assembly, I turned to truncated forms that contains only the ACD, termed cHSP27 (Fig. 5a), which forms dimers whose structures are essentially independent of oxidation state (Baranova et al. 2011, Hochberg et al. 2014, Rajagopal et al. 2015a). In addition to the wild-type cHSP27 sequence, I produced two disulfide-incompetent variants, C137S and H124K/C137S (Clouser & Klevit 2017). The additional H124K mutation was introduced following the observations of Rajagopal et al. that determined mutations mimicking auto-protonation of the H124 side-chain can destabilize the dimeric form (Clouser & Klevit 2017). Native mass spectra of these constructs at 5 μ M monomer concentration revealed pure dimers (oxidized), pure monomers (H124K/C137S), or mixtures of the two (C137S, reduced) (Fig. 5b). This redox-dependent monomerization is consistent with the major difference I observed in the oligomeric distributions of the full-length HSP27 (Fig. 1a).

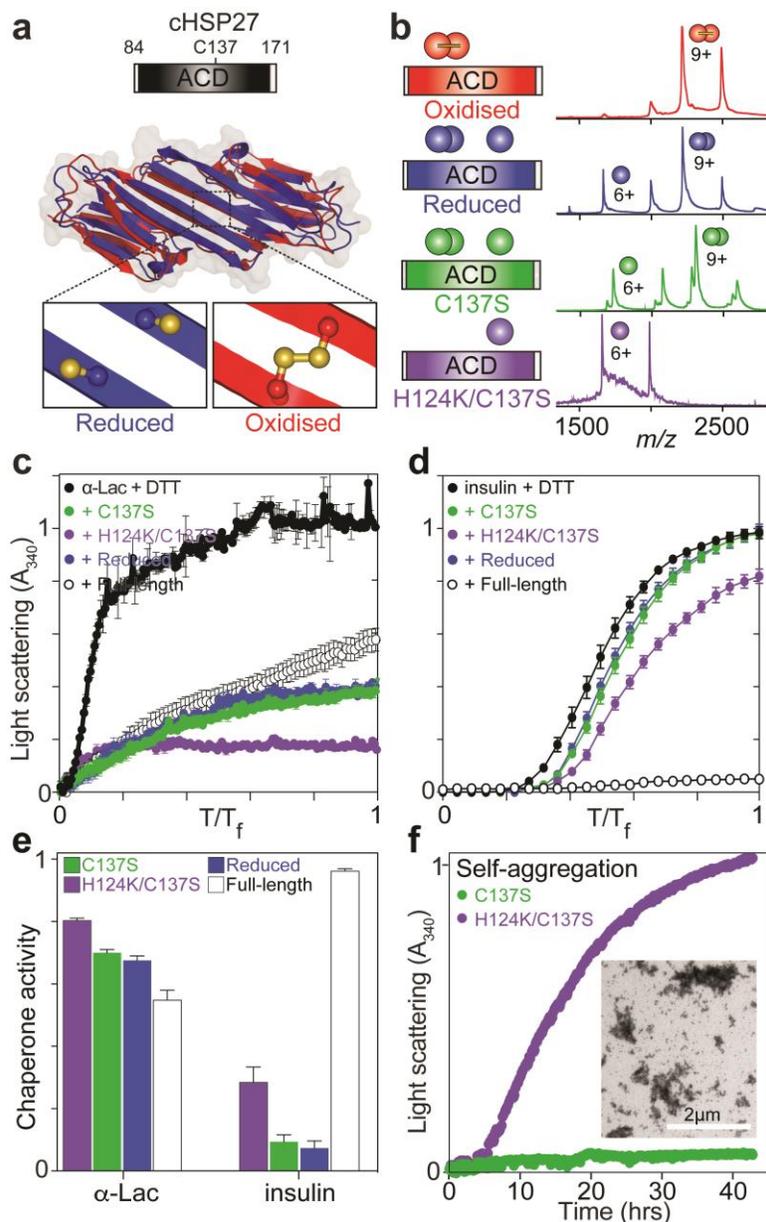


Fig. 5: cHSP27 monomers are potent chaperones *in vitro*. (a) The central α -crystallin domain (ACD) exists as stable dimers in which residue C137 at the dimer interface can access both reduced (blue, PDB 4mjh) and oxidized (red, PDB 2n3j) states. (b) The cHSP27 variants used in this study together with their native mass spectra at 5 μ M, showing the relative populations of monomers and dimers. To generate the reduced cHSP27 spectrum, 250 μ M DTT was added to the oxidized protein. (c) Aggregation of 300 μ M α -lactalbumin (α Lac; black) at 37 $^{\circ}$ C and (d) 80 μ M insulin at 40 $^{\circ}$ C following the addition of 2 mM DTT, both monitored by light scattering. The values of T_f are 10 (c) and 0.3 hrs (d). Aggregation of both proteins is inhibited by C137S (green), reduced cHSP27 (blue), H124K/C137S (purple), and full length HSP27 (empty). For α Lac and insulin, the sHSP concentration was 70 μ M and 40 μ M, respectively. The traces represent the average of three experiments with errors bars indicating \pm one SD. Under these conditions, C137S and reduced cHSP27 are predominantly dimeric whereas H124K/C137S is monomeric (e) Relative chaperone activity of the cHSP27 variants from panels (c) and (d), with activity defined as the ratio of the average signal with and without chaperone. For both substrates, the core domains are chaperone-active, with the monomeric form exhibiting more activity than the dimer. (f) The self-aggregation of 800 μ M C137S and H124K/C137S was quantified by light scattering at 340 nm at 37 $^{\circ}$ C over the course of two days. The traces represent the average of six (C137S) or three (H124K/C137S) experiments with error bars indicating \pm one

SD. Inset: A representative transmission electron microscopy (TEM) micrograph at 40 hours reveals large, non-fibrillar aggregates of H124K/C137S.

To directly compare the chaperone activity of the cHSP27 dimer and monomer, I assayed the ability of C137S and H124K/C137S to suppress the aggregation of α -lactalbumin (α Lac) (Hochberg et al. 2014) and insulin (Jovcevski et al. 2015), two clients whose aggregation is initiated by the addition of DTT (Fig. 5b). For both aggregating clients, at a concentration where H124K/C137S is predominantly a monomer and reduced C137S are dimers, the monomeric form (H124K/C137S) is a more active chaperone (Fig. 5b). Reduced cHSP27 and C137S prevented aggregation nearly identically, further reflecting the similarity between the two dimeric forms. In the case of insulin, the activity of full-length HSP27 was higher than cHSP27, but α Lac aggregation was inhibited similarly by the two HSP27 forms. These results support that HSP27 exhibits client-dependent chaperone activity to some degree (Mymrikov et al. 2016), and that, while the ACD can be an active chaperone, other components such as the NTD are also important for function (Jovcevski et al. 2015). Nevertheless, my results directly show that, against these aggregating proteins, the monomeric ACD is more active than the dimer.

Interestingly, at elevated concentration (>800 μ M) and neutral pH, the H124K/C137S monomer showed a greater propensity than C137S to self-aggregate. H124K/C137S forms large amorphous aggregates (Fig. 5f, Fig. 6) that did not display the Thioflavin-T (ThT) binding characteristic of amyloid fibrils (Fig. 6). In addition, the thermal stabilities of both reduced cHSP27 and C137S were markedly reduced at low concentrations that favor monomer release (Fig. 6). Taken together, these results suggest that, in addition to being more chaperone active, the cHSP27 monomer is also kinetically unstable and aggregation prone. I note that self-aggregation of H124K/C137S was not evident during the chaperone activity assays (Fig. 5), due to the lower concentrations employed therein.

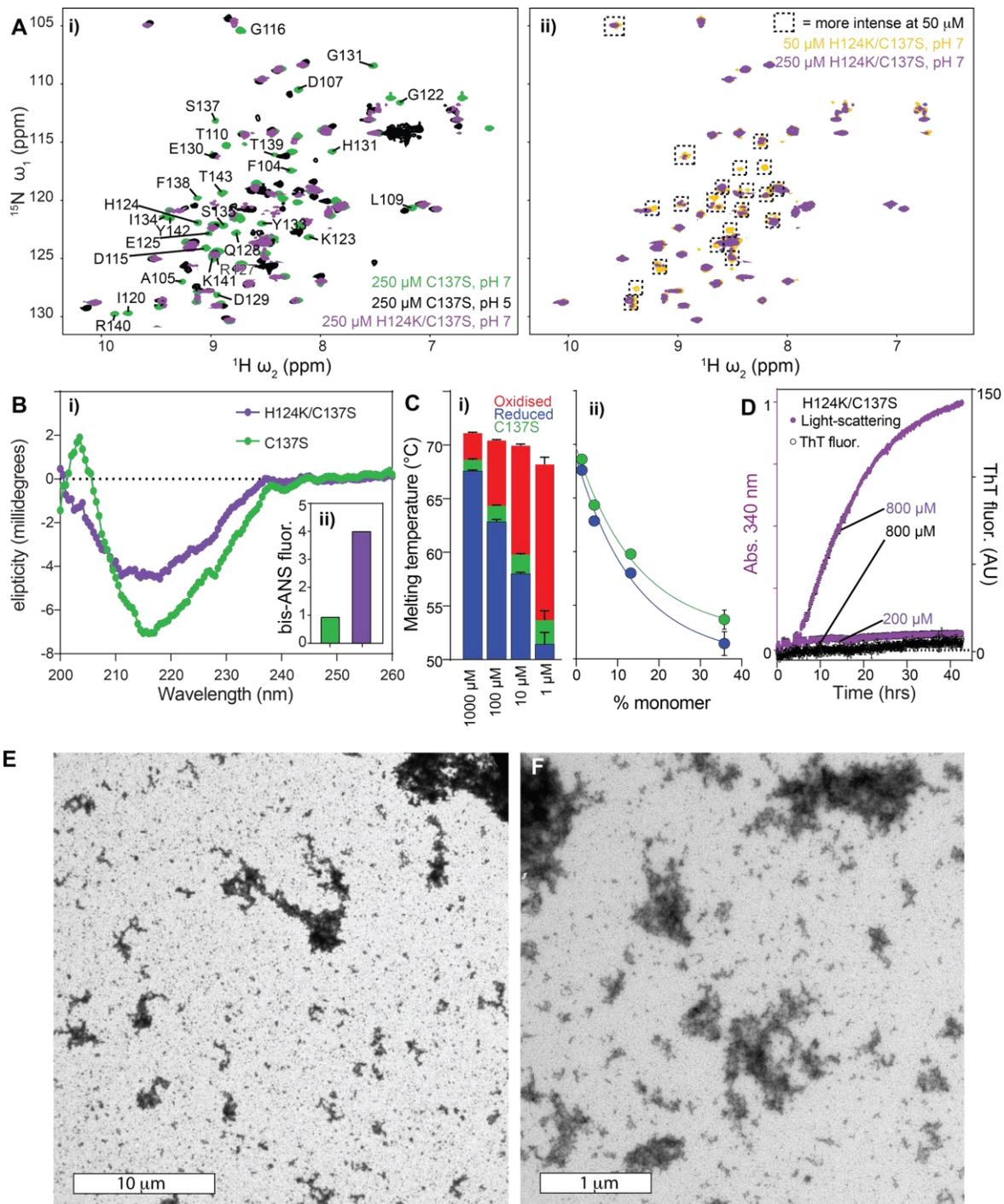


Fig. 6: The H124K/C137S monomeric mutant is destabilized and has a significantly heightened tendency to aggregate. (Ai) 2D ^1H - ^{15}N HSQC spectra of C137S at 250 μM pH 7 where the protein is substantially dimeric (green) and pH 5 where it is predominately monomeric (black). The spectrum of H124K/C137S at pH 7 (purple) closely mimics the monomer, but many peaks are exchange-broadened and not visible. (ii) Decreasing the total concentration of H124/C137S from 250 μM to 50 μM at pH 7 significantly increases the relative signal intensities of many resonances, consistent with the stabilization of the monomeric state. (B i) CD spectra of cHSP27(H124K/C137S) and C137S at 20 μM and pH 7 reveal increased propensity for disordered conformations in the monomeric mutant. (ii) Enhanced fluorescence in

the presence of the dye bis-ANS indicates that the H124K/C137S monomer has increased exposed hydrophobic sites with respect to C137S at pH 7. (C) The thermal stability of cHSP27 shown as a function of redox state and total protein concentration. (i) Nano-DSF was employed to measure the melting temperature (T_m) of oxidised (*red*), reduced (*blue*), and C137S (*green*) variants of cHSP27. The protein concentration varied logarithmically from 1000 μM to 1 μM . Data were measured in NMR buffer (methods) at pH 7 with 5 mM BME added to the reduced sample. Reduced and C137S cHSP27 showed a marked reduction in T_m at low concentration where monomers are favored. (ii) The change in T_m is shown as a function of monomer population for reduced (*blue*) and C137S (*green*) cHSP27. A K_d of 0.5 μM is assumed in these calculations, which was the average values obtained from CPMG RD data sets measured at two different protein concentrations. The solid line depicts a fit to the equation $\exp(-k x)$. (D) Light scattering at 340 nm was monitored as a function of time to observe the self-aggregation of H124K/C137S at high concentration. At 800 μM , the double mutant is highly aggregation-prone, whereas the C137S is kinetically stable (Fig 6). By contrast, at 200 μM , the aggregation propensity of the double mutant is substantially reduced under these conditions (A i, above). No significant fluorescence in the presence of ThioflavinT suggests that amorphous aggregates are present. (E, F) Transmission electron microscopy (EM) images of amorphous aggregates of H124K/C137S. In both panels, the scale bar is shown in the lower left corner.

3.3.3 Dynamics at the HSP27 dimer interface are redox sensitive

To obtain insight into the structural rearrangements that trigger the enhanced chaperone activity and aggregation propensity of the cHSP27 monomer, I turned to solution-state NMR (Fig. 7a). Backbone NMR chemical shifts from the oxidized ^{13}C , ^{15}N -labeled cHSP27 dimer were assigned using standard triple-resonance experiments, including 3D HNC(O), HN(CA)CO, HNCA, HN(CO)CA, and CC(CO)NH spectra (Fig. 7b) (Sattler et al. 1999). A quantitative analysis of the NMR data confirmed that both the secondary structure and hydrogen-bonding network (Fig. 7c) in my construct were consistent with published structures (Baranova et al. 2011, Hochberg et al. 2014, Rajagopal et al. 2015a).

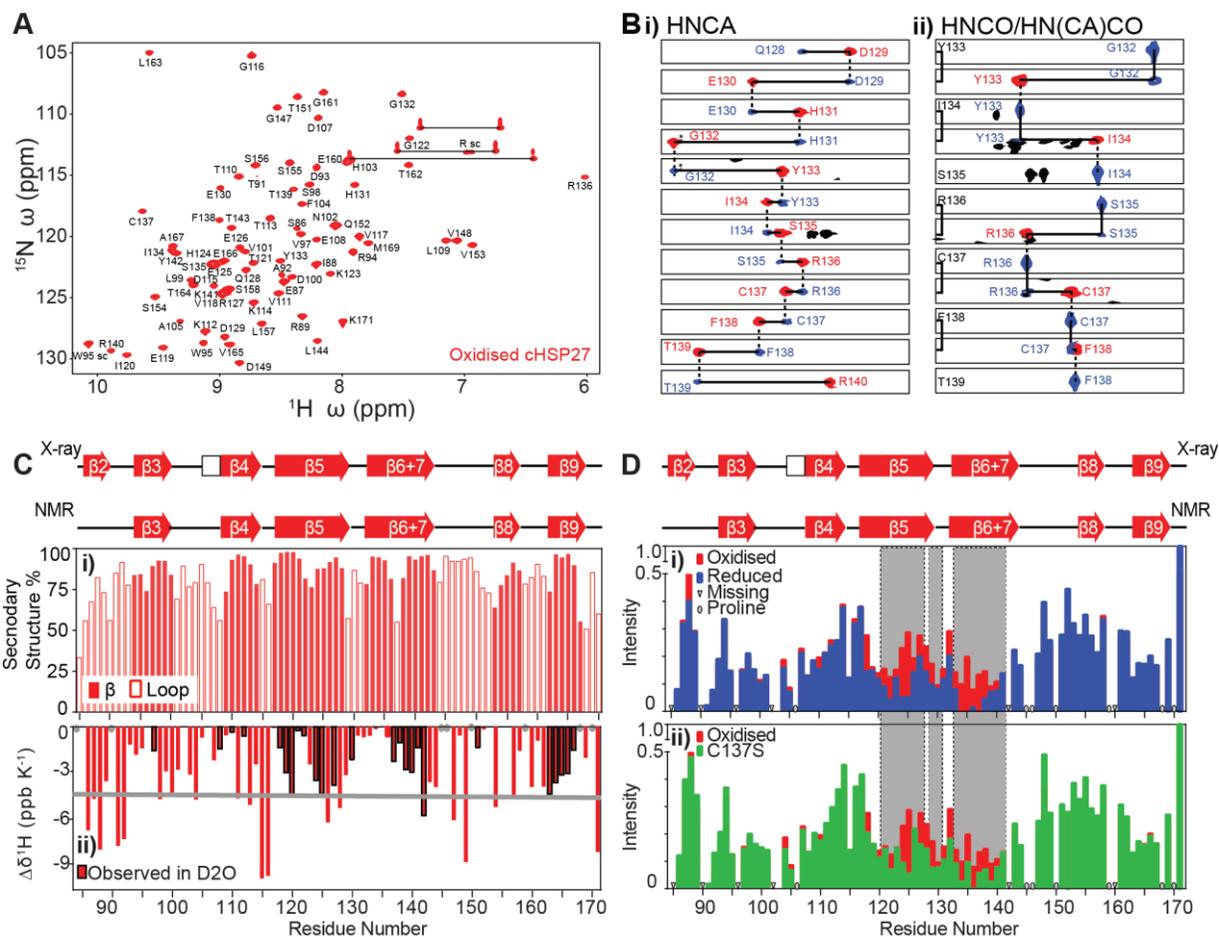


Fig. 7: Characterization of cHSP27 by NMR. (A) The assigned 2D ^1H - ^{15}N HSQC spectrum of 1 mM [U - ^{13}C , ^{15}N]-oxidized cHSP27 at 25 °C, pH 7.0 in 30 mM NaH_2PO_4 , 2 mM EDTA, 2 mM NaN_3 . Horizontal lines indicate resonances originating from the same side-chain. (B) 2D ^1H - ^{13}C strips taken from 3D HNCA (i) and HNCO/HN(CA)CO (ii) experiments recorded on [U - ^{13}C , ^{15}N]-oxidized cHSP27 at 1 mM used for resonance assignments. Correlations in ^{13}C define the sequential relationship between residues (i , red $i-1$ blue). Strip plots for residues 130-139 are shown, revealing unambiguous assignment. (C) i) TALOS-N-derived secondary structure of oxidized cHSP27 using ^1H , ^{15}N , $^{13}\text{C}\alpha$, $^{13}\text{C}\beta$, and $^{13}\text{C}\text{O}$ chemical shifts. The secondary structure is displayed above in a ribbon format. The secondary structure of cHSP27 largely agrees with previously published crystal and solution structures. I note that here the β_2 strand near the N-terminus is disordered and the α -helix between the β_3 and β_4 strands resembles a loop, with some helical tendencies. (ii) ^1H chemical shift temperature coefficients ($\Delta\delta^1\text{H}/\Delta T$) for oxidized cHSP27. Those below -4.6 ppb/K (grey line) indicate residues that are solvent-exposed and hydrogen bonded to water, whereas values above this are suggestive of hydrogen bonding to atoms within the protein. Bars that are outlined in black are resonances that are highly protected from solvent, and remain observable after exchanging into 99.9% D_2O after 40 minutes (Observed in D_2O). (D) 2D ^1H - ^{15}N HSQC spectra of 1 mM samples of oxidized (red), reduced (blue), or C137S (green) cHSP27 acquired in the presence of 5 mM β -mercaptoethanol (BME) (for the reduced cHSP27 sample only) and otherwise identical solution conditions (pH 7, 25 °C, buffer as in A). The relative HSQC peak intensities as a function of residue number for the three samples was determined, normalized to the C-terminus (K171), which was unaffected by changes in redox state. The significant attenuation of peak intensity in reduced cHSP27 and C137S could arise from either elevated solvent exchange or conformational exchange between two (or more) states on the μs - ms timescale. Elevated ^{15}N R_2 values in the region encompassing β_5 and β_6+7 , and the loop between them, $L_{5,6+7}$ supports the latter.

The above described NMR data were collected under oxidizing conditions whereby the ACD dimer is reinforced by an inter-molecular disulfide bond. To assess potential structural and dynamical changes to the ACD dimer under reducing conditions, 2D ^1H - ^{15}N HSQC NMR spectra of oxidized and reduced cHSP27 were recorded at 1 mM, a concentration that favors dimer formation (Fig. 8). The NMR spectra of cHSP27 in both redox states were highly similar, a finding that is consistent with the 2.5-Å backbone RMSD between the two forms (Fig. 5a). Moreover, ^{15}N spin relaxation experiments that probe backbone motions revealed that the ps-ns dynamics in cHSP27 were essentially unaltered by changes in redox state (Fig. 9). Similarly, I confirmed that C137S effectively mimics the reduced form, as their NMR spectra revealed very similar CSPs to the oxidized state (Fig. 8b), apart from the residues immediately adjacent to the C137S mutation.

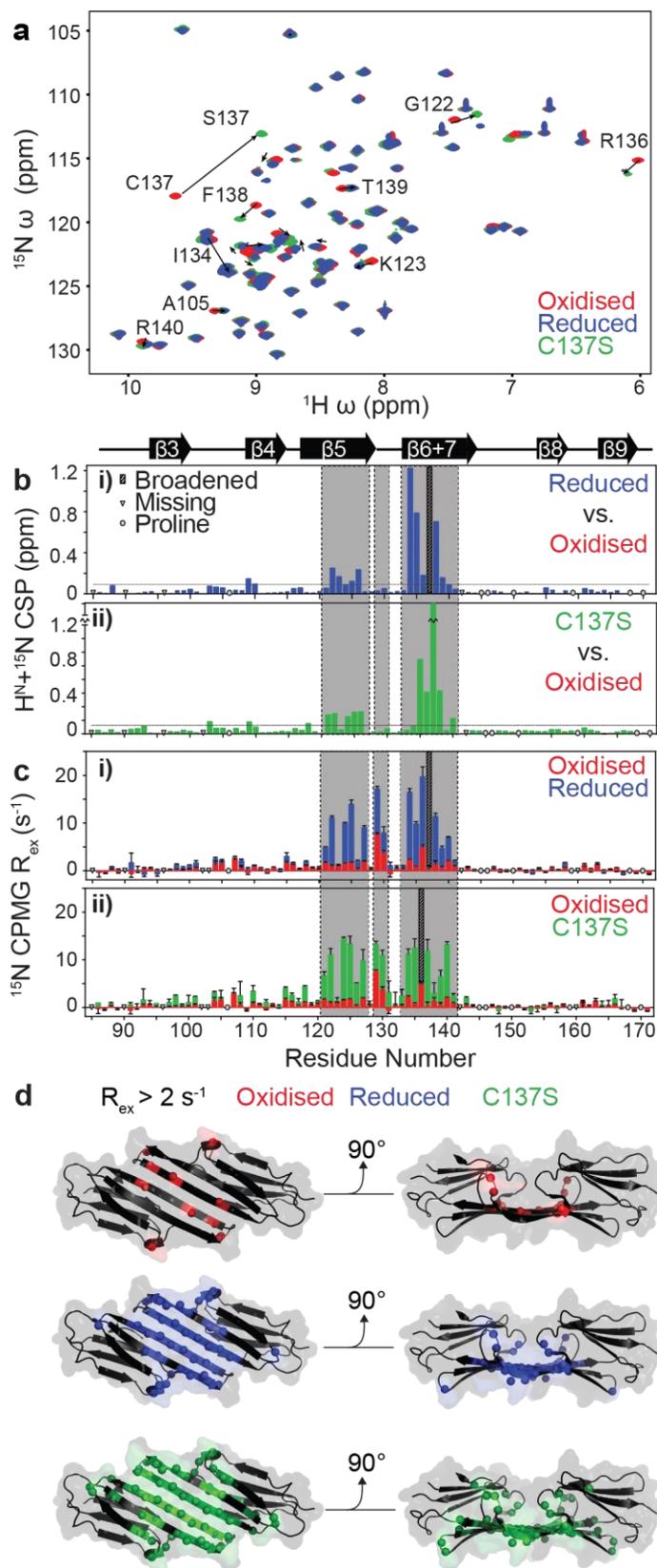


Fig. 8: Redox-induced perturbations to the structure and dynamics of cHSP27. (a) Overlaid 2D ^1H - ^{15}N HSQC spectra of oxidized (red), reduced (blue), and C137S (green) cHSP27 under identical conditions reveal their similarity. The only difference in buffer comprises 5 mM BME added to the reduced sample. Significant CSPs are indicated with arrows. (b) CSPs for individual residues between reduced (i) or C137S (ii) and the oxidized state reveal differences that localize to the $\beta 5$ and $\beta 6+7$ strands. The resonance from C137 was broadened beyond detection in reduced cHSP27 (cyan). Proline residues that do not contribute to this spectrum and unassigned residues are indicated. (c) Where R_{ex} , approximated here as the difference in $R_{2,eff}$ at the lowest and highest CPMG pulse frequencies in ^{15}N CPMG experiments, is greater than zero, there is motion on the μs -ms timescale. Residues in the vicinity of $L_{5,6+7}$ show large R_{ex} values in oxidized, reduced, and C137S cHSP27. The motion extends into the $\beta 5$ and $\beta 6+7$ strands for reduced and C137S cHSP27. (d) R_{ex} values $> 2 \text{ s}^{-1}$ are mapped onto the structure of cHSP27 (PDB 4mjh), indicating that residues with slow dynamics cluster near the dimer interface.

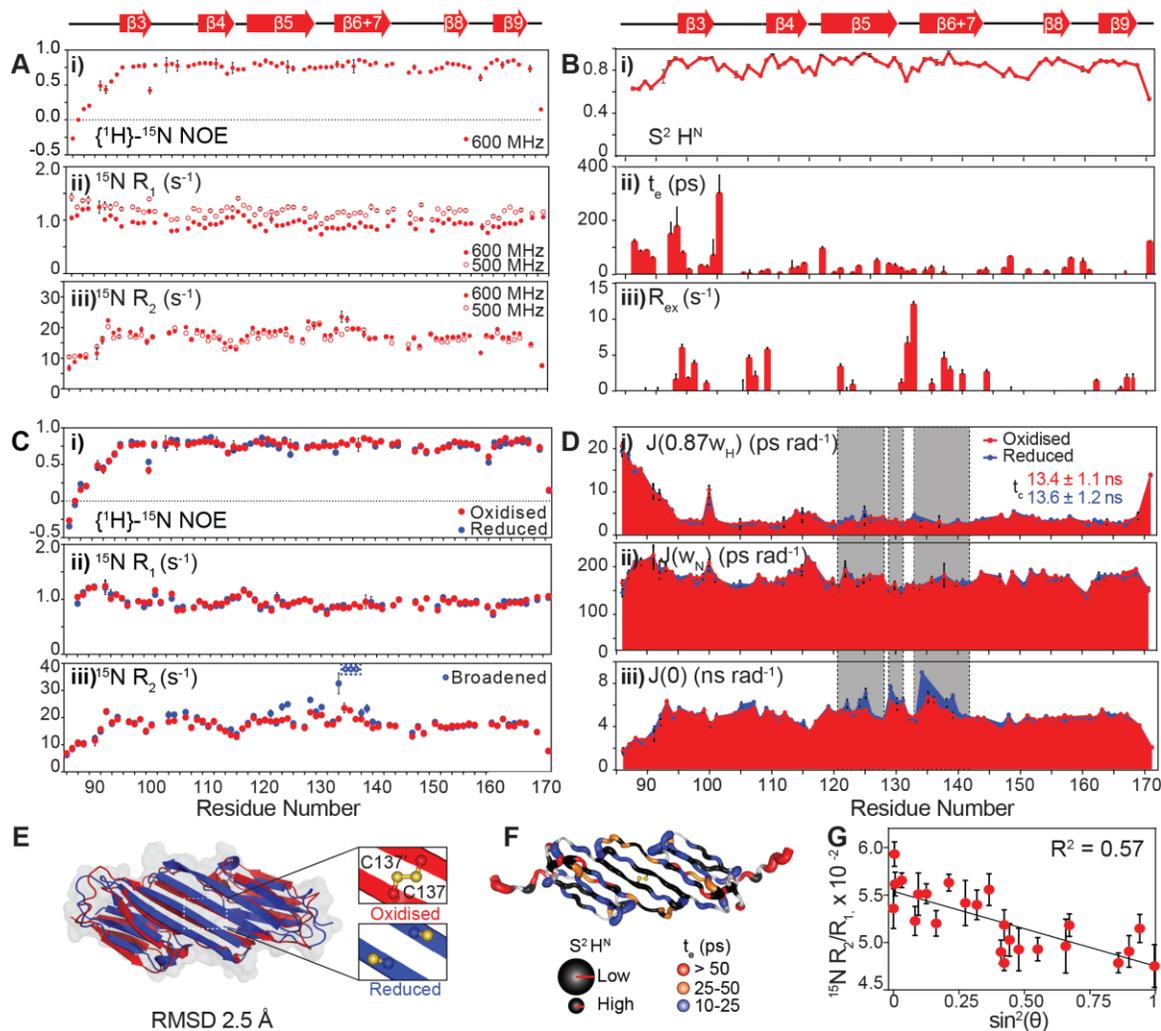


Fig. 9: ^{15}N spin relaxation experiments probe fast dynamics in reduced and oxidised cHSP27. (A) Data acquired for a model-free analysis of ps-ns dynamics in oxidised cHSP27. (i) $\{^1\text{H}\}$ - ^{15}N NOE recorded at 14.1 T (600 MHz ^1H Larmor frequency), (ii) ^{15}N R_1 data recorded at 11.7 T (500 MHz; open circles) and 14.1 T (closed circles), and (iii) ^{15}N R_2 data recorded at 11.7 T and 14.1 T at 25 °C. (B) ^{15}N R_1 , ^{15}N R_2 , and heteronuclear NOE data acquired at 11.7 and 14.1 T were used to calculate (i) generalized order parameters (S^2), (ii) effective local correlation times (t_e), and (iii) exchange contributions to R_2 (R_{ex}) for backbone H–N bond vectors. An axially symmetric diffusion tensor ($D_{\text{par}}/D_{\text{perp}} = 1.52$) was used; the determined τ_c was 13.65 ns (see Methods). (C) Comparison of ^{15}N relaxation measurements on 1 mM $[\text{U-}^{13}\text{C}, ^{15}\text{N}]$ -oxidised (red) and reduced (blue) cHSP27. (i) $\{^1\text{H}\}$ - ^{15}N NOE, (ii) ^{15}N R_1 , and (iii) ^{15}N R_2 . All data were recorded at 14.1 T and 25 °C. The open circles in panel (iii) correspond to resonances that were too broad to accurately quantify R_2 rates, suggestive of chemical exchange on the μs -ms timescale in this region ($\beta 6+7$ strand). (D) Calculation of reduced spectral density functions from the data in panel (C) enables a comparison of fast (ps–ns) and slow (μs –ms) dynamics in reduced and oxidised cHSP27, and provides the spectral density function at three frequencies, (i) $J(0.87\omega_{\text{H}})$, (ii) $J(\omega_{\text{N}})$, and (iii), $J(0)$. (E) Crystal and NMR structures of reduced (blue, PDB 4mjh) and oxidised (red, PDB 2n3j) cHSP27 dimers are highly similar (backbone RMSD of 2.5 Å). (F) S^2 and t_e values are mapped onto the structure of oxidised cHSP27 (PDB 2n3j). Low (high) S^2 values are encoded by a large (small) tube radius and t_e values are indicated by a color gradient from red (high) to blue (low). (G) ^{15}N R_2/R_1 ratios for oxidised cHSP27 plotted as a function of the angle between a

given N-H bond vector and the principal component of the cHSP27 diffusion tensor. These calculations used the X-ray structure of cHSP27 (PDB 4mjh), and the parameters were used for the local correlation time calculation (see B).

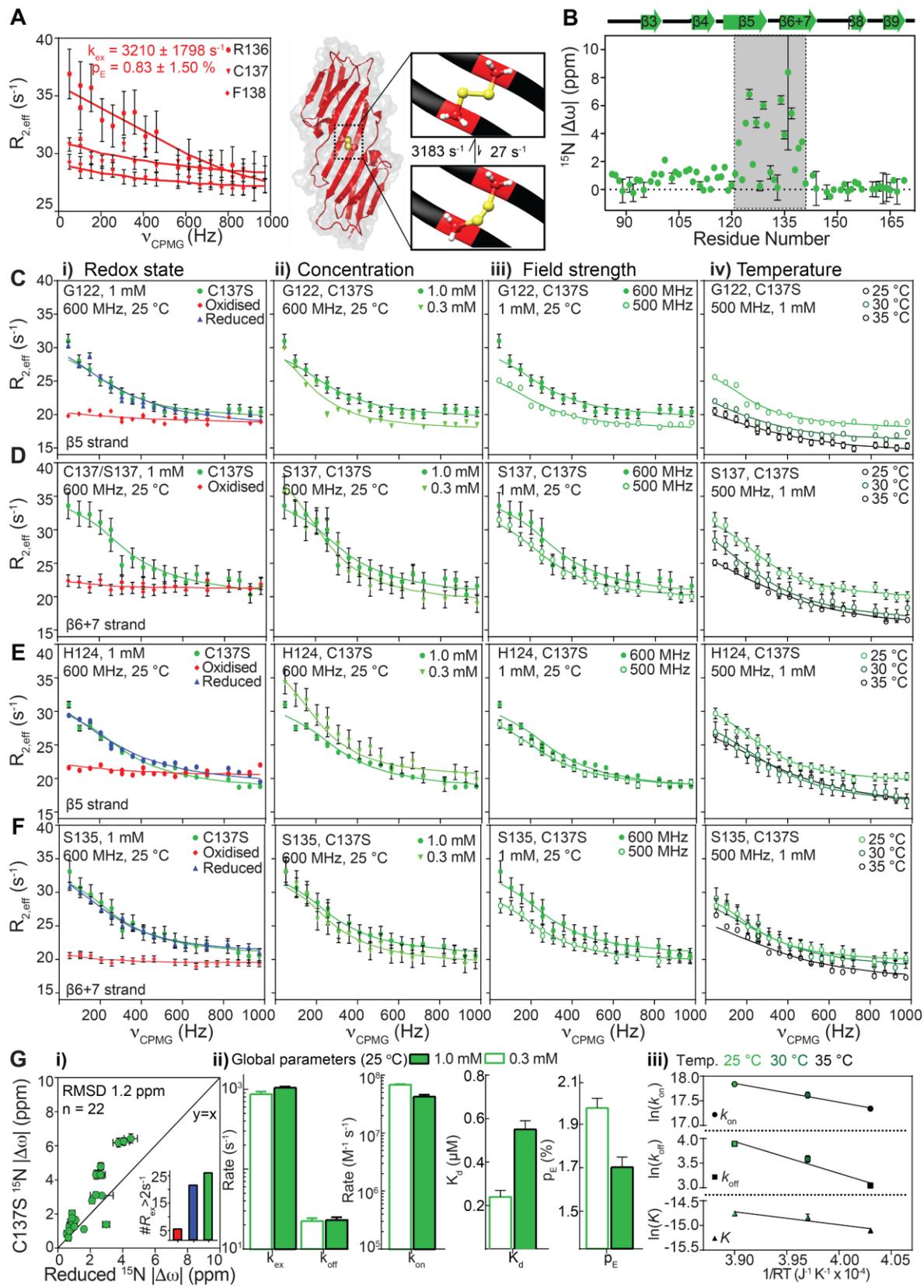
Although the structure of the underlying dimer (Hochberg et al. 2014, Rajagopal et al. 2015a) and fast backbone dynamics were redox-independent, when comparing either reduced or C137S to the oxidized form of cHSP27, the NMR signal intensities for residues in the vicinity of the dimer interface were substantially reduced (Fig. 7d), with elevated ^{15}N transverse relaxation rates (Fig. 9c). These observations demonstrate that reduction of the disulfide bond leads to dynamics on the μs -ms timescale near the dimer interface of cHSP27. Conformational fluctuations between multiple states on this timescale can be characterized using CPMG RD NMR spectroscopy experiments (Baldwin & Kay 2009, Palmer et al. 2001), which employ a variable pulse frequency, ν_{CPMG} , to measure effective transverse (R_2) relaxation rates (Carver & Richards 1972, J. Patrick Loria et al. 1999, Tollinger et al. 2001). The relative populations of the states that are interconverting (p_G, p_E), their rate of interconversion (k_{ex}), and the chemical shift differences ($|\Delta\omega|$) associated with the structural changes can be obtained through quantitative analysis of CPMG RD data.

With ^{15}N CPMG RD, I observed motions on the μs -ms timescale in the $\beta 5$, $L_{5,6+7}$, and $\beta 6+7$ regions of reduced cHSP27 and C137S, with oxidized cHSP27 only exhibiting motions in $L_{5,6+7}$ and near the disulfide bond (Fig. 8c, Fig. 8d, Fig. 10). H124K/C137S was insufficiently stable to perform the CPMG RD measurements (Fig. 5f). To allow us to characterize the motions in detail I collected an extensive set of CPMG RD data on C137S spanning multiple temperatures (25, 30, and 35°C), concentrations (0.3 mM and 1 mM), and magnetic field strengths (11.7 and 14.1 T) (Fig. 10). The dispersions were consistent with a global, two-state exchange between the principally populated dimer and a sparsely populated, but partially disordered monomeric state (Fig. 10, Fig. 11c iii). From the CPMG RD data, I calculated a K_d for the C137S monomer-dimer equilibrium of 0.5 μM (Fig. 10), a result qualitatively consistent with results obtained by native MS

(Fig. 5b). Activation and thermodynamic parameters of the monomer-dimer inter-conversion were obtained by analyzing the variation in CPMG RD data with temperature. Dissociation of the dimer was endothermic, consistent with a disruption of stabilizing interactions, and entropically favored, consistent with an increase in structural disorder. The transition state for dissociation was more disordered than the dimer as evidenced by a positive activation enthalpy and entropy (Fig. 10).

(Figure 10 is located on the next page)

Fig. 10: μ s-ms motions in oxidised, reduced, and C137S cHSP27 characterized using CPMG relaxation dispersion (RD) NMR methods. ^{15}N CPMG RD data were acquired at 11.7 and 14.1 T over a range of temperatures (25, 30, 35 °C) and concentrations (0.3 and 1 mM). CPMG RD curves were globally fit to a model of two-site chemical exchange using CATIA. **(A)** In a 1.5 mM sample of oxidised cHSP27, residues R136, C137, and F138 yielded ^{15}N CPMG RD curves that were not well explained by the thermodynamic parameters used to fit the group of residues in $L_{5,6+7}$ (Fig 3). Analyzing R136, C137, and F138 as a group yielded $k_{\text{ex}} = \sim 3200 \text{ s}^{-1}$ with $p_{\text{B}} = \sim 1 \%$. Such motions likely arise from *gauche-trans* rotations of the inter-molecular disulphide bond. **(B)** The ^{15}N chemical shift differences from C137S obtained from a global analysis of ^{15}N CPMG RD data at 25 °C. **(C-F)** ^{15}N CPMG RD data for G122 **(C)**, C137/S137 **(D)**, H124 **(E)** and S135 **(F)** in oxidised (*red*), reduced (*blue*) and C137S (*green*) cHSP27. Note that the resonance for C137 in the reduced state is too broad to yield reliable measurements. These residues are representative of the major secondary structural elements (β_5 and β_{6+7} strands). Residues in the loop between them ($L_{5,6+7}$) are shown in Fig 3. CPMG RD curves at 600 MHz, 1 mM and 25 °C **(i)** reveal the similarities in dynamics between reduced cHSP27 and C137S. For C137S, the variation of the curves with concentration **(ii)** 1 mM and 0.3 mM field strength **(iii)** 11.7 and 14.1 T and temperature **(iv)** 25 °C, 30 °C and 35 °C are shown. These data were analyzed globally using CATIA to simultaneously provide local chemical shift differences and global thermodynamic and kinetic parameters. **(G i)** Correlation between CPMG RD-derived ^{15}N chemical shift changes for C137S and reduced cHSP27. The RMSD and number of residues are listed. Note that CPMG RD data on reduced cHSP27 were available at only one magnetic field strength (14.1 T), whereas two magnetic field strengths were used to fit the C137S data (14.1, 11.7 T). However, the overall similarity between the two datasets (G i, C-F) confirms that C137S mimics the reduced state. *Inset:* the number of residues with $R_{\text{ex}} > 2 \text{ s}^{-1}$ at 14.1 T, where R_{ex} is the difference in $R_{2,\text{eff}}$ at the lowest and highest CPMG pulse frequencies. **(ii)** Kinetic parameters derived from the concentration-dependent CPMG RD data shown above for C137S. k_{on} refers to the association rate of monomers ($\text{M}^{-1} \text{ s}^{-1}$) and k_{off} to the dissociation rate of C137S dimers (s^{-1}) at 25 °C. See the text below for mathematical descriptions of their derivations. **(iii)** Temperature variation in the kinetic and thermodynamic parameters reveals activation and thermal properties of the exchange process.



Similar to reduced cHSP27 and C137S, oxidized cHSP27 exhibited μ -ms dynamics in $L_{5,6+7}$ (Fig. 11a iii, Fig. 10), but motions were not observed in $\beta 5$ or $\beta 6+7$. Detailed analysis revealed that the motions in $L_{5,6+7}$ involve unfolding of the loop, thereby disrupting the intermolecular salt bridge between D129 in $L_{5,6+7}$ and R140 in $\beta 6+7$ from the adjacent monomer (Fig. 11b), an order-to-disorder transition on the μ -ms timescale that is independent of oxidation state. In addition to the local unfolding of $L_{5,6+7}$, the oxidized form of cHSP27 showed μ -ms motions in the vicinity of residue C137 on a faster timescale, consistent with isomerisation of the disulfide bond (Fig. 10).

3.3.4 Partially disordered monomers of cHSP27 characterized by RD NMR

Given the increased chaperone activity of the monomeric ACD (Fig. 5b, 5c) I pursued a structural characterization of the C137S monomer. While unable to observe the C137S monomer directly, the ^{15}N chemical shift differences obtained from the CPMG RD experiments ($|\Delta\omega|$) report on the structure of the monomeric state. The $|\Delta\omega|$ values that I obtained indicate that residues at the dimer interface ($\beta 6+7$, $L_{5,6+7}$) adopt 'random-coil-like', disordered conformations (Fig. 11c iii). Consistent with this finding, I observed that the H124K/C137S monomer displayed characteristics of a partially disordered protein, as evidenced by its 2D ^1H - ^{15}N HSQC spectrum, circular dichroism spectrum, and bis-ANS (4,4'-dianilino-1,1'-binaphthyl-5,5'-disulfonic acid) fluorescence (Fig. 6). While the majority of the molecule retains its fold upon monomer release, as noted by small $|\Delta\omega|$ values (Fig. 10b), residues in $L_{5,6+7}$ and $\beta 6+7$, the region responsible for the dimer interface, become disordered.

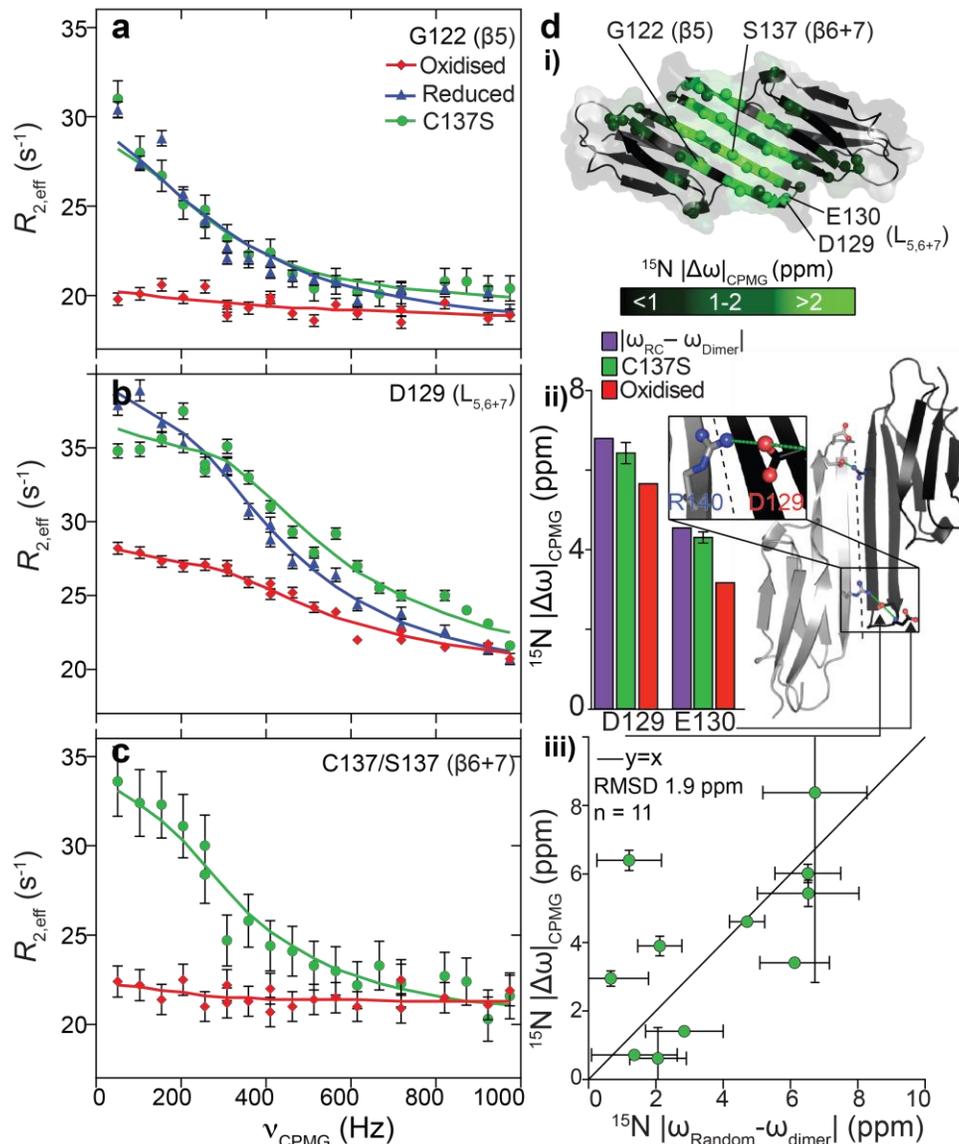


Fig. 11: Relaxation dispersion reveals partial unfolding upon transient monomer formation. ^{15}N CPMG RD experiments quantify μs - ms motions in oxidized (*red*), reduced (*blue*), and C137S (*green*) chHSP27. Fitted curves from a global analysis are shown as solid lines. Significant CPMG RD curves were observed in the $\beta 5$ strand (**a**), $L_{5,6+7}$ loop (**b**), and the $\beta 6+7$ strand (**c**). Redox-independent motions were observed in $L_{5,6+7}$, which arise from unfolding of the loop, whereas only the non-covalent dimers in C137S and reduced chHSP27 show motions throughout $\beta 5$ and $\beta 6+7$. (**d i**) CPMG RD-derived ^{15}N chemical shift changes in C137S ($|\Delta\omega|$) plotted onto the structure (PDB: 4mjh) revealing significant changes that are localized to $L_{5,6+7}$ and the $\beta 5$ and $\beta 6+7$ strands. (**d ii**) The ^{15}N $|\Delta\omega|$ values in $L_{5,6+7}$ are similar in both C137S and oxidized chHSP27, and correlate with those expected for a transition to a random coil, indicating unfolding of $L_{5,6+7}$ that is independent of oxidation state. In $L_{5,6+7}$, D129 forms an intermolecular salt bridge with R140 from an adjacent subunit, and the amide nitrogen from E130 forms a hydrogen bond with the carbonyl of D129 within the same subunit. (**d iii**) ^{15}N $|\Delta\omega|$ values in $L_{5,6+7}$ and $\beta 6+7$ upon monomerization are compared to the changes expected for random coil formation. The agreement is reasonable, indicating the monomer is substantially disordered in these regions. Outliers are indicated and mainly residue in the end of $\beta 6+7$.

3.3.5 High pressures and acidic conditions enable direct detection of the cHSP27 monomer

While CPMG RD enables a direct characterization of the HSP27 monomer under near physiological conditions, its low population (ca. 1.5% free monomer at 1 mM total monomer concentration) renders further high-resolution analysis challenging. I thus sought to stabilize the monomeric fold in order to directly observe it by NMR. A well-resolved resonance (G116) provided a straightforward marker for distinguishing between the monomeric and dimeric states. Two resonances from this residue were observed in slow exchange at low concentrations for reduced cHSP27 and C137S (Fig. 12), and the $|\Delta\omega|$ between the two resonances (1 ppm) matched the value obtained from the CPMG analysis. Following the intensities of these resonances allowed us to determine a K_d for C137S of 0.7 μM , a value consistent with the CPMG analysis. Similarly, the concentration dependence of the intensity of the G116 monomer and dimer resonances for H124K/C137S revealed an increase in the K_d by 3 orders of magnitude to ca. 1.1 mM.

To preserve the monomeric form at sufficiently high concentrations to render it amenable for atomic-resolution characterization by NMR spectroscopy, increased hydrostatic pressure (Akasaka 2006) was employed. NMR spectra of 200 μM C137S as a function of pressure from 1 bar to 2500 bar were recorded at pH 7 in a baroresistant buffer (Quinlan & Reinhart 2005) whose pH does not change with pressure. These spectra revealed a shift in the equilibrium from folded dimer at low pressure to entirely unfolded monomeric C137S at high pressure (Fig. 12a, Fig. 13) via an intermediate species that was maximally populated at 1600 bar (Fig. 12b, Fig. 12c, Fig. 13). The variation in populations of dimer, monomer, and unfolded monomer as a function of pressure were explained quantitatively by a three-state linear equilibrium model (Fig. 12b).

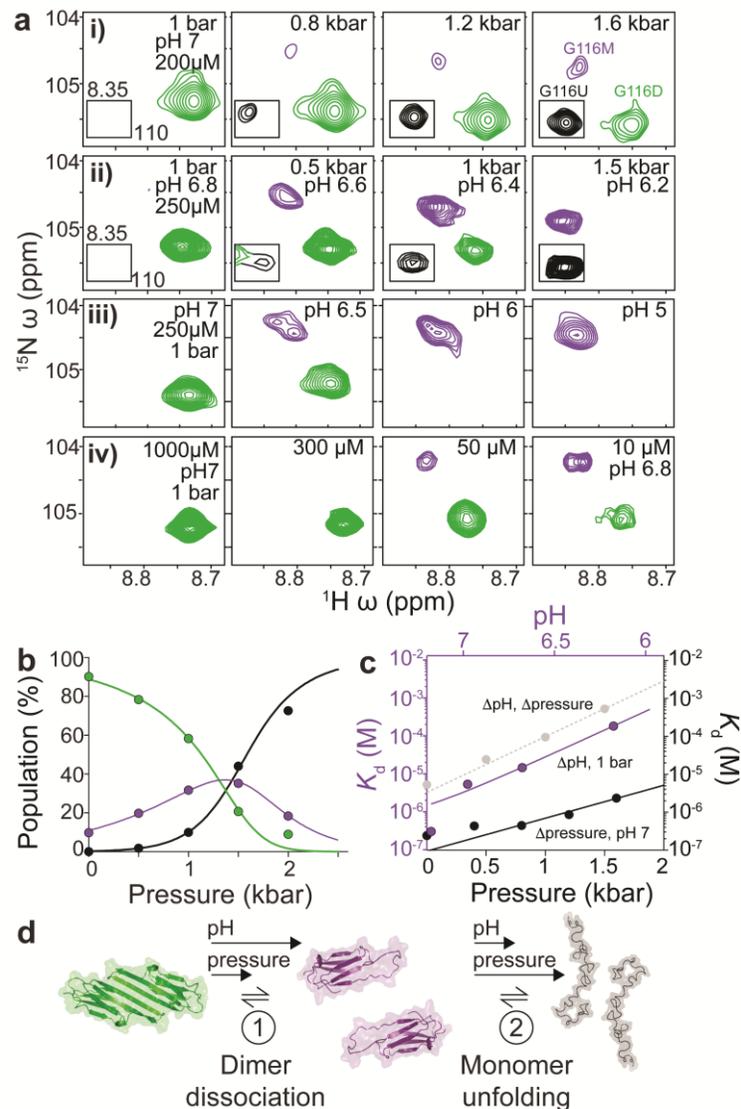


Fig. 12: High pressure and low pH stabilize the cHSP27 monomer. (a) NMR spectra of C137S, focusing on residue G116 as a function hydrostatic pressure, pH, and concentration. Shown here are (i) increasing pressures at pH 7 in a baroresistant buffer, (ii) increasing pressures in phosphate buffer at pH 6.8 at 1 bar, with the pH of phosphate buffer varying with pressure and decreasing by ca. 1 unit over 2 kbar, (iii) decreasing pH at 1 bar, and (iv) decreasing concentration at pH 7 at 1 bar. Resonances from the dimer (green), monomer (purple), and unfolded (black) state are readily distinguishable. Decreasing concentration and increasing either pH or pressure favor the monomer over the dimer. At pressures greater than approximately 1.5 kbar, the unfolded form becomes the principally populated state. (b) Variations in NMR signal intensities with pressure from four residues that were unambiguously assigned in all three conformations were well explained (solid lines) by the quantitative 3-state equilibrium model shown. (c) The K_d for dimerisation is shown as a function of pressure (black), pH (purple), and a combination of the two, acquired in phosphate buffer, where the pH of the buffer varies with pressure (grey). (d) The three-state mechanism of C137S dimer dissociation, where pH and pressure favor the partially disordered monomer, before ultimately the monomers completely unfold.

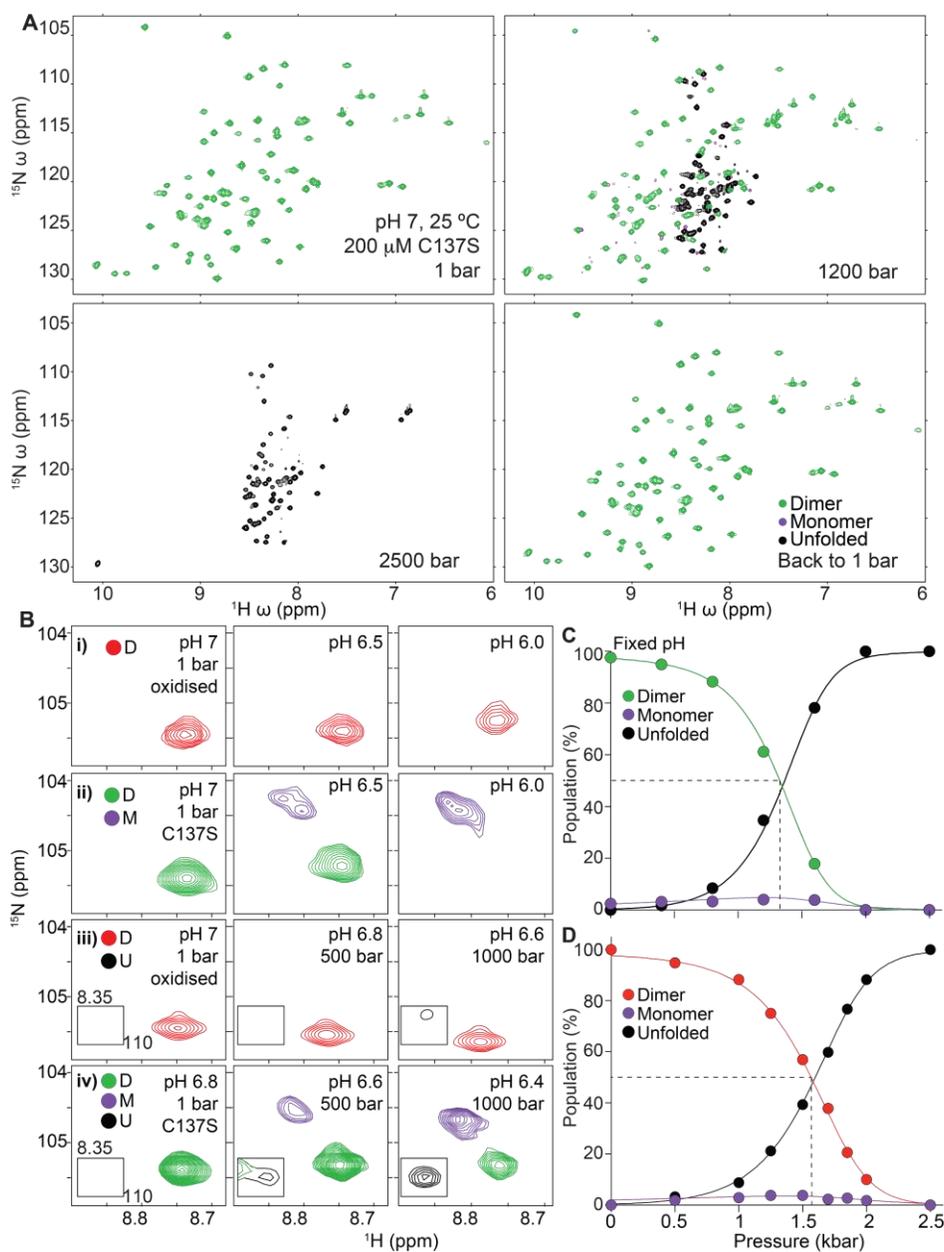


Fig. 13: NMR spectra of C137S as a function of pressure. (A) 2D ^1H - ^{15}N HSQC spectra of ^2H , ^{13}C , ^{15}N -labeled C137S at 200 μM , pH 7, 25 °C, as a function of hydrostatic pressure. Peaks from the dimer are colored *green*, the monomer *purple*, and the unfolded monomer *black*. Minor peaks from the intermediate monomeric state are evident in the 1200 bar spectrum. The pressure-induced unfolding of C137S is reversible, as demonstrated by the recovery of the folded spectrum at 1 bar after 2500 bar. The minor peaks in the spectrum at 2500 bar arise from *cis*-Pro formation at the seven X-Pro bonds in C137S. (B) Zoomed-in region from panel (A) showing the multiple resonances from G116: monomer (G116M), dimer (G116D), and unfolded (G116U). Panels (i) and (iii) depict data from oxidized cHSP27 as a function of pH and pressure, respectively, and panels (ii) and (iv) contain the corresponding data from C137S. (C) Three-state fit to the data recorded as a function of pressure at a constant pH value. See Figure 5b for the same fit performed on a sample in phosphate buffer, which varies pH with pressure. (D) Same as (C), but for oxidized cHSP27 in phosphate buffer.

Volumetric changes upon application of pressure were obtained, together with the equilibrium constant of monomer unfolding, K_u , at 1 bar, revealing a free energy difference ($\Delta_u G$) of 5 ± 0.4 kJ mol⁻¹ between the monomer and unfolded species at 1 bar and pH 7. The K_d for dimerization increased ten-fold at 1600 bar (Fig. 12c, Fig. 13), and was further increased by three orders of magnitude upon lowering from pH 7 to 5 at 1 bar (Fig. 12a). I was able to combine the effects using a phosphate buffer whose pH decreases with pressure, to maximally stabilize the monomeric form of C137S (Fig. 12b ii). By contrast, no significant change in the spectrum of oxidized cHSP27 was observed with pH, consistent with its adoption of a rigid dimer (Fig. 13b i). Interestingly, however, the oxidized protein still undergoes pressure-induced unfolding (Fig. 13b, Fig. 13c), with the unfolding pressure shifted by about +200 bar relative to the C137S dimer.

While the C137S monomer aggregated under acidic conditions at elevated protein concentrations, it remained stable up to 100 μ M at pH 4.1 at 1 bar (Fig. 14). The monomeric nature of this sample was confirmed by NMR translational diffusion measurements (Fig. 14a) wherein the monomeric state diffused *ca.* 20% faster than the dimer. Triple-resonance NMR spectra for backbone resonance assignments of the monomer were acquired under these conditions (Fig. 15a). All observable H^N, N, C α , and CO nuclei were assigned (Fig. 14c, Fig. 15a) and resonance assignments were transferred to a 2D HSQC spectrum obtained at 20 μ M concentration, pH 4.5. Similar to observations by CPMG RD, the largest CSPs relative to the C137S dimer fell in L_{5,6+7} and β ₆₊₇ (Fig. 14e, Fig. 15b i, Fig. 16). A reasonable correlation was observed (RMSD 1.6 ppm, Fig. 15d, Fig. 15e) when the ¹⁵N chemical shift differences from CPMG RD acquired at pH 7 were compared to those measured directly at pH 4.5, indicating that the monomeric conformation is similar in both cases. Further confirming their structural similarity despite a nearly ~3 unit change in pH, minor resonances from the monomeric protein could be observed in a sample of C137S at 20 μ M at pH 7, and the observed monomeric chemical shifts were close to the values obtained directly under acidic conditions (Fig. 14c, Fig. 15a). I transferred the majority of resonance

assignments from the C137S monomer at pH 4.1 to the H124K/C137S monomer at pH 7 (Fig. 14e). The CSPs between the C137S dimer and H124K/C137S monomer were consistent with those determined for the dimer-to-monomer transition at low pH (Fig. 14e, Fig. 14f), further confirming that the H124K/C137S monomeric variant resembles the C137S monomer that is present predominantly at low pH, but also transiently at neutral pH.

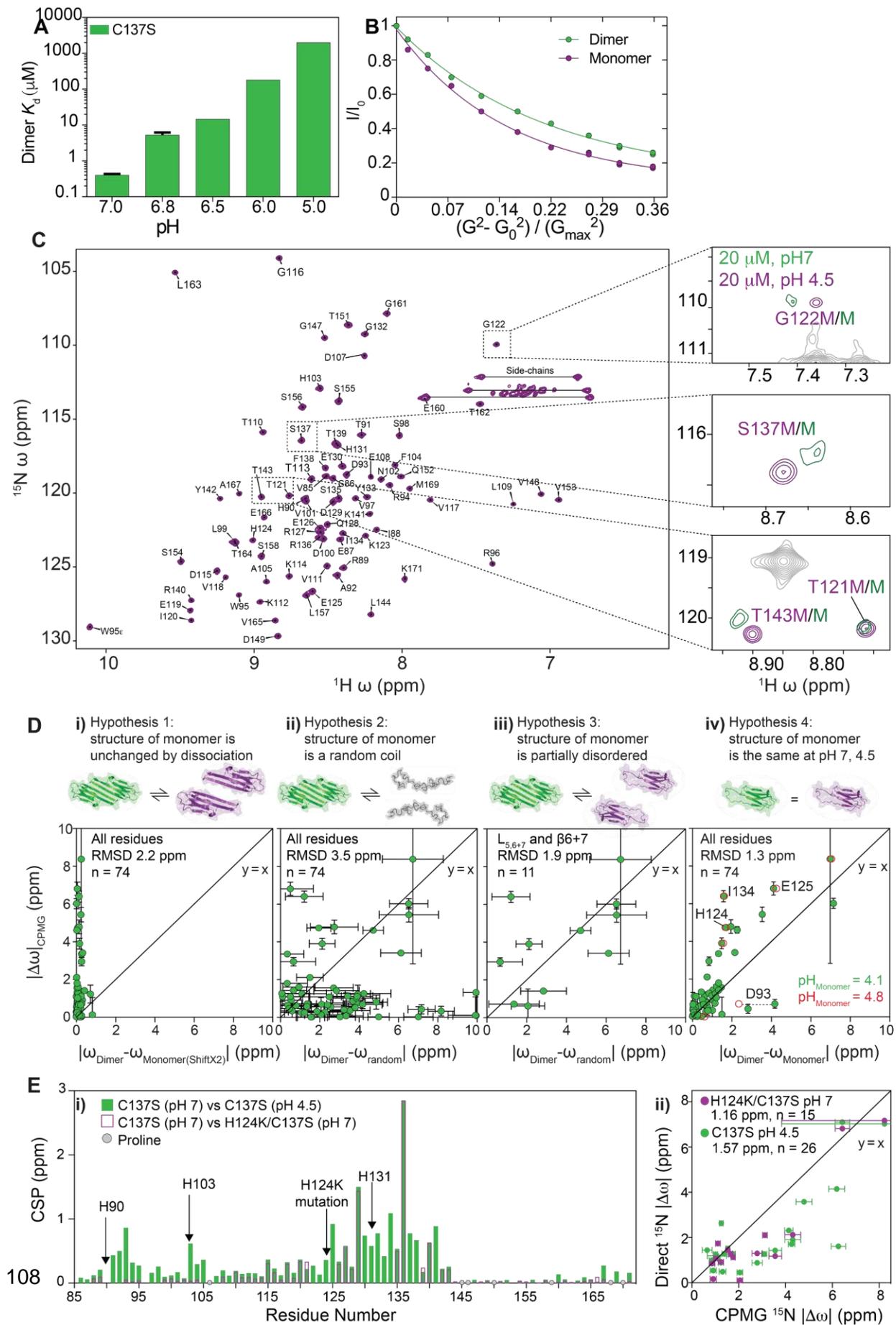


Fig. 14: Comparison of the monomer directly observed at low pH to that observed with CPMG RD at pH 7. (A) The K_d for C137S dimer dissociation shown as a function of pH. Note that the K_d is raised by over four orders of magnitude upon lowering the pH from 7 to 5. **(B)** NMR pulsed field gradient (PFG)-based translational diffusion measurements on the C137S dimer at pH 7 (*green*) and monomer at pH 4.1 (*purple*) confirm that the sample at acidic pH is indeed monomeric. The diffusion constant for the monomeric sample is ca. 25% larger than the dimer, consistent with theoretical expectations. **(C)** Assigned 2D ^1H - ^{15}N HSQC spectrum of the C137S monomer at pH 4.5 and 20 μM total protein concentration. Resonance assignments were transferred from data collected at pH 4.1 and 100 μM total protein concentration. The sample aggregates under acidic conditions at elevated protein concentration. Regions indicated **(i)** to **(vi)** contain zoomed-in views of the spectrum of the C137S monomer (*purple*) and dimer (*green*) at identical total protein concentration. Minor peaks from the C137S monomer at pH 7 are observable under these conditions, as the monomer-dimer equilibrium exists in slow exchange. The close correspondence between C137S monomer resonances at pH 7 and pH 4.5 indicates that the overall structures are similar. The grey resonances arise from the major dimer peaks in the pH 7 sample. **(D)** To identify the conformation of the dissociated, free monomer, four hypotheses were tested by comparing the expected (ShiftX2) or measured (experimental) chemical shifts to those obtained from CPMG measurements under physiological conditions. ^{15}N $|\Delta\omega|$ values were calculated for all residues ($n=74$, see Methods) using the CPMG RD data acquired at multiple fields at 25 °C and 1 mM protein concentration. **(i)** Hypothesis 1: the conformation of the monomer is essentially identical to its conformation in the dimer. **(ii)** Hypothesis 2: the conformation of the monomer is fully unfolded. The correlation between is relatively poor, indicating that the monomer retains (at least some) structured elements. **(iii)** Hypothesis 3: the conformation of the monomer is partially unfolded in the regions $L_{5,6+7}$ and β_{6+7} . This reasonable correlation indicates that this region of C137S has likely unfolded in its monomeric state. **(iv)** Hypothesis 4: the conformation of the monomer at pH 7 is similar to the species directly observed at acidic pH and dilute protein concentrations. The reasonable correlation between the CPMG chemical shift changes and those measured directly from the stabilized intermediate supports this hypothesis. In addition, the observed chemical shifts of the monomer at pH 7 (see A, above) further support this notion. Some of the outliers on this correlation plot are due to ionization at low pH, as revealed by some ^{15}N chemical shift changes between pH 4.1 (*green*) and pH 4.8 (*red*) (e.g. D93). Only the handful of resonances that showed ^{15}N chemical shift changes between pH 4.1 and pH 4.8 are indicated here. As compared to pH 7, His residues ($pK_a \sim 6$) will become protonated and Asp/Glu residues will depend on their local pK_a values ($pK_a \sim 3.8-4.1$). The resonance from D93 shows a large, pH-dependent ^{15}N chemical shift change in the 4.1 – 4.8 range. Other residues near His (e.g. H90, H103, H124, H131) may show ^{15}N chemical shift changes closer to pH 6 and above. **(E)** Comparison of the C137S dimer at pH 7 to the C137S monomer at pH 4.5 and the H124K/C137S monomer at pH 7. **(i)** Combined, weighted ^1H and ^{15}N CSPs between (*green*) the C137S dimer at pH 7 and C137S monomer at pH 4.5 and (*purple*) the C137S dimer at pH 7 and H124K/C137S monomer at pH 7. Resonance assignments for the H124K/C137S monomer were obtained by transferring those from the fully assigned C137S monomer at pH 4.5, yielding 56 assignable resonances. In general, the pattern of CSPs between the C137S dimer and H124K/C137S monomer resemble those for the C137S dimer (pH 7) and C137S monomer (pH 4.5). Differences are observed near H90, H103, and H124, which either have fully protonated side-chains that affect the backbone N-H chemical shift (H90, H103) or have been mutated and thus resonate in a new position (H124). **(ii)** Comparison of ^{15}N $|\Delta\omega|$ values measured on C137S by CPMG RD, which report on monomer-dimer structural transitions, and ^{15}N $|\Delta\omega|$ values between the C137S dimer and the C137S monomer at pH 4.5 (*green*) or H124K/C137S monomer at pH 7 (*purple*). The directly observed values were obtained through comparison of HSQC spectra. The RMSD and number of analyzed residues are indicated in the legend.

3.3.6 Structural and dynamical characterization of the partially disordered cHSP27 monomer

To characterize the monomeric state of C137S structurally, I used the observed chemical shifts to determine β -strand formation in the C137S dimer and monomer using the chemical shift (Hafsa &

Wishart 2014) and random coil indices (CSI, RCI, Fig. 15b ii, Fig. 16g). The raw chemical shift differences and secondary chemical shifts are shown in Fig. 16c-f. This analysis confirmed that, while the disordered L_{5,6+7} spans from Q128 to Q132 in the dimer, it becomes substantially elongated in the monomer and includes residues between K123 and S137, thereby shortening the β 5 and β 6+7 strands. Notably, the second half of the β 5 and first half of the β 6+7 strands are not formed in the monomer, implying that these regions fold upon dimerization. In addition, ¹H chemical shift temperature coefficients were measured, which revealed a significant loss of hydrogen bonding near the interfacial region of the monomer (Fig. 16b)

Finally, I recorded the {¹H}-¹⁵N heteronuclear nuclear Overhauser effect (hetNOE) for each backbone H-N spin pair in the monomer and the dimer, which allowed a direct comparison of fast motions on the ps-ns timescale (Kay et al. 1989). In the monomer, residues including L_{5,6+7}, the C-terminal portion of β 5, and the N-terminal portion of β 6+7 were highly dynamic (Fig. 15b iii, Fig. 16a, Fig. 16h), consistent with the random-coil-like chemical shifts observed in this area by CPMG at pH 7 and directly at pH 4.1. The rigid dimer interface in cHSP27 therefore partially unfolds and becomes highly dynamic in the monomeric state.

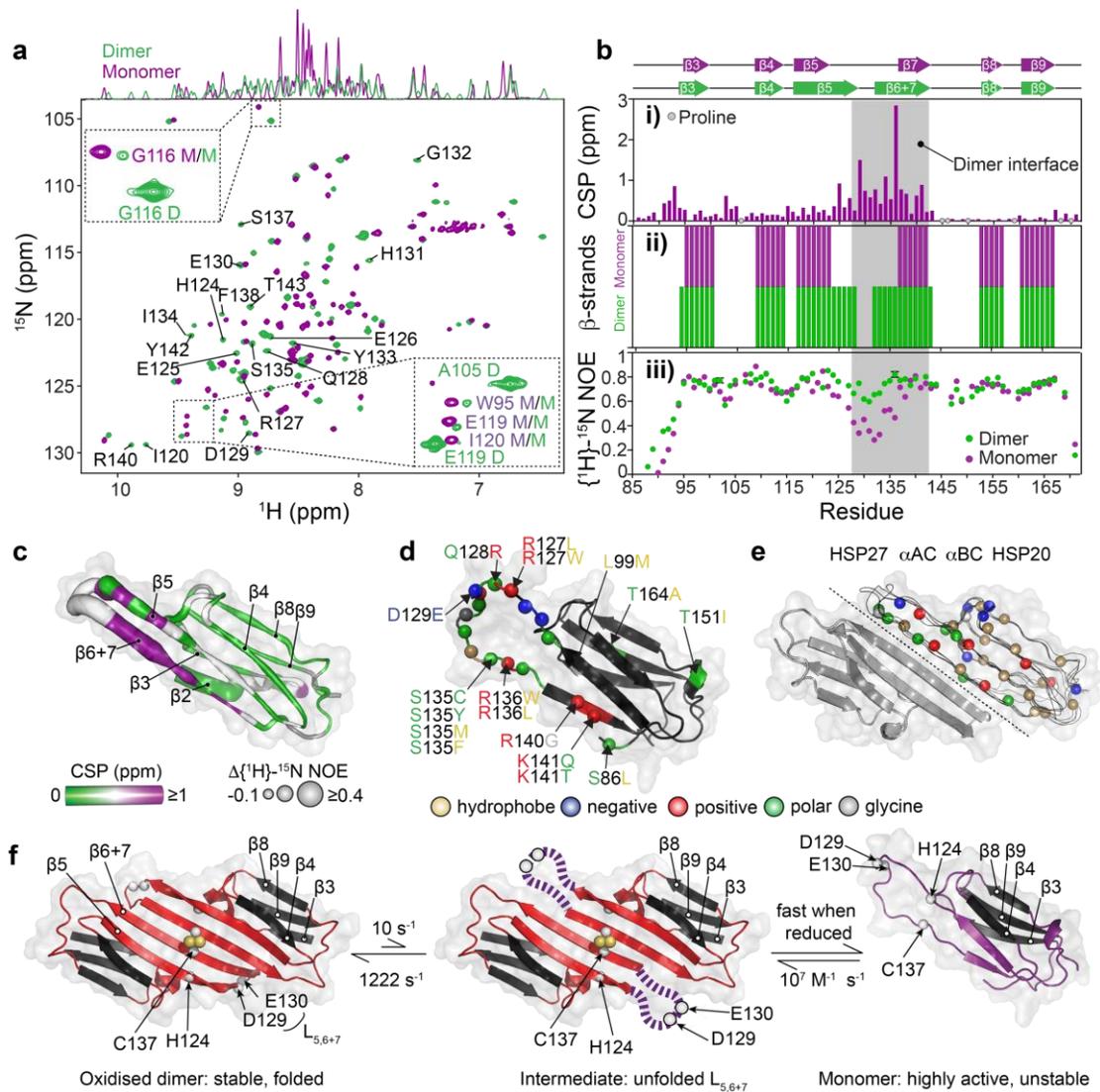


Fig. 15: Structural and dynamical characterization of the partially disordered cHSP27 monomer. (a) *i.* Overlaid NMR spectra of C137S at 20 μ M and pH 7 (green) where it is predominantly dimeric or pH 4.5 (purple) where it is monomeric. The 1D projection onto the ^1H dimension reveals significant disorder in the monomer. *Insets:* minor resonances observable in the pH 7 sample correspond to the directly observed monomer species at pH 4.5. D/M corresponds to dimer/monomer, respectively. (b) *i.* Combined and weighted ^1H , ^{15}N chemical shift perturbations (CSPs) between the C137S dimer at pH 7 and monomer at pH 4.5. (ii) β -strands in the C137S monomer (purple) and dimer (green) as identified by RCI (Berjanskii & Wishart 2006). (iii) $\{^1\text{H}\}$ - ^{15}N NOEs for the C137S dimer and monomer. (c) The ACD monomeric fold from PDB 4mjh with the difference (dimer-monomer) in $\{^1\text{H}\}$ - ^{15}N NOE values (tube thickness) and CSPs (color). (d) Inherited mutations that are implicated in the onset of CMT or dHMN disease are indicated in the ACD of HSP27. The mutations cluster to the regions that become solvent exposed upon monomer formation and tend to lower the charge density in the region. (e) Overlaid dimer structures of human HSP27 (PDB 4mjh), human α B-crystallin (PDB 4m5s), bovine α A-crystallin (PDB 3I1f), and rat HSP20 (PDB 2wj5) are shown in ribbon format for one subunit of each dimer, and indicate the highly conserved fold of vertebrate ACDs. The second subunit of HSP27 is shown in cartoon format. Highly conserved residues among human sHSPs (HSPB1-HSPB6) are shown as spheres with the same color format as (d). (f) The combined results from the CPMG RD and high pressure NMR experiments allow us to propose a hierarchical mechanism for monomer formation. The oxidized, reduced, and C137S forms of cHSP27 exhibit similar dynamics in $L_{5,6+7}$

and form a disordered loop. In the absence of a disulfide bond, this motion in L_{5,6+7} propagates, resulting in the eventual unfolding of the β 5 and β 6+7 strands in the free monomer.

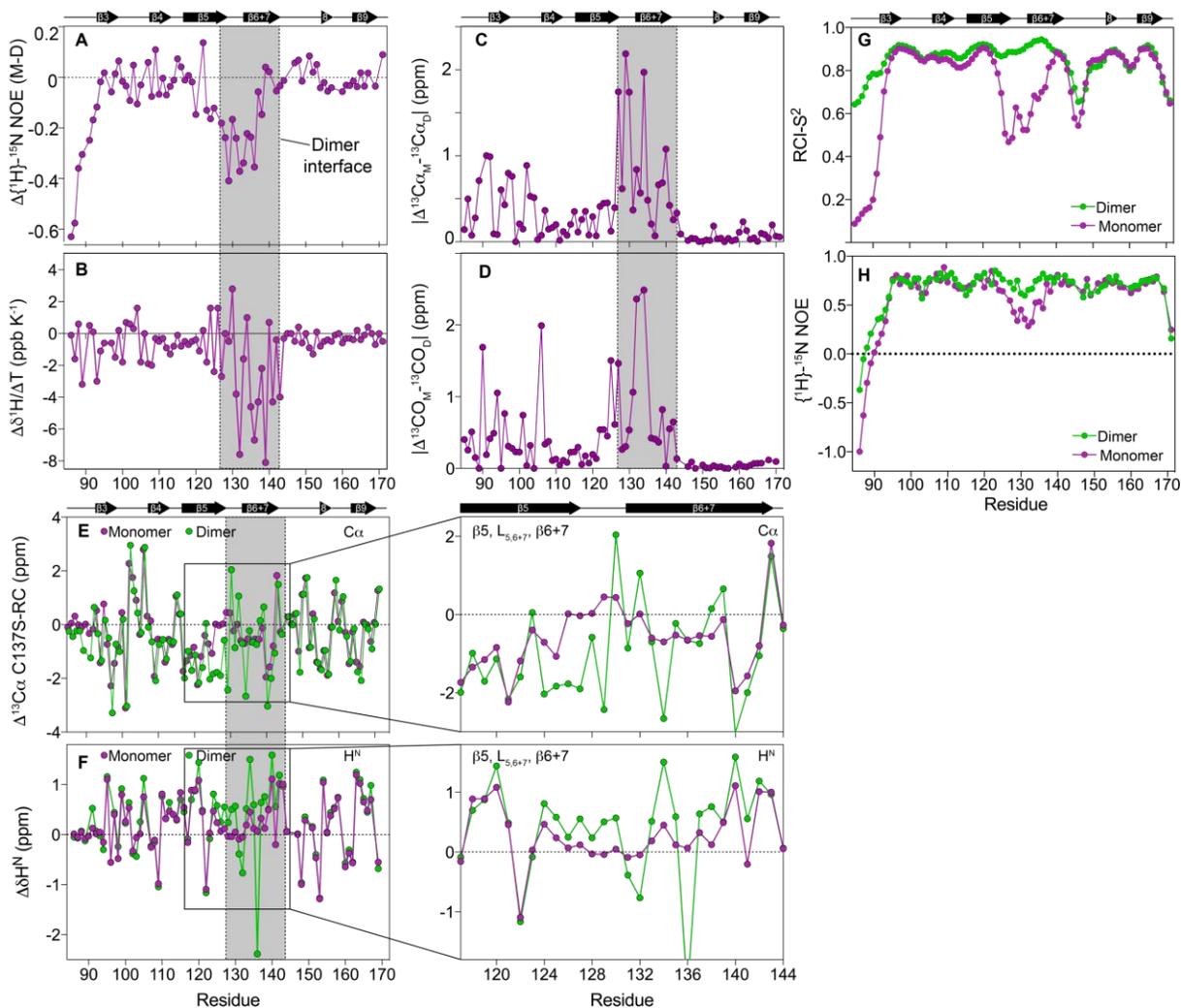


Fig. 16: Structural and dynamical comparison of the C137S monomer and dimer. (A) The difference in $\{^1\text{H}\}$ - ^{15}N NOE values between the C137S monomer and dimer (monomer-dimer). Negative values indicate enhanced backbone dynamics in the monomeric state. (B) The difference in ^1H chemical shift temperature coefficients ($\Delta\delta^1\text{H}/\Delta\text{T}$) for the C137S monomer and dimer (monomer-dimer). More negative values indicate a progressively lower likelihood that a given amide proton is involved in a hydrogen bond to other atoms in the protein. The region that comprises the dimer interface shows many residues with a lower likelihood of hydrogen bond formation, consistent with the less secondary structure. Residues G147-K171 show very little differences between the monomer and dimer. (C) and (D) show the absolute value of the difference in $^{13}\text{C}\alpha$ (C) and ^{13}CO (D) chemical shifts for the C137S monomer and dimer. From G147 to K171, there are essentially no structural differences between the C137S monomer and dimer, indicating that the dimeric conformation is preserved in this region. Notably, these residues yield small ^{15}N $|\Delta\omega|$ values in the CPMG RD experiments at pH 7. (E, F) Secondary chemical shifts ($\Delta\delta$), or the difference between the measured chemical shift for a given nucleus within a residue and the same nucleus in a random coil conformation, are shown for $^{13}\text{C}\alpha$ and $^1\text{H}^{\text{N}}$. A value of zero indicates a random coil conformation, and positive or negative values indicate secondary structure, with the sign and magnitude depending on the nucleus and secondary structure element. Secondary structure motifs of the cHSP27 dimer, as determined from PDB 4mjh, are shown above for reference. (G) RCI-determined N-H S^2 values for the C137S dimer (green) and monomer

(purple). (H) $\{^1\text{H}\}$ - ^{15}N NOE values for the C137S monomer (purple) and dimer (green). The data in Figure 15 b iii are truncated at 0, whereas here the full data are shown.

3.4 Discussion

The function and monomerization of the molecular chaperone HSP27 is regulated by its redox state through an inter-dimer disulfide bond (Chalova et al. 2014a, Diaz-Latoud et al. 2005, Pasupuleti et al. 2010). Here, I identified the important role of monomerization in regulating the chaperone activity of HSP27 (Fig. 1, Fig. 2, Fig. 5) and determined the structural basis for this regulation (Fig. 8, Fig. 11, Fig. 15). Remarkably, the oligomeric distribution and ps-ns dynamics within the flexible CTR in the context of full length HSP27, and both the structure ps-ns dynamics of cHSP27 dimers were invariant to formation of the disulfide bond in the dimer interface. Reduction of full-length HSP27 and cHSP27 leads predominantly to the release of the free monomers (Fig. 1b). I observe enhanced chaperone activity *in vitro* against multiple aggregating proteins when I increase the quantity of free monomers (Fig. 1, Fig. 2, Fig. 5), which was achieved by manipulating the full-length HSP27 concentration to favor monomer release and by comparing monomeric and dimeric ACDs of HSP27. I note that the core domains do not entirely recapitulate the activity of the full-length chaperone. In the case of insulin, the activity of the full length chaperone was significantly greater than the ACDs, likely suggesting an important role for the NTD of the protein in chaperone activity (Jovcevski et al. 2015), whereas against αLac , both monomeric and dimeric ACDs were more efficient chaperones than the full-length protein. Although in general, the specific activities of HSP27 and cHSP27 vary depending on the specific choice of aggregating protein (Mymrikov et al. 2016), from direct comparisons between the monomeric ACD mutant H124K/C137S and the dimeric ACD (C137S and wild-type) (Fig. 5), I conclude that the monomeric ACD is a significantly more effective chaperone than the dimer against the aggregating proteins presented in this work.

Using a combination of CPMG RD and high pressure NMR, I established that the monomeric ACD of HSP27 partially unfolds upon dissociation (Fig. 10, Fig. 11, Fig. 12, Fig. 13), such that the region responsible for the rigid interface in the dimer becomes highly dynamic. I can attribute the difference in activity between the monomeric and dimeric forms to this structural rearrangement. The link between heightened disorder and enhanced activity have been previously observed for α B-crystallin (ABC). Under acidic conditions (pH 2.5), ABC exists as a predominantly unfolded monomer, where it can prevent the aggregation of β 2-microglobulin (Raman et al. 2005). Moreover, unstructured peptides from both ABC and α A-crystallin (AAC), including an 8-mer comprising the β 6+7 strand, are able to prevent substrate aggregation (Ghosh et al. 2005) and can interact with actin (Ghosh et al. 2007). Consistent with these observations, both the acid-induced unfolding of HdeA and HdeB (Tapley et al. 2009) and the oxidation dependent unfolding of HSP33 (Groitel et al. 2016) heighten the anti-aggregation activity of these molecular chaperones. Similarly, the homodimeric chaperone CesAB exists in a molten globule-like state with only residual helical structure (Chen et al. 2011), but undergoes a disorder-to-order transition upon binding to its substrate (Chen et al. 2013). More generally, the plastic nature of intrinsically disordered proteins (IDPs) is thought to aid their ability to bind a wide variety of partners (Demarest et al. 2002, Ferreon et al. 2013, Shammass et al. 2014, Sugase et al. 2007) via specific, yet transient interactions (Wright & Dyson 2014). It is interesting to speculate that the same mechanism for rapid, promiscuous recognition of binding partners by IDPs is responsible for the heightened activity of partially unfolded chaperones. Recent reports have indicated that HSP27 interacts with substrates α S and Tau in a dynamic, transient manner (Baughman et al. 2018). Interestingly, many of the residues that are unfolded in the HSP27 monomer are charged or polar (Fig. 15d), suggesting that electrostatics may play a significant role in substrate-recognition (Fu et al. 2014), in addition to hydrophobic interactions (Carver & Lindner 1998). This situation is similar to the electrostatically mediated binding for the molecular chaperone Spy (Koldewey et al. 2016,

Salmon et al. 2016), but contrasts with the chaperones that predominantly recognize exposed hydrophobic residues such as SecB (Huang et al. 2016), trigger factor (Saio et al. 2014), DnaK/HSP70 (Rosenzweig et al. 2017), and ClpB/HSP100 (Rosenzweig et al. 2015).

My CPMG RD data inform on a specific mechanism for monomer release. The loop L_{5,6+7} located at the dimer interface of HSP27 undergoes redox-independent motions on the ms timescale, and exists in a minor conformation that is unfolded (Fig. 11). When the disulfide bond is present, the local unfolding does not propagate further. However, in the reduced form and C137S variant, the disordering process extends, on the same timescale, into both the end of the β 5 and beginning of the β 6+7 strands, effectively destabilizing the interface and facilitating monomer release (Fig. 15). Given the conservation of sHSP residues in L_{5,6+7} (Figure 15, Fig. 17) and the previously observed ms motions in this region of cABC (Mainz et al. 2012, Rajagopal et al. 2015b), transient unfolding of L_{5,6+7} and the adjacent strands upon monomerization is likely a common property of mammalian sHSPs.

I analyzed the positions of 28 mutations in HSP27 that cause either CMT or dHMN (Fig. 15a), including the 17 missense mutations that reside in the ACD (Nefedova et al. 2015). My structural and dynamical analysis of the cHSP27 monomer reveals that 14 of the 17 mutations in the ACD are located in or near the disordered region within L_{5,6+7} and the β 5 and β 6+7 strands. As previously noted, a number of these mutations cluster to the ACD dimer interface. My NMR data further reveal that the mutations that occur in regions beyond the dimer interface, predominantly fall in regions that are highly disordered in the monomer, suggesting that the behavior of the monomer is important for understanding the molecular bases of CMT disease and dHMN (Nefedova et al. 2015). Such significance could manifest perhaps in terms of altered activity, abundance, or through causing uncontrolled self-aggregation (Fig. 5).

While certain mutations decrease chaperone activity, some disease-related HSP27 variants that are more monomeric (e.g. R127W, S135F) exhibit significantly elevated chaperone

activity both *in vitro* and *in vivo* (Almeida-Souza et al. 2010, 2011), with increased affinity for substrate proteins (e.g. R140 mutation) (Koteiche & Mchaourab 2006). Conversely, disease-related mutants that did not impact monomerization have shown either no change (T151I) or a decrease (P182S/L) in activity (Almeida-Souza et al. 2010, 2011; Chalova et al. 2014b). Recent *in vitro* work on CMT-related HSP27 variants has demonstrated that ACD mutations can both enhance the dissociation of oligomers into smaller species and increase the overall size of the oligomers (Elliott et al. 2013, Nefedova et al. 2013, Weeks et al. 2018). As the concentration of HSP27 in healthy human cells under basal conditions should be in the high nM to low μ M range (Nagaraj et al. 2014), free monomers would therefore be readily populated. These observations suggest that the strength of the monomer/dimer interface, and the monomer/dimer/oligomer equilibria are important factors for understanding neuropathies associated with variants of HSP27.

In light of my findings, I hypothesize that partial unfolding of sHSP ACDs upon monomer release may be a general feature of this class of chaperone. A recent study of cABC showed that the chemical shift changes upon monomer formation are larger for residues located at the dimer interface (Rajagopal et al. 2015b). I analyzed the data and found a strong correlation between ^{15}N chemical shift changes in cABC upon monomer formation with those expected for the formation of a random coil (Fig. 17). More generally, the dimeric structure and sequence composition of residues at the ACD dimer interface are highly conserved in the mammalian sHSPs HSP27, ABC, AAC, and HSP20 (Fig. 15e, Fig. 17). These results suggest that partial unfolding of monomers upon dissociation is a common property of human sHSPs and that the dimeric building block of sHSP oligomers is assembled first through partly unfolded monomers. Odd-numbered sHSP oligomers (Baldwin et al. 2011), as encountered in both human sHSPs ABC and HSP27, will have at least one monomer without a complete dimer interface, indicating that the unstructured monomer can also exist within larger oligomers. Interestingly, for the related ABC, I observed that the dimeric form was more chaperone-active than the monomer, particularly for an amyloidogenic

substrate (Hochberg et al. 2014). These differences between the two sHSPs could reflect their contrasting substrate profiles (Mymrikov et al. 2016), or multiple binding modes (Mainz et al. 2015).

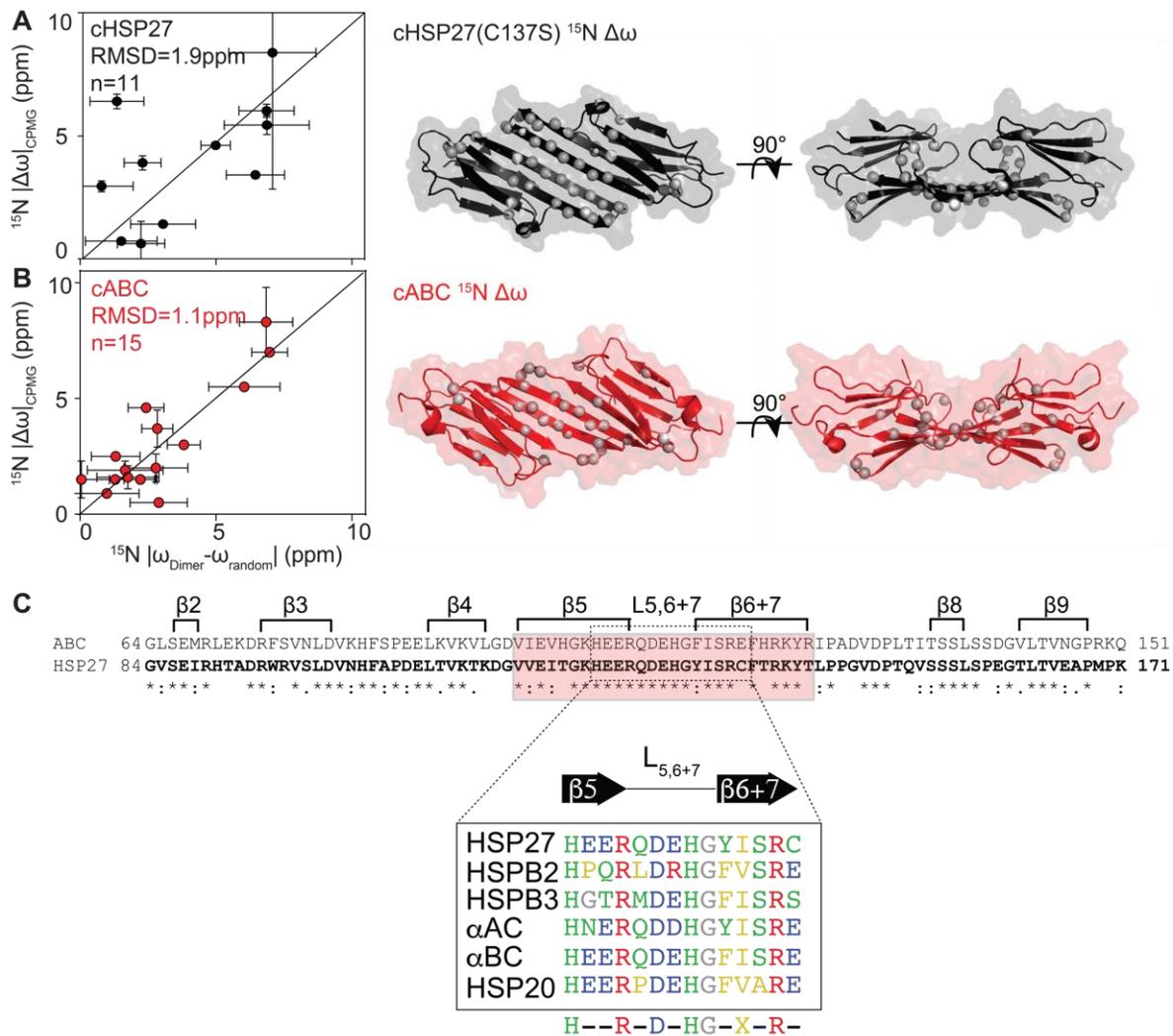


Fig. 17. The human paralog α B-crystallin (ABC) likely also adopts a partially disordered conformation in its monomeric form. (A/B) ^{15}N chemical shift changes that were measured by CPMG (C137S and cABC) or directly from a pH titration (cABC) are correlated with expected ^{15}N chemical shift changes upon the formation of random coil conformations. The nCIDP database was used to calculate sequence-corrected ^{15}N random coil chemical shifts, shown here as the mean \pm 1 SD. The agreement between experimentally measured ^{15}N chemical shift changes upon monomerization with chemical shift

changes expected for random coil conformations is reasonable (C137S RMSD: 1.6 ppm, n = 14; cABC RMSD: 1.1 ppm, n = 15), suggesting the change in conformation reflects the formation of a disordered monomer for both cHSP27 (A) and α B-crystallin (B). The locations of the residues that show significant relaxation dispersion curves are indicated (cHSP27 PDB 4mjh, black; cABC PDB 4m5s, red). Data from α B-crystallin were extracted from Rajagopal *et al.* and originate from ^{15}N CPMG RD data (residues 102, 106, 109, 117, 118, and 136) and ^{15}N chemical shift changes measured in a pH titration (residues 104, 108, 113, 114, 115, 116, 120, 121, and 126). (C) Amino acid sequence comparison between the ACDs of α B-crystallin and HSP27. The β -strands are indicated, and residues in the β 5 and β 6+7 strands are indicated (box). There is significant sequence conservation in this area. The inset depicts the sequence alignment of human sHSPs HSPB1-HSPB6 (i.e. HSP27, HSPB2, HSPB3, α A-crystallin, α B-crystallin, and HSP20) in the region β 5-L_{5,6+7}- β 6+7.

In the context of the isolated ACD, my data suggest that increased disorder in the HSP27 monomer renders it a more potent chaperone *in vitro*. In addition to the partially unfolded ACD, full-length HSP27 contains a disordered 80-residue NTD (McDonald *et al.* 2012) and a highly flexible CTR comprising 28 residues (Alderson *et al.* 2017, Carver *et al.* 1995). More generally, these disordered regions are important for stabilizing the oligomeric forms (Carver *et al.* 2017) and also contribute to chaperone activity. The inherent plasticity in these disordered regions, combined with the disorder in the monomeric ACD would, in principle, allow for the sampling of a wide range of conformational space and thereby facilitate its ability to interact with a diverse set of misfolded target proteins. As the monomer is itself prone to aggregation (Fig. 5d), I speculate that the aggregation-prone contacts in HSP27 are largely responsible for detecting misfolded proteins (Bardwell & Jakob 2012, Carver *et al.* 2002, Jacobs *et al.* 2016, Meehan *et al.* 2007). In the context of the cell, it would seem undesirable to have high concentrations of aggregation-prone monomers, making it advantageous to store them in oligomers that are sensitive to environmental conditions. By holding monomers in this 'storage' form, the population of the active but unstable monomeric form is kept both transiently low and highly available (Hilton *et al.* 2013, Van Montfort *et al.* 2001).

In vivo, the concentration of HSP27 can be estimated in the low μM range, e.g. $\sim 10 \mu\text{M}$ in HeLa cells (Nagaraj *et al.* 2014). As HeLa cells are known to express constitutively high levels of

HSP27 ([Paul et al. 2002](#)), the concentration in non-cancerous human cells should be significantly less under basal conditions. Therefore, in healthy human cells, a significant fraction of HSP27 monomer will exist in equilibrium with dimers and large oligomers, as the fractional population of monomeric HSP27 will increase with decreasing total concentration. I anticipate that the chaperone will retain its quaternary dynamics and populate sub-oligomeric states (monomers, dimers, etc.) even under conditions of molecular crowding.

If the reduced, monomeric form of HSP27 is more potent than its oxidized counterpart, then how can the chaperone be a redox sensor inside living cells? I speculate that the monomeric form of HSP27 acts as a homeostatic chaperone under basal conditions, in which this highly chaperone-active state can efficiently recognize misfolded proteins and prevent their aggregation ([Fig. 1b](#), [Fig. 2](#)). Upon exposure to oxidative stress, the cell rapidly upregulates the expression of HSP27, increasing the total concentration of HSP27 and thereby shifting the equilibrium toward large oligomers. However, the binding of monomeric HSP27 to client substrate proteins would prevent disulfide formation and deplete the amount of free HSP27, and lower the effective concentration of the chaperone. Finally, the lone cysteine residue in HSP27 has been shown to be essential for its anti-apoptotic interaction with cytochrome C ([Bruey et al. 2000](#)), suggesting that redox changes may play a role in the ability of HSP27 to influence cell death pathways.

In conclusion, my analysis combining CPMG RD and high-pressure NMR with chaperone and aggregation assays provides the first structural characterization of the sparsely populated and experimentally elusive monomeric form of HSP27. I demonstrate through redox- and concentration-dependent aggregation assays that the monomeric state is particularly active *in vitro*. Part of the region that forms the rigid interface in the dimer unfolds and becomes highly dynamic in the monomer, with the additional structural plasticity in the monomer rendering it a more effective chaperone. The monomer itself is more aggregation-prone, whereby off-pathway self-assembly culminates in uncontrolled aggregation. A survey of extant sHSP structures

suggests that the exposure of the disordered, yet functional dimer interface, either in the context of oligomers or free in solution, is widespread across all sHSPs in all kingdoms of life.

3.5 Materials and Methods

3.5.1 Protein expression and purification

HSP27

All media for growth of cells containing pET(HSP27) plasmids contained 100 mg/L of ampicillin (Amp). Glycerol stocks of Amp-resistant (Amp^R) pET(HSP27) plasmids were used to inoculate 5 mL cultures for expression of full-length HSP27. Following growth at 37 °C for 6 hours, the 5 mL cultures were transferred to 100 mL (LB medium) or 50 mL (M9 minimal medium) cultures and grown overnight at 37 °C, which were then used to inoculate 1 L (LB medium) or 500 mL (M9 minimal medium) cultures. When the optical density at 600 nm (OD₆₀₀) of these cultures reached between 0.6 and 0.8 units, IPTG was added to a final concentration of 1 mM and protein expression ensued for 3 hours at 37 °C. Cells were pelleted and frozen at -80 °C until use. For preparation of uniformly-¹³C, ¹⁵N-labeled ([U-¹³C,¹⁵N]-) HSP27, the M9 minimal medium contained 2 g L⁻¹ of [U-¹³C]-glucose and 1 g L⁻¹ of ¹⁵NH₄Cl.

HSP27 was purified in a similar manner as described previously (Muranova et al. 2015) using anion exchange chromatography (AEX) with HiTrapQ HP columns (GE Healthcare). The column was equilibrated in 20 mM Tris-HCl, 1 mM EDTA, pH 7 (AEX buffer A), after which the lysate was applied. The column was washed with 5% of IEX buffer B (AEX buffer A with 1 M NaCl), and a linear gradient of 5-40% AEX Buffer B followed. HSP27 eluted around 15% AEX Buffer B (150 mM NaCl), after which all HSP27-containing fractions were pooled and concentrated for size exclusion chromatography (SEC). A Superdex S200 26/60 column (GE Healthcare) equilibrated in 20 mM Tris-HCl, 150 mM NaCl, 2 mM EDTA at pH 7 was used to separate HSP27 from the remaining contaminants. HSP27 eluted off the Superdex S200 26/60 near 200 mL, or 0.63 column volumes. After pooling HSP27-containing fractions after SEC, the samples were further purified with a second AEX step using a HiTrap Capto Q ImpRes column (GE Healthcare), with AEX

Buffers A and B as described previously. Purified HSP27 was then buffer exchanged using 10K MWCO Amicon spin filters into 30 mM NaH₂PO₄, 2 mM EDTA, 2 mM NaN₃, pH 7.0 with 6% D₂O (NMR buffer), with an additional 100 mM NaCl for the oxidized state. To study the reduced state, 5 mM β-mercaptoethanol (BME) was added (Fig. 1, S1). Oxidation was confirmed by both SDS-PAGE and native MS under non-reducing conditions.

cHSP27

DNA encoding the residues 84–171 of HSP27 (cHSP27) was inserted into kanamycin-resistant (Kan^R) pET28-b plasmids, which contained an N-terminal hexahistidine (His₆)-tag followed by a TEV protease recognition site (Hochberg et al. 2014). The residual glycine that remains after TEV protease cleavage corresponds to G84 in the HSP27 amino acid sequence. All media for growth of cells containing pET28-b(cHSP27) plasmids contained 30 μg/mL of Kan. cHSP27 cultures were grown as described for full length HSP27. Upon inoculation of the 1 L (LB medium) or 500 mL (M9 minimal medium) cultures, A₆₀₀ was allowed to reach between 0.6 and 0.8 units, upon which IPTG was added to a final concentration of 1 mM and protein expression continued for 3 hours at 37 °C. Cells were pelleted and frozen at -80 °C until use. For preparation of isotopically labeled cHSP27, cells were grown in H₂O-based M9 minimal medium as above, supplemented with 4 g L⁻¹ of unlabeled glucose, 2 g L⁻¹ of [*U*-¹³C]-glucose, 1 g L⁻¹ of ¹⁵NH₄Cl, or grown in 99.9% D₂O, respectively, for [*U*-¹⁵N]-, [*U*-¹³C,¹⁵N]-, or fractionally deuterated (frac²H), [*U*-¹⁵N]-cHSP27.

cHSP27 was purified in a similar manner as described previously (Hochberg et al. 2014), with initial separation on a Ni²⁺ column that was equilibrated in 20 mM Tris, 150 mM NaCl, 30 mM imidazole, pH 8.0 (IMAC Buffer A). Elution of bound protein from the column was achieved using IMAC Buffer A with 300 mM imidazole, and cHSP27-containing fractions were pooled and incubated with His₆-tagged TEV protease in IMAC Buffer A at room temperature overnight with 5 mM BME to cleave the His₆-tag from cHSP27. Cleaved cHSP27 was then separated from His₆-

tagged TEV protease, the cleaved His₆-tag, and other contaminants using reverse Ni²⁺ affinity separation and then subjected to SEC on a Superdex S75 26/60 column (GE Healthcare), eluting at 170-180 mL, or 0.53-0.56 column volumes. The final SEC buffer contained 20 mM Tris-HCl, 150 mM NaCl, and 5 mM BME at pH 7.0. Using 3K MWCO Amicon spin filters, protein samples were buffer exchanged into and concentrated into NMR buffer for the oxidized state at pH 7 unless otherwise specified. 5 mM BME was added when studying the reduced state.

3.5.2 Site-directed mutagenesis in cHSP27

Primers were designed to encode the single-point mutations H124K and C137S in cHSP27. Site-directed mutagenesis was performed using the QuikChange II Mutagenesis Kit (Agilent), and the correct mutations were verified by DNA sequencing. Protein expression and purification was carried out as above. The final yield of the double mutant, cHSP27(H124K/C137S) (~10 mg/L of *E. coli*) was significantly lower than either WT cHSP27 or C137S (30-40 mg/L of *E. coli*).

3.5.3 Aggregation of substrate proteins and chaperone activity assays

All chaperone activity assays were completed using a 96-well plate and a FLUOstar Omega Microplate Reader or Tecan Infinite M200 PRO plate reader. Chaperone activity was defined as one minus the ratio of signal observed in the presence of chaperone to that in the absence of chaperone, e.g. values of 1 and 0 respectively indicate complete protection against aggregation and incapability to prevent aggregation. To avoid selecting a single time point or averaging a few time points for the analysis of chaperone activity, I calculated the integral under the observed curves of absorbance or fluorescence versus time. The integral under the curve was determined for each of the replicate wells for a given substrate and substrate/chaperone combination, and the reported chaperone activity values reflect the average and one standard deviation. The y-axes for plots of aggregation versus time were normalized on a scale of 0–1, with 1 set to the maximum

value obtained for substrate alone. The x axes were likewise normalized on a scale of 0–1 by dividing time by a factor T_i , which varied for each substrate.

Porcine heart CS (Sigma-Aldrich) was buffer exchanged into NMR buffer supplemented with 100 mM NaCl. The aggregation of CS is initiated at high temperatures and does not require the addition of a chemical denaturant (e.g. reducing agents), which enables a direct comparison of the chaperone activity of reduced and oxidized HSP27. However, CS aggregates faster under reducing conditions (Fig. 2), complicating interpretation. The aggregation of 10 μ M CS was monitored at 43 °C by following light scattering at 340 nm as a function of time. HSP27 was added to a final concentration of 0.5 μ M, a concentration low enough for a substantial concentration of monomers to be expected (Jovceviski et al. 2015). The assays were conducted in the absence or presence of 5 mM BME. Assays were completed in duplicate and the mean \pm one standard deviation is reported. The light scattering traces in the presence of reduced and oxidized HSP27 shown in Fig. 1 are normalized to the respective aggregation of reduced and oxidized CS alone, and plotted on the same axis.

Porcine heart MDH and rabbit muscle GAPDH (Sigma-Aldrich) were respectively buffer exchanged into or dissolved in NMR buffer, and aggregation assays were conducted at 40 °C at 0.2 μ M (MDH) or 2 μ M (GAPDH) final concentration. Aggregation was monitored by measuring light scattering at 340 nm as a function of time, in the presence or absence of 5 mM BME. The redox state of MDH and GAPDH did not affect aggregation (Fig. 2), enabling a direct comparison between the chaperone activity of reduced and oxidized HSP27, which was added in final concentrations of 0.5, 2, or 20 μ M. Assays were conducted in triplicate and the mean \pm one standard deviation is reported.

Human α S was recombinantly expressed and purified as previously described (Cox et al. 2016). All samples were prepared in PBS with a final concentration of 200 μ M α S, 4 μ M ThT, and

10 μM αS fibril seeds in the presence or absence of 1 mM DTT. The αS fibril seeds were prepared through previous ThT-based fibrillation assays and confirmed by EM. Amyloid fibril formation in the seeded assays was determined by ThT fluorescence with excitation at 440 nm and emission at 485 nm at 37 °C over one week. All fluorescence experiments were performed using a Corning 3881 96-well plate.

Calcium-depleted bovine $\alpha\text{-Lac}$ and human insulin (Sigma-Aldrich) were dissolved in NMR buffer, and aggregation was initiated by the addition of 2 mM DTT followed by incubation at high temperatures. The aggregation of αLac (300 μM) at 37 °C and insulin (80 μM) at 40 °C was initiated upon the addition of 1 mM DTT, and aggregation was monitored by measuring light scattering at 340 nm as a function of time. cHSP27 variants were added to a final concentration of 70 μM (αLac) or 40 μM (insulin), a concentration where C137S and reduced cHSP27 are predominantly dimers and H124K/C137S a monomer. Assays were conducted in triplicate and the mean \pm one standard deviation is reported.

3.5.4 Aggregation of H124K/C137S

The aggregation of H124K/C137S was monitored by following the absorbance at 340 nm at 37 °C in NMR buffer. The total volume was 100 μL and the protein concentration was either 200 μM or 800 μM . For comparison, C137S was prepared at 800 μM and subjected to identical treatment. Notably, the sequence-based aggregation propensity predictors Tango ([Linding et al. 2004](#)) and Zyggregator ([Tartaglia & Vendruscolo 2008](#)) do not predict any notable change owing to the choice of mutation. All concentrations had either three or six replicates, and the mean \pm SD is reported.

3.5.5 Native MS

For native MS data acquisition, nanoelectrospray ionization (nESI) experiments were executed according to previously published protocols using instrumental settings optimized to transmit intact protein complexes (Hernández & Robinson 2007). A 25 μM (monomer concentration) sample of HSP27, which was purified in the absence of reducing agent, was prepared in 200 mM ammonium acetate (pH 6.9) in the presence and absence of 250 μM DTT. cHSP27 samples (5 μM) were prepared in the same buffer with and without 250 μM DTT. Data were collected on Q-ToF 2 (Micromass/Waters Corp.) and a Synapt G1 (Waters Corp.) instruments.

3.5.6 Thermal denaturation of cHSP27

A nanoDSF instrument (NanoTemper) was used to monitor the intrinsic fluorescence of cHSP27 and C137S as a function of temperature. Capillaries contained ~ 10 μL of cHSP27 that had been prepared in NMR buffer without (oxidized) or with 5 mM BME (reduced). C137S was prepared in NMR buffer without BME. The concentrations of cHSP27 and C137S were 1000, 100, 10, and 1 μM , with three replicates recorded at each concentration. The initial temperature was 20 $^{\circ}\text{C}$ and was set to increase by 1 $^{\circ}\text{C}$ per minute. Fluorescence readings were recorded at 330 and 350 nm, and the melting temperature (T_m) determined through a plot of the ratio of fluorescence (350/330 nm) versus temperature. The T_m is at the inflection point, and the reported values reflect the average \pm one standard deviation from three replicates.

3.5.7 NMR spectroscopy

Backbone and side-chain resonance assignments for dimeric cHSP27

All NMR spectroscopy experiments for resonance assignments of cHSP27 and HSP27 at ambient pressure were recorded on a 14.1 T Varian Inova spectrometer equipped with a 5 mm z-axis gradient, triple resonance room temperature probe. 2D ^1H - ^{15}N sensitivity-enhanced HSQC spectra

([Kay et al. 1992](#)) at 14.1 T were typically acquired with ^1H (^{15}N) 512 (64) complex points, spectral widths of 8012 Hz (1800 Hz), maximum acquisition times of 64 ms (35.6 ms), an inter-scan delay of 1 s, and four scans per FID for a total acquisition time of 19 minutes. Uniformly ^{15}N - and ^{13}C -labeled cHSP27 ($[U\text{-}^{13}\text{C},^{15}\text{N}]\text{-cHSP27}$) was prepared at a final concentration of 1 mM in NMR buffer. Standard backbone and side-chain assignment experiments ([Sattler et al. 1999](#)) were collected, specifically:

HNCO: 766/50/25 complex points and 85.1/39.8/23.6 ms maximum acquisition times for the H/C/N dimensions respectively, with 8 scans per FID for a total acquisition time of 12.76 hours.

HN(CA)CO: 1024/40/30 complex points and 85.1/10.6/13.3 ms maximum acquisition times for the H/C/N dimensions, with 16 scans per FID for a total acquisition time of 23.66 hours

HNCA: 766/40/30 complex points and 85.1/14.8/28.4 ms maximum acquisition times for the H/C/N dimensions, with 16 scans per FID for a total acquisition time of 24.07 hours.

HN(CO)CA: 1024/50/20 complex points sampled at 25% sparsity, 85.1/18.5/18.9 ms maximum acquisition times for the H/C/N dimensions, with 32 scans per FID for a total acquisition time 8.8 hours.

C(CO)NH: 577/40/20 complex points and 64/4.4/18.9 ms maximum acquisition times for the H/C/N dimensions, with 64 scans per FID for a total acquisition time of 61.86 hours.

When NUS was employed, an exponentially-weighted sampling scheme was employed in the indirect dimensions and time-domain data were reconstructed with MddNMR ([Kazimierczuk & Orekhov 2011](#)). All 3D NMR spectra at ambient pressure were acquired at 25 °C, processed with NMRPipe ([Delaglio et al. 1995](#)), and visualized with NMRFAM-Sparky ([Lee et al. 2015](#)). The resultant $^1\text{H}^{\text{N}}$, ^{15}N , ^{13}CO , $^{13}\text{C}\alpha$, and $^{13}\text{C}\beta$ chemical shifts were analyzed with TALOS-N ([Shen &](#)

[Bax 2013](#)) and RCI ([Berjanskii & Wishart 2006](#)) to respectively estimate the secondary structure and N-H order parameters.

Backbone resonance assignments for monomeric C137S

Resonance assignments for monomeric C137S were obtained on a 14.1 T Bruker Avance-III spectrometer equipped with a cryogenic probe. A [U - ^2H , ^{13}C , ^{15}N]-labelled sample at pH 4.1, 2 mM sodium phosphate, 25 °C was prepared at 100 μM . 3D HNCA and HNCO spectra were acquired under these conditions. The HNCO spectrum was acquired with 10% NUS, an inter-scan delay of 1.5 s, 1024/170/140 complex points (H/C/N) with maximum acquisition times of 104/70/85.5 ms, and 4 scans per FID for a total acquisition time of 16 hours. For the HNCA spectrum, 1024/115/84 complex points were acquired for H/C/N with maximum acquisition times of 104/20/58 ms, with 8 scans per FID and an inter-scan delay of 1.4 s for a total acquisition time of 81 hours. A similar set of experiments was recorded on [^2H , ^{13}C , ^{15}N]-labelled C137S dimer at pH 7 in order to compare ^{13}C chemical shifts. The NUS spectra were reconstructed with SMILE ([Ying et al. 2016](#)).

Indirectly probing hydrogen bonds in oxidized cHSP27

^1H - ^{15}N HSQC spectra of [U - ^{15}N]-oxidized cHSP27 were recorded at 293, 295, 298, and 303 K, and the ^1H chemical shift temperature coefficients ($d\Delta^1\text{H}/dT$) were calculated. $d\Delta^1\text{H}/dT$ values that are more negative than -4.6 ppb/K are more likely to be solvent exposed and hydrogen bonded to water ([Baxter & Williamson 1997](#)). Residues with temperature coefficients more positive than -4.6 ppb/K are more likely to be involved in intra- or inter-protein hydrogen bonds. However, it should be noted that residues that are near aromatic rings can yield false positives with values less than -4.6 ppb/K ([Baxter & Williamson 1997](#)).

To provide an independent NMR dataset that also indirectly probes hydrogen bonds, I buffer exchanged a sample of 1 mM [U - ^{15}N]-oxidized cHSP27 into 99.9% D_2O and recorded a 2D ^1H - ^{15}N HSQC spectrum at 25 °C (19-minute experimental acquisition). The dead time was ca. 40

min for the buffer exchange process and the sample was kept at 4 °C during this time. Intra- and inter-cHSP27 hydrogen bonds involving amide protons were assessed by the presence or absence of signals in the D₂O ¹H-¹⁵N HSQC spectrum.

¹⁵N spin relaxation experiments (T_1 , T_2 , NOE) on cHSP27 and full-length HSP27

Standard pulse sequences to measure ¹⁵N heteronuclear nuclear Overhauser enhancements (hetNOE), longitudinal (T_1), and transverse (T_2) relaxation times were employed (Palmer 1993), using 14.1 T and 11.7 T Varian Inova spectrometers equipped with 5 mm z-axis gradient triple resonance, room temperature probes. All experiments at 14.1 T were recorded with 576 (32) complex points, 8992 Hz (1800 Hz) sweep widths, and 64 ms (18 ms) acquisition times, for ¹H (¹⁵N). All experiments at 11.7 T were recorded with 512 (32) complex points, 8000 Hz (1519 Hz) sweep widths, and 64 ms (21 ms) acquisition times, for ¹H (¹⁵N). T_1 and T_2 values are reported as their respective inverses, *i.e.* rates (R_1 and R_2), for convenience. The spectrometer temperature was calibrated with d₄-methanol. I recorded ¹⁵N R_1 (11.7, 14.1 T), R_2 (11.7, 14.1 T), and hetNOE (14.1 T only) values at 25 °C on a 1 mM sample of [*U*-¹³C,¹⁵N]-labeled oxidized cHSP27 in NMR buffer. Upon completion of these measurements, 5 mM β-mercaptoethanol (BME) was added to the NMR tube, and ¹⁵N R_1 , R_2 , and hetNOE values were recorded on reduced cHSP27 at 600 MHz only (14.1 T). Similarly, ¹⁵N R_1 , R_2 , and hetNOE experiments at 14.1 T were recorded on oxidized [*U*-¹³C,¹⁵N]-labeled full-length HSP27, after which 5 mM BME was added to obtain relaxation measurements on the reduced species. For data acquired exclusively at 14.1 T, reduced spectral density functions were calculated as described previously (Lefevre et al. 1996). For data acquired at multiple magnetic field strengths (oxidized cHSP27), a model-free analysis was conducted (see below). All data sets mentioned here were processed with NMRPipe (Delaglio et al. 1995), visualized with NMRFAM-Sparky (Lee et al. 2015), and peak shapes were fit and analyzed with FuDA (Vallurupalli et al. 2008). hetNOE experiments on the C137S monomer and

dimer were performed on a 14.1 T Bruker Avance-III spectrometer equipped with a cryogenically cooled probe using a previously published pulse sequence (Lakomek et al. 2012).

Model-free analysis of ^{15}N spin relaxation data from oxidized cHSP27

I analyzed the ^{15}N relaxation data from oxidized cHSP27 at 11.7 T (500 MHz) and 14.1 T (600 MHz) using Lipari-Szabo models (Clore et al. 1990, Lipari & Szabo 1982). Prior to fitting the relaxation data with various Lipari-Szabo models, I first determined the rotational diffusion behavior of cHSP27. Residues with evidence of enhanced ps–ns or μs -ms dynamics, as respectively determined by hetNOE values < 0.65 and R_2 values > 1 S.D. from the mean, were excluded from the diffusion tensor analysis according to Tjandra et al. (Tjandra et al. 1995). After removing dynamic residues, the remaining ^{15}N R_2/R_1 ratios were fit to isotropic ($D_{xx} = D_{yy} = D_{zz}$), axially symmetric ($D_{xx} = D_{yy} \neq D_{zz}$), and anisotropic ($D_{xx} \neq D_{yy} \neq D_{zz}$) diffusion tensors, which were statistically validated by comparing values of χ^2_{red} and by using F-tests. A crystal structure of cHSP27 (Hochberg et al. 2014) was rotated into the frame of the diagonalized diffusion tensor to determine the angle of the N-H bond vectors with respect to the principle component of the rotational diffusion tensor. This analysis indicated that axial symmetry provided the most appropriate diffusion model for cHSP27, which was used to fit ^{15}N R_2/R_1 ratios with ModelFree4.15 (Mandel et al. 1995) to determine a global correlation time (τ_c) of 13.65 ns. This value was consistent with a ~20 kDa dimer that exhibits slower rotational diffusion of the axis parallel to the principal component of the diffusion tensor ($D_{\parallel}/D_{\perp} = 1.52$). Values for τ_c and D_{\parallel}/D_{\perp} were fixed in the subsequent model-free analysis, in which the N–H bond length was set to 1.02 Å and the ^{15}N CSA was set to –160 ppm. Using the five aforementioned values of ^{15}N relaxation rates from two magnetic fields, I fit these data using ModelFree4.15 (Mandel et al. 1995) to four models: model 1 (optimized \mathbf{S}^2), model 2 (\mathbf{S}^2 and τ_e), model 3 (\mathbf{S}^2 and R_{ex}), and model 4 (\mathbf{S}^2 , τ_e , and R_{ex}). No further benefit was obtained when \mathbf{S}^2_f was included in the fitting, and thus was not required for this analysis.

Analysis of ¹⁵N CPMG RD data

¹⁵N CPMG relaxation dispersion (RD) experiments were recorded on 11.7 and 14.1 T NMR spectrometers at 25, 30, and 35 °C using standard, relaxation-compensated pulse sequences (Sauerwein & Hansen 2015). Each data set was recorded with ¹H (¹⁵N) 615 (35) complex points (512/29 at 11.7 T), sweep widths of 9615 Hz (1800 Hz) (8000 Hz/1519 Hz), acquisitions times of 64 ms (19 ms) (64 ms/19 ms), and an inter-scan delay of 3 seconds. Each experiment was recorded as a pseudo-3D spectrum with the third dimension encoded by the variable delay between pulses in the CPMG pulse train ($\nu_{\text{CPMG}} = 4\tau_{\text{CPMG}}^{-1}$). For these measurements, ²H, [^{U-¹⁵N}]-labeled samples of cHSP27 (1 mM) or C137S (1 mM and 0.3 mM) were prepared. The experiments contained a fixed constant time of 39 ms for the CPMG period and 20 values of ν_{CPMG} ranging from 54 Hz to 950 Hz. Four scans (1 mM) or eight scans (0.3 mM) per FID were recorded for a total acquisition time of ~10 hours (1 mM) or ~20 hours (0.3 mM). Peak shapes were fit with FuDA (Vallurupalli et al. 2008) to extract peak intensities, which were then converted into $R_{2,\text{eff}}$ values using the following relation: $R_{2,\text{eff}}(\nu_{\text{CPMG}}) = -\ln(I_{\text{CPMG}}/I_0) * 1/T_{\text{relax}}$, where I_{CPMG} is the intensity of a peak at ν_{CPMG} , T_{relax} is the constant relaxation delay of 39 ms that was absent in the reference spectrum, and I_0 is the intensity of a peak in the reference spectrum. Two duplicate ν_{CPMG} points were recorded in each dispersion data set for error analysis, and uncertainties in $R_{2,\text{eff}}$ were calculated using the standard deviation of peak intensities from such duplicate measurements. From plots of $R_{2,\text{eff}}$ as a function of ν_{CPMG} , R_{ex} was estimated by taking the difference of $R_2(54 \text{ Hz})$ and $R_2(950 \text{ Hz})$. The program CATIA (Vallurupalli et al. 2008) was used for analysis, which accounts for all relevant spin physics including imperfect 180° pulses and differential relaxation between spin states, factors that are not accounted for when using closed form analytical solutions (Baldwin 2014). For each combination of concentration and temperature, data were analyzed over multiple field strengths for all residues where $R_{\text{ex}} > 2 \text{ s}^{-1}$. All residues were assumed to experience a ‘two-state’ equilibrium between a majorly populated ‘ground’ state

and a sparsely populated 'excited' conformational state, and so the same interconversion rates were applied to all residues during the analysis. Uncertainties in fitted parameters were estimated by a boot-strapping procedure where synthetic datasets were created by sampling residues with replacement, re-analyzing the new dataset and storing the result. The distribution of parameters achieved from 1,000 such operations provided a measure of experimental uncertainty.

Twenty-six residues with $R_{ex} > 2 \text{ s}^{-1}$ at 600 MHz were selected for further analysis, and their ^{15}N CPMG RD data at 14.1 T and 11.7 T were globally fit to a model of two-site chemical exchange using the program CATIA (Vallurupalli et al. 2008). This analysis enabled extraction of k_{ex} , p_E , and $|\Delta\omega|$ values. The global fit from all 25 residues yielded a reduced χ^2 of 1.46. These residues comprised V101, D107, T110, D115, G116, V117, T121, G122, H124, E125, E126, R127, D129, E130, G132, Y133, I134, S135, R136, S137, F138, T139, R140, K141, and T143. The resultant ^{15}N $|\Delta\omega|_{\text{CPMG}}$ values from C137S were compared to ^{15}N chemical shift changes between the folded C137S dimer and random coil conformations (^{15}N $|\Delta\omega|_{\text{RC}}$). The four random coil ^{15}N chemical shift databases within the neighbor-corrected intrinsically disordered library (ncIDP (Tamiola et al. 2010)) were employed, and the average ^{15}N random coil chemical shift and standard deviation was used as the ^{15}N $|\Delta\omega|_{\text{RC}}$, which was then correlated with the measured values of ^{15}N $|\Delta\omega|_{\text{CPMG}}$.

For oxidized cHSP27, residues D129, E130, R136, C137 and F138 showed significant variation in $R_{2,eff}$ with ν_{CPMG} ($R_{ex} > 2\text{s}^{-1}$). Analysis assuming each residue had independent exchange rates revealed that D129 and E130 clustered with a $k_{ex} \sim 1250\text{s}^{-1}$, whereas R136, C137 and F138 formed a cluster with $k_{ex} > 3000\text{s}^{-1}$. This distinction was apparent from the raw data as CPMG RD curves from the latter group showed variation in $R_{2,eff}$ with ν_{CPMG} at high pulse frequencies, whereas those from the first group were effectively in the fast exchange limit at much lower CPMG frequencies. The reduced chi squared value, χ^2_{red} of the first group was 1.01, and that of the second group was 1.02.

For reduced cHSP27, CPMG RD data were acquired at 14.1 T exclusively. Residues with $R_{\text{ex}} > 2 \text{ s}^{-1}$ were globally fit to a model of two-site chemical exchange, assuming a monomer-dimer exchange event, as above. Twenty-three residues were included in the fit, comprising V101, D107, T110, D115, G116, V118, T121, G122, K123, H124, E125, E126, R127, D129, E130, Y133, I134, S135, F138, T139, R140, K141, and T143. Residues R136 and C137 were not analyzed because these resonances were too broad to provide reliable measurements. The χ^2_{red} value from the global fit at one magnetic field strength was 1.45, with k_{ex} of $\sim 1500 \text{ s}^{-1}$ and p_E of $\sim 2\%$.

Determination of the dimerization dissociation constant (K_d) from ^{15}N CPMG RD data

The CPMG data were analyzed according to a two-state equilibrium scheme where $G \xrightleftharpoons[k_{\text{eg}}]{k_{\text{ge}}} E$.

As described in the text, k_{eg} was found to be concentration dependent, allowing identification of the minor state E to be monomeric, suggesting the equilibrium has the form $A_2 \xrightleftharpoons[k_{\text{on}}]{k_{\text{off}}} 2A$. The two-state rate constants derived from CPMG measurements, k_{eg} and k_{ge} , were converted to K_d , k_{on} and k_{off} measurements. The K_d values obtained from CPMG analysis from C137S using data acquired at 1 mM and 0.3 mM were highly similar, supporting the identification of the equilibrium to be monomer/dimer exchange.

Thermal equilibrium and activation analysis of the cHSP27 dimer-to-monomer transition

^{15}N CPMG RD data were recorded on 1 mM ^2H , [$U\text{-}^{15}\text{N}$]-C137S at 25, 30, and 35 °C at both 11.7 T and 14.1 T and analyzed globally, assuming the forward and backward rates can be described by the Eyring equation such that $k_{\text{eg}} = AT \exp\left(-(\text{DH}_{\text{eg}} - T\text{DS}_{\text{eg}})/RT\right)$ and $k_{\text{ge}} = AT \exp\left(-(\text{DH}_{\text{ge}} - T\text{DS}_{\text{ge}})/RT\right)$. A is the reaction frequency, set here to be $3000 \text{ s}^{-1} \text{ K}^{-1}$ and T is the thermodynamic temperature (Fersht 1985). Similarly, the equilibrium constant is given by

$K_{eq} = [E]/[G] = \exp(-(DH - TDS)/RT)$. Justifying this, locally obtained values of k_{eg} , k_{ge} , and K_{eq} vary with temperature in a manner consistent with these equations.

pH-induced dissociation and unfolding of C137S dimers

A 2D ^1H - ^{15}N HSQC spectrum was recorded on a sample of [U - ^{15}N]-C137S at 250 μM in NMR buffer at pH 7 at 25 °C. Separate samples were independently prepared in NMR buffer at pH 6.5, 6.0, and 5.0 and NMR spectra were recorded to assess the effect of pH on dimerization. The well-resolved peak from G116 was used to calculate the dimerization K_d as a function of pH. At pH 5, no dimer was observed, and thus the K_d at pH 5 (~5 mM) is four orders of magnitude larger than that at pH 7 (0.5 μM). Below pH 6.5, the sample was highly unstable and white precipitant was evident by the end of the NMR experiments (20-40 minutes). Similar pH titrations were performed on oxidized cHSP27.

Analysis of the pressure-induced dissociation and unfolding of C137S dimers

A sample of 200 μM [U - ^{15}N]-C137S was prepared in a baroresistant buffer at pH 7 as described previously (Quinlan & Reinhart 2005) in order to prevent changes in pH with increasing pressure. The buffer contained a mixture of 100 mM Tris-HCl and 100 mM phosphate buffer, both at pH 7, and yields a negligible change in pH between 1 bar and 2500 bar. The standard NMR buffer used throughout this work (30 mM sodium phosphate, 2 mM EDTA, pH 7) experiences a change in pH of -0.9 units between 1 bar and 2500 bar due to the change in the pK_a of phosphate with pressure. Similarly, the change in the pK_w of water with pressure amounts to a ~5.4-fold increase in the concentration of hydroxide ions at 2500 bar relative to 1 bar, leading to faster hydrogen exchange rates. This is evident in the unfolded spectrum at 2500 bar where some peaks have become weaker than anticipated for an unfolded protein. Spectra recorded in phosphate buffer at 2500 bar, where the pH change of -0.9 units somewhat offsets the increase in hydroxide ions, reveal more uniform intensities, as expected for an unfolded protein (data not shown). 2D ^1H - ^{15}N HSQC

spectra were recorded with ^1H (^{15}N) 1024 (100) complex points, 8417 Hz (2083 Hz) sweep widths, 121.7 ms (48 ms) acquisition times at 14.1 T at 25 °C as a function of hydrostatic pressure between 1 bar to 2500 bar. Manipulation of the hydrostatic pressure was carried out using a commercial ceramic high-pressure NMR cell and an automatic pump system (Daedalus Innovations, Philadelphia, PA). Peak intensities from non-overlapping signals arising from the dimer, monomer, and unfolded species were quantified at each pressure and fit to a model of three-state unfolding. Similar pressure titrations were performed on oxidized cHSP27, with no evidence for population of a monomeric intermediate, although complete and reversible unfolding was observed.

3.5.8 CD spectroscopy and bis-ANS fluorescence of C137S and H124K/C137S

Samples for CD spectroscopy were prepared at 20 μM in NMR buffer and placed into cuvettes with a 1 cm path length. Data were recorded using a Jasco Model J720 CD spectrophotometer at wavelengths between 200 and 260 nm. The fluorescent probe bis-ANS (Sigma Aldrich) was dissolved in NMR buffer and added to a final concentration of 10 μM to wells containing a final volume of 100 μL of 40 μM C137S or H124K/C137S in NMR buffer. Under these conditions, C137S is primarily a dimer and H124K/C137S is primarily a monomer. Fluorescence at 500 nm was recorded following excitation at 350 nm.

3.5.9 Random Coil Index

To identify β strands in the C137S monomer and dimer, I utilized the software Random Coil Index ([Berjanskii & Wishart 2006](#)) and Chemical Shift Index ([Hafsa & Wishart 2014](#)) with default settings. $^1\text{H}^{\text{N}}$, ^{13}CO , $^{13}\text{C}\alpha$, and ^{15}N chemical shifts from the C137S dimer (pH 7) and monomer (pH 4.1) were used as input values. The deuterium isotope effect was corrected for using established values ([Cavanagh et al. 2007](#)).

3.6 Acknowledgements

The work outlined in this chapter has been published in *Nature Communications* in 2019 under the title, “Local unfolding of the HSP27 monomer regulates chaperone activity,” by the authors T. Reid Alderson, Julien Roche, Heidi Y. Gastall, David M. Dias, Iva Pritišanac, Jinfa Ying, Ad Bax, Justin L. P. Benesch, and Andrew J. Baldwin with the DOI 10.1038/s41467-019-08557-8. It was also published as a preprint article in the *bioRxiv* (article identifier 345751) repository with the DOI 10.1101/345751.

The author expressed and purified all indicated proteins in the study, and performed all experiments and analyses except for the aggregation assay of α -synuclein (Fig. 2g-i, Fig. 3d), which was performed by David M. Dias, collection of native MS on cHSP27 (Fig. 3b), which was performed by Justin L. P. Benesch, and creation of the cHSP27 and C137S expression plasmids, which Heidi Y. Gastall had previously performed. H. Y. G. also assisted with native MS collection on full-length HSP27. Ambient pressure NMR experiments were collected by the author with Andrew J. Baldwin or Jinfa Ying; high-pressure NMR data were collected by the author or with help from Julien Roche. NMR processing and analysis was performed by the author, with input from J. Y., A. J. B., and Ad Bax. Data analysis tools were contributed by Iva Pritišanac. The original draft was written by the author, with contributions added by all listed authors.

I thank Prof. Heath Ecroyd (University of Wollongong) for the WT HSP27 expression plasmid, Prof. Arthur Laganowsky (Texas A&M University) for assistance in constructing the cHSP27 expression plasmid, Prof. Christina Redfield (University of Oxford) for usage of the CD spectrophotometer, Dr. Errin Johnson (University of Oxford) for assistance with electron microscopy, and Dr. David Staunton (University of Oxford) for assistance with thermal denaturation experiments.

3.7 References

- Akasaka K. 2006. Probing conformational fluctuation of proteins by pressure perturbation. *Chem. Rev.* 106(5):1814–35
- Alderson TR, Benesch JLP, Baldwin AJ. 2017. Proline isomerization in the C-terminal region of HSP27. *Cell Stress Chaperones.* 1–13
- Almeida-Souza L, Asselbergh B, d'Ydewalle C, Moonens K, Goethals S, et al. 2011. Small heat-shock protein HSPB1 mutants stabilize microtubules in Charcot-Marie-Tooth neuropathy. *J. Neurosci.* 31(43):15320–28
- Almeida-Souza L, Goethals S, De Winter V, Dierick I, Gallardo R, et al. 2010. Increased monomerization of mutant HSPB1 leads to protein hyperactivity in Charcot-Marie-Tooth neuropathy. *J. Biol. Chem.* 285(17):12778–86
- Aquilina JA, Shrestha S, Morris AM, Ecroyd H. 2013. Structural and functional aspects of hetero-oligomers formed by the small heat shock proteins α B-crystallin and HSP27. *J. Biol. Chem.* 288(19):13602–9
- Arrigo A-P, Viot S, Chaufour S, Firdaus W, Kretz-Remy C, Diaz-Latoud C. 2005. Hsp27 consolidates intracellular redox homeostasis by upholding glutathione in its reduced form and by decreasing iron intracellular levels. *Antioxid. Redox Signal.* 7(3–4):414–22
- Arrigo AP. 2001. HSP27: novel regulator of intracellular redox state. *IUBMB Life.* 52(6):303-07
- Baldwin AJ. 2014. An exact solution for R2,eff in CPMG experiments in the case of two site chemical exchange. *J. Magn. Reson.* 244:114–24
- Baldwin AJ, Kay LE. 2009. NMR spectroscopy brings invisible protein states into focus. *Nat. Chem. Biol.* 5(11):808–14
- Baldwin AJ, Lioe H, Robinson C V, Kay LE, Benesch JLP. 2011. α B-crystallin polydispersity is a consequence of unbiased quaternary dynamics. *J. Mol. Biol.* 413(2):297–309
- Baranova EV, Weeks SD, Beelen S, Bukach OV, Gusev NB, Strelkov SV. 2011. Three-Dimensional Structure of α -Crystallin Domain Dimers of Human Small Heat Shock Proteins HSPB1 and HSPB6. *J. Mol. Biol.* 411(1):110–22
- Bardwell JCA, Jakob U. 2012. Conditional disorder in chaperone action. *Trends Biochem. Sci.* 37(12):517–25
- Basha E, O'Neill H, Vierling E. 2012. Small heat shock proteins and α -crystallins: dynamic proteins with flexible functions. *Trends Biochem. Sci.* 37(3):106–17
- Baughman HER, Clouser AF, Klevit RE, Nath A. 2018. HspB1 and Hsc70 chaperones engage distinct tau species and have different inhibitory effects on amyloid formation. *J. Biol. Chem.* 293(8):2687–2700
- Baxter NJ, Williamson MP. 1997. Temperature dependence of ^1H chemical shifts in proteins. *J. Biomol. NMR.* 9(4):359–69
- Bepperling A, Alte F, Kriehuber T, Braun N, Weinkauff S, et al. 2012. Alternative bacterial two-component small heat shock protein systems. *Proc. Natl. Acad. Sci. U. S. A.* 109(50):20407–12
- Berjanskii M, Wishart DS. 2006. NMR: prediction of protein flexibility. *Nat. Protoc.* 1(2):683–88
- Bova MP, Yaron O, Huang Q, Ding L, Haley DA, et al. 1999. Mutation R120G in alphaB-crystallin, which is linked to a desmin-related myopathy, results in an irregular structure and defective

- chaperone-like function. *Proc. Natl. Acad. Sci. U. S. A.* 96(11):6137–42
- Bruey JM, Ducasse C, Bonniaud P, Ravagnan L, Susin SA, et al. 2000. Hsp27 negatively regulates cell death by interacting with cytochrome c. *Nat. Cell Biol.* 2(9):645–52
- Buchner J, Grallert H, Jakob U. 1998. Analysis of chaperone function using citrate synthase as nonnative substrate protein. *Methods Enzymol.* 290:323–38
- Cabrita LD, Cassaignau AME, Launay HMM, Waudby CA, Wlodarski T, et al. 2016. A structural ensemble of a ribosome–nascent chain complex during cotranslational protein folding. *Nat. Struct. Mol. Biol.* 23(4):278–85
- Carver J., Richards R. 1972. A general two-site solution for the chemical exchange produced dependence of T2 upon the carr-Purcell pulse separation. *J. Magn. Reson.* 6(1):89–105
- Carver J a, Esposito G, Schwedersky G, Gaestel M. 1995. ¹H NMR spectroscopy reveals that mouse Hsp25 has a flexible C-terminal extension of 18 amino acids. *FEBS Lett.* 369(2–3):305–10
- Carver J a, Lindner R a. 1998. NMR spectroscopy of alpha-crystallin. Insights into the structure, interactions and chaperone action of small heat-shock proteins. *Int. J. Biol. Macromol.* 22(3–4):197–209
- Carver JA, Grosas AB, Ecroyd H, Quinlan RA. 2017. The functional roles of the unstructured N- and C-terminal regions in α B-crystallin and other mammalian small heat-shock proteins. *Cell Stress Chaperones.* 22(4):627–38
- Carver JA, Lindner RA, Lyon C, Canet D, Hernandez H, et al. 2002. The interaction of the molecular chaperone alpha-crystallin with unfolding alpha-lactalbumin: a structural and kinetic spectroscopic study. *J. Mol. Biol.* 318(3):815–27
- Cavanagh J, Fairbrother WJ, Palmer AG, Rance M, Skelton NJ. 2007. *Protein NMR Spectroscopy: Principles and Practice.* Academic Press
- Chalova AS, Sudnitsyna M V, Semenyuk PI, Orlov VN, Gusev NB. 2014a. Effect of disulfide crosslinking on thermal transitions and chaperone-like activity of human small heat shock protein HspB1. *Cell Stress Chaperones.* 19(6):963–72
- Chalova AS, Sudnitsyna M V., Strelkov S V., Gusev NB. 2014b. Characterization of human small heat shock protein HspB1 that carries C-terminal domain mutations associated with hereditary motor neuron diseases. *Biochim. Biophys. Acta - Proteins Proteomics.* 1844(12):2116–26
- Chen L, Ai X, Portaliou AG, Minetti CASA, Remeta DP, et al. 2013. Substrate-Activated Conformational Switch on Chaperones Encodes a Targeting Signal in Type III Secretion. *Cell Rep.* 3(3):709–15
- Chen L, Balabanidou V, Remeta DP, Minetti CASA, Portaliou AG, et al. 2011. Structural Instability Tuning as a Regulatory Mechanism in Protein-Protein Interactions. *Mol. Cell.* 44(5):734–44
- Christodoulou J, Larsson G, Fucini P, Connell SR, Pertinhez TA, et al. 2004. Heteronuclear NMR investigations of dynamic regions of intact Escherichia coli ribosomes. *Proc. Natl. Acad. Sci.* 101(30):10949–54
- Ciocca DR, Calderwood SK. 2005. Heat shock proteins in cancer: diagnostic, prognostic, predictive, and treatment implications. *Cell Stress Chaperones.* 10(2):86–103
- Clore GM, Szabo A, Bax A, Kay LE, Driscoll PC, Gronenborn AM. 1990. Deviations from the simple two-parameter model-free approach to the interpretation of nitrogen-15 nuclear magnetic relaxation of proteins. *J. Am. Chem. Soc.* 112(12):4989–91

- Clouser AF, Klevit RE. 2017. pH-dependent structural modulation is conserved in the human small heat shock protein HSBP1. *Cell Stress Chaperones*. 1–7
- Cox D, Selig E, Griffin MDW, Carver JA, Ecroyd H. 2016. Small Heat-shock Proteins Prevent α -Synuclein Aggregation via Transient Interactions and Their Efficacy Is Affected by the Rate of Aggregation. *J. Biol. Chem.* 291(43):22618–29
- Deckert A, Waudby CA, Wlodarski T, Wentink AS, Wang X, et al. 2016. Structural characterization of the interaction of α -synuclein nascent chains with the ribosomal surface and trigger factor. *Proc. Natl. Acad. Sci.* 113(18):5012–17
- Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A. 1995. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR.* 6(3):277–93
- Demarest SJ, Martinez-Yamout M, Chung J, Chen H, Xu W, et al. 2002. Mutual synergistic folding in recruitment of CBP/p300 by p160 nuclear receptor coactivators. *Nature.* 415(6871):549–53
- Diaz-Latoud C, Buache E, Javouhey E, Arrigo A-P. 2005. Substitution of the unique cysteine residue of murine Hsp25 interferes with the protective activity of this stress protein through inhibition of dimer formation. *Antioxid. Redox Signal.* 7(3–4):436–45
- Elliott JL, Der Perng M, Prescott AR, Jansen KA, Koenderink GH, Quinlan RA. 2013. The specificity of the interaction between α B-crystallin and desmin filaments and its impact on filament aggregation and cell viability. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 368(1617):20120375
- Evgrafov O V, Mersiyanova I, Irobi J, Van Den Bosch L, Dierick I, et al. 2004. Mutant small heat-shock protein 27 causes axonal Charcot-Marie-Tooth disease and distal hereditary motor neuropathy. *Nat. Genet.* 36(6):602–6
- Ferreon ACM, Ferreon JC, Wright PE, Deniz AA. 2013. Modulation of allostery by protein intrinsic disorder. *Nature.* 498(7454):390–94
- Fersht A. 1985. *Enzyme Structure and Mechanism*. W.H. Freeman
- Fleckenstein T, Kastenmüller A, Stein ML, Peters C, Daake M, et al. 2015. The Chaperone Activity of the Developmental Small Heat Shock Protein Sip1 Is Regulated by pH-Dependent Conformational Changes. *Mol. Cell.* 58(6):1067–78
- Franzmann TM, Wühr M, Richter K, Walter S, Buchner J. 2005. The Activation Mechanism of Hsp26 does not Require Dissociation of the Oligomer. *J. Mol. Biol.* 350(5):1083–93
- Fu X, Chang Z, Shi X, Bu D, Wang C. 2014. Multilevel structural characteristics for the natural substrate proteins of bacterial small heat shock proteins. *Protein Sci.* 23(2):229–37
- Ghosh JG, Estrada MR, Clark JI. 2005. Interactive Domains for Chaperone Activity in the Small Heat Shock Protein, Human α B Crystallin †. *Biochemistry.* 44(45):14854–69
- Ghosh JG, Houck SA, Clark JI. 2007. Interactive sequences in the stress protein and molecular chaperone human α B crystallin recognize and modulate the assembly of filaments. *Int. J. Biochem. Cell Biol.* 39(10):1804–15
- Groitel B, Horowitz S, Makepeace KAT, Petrotchenko E V, Borchers CH, et al. 2016. Protein unfolding as a switch from self-recognition to high-affinity client binding. *Nat. Commun.* 7:10357
- Hafsa NE, Wishart DS. 2014. CSI 2.0: a significantly improved version of the Chemical Shift Index. *J. Biomol. NMR.* 60(2–3):131–46
- Haslbeck M, Vierling E. 2015. A first line of stress defense: small heat shock proteins and their

- function in protein homeostasis. *J. Mol. Biol.* 427(7):1537–48
- Hernández H, Robinson C V. 2007. Determining the stoichiometry and interactions of macromolecular assemblies from mass spectrometry. *Nat. Protoc.* 2(3):715–26
- Hilton GR, Lioe H, Stengel F, Baldwin AJ, Benesch JLP. 2013. Small heat-shock proteins: paramedics of the cell. *Top. Curr. Chem.* 328(May 2012):69–98
- Hochberg GK a., Ecroyd H, Liu C, Cox D, Cascio D, et al. 2014. The structured core domain of β -crystallin can prevent amyloid fibrillation and associated toxicity. *Proc. Natl. Acad. Sci.* 111(16):E1562–70
- Hochberg GKA, Benesch JLP. 2014. Dynamical structure of α B-crystallin. *Prog. Biophys. Mol. Biol.* 115(1):11–20
- Houlden H, Laura M, Wavrant-De Vrièze F, Blake J, Wood N, Reilly MM. 2008. Mutations in the HSP27 (HSPB1) gene cause dominant, recessive, and sporadic distal HMN/CMT type 2. *Neurology.* 71(21):1660–68
- Huang C, Rossi P, Saio T, Kalodimos CG. 2016. Structural basis for the antifolding activity of a molecular chaperone. *Nature.* 537(7619):202–6
- Loria JP, Rance R, and Palmer AGIII. 1999. A Relaxation-Compensated Carr-Purcell-Meiboom-Gill Sequence for Characterizing Chemical Exchange by NMR Spectroscopy
- Jacobs WM, Knowles TPJ, Frenkel D. 2016. Oligomers of Heat-Shock Proteins: Structures That Don't Imply Function. *PLoS Comput. Biol.* 12(2):e1004756
- Jehle S, van Rossum B, Stout JR, Noguchi SM, Falber K, et al. 2009. α B-Crystallin: A Hybrid Solid-State/Solution-State NMR Investigation Reveals Structural Aspects of the Heterogeneous Oligomer. *J. Mol. Biol.* 385(5):1481–97
- Jovceviski B, Kelly MA, Rote AP, Berg T, Gastall HY, et al. 2015. Phosphomimics Destabilize Hsp27 Oligomeric Assemblies and Enhance Chaperone Activity. *Chem. Biol.* 22(2):186–95
- Kampinga HH, de Boer R, Beerstra N. 2015. *The Multicolored World of the Human HSPB Family*, Vol. 8. Cham: Springer International Publishing
- Kay L, Keifer P, Saarinen T. 1992. Pure absorption gradient enhanced heteronuclear single quantum correlation spectroscopy with improved sensitivity. *J. Am. Chem. Soc.* 114(26):10663–65
- Kay LE, Torchia D a, Bax a. 1989. Backbone dynamics of proteins as studied by ^{15}N inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease. *Biochemistry.* 28(23):8972–79
- Kazimierczuk K, Orekhov VY. 2011. Accelerated NMR spectroscopy by using compressed sensing. *Angew. Chem. Int. Ed. Engl.* 50(24):5556–59
- Kirstein J, Morito D, Kakahana T, Sugihara M, Minnen A, et al. 2015. Proteotoxic stress and ageing triggers the loss of redox homeostasis across cellular compartments. *EMBO J.* 34(18):2334–49
- Koldewey P, Stull F, Horowitz S, Martin R, Bardwell JCA. 2016. Forces Driving Chaperone Action. *Cell.* 166(2):369–79
- Koteiche HA, Mchaourab HS. 2006. Mechanism of a Hereditary Cataract Phenotype. *J. Biol. Chem.* 281(20):14273–79
- Lakomek N-A, Ying J, Bax A. 2012. Measurement of ^{15}N relaxation rates in perdeuterated proteins by TROSY-based methods. *J. Biomol. NMR.* 53(3):209–21

- Lee W, Tonelli M, Markley JL. 2015. NMRFAM-SPARKY: enhanced software for biomolecular NMR spectroscopy. *Bioinformatics*. 31(8):1325–27
- Lefevre JF, Dayie KT, Peng JW, Wagner G. 1996. Internal mobility in the partially folded DNA binding and dimerization domains of GAL4: NMR analysis of the N-H spectral density functions. *Biochemistry*. 35(8):2674–86
- Lin MT, Beal MF. 2006. Mitochondrial dysfunction and oxidative stress in neurodegenerative diseases. *Nature*. 443(7113):787–95
- Linding R, Schymkowitz J, Rousseau F, Diella F, Serrano L. 2004. A Comparative Study of the Relationship Between Protein Structure and β -Aggregation in Globular and Intrinsically Disordered Proteins. *J. Mol. Biol.* 342(1):345–53
- Lipari G, Szabo A. 1982. Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and range of validity. *J. Am. Chem. Soc.* 104(17):4546–59
- Mainz A, Bardiaux B, Kuppler F, Multhaupt G, Felli IC, et al. 2012. Structural and mechanistic implications of metal binding in the small heat-shock protein α B-crystallin. *J. Biol. Chem.* 287(2):1128–38
- Mainz A, Peschek J, Stavropoulou M, Back KC, Bardiaux B, et al. 2015. The chaperone α B-crystallin uses different interfaces to capture an amorphous and an amyloid client. *Nat. Struct. Mol. Biol.* 22(11):898–905
- Mandel AM, Akke M, Palmer AG. 1995. Backbone dynamics of Escherichia coli ribonuclease HI: correlations with structure and function in an active enzyme. *J. Mol. Biol.* 246(1):144–63
- McDonald ET, Bortolus M, Koteiche HA, Mchaourab HS. 2012. Sequence, Structure, and Dynamic Determinants of Hsp27 (HspB1) Equilibrium Dissociation Are Encoded by the N-Terminal Domain. *Biochemistry*. 51(6):1257–68
- Mchaourab HS, Berengian AR, Koteiche HA. 1997. Site-Directed Spin-Labeling Study of the Structure and Subunit Interactions along a Conserved Sequence in the α -Crystallin Domain of Heat-Shock Protein 27. Evidence of a Conserved Subunit Interface [†]. *Biochemistry*. 36(48):14627–34
- Meehan S, Knowles TPJ, Baldwin AJ, Smith JF, Squires AM, et al. 2007. Characterisation of amyloid fibril formation by small heat-shock chaperone proteins human α A-, α B- and R120G α B-crystallins. *J. Mol. Biol.* 372(2):470–84
- Muranova LK, Weeks SD, Strelkov S V., Gusev NB. 2015. Characterization of mutants of human small heat shock protein HspB1 carrying replacements in the N-terminal domain and associated with hereditary motor neuron diseases. *PLoS One*. 10(5):
- Mymrikov E V, Daake M, Richter B, Haslbeck M, Buchner J. 2016. The Chaperone Activity and Substrate Spectrum of Human Small Heat Shock Proteins. *J. Biol. Chem.* jbc.M116.760413
- Nagaraj N, Wisniewski JR, Geiger T, Cox J, Kircher M, et al. 2014. Deep proteome and transcriptome mapping of a human cancer cell line. *Mol. Syst. Biol.* 7(1):548–548
- Nefedova V V., Datskevich PN, Sudnitsyna M V., Strelkov S V., Gusev NB. 2013. Physico-chemical properties of R140G and K141Q mutants of human small heat shock protein HspB1 associated with hereditary peripheral neuropathies. *Biochimie*. 95(8):1582–92
- Nefedova V V., Muranova LK, Sudnitsyna M V., Ryzhavskaia AS, Gusev NB. 2015. Small Heat Shock Proteins and Distal Hereditary Neuropathies. *Biochem.* 80(13):1734–47
- Niemann A, Huber N, Wagner KM, Somandin C, Horn M, et al. 2014. The Gdap1 knockout mouse mechanistically links redox control to Charcot–Marie–Tooth disease. *Brain*. 137(3):668–82

- Outeiro TF, Klucken J, Strathearn KE, Liu F, Nguyen P, et al. 2006. Small heat shock proteins protect against α -synuclein-induced toxicity and aggregation. *Biochem. Biophys. Res. Commun.* 351(3):631–38
- Palmer AG. 1993. Dynamic properties of proteins from NMR spectroscopy. *Curr. Opin. Biotechnol.* 4(4):385–91
- Palmer AG, Kroenke CD, Loria JP. 2001. Nuclear magnetic resonance methods for quantifying microsecond-to-millisecond motions in biological macromolecules. *Methods Enzymol.* 339(1997):204–38
- Pasupuleti N, Gangadhariah M, Padmanabha S, Santhoshkumar P, Nagaraj RH. 2010. The role of the cysteine residue in the chaperone and anti-apoptotic functions of human Hsp27. *J. Cell. Biochem.* 110(2):408–19
- Paul C, Manero F, Gonin S, Kretz-Remy C, Virost S, Arrigo A-P. 2002. Hsp27 as a negative regulator of cytochrome C release. *Mol. Cell. Biol.* 22(3):816–34
- Quinlan RJ, Reinhart GD. 2005. Baroresistant buffer mixtures for biochemical analyses. *Anal. Biochem.* 341(1):69–76
- Rajagopal P, Liu Y, Shi L, Clouser AF, Klevit RE. 2015a. Structure of the α -crystallin domain from the redox-sensitive chaperone, HSPB1. *J. Biomol. NMR.* 63(2):223–28
- Rajagopal P, Tse E, Borst AJ, Delbecq SP, Shi L, et al. 2015b. A conserved histidine modulates HSPB5 structure to trigger chaperone activity in response to stress-related acidosis. *Elife.* 4:1–21
- Raman B, Ban T, Sakai M, Pasta SY, Ramakrishna T, et al. 2005. α B-crystallin, a small heat-shock protein, prevents the amyloid fibril growth of an amyloid β -peptide and β 2 - microglobulin. *Biochem. J.* 392:573–81
- Rogalla T, Ehrnsperger M, Preville X, Kotlyarov A, Lutsch G, et al. 1999. Regulation of Hsp27 oligomerization, chaperone function, and protective activity against oxidative stress/tumor necrosis factor alpha by phosphorylation. *J. Biol. Chem.* 274(27):18947–56
- Rosenzweig R, Farber P, Velyvis A, Rennella E, Latham MP, Kay LE. 2015. ClpB N-terminal domain plays a regulatory role in protein disaggregation. *Proc. Natl. Acad. Sci.* 112(50):E6872–81
- Rosenzweig R, Sekhar A, Nagesh J, Kay LE. 2017. Promiscuous binding by Hsp70 results in conformational heterogeneity and fuzzy chaperone-substrate ensembles. *Elife.* 6:
- Saio T, Guan X, Rossi P, Economou A, Kalodimos CG. 2014. Structural Basis for Protein Antiaggregation Activity of the Trigger Factor Chaperone. *Science.* 344(6184):1250494–1250494
- Salmon L, Ahlstrom LS, Horowitz S, Dickson A, Brooks CL, Bardwell JCA. 2016. Capturing a Dynamic Chaperone-Substrate Interaction Using NMR-Informed Molecular Modeling. *J. Am. Chem. Soc.* 138(31):9826–39
- Sattler M, Schleucher J, Griesinger C. 1999. Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. *Prog. Nucl. Magn. Reson. Spectrosc.* 34(2):93–158
- Sauerwein AC, Hansen DF. 2015. Relaxation Dispersion NMR Spectroscopy. , pp. 75–132. Springer US
- Schreiber G, Fersht AR. 1996. Rapid, electrostatically assisted association of proteins. *Nat. Struct. Biol.* 3(5):427–31

- Shammas SL, Travis AJ, Clarke J. 2014. Allostery within a transcription coactivator is predominantly mediated through dissociation rate constants. *Proc. Natl. Acad. Sci. U. S. A.* 111(33):12055–60
- Shen Y, Bax A. 2013. Protein backbone and sidechain torsion angles predicted from NMR chemical shifts using artificial neural networks. *J. Biomol. NMR.* 56(3):227–41
- Stengel F, Baldwin AJ, Painter AJ, Jaya N, Basha E, et al. 2010. Quaternary dynamics and plasticity underlie small heat shock protein chaperone function. *Proc. Natl. Acad. Sci. U. S. A.* 107(5):2007–12
- Sugase K, Dyson HJ, Wright PE. 2007. Mechanism of coupled folding and binding of an intrinsically disordered protein. *Nature.* 447(7147):1021–25
- Tamiola K, Acar B, Mulder FAA. 2010. Sequence-Specific Random Coil Chemical Shifts of Intrinsically Disordered Proteins. *J. Am. Chem. Soc.* 132(51):18000–3
- Tapley TL, Körner JL, Barge MT, Hupfeld J, Schauerte JA, et al. 2009. Structural plasticity of an acid-activated chaperone allows promiscuous substrate binding. *Proc. Natl. Acad. Sci. U. S. A.* 106(14):5557–62
- Tartaglia GG, Vendruscolo M. 2008. The Zyggregator method for predicting protein aggregation propensities. *Chem. Soc. Rev.* 37(7):1395
- Tjandra N, Feller SE, Pastor RW, Bax A. 1995. Rotational diffusion anisotropy of human ubiquitin from ¹⁵N NMR relaxation. *J. Am. Chem. Soc.* 117(50):12562–66
- Tollinger M, Skrynnikov NR, Mulder FA, Forman-Kay JD, Kay LE. 2001. Slow dynamics in folded and unfolded states of an SH3 domain. *J. Am. Chem. Soc.* 123(46):11341–52
- Treweek TM, Rekas A, Lindner RA, Walker MJ, Aquilina JA, et al. 2005. R120G alphaB-crystallin promotes the unfolding of reduced alpha-lactalbumin and is inherently unstable. *FEBS J.* 272(3):711–24
- Ungelenk S, Moayed F, Ho C-T, Grousl T, Scharf A, et al. 2016. Small heat shock proteins sequester misfolding proteins in near-native conformation for cellular protection and efficient refolding. *Nat. Commun.* 7:13673
- Vallurupalli P, Hansen DF, Kay LE. 2008. Structures of invisible, excited protein states by relaxation dispersion NMR spectroscopy. *Proc. Natl. Acad. Sci.* 105(33):11766–71
- Van Montfort R, Slingsby C, Vierling E. 2001. Structure and function of the small heat shock protein/alpha-crystallin family of molecular chaperones. *Adv. Protein Chem.* 59:105–56
- Vos MJ, Kanon B, Kampinga HH. 2009. HSPB7 is a SC35 speckle resident small heat shock protein. *Biochim. Biophys. Acta - Mol. Cell Res.* 1793(8):1343–53
- Walther DM, Kasturi P, Zheng M, Pinkert S, Vecchi G, et al. 2015. Widespread Proteome Remodeling and Aggregation in Aging *C. elegans*. *Cell.* 161(4):919–32
- Weeks SD, Muranova LK, Heirbaut M, Beelen S, Strelkov S V., Gusev NB. 2018. Characterization of human small heat shock protein HSPB1 α -crystallin domain localized mutants associated with hereditary motor neuron diseases. *Sci. Rep.* 8(1):688
- Weindruch R, Prolla TA, Lee C-K. 2000. Gene-expression profile of the ageing brain in mice. *Nat. Genet.* 25(3):294–97
- Wright PE, Dyson HJ. 2009. Linking folding and binding. *Curr. Opin. Struct. Biol.* 19(1):31–38
- Wright PE, Dyson HJ. 2014. Intrinsically disordered proteins in cellular signalling and regulation. *Nat. Rev. Mol. Cell Biol.* 16(1):18–29

- Ye H, Huang H, Cao F, Chen M, Zheng X, Zhan R. 2016. HSPB1 Enhances SIRT2-Mediated G6PD Activation and Promotes Glioma Cell Proliferation. *PLoS One*. 11(10):e0164285
- Ying J, Delaglio F, Torchia DA, Bax A. 2016. Sparse multidimensional iterative lineshape-enhanced (SMILE) reconstruction of both non-uniformly sampled and conventional NMR data. *J. Biomol. NMR*
- Yu AL, Fuchshofer R, Birke M, Kampik A, Bloemendal H, Welge-Lüssen U. 2008. Oxidative Stress and TGF- β 2 Increase Heat Shock Protein 27 Expression in Human Optic Nerve Head Astrocytes. *Investig. Ophthalmology Vis. Sci.* 49(12):5403
- Zavialov A, Benndorf R, Ehrnsperger M, Zav'yalov V, Dudich I, et al. 1998. The effect of the intersubunit disulfide bond on the structural and functional properties of the small heat shock protein Hsp25. *Int. J. Biol. Macromol.* 22(3–4):163–73

4

Dysregulation of HSP27 by a neuropathy-associated mutation

4.1 Abstract

In humans, HSP27 is a systemically expressed small heat-shock protein with myriad cellular roles under both stressful and basal conditions. Numerous heritable mutations in HSP27 have been identified that contribute to the onset of neuromuscular maladies, including Charcot-Marie-Tooth (CMT) disease, the most commonly inherited disorder of the peripheral nervous system. At the molecular level, HSP27 assembles into large, dynamic oligomers, in part, through intermolecular contacts between the structured α -crystallin domain (ACD) and a highly conserved IxI/V motif (IPV) in the disordered C-terminal region. Here, I investigated the structural, dynamical, and functional impact of the CMT-associated P182L mutation, which involves the central Pro residue of the IxI/V motif of HSP27. The P182L mutation increased the average oligomeric mass of full-length HSP27 by nearly 30-fold, from *ca.* 500 kDa to 15 MDa. Accompanying this large increase in molecular mass, the mutant protein was unable to prevent the heat-induced aggregation of a substrate. Using NMR spectroscopy, I determined that a peptide bearing the mutant ILV motif from the P182L variant binds to the isolated ACD markedly weaker than the corresponding wild-type peptide. Peptide binding kinetics were determined by relaxation dispersion NMR studies on the ACD in the presence of sub-stoichiometric amounts of P182L peptide, and reveal an attenuated association rate with altered enthalpic and entropic contributions to binding. My results indicate that the kinetics and thermodynamics of the sparsely bound IxI/V motif significantly modulate the oligomerization state, quaternary structure, and chaperone function of HSP27.

4.2 Introduction

HSP27 is a systemically expressed, human small heat-shock protein (sHSP) that performs diverse functions under basal and stressful cellular conditions (Kampinga et al. 2015). With significant roles in the maintenance of protein homeostasis, regulation of the redox environment, prevention of apoptosis, and stabilization of the cytoskeletal network, the biological activity of HSP27 is critical to overall cellular health (Arrigo 2001, Basha et al. 2012, Bruey et al. 2000, Gusev et al. 2002, Haslbeck & Vierling 2015). The malfunction or dysregulation of HSP27 activity can result in debilitating diseases: numerous heritable mutations are implicated in the onset of neuromuscular disorders (Nefedova et al. 2015), including Charcot-Marie-Tooth (CMT) disease, the most commonly inherited disorder involving the peripheral nervous system (Barisic et al. 2008). CMT disease is a group of neuropathies that affects approximately 1 in 2500 individuals and is characterized by progressive demyelination (type 1 CMT), axonal loss (type 2 CMT), or their combination (Rossor et al. 2013). When the affected axons exclusively include motor neurons, the disease is referred to as distal hereditary motor neuropathy (dHMN). Over 30 heritable mutations in the *hspB1* gene that encodes for HSP27 have been identified in patients with CMT (Echaniz-Laguna et al. 2017, Nefedova et al. 2015), rendering it among the most commonly mutated genes in CMT (DiVincenzo et al. 2014, Dohrn et al. 2017). In addition, transgenic mouse models bearing clinically observed mutations in *hspB1* develop CMT disease-like symptoms (D'Ydewalle et al. 2011), thereby implicating HSP27 as a direct driver of CMT disease onset.

At the molecular level, HSP27 is comprised of three domains: a conserved, ca. 80-residue α -crystallin domain (ACD) that is flanked by a ca. 90-residue N-terminal region (NTR) and a 30-residue C-terminal region (CTR), which contains the highly conserved IxI/V tripeptide motif (Fig. 1A). The CMT- and dHMN-associated mutations are dispersed throughout the amino acid sequence of HSP27 (Fig. 1A). While the ACD folds into a β -strand-rich immunoglobulin-like domain (Fig. 1B) (Hochberg et al. 2014, Rajagopal et al. 2015), the CTR is intrinsically disordered

and the NTR is predicted to lack a fixed structure. Similar to other metazoan sHSPs (Haslbeck & Vierling 2015, Hochberg & Benesch 2014), intermolecular contacts between all three domains of HSP27 facilitate its assembly into large oligomers (Hochberg et al. 2014, McDonald et al. 2012). In a related protein, α B-crystallin, results from solution- and solid-state NMR spectroscopy have revealed transient interactions in the full-length protein between the NTR and ACD (Jehle et al. 2010) and the IxI/V motif of one CTR and the ACD of an adjacent subunit (Baldwin et al. 2011a, 2012). HSP27 oligomers display an average mass of ca. 500 kDa, but dynamically exchange subunits, and these quaternary dynamics have hindered atomic-level investigations of the structure and dynamics of full-length HSP27.

High-resolution structures have been obtained on the isolated ACD (Hochberg et al. 2014, Rajagopal et al. 2015), which forms stable dimers (Fig. 1B, Fig. 1C). A crystal structure of the HSP27 ACD bound to a fragment of the CTR containing the highly conserved IxI/V motif (I181-P182-V183) provided detailed molecular insight, revealing the formation of hydrogen bonds between the peptide and ACD (Fig. 1D) and also penetration of the Ile and Val residues into a hydrophobic groove created by the β 4/ β 8 strands of the ACD (Hochberg et al. 2014). In α B-crystallin, an allosteric network exists between the β 4/ β 8 groove and the ACD dimer interface (Delbecq et al. 2012), with IxI/V binding consequently increasing the K_d of the dimer (Hilton et al. 2013). Moreover, combined insights from NMR and native mass spectrometry (MS) experiments performed on full length α B-crystallin, observed that the rate associated with the binding of two IxI/V motifs to an ACD is correlated with the rate of subunit exchange (Baldwin et al. 2011b). Therefore, when two IxI/V motifs dock into the hydrophobic β 4/ β 8 groove in the ACD, an allosteric signal weakens the dimer interface and culminates in ejection of the CTR-bound subunit from the oligomer (Baldwin et al. 2011b).

Notably, two dHMN-causing mutations in HSP27 reside within the highly conserved IxI/V motif (P182L, P182S) (Evgrafov et al. 2004), with the P182L variant comprising one of the most

clinically severe HSP27 variants. A transgenic mouse model expressing the mutant P182L protein in neurons progressively displayed motor defects and axonal loss in motor neurons (D'Ydewalle et al. 2011), while an *in vitro* study demonstrated that the P182S mutation led to a large increase in the average molecular mass of HSP27 with diminished chaperone activity (Chalova et al. 2014). The molecular basis for the disease-causing mutations involving P182, however, remains unknown. When bound to the ACD, the affected Pro residue (P182) does not make direct intermolecular contacts (Hochberg et al. 2014) (Fig. 1E), thus the impact of its mutation to Leu remains unclear and may arise through indirect effects.

In this chapter, I summarize my efforts to determine the biophysical significance of the P182L mutation and its impact on the interaction of the CTR with the ACD. I observed that the P182L mutation increases the average molecular mass of soluble HSP27 oligomers to near 15 MDa, an increase over the WT protein by a factor of 30, and nearly five times the mass of the eukaryotic 80S ribosome. With a minimalistic system comprising the ACD and a peptide containing the IxI/V motif, I determined that the binding affinity of the IxI/V motif is drastically attenuated in the P182L variant, with the K_d increased by an order of magnitude. The origin of this weakened affinity was investigated using Carr-Purcell-Meiboom-Gill (CPMG) relaxation dispersion (RD) experiments, providing detailed insight into the kinetics and thermodynamics of IxI/V binding. Given the highly conserved nature of the IxI/V motif, my findings should be generally relevant for sHSPs that harbor this tripeptide motif.

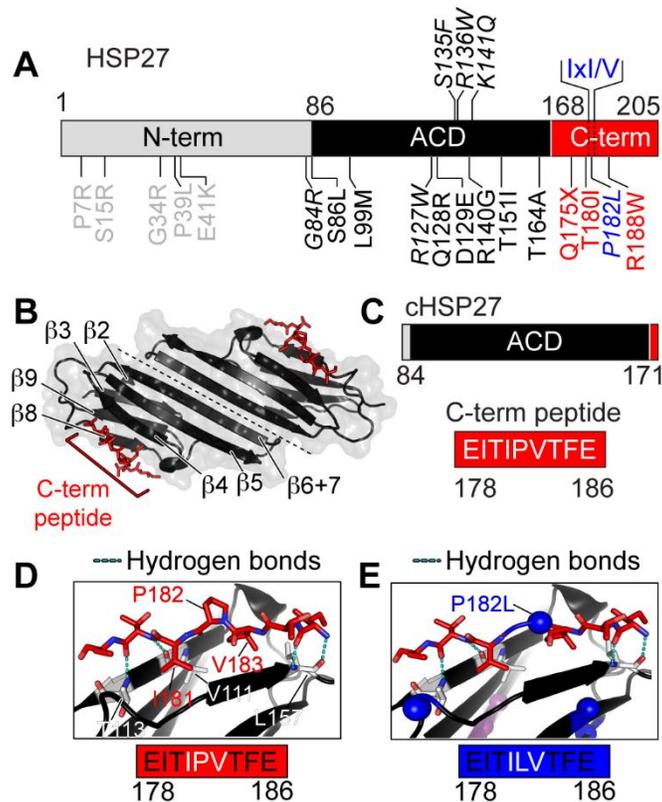


Figure 1. (A) Domain architecture of human HSP27: N-terminal domain (N-term), α -crystallin domain (ACD), and C-terminal region (C-term). The conserved IxI/V motif in the C-terminal region is indicated. Mutations associated with CMT disease, dHMN, and other motor neuropathies are shown. The P182L mutation in the IxI/V motif is blue. Residues with multiple mutations are italicized (e.g. P182S has also been reported). (B) Three-dimensional structure of the ACD dimer (black) bound to a peptide from the C-terminal region (red) containing the IxI/V motif. The β -strands of the ACD are numbered. (C) Domain boundaries of the ACD used in this study, cHSP27, and residues in the C-terminal peptide. (D, E) Zoomed-in region from (B) showing the contacts made between the C-terminal peptide and the ACD for the wild-type peptide (D) and if P182 was mutated to Leu (E). Other disease-causing mutations in the ACD are indicated with blue spheres (e.g. T164A, T151I).

4.3 Results

4.3.1 P182L forms very large, soluble oligomers that are chaperone-inactive

I purified recombinant human HSP27 and the dHMN-implicated P182L variant from *E. coli*, and characterized both proteins by size exclusion chromatography coupled to multi-angle light scattering (SEC-MALS). For WT HSP27, the SEC-MALS data revealed an average mass near 500 kDa (Fig. 2A), consistent with previous data. In contrast, the SEC-MALS profile of the P182L variant revealed significantly larger oligomers that exhibited an average mass near 15 MDa (Fig. 2A), a 30-fold increase over the WT protein. With a monomeric mass of 22.7 kDa, the P182L

variant therefore contains an average of *ca.* 660 subunits, whereas the WT protein comprises *ca.* 24 subunits. The average radius of hydration of the P182L variant increased from *ca.* 9 nm to *ca.* 38 nm, indicating that the single point-mutation P182L in the CTR drastically shifts the oligomeric distribution of HSP27 toward larger species. The isolated ACD, in contrast, exists as a dimer with a molecular mass of *ca.* 20 kDa (Fig. 2A).

In support of these SEC-MALS data, negative-stain electron microscopy (EM) demonstrated that the P182L variant assembles into much larger oligomeric forms than the WT protein (Fig. 2B). Negative-stain EM data were collected at a protein concentration of *ca.* 4 μ M, 10-fold lower than the SEC-MALS experiments, indicating that both WT and P182L HSP27 can form large oligomers in this concentration regime. Interestingly, the polydispersity of both the P182L variant and the WT protein was also evident from the negative-stain EM images (Fig. 2B).

To compare the chaperone activity of the WT and P182L forms of HSP27, I assayed their ability to prevent the aggregation of two model substrate proteins, malate dehydrogenase (MDH) and insulin. The WT form of HSP27 proved highly effective at inhibiting the aggregation of both substrates (Fig. 2C, Fig. 2D), indicative of its potent molecular chaperone activity. However, the P182L variant did not prevent the aggregation of either substrate (Fig. 2C, Fig. 2D), revealing a drastic loss of function. Indeed, the P182L form of HSP27 appeared to co-aggregate with MDH and insulin, based on the elevated absorbance signal in the presence of the chaperone. Control aggregation assays using P182L or HSP27 alone indicated the absence of appreciable aggregation, suggesting that the loss of P182L chaperone activity is not due to self-aggregation. The P182L mutation thus drastically alters the oligomeric properties of HSP27, shifts the oligomeric equilibrium to very large states, and is coupled to a concomitant loss of chaperone activity.

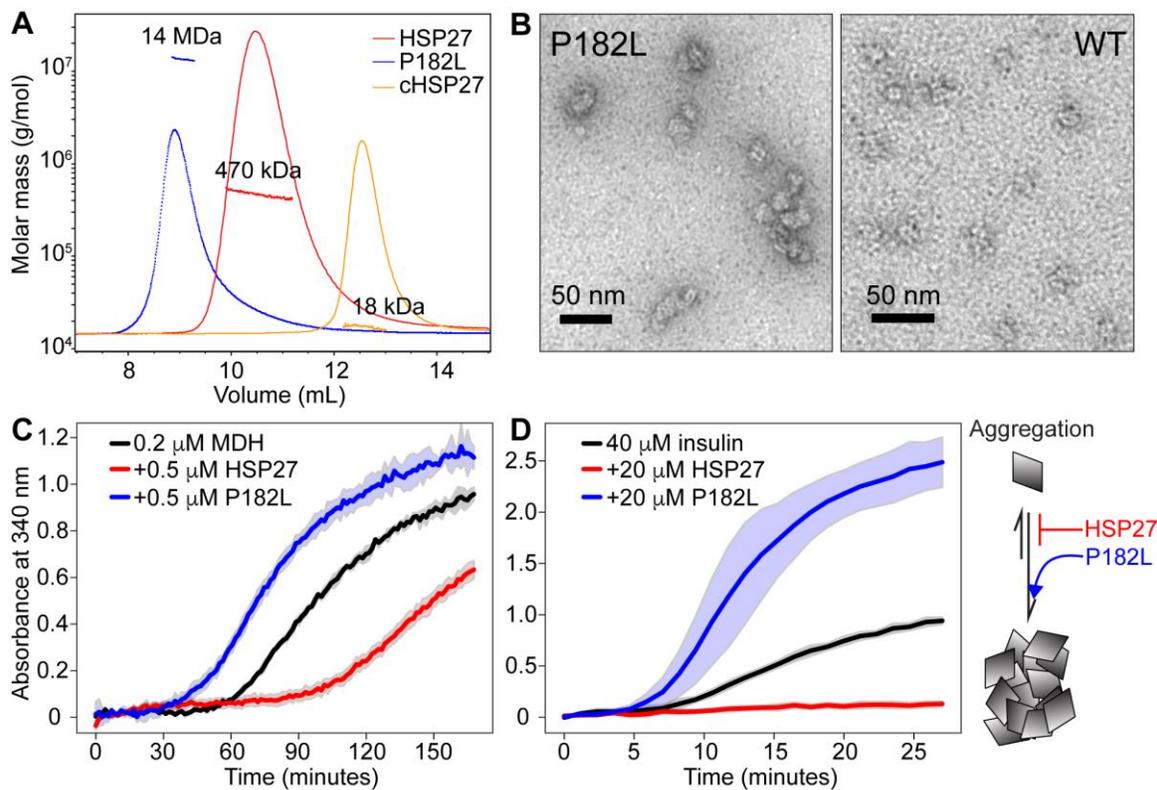


Figure 2. The P182L variant of HSP27 forms large oligomers that are chaperone-inactive. (A) SEC-MALS data from WT HSP27 and the P182L variant. Both samples were injected at 40 μ M concentration (monomer) in 30 mM sodium phosphate, 2 mM EDTA, 100 mM NaCl buffer at pH 7. The UV signals are plotted on the right y-axis as solid lines and the molecular masses are plotted on the left y-axis as circles. (B) Negative-stain EM images show that the P182L variant (left) forms much larger oligomeric assemblies than the WT form (right). Both samples were loaded on the grids at the same concentration (4.5 μ M). (C, D) Chaperone activity assays monitor the increase in light scattering at 340 nm due to the formation of large aggregates. The solid lines represent the average of three replicates with the filled area reflecting \pm one standard deviation. In (C), MDH (0.2 μ M) was incubated at 40 $^{\circ}$ C in the absence (black) or presence of 0.5 μ M WT HSP27 (red) or 0.5 μ M P182L HSP27 (blue). In (D), 40 μ M insulin was incubated at 40 $^{\circ}$ C in the absence (black) or presence of 20 μ M WT HSP27 (red) or 20 μ M P182L HSP27 (blue). The buffer used for both (C) and (D) was the same as in (A).

4.3.2 The P182L mutation significantly lowers binding affinity for the IxI motif

Because the CTRs of sHSPs are implicated in oligomerization, I sought to compare the binding affinities of IxI/V motifs in HSP27 (IPV) and the P182L variant (ILV). A previous solution-state NMR study of WT HSP27 only observed resonances from the disordered CTR, and thus could not probe CTR binding to the ACD in the context of the full-length protein (Alderson et al. 2017) (e.g. see Chapter 7). I thus turned to the isolated ACD, which forms stable dimers (2 x 10 kDa) that are known to bind to a peptide encompassing the IxI/V motif of the CTR (Hochberg et al. 2014). cHSP27 contains an intermolecular disulfide bond that locks the dimer in place. I employed

solution-state NMR spectroscopy and first determined the K_d of the interaction between the ACD of HSP27 and a peptide comprising the WT IPV motif (Fig. 3A). Since the ACD was uniformly ^{15}N labeled, each amide bond in the protein contributed a signal in the 2D ^1H - ^{15}N heteronuclear single quantum coherence (HSQC) spectrum (Fig. 3A, bottom). During the peptide titration, an increase in concentration of the IPV-containing peptide with respect to the ACD led to progressive changes in the chemical shifts of specific ACD residues (Fig. 3B). The chemical shift perturbations (CSPs) from resonances in fast exchange were globally fit to equation 1 to determine the K_d of this binding interaction, yielding a value of ca. 125 μM (Fig. 3C).

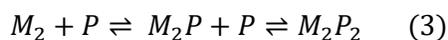
$$\Delta\delta = \Delta\delta_{\text{max}} \frac{\left(([P]_t + [L]_t + K_d) - \sqrt{([P]_t + [L]_t + K_d)^2 - 4[P]_t[L]_t} \right)}{2[P]_t} \quad (1)$$

Similar results were obtained with the software TITAN (Fig. 4) (Waudby et al. 2016). This K_d is consistent with, though slightly higher than, previous NMR titration data on αB -crystallin (80 μM) (Delbecq et al. 2012), perhaps due to the absence of a palindromic sequence in HSP27 that enables bi-directional binding of the IxI/V motif in αB -crystallin, as noted by Delbecq et al.

The high-quality of the fits of the NMR data to a two-state model suggests that two IxI/V-containing peptides bind concurrently to the ACD dimer in the following manner:



where M stands for an ACD monomer and P for a peptide. Note that the reaction rates are not indicated, but k_{on} and k_{off} would appear above and below the equilibrium arrow, respectively. I cannot rule out a sequential binding model, however, as the observed NMR signals could also contain contributions from singly bound peptide forms of the ACD. For resonances in fast exchange, the chemical shift would reflect an ensemble-averaged position of the free, singly bound, and double bound forms of the ACD



My CSP data cannot differentiate between concurrent and sequential binding models; however, I note that previous native mass spectrometry experiments, which can resolve individual

components in a heterogeneous mixture, detected singly-bound forms of the ADC from ABC (Hilton et al. 2013).

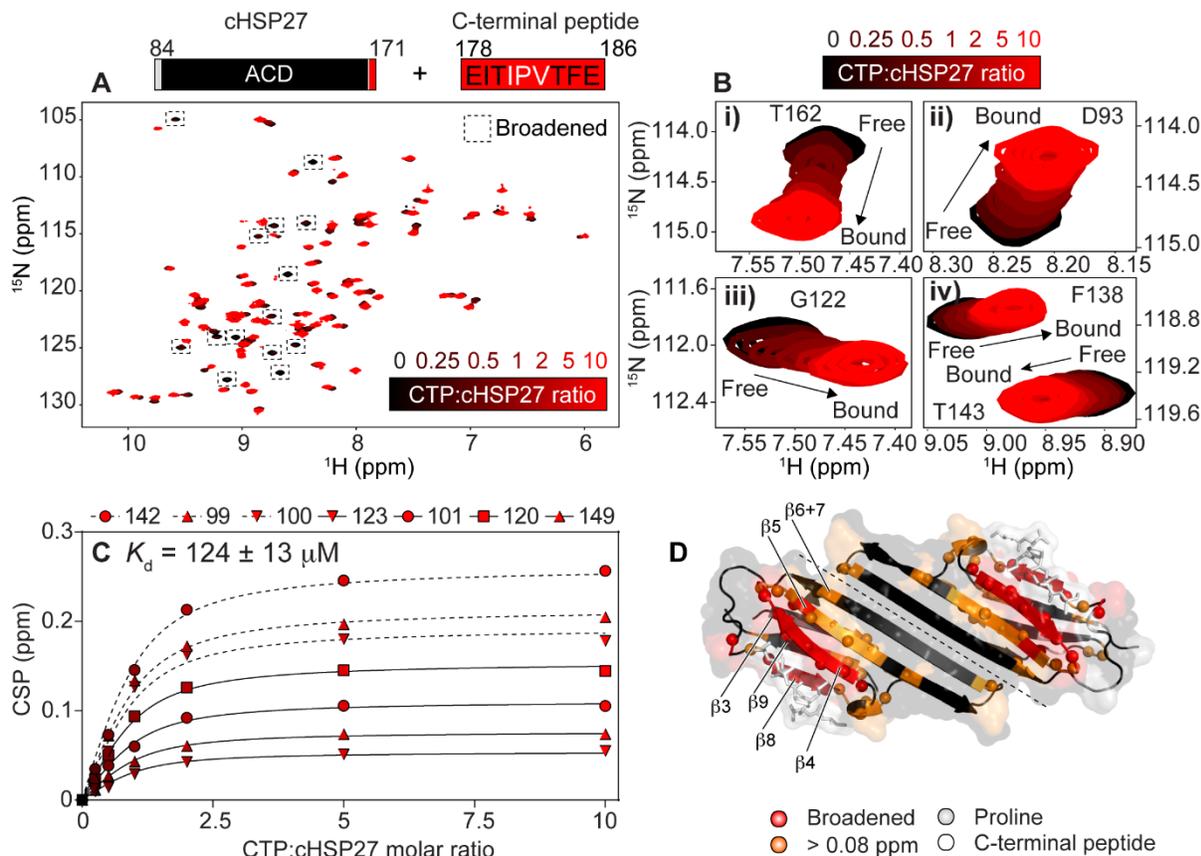


Figure 3. NMR characterization of IxI/V binding to the ACD of HSP27. (A) 2D ^1H - ^{15}N HSQC spectra of ^{15}N -labeled cHSP27 (black) in the presence of increasing amounts of IxI/V-containing CTP (red). The ratio of CTP:cHSP27 is indicated by the black-to-red color bar. (B, i-iv) Zoomed-in regions of (A) displaying cHSP27 resonances that are in fast exchange. (C) The combined ^1H and ^{15}N chemical shift changes, denoted as chemical shift perturbation (CSP), are plotted as a function of the CTP:cHSP27 molar ratio. The concentration of cHSP27 was 0.4 mM for all points except the 10:1 ratio, where a cHSP27 concentration of 0.2 mM was employed. Lines indicate the best fit to Eqn 1. The average $K_D \pm$ one standard deviation is shown in the upper left. (D) X-ray structure of cHSP27 bound to its CTP (PDB: 4mjh), shown in white sticks. At a 10:1 molar ratio of CTP:cHSP27, residues experiencing CSPs > 0.08 ppm are colored orange, whereas residues whose resonances have become broadened are red. Proline residues are colored grey.

In addition to providing information about the strength of the binding interaction between the IPV peptide and the ACD, the NMR data also indicated residues whose chemical environments changed upon peptide binding. As expected from the crystal structure of the ACD-CTP complex (Hochberg et al. 2014), resonances from residues near the β_4/β_8 groove are significantly impacted by CTP binding. Notably, these resonances became severely broadened due to intermediate

exchange and their large chemical shift differences between the free and bound states (Fig. 3D). However, residues in the $\beta 5$ and $\beta 6+7$ strands also displayed significant CSPs (Fig. 3D), which is likely due to structural changes upon peptide binding. In particular, E130 and Y142, both of which make intermolecular contacts that stabilize the dimer interface, exhibited relatively large CSPs (Fig. 3D).

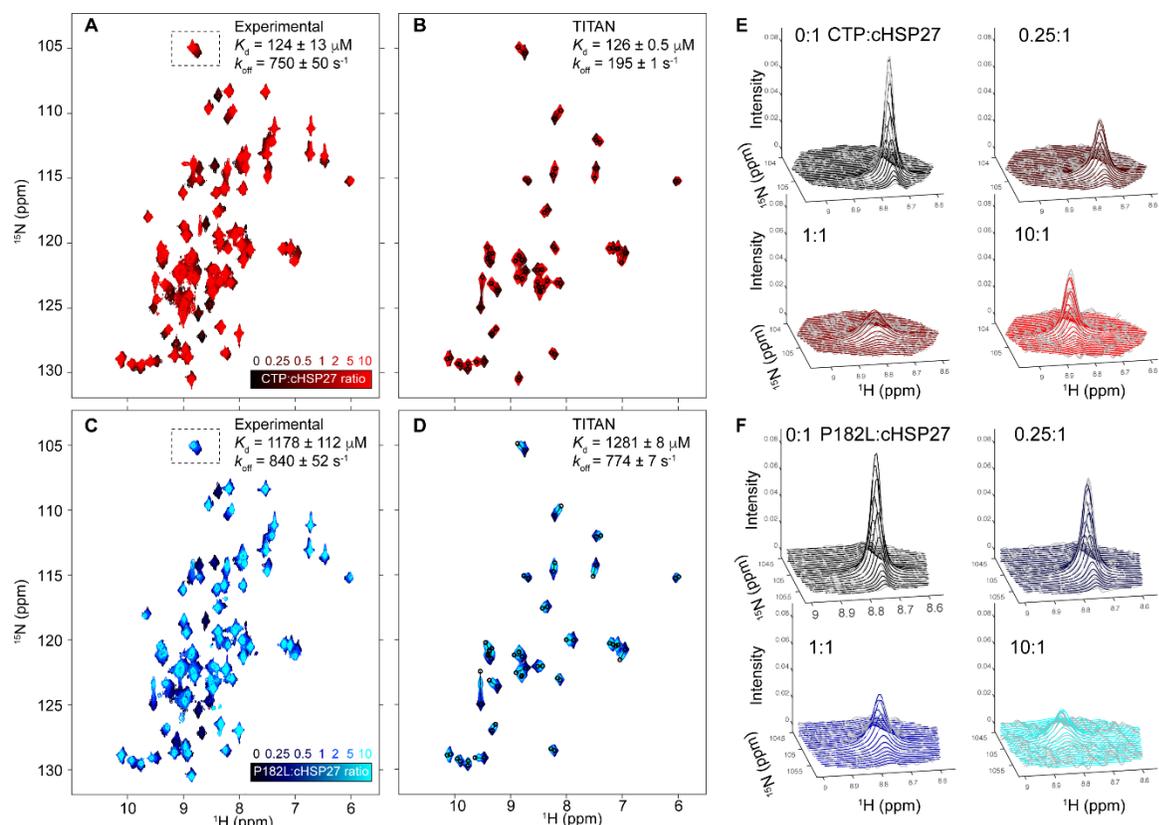


Figure 4. Analysis of the peptide binding titration data with the TITAN software. (A), Experimental 2D ^1H - ^{15}N HSQC spectra of ^{15}N -ACD (black) in the presence of increasing amounts of unlabeled CTP (red color gradient). The globally determined K_d from fitting resonances in fast exchange to Eqn 1 is listed in the upper-right corner. The k_{off} listed is derived from CPMG RD experiments (*vide infra*). (B) Using resonances in fast, intermediate, and slow exchange, the software TITAN fits NMR titration data to the Bloch-McConnell equations. For the selected resonances in the ACD, the reconstructed spectra is shown as fitted by the software. The K_d and k_{off} fitted by TITAN are listed in the upper-right corner. The color scheme is the same as panel A. (C) Same as panel A except for the peptide for the P182L variant. (D) Same as panel B for the peptide P182L. (E, F) For the resonances in panels A and C, selected peak shapes (grey lines) and the corresponding fits (colored lines) by TITAN are shown. Note that the intensities of the resonances at 10:1 peptide:chHSP27 ratio are significantly lower than those at 0:1 ratio.

I next compared the binding affinity and the uncovered allosteric changes in the ACD to the effects of the ILV motif binding within the P182L mutant peptide that is implicated in the onset

of dHMN (Fig. 5). Surprisingly, even though the central Pro residue in the IPV motif does not contact the ACD, its mutation to Leu (P182L) lowers the binding affinity for the ACD by nearly one order of magnitude (from *ca.* 125 μ M to 1.2 mM, Fig. 3C, Fig. 5B). A comparison of CSPs between the WT and mutant peptides revealed that many of the same residues were impacted by peptide binding (Fig. 3D), with residues in the β 4/ β 8 groove also becoming broadened in the P182L mutant (Fig. 5D-F). The overall similarity of the CSPs between the two peptides suggests that the P182L peptide binds in a similar site and conformation as the WT peptide (Fig. 5E, F), with common allosteric changes that manifest upon binding. Therefore, the main difference between the WT and P182L mutant peptides corresponds to a drastically lowered binding affinity of the mutant. Unfortunately, the weak interaction between the P182L peptide and the ACD, coupled with relatively low solubility of the peptide, prevented the preparation of a nearly-saturated complex. Consequently, 15 N spin relaxation experiments that probe dynamical changes on the fast (ps-ns) timescale could not be recorded.

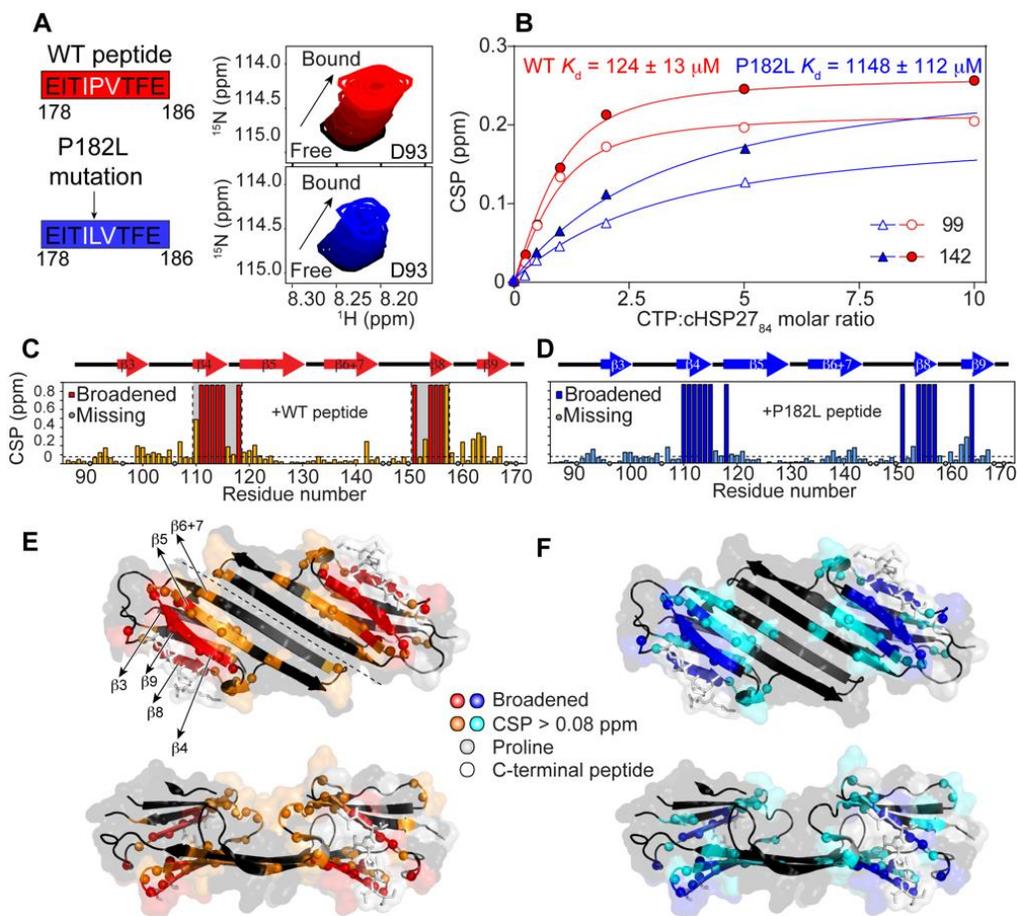


Figure 5. Impact of the P182L mutation on binding of the IxI/V motif to the ACD. (A) *Left*, Primary sequences from the WT (red) and P182L (blue) peptides. The IxI/V motif is indicated with white text. *Right*, example NMR titration data. The resonance from D93 in ¹⁵N-cHSP27 is shown as a function of increasing concentration of WT (top) or P182L (bottom) peptides. (B) CSPs in cHSP27 residues L99 and Y142 as a function of increasing peptide concentration. The CSPs were fit to Eqn 1 in order to derive a K_D for the peptide-cHSP27 interaction. (C, D) CSPs shown as a function of residue number for the WT (red) and P182L (blue) peptides at a 10:1 peptide:cHSP27 molar ratio. (E, F) The results from (C, D) are plotted onto the structure of IxI/V-bound cHSP27 (PDB ID: 4mjh).

4.3.3 Mutant and WT IxI/V binding kinetics probed by relaxation dispersion NMR

To determine the origin of the weakened binding interaction in the P182L mutant, I instead turned to Carr-Purcell-Meiboom-Gill (CPMG) relaxation dispersion (RD), which enables quantification of the binding kinetics and thermodynamics involved in a protein-ligand interaction if k_{ex} is in the micro-millisecond timescale (Demers & Mittermaier 2009, Hansen et al. 2008b, Palmer et al. 2001, Sugase et al. 2007). Importantly, if the interconversion rates are favorable, CPMG RD experiments can provide such information when as little as 0.5% of a peptide-ligand complex is present (Palmer et al. 2001). In CPMG RD experiments, the rate of interconversion between a ground state (p_A)

and a sparsely populated state ($p_B > ca. 0.5\%$) can be determined by fitting the measured effective R_2 relaxation rates ($R_{2,eff}$) as a function of the number of 180° refocusing pulses employed during a fixed relaxation delay in the experiment (Hansen et al. 2008a, Palmer et al. 2001). In a reaction involving the association of a protein (A) and a ligand (B), the values obtained from fitting the CPMG data ($k_{ex} = k_{AB} + [L]_{free}k_{BA}$, p_B , and $|\Delta\omega|$) together provide a detailed characterization of the kinetics (k_{ex}), thermodynamics (p_B), and structural changes ($|\Delta\omega|$) that occur upon the transition from the visible ground state (*i.e.* the free ACD) to the sparsely populated state (*i.e.* the ACD-peptide complex).

I interrogated peptide binding kinetics and thermodynamics with ^{15}N CPMG RD by adding a small amount of WT or P182L peptide to a solution of ^2H , ^{15}N -labeled ACD. With *ca.* 2% of the WT peptide-ACD complex present, large R_{ex} ($R_{2,eff}(0 \text{ Hz}) - R_{2,eff}(\infty)$) values were observed for residues in the $\beta 4/\beta 8$ groove (Fig. 6). Importantly, in the absence of peptide, these residues did not yield significant R_{ex} (Fig. 6A, Fig. 6B *i.*), indicating that the observed effects are specific to the peptide-bound state. In addition, the magnitude of R_{ex} depended on the amount of added peptide (Fig. 7), which revealed that the conformational exchange arises from an intermolecular binding event, *i.e.* the minor state is the ACD-peptide complex. In conclusion, the observed dispersions in the presence of *ca.* 2% of the ACD-peptide complex reflect chemical exchange between the peptide-free and peptide-bound states. Similar data were obtained for the P182L peptide-ACD complex, albeit with more peptide added to reach *ca.* 2% of the complex, as expected from the significantly elevated K_d (Fig. 4, Fig. 5).

CPMG RD data for the WT and P182L peptides (Fig. 6A, Fig. 6B, *ii.* and *iii.*) were recorded at two static magnetic field strengths and analyzed using a two-state model of chemical exchange. Similar ACD regions experienced conformational exchange in both peptide-bound states (Fig. 6C *ii.*, *iii.*), further supporting the notion that both the WT and P182L peptides recognize and bind to the same site on the ACD. The dissociation rates of the bound peptides from the ACD-peptide complexes were highly similar and differed by less than a factor of two (Fig. 6C *i.*). However, the

association rates (k_{on}) for the two peptides differed by a factor of 10 (Fig. 6C i), suggesting that the nearly 10-fold difference in binding affinity largely originates from an attenuated k_{on} rate for the P182L peptide binding.

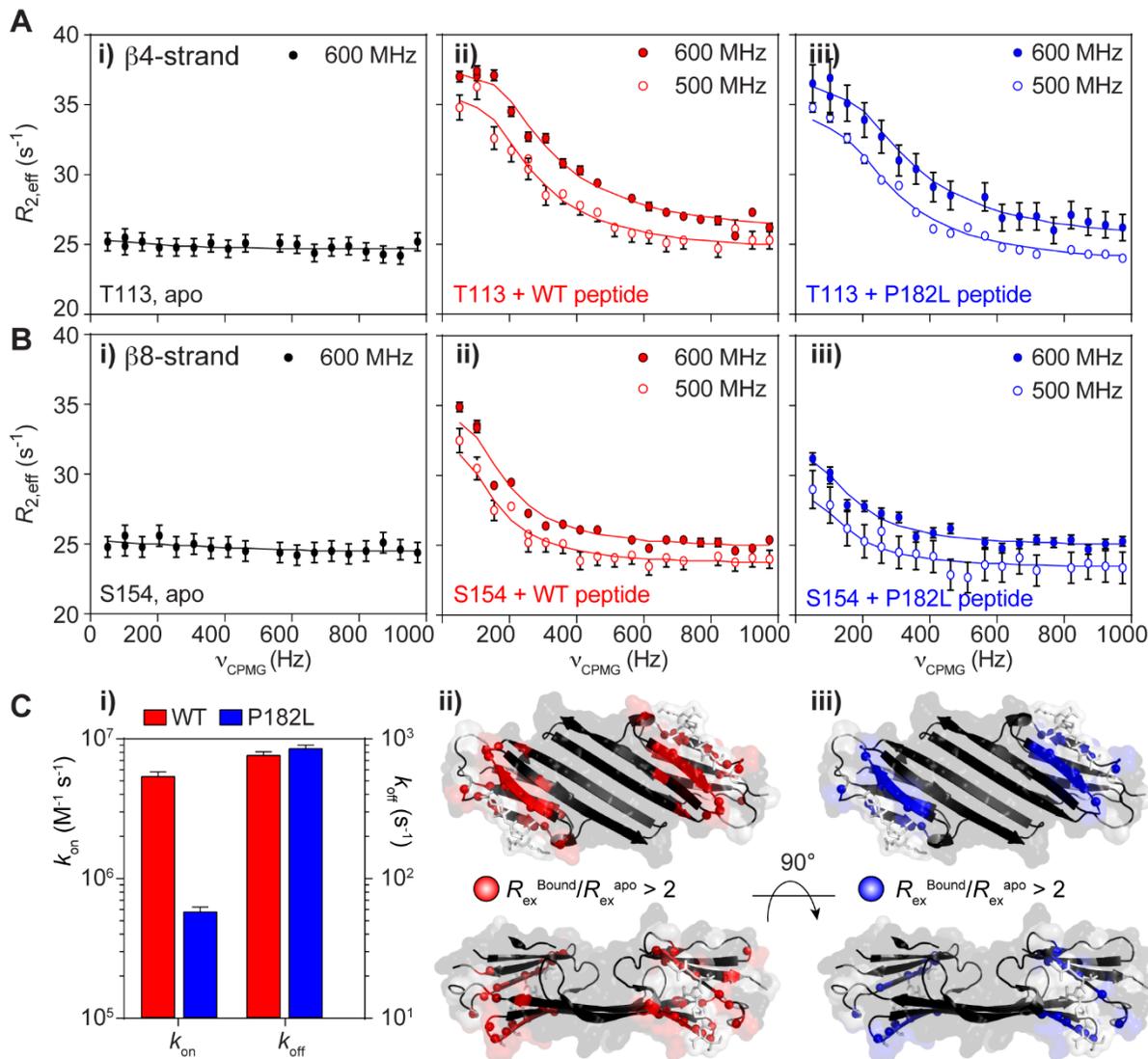


Figure 6. CPMG RD-based measurement of IxI/V binding kinetics. (A, B) *i.* ¹⁵N CPMG RD data for a residue in the $\beta 4$ strand (T113) and the $\beta 8$ strand (S154) in the absence of added peptide. CPMG data are shown for the same residues in the presence of *ii.* ca. 2% chSP27-WT peptide complex or *iii.* ca. 2% of the chSP27-P182L peptide complex. (C) *i.* The association (k_{on}) and dissociation (k_{off}) rates for IxI/V binding derived from the CPMG RD data collected on the WT (red) and P182L (blue) peptides. Note that k_{off} was calculated directly from the CPMG RD data ($p_A k_{ex}$), whereas k_{on} was calculated using the K_d derived from the CSP titration. Residues that are impacted by the presence of a small amount of peptide are indicated by red (*ii.*, WT) and blue (*iii.*, P182L) spheres. In both cases, residues in similar ACD regions show enhanced R_{ex} .

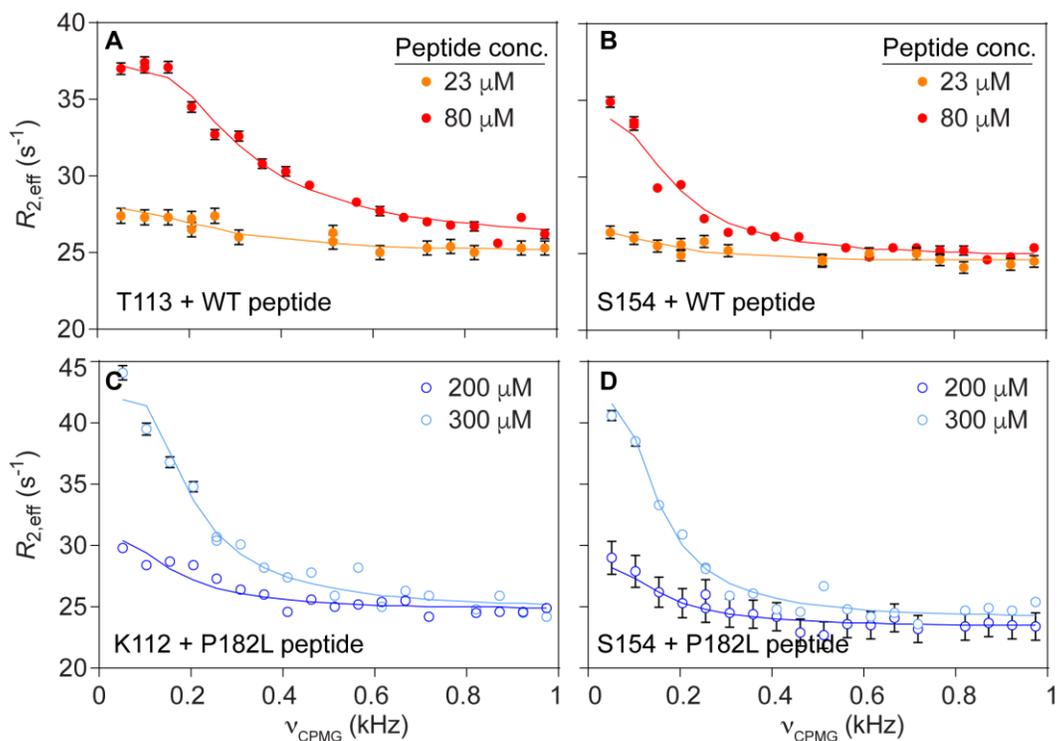


Figure 7. Concentration-dependence of dispersions from peptide-bound cHSP27 (A, B) *i.* ^{15}N CPMG RD data for a residue in the $\beta 4$ strand (T113) and the $\beta 8$ strand (S154) in the presence of *ii.* *ca.* 2% (red) or *ca.* 0.2% cHSP27-WT peptide complex. (C, D) Dispersions from residues in the $\beta 4$ (C) or $\beta 8$ strand (D) as a function of added P182L peptide. In dark blue is *ca.* 2% of the cHSP27-P182L peptide complex and in light blue is *ca.* 4% of the cHSP27-P182L peptide complex.

4.3.4 Altered thermodynamic properties of the P182L variant

In order to obtain a holistic description of the binding interactions, I analyzed the thermodynamics governing the WT and P182L peptide dissociation reactions as a function of temperature over the range from 20 to 45 °C in 5 °C increments (Fig. 8A, 8B). Under identical buffer conditions, the melting temperature for the ACD is near 70 °C (see Chapter 3), ensuring that the protein is stable up to 45 °C. The multi-temperature CPMG RD data provide the change in activation enthalpy (ΔH^\ddagger) and entropy (ΔS^\ddagger) for the dissociation event, enabling a direct comparison of the thermodynamics. An Eyring plot of these data (Fig. 8C) revealed only small differences in the ΔH^\ddagger and ΔS^\ddagger of dissociation for the WT and P182L peptides, indicating that the primary differences arise from the association reaction.

Together with the CSP and kinetic analyses above, these thermodynamic data suggest that the WT and P182L peptides behave nearly identically when bound to the $\beta 4/\beta 8$ groove in the ACD. Remarkably, however, I observed opposite behavior regarding the fractional change of the two peptide-ACD complexes as a function of temperature (Fig. 8D). The fraction of the WT peptide-ACD complex increased as the temperature increased, whereas the P182L peptide-ACD complex decreased with increasing temperature. These distinct thermodynamics therefore modulate the extent of IxI/V binding and overall oligomerization of HSP27, and further investigation will unravel which thermodynamic parameters modulate the distinct behavior. Likely there involves a change in the sign of the difference in entropy (ΔS) upon binding.

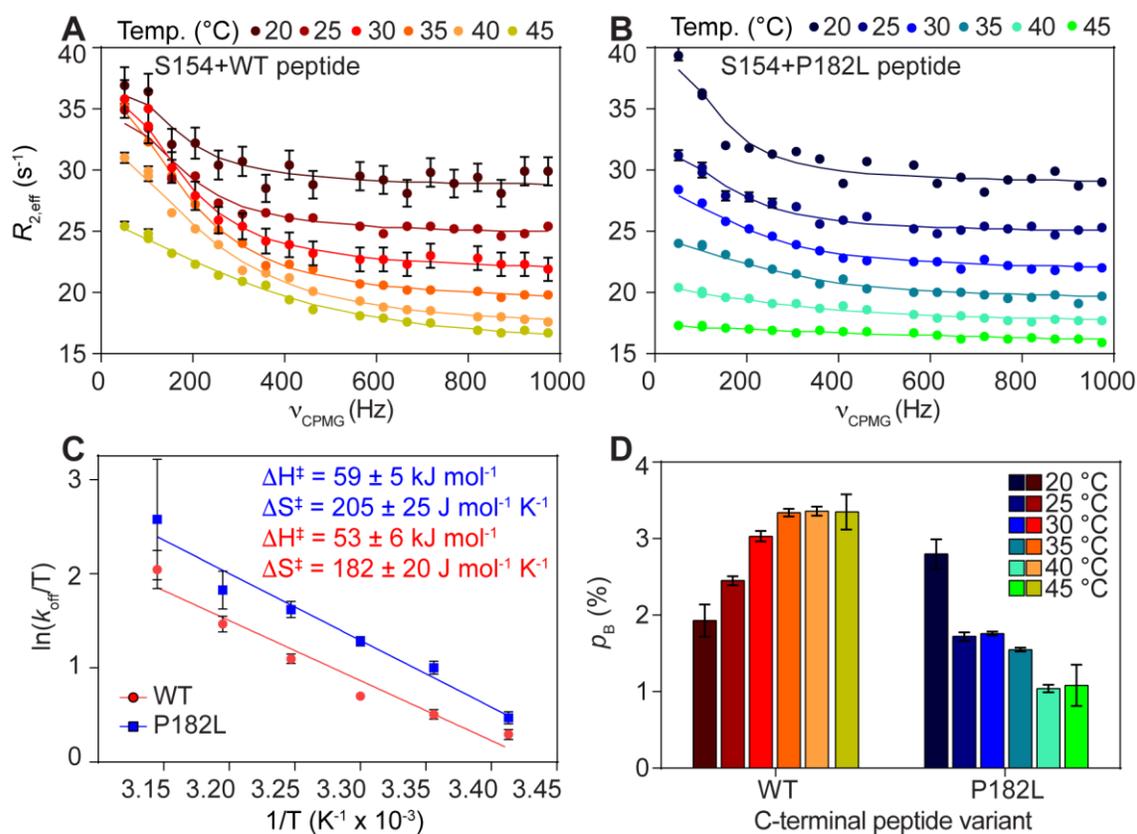


Figure 8. CPMG RD-based measurement of IxI/V binding thermodynamics. (A, B) *i.* ^{15}N CPMG RD data for a residue in the $\beta 8$ strand (S154) in the presence of a small amount of added WT (A) or P182L (B) peptide shown as a function of temperature. (C) Eyring plot ($\ln(k_{\text{off}}/T)$ vs $1/T$) showing the temperature dependence of the off-rate of WT (red) and P182L (blue) peptides. The activation enthalpy and activation entropy are indicated in the upper-right corner. (D) At a fixed concentration of added peptide, the CPMG RD-derived percentage of the peptide-bound chSP27 state (*i.e.* ρ_B) is shown as a function of temperature for both the WT and P182L peptides.

4.3.5 Summary and Conclusions

The work presented in this chapter revealed that the clinically severe P182L mutation in HSP27 significantly increases the average molecular mass of the oligomers (Fig. 2). In my *in vitro* assay, the *ca.* 15 MDa oligomer formed by the P182L variant is also devoid of chaperone activity (Fig. 2). This result is consistent with previous data on the P182S variant (Chalova et al. 2014), although the P182L variant appears to form much larger oligomers. To my knowledge, the P182L variant of HSP27 therefore assembles into the largest sHSP oligomers to-date, surpassing the mass of *E. coli* IbpB (2-3 MDa) (Stróżecka et al. 2012) and sHSP-substrate complexes (several MDa) (Stróżecka et al. 2012).

To provide molecular details regarding the impact of this single amino acid mutation, I utilized NMR spectroscopy (Fig. 3, Fig. 4, Fig. 5) and CPMG RD (Fig. 6, Fig. 7) to quantify the kinetics and thermodynamics of the IxI/V binding to the ACD. The P182L mutation drastically lowers the binding affinity of the IxI/V motif for the ACD (Fig. 5), largely through attenuation of the association rate (Fig. 6). Interestingly, the dissociation rate and CSPs induced by peptide binding to the ACD were similar for both the WT and P182L forms (Fig. 5), suggesting that the mutant ILV motif behaves like the WT IPV motif once bound to the ACD. As compared to the WT form, IxI/V binding in the P182L variant is thermodynamically reversed (Fig. 8), with negative enthalpic and positive entropic contributions to binding. These results, combined with the significantly slower association rate, could indicate that the P182L variant exposes its dynamic CTR to solution more frequently than the WT protein, especially at physiological temperatures.

If the P182L variant decreases the overall binding affinity for the IxI/V motif, how then does the protein form larger oligomers than the WT protein? Previous work on the similar protein ABC established that concomitant binding of two IxI/V motifs facilitates subunit exchange via ejection of a monomeric subunit from the oligomer (Baldwin *et al.* 2011a, 2012). Thus, were the IxI/V binding affinity to be lowered, subunit exchange would occur more slowly and the average oligomer size would increase (Baldwin *et al.* 2011a, 2012). In addition, other HSP27 mutants in the IxI/V motif have been shown to increase the molecular mass, including P182S (Chalova *et al.* 2014) and GPG variants (Gestwicki *et al.* unpublished). Together, these data imply that alteration of the binding properties of the IxI/V motif significantly affects the oligomeric landscape of HSP27. Finally, multiple lines of evidence point to the significance of the NTR in regulating oligomeric assembly, with oligomerization still observed for CTR-truncated forms of HSP27, but not NTR-truncated forms (Hochberg & Benesch 2014).

Given the high level of conservation in the IxI/V motif across various mammalian sHSPs, I anticipate that my results contained herein will be generally applicable to other sHSP systems. Finally, I note that the BAG3 co-chaperone interacts with HSP27 by binding to the latter's IxI/V

motif (Rauch et al. 2017). This chaperone/chaperone interaction promotes new protein-protein interactions, in part by bringing together HSP27 with other BAG3-bound chaperones, including HSP70 (Rauch et al. 2017). Thus, the P182L mutation in HSP27 is expected to also lower the affinity of HSP27 binding to BAG3, and potentially disrupt this chaperone network of interactions. Finally, it is interesting to note that BAG3 itself contains two IxI/V motifs (IPV), one of which contains a Pro-to-Leu mutation implicated in myofibrillar myopathy (Selcen et al. 2008).

4.4 Methods

4.4.1 Proteins and peptides

Peptides were synthesized by Biomatik with N-terminal acetylation and C-terminal amidation. The peptides were dissolved in NMR buffer (30 mM sodium phosphate, 2 mM EDTA, pH 7), with solubility levels around the *ca.* 5 mM range. Despite the addition of pH 7 buffer, the pH of the peptide solutions was found to be in the 4-5 range. This is likely caused by residual trifluoroacetic acid (TFA) used in the peptide purification steps; such low pH values cause complications for titrations into cHSP27, as the protein is highly pH-sensitive (*e.g.* see Chapter 3). As such, the pH of all peptide solutions was corrected to 7 using NaOH.

All proteins were expressed and purified as described in Chapter 3 of this work. WT HSP27 comprises residues 1–205, encompassing the entire amino acid sequence, whereas cHSP27 contains residues 84–171. Importantly, cHSP27 contains an intermolecular disulfide bond involving C137 from adjacent subunits. This disulfide was left intact, so as to minimize contributions from exchange between dimer and monomer. The formation of the disulfide bond is readily identifiable in the 2D ¹H-¹⁵N HSQC spectra of ¹⁵N-cHSP27 via the resonances from R136 and C137; upon reduction, these resonances disappear (*e.g.* see Chapter 3). If cHSP27 were to be reduced or contain the C137S mutation, exchange between the non-covalent dimer and free monomer (*e.g.* see Chapter 3) would lead to additional signal broadening. In addition, the oxidized

dimer simplifies the analysis of titration and CPMG RD data, as only a single state of the protein is present, rather than a mixture of two forms.

Natural abundance proteins were used for chaperone activity assays, EM, and SEC-MALS. ¹⁵N-labeled cHSP27 was used for the CSP studies with WT and P182L peptides. ²H, ¹⁵N-labeled cHSP27 was used for all CPMG studies, with isotope incorporation as described in Chapter 3. The full-length P182L variant of HSP27 was expressed as the WT version, but went into inclusion bodies. The inclusion bodies were solubilized in 8M urea, spun at 20k x g for 20 minutes, and then dialyzed into 30 mM sodium phosphate, 100 mM NaCl, 2 mM EDTA, pH 7 buffer. Some precipitation was evident, and thus the solution was filtered with a 0.22 μm filter to remove particulate matter. The dialysis step could likely be further optimized to minimize losses due to precipitation. The filtrate was then loaded onto a S200 26/60 gel filtration column equilibrated in 30 mM sodium phosphate buffer, 100 mM NaCl, 2 mM EDTA at pH 7. The P182L protein eluted in the column void, earlier than WT HSP27, already indicating its large oligomeric size.

4.4.2 SEC-MALS, negative-stain EM, and chaperone activity assays

SEC-MALS data were conducted and analyzed by Dr. J. M. Louis (NIH) on a Wyatt Technologies instrument using a S200 column (P182L, HSP27) or an S75 column (cHSP27 with the C137S mutation in this case). The instrument was equipped with ASTRA 6 software. In all cases, 250 μg of protein was injected onto the column, which had been equilibrated in 30 mM sodium phosphate, 100 mM NaCl, 2 mM EDTA buffer at pH 7. The calculated molar masses of HSP27, P182L, and cHSP27(C137S) were 470 kDa, 13600 kDa (*i.e.* 13.6 MDa), and 17.5 kDa, respectively. The theoretical mass of the cHSP27(C137S) dimer is 19.8 kDa, so the slightly lower value as determined by SEC-MALS likely reflects contributions from exchange with the free monomer. Note that the C137S mutation does not impact the overall structure of the cHSP27 dimer (*e.g.* see

Chapter 3). The molecular mass of HSP27 is in good agreement with previous publications (Jovcevski et al. 2015, Mishra et al. 2018), with the former reference indicating 670 kDa by analytical SEC and the latter 400 kDa by SEC-MALS. Some sample heterogeneity is therefore to be expected.

Negative-stain EM data were collected by Dr. Marielle Wälti (NIH) with assistance from the author. Samples were prepared at 4.5 μM in 30 mM sodium phosphate, 100 mM NaCl, 2 mM EDTA at pH 7. Samples were loaded onto carbon-coated copper EM grids (Ultrathin Carbon Film/Holey Carbon; Ted Pella) and incubated there for 60 seconds. Excess sample was blotted off the grid, and then the grids were stained with 4% uranyl formate for 15 seconds. Excess stain was removed by blotting. Negative-stain EM images were then collected on a FEI Tecnai T12 electron microscope operating at 120 kV, which was equipped with a Gatan US1000 CCD camera.

Chaperone activity assays were conducted as in Chapter 3. Briefly, all chaperone activity assays were completed in triplicate, with the mean and one standard deviation reported. Assays were performed using a 96-well plate and a FLUOstar Omega Microplate Reader or Tecan Infinite M200 PRO plate reader. 0.2 μM porcine heart MDH (Sigma-Aldrich) was incubated at 40 °C in 30 mM sodium phosphate, 100 mM NaCl, 2 mM EDTA at pH 7. The increase in absorbance at 340 nm was monitored over time in the absence and presence of 0.5 μM WT or P182L HSP27. Likewise, the aggregation of 40 μM human insulin (Sigma-Aldrich) at 40 °C in the same buffer as above was monitored by the increase in absorbance at 340 nm upon the addition of 1 mM DTT. Aggregation reactions were monitored in the absence or presence of 20 μM WT or P182L HSP27.

4.4.3 NMR spectroscopy

All NMR spectra were recorded on a 14.1 T Varian Inova spectrometer equipped with a 5 mm z-axis gradient, triple resonance room temperature probe. Experimental parameters for 2D ^1H - ^{15}N HSQC, ^{15}N T_1 relaxation, ^{15}N T_2 relaxation, $\{^1\text{H}\}$ - ^{15}N heteronuclear NOE, and ^{15}N CPMG RD

spectra were identical to Chapter 3. All NMR data were processed with NMRPipe (Delaglio et al. 1995), visualized with NMRFAM-Sparky (Lee et al. 2015), and peak shapes were fit with FuDA (Vallurupalli et al. 2008).

^{15}N CPMG RD data were recorded on ^2H , ^{15}N -cHSP27 (oxidized) at both 500 and 600 MHz at 298 K, and analyzed as in Chapter 3. WT peptide was added to a final concentration of 80 μM , whereas P182L peptide was added at 200 μM . The expected percentage of bound peptides were calculated as follows:

$$p_{\text{B}}^{\text{calc}} = \frac{K_{\text{d}} + [L]_{\text{t}} + [P]_{\text{t}} - \sqrt{(K_{\text{d}} + [L]_{\text{t}} + [P]_{\text{t}})^2 - 4[L]_{\text{t}}[P]_{\text{t}}}}{2[P]_{\text{t}}} \quad (3)$$

where $[L]_{\text{t}}$ and $[P]_{\text{t}}$ refer to the total concentration of ligand (peptide) and protein (cHSP27), respectively. Thus, at 80 μM of added WT peptide in the presence of 1.5 mM protein for a K_{d} of 125 μM , approximately 4.9% of cHSP27 should be bound. For the P182L peptide at a concentration of 200 μM in the presence of 1.5 mM protein for a K_{d} of 1200 μM , approximately 7.2% of cHSP27 should be bound. These calculations are significantly larger than what was experimentally measured by CPMG RD: 2.0% and 1.7% for WT and P182L, respectively. See below for a possible discussion on why.

Dispersions were fit to a model of two-state chemical exchange using the CATIA software (Vallurupalli et al. 2008). A total of 10 residues each from the WT and P182L peptide datasets were included in the global fit: T110, K112, T113, K114, T151, V153, S154, S156, L157, and S158 (Table 1). In oxidized cHSP27, residues D129, E130, R136, C137, and F138 also show dispersions (e.g. see Chapter 3), but these residues were not included in the fit, as they arise from separate, non-peptide-binding related processes. Between the two CPMG RD datasets derived from WT and P182L peptides, a large degree of similarity in the fitted $|\Delta\omega|$ values is observed (Fig. 9), with an RMSD of 0.27 ppm. Since the $|\Delta\omega|$ values reflect the chemical shift changes upon

binding to peptide, this agreement between the two independently fit datasets further highlights the similarity of the conformations of the peptide-bound states.

Table 1. ^{15}N $|\Delta\omega|$ values in cHSP27 derived from CPMG RD analysis in the presence of a small amount of WT or P182L peptide. Dispersions at 298K at 500 and 600 MHz were included in the global fit; WT and P182L datasets were analyzed independently. The superscripts ^{WT} and ^{mut} refer to values obtained in the presence of WT peptide and P182L peptide, respectively.

Residue	^{15}N $ \Delta\omega^{\text{WT}} $ (ppm)	Error ^{WT} (ppm)	^{15}N $ \Delta\omega^{\text{mut}} $ (ppm)	Error ^{mut} (ppm)
T110	1.45	0.09	1.58	0.07
K112	2.32	0.10	2.40	0.08
T113	4.61	0.16	4.96	0.19
K114	2.51	0.09	2.42	0.09
T151	3.16	0.12	3.03	0.22
V153	1.11	0.08	0.82	0.14
S154	2.69	0.10	2.15	0.10
S156	3.96	0.14	4.32	0.15
L157	2.10	0.09	2.17	0.09
S158	1.27	0.08	1.08	0.07

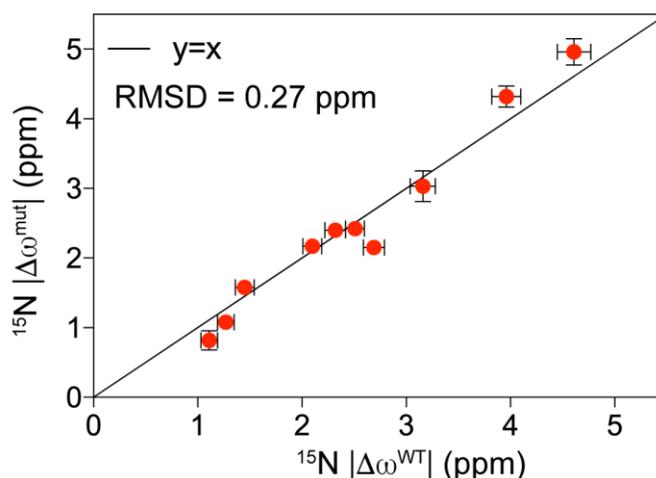


Figure 9. The overall similarity between the independently fit dispersions at 298K, RMSD = 0.27 ppm, suggests that cHSP27 adopts a relatively similar conformation when bound to the WT or P182L peptide. The superscript ^{mut} refers to P182L. Dispersions from 500 and 600 MHz were analyzed.

^{15}N CPMG RD data were then acquired at a range of temperatures on cHSP27 sparsely bound to WT and P182L peptides. These data were obtained at 14.1 T (600 MHz) exclusively at 293, 303, 308, 313, and 318 K. Since only a single B_0 field was employed for the multi-temperature study, $|\Delta\omega|$ values were fixed to their values obtained from the 298 K dataset, which utilized

dispersions at 500 and 600 MHz. This ensures higher accuracy in the fitted k_{ex} and ρ_{B} values (e.g. see Chapter 2), as relatively larger errors can be introduced from fitting CPMG RD data at a single field.

Table 2. Fitted values of k_{ex} and p_{B} derived from CPMG RD data recorded on the sparsely bound states of cHSP27 in the presence of WT or P182L (Mut) peptides. Data at 25 °C were included dispersions at both 500 and 600 MHz, whereas all other temperatures utilized 600 MHz data only. However, the $|\Delta\omega|$ values determined at 25 °C were then fixed in the analysis of all one-field data. Errors indicate +/- one standard deviation from the mean.

T (°C)	WT k_{ex} (s ⁻¹)	WT p_{B} (%)	Mut k_{ex} (s ⁻¹)	Mut p_{B} (%)
20	564 ± 110	1.82 ± 0.24	489 ± 56	2.74 ± 0.24
25	774 ± 44	1.97 ± 0.06	854 ± 48	1.69 ± 0.05
30	685 ± 31	3.03 ± 0.07	1111 ± 40	1.76 ± 0.03
35	1035 ± 51	3.34 ± 0.05	1580 ± 81	1.55 ± 0.03
40	1466 ± 67	3.36 ± 0.06	1967 ± 190	1.04 ± 0.05
45	2542 ± 184	3.35 ± 0.23	4225 ± 817	1.08 ± 0.27

4.4.4 Two- or three-state binding model

While both the CSP (Fig. 3, Fig. 4, Fig. 5) and CPMG RD (Fig. 6, Fig. 7, Fig. 8) data fit well to a model of two-state binding, the potential presence of a third-state is evident from both the discrepancy between p_{B} calculated from theory (eqn 3) and that measured by CPMG RD, and also from incomplete recovery of peak intensities from many resonances in cHSP27 during the peptide titrations. For the CPMG RD data, the measured p_{B} values are lower than the expected values given the K_{d} values and the concentrations of ligand and protein. The back-calculated K_{d} values are significantly higher than those measured by CSPs: namely, for the WT peptide the CPMG-derived K_{d} is 2.51 mM and for the P182L peptide the K_{d} is 10.2 mM. The respective k_{off} and k_{on} values for WT (P182L) are 759 s⁻¹ (840 s⁻¹) and 3.02 x 10⁵ M⁻¹ s⁻¹ (8.26 x 10⁴ M⁻¹ s⁻¹). To convert the fitted k_{ex} values into k_{on} and k_{off} , and calculate the K_{d} , the following equations were used:

$$k_{\text{ex}} = k_{\text{on}}[L]_{\text{free}} + k_{\text{off}} \quad (4)$$

$$[L]_{\text{free}} = [L]_{\text{t}} - p_{\text{B}}[P]_{\text{t}} \quad (5)$$

$$K_d = \frac{k_{\text{off}}}{k_{\text{on}}} \quad (6)$$

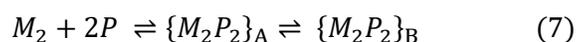
A table of k_{on} , k_{off} , and K_d values derived from CPMG RD at all temperatures is below. As seen, the K_d for the WT peptide decreases with increasing temperature, seemingly plateauing at a value of ca. 850 μM . This value remains ca. 7-fold higher than that determined by CSPs, whereas the CPMG RD-derived K_d is ca. 20-fold larger at lower temperatures. Likewise, the P182L K_d determined by CPMG RD is ca. 8-fold larger than the CSP-derived K_d at 25 °C, further increasing to near 15-fold larger at 40 °C. It should be noted, however, that the CSP-derived K_d values were only obtained at 25 °C, and therefore a full temperature dependence of the K_d is not available.

Table 3. Values of k_{on} ($\text{M}^{-1} \text{s}^{-1}$), k_{off} (s^{-1}), and K_d (μM) as derived from CPMG RD studies on sparsely peptide-bound cHSP27. The K_d refers to the peptide-cHSP27 equilibrium, with k_{off} and k_{on} the dissociation and association rates, respectively. Note that the CSP-derived K_d for WT and P182L peptides are 125 μM and 1200 μM , respectively.

T (°C)	$k_{\text{on}}^{\text{WT}}$ ($\text{M}^{-1} \text{s}^{-1}$)	$k_{\text{off}}^{\text{WT}}$ (s^{-1})	K_d^{WT} (μM)	$k_{\text{on}}^{\text{mut}}$ ($\text{M}^{-1} \text{s}^{-1}$)	$k_{\text{off}}^{\text{mut}}$ (s^{-1})	K_d^{mut} (μM)
20	1.95×10^5	554	2,843	8.43×10^4	475	5,640
25	3.02×10^5	759	2,510	8.26×10^4	840	10,160
30	6.00×10^5	664	1,106	1.12×10^5	1091	9,690
35	1.15×10^6	1000	856	1.39×10^5	1555	11,226
40	1.66×10^6	1416	851	1.10×10^5	1946	17,546
45	2.86×10^6	2457	858	2.48×10^5	4179	16,835

From the CSP titration standpoint, chemical shifts from cHSP27 do not recover their intensities in the presence of 10-fold excess peptide (data not shown). Another potential source of broadening could arise from transient oligomerization. This was ruled out by recording ^{15}N relaxation data, namely measuring T_1 and T_2 relaxation times, in both the absence and presence of near-saturating peptide (95% bound). The fitted τ_c values from ^{15}N T_1/T_2 ratios (Kay et al. 1989) in both cases are consistent with a dimer of 20 kDa in mass at 310 K (ca. 10 ns) (Fig. 10). As such, I speculate that multiple conformations of the peptide-bound form may exist, with significant conformational interconversion on the millisecond timescale, as discussed below.

This discrepancy in binding affinity as measured by CSPs and CPMG RD could arise from a third state. The source of the broadening and lower than expected p_B values could be significant heterogeneity of the peptide when bound to cHSP27, leading to multiple interconverting states on the millisecond timescale. However, the fact that both CSP and CPMG RD data are well-fit to a model of two-state chemical exchange thus implies that the third state rapidly interconverts with one of the two states probed here. Therefore, a plausible binding model may look like one of the schemes below:



where the second arrows would have very fast rates associated with them.

To confirm or reject the two-state model, I plan to globally fit the CSP, linewidth broadening, and CPMG RD data together, as outlined previously ([Korzhev et al. 2009](#)), to a model of three-state chemical exchange, and examine if the discrepancy can be resolved.

4.5 Acknowledgements

I thank Ad Bax for valuable input on the project and Jenny E. Hinshaw (NIH) for allowing usage of the electron microscope. I also thank my collaborators Bob Asselbergh, Elias Adriaenssens, and Vincent Timmerman (University of Antwerp) who collected expansion microscopy data on the P182L and WT variants of HSP27, and confirmed that P182L forms large oligomers *in vivo* (not shown here).

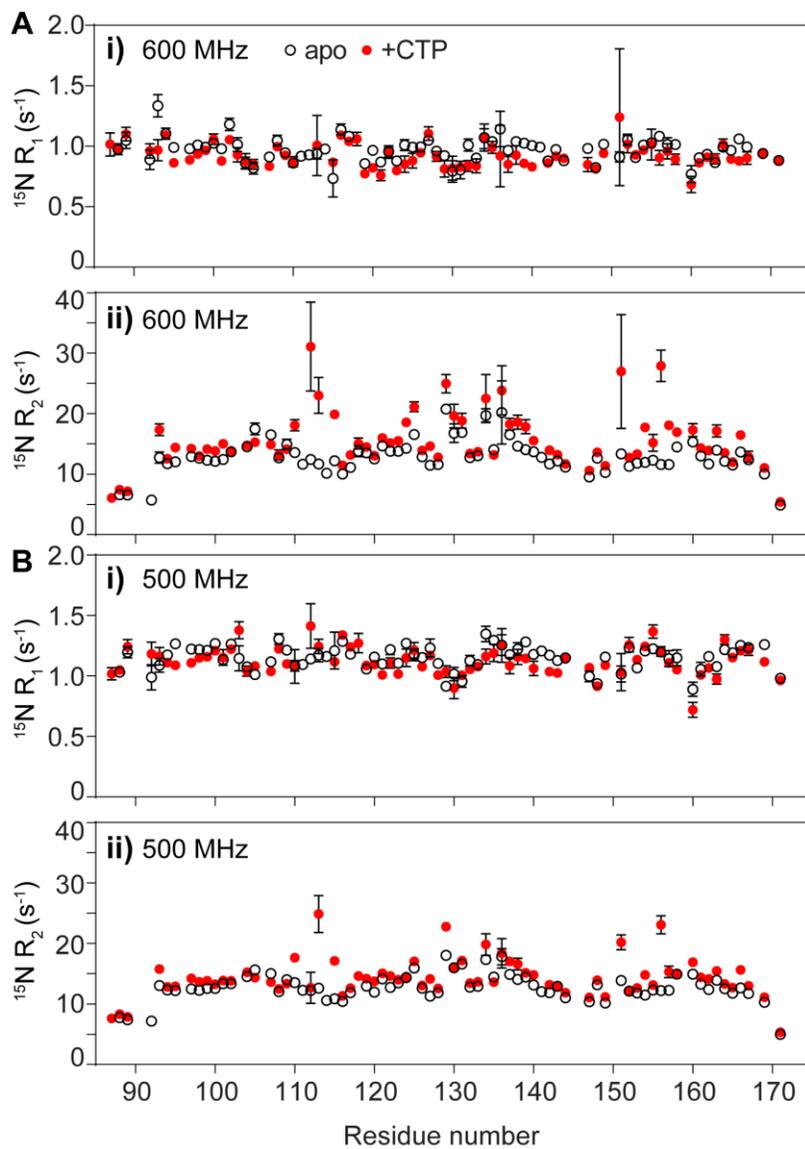


Figure 10. ^{15}N relaxation of peptide-free and peptide-bound cHSP27. ^2H , ^{15}N -labeled cHSP27 was prepared at a concentration of 0.3 mM in 30 mM sodium phosphate, 2 mM EDTA buffer at pH 7 and 37 °C. ^{15}N R_1 (i) and ^{15}N R_2 (ii) relaxation rates were measured at 600 MHz (A) and 500 MHz (B). In the figure, apo means peptide-free cHSP27 and +CTP refers to the peptide-bound form. To generate the peptide bound form, 3.8 mM WT peptide (CTP) was added to 0.3 mM cHSP27. For a K_d of 125 μM , this corresponds to 96.8% of bound cHSP27. The elevated R_2 values in A i and B i reflect contributions due to chemical exchange.

4.6 References

- Alderson TR, Benesch JLP, Baldwin AJ. 2017. Proline isomerization in the C-terminal region of HSP27. *Cell Stress Chaperones*. 22(4):639–51
- Arrigo AP. 2001. HSP27: novel regulator of intracellular redox state. *IUBMB Life*. 52(6):303-07
- Baldwin AJ, Hilton GR, Lioe H, Bagn ris C, Benesch JLP, Kay LE. 2011a. Quaternary dynamics of α B-crystallin as a direct consequence of localised tertiary fluctuations in the C-terminus. *J. Mol. Biol.* 413(2):310–20
- Baldwin AJ, Lioe H, Robinson C V, Kay LE, Benesch JLP. 2011b. α B-crystallin polydispersity is a consequence of unbiased quaternary dynamics. *J. Mol. Biol.* 413(2):297–309
- Baldwin AJ, Walsh P, Hansen DF, Hilton GR, Benesch JLP, et al. 2012. Probing dynamic conformations of the high-molecular-weight α B-crystallin heat shock protein ensemble by NMR spectroscopy. *J. Am. Chem. Soc.* 134(37):15343–50
- Barisic N, Claeys KG, Sirotkovi -Skerlev M, L fgren A, Nelis E, et al. 2008. Charcot-Marie-Tooth Disease: A Clinico-genetic Confrontation. *Ann. Hum. Genet.* 72(3):416–41
- Basha E, O’Neill H, Vierling E. 2012. Small heat shock proteins and α -crystallins: dynamic proteins with flexible functions. *Trends Biochem. Sci.* 37(3):106–17
- Bruey JM, Ducasse C, Bonniaud P, Ravagnan L, Susin SA, et al. 2000. Hsp27 negatively regulates cell death by interacting with cytochrome c. *Nat. Cell Biol.* 2(9):645–52
- Chalova AS, Sudnitsyna M V, Strelkov S V, Gusev NB. 2014. Characterization of human small heat shock protein HspB1 that carries C-terminal domain mutations associated with hereditary motor neuron diseases. *Biochim. Biophys. Acta.* 1844(12):2116–26
- D’Ydewalle C, Krishnan J, Chiheb DM, Van Damme P, Irobi J, et al. 2011. HDAC6 inhibitors reverse axonal loss in a mouse model of mutant HSPB1-induced Charcot-Marie-Tooth disease. *Nat. Med.* 17(8):968–74
- Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A. 1995. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR.* 6(3):277–93
- Delbecq SP, Jehle S, Klevit R. 2012. Binding determinants of the small heat shock protein, α B-crystallin: recognition of the ‘Ixl’ motif. *EMBO J.* 31(24):4587–94
- Demers J-P, Mittermaier A. 2009. Binding Mechanism of an SH3 Domain Studied by NMR and ITC. *J. Am. Chem. Soc.* 131(12):4355–67
- DiVincenzo C, Elzinga CD, Medeiros AC, Karbassi I, Jones JR, et al. 2014. The allelic spectrum of Charcot-Marie-Tooth disease in over 17,000 individuals with neuropathy. *Mol. Genet. Genomic Med.* 2(6):522–29
- Dohrn MF, Gl ckle N, Mulahasanovic L, Heller C, Mohr J, et al. 2017. Frequent genes in rare diseases: panel-based next generation sequencing to disclose causal mutations in hereditary neuropathies. *J. Neurochem.*
- Echaniz-Laguna A, Geuens T, Petiot P, P r on Y, Adriaenssens E, et al. 2017. Axonal Neuropathies due to Mutations in Small Heat Shock Proteins: Clinical, Genetic, and Functional Insights into Novel

Mutations. *Hum. Mutat.* 38(5):556–68

- Evgrafov O V, Mersyanova I, Irobi J, Van Den Bosch L, Dierick I, et al. 2004. Mutant small heat-shock protein 27 causes axonal Charcot-Marie-Tooth disease and distal hereditary motor neuropathy. *Nat. Genet.* 36(6):602–6
- Gusev NB, Bogatcheva N V., Marston SB. 2002. Structure and Properties of Small Heat Shock Proteins (sHsp) and Their Interaction with Cytoskeleton Proteins. *Biochem.* 67(5):511–19
- Hansen DF, Vallurupalli P, Kay LE. 2008a. An Improved 15N Relaxation Dispersion Experiment for the Measurement of Millisecond Time-Scale Dynamics in Proteins. *J. Phys. Chem. B.* 112(19):5898–5904
- Hansen DF, Vallurupalli P, Lundström P, Neudecker P, Kay LE. 2008b. Probing Chemical Shifts of Invisible States of Proteins with Relaxation Dispersion NMR Spectroscopy: How Well Can We Do? *J. Am. Chem. Soc.* 130(8):2667–75
- Haslbeck M, Vierling E. 2015. A first line of stress defense: small heat shock proteins and their function in protein homeostasis. *J. Mol. Biol.* 427(7):1537–48
- Hilton GR, Hochberg GKA, Laganowsky A, McGinnigle SI, Baldwin AJ, Benesch JLP. 2013. C-terminal interactions mediate the quaternary dynamics of α B-crystallin. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 368(1617):20110405
- Hochberg GK a., Ecroyd H, Liu C, Cox D, Cascio D, et al. 2014. The structured core domain of β -crystallin can prevent amyloid fibrillation and associated toxicity. *Proc. Natl. Acad. Sci.* 111(16):E1562–70
- Hochberg GK a, Benesch JLP. 2014. Dynamical structure of α B-crystallin. *Prog. Biophys. Mol. Biol.* 115(1):11–20
- Jehle S, Rajagopal P, Bardiaux B, Markovic S, Kühne R, et al. 2010. Solid-state NMR and SAXS studies provide a structural basis for the activation of α B-crystallin oligomers. *Nat. Struct. Mol. Biol.* 17(99):1037–42
- Jovcevski B, Kelly MA, Rote AP, Berg T, Gastall HY, et al. 2015. Phosphomimics Destabilize Hsp27 Oligomeric Assemblies and Enhance Chaperone Activity. *Chem. Biol.* 22(2):186–95
- Kampinga HH, de Boer R, Beerstra N. 2015. *The Multicolored World of the Human HSPB Family*, Vol. 8. Cham: Springer International Publishing
- Kay LE, Torchia DA, Bax A. 1989. Backbone dynamics of proteins as studied by 15N inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease. *Biochemistry.* 28(23):8972–79
- Korzhev DM, Bezsonova I, Lee S, Chalikian T V., Kay LE. 2009. Alternate Binding Modes for a Ubiquitin–SH3 Domain Interaction Studied by NMR Spectroscopy. *J. Mol. Biol.* 386(2):391–405
- Lee W, Tonelli M, Markley JL. 2015. NMRFAM-SPARKY: enhanced software for biomolecular NMR spectroscopy. *Bioinformatics.* 31(8):1325–27
- McDonald ET, Bortolus M, Koteiche HA, Mchaourab HS. 2012. Sequence, structure, and dynamic determinants of Hsp27 (HspB1) equilibrium dissociation are encoded by the N-terminal domain. *Biochemistry.* 51(6):1257–68
- Mishra S, Chandler SA, Williams D, Claxton DP, Koteiche HA, et al. 2018. Engineering of a Polydisperse Small Heat-Shock Protein Reveals Conserved Motifs of Oligomer Plasticity. *Structure.* 26(8):1116–1126.e4

- Nefedova V V., Muranova LK, Sudnitsyna M V., Ryzhavskaya AS, Gusev NB. 2015. Small Heat Shock Proteins and Distal Hereditary Neuropathies. *Biochem.* 80(13):1734–47
- Palmer AG, Kroenke CD, Loria JP. 2001. Nuclear magnetic resonance methods for quantifying microsecond-to-millisecond motions in biological macromolecules. *Methods Enzymol.* 339(1997):204–38
- Rajagopal P, Liu Y, Shi L, Clouser AF, Klevit RE. 2015. Structure of the α -crystallin domain from the redox-sensitive chaperone, HSPB1. *J. Biomol. NMR.* 63(2):223–28
- Rauch JN, Tse E, Freilich R, Mok S-A, Makley LN, et al. 2017. BAG3 Is a Modular, Scaffolding Protein that physically Links Heat Shock Protein 70 (Hsp70) to the Small Heat Shock Proteins. *J. Mol. Biol.* 429(1):128–41
- Rossor AM, Polke JM, Houlden H, Reilly MM. 2013. Clinical implications of genetic advances in Charcot–Marie–Tooth disease. *Nat. Rev. Neurol.* 9(10):562–71
- Selcen D, Muntoni F, Burton BK, Pegoraro E, Sewry C, et al. 2008. Mutation in BAG3 causes severe dominant childhood muscular dystrophy. *Ann. Neurol.* 65(1):83–89
- Stróżecka J, Chrusciel E, Górna E, Szymanska A, Ziętkiewicz S, Liberek K. 2012. Importance of N- and C-terminal regions of IbpA, Escherichia coli small heat shock protein, for chaperone function and oligomerization. *J. Biol. Chem.* 287(4):2843–53
- Sugase K, Lansing JC, Dyson HJ, Wright PE. 2007. Tailoring Relaxation Dispersion Experiments for Fast-Associating Protein Complexes. *J. Am. Chem. Soc.* 129(44):13406–7
- Vallurupalli P, Hansen DF, Kay LE. 2008. Structures of invisible, excited protein states by relaxation dispersion NMR spectroscopy. *Proc. Natl. Acad. Sci.* 105(33):11766–71
- Waudby CA, Ramos A, Cabrita LD, Christodoulou J. 2016. Two-Dimensional NMR Lineshape Analysis. *Sci. Rep.* 6(1):24826

5

Development of pressure-jump hydrogen exchange for protein folding

5.1 Abstract

A method is introduced that permits direct observation of the rates at which backbone amide hydrogens become protected from solvent exchange after rapidly dropping the hydrostatic pressure inside the NMR sample cell from denaturing (2.5 kbar) to native (1 bar) conditions. The method is demonstrated for a ubiquitin variant that contains two Val to Ala mutations to create small hydrophobic cavities, which make the protein more susceptible to pressure-induced unfolding while retaining its native structure. Increased protection against hydrogen exchange with solvent is monitored as a function of time during the folding process. Results for 53 backbone amides show narrow clustering with protection occurring with a time constant of *ca.* 85 ms, but slower protection is observed around a reverse turn near the C-terminus of the protein. Remarkably, the native NMR spectrum returns with this slower time constant of *ca.* 150 ms, indicating that the almost fully folded protein retains molten globule characteristics with severe NMR line broadening until the final hydrogen bonds are formed. Prior to crossing the transition state barrier, hydrogen exchange protection factors are close to unity, but with slightly elevated values in the $\beta 1$ – $\beta 2$ hairpin, previously shown to be already lowly populated in the urea-denatured state.

5.2 Introduction

Rapid mixing experiments, where proteins are switched from a denaturing to a folding environment, have long been coupled with studies of the rates at which amide protons exchange with solvent deuterons, or vice versa, to capture information on the time sequence involved in hydrogen bond (H-bond) formation during the protein folding process (Dyson & Wright 2017, Englander & Mayne 1992). In these so called quenched-flow hydrogen exchange (HX) experiments, samples are harvested and subsequently analyzed by NMR or mass spectrometry (MS) in order to determine the degree to which a given residue (NMR) or group of residues (MS) has undergone hydrogen exchange with solvent during the mixing process (Bai et al. 1995, Baldwin 1993, Miranker et al. 1993). These experiments have provided valuable insights into the time course at which individual amides become engaged in H-bond formation, and thereby protected from exchange with solvent, revealing remarkable details on kinetic intermediates that can occur during the folding pathway (Briggs & Roder 1992, Dyson & Wright 2017, Englander & Mayne 1992, Liu et al. 2014, Panick et al. 1999). Recently, a ^{13}C -detected NMR experiment to monitor HX rates in arginine guanidinium groups has been introduced (Mackenzie & Flemming Hansen 2018), as well as methods to measure HX rates in invisible protein states (Long et al. 2014, Yuwen et al. 2018).

Here, I introduce a more direct NMR method for monitoring the time course of HX protection during the folding process. By unfolding the protein under high hydrostatic pressure inside the NMR sample cell, followed by a rapid drop in pressure to initiate the protein folding process, the rate at which hydrogens exchange with solvent can be monitored directly by exploiting the resonance frequency difference between amide protons and bulk solvent. Conceptually, these experiments are analogous to pressure-jump fluorescence measurements (Liu et al. 2014, Panick et al. 1999), but the combination with two-dimensional NMR spectroscopy provides a separate spectroscopic probe at every backbone amide moiety. Here, I provide the

first such example of following the switch between a fully pressure-denatured state to the fully folded state with millisecond resolution, using the change in HX rate during the folding process as the experimentally probed parameter. I note, however, that it is also feasible to probe multiple other parameters during the folding process, including NOEs, chemical shifts, relaxation rates, and possibly residual dipolar couplings.

Over the past 25 years, NMR spectroscopy has explored extensively proteins under high hydrostatic pressure, and has provided site-specific information on folding intermediates as well as thermodynamic parameters ([Akasaka et al. 2000](#), [Nash & Jonas 1997](#), [Royer et al. 1993](#), [Samarasinghe et al. 1992](#), [Yamaguchi et al. 1995](#)). Most of these studies were carried out under static pressure conditions, but more recent studies have also utilized a sudden change in pressure to explore kinetic behavior ([Kremer et al. 2011](#), [Roche et al. 2013](#)). Here, I show that rapidly and repeatedly switching the pressure in the NMR sample cell between 2.5 kbar and 1 bar allows atomic-level characterization of the protein at millisecond resolution during the actual folding process.

I measure τ_f , the mean time between the initiation of protein folding (when pressure drops from 2.5 kbar to 1 bar) and the time at which individual backbone amides become protected from HX. Protection from HX typically occurs because a given amide has either engaged in a stable H-bond or has otherwise become solvent inaccessible due to the formation of locally folded structure. The ability to denature a protein at high pressure is critical to the experiments outlined below. The negative change in volume between the folded and unfolded states is the primary cause of pressure-induced protein unfolding. This mechanism can be magnified by introducing small hydrophobic cavities in the interior of the protein through mutagenesis. Here, two Val to Ala mutations (V17A/V26A) are used to lower the mid-point of ubiquitin pressure denaturation from >5 kbar to *ca* 1.4 kbar at 298 K. The structure of ubiquitin is minimally affected by these mutations, as indicated by the very similar values of the structurally sensitive $^{13}\text{C}^\alpha$ and $^{13}\text{C}'$ chemical shifts of the wild type and double mutant protein ([Charlier et al. 2018](#)). The midpoint of pressure

denaturation can also be lowered by decreasing the temperature (Panick et al. 1999), but this results in much slower kinetics of both folding and unfolding.

5.3 Results and Discussion

Following the rapid drop in pressure from 2.5 kbar to 1 bar, and a concomitant 3 K drop in temperature due to adiabatic solvent expansion, the return of the pressure-denatured protein to its native, folded state can be monitored by simply recording 1D NMR spectra (Fig. 1, Fig. 2). A plot of the intensity of the most downfield shifted amides (I13 and L67) as a function of refolding time shows bi-exponential behavior with about two thirds refolding with a time constant of *ca* 150 ms, and one third refolding more than an order of magnitude slower (Fig. 1). The same time constants apply to all other well resolved amide resonances, indicating that the I13 and L67 intensities are representative of the final, native state of the protein. The slowly recovering component is attributable to an off-pathway oligomeric intermediate, with subsequent work demonstrating a clear concentration-dependence to the fraction of the slow phase. The off-pathway oligomeric species contributes minimally to my measurements of HX protection, which only observe signals from proteins that fold on the faster time scale.

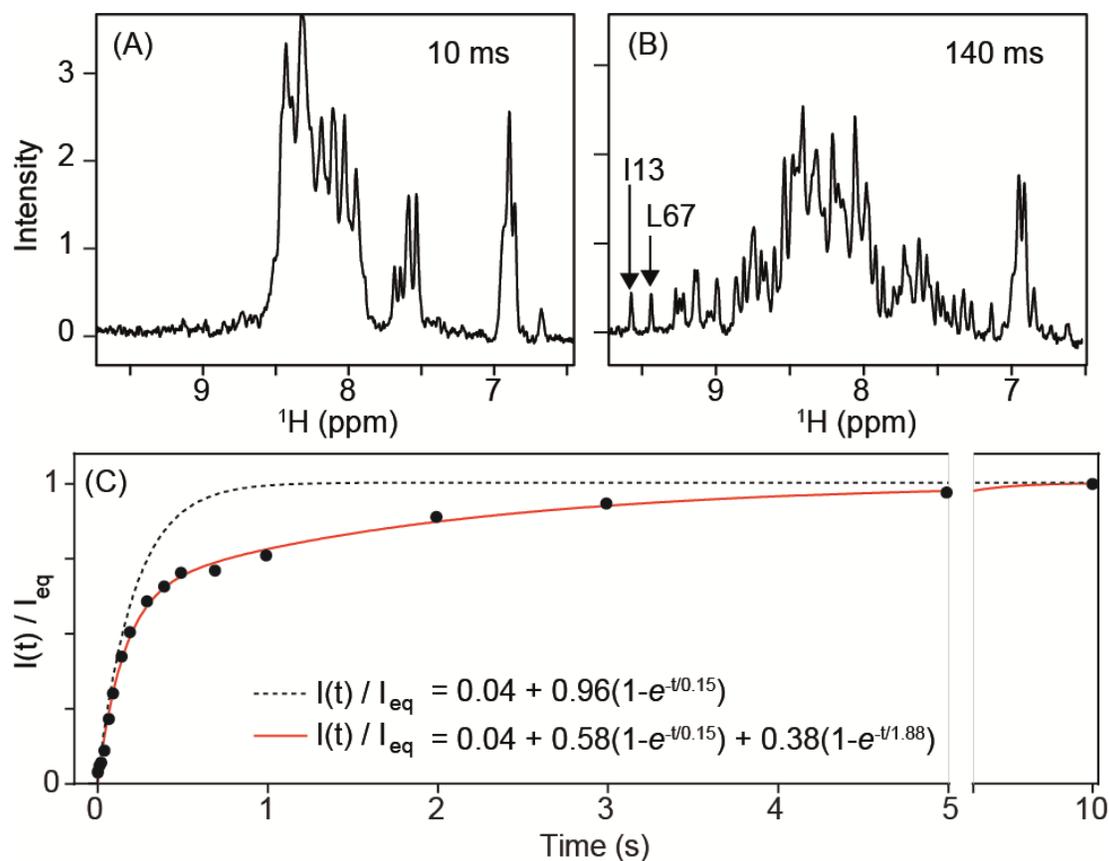


Figure 1. Recovery of the ubiquitin amide signals after dropping the pressure from 2.5 kbar to 1 bar. (A,B) 1D NMR spectra recorded at different times after the pressure drop. A larger set of such spectra is shown in Figure 1. (C) Plot of the ratio of the time-dependent intensity ($I(t)$) of the two most downfield shifted amides (I13 and L67) over their equilibrium value (measured 10 s after pressure drop, I_{eq}). The 0.04 offset in the exponential fits is the fraction of protein, measured in a separate high S/N HSQC spectrum, that remains folded at 2.5 kbar and 298 K. Reproduced from the American Chemical Society with permission.

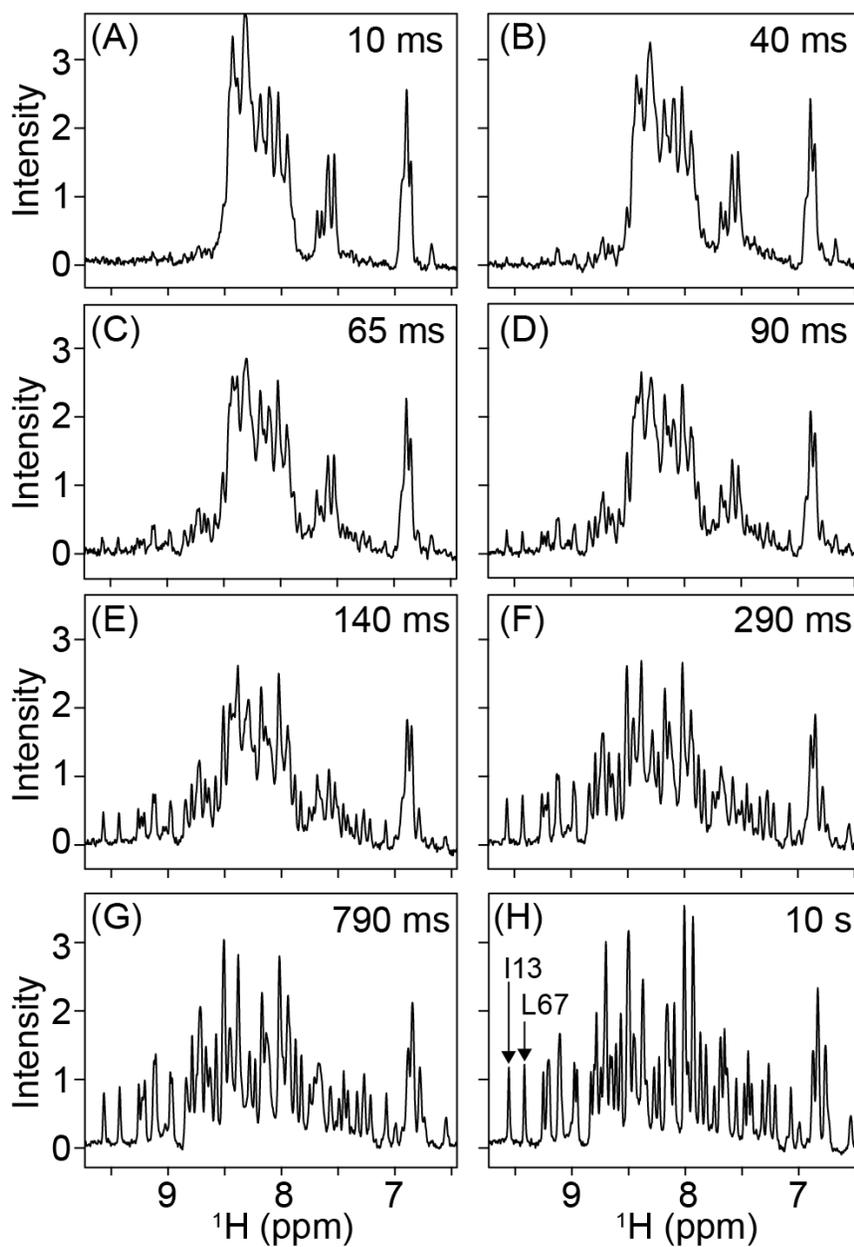


Figure 2. 1D NMR spectra of ubiquitin after dropping the pressure from 2.5 kbar to 1 bar to initiate folding. Sample: 0.3 mM $^{15}\text{N}/^{13}\text{C}/^2\text{H}$ V17A/V26A ubiquitin in 250 μL $\text{H}_2\text{O}/\text{D}_2\text{O}$ (93/7%), 25 mM phosphate buffer, pH 7.5, 295 K (at 1 bar). (A-H) 1D NMR spectra recorded at different times after the pressure drop. Reproduced from the American Chemical Society with permission.

In the fully unfolded state, the intrinsic HX rates (k_{intr}) depend strongly on residue type and that of the immediate neighbors (Bai et al. 1993). Values of k_{intr} at 2.5 kbar were measured using

0.3 mM uniformly $^{13}\text{C}/^{15}\text{N}/^2\text{H}$ -enriched ubiquitin. Perdeuteration avoided any complications from potential NOE transfer of magnetization from H^α protons, and the absence of NOE effects through rapidly exchanging hydroxyl protons was verified by carrying out measurements at three pH values: 6.4, 7.1, and 7.5.

To measure HX with bulk solvent during the folding process, I essentially follow the procedure of Grzesiek and Fitzkee (Fitzkee et al. 2011, Grzesiek & Bax 1993). In my experiments, a duration $T_{\text{ex}} = 125$ ms, where water magnetization is either inverted or left unchanged, precedes the start of the 2D ^1H - ^{15}N heteronuclear single quantum correlation (HSQC) experiment. A refolding delay, τ_r , separates the start of the T_{ex} period from the time when the pressure suddenly drops to 1 bar (Fig. 3). Each pulse sequence is executed twice: once with a water-selective inversion pulse and a subsequent water-flip-back pulse immediately prior to the start of the HSQC, and a second, reference experiment where these water inversion pulses are not applied. The ratio between corresponding intensities in the two HSQC spectra therefore provides a measure for the fraction of any amide hydrogen in the ensemble that exchanged with solvent during T_{ex} . Amide hydrogens that rapidly exchange with solvent ($k_{\text{ex}} \geq \sim 1/T_{\text{ex}}$) will appear inverted relative to the reference spectrum, while those that exchange more slowly will simply be attenuated (Fig. 4, Fig. 5).

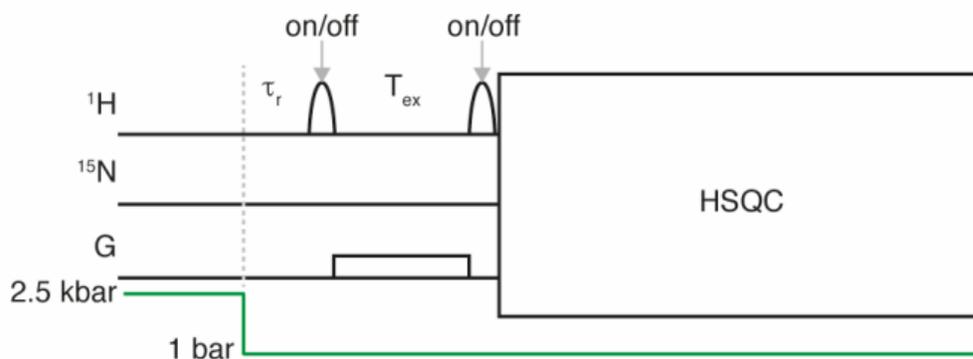


Figure 3. Schematic diagram of the pressure jump NMR experiment used to measure HX rates during protein folding. Hydrostatic pressure is marked in green. G marks a weak magnetic field gradient (0.5 G/cm), needed to prevent radiation damping of the inverted H_2O magnetization. HSQC refers to the standard,

gradient-enhanced HSQC experiment,¹¹ including water flip-back to +z. A long delay (12 s at 298 K and 2.5 kbar) between scans was used to unfold the protein prior to the pressure drop. Reproduced from the American Chemical Society with permission.

Using the experiment of Figure 2 ($\tau_r = 0$), the difference in intensity of a given amide signal, recorded without and with water inversion, normalized to the signal without water inversion, $S_f(T_{ex})$, was derived by D.A. Torchia by extending the derivations of Grzesiek and Fitzkee (Fitzkee et al. 2011, Grzesiek & Bax 1993) to include non-equilibrium chemical exchange (Helgstrand et al. 2000). In the limit that R_{1u} (unfolded state R_1), R_{1f} (folded state R_1), R_{1W} (water R_1), and k_f (HX rate in the folded state) are all much smaller than $1/T_{ex}$, $S_f(T_{ex})$ simplifies to:

$$S_f(T_{ex}) = \Delta f \{1 - [\lambda / (\lambda + k_u)] [(1 - e^{-(\lambda + k_u)T_{ex}}) / (1 - e^{-\lambda T_{ex}})]\} \quad (1)$$

where Δf denotes the difference in H₂O magnetization before and after the selective 180° pulse ($\Delta f \approx 1.8$); λ is the inverse of τ_i , the mean time at which a given amide becomes HX-protected; k_u is the HX rate at 1 bar, prior to crossing the transition state; and T_{ex} is the duration that the water remains inverted. Note that extracting λ from eq 1 requires knowledge of k_u , which may well differ from k_{intr} , the intrinsic, random coil exchange rate (see below).

Nevertheless, the scheme of Fig. 3 can be used to detect τ_f provided that $k_u \gg k_f$. By inserting a refolding delay, τ_r , prior to the first water inversion pulse, a fraction $e^{-\tau_r/\tau_f}$ of the amides remain in the HX-unprotected state prior to the first water inversion pulse. S_f therefore scales as $S_f(\tau_r, T_{ex}) = S_f(0, T_{ex})e^{-\tau_r/\tau_f}$ and since it must approach the value obtained at 1 bar for $\tau_r \rightarrow \infty$ $S_f(\tau_r, T_{ex})$ is well described by the function $Ae^{-\tau_r/\tau_f} + B$.

Spectra were collected in an interleaved manner for six τ_r values: 0, 50, 100, 200, 300, and 400 ms. For $\tau_r = 0$, a number of amide signals appear inverted due to rapid exchange with solvent during the T_{ex} period when water is inverted, whereas others are strongly attenuated (Fig. 5). However, when HX is impeded by protein folding during the delay τ_r , attenuation due to HX with solvent becomes progressively less (Fig. 5), and $S_f(\tau_r, T_{ex})$ follows the expected $Ae^{-\tau_r/\tau_f} + B$

dependence (Fig. 5, Fig. 6). Residues that are not significantly protected in the folded protein (e.g., G10-T12) will show very little change in intensity when τ_r is increased.

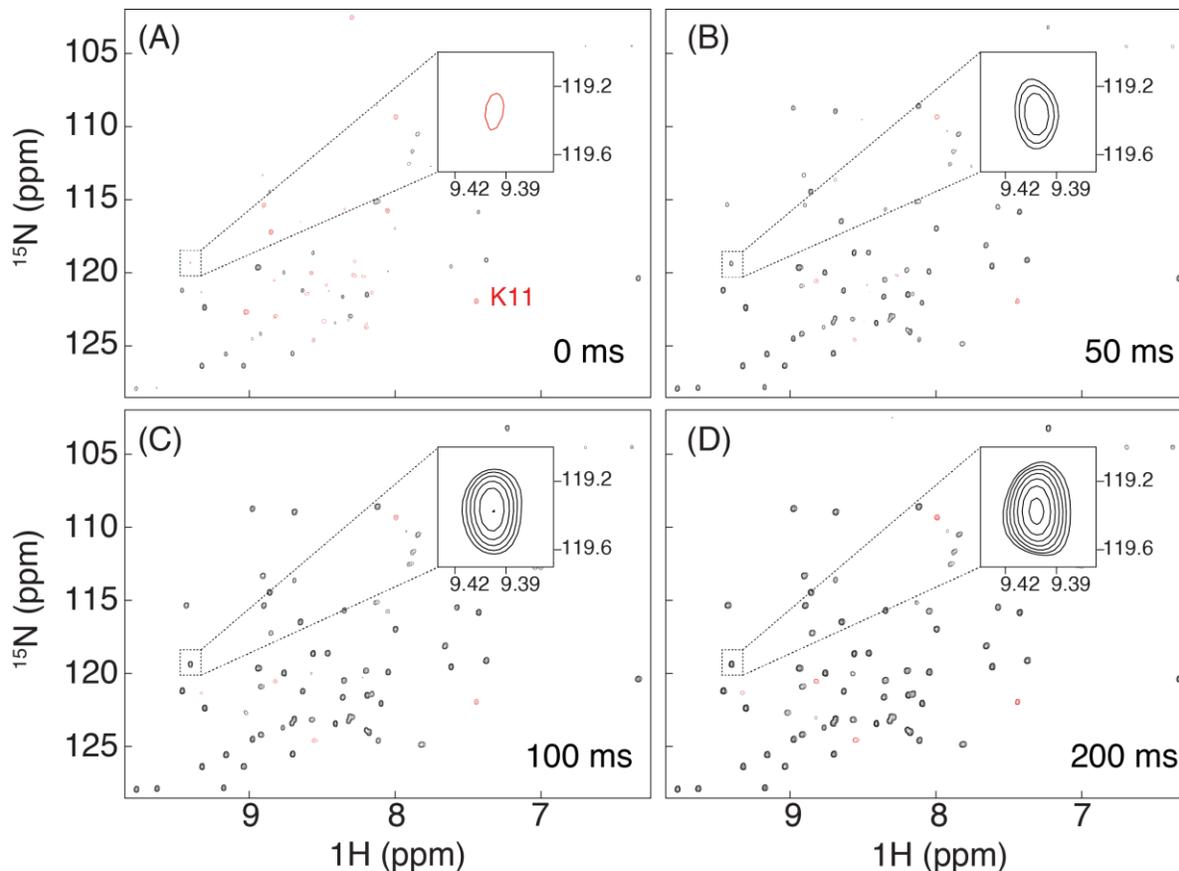


Figure 4. Ubiquitin 2D HSQC NMR spectra, measured at various refolding intervals, τ_r , with application of the water inversion pulses. The inset corresponds to the cross peak of H68. Negative signals are in red. The assignment of K11, representative of a residue with fast HX in the folded state ($k_f = 5.4 \text{ s}^{-1}$), is marked in (A). Reproduced with permission from the American Chemical Society.

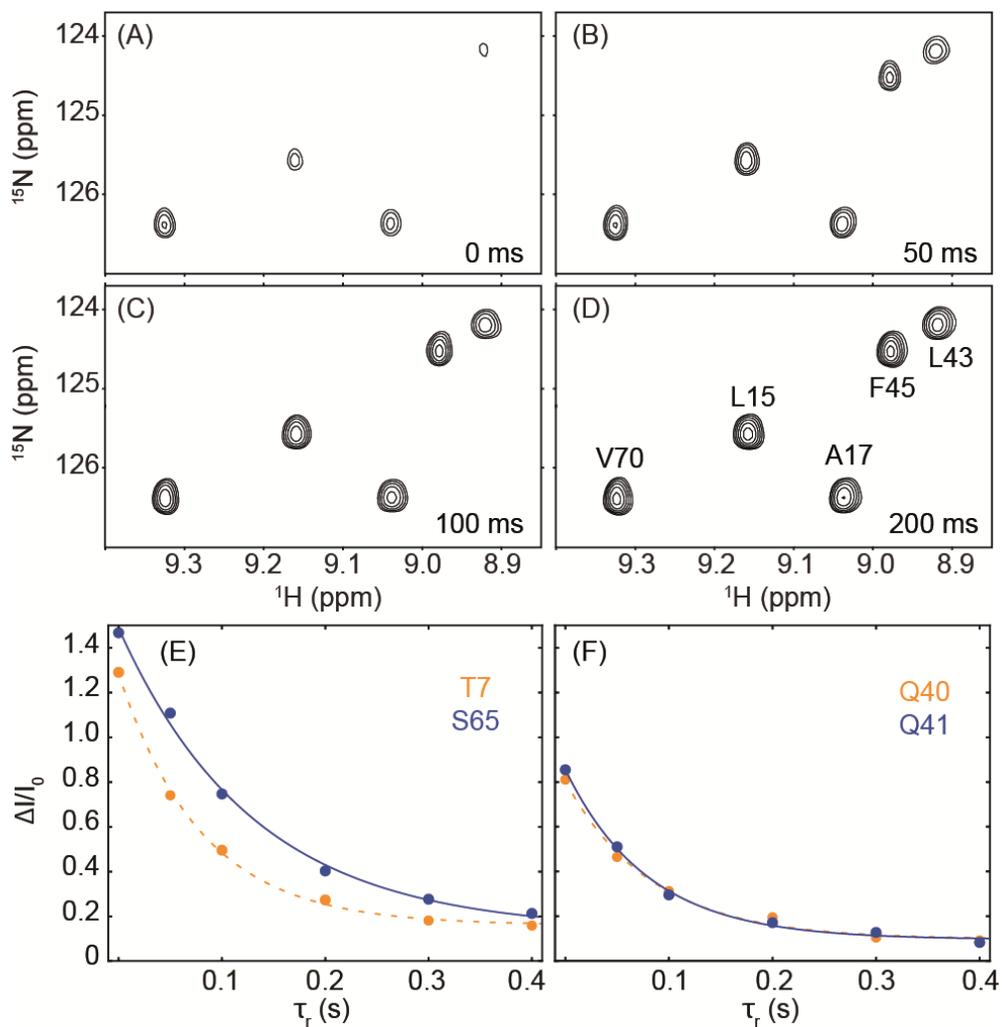


Figure 5. Refolding time (τ_r) dependence of NMR intensities. (A-D) Sections of ubiquitin 2D HSQC NMR spectra (with application of the water inversion pulses), measured at various τ_r values. Sample conditions are as in Fig. 1 and Fig. 2. The sample temperature was set at 298 K during the high pressure (2.5 kbar) period (12 s). (E,F) Fits of the $S_i(\tau_r, T_{ex})$ to $Ae^{-\tau_r/\tau_f} + B$, for T7, S65, Q40, and Q41. Reproduced with permission from the American Chemical Society.

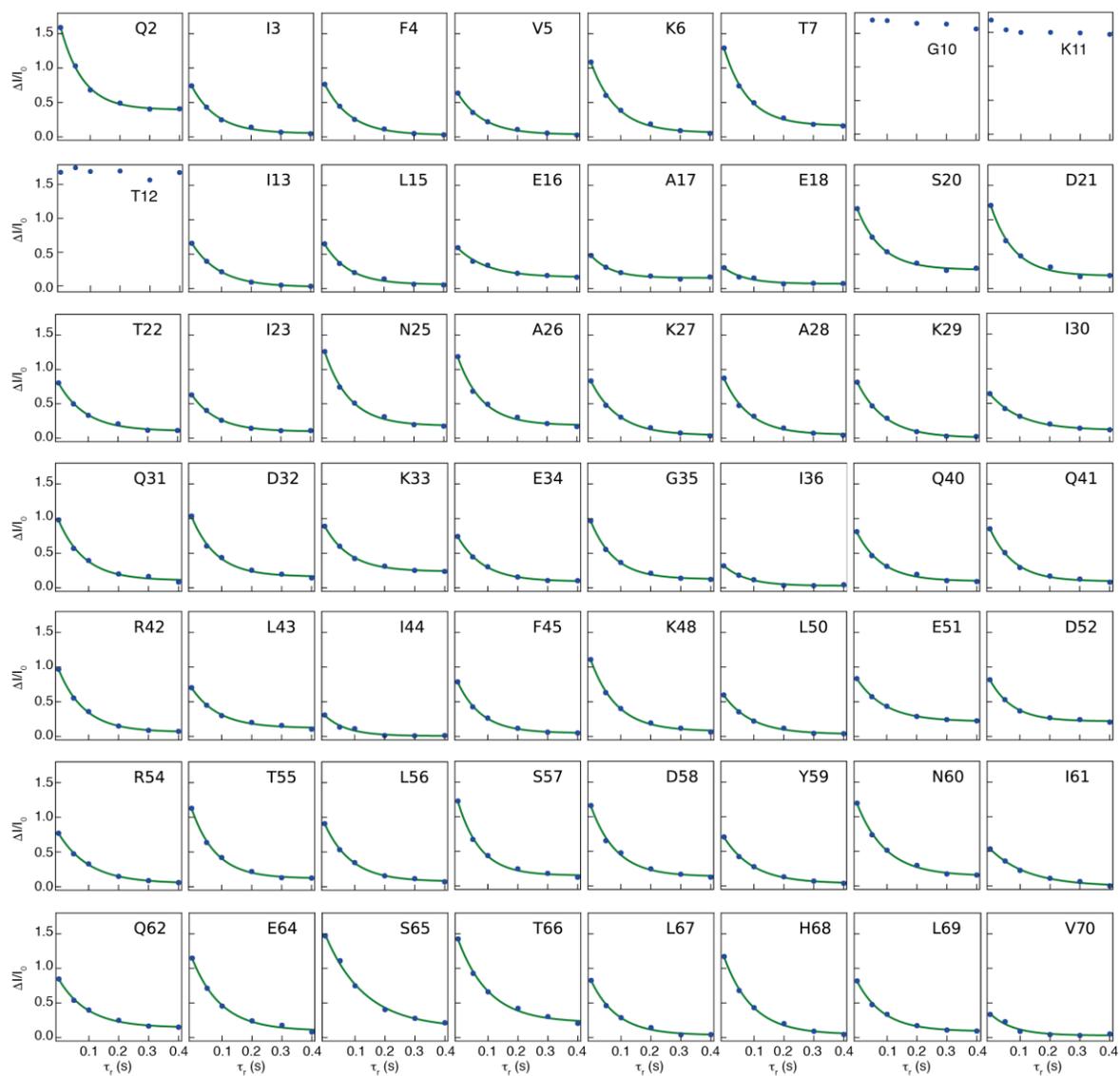


Figure 6. Fits of $S_f(\tau_r, T_{ex}) = \Delta I(\tau_r)/I_0$ to $Ae^{-\tau_r/\tau_f} + B$ for all of the residues analyzed. The sample contained 300 μ M of V17A/V26A ubiquitin (^2H , ^{13}C , ^{15}N) dissolved at pH 7.5 thermostated at 298 K when at 2.5 kbar. An exchange time (T_{ex}) of 125 ms was used in this experiment, with τ_r values of 0, 50, 100, 200, 300, and 400 ms. Intensities of residues with fast HX in both the unfolded and folded states (G10, K11, T12) show no significant decay and were not fitted. Reproduced with permission from the American Chemical Society.

Values for τ_f could be extracted for 53 amides whose HX rates differ substantially between the unfolded and folded states of the protein (Fig. 6, Fig. 7). Time constants for the entire β -sheet (86 ± 8 ms; $N = 18$) and the long α -helix (88 ± 6 ms; $N = 11$) show narrow clustering (Fig. 7, Fig. 8). The highly homogeneous set of time constants, seen for most of the protein, gives it the

appearance of being a classic two-state folder, in agreement with previous denaturant-based studies (Krantz & Sosnick 2000, Went et al. 2004). However, the loop/turn region (I61–T66) shows longer and more heterogeneous time constants, most notably for I61 ($\tau_f = 126$ ms) and S65 ($\tau_f = 131$ ms) (Fig. 7, Fig. 8). The amide of I61 shares an H-bond to the carbonyl of L56, and S65 H-bonds to Q62, stabilizing a type-II β -turn (Fig. 8). These observations indicate that protein folding is not fully complete until the final type-II β -turn, centered at residues K63-E64, has formed. Remarkably, prior to the formation of this turn, the NMR spectrum of the entire protein is impacted by dynamic processes that cause severe line broadening, essentially rendering it invisible, until this last native element has formed (Fig. 1, Fig. 2).

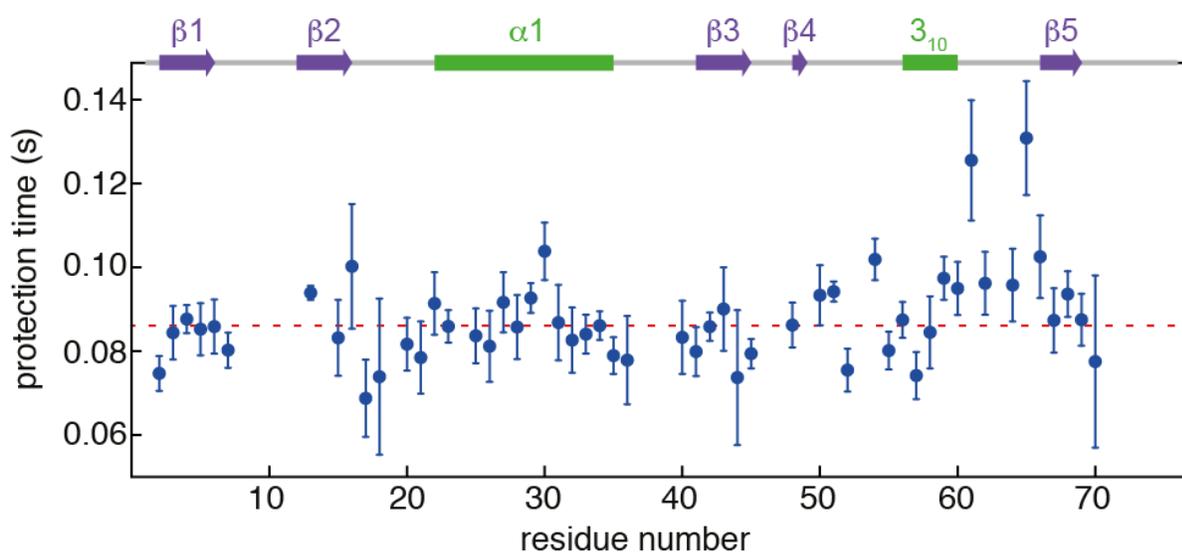


Figure 7. Protection times, $\tau_f = 1/\lambda$ for HX during protein folding. Fitted time constants as a function of residue number in the V17A/V26A ubiquitin mutant at 295 K. Error bars reflect the uncertainty as obtained from the covariance matrix of the fit. A color map of the rates on the ubiquitin ribbon diagram highlights spatial proximity of I61 and S65 (Fig. 8). Reproduced with permission from the American Chemical Society.

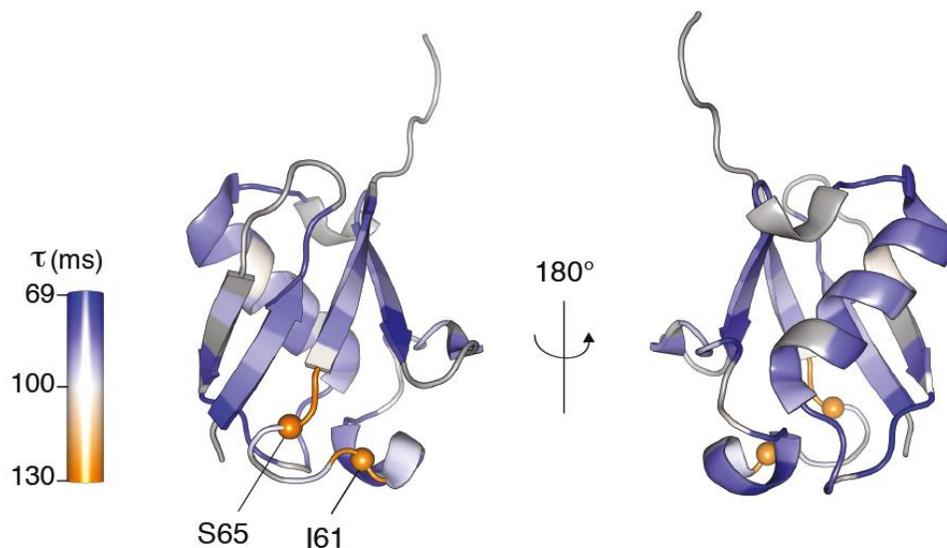


Figure 8. Protection times, $\tau_f = 1/\lambda$ for HX during protein folding mapped on the ribbon diagram of the ubiquitin structure (PDB entry 2MJB). Grey indicates residues for which no protection times were obtained. Reproduced with permission from the American Chemical Society.

This penultimate state during the folding process therefore resembles what is sometimes called a “dry molten globule”, where the vast majority of backbone H-bonds are properly in place, but dynamic conformational exchange prevents coalescence to a sharply defined folded structure (Baldwin et al. 2010, Neumaier & Kiefhaber 2014, Shakhnovich & Finkelstein 1989). The NMR resonances are exceptionally sensitive to such dynamic processes, effectively broadening them to an extent that they escape detection, in particular when recording 2D HSQC spectra that involve fixed magnetization transfer delays.

Besides the fast folding component, the double-mutant ubiquitin also shows a much slower folding component (Fig. 1): about two thirds fold at a relatively rapid rate of ca 7 s^{-1} , whereas one-third folds at a rate that is more than an order of magnitude slower and is not significantly probed by my HX experiments. A similar biphasic recovery in ubiquitin has previously been attributed to the fraction of unfolded chains for which one or more of the three Pro residues starts in the *cis* conformation (Briggs & Roder 1992). However, as shown later, the fraction of *cis*-Pro in

unfolded VA2-ubq is 2% and 10% for the bonds involving Pro38 and Pro39 (Alderson et al. 2018); thus, the slower time constant may be caused by an oligomeric off-pathway intermediate (Went et al. 2004), caused by aggregation of the unfolded chain that exposes a large number of hydrophobic residues. Indeed, a pronounced increase in the fraction of slowly recovering protein is observed when the sample concentration is raised from 0.3 mM to 1 mM, confirming the presence of an oligomeric species. This was investigated in more detail in a follow-up publication (Charlier et al. 2018). This off-pathway oligomerization therefore limits the concentration at which pressure jump NMR experiments of the type described here can be carried out. Alternatively, working at a pH where the protein carries a large net charge, or the use of a small volume fraction of a mild organic co-solvent such as ethanol, can mitigate the self-association process (Eliezer et al. 1998, Redfield 2004).

My results provide the residue-specific time constants at which the HX rate switches from that of the unfolded to that of the folded state. Their measurements therefore do not require a high degree of protection in the folded state, and indeed time constants are also obtained for residues that do not engage in H-bonding in the final structure, simply reflecting the change in accessibility to OH⁻ catalyzed hydrogen exchange. For example, the solvent-exposed amides of the β 2 strand yield time constants very similar to their immediate H-bonded neighboring residues (Fig. 7, Fig. 8). However, the smaller difference in HX rate between the folded and unfolded states makes their uncertainty somewhat larger. The pressure jump approach therefore provides access to a larger number of probes than can be evaluated by quenched-flow rapid mixing experiments.

My pressure jump experiment also yields information on the average degree of HX protection prior to crossing the transition state: to a first approximation, for $\tau_r = 0$ and k_f , R_{1W} , R_{1u} , and $R_{1f} \ll 1/T_{ex}$, the HX rate of the protein at 1 bar, k_u , can be extracted from eq 1. Comparison of k_u with the intrinsic HX rates, k_{intr} , measured at high pressure in the fully unfolded state, shows a close correlation with a slope of ca 1 (Fig. 9), indicative of essentially no HX protection prior to

the point where the protein crosses the transition state. A possible exception to this observation is seen for residues in the first two β -strands, which account for five of the six highest $k_{\text{intr}}/k_{\text{u}}$ ratios, ranging from 1.75 to 1.85 (Fig. 9). A caveat in determining these ratios is that k_{intr} values were measured at 2.5 kbar, and conversion to 1 bar requires scaling to account for the pressure dependence of the phosphate buffer pH, lowering it by approximately 1 unit, i.e. 10-fold lower $[\text{OH}^-]$. This effect is largely, but not completely, offset by the five-fold increased water dissociation at 2.5 kbar which, combined with the pH effect, results in a net lowering of the effective $[\text{OH}^-]$ concentration, and therefore HX rates, by about 2-fold relative to 1 bar (Fuentes & Wand 1998, Zhang et al. 1995).

The small degree of HX protection seen for amides involved in formation of the first β -hairpin is consistent with the transient presence of this structural element even in the urea-denatured protein, albeit to substantially less than 25% (Meier et al. 2007). Its formation is expected to be essentially instantaneous ($\sim 6 \mu\text{s}$) on the time scale of my NMR experiment, judging by measurements on the analogous hairpin of the structurally homologous GB1 protein (Muñoz et al. 1997). Regardless, my observation indicates that in my ubiquitin mutant, prior to crossing the transition state, dynamic H-bonds are remarkably lowly populated, in agreement with prior work by Gladwin and Evans (Gladwin & Evans 1996).

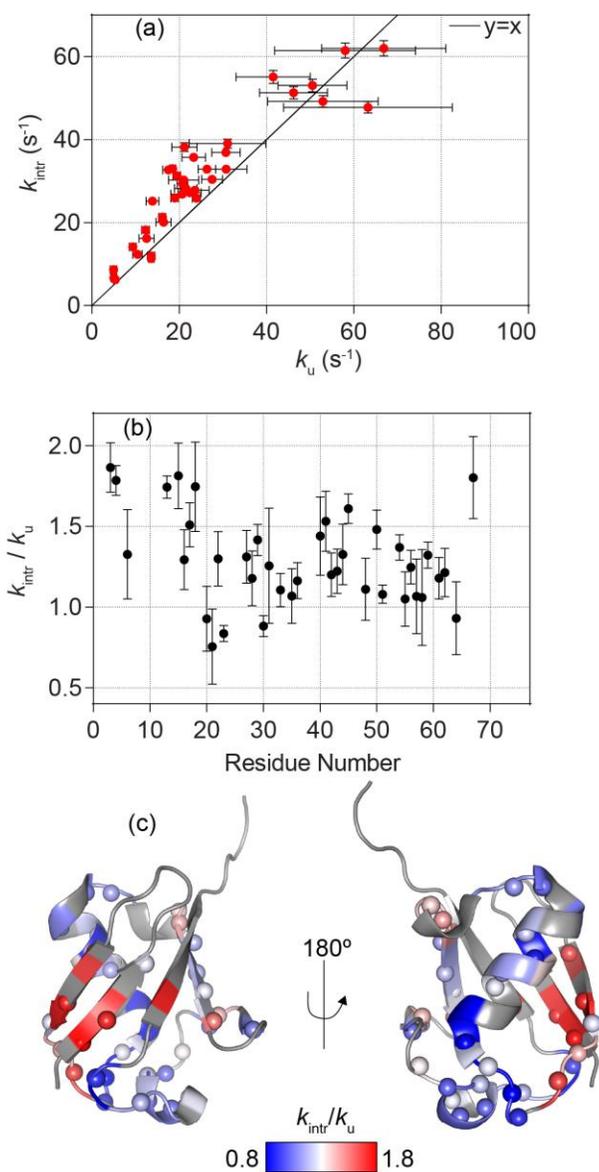


Figure 9. Comparison of the intrinsic HX rates (k_{intr}) in the pressure-denatured state of V17A/V26A ubiquitin with the HX rates in the pre-transition state (k_u) measured during the pressure-jump experiments. (a) Correlation plot of the pressure-corrected k_{intr} rates (rates multiplied by 2.0 (to account for the 1 unit decrease in the pH of phosphate buffer, prepared at pH 7.5, when pressure is raised to 2.5 kbar, an effect that is partially offset by a five-fold increase in OH⁻ concentration resulting from pressure-induced water ionization) and k_u rates determined from the pressure-jump experiments. In both cases, the sample contained 300 μM of ²H, ¹³C, ¹⁵N-labeled sample, in 25 mM sodium phosphate buffer, pH 7.5 at 1 bar. Static experiments at 2.5 kbar were carried out at 295 K (k_{intr}). Pressure jump experiments (k_u) were carried out with the 2.5 kbar sample thermostated at 298 K, corresponding to 295K after the pressure drop to 1 bar. (b) The ratio of k_{intr} over k_u shown as a function of residue number. (c) Data from panel (b) shown on the structure of ubiquitin (PDB 2mjb). Residues that are colored gray have no data available. Reproduced with permission from the American Chemical Society.

5.4 Methods

5.4.1 Protein expression and purification

A codon-optimized plasmid encoding residues 1-76 of human ubiquitin bearing the V17A/V26A mutations (ATUM) was transformed into *E. coli* BL21(DE3) competent cells (New England Biolabs). All bacterial media for protein expression contained 100 µg/L of ampicillin. Following overnight growth at 37 °C in 5 mL of LB media, the culture was spun down and transferred to 100 mL of M9 minimal medium. This culture grew at 37 °C for 3-5 hours, and then the cells were pelleted and transferred to 1 L of M9 minimal media containing 1 g L⁻¹ of ¹⁵NH₄Cl and either 2 g L⁻¹ of [*U*-¹³C]-glucose or 6 g L⁻¹ of natural abundance glucose. For perdeuteration, the M9 minimal media was prepared in 99.8% D₂O and [*U*-²H,¹³C]-glucose was used. All media for isotopic labeling was supplemented with 1 g L⁻¹ of appropriately isotopically labeled ISOGRO (Sigma-Aldrich). When the optical density of these cultures reached ca 0.6-0.8 units, protein expression was initiated by addition of 1 mM IPTG and allowed to proceed for 4 hours at 37 °C. The cells were then harvested and stored at -80 °C until used.

Cell pellets were resuspended in 50 mM Tris-HCl buffer, pH 7.6, and then lysed using a Cell Disruptor (Constant Systems Ltd), with multiple passages of the cells at 2.2 kbar. The lysed cells were then spun at 50,000 x g for 20 minutes at 4 °C. The supernatant was collected and the pH was brought to 3.3 using acetic acid in order to precipitate contaminants. The pH 3.3 sample was then spun at 50,000 x g for 20 minutes at 4 °C, and the supernatant was concentrated and loaded onto a Superdex S75 26/60 column (GE Healthcare) equilibrated in 10 mM potassium buffer at pH 6.8. The ubiquitin-containing fractions were collected, concentrated, and then subjected to reverse-phase HPLC using a Vydac 214TP C4 column. Ubiquitin V17A/V26A eluted near 35% acetonitrile with 0.1% trifluoroacetic acid, and this sample was then dialyzed twice into four liters of 25 mM potassium phosphate buffer (pH 6.4), concentrated, and stored until use. All samples for NMR spectroscopy contained 7% D₂O in 25 mM potassium phosphate buffer at pH

6.4, 7.1, or 7.5. The ionic strength was slightly changed in the elevated pH samples due to the addition of NaOH.

5.4.2 NMR spectroscopy

The unfolded state of ubiquitin V17A/V26A was assigned using non-uniformly sampled 3D HNCA and HNCO spectra acquired on a Bruker Avance III 800 MHz spectrometer equipped with a cryogenically cooled probe. The pressure inside the NMR tube was maintained at 2.5 kbar (Daedalus Innovations, LLC), and NMR spectra were processed and NUS-reconstructed with NMRPipe (Delaglio et al. 1995) and SMILE (Ying et al. 2016). All hydrogen exchange (HX) NMR spectra were collected on a Bruker Avance III 600 MHz spectrometer equipped with a 3-axis gradient, room temperature probe, modified with a home-built apparatus enabling fast pressure changes (see below). The pulse sequence to measure HX during the refolding process is analogous to that of Fitzkee et al. (Fitzkee et al. 2011).

5.4.3 Pressure jump apparatus

The detailed design and performance characteristics of a home-built pressure-jump apparatus will be reported elsewhere, with some details already published (Charlier et al. 2018). Briefly, a ceramic NMR tube is connected via a high-pressure transfer tube to a “T” fitting mounted above the magnet. Pneumatic valves connected to the “T” fitting open to an adjustable high-pressure oil reservoir (≤ 3 kbar) or an atmospheric pressure vessel. Spectrometer-generated TTL signals trigger the valve openings to alternately pressurize and depressurize the protein sample. Pressure rise and fall times can be adjusted from ca 1 ms to several ms by restricting flow to the NMR tube. Reported data used switching times of ca. 4 ms.

5.4.4 Pressure-induced changes to the pK_w of water

To calculate the pK_w of water analytically, such that the pK_w can be obtained at any pressure and temperature, the following equation was used (Bandura & Lvov 2006, Dooley 2007):

$$pK_w^G = \gamma_0 + \gamma_1 T^{-1} + \gamma_2 T^{-2} + \gamma_3 T^{-3} \quad (2)$$

$$pK_w = -2n [\log_{10} 1 + Q] - \frac{Q}{Q+1} \rho (\beta_0 + \beta_1 T^{-1} + \beta_2 \rho) + pK_w^G + 2 \log \frac{m_0 M_w}{G} \quad (3)$$

$$Q = \frac{\rho}{\rho_0} \exp \left(\alpha_0 + \alpha_1 T^{-1} + \alpha_2 T^{-2} \rho^{\frac{2}{3}} \right) \quad (4)$$

where γ_n , α_n , and β_n ($n = [0,3]$) are empirically derived factors, T is the temperature, n is the ion coordination number, ρ is the mass density, ρ_0 is the normalizing mass density, m_0 is the absolute molality, M_w is the molar mass of water, and G is a conversion factor for g to kg. The actual values utilized for the calculation are presented in Table 1.

The density of water (ρ) changes with temperature and pressure, and thus must be calculated explicitly to avoid errors that propagate into the pK_w of water. The dependence on ρ with pressure and temperature is:

$$\rho(P) = \frac{\rho_0}{1 - \frac{\Delta P}{K}} \quad (5)$$

$$\rho(T) = \frac{\rho_0}{1 + \Delta T \alpha} \quad (6)$$

where ΔP is the difference in pressure ($P - P_0$), K is the bulk modulus of water, ΔT is the difference in temperature ($T_0 - T$), and α is the coefficient of thermal expansion. For a given P and T combination, the density of water is then calculated as $\rho(T)\rho(P)$. The variation in the bulk modulus of water with pressure and the coefficient of thermal expansion with temperature were determined empirically by the author through fitting tabulated values (Fig. 10). The solid lines in Fig. 10 represent best-fit lines to a linear equation ($y = mx + b$, panel A) or an exponential ($y = A \exp(-x/T) + B$, panel B) for K and α , respectively.

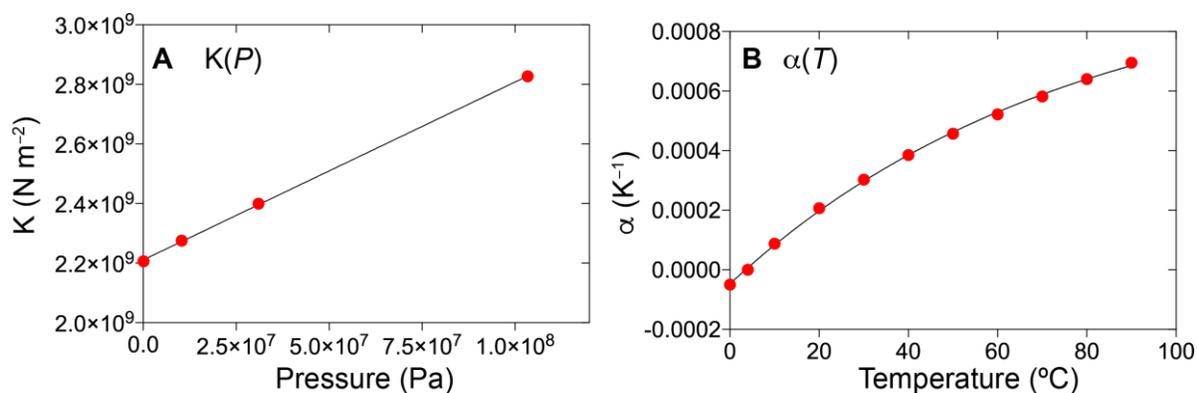


Figure 10. Variations in the bulk modulus of water (A) and coefficient of thermal expansion (B) with pressure and temperature, respectively. The red data points represent tabulated values and black lines are best-fit lines derived from a linear equation ($K = 5.971 P + 2210681814$) or an exponential equation ($\alpha = -0.001053 \exp(-P/75.63) + 0.001006$).

Failure to account for the changes in K and α yielded incorrect values of the pK_w versus pressure at $T=298$ K (Fig. 11, top). The black dotted line represents eqn 3 calculated with $\rho(T)$ and $\rho(P)$ using $K = K_0$ and $\alpha = \alpha_0$, neglecting other changes. The correct value for the pK_w of water at 2500 bar is 13.265, and failure to correct for changes in K and α yielded a pK_w of 13.05. Inclusion of the empirically derived parameters in turn yields a pK_w of 13.265 (Fig. 11, bottom).

5.4.5 Pressure-induced changes to the pK_a of phosphoric acid

In this study, phosphate buffer was employed, which is a polyprotic acid where the second protonation event has a pK_a value (7.21) in the range of physiological pH values. The first and third protonation events have pK_a values far outside this regime, 2.15 and 12.35, respectively, are neglected in the henceforth analysis. pK_a values are the negative logarithm of the associated equilibrium constant (K), with ΔG related to an equilibrium constant through the following expression:

$$\Delta G = -RT \ln K \quad (7)$$

For a two-state transition at fixed temperature, changes in pressure, compressibility, and volume impact the free energy as follows:

$$\Delta G = \frac{\Delta\beta}{2}(P - P_0)^2 + \Delta V(P - P_0) + \Delta G_0 \quad (8)$$

where ΔG stands for the difference in free energy at pressure P , ΔG_0 for the difference in free energy at the initial pressure P_0 , $\Delta\beta$ for the difference in compressibility, and ΔV for the difference in volume. Thus, in the presence of an appreciably large negative change in volume or compressibility, high pressures can alter the thermodynamic equilibrium and therefore the relative proportions of two different states. Typically for protein folding/unfolding studies, the ΔV associated with unfolding ranges from -20 to $-100 \text{ cm}^3 \text{ mol}^{-1}$ due to the collapse of hydrophobic cavities and $\Delta\beta$ has a much smaller negative value, on the order of 10^{-3} to $10^{-4} \text{ cm}^3 \text{ mol}^{-1} \text{ bar}^{-1}$. The pK_a value for the second protonation event of phosphoric acid ($\text{HPO}_4^{2-} \leftrightarrow \text{H}_2\text{PO}_4^-$) has a ΔV of $-25.85 \text{ cm}^3 \text{ mol}^{-1}$ and $\Delta\beta$ of $-40.9 \times 10^{-4} \text{ cm}^3 \text{ mol}^{-1} \text{ bar}^{-1}$. This means that the equilibrium constant associated with the protonation of phosphoric acid is pressure-dependent, resulting in a change in pH of the buffer with increasing pressures. The pK_a for the second protonation event of phosphoric acid is defined as pK_2 :

$$pK_2 = -\log\left(\frac{[\text{H}^+][\text{HPO}_4^{2-}]}{[\text{H}_2\text{PO}_4^-]}\right) \quad (9)$$

where the equilibrium constant K_2 has a value of 6.2×10^{-8} at 1 bar and 298 K, yielding a ΔG value of 41.1 kJ mol^{-1} . At 2500 bar, however, the ΔG value is lowered to 35.9 kJ mol^{-1} , resulting in an increase of K_2 to 5.0×10^{-7} . This manifests as a decrease in the pH of the phosphate buffer with pressure, as the pH can be calculated from the Henderson-Hasselbalch equation:

$$pH = pK_2 + \log\left(\frac{[\text{HPO}_4^{2-}]}{[\text{H}_2\text{PO}_4^-]}\right) \quad (10)$$

for a phosphate buffer prepared at $pH_0 = 6.40$ at 1 bar, where the amount of added $[\text{HPO}_4^{2-}]:[\text{H}_2\text{PO}_4^-]$ is equal to 0.15488, the effective pH at 2500 bar becomes 5.49 – a decrease of 0.89 pH units. This calculation is consistent with previous analyses, which indicated that phosphate buffer decreases by ca. -0.4 pH units every 100 MPa (1000 bar) (Gayán et al. 2013).

This decrease in pH with increasing pressure impacts the measured HX rates accordingly. In the pH regime where HX is catalyzed by hydroxide ions, the measured rates scale according to the following changes in pH and hydroxide concentration:

$$HX_B = HX_A \Delta[OH^-]10^{\Delta pH} \quad (11)$$

where Δ stands for the difference between condition B and condition A. Therefore, HX scales linearly with hydroxide concentration and logarithmically with pH.

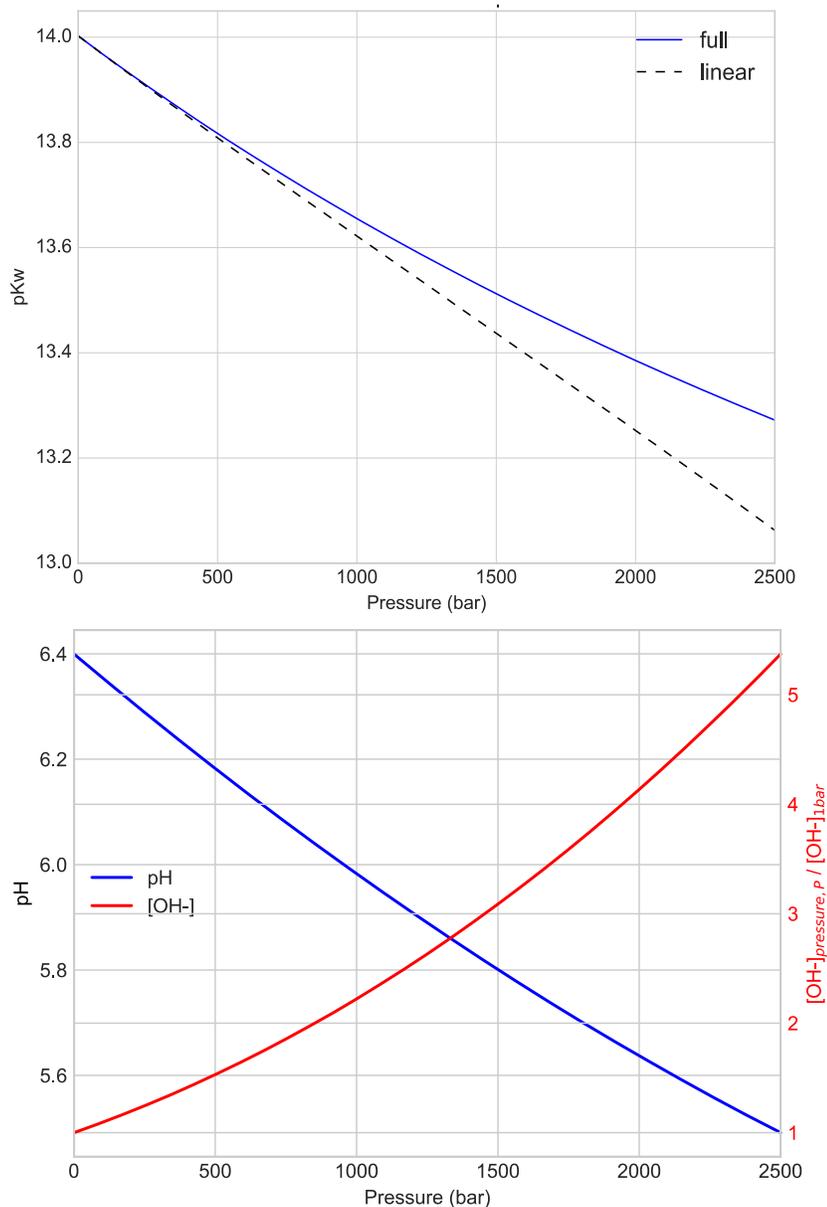


Figure 11. Variation in the pKw of water, pH of a phosphate buffer, and relative OH- concentration as a function of pressure at 298 K. (*Top*) Calculation of the pKw of water using eqn 3 and correcting for the change in K and α with pressure and temperature (blue line) or by scaling relative to K_0 and α_0 (black line, dotted). The correct value for the pKw of water at 298 K and 2500 bar is 13.265 (Dooley 2007). (*Bottom*) Change in the pH of phosphate buffer with pressure. Given the pKw of water at each pressure and the calculate change in pH of a phosphate buffer ($\text{pH}_0 = 6.4$ at 1 bar), the OH- concentration is calculated is relative to that at 1 bar. At 2500 bar, the concentration of hydroxide ions is 5.37-fold increased relative to 1 bar. Values for pKw are plotted on the left y-axis, whereas values of relative OH- concentration are shown on the right y-axis.

Table 1. Values and units used to calculate the pK_w of water

Parameter	Value	Units
α_0	-0.864671	-
α_1	8659.19	K
α_2	-22786.2	$(\text{g cm}^{-3})^{-2/3} \text{K}^2$
β_0	0.642044	$(\text{g cm}^{-3})^{-1}$
β_1	-56.8534	$(\text{g cm}^{-3})^{-1} \text{K}$
β_2	-0.375754	$(\text{g cm}^{-3})^{-2}$
γ_0	6.141500×10^{-1}	-
γ_1	4.825133×10^4	K
γ_2	-6.770793×10^4	K^2
γ_3	1.010210×10^7	K^3
n	6	-
ρ_0	1.0000	g cm^{-3}
$\rho^{298\text{K}}$	0.9971	g cm^{-3}
G	1000	g kg^{-1}
m_0	1	mol kg^{-1}
M_w	18.015268	g mol^{-1}

Table 2. Protection times ($\lambda = 1/\tau_f$) for HX during protein folding. The time constants below were obtained at 295 K for the refolding of V17A/V26A ubiquitin using a sample prepared at pH 7.5 (at 1 bar) in 25 mM potassium phosphate buffer. Refolding delays (τ_r) of 0, 50, 100, 200, 300, and 400 ms were employed with a constant exchange time (T_{ex}) of 125 ms. Uncertainties in the reported values are those obtained from the covariance matrix of the fit.

Q2	74.6	±	4.2
I3	84.4	±	6.4
F4	87.6	±	3.3
V5	85.2	±	6.2
K6	85.8	±	6.5
T7	80.2	±	4.2
I13	93.9	±	1.7
L15	83.2	±	9.1
E16	100.2	±	14.9
A17	68.7	±	9.2
E18	73.9	±	18.6
S20	81.6	±	6.3
D21	78.4	±	8.6
T22	91.4	±	7.5
I23	85.9	±	3.9
N25	83.6	±	6.5
A26	81.1	±	8.5
K27	91.6	±	7.2
A28	85.7	±	7.7
K29	92.7	±	3.6
I30	103.8	±	6.9
Q31	86.8	±	9.0
D32	82.6	±	7.8
K33	84.0	±	4.6
E34	86.1	±	3.4
G35	78.9	±	4.4
I36	77.8	±	10.5

Q40	83.3	±	8.7
Q41	79.9	±	5.8
R42	85.8	±	3.4
L43	90.3	±	10.0
I44	73.7	±	16.1
F45	79.4	±	3.5
K48	86.2	±	5.4
L50	93.3	±	7.2
E51	94.2	±	2.4
D52	75.4	±	5.1
R54	101.9	±	5.0
T55	80.1	±	4.5
L56	87.4	±	4.3
S57	74.2	±	5.6
D58	84.5	±	8.6
Y59	97.3	±	5.1
N60	95.0	±	6.3
I61	125.5	±	14.4
Q62	96.2	±	7.5
E64	95.7	±	8.7
S65	130.8	±	13.6
T66	102.5	±	9.9
L67	87.3	±	7.7
H68	93.6	±	5.5
L69	87.5	±	6.2
V70	77.5	±	20.5

Table 3. Intrinsic HX rates (k_{intr}) at 2.5 kbar, 295 K, of the unfolded state of V17A/V26A ubiquitin for a sample prepared at pH 7.5 (at 1 bar) in 25 mM potassium phosphate buffer.^a

I3	17	±	1
F4	17	±	1
V5 ^b	30	±	1
K6	29	±	2
T7	44	±	3
L8	54	±	3
T9	77	±	4
G10	88	±	5
K11	53	±	3
T12	53	±	3
I13	13	±	0.6
T14	19	±	0.9
L15	13	±	0.6
E16	8.3	±	0.4
A17	7.2	±	0.4
E18	4.4	±	0.3
S20	32	±	2
D21	25	±	1.2
T22	14	±	0.6
I23	5.8	±	0.3
E24	8.1	±	0.4
N25	42	±	3
A26 ^b	12.2	±	0.3
K27	15	±	1
A28	14	±	1
K29	15	±	1
I30	6.1	±	0.3
Q31	20	±	1
D32	23	±	1
K33	16	±	1
E34 ^b	12	±	1
G35	17	±	1
I36	3.2	±	0.2
D39	14	±	1
Q40	16	±	1
Q41	18	±	1
R42	19	±	1
L43	10.3	±	0.5
I44	3.4	±	0.2
F45	16	±	1
A46	39	±	2
G47	32	±	2
K48	27	±	2
Q49	20	±	1
L50	9.3	±	0.5
E51	13.3	±	0.6
G53	24	±	1
R54	13	±	1
T55	27	±	1
L56	17	±	1
S57	40	±	2
D58	32	±	2
Y59	10.9	±	0.5
N60	39	±	2
I61	6.3	±	0.3
Q62	14	±	0.6
K63	26	±	1
E64	25	±	1
S65	71	±	4
T66	80	±	5
L67	20	±	1
H68 ^b	183	±	5
L69	36	±	2
V70	6.6	±	0.3
L71	10	±	0.5
R72	26	±	1
L73	23	±	1
R74	40	±	2
G75	82	±	5
G76	3.9	±	0.3

^a Note that these values have not been corrected for the effect of pressure, and the 1 bar rates will be ca 2-fold faster, due to the lower pH at 2.5 kbar, partially offset by the increase in $[\text{OH}^-]$.⁷ Measurements were carried out at three pH values (6.4, 7.1, and 7.5), with reported values pertaining to pH 7.5. Residues for which accurate values were only available at pH 7.1 and 6.4, or 6.4 only, were converted to pH 7.5 using the known dependence of k_{intr} on pH.

^b Values measured only at pH 6.4 but converted to pH 7.5 using the known pH dependence of HX.

Table 4. HX rates (k_f) for the folded V17A/V26A ubiquitin mutant at 1 bar, pH 7.5, 295 K.^a Values measured only at pH 6.4 but converted to pH 7.5 using the known pH dependence of HX

Q2	1.5 ± 0.1
I3	0.06 ± 0.1
F4	0.03 ± 0.1
V5	0.03 ± 0.1
K6	0.06 ± 0.1
T7	0.5 ± 0.1
L8	4.0 ± 1.3
T9 ^a	56 ± 3
G10	4.5 ± 1
K11	5.4 ± 1
T12	4.1 ± 1
I13	0.11 ± 0.1
T14	1.6 ± 0.1
L15	0.07 ± 0.1
E16	0.55 ± 0.1
V17	0.57 ± 0.1
E18	0.07 ± 0.1
S20	0.82 ± 0.1
D21	0.13 ± 0.1
T22	0.27 ± 0.1
I23	0.15 ± 0.1
N25	0.43 ± 0.1
V26	0.06 ± 0.1
K27	0.05 ± 0.1
A28	0.06 ± 0.1
K29	0.02 ± 0.1
I30	0.04 ± 0.1
Q31	0.07 ± 0.1
D32	0.24 ± 0.1
K33	0.72 ± 0.1
E34	0.19 ± 0.1
G35	0.17 ± 0.01
I36	0.08 ± 0.1
D39	2.4 ± 0.2
Q40	0.32 ± 0.1
Q41	0.16 ± 0.1
R42	0.11 ± 0.1
L43	0.39 ± 0.1
I44	0.05 ± 0.1
F45	0.08 ± 0.1
A46	2.8 ± 2.2

G47	3.3 ± 0.3
K48	0.21 ± 0.1
Q49	3.5 ± 0.3
L50	0.12 ± 0.1
E51	0.79 ± 0.1
D52	0.53 ± 0.1
R54	0.20 ± 0.1
T55	0.30 ± 0.1
L56	0.12 ± 0.1
S57	0.29 ± 0.1
D58	0.50 ± 0.1
Y59	0.14 ± 0.1
N60	0.41 ± 0.1
I61	0.07 ± 0.1
Q62	0.50 ± 0.1
K63	2.31 ± 0.1
E64	0.23 ± 0.1
S65	0.60 ± 0.1
T66	0.58 ± 0.1
L67	0.05 ± 0.1
H68	0.03 ± 0.1
L69	0.06 ± 0.1
V70	0.04 ± 0.1
L71	0.34 ± 0.1
R72	2.3 ± 0.2
L73	4.8 ± 0.8
R74	3.2 ± 2.3
G75 ^a	59 ± 15
G76	4 ± 0.2

5.5 Acknowledgements

The work outlined in this chapter has been published in *Journal of the American Chemical Society* in 2017 under the title, “Monitoring hydrogen exchange during protein folding by fast pressure jump NMR spectroscopy,” by the authors T. Reid Alderson, Cyril Charlier, Dennis A. Torchia, Philip Anfinrud, and Ad Bax with the DOI 10.1021/jacs.7b066076.

The author expressed and purified all indicated proteins in the study and analyzed the data with C.C. Data were collected by A.B. with assistance from the author, data were analyzed by the author and C.C., and equations were derived by D.A.T. with P.A. providing instrumentation support and data analysis.

5.6 References

- Akasaka K, Kalbitzer HR, Inoue K, Yamada H, Herrmann C, et al. 2000. Pressure-induced local unfolding of the Ras binding domain of RalGDS. *Nat. Struct. Biol.* 7(7):547–50
- Alderson TR, Lee JH, Charlier C, Ying J, Bax A. 2018. Propensity for cis-Proline Formation in Unfolded Proteins. *ChemBioChem.* 19(1):37–42
- Bai Y, Milne JS, Mayne L, Englander SW. 1993. Primary structure effects on peptide group hydrogen exchange. *Proteins Struct. Funct. Genet.* 17(1):75–86
- Bai Y, Sosnick TR, Mayne L, Englander SW. 1995. Protein folding intermediates: native-state hydrogen exchange. *Science.* 269(5221):192–97
- Baldwin RL. 1993. Pulsed H/D-exchange studies of folding intermediates. *Curr. Opin. Struct. Biol.* 3(1):84–91
- Baldwin RL, Frieden C, Rose GD. 2010. Dry molten globule intermediates and the mechanism of protein unfolding. *Proteins Struct. Funct. Bioinforma.* 78(13):2725–37
- Bandura A V., Lvov SN. 2006. The Ionization Constant of Water over Wide Ranges of Temperature and Density. *J. Phys. Chem. Ref. Data.* 35(1):15–30
- Briggs MS, Roder H. 1992. Early hydrogen-bonding events in the folding reaction of ubiquitin. *Proc. Natl. Acad. Sci. U. S. A.* 89(6):2017–21
- Charlier C, Alderson TR, Courtney JM, Ying J, Anfinrud P, Bax A. 2018. Study of protein folding under native conditions by rapidly switching the hydrostatic pressure inside an NMR sample cell. *Proc. Natl. Acad. Sci.* 115(18):201803642
- Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A. 1995. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR.* 6(3):277–93
- Dooley RB. 2007. The International Association for the Properties of Water and Steam Release on the Ionization Constant of H₂O
- Dyson HJ, Wright PE. 2017. How Does Your Protein Fold? Elucidating the Apomyoglobin Folding Pathway. *Acc. Chem. Res.* 50(1):105–11
- Eliezer D, Yao J, Dyson HJ, Wright PE. 1998. Structural and dynamic characterization of partially folded states of apomyoglobin and implications for protein folding. *Nat. Struct. Biol.* 18(2):148–55
- Englander SW, Mayne L. 1992. Protein Folding Studied Using Hydrogen-Exchange Labeling and Two-Dimensional NMR. *Annu. Rev. Biophys. Biomol. Struct.* 21(1):243–65
- Fitzkee NC, Torchia DA, Bax A. 2011. Measuring rapid hydrogen exchange in the homodimeric 36 kDa HIV-1 integrase catalytic core domain. *Protein Sci.* 20(3):500–512
- Fuentes EJ, Wand AJ. 1998. Local Stability and Dynamics of Apocytochrome b562 Examined by the Dependence of Hydrogen Exchange on Hydrostatic Pressure
- Gayán E, Condón S, Álvarez I, Nabakabaya M, Mackey B. 2013. Effect of pressure-induced changes in the ionization equilibria of buffers on inactivation of *Escherichia coli* and *Staphylococcus aureus* by high hydrostatic pressure. *Appl. Environ. Microbiol.* 79(13):4041–47
- Gladwin ST, Evans PA. 1996. Structure of very early protein folding intermediates: new insights through a variant of hydrogen exchange labelling. *Fold. Des.* 1(6):407–17

- Grzesiek S, Bax A. 1993. Measurement of amide proton exchange rates and NOEs with water in $^{13}\text{C}/^{15}\text{N}$ -enriched calcineurin B. *J. Biomol. NMR.* 3(6):627–38
- Helgstrand M, Härd T, Allard P. 2000. Simulations of NMR pulse sequences during equilibrium and non-equilibrium chemical exchange. *J. Biomol. NMR.* 18(1):49–63
- Krantz BA, Sosnick TR. 2000. Distinguishing between Two-State and Three-State Models for Ubiquitin Folding
- Kremer W, Arnold M, Munte CE, Hartl R, Erlach MB, et al. 2011. Pulsed Pressure Perturbations, an Extra Dimension in NMR Spectroscopy of Proteins. *J. Am. Chem. Soc.* 133(34):13646–51
- Liu Y, Prigozhin MB, Schulten K, Gruebele M. 2014. Observation of Complete Pressure-Jump Protein Refolding in Molecular Dynamics Simulation and Experiment. *J. Am. Chem. Soc.* 136(11):4265–72
- Long D, Bouvignies G, Kay LE. 2014. Measuring hydrogen exchange rates in invisible protein excited states. *Proc. Natl. Acad. Sci.* 111(24):8820–25
- Mackenzie HW, Flemming Hansen D. 2018. Arginine Side-Chain Hydrogen Exchange: Quantifying Arginine Side-Chain Interactions in Solution. *ChemPhysChem*
- Meier S, Strohmeier M, Blackledge M, Stephan G. 2007. Direct Observation of Dipolar Couplings and Hydrogen Bonds across a β -Hairpin in 8 M Urea
- Miranker A, Robinson C V, Radford SE, Aplin RT, Dobson CM. 1993. Detection of transient protein folding populations by mass spectrometry. *Science.* 262(5135):896–900
- Muñoz V, Thompson PA, Hofrichter J, Eaton WA. 1997. Folding dynamics and mechanism of β -hairpin formation. *Nature.* 390(6656):196–99
- Nash DP, Jonas J. 1997. Structure of the Pressure-Assisted Cold Denatured State of Ubiquitin. *Biochem. Biophys. Res. Commun.* 238(2):289–91
- Neumaier S, Kiefhaber T. 2014. Redefining the Dry Molten Globule State of Proteins. *J. Mol. Biol.* 426(13):2520–28
- Panick G, Gediminas JAV, Malessa R, Rapp G, Winter R, Royer CA. 1999. Exploring the Temperature–Pressure Phase Diagram of Staphylococcal Nuclease
- Redfield C. 2004. Using nuclear magnetic resonance spectroscopy to study molten globule states of proteins. *Methods.* 34(1):121–32
- Roche J, Dellarole M, Caro JA, Norberto DR, Garcia AE, et al. 2013. Effect of Internal Cavities on Folding Rates and Routes Revealed by Real-Time Pressure-Jump NMR Spectroscopy. *J. Am. Chem. Soc.* 135(39):14610–18
- Royer CA, Hinck AP, Loh SN, Prehoda KE, Peng X, et al. 1993. Effects of amino acid substitutions on the pressure denaturation of staphylococcal nuclease as monitored by fluorescence and nuclear magnetic resonance spectroscopy. *Biochemistry.* 32(19):5222–32
- Samarasinghe SD, Campbell DM, Jonas A, Jonas J. 1992. High-resolution NMR study of the pressure-induced unfolding of lysozyme. *Biochemistry.* 31(34):7773–78
- Shakhnovich EI, Finkelstein A V. 1989. Theory of cooperative transitions in protein molecules. I. Why denaturation of globular protein is a first-order phase transition. *Biopolymers.* 28(10):1667–80
- Went HM, Benitez-Cardoza CG, Jackson SE. 2004. Is an intermediate state populated on the folding pathway of ubiquitin? *FEBS Lett.* 567(2–3):333–38

- Yamaguchi T, Yamada H, Akasaka K. 1995. Thermodynamics of Unfolding of Ribonuclease A under High Pressure. A study by Proton NMR. *J. Mol. Biol.* 250(5):689–94
- Ying J, Delaglio F, Torchia DA, Bax A. 2016. Sparse multidimensional iterative lineshape-enhanced (SMILE) reconstruction of both non-uniformly sampled and conventional NMR data. *J. Biomol. NMR*
- Yuwen T, Bah A, Brady JP, Ferrage F, Bouvignies G, Kay LE. 2018. Measuring Solvent Hydrogen Exchange Rates by Multifrequency Excitation 15N CEST: Application to Protein Phase Separation. *J. Phys. Chem. B.* acs.jpcc.8b06820
- Zhang J, Peng X, Jonas A, Jonas J. 1995. NMR Study of the Cold, Heat, and Pressure Unfolding of Ribonuclease A. *Biochemistry.* 34(27):8631–41

6

Monitoring protein structure by high-pressure, pulsed field gradient diffusion

NMR spectroscopy

6.1 Abstract

Pulsed-field gradient (PFG) NMR spectroscopy is used extensively to measure translational diffusion of biomolecules in solution. From such measurements, their hydrodynamic radius (R_h) can be calculated, often using a small amount of dioxane as an internal reference. For proteins, R_h provides important structural information regarding the oligomeric state and the extent to which an unfolded chain may have collapsed. The latter often requires the measurement of R_h throughout a titration of denaturant, requiring different sample preparations for each measurement. Hydrostatic pressure provides a convenient and relatively benign alternative method for denaturation of proteins and dissociation of oligomers. I employed PFG NMR to study the translational diffusion of wild-type (WT) and mutant ubiquitin as a function of hydrostatic pressure. The R_h of the commonly used dioxane reference was found to increase with pressure, by up to 8% at 3000 bar. Somewhat smaller increases in R_h values of methanol, ethanol, and DMSO suggest that the increased hydration of such molecules is a general phenomenon, which may be correlated with the strong non-linearity in the pressure dependence of their chemical shifts. Using the newly calibrated pressure dependence of the R_h of dioxane, the R_h values of both the folded and the unfolded state of a pressure-sensitized ubiquitin mutant remain nearly invariant, either by the application of pressure or addition of urea. At 24.2 Å, the R_h of the pressure-denatured state is only slightly smaller than that of the urea-denatured state, and close to the value expected for an idealized random coil of 76 residues. Translational diffusion of α -synuclein confirms the previously reported compaction relative to an idealized random coil but shows no detectable pressure dependence of R_h .

6.2 Introduction

Pulsed-field gradient (PFG) NMR spectroscopy enables the measurement of translational diffusion of biomolecules in solution (Johnson 1999, Price 1997, Price 1998, Price 2009). Through the Stokes-Einstein equation (eq 1, Methods), the translational diffusion coefficient is inversely related to the hydration radius (R_h) of a molecule. Therefore, R_h provides insight into molecular size, including alterations upon oligomerization (Altieri et al. 1995, Dingley et al. 1995, Bernado et al. 2003, Jansma et al. 2010, Huang et al. 2016, Price et al. 1999), ligand binding (Weljie et al. 2003), or folding and unfolding (Wilkins et al. 1999, Jones et al. 1997, Kamatari et al. 2001, Choy et al. 2002, Buevich and Baum 2002). As the initial biophysical characterization of a protein often entails determining its oligomeric state in solution, PFG NMR provides a simple and robust method to discriminate between monomers and oligomers, without the need for isotope labeling and measurement of spin relaxation to derive a measure of the rotational correlation time (Kay et al. 1989, Lee et al. 2006). Pulsed field gradients are widely used in modern NMR spectroscopy, and availability of the requisite hardware now enables routine measurement of translational diffusion constants for a wide range of systems, spanning molecules as small as tens of Da [19] to complex protein assemblies that approach 1 MDa (Huang et al. 2016, Baldwin et al. 2011), amyloid fibrils (Baldwin et al. 2008), virus-like particles (Leung et al. 2017), and phase-separated membrane-less organelles (Brady et al. 2017, Huang et al. 2017).

In addition to determining the oligomeric state of a protein in solution, PFG translational diffusion measurements offer insight into structural ensembles populated by denatured, intrinsically disordered, or partially folded proteins (Weljie et al. 2003, Jones et al. 1997, Choy et al. 2003, Wilkins et al. 1999). Empirical measurements have established a relationship between the number of amino acids in a fully denatured protein and its expected R_h (Wilkins et al. 1999), and comparison to the R_h measured for the protein of interest then reports on the overall compaction or, in rare cases, expansion of the polypeptide chain. Compaction or expansion of the

chain can result from the presence of transiently structured elements or weak hydrophobic or electrostatic interactions within the unfolded polypeptide. While studies performed under high denaturant conditions (*e.g.* 8 M urea or 6 M guanidine hydrochloride) generally yield R_h values that closely follow the empirical relation between chain length and R_h , the magnitude of the change in the structural ensemble populated by unfolded proteins at lower denaturant conditions, approaching the folding threshold, remains controversial: Under such conditions, both significantly collapsed and fully expanded conformations were deduced from Förster resonance energy transfer (FRET) and small-angle X-ray scattering (SAXS) methods (Yoo et al. 2012, Watkins et al. 2015, Borgia et al. 2016, Riback et al. 2017). Previous solution-state NMR studies demonstrated that R_h of the unfolded state of the N-terminal SH3 domain of drk, which exists in a slow equilibrium with its folded state, is *ca.* 30% larger than its folded state, but *ca.* 10% smaller than the R_h of the guanidine-hydrochloride, fully denatured state (Choy et al. 2003, Mok et al. 1999). Following a recent, renewed debate about discrepancies between the denaturant dependence of the unfolded state measured by SAXS and FRET (Best et al. 2018, Riback et al. 2018), another group confirmed that such differences are consistent with a decoupling of size and shape fluctuations in finite-sized, heteropolymeric disordered systems (Fuentes et al. 2018); *i.e.*, the FRET-derived distance distributions, SAXS-derived radii of gyration, and diffusion-derived R_h contain semi-independent constraints on the ensemble distribution (Choy et al. 2003).

Here, I set out to further investigate the effect of pressure on protein structure near its pressure denaturation transition. Under conditions where there is an appreciably negative volume change between the folded and unfolded states, on the order of ≥ 50 mL/mol, high-pressure can drive the thermodynamic equilibrium of a folded protein to that of the unfolded state. Hydrophobic cavities present in the folded protein but absent in the unfolded state often provide an important component of this volume difference (Roche et al. 2012, Roche et al. 2017). Similarly, elevated pressures can shift an oligomeric equilibrium toward smaller subunits or monomers (Silva et al.

1993). In comparison to high temperatures and chemical denaturation, elevated hydrostatic pressure represents a relatively benign and often reversible denaturing mechanism (Caro and Wand 2018), which has been extensively applied to the investigation of protein folding (Roche et al. 2017, Royer et al. 1993, Nucci et al. 2014). Proteins that unfold at high pressure typically exist in slow exchange with the folded state when studied at intermediate pressures, enabling observation of signals from both states. This situation applies for the previously characterized pressure-sensitized double mutant of ubiquitin, V17A/V26A (VA2-ubiquitin), used as a model system in the present study (Alderson et al. 2017, Charlier et al. 2018a, Charlier et al. 2018b). Therefore, high-pressure PFG NMR provides simultaneous access to the measurement of translational diffusion coefficients for both the folded and unfolded states that are in equilibrium with one another over a range of conditions.

PFG translational diffusion rates are commonly extracted from an array of 1D or 2D NMR spectra in which signal attenuation resulting from the application of increased gradient strengths enables quantification of the diffusion coefficient (Price et al. 1999, Stilbs et al. 1987, Stejskal and Tanner 1965, Tanner 1970). In such experiments, the magnetization of interest is stored along the z-axis for a fixed diffusion delay time, which is preceded and followed by the respective application of encoding and decoding gradients of increasing strengths (Dingley et al. 1995, Wu et al. 1995, Chou et al. 2004). Translational diffusion manifests as incomplete gradient decoding and concomitant signal attenuation, from which the diffusion coefficient can be calculated based on the Stejskal-Tanner equation (eq 2, Methods). Large proteins that suffer from rapid signal decay or insufficient signal attenuation during the diffusion delay can be studied using relaxation-optimized pulse sequences (Horst et al. 2011, Didenko et al. 2011) or newly developed methyl-based triple quantum experiments (Brady et al. 2017, Huang et al. 2017), respectively.

Determination of R_h from PFG NMR experiments requires accurate knowledge of the solution viscosity and gradient strength in $G\text{ cm}^{-1}$ (if using the Stejskal-Tanner equation, eq 2).

With water viscosity being strongly dependent on temperature, and inhomogeneity in the gradient across the sample making this parameter sensitive to the sample size and position, the use of an internal standard is often preferred (Wilkins et al. 1999). In this case, the diffusion of a reference molecule of known R_h is determined alongside that of a protein of unknown R_h (eqn 3, Methods). Dioxane is a commonly used reference molecule ($R_h = 2.12 \text{ \AA}$ under ambient conditions) as it conveniently resonates near 3.5 ppm, upfield of most H^α resonances and downfield of most H^β signals, and was found not to interact with folded or unfolded proteins [12].

Here, I employ high-pressure PFG NMR spectroscopy to measure the translational diffusion coefficients and R_h values at hydrostatic pressures ranging from 1 to 3000 bar. Although the diffusivity of water follows the known pressure dependence of its viscosity (<https://webbook.nist.gov/chemistry/fluid/>), dioxane and other small molecules show a significant expansion in their R_h values with increasing pressure. When using dioxane as an internal reference to derive the R_h of a protein, failure to correct for this effect results in underestimates in the protein R_h . Taking the newly calibrated pressure dependence of R_h into account restores accuracy of the internal referencing method, and the unfolded chain of VA2-ubiquitin is then found to compact only slightly near its pressure denaturation transition, corresponding to a ca. 10% smaller R_h relative to its fully denatured state, with similar results also obtained when using urea to denature the protein. By contrast, wild type ubiquitin, which requires much higher pressures for denaturation, shows virtually no change in its R_h for pressures up to 3 kbar.

6.3 Results and Discussion

6.3.1 Pressure-induced changes to chemical shifts and hydration radii of tracer molecules

Translational diffusion measurements as a function of pressure were performed using a commercially available high-pressure NMR tube rated for pressures up to 3000 bar. Simultaneous measurement of the diffusion constants for a protein and reference molecule of known R_h

eliminates the need for accurate knowledge of solution viscosity as R_h of the protein can then be obtained directly from the ratio of the diffusion constants. However, this procedure requires that R_h of the reference molecule is independent of pressure, which I find not to be the case. However, as described below for dioxane and several other reference compounds, the pressure dependence of the reference R_h can easily be calibrated, thereby resolving this issue.

The substantial variation in translational diffusivity as a function of pressure for various small molecules can be visualized by comparing the levels of attenuation seen in the NMR signal intensities obtained at the maximum PFG strength employed in the diffusion experiments (Fig. 1A). Notably, all tracer molecules also exhibit significant non-linear ^1H chemical shift changes with pressure (Fig. 1A, Fig. 1B). Chemical shift changes with pressure are typically described by a second-order polynomial function:

$$\delta = \delta_0 + B_1(P - P_0) + B_2(P - P_0)^2 \quad (4)$$

where δ refers to the observed chemical shift, δ_0 to the chemical shift at atmospheric pressure (P_0), P to the applied pressure, and B_1 and B_2 to coefficients that respectively scale the first- and second-order terms. Prior NMR studies on a wide range of proteins and peptides have documented pressure-induced, non-linear chemical shift changes that vary in magnitude and sign depending on the type of nucleus and the location of the atom in the peptide or protein structure (Akasaka 2006, Akasaka & Li 2001, Kalbitzer 2015, Kitahara et al. 2013, Li & Kamatari 2015). The second order term is often attributed to large conformational fluctuations associated with an increased population of a low-lying conformationally excited state or ensemble of states (Nucci et al. 2014, Kitahara et al. 2013, Kalbitzer et al. 2013, Maeno et al. 2015). For a two-state transition at fixed temperature, pressure, and compressibility, volume changes impact the free energy as follows:

$$\Delta G = \frac{\Delta\beta}{2}(P - P_0)^2 + \Delta V(P - P_0) + \Delta G_0 \quad (5)$$

where Δ indicates the difference, β is the isothermal compressibility, V the partial molar volume, and G_0 the free energy at a reference pressure. For a two-state reaction, the B_2/B_1 ratio was previously related to the ratio of $\Delta\beta$ and ΔV , with negative B_2/B_1 ratios indicative of a positive value of $\Delta\beta/\Delta V$ (Beck et al. 2014). The small molecules used in my study also yield negative B_2/B_1 ratios of comparable values to proteins and peptides analyzed previously (Beck et al. 2014). However, it is more difficult to envision such molecules being subject to an equilibrium between two conformational states of different compressibility. Alternatively, it is likely that pressure impacts details of the hydration/packing of water molecules around the solute, and that such differentially hydrated forms have different compressibility and total volume as well as different chemical shifts. My results therefore imply that factors other than traditional volume or compressibility changes can cause the non-linear chemical shift dependencies on pressure, and the structural interpretation of non-linear chemical shift changes in proteins in the absence of other diagnostic features (e.g. intensity loss related to exchange line broadening) is cautioned.

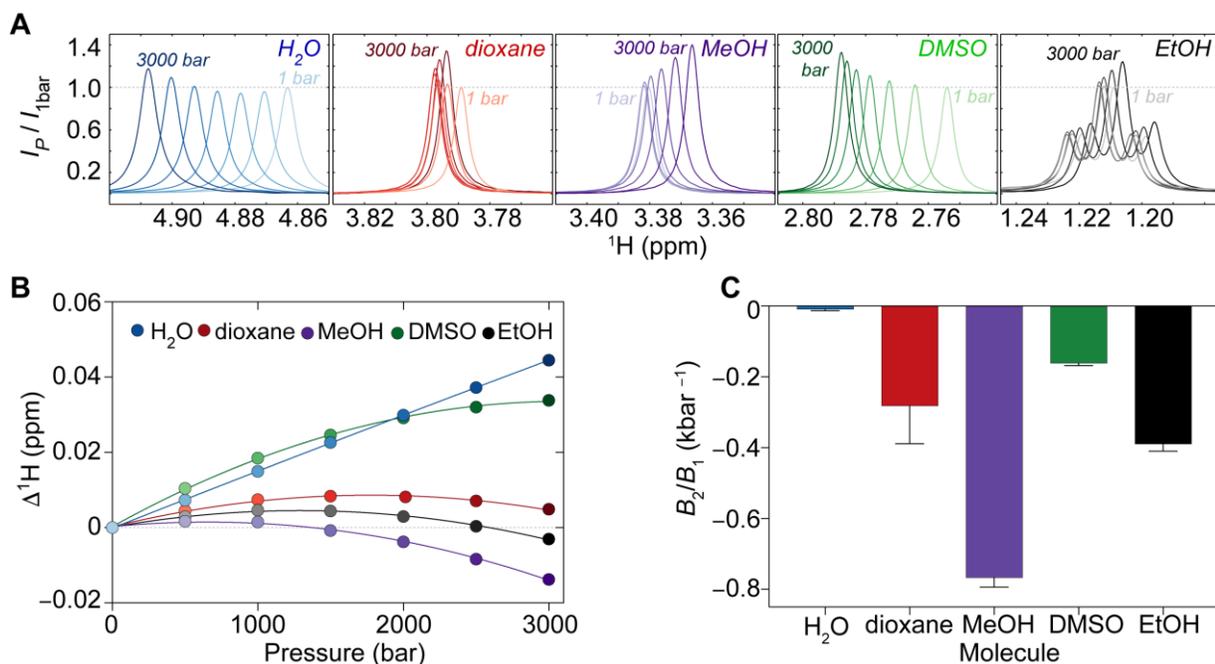


Figure 1 | 1D PFG NMR spectra of tracer molecules reveal pressure dependence of chemical shifts at 293 K. (A) For each of the five molecules shown here, seven 1D PFG NMR spectra are shown at hydrostatic pressures ranging from 1 bar (light) to 3000 bar (dark) in 500 bar steps. The spectra are shown at the maximum gradient strength used in the PFG diffusion experiments and scaled such that the intensity at 1 bar equals 1. R_h values extracted from decay rates as a function of gradient strength are shown in Fig. 2. **(B)** ^1H chemical shift changes with pressure for water, dioxane, and the methyl groups of methanol, ethanol, and dimethylsulfoxide (DMSO). Solid lines represent best-fits to eq 4. The water chemical shift follows reference [63]. **(C)** The ratios of fitted B_2 and B_1 values from the data in panel B. Error bars represent one standard deviation.

The diffusivity of water as a function of pressure at a fixed temperature is readily extracted from the data shown in Fig. 1. Note, however, that strong detuning of the ^1H coil as well as a weak (ca 1 G/cm) z gradient during the diffusion delay were needed to prevent radiation damping. The relative diffusivity as a function of pressure ($D_{\text{H}_2\text{O}}^{x\text{ bar}} / D_{\text{H}_2\text{O}}^{1\text{ bar}}$) shows the maximum near 1200 bar, expected on the basis of the known pressure dependence of its viscosity (Fig. 2A). Although the agreement is much closer than seen in early NMR PFG studies of water diffusion (Benedek and Purcell 1954, Defries and Jonas 1977), a small but systematic difference, approaching 1.5% at 3 kbar is reproducibly seen. This difference relative to the known pressure dependence of water viscosity (<https://webbook.nist.gov/chemistry/fluid/>) does not arise from the presence of 4% D_2O and 4% tracer molecules (dioxane, methanol, ethanol, and DMSO), as repeating the measurement

with 2% D₂O and 2% total volume for the tracer molecules yielded the same, small systematic difference (Fig. 2A). It is conceivable that a small increase in the H-O-H bond angle with pressure, resulting from the pressure dependence of water H-bonding (Soper and Ricci 2000), could increase the effective water diameter, and thereby be responsible for this effect.

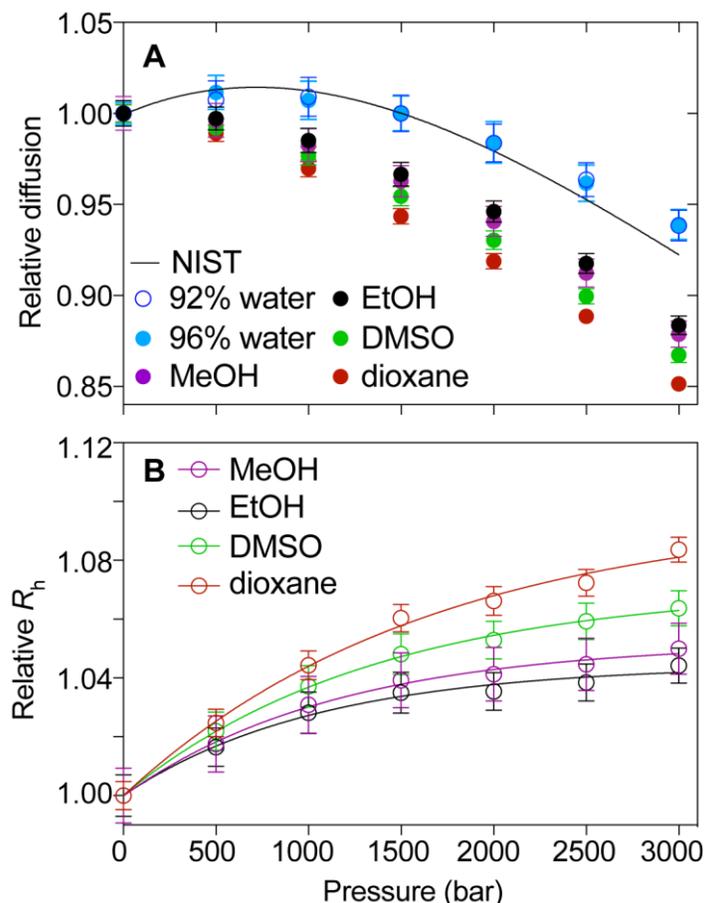


Figure 2 | Pressure-dependent corrections to the hydration radii of tracer molecules. (A) The relative diffusion ($D_{x\text{-bar}} / D_{1\text{-bar}}$) at 293 K for water and the four tracer molecules plotted as a function of the hydrostatic pressure. Solid symbols correspond to a sample of 92% H₂O, 4% D₂O, 1% each (v/v) of dioxane, DMSO, methanol and ethanol. The solid line depicts the inverse relative viscosity ($\eta_{1\text{bar}}/\eta_{x\text{bar}}$) as obtained from the NIST website for water at 293 K (<https://webbook.nist.gov/chemistry/fluid/>). The relative diffusion of water follows the expected pressure dependence of the water viscosity. Slight differences do not result from the use 4% D₂O and 4% tracers (v/v), as a repeat using a sample with 2% D₂O and 2% tracers (v/v) yielded essentially indistinguishable results (cyan circles). (B) The relative R_h values ($R_h^P / R_h^{1\text{bar}}$) of dioxane (red), ethanol (black), methanol (purple), and DMSO (green) increase as a function of pressure. R_h values were derived relative to dioxane, which was scaled at each pressure to correct for the pressure dependence of the water viscosity known from the NIST website. The solid lines indicate best fit lines to second-order polynomial functions. The R_h values at 1 bar for methanol, ethanol, DMSO, and dioxane are respectively 1.50, 1.95, 2.01, and 2.12 Å.

Interestingly, all tracer molecules exhibit a decrease in diffusivity with increasing pressure that is far greater than expected based on the change in solvent viscosity (Fig. 2, Fig. 3). These results therefore point to increases in their hydration radii with pressure. For dioxane, the best fit of R_h as a function of pressure yields:

$$R_h^{\text{dioxane}} = 2.12 \text{ \AA} + (1.028 \times 10^{-4}) \frac{P}{P_0} - (1.566 \times 10^{-8}) \left(\frac{P}{P_0}\right)^2 \quad (6)$$

$$R_h^{\text{DMSO}} = 2.01 \text{ \AA} + (8.181 \times 10^{-5}) \frac{P}{P_0} - (1.377 \times 10^{-8}) \left(\frac{P}{P_0}\right)^2$$

$$R_h^{\text{EtOH}} = 1.95 \text{ \AA} + (5.608 \times 10^{-5}) \frac{P}{P_0} - (1.017 \times 10^{-8}) \left(\frac{P}{P_0}\right)^2$$

$$R_h^{\text{MeOH}} = 1.51 \text{ \AA} + (4.956 \times 10^{-5}) \frac{P}{P_0} - (8.88 \times 10^{-9}) \left(\frac{P}{P_0}\right)^2$$

where P is the applied pressure, and $P_0 = 1$ bar.

If the pressure-dependence of R_h of the internal dioxane reference were neglected, both folded wild-type ubiquitin and the intrinsically disordered protein α S would show substantial decreases in their R_h with increasing pressure (Fig. 4, red symbols), which could be misinterpreted as a dramatic compaction or loss of hydration for folded ubiquitin, or a collapse of the disordered α S chain with pressure. Although folded proteins are slightly compressible (Gavish et al. 1983), resulting in a volume loss of ca. 3% at 3 kbar, this would yield a reduction of R_h by only ~1%. Indeed, when accounting for the pressure dependence of the dioxane R_h , the R_h values obtained for ubiquitin and α S remain nearly independent of pressure up to 2.5 kbar (Fig. 4, black symbols). For ubiquitin, the absence of any significant structural change at pressures ≤ 2.5 kbar is consistent with excellent fits of the backbone residual dipolar couplings, measured at these elevated pressures, to the solution NMR structure derived at 1 bar (Roche et al. 2015). The onset of a small increase in R_h seen at 3 kbar, well below the unfolding pressure threshold, may reflect a structural change previously deduced from substantial changes in chemical shifts and measurement of NOEs (Kitahara et al. 2005). For α S, R_h is found to remain essentially unchanged

over the entire 1–3000 bar range, provided that the change of the dioxane R_h is taken into account. In the absence of such a correction, a substantial decrease in the radius of hydration for this intrinsically disordered protein with increasing pressure would be obtained. The latter would appear counter-intuitive considering that at 1 bar the opposing charges of the N- and C-terminal regions of the protein are responsible for a modest degree of electrostatic compaction ([Bertoncini et al. 2005](#)). The *ca.* 10% increase in the water dielectric constant at 3 kbar would then be expected to perhaps slightly decrease this electrostatic compaction, counter to observation.

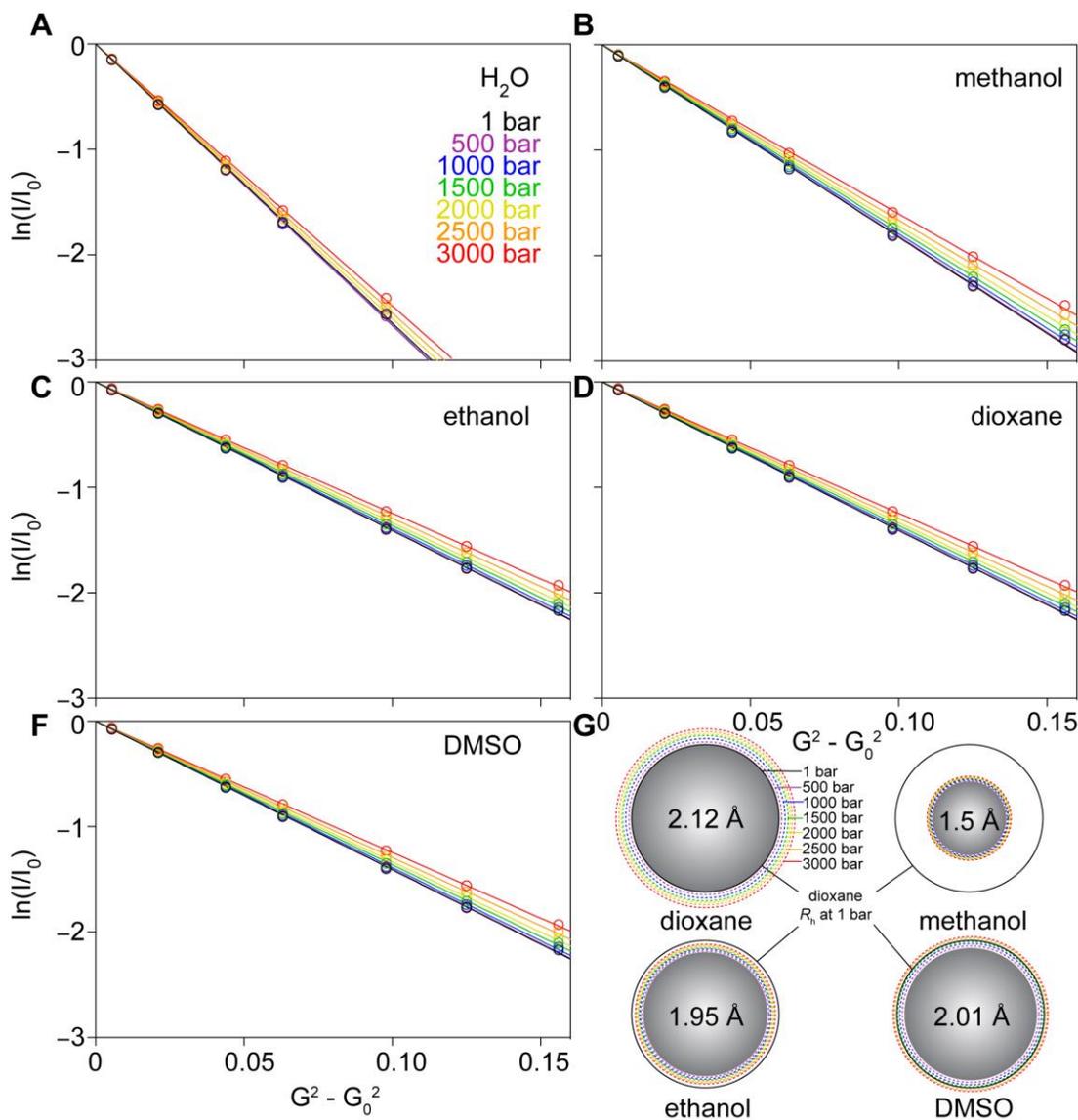


Figure 3 | PFG NMR data used to determine the solution viscosity and pressure-induced changes to the hydration radii of tracer molecules. (A-F) Natural logarithm of the intensity ratio (I/I_0) and best-fit line to $y = \exp(-Ax)$ through the data points. Note that the x-axis depicts the normalized gradient strength squared, where G_0 is the weakest gradient. Panels A-F show data measured at seven different pressures for five different molecules, including H₂O, methanol, ethanol, dioxane, and DMSO, respectively. I is the peak intensity recorded in a given spectrum and I_0 is the peak intensity at the weakest gradient strength. See Methods for experimental details. The relative diffusion of H₂O was used to calculate relative changes in solution viscosity with pressure. **(G)** The R_h value of dioxane at 1 bar is 2.12 Å, which was used to calculate R_h values for all other tracer molecules at atmospheric pressure (indicated with text and sphere size). The increases in R_h with pressure are shown as dotted spheres with colors matching panels A-F. The black dotted spheres for methanol, ethanol, and DMSO indicate the R_h of dioxane at 1 bar for comparison.

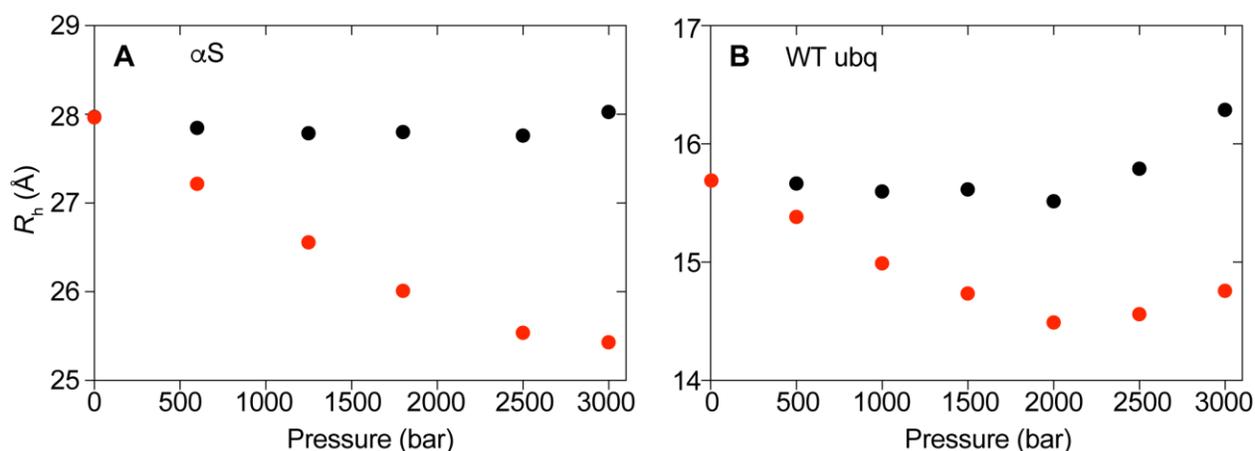


Figure 4 | Pressure-dependent corrections to the hydration radii of proteins. R_h values of (A) the IDP αS and (B) folded, wild-type ubiquitin are plotted as a function of hydrostatic pressure. The ubiquitin mid-point of pressure-induced unfolding of ubiquitin is 5.4 kbar [71], and my measurements therefore pertain to the folded state. When using a fixed 2.12-Å radius for the internal dioxane reference (red symbols), systematically smaller R_h values are obtained due to the increased hydration radius of dioxane with pressure (Fig. 2). Accounting for the pressure dependence of the dioxane R_h yields protein R_h values (black symbols) that are nearly invariant between 1 bar and 2500 bar. The raw data are included in Fig. 5.

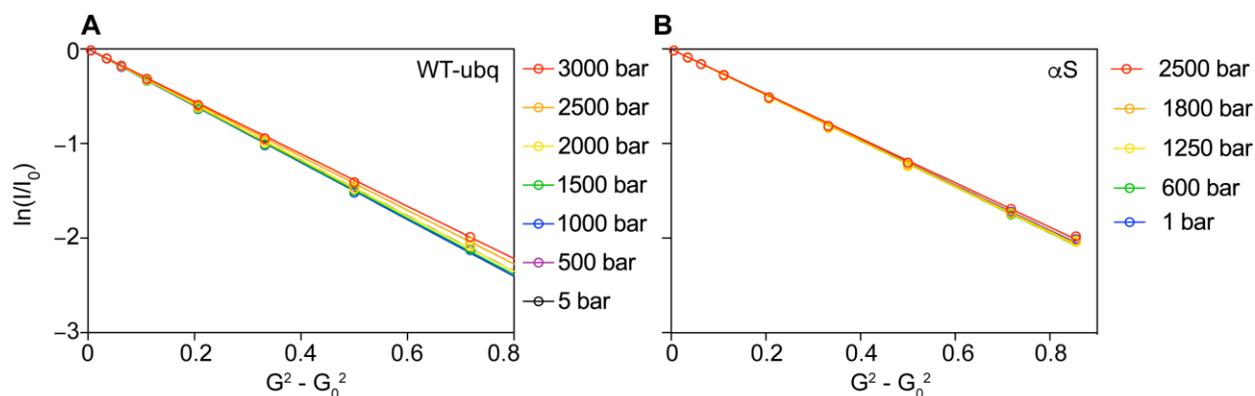


Figure 5 | Diffusion measurements on model proteins as a function of pressure. The natural logarithm of the intensity ratio (I/I_0), where I_0 refers to the intensity in the BPP-LED spectrum acquired at the weakest gradient, plotted vs the squared gradient strength, with G_0 referring to the weakest gradient strength. Shown here are results for (A) WT-ubq and (B) αS as a function of pressure. Note that WT-ubq experiences a more pronounced change in diffusion at higher pressures in comparison to αS , likely reflecting a small increase in R_h of WT-ubq at very high pressures.

6.3.2 Pressure- and urea-denatured VA2-ubiquitin have similar hydrodynamic properties

To investigate the ubiquitin R_h near its mid-point of unfolding, I utilized a previously introduced cavity mutant (V17A/V26A), VA2-ubiquitin, which unfolds at lower pressures due to a decreased volume of the unfolded state, but a volume of the folded state that is very close to that of the wild-type protein (Alderson et al. 2017).

At 1 bar, VA2-ubiquitin and the wild-type protein show indistinguishable translational diffusion (Fig. 6), confirming their very similar 3D structures. At 3 kbar, where folded VA2-ubiquitin is no longer appreciably populated, the diffusion-derived R_h of the unfolded chain is 24.2 Å (Fig. 7). Both the experimentally derived R_h of the folded protein at 1 bar (15.7 Å) and the R_h of the unfolded chain are in good agreement with the empirical equations of Wilkins et al. (Wilkins et al. 1999), which predict values of 16.7 and 26.1 Å, respectively for this 76-residue protein. If instead of high pressure, 8M urea is used to denature VA2-ubiquitin, translational diffusion measurements at 1 bar yield $R_h = 25.3$ Å, very close to the value seen in the pressure-denatured state (Fig. 7).

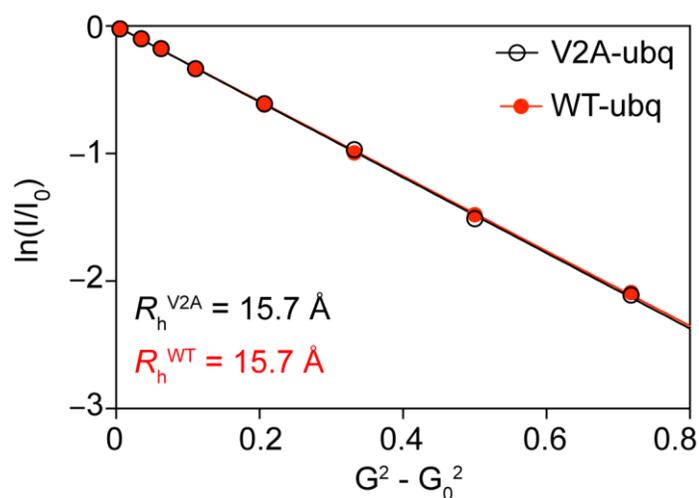


Figure 6 | Comparison of the diffusion of V2A- (black) and WT-ubq (red). The R_h values obtained from comparison to dioxane are listed in the lower left corner. BPP-LED measurements were recorded at 288 K and 1 bar.

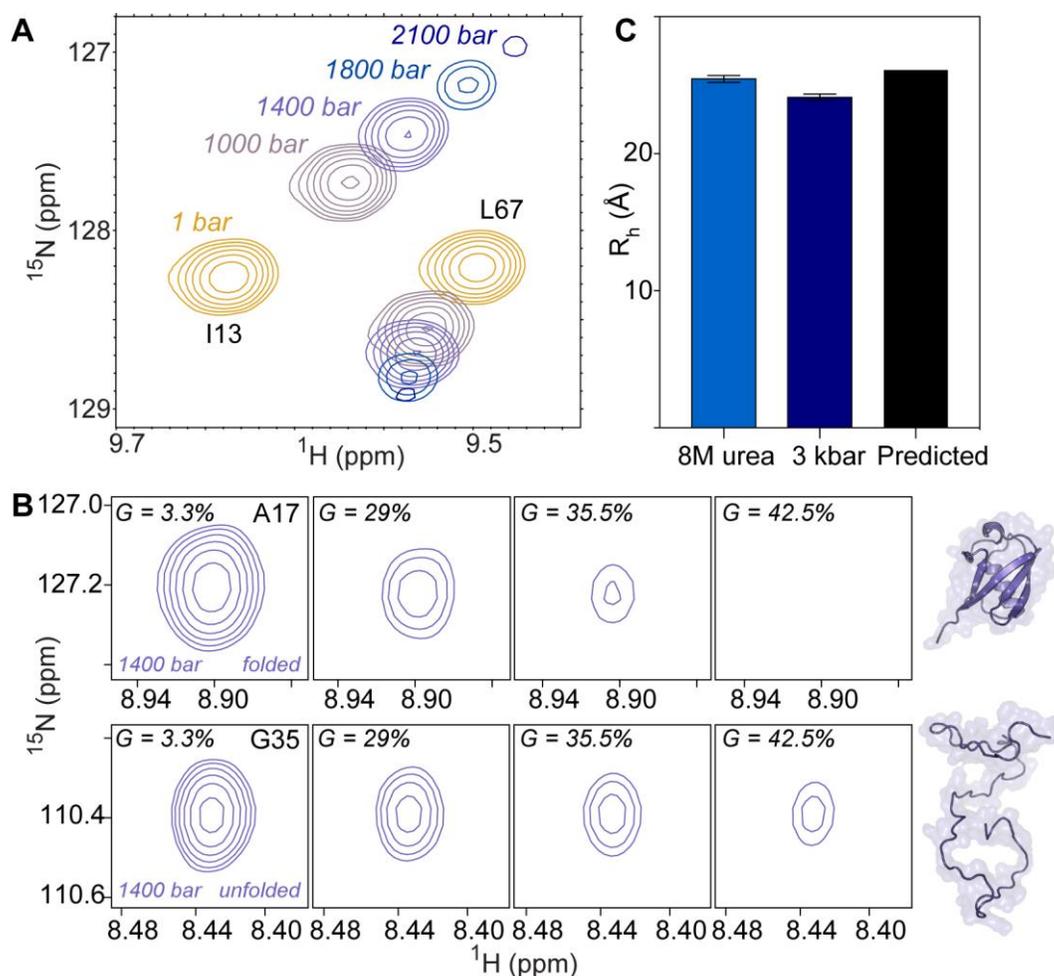


Figure 7 | Urea- and pressure-denatured VA2-ubiquitin have similar hydration radii. (A) Small regions of 2D ^1H - ^{15}N HSQC spectra of ^{15}N -VA2-ubiquitin recorded as a function of hydrostatic pressure. Shown are resonances from the folded state (I13, L67) at 1, 1000, 1400, 1800, and 2100 bar. (B) Example of the decrease in cross peak intensity with increasing gradient strength, G , in 2D PFG NMR spectra of folded (top, A17 resonance) and unfolded (bottom, G35 resonance) VA2-ubiquitin recorded at 1400 bar where both states are about equally populated. Corresponding 2D PFG NMR data from the urea-denatured state are shown in Fig. 8. (C) R_h values determined for unfolded VA2-ubiquitin in 8M urea (light blue) and at 3000 bar (dark blue) are close to the value predicted by the empirical equation of Wilkins et al.[12] for a 76-residue unfolded protein (black).

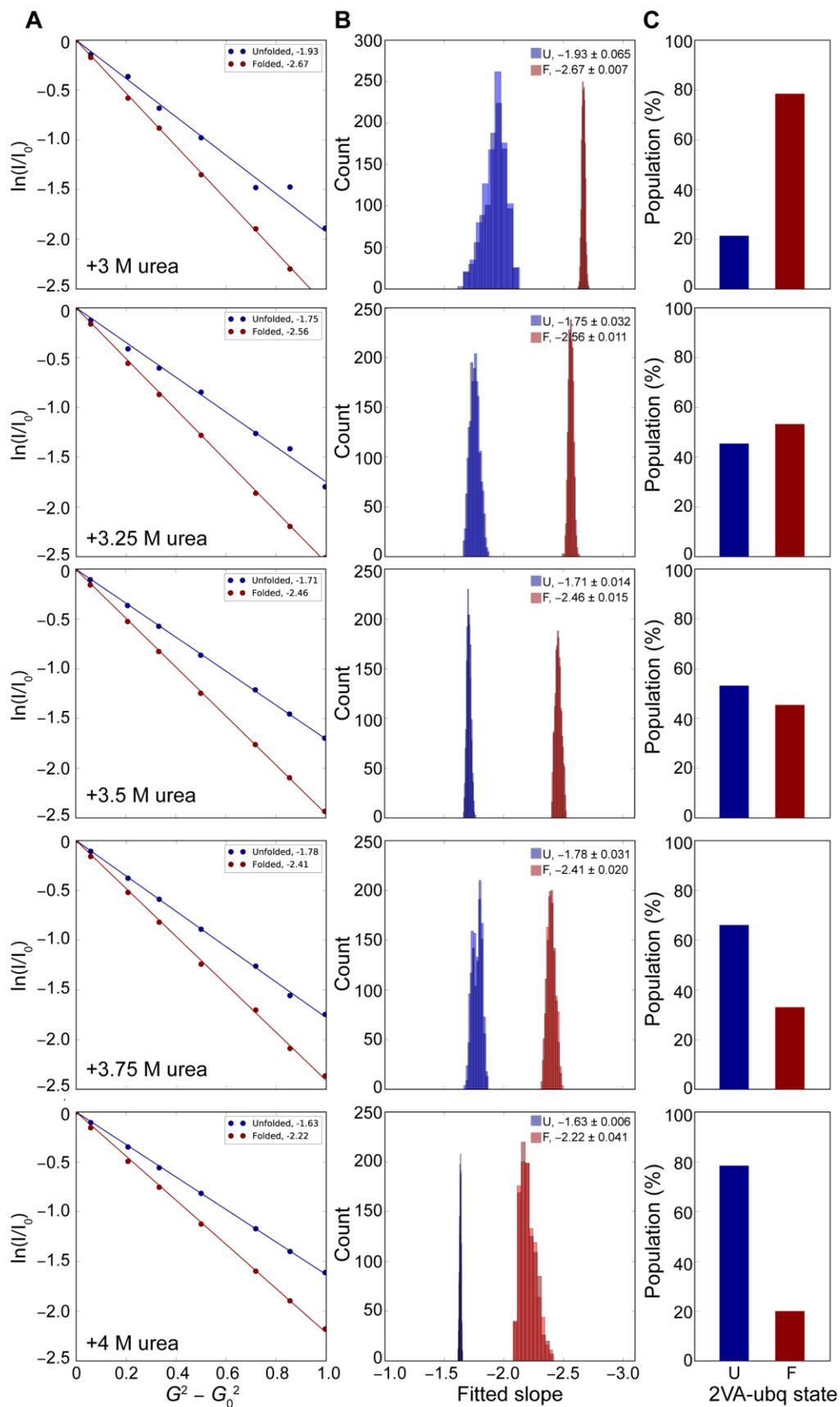


Figure 8 | 2D PFG NMR translational diffusion measurements in the presence of increasing urea concentrations. Pseudo-3D PFG NMR experiments were employed to measure the translational diffusion of the unfolded and folded states of 2VA-ubq simultaneously in the same sample. **(A)** The decay of signal intensity, $\ln(I/I_0)$, for both the folded (red) and unfolded (blue) states, where I_0 is the intensity in the plane with the weakest gradient. Resonances from each state with signal-to-noise ratios > 10 were selected and pooled, with a single exponential fit to the summed intensities at each gradient strength. The solid lines represent best-fit lines to the data points. **(B)** A jackknife procedure was employed to estimate the errors associated with the fitted slopes in panel A. For each state, resonances were randomly split into two groups, with the intensities from the two groups then separately summed and fit to a best-fit line, as in panel A. This procedure was repeated 1000 times to generate 2×1000 best-fit lines for both the folded and unfolded states. The error was derived by computing the root-mean-square deviation between the two sets of best-fit lines and dividing by the square root of two. The results are plotted in a histogram format with the two groups depicted for each state in darker and lighter shades of blue (unfolded) and red (folded). The value of the slope derived from fitting all summed intensities is indicated in the upper right, with the errors derived as described. **(C)** The population of unfolded (blue) and folded (red) 2VA-ubq at each urea concentration, as estimated from resolved resonances in both states. The population of unfolded was computed as $[U] / ([U]+[F])$, with $[F]$ then equal to $1-[U]$, where the square brackets represent the peak intensity associated with a resonance from either the unfolded (U) or folded (F) state of 2VA-ubq.

6.3.3 PFG NMR diffusion experiments indicate minimal compaction of the unfolded state in low denaturant conditions

To determine if the R_h of unfolded VA2-ubiquitin decreases with decreasing denaturant levels, I measured its diffusion across a range of pressures and urea concentrations where both the folded and unfolded states are appreciably populated. The two forms exist in slow exchange on the NMR timescale, and thus separate resonances for each state are observed. The presence of both the folded and unfolded state in a single sample complicates analysis of PFG NMR diffusion data due to extensive resonance overlap of their one-dimensional ^1H NMR spectra. This can be safely mitigated by utilizing a 2D ^1H - ^{15}N HSQC version of the PFG diffusion experiment (Choy et al. 2003), providing enhanced resolution to separate folded from unfolded resonances. A second complication arises from potential exchange between the two states during the long diffusion delay, which could introduce cross-talk between the measured diffusion rates. I ruled out this possibility by measuring high signal-to-noise ^{15}N ZZ-exchange spectra (Farrow et al. 1994) at 288 K with the exchange time set equal to the diffusion delay. The absence of cross-peaks between folded and unfolded resonances, even when the spectra are recorded at a very high signal-to-

noise ratio (>100:1) rules out the presence of appreciable exchange at this temperature. Moreover, by encoding the ^{15}N t_1 chemical shift evolution right after the gradient encoding, the presence of exchange between the folded and unfolded states, to first order, would not impact the decay of the auto peaks even if exchange were present (Choy et al. 2003). Thus, each resonance for folded and unfolded VA2-ubiquitin reports faithfully on the respective state without contamination from exchange.

Analogous diffusion measurements were also carried out under urea concentrations where both the folded and unfolded states of VA2-ubiquitin remain significantly populated. With the denaturation mid-point at 3.5 M urea, I collected diffusion data in 250 mM increments around this concentration. At 4 M urea, where the protein is predominantly unfolded, the R_h of the unfolded state (24.7 Å), remains very close to the value measured at 8 M urea (25.3 Å). Likewise, at 3 M urea where the unfolded state is populated to *ca.* 10%, the R_h only slightly decreases to 23.5 Å while the experimental uncertainty in the measurement increases due to the lower concentration of the unfolded species (Fig. 9B). Overall, only an 8% decrease in R_h between 8 M and 3 M urea was observed, whereas the unfolded state population drops from 100% to 20% over this range. Performing the same experiments on VA2-ubiquitin at 1000 bar (*ca.* 10% unfolded) and 3000 bar (100% unfolded) reveals a slightly larger, *ca.* 12% decrease in R_h at the lower pressure, where the folded state is favored (Fig. 9A). My results therefore indicate that the unfolded state of VA2-ubiquitin shows only a small degree of collapse when denaturing conditions are adjusted such that both folded and unfolded states are significantly populated (Fig. 9).

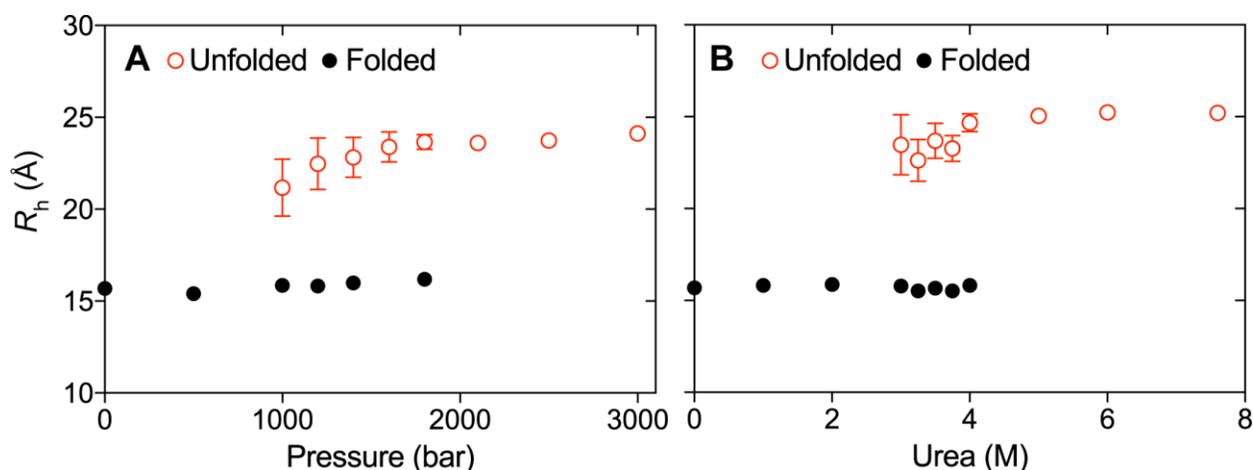


Figure 9 | Determination of the R_h of folded and unfolded VA2-ubiquitin under a range of denaturant conditions. The R_h values for both folded (black) and unfolded (red) VA2-ubiquitin were measured over a range of hydrostatic pressures (A) and urea concentrations (B). In general, only a small decrease in the R_h of the unfolded state is observed in low denaturant conditions (<10%). The folded protein does not exhibit significant changes in R_h , as its diffusion was only measurable to ca. 1800 bar and 4 M urea.

6.3.4 Diffusion measurements on equilibria of folded and unfolded protein by 1D PFG NMR

As a result of relaxation losses during the relatively long diffusion delay, coupled with the loss associated with the stimulated echo which refocuses only half of the initial magnetization, 2D PFG NMR diffusion experiments are intrinsically of rather low sensitivity. Coupled with the requirement to measure the weakest spectrum, *i.e.*, recoded with the strongest gradient, at adequate signal to noise, and reduced sample volumes available in high pressure measurements that require thick wall sample tubes, the above 2D PFG measurements were carried out at relatively high protein concentrations. For small molecules, conventional DOSY measurements can be applied to mixtures of molecules that contain overlapping resonances, or at selected resonances free of overlap between molecular species. For proteins, I find that even though this approach works well for the folded fraction of the sample, which typically contains well dispersed methyl resonances that resonate upfield of 0.5 ppm, no unique resonances of the unfolded protein, free of overlap with folded protein resonances, are available for diffusion measurements. In particular when the

unfolded fraction of protein is small, e.g. ca 10% at 1000 bar (0 M urea) or 3 M urea (1 bar) for VA2-ubiquitin, the effective concentration of unfolded protein is low even while using a quite concentrated sample (e.g. 1 mM). Consequently, such 2D PFG diffusion measurements can become prohibitively long. As such, I explored the feasibility of faster 1D NMR alternatives for diffusion measurements of such mixtures. The method, outlined below, requires a reference 1D NMR spectrum for the folded state (e.g., 1 bar, 0 M urea) to measure diffusion for samples where both folded and unfolded states are populated. Except for the number of scans, all experimental parameters were otherwise kept constant. An in-house Python script that utilizes the software NmrGlue [54] was then used to automatically execute the steps outlined below.

Commonly, folded proteins contain upfield shifted methyl resonances that are free of overlap with unfolded protein resonances. If not, invariably weaker, downfield shifted amide protons can be used. Shifting of the reference spectrum to match the resonance position of a unique resonance in the mixed sample (Fig. 10A), followed by downscaling of the reference spectrum such that the selected resonance (L50-C³²H₃ for my VA2-ubiquitin sample) in the folded reference spectrum matches the intensity in the mixed sample (Fig. 10B), then allows generation of a difference spectrum (Fig. 10C) that no longer contains a net integral of folded protein resonance intensity. Of course, it is critical that relative resonance intensities in the 1D reference spectrum are the same as those recorded with the 1D BPP-LED experiment for the folded component of the mixed sample. For this reason, the reference spectrum of the fully folded protein is also recorded with the 1D BPP-LED pulse sequence, using a weak encoding/decoding gradient combination. Even though resonances of the folded protein may have shifted somewhat relative to the reference spectrum, either due to the effect of pressure or weak interactions of denaturant with the folded protein, thereby generating a difference spectrum that contains negative components (Fig. 10C), the integral of the methyl region is minimally affected by such shifting, provided no resonances move outside of the integrated region. Integration regions were chosen

to include the methyl signals for the artificially reconstructed unfolded state, typically 0.7 – 0.9 ppm but results were found rather insensitive to the limits of the chosen region. Using this procedure to generate difference spectra for the unfolded state (Fig. 10D) then permits measuring the integrated unfolded protein intensity over the methyl region, at rather high sensitivity. The clean, single exponential decay of the integrated intensity with increased gradient strength (Fig. 10E) implies the method is reliable. Close agreement of the diffusion decay ratios with corresponding values extracted from *ca.* 7.5- to 15-fold lengthier 2D PFG diffusion measurements confirms robustness of the method (Fig. 10F).

Application of the 1D BPP-LED experiment to mixed samples above is demonstrated for the case of pressure-induced unfolding, where changes in pressure cause non-negligible shifts in peak positions. This application is more challenging than the study of, for example, denaturation by the addition of urea, where shifts in peak positions of the folded protein tend to be much smaller, thereby allowing a cleaner difference spectrum to be generated for the unfolded component. This was also applied on 2VA-ubq at 1400 bar at a total concentration of *ca.* 100 μM , where the amount of unfolded protein is near 50 μM . The higher sensitivity afforded by 1D experiments enabled acquisition of high S:N diffusion data at this low protein concentration (data not shown). The fitted R_h agrees with data obtained at 0.75 mM. Notably, a 50 μM total 2VA-ubq concentration (*ca.* 0.6 mg/mL) is lower than that typically encountered in SAXS (*ca.* 1 mg/mL).

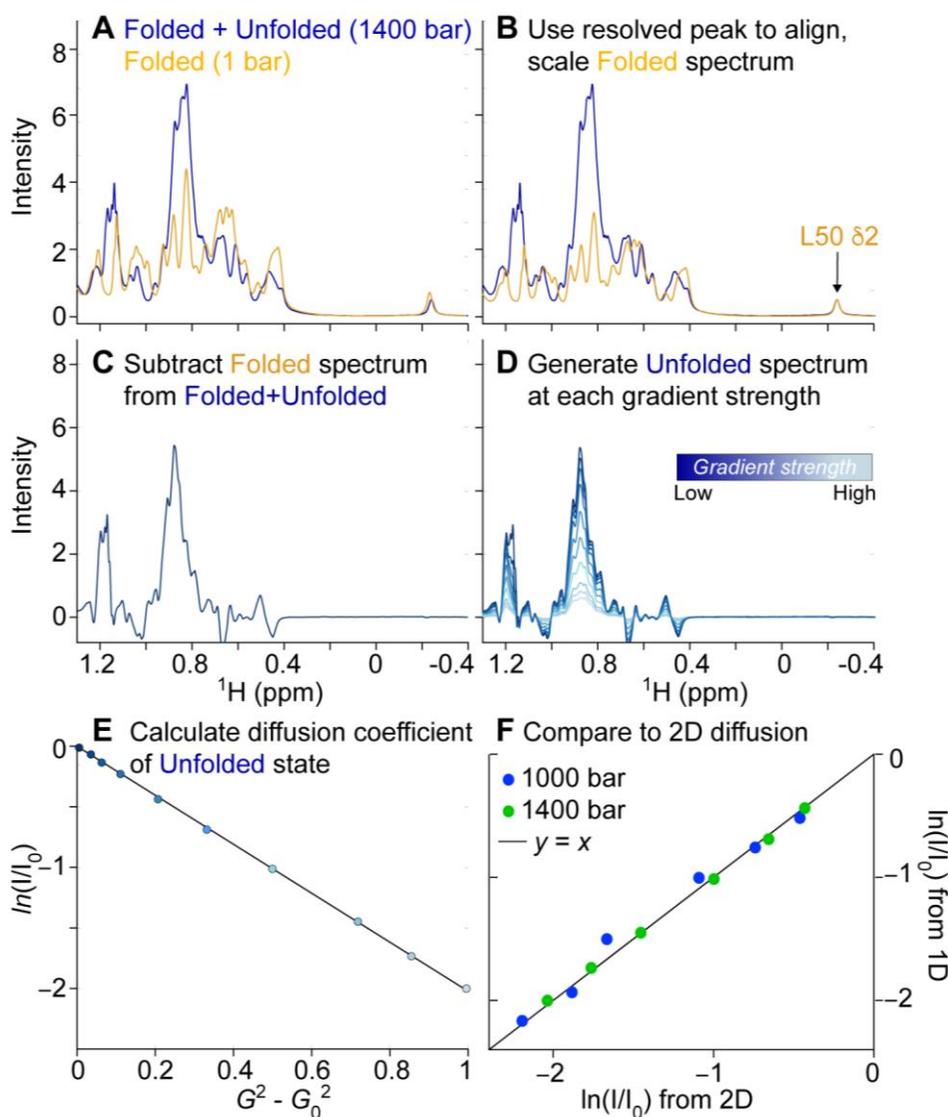


Figure 10 | Measurement of translational diffusion of the unfolded VA2-ubiquitin under conditions where a fraction is stably folded (A) The 1D NMR spectrum at 1400 bar reveals signals from both folded and unfolded protein (blue). The L50 $C^{\delta 2}H_3$ resonance at -0.3 ppm serves as a reference for the folded protein spectrum (orange). (B) The spectrum of the folded state is aligned and scaled such that L50 $C^{\delta 2}H_3$ matches that of the mixed sample. (C) Subtraction of the scaled, folded spectrum then yields an artificial spectrum of the unfolded state. (D) Repeating the procedure for spectra generated at different gradient strengths yields the series of 1D spectra for which integrals are over the methyl region are obtained. (E) Natural log plot of the decay in integrated resonance intensity is found to be highly linear. (F) Correlation plot showing the agreement between $\ln(I/I_0)$ values calculated from the 2D PFG NMR diffusion data set (x-axis) and those determined from the 1D method outlined above (y-axis). Data collected at 1000 bar (blue) and 1400 bar (green) are presented, and the solid line indicates $y = x$. Pairwise RMSD values between the respective $\ln(I/I_0)$ for the two datasets are 0.017 and 0.001, respectively, with the higher RMSD at 1000 bar dominated by the low S/N ratio of the 2D PFG spectra.

6.4 Conclusions

My work confirms that precise measurement of translational diffusion of proteins under elevated hydrostatic pressures is readily feasible, but that care needs to be exercised when following the common procedure of using an internal reference such as dioxane to derive the radius of hydration. My results for dioxane and a number of other small molecules indicate that their effective R_h values vary substantially with pressure, an effect that may be related to the strongly non-linear pressure dependence of their ^1H chemical shifts.

Probing the translational diffusion of both the folded and unfolded states of VA2-ubiquitin across a range where both states are significantly populated reveals only minor variation in their respective R_h values, both for the cases where high pressure and urea are used to induce the unfolding of the protein. Wild-type ubiquitin, which requires pressures much higher than 3 kbar to unfold, retains a nearly invariant R_h up to 2.5 kbar, but show a minor expansion at the highest pressure (3 kbar) attainable in the instrument, consistent with earlier structural work that showed the onset of large chemical shift changes at this pressure ([Kitahara et al. 2005](#)). The absence of any similarly large chemical shift changes in the folded VA2-ubiquitin spectrum upon traversing the unfolding range of conditions suggests the absence of any analogous pre-unfolding structural transition. The absence of any detectable change in R_h for the folded fraction of this protein therefore is not surprising, in particular in view of the low compressibility of folded globular proteins ([Gavish et al. 1983](#)).

More remarkable is the absence of significant compaction of the unfolded protein when conditions start favoring the folded state. This contrasts with earlier work on an SH3 domain, where PFG NMR diffusion found significant compaction relative to the fully denatured chain in the region where both folded and unfolded proteins co-exist ([Choy et al. 2003](#)). A number of single molecule FRET studies also provide strong evidence for compaction when the concentration of denaturant is reduced ([Deniz et al. 2000](#), [Schuler et al. 2002](#), [Michalet et al. 2006](#), [Moglich et al.](#)

2006) as did two-focus correlation spectroscopy (Hofmann et al. 2012), an effect that potentially could be attributed to transient H-bond formation (Holthauzen et al. 2010). However, whereas a small degree of compaction, as indicated by a 10-12% reduction in R_h , is observed for VA2-ubiquitin in the transition region, no further expansion of R_h is seen when moving further from the denaturing threshold. This latter observation suggests that transient sampling of states that might lead to a folded structure does not significantly impact the overall shape of the unfolded VA2-ubiquitin. Similar conclusions were reached for several other systems using small angle X-ray scattering (SAXS) measurements (Yoo et al. 2012, Watkins et al. 2015). Analogous to a prior study that found no effect of urea on the SAXS-derived R_g of an intrinsically disordered protein (IDP), I also find no significant effect of pressure on the R_h of α S. I note, however, that measurements for α S needed to be carried out at a relatively low concentration of 250 μ M. Above 1 mM, a concentration-dependent increase in R_h is seen, indicative of transient intermolecular interaction, an effect previously noted for other disordered systems (Noppert et al. 1996).

As recently discussed, differences in conclusions regarding the degree of collapse of unfolded proteins derived from FRET and SAXS measurements results from the decoupling of size and shape fluctuations, and such measurements therefore contain complementary, semi-independent information. The same applies for SAXS- and PFG NMR-derived R_g and R_h values, an effect previously exploited to generate improved ensembles for an unfolded SH3 domain (Choy et al. 2003).

6.5 Methods

6.5.1 Protein expression and purification

¹⁵N-labeled VA2-ubiquitin and α S were expressed and purified as described previously (Alderson et al. 2018). For VA2-ubiquitin, a final HPLC purification step was essential to avoid protease contamination, which efficiently degrades sample at high pressures where VA2-ubiquitin unfolds. WT ubiquitin was purchased from Sigma-Aldrich (USA). All protein NMR samples were prepared in 30 mM sodium phosphate at pH 6.4. Protein concentrations ranged from 0.25 mM (α S) to 0.75 mM (WT-ubiquitin, and VA2-ubiquitin used for urea studies) or 1.35 mM (VA2-ubiquitin for analysis of diffusion variation with pressure). Samples were doped with 0.1% dioxane (v/v) to facilitate the comparison of diffusion coefficients under otherwise identical conditions, enabling calculation of the hydration radius for the protein (eqn 3), as outlined previously (Wilkins et al. 1999):

$$D = \frac{k_B T}{6 \pi \eta R_h} \quad (1)$$

$$I_j = I_0 e^{-[\gamma^2 G_j^2 \delta^2 (\Delta - \frac{\delta}{3} - \frac{\tau}{2}) D]} \quad (2)$$

$$R_h^{unknown} = \frac{D_{reference}}{D_{unknown}} \times R_h^{known} \quad (3)$$

where in eqn 1, D refers to the diffusion coefficient, k_B to Boltzmann's constant, T to the temperature, η to the solution viscosity; in eq 2, I_j refers to the intensity in spectrum j , I_0 to the intensity in the reference spectrum, γ to the gyromagnetic ratio, G_j to the gradient strength in spectrum j , δ to the duration of the gradient pulse, τ to the gradient recovery delay, and Δ is defined as the diffusion time, T , plus 2τ and δ as in Fig. 1B of reference (Chou et al. 2004); and in eqn 3,

$D_{\text{reference}}$ refers to the diffusion coefficient of the molecule of known R_h , with D_{unknown} and R_h^{unknown} respectively referring to the diffusion coefficient and hydration radius of the target molecule.

6.5.2 NMR spectroscopy

All NMR spectra were recorded on a 700-MHz Bruker Avance-III spectrometer equipped with a room-temperature, triple-axis gradient TXI probe, using a ceramic NMR tube rated for pressures up to 3000 bar (Daedalus Innovations, Philadelphia, PA) that was linked to an automatic pump system to regulate the hydrostatic pressure inside the sample cell [50].

To measure self-diffusion coefficients (D), the 1D stimulated echo longitudinal encode-decode pulse sequence with bipolar pulse paired gradients (BPP-LED) pulse sequence (Wu et al. 1995, Chou et al. 2004) was used. BPP gradients (Wider et al. 1994) strongly reduce eddy currents and thereby the requisite gradient recovery delays, while also minimizing ^1H - ^1H NOE losses during the diffusion delay (Chou et al. 2004). In my BPP-LED experiments, the strongest encode and decode gradient strengths were respectively set to 53% and -47% (VA2-ubiquitin, WT-ubiquitin, tracers) or 60% and -56% (αS) from their maximum value, with the x and y gradients applied at identical percentages. The maximum gradient strengths were measured at 52 G/cm for x and 55 G/cm for y , yielding a combined total of 76.4 G/cm. Transverse gradients were employed to minimize complications that arise from convection when using the z -axis gradient in isolation (Kiraly et al. 2016]. The encoding and decoding bipolar gradients were applied for 2×1.5 ms, with the diffusion delay typically set to 375 ms, 400 ms, or 200 ms to ensure sufficient signal attenuation of the unfolded state of WT- or VA2-ubiquitin, αS , and tracer molecules, respectively, with the BPP pulse pair slightly (ca. 12%) imbalanced (Pelta et al. 2002).

For diffusion measurement of the tracer molecules at 293 K, a solution of 92% H_2O , 4% D_2O , and 1% each (v/v) of dioxane, ethanol, methanol, and DMSO was prepared, to which 0.4 mg/mL of CuCl_2 was added. In addition, a second sample was prepared with 96% H_2O , 2% D_2O ,

and 0.5% each (v/v) of dioxane, ethanol, methanol, and DMSO. The Cu^{2+} ions served to paramagnetically shorten the T_1 time of water to ca. 1 s and the other molecules to ca. 2 s. Eight 1D spectra with different encoding and decoding gradient strengths were recorded in an interleaved manner, with the diffusion delay set to 200 ms. The average of the 12% imbalanced encoding/decoding gradients strengths were set to 6.7, 10, 16, 22, 26, 32, 36, and 40% of the maximum value (76.4 G/cm combined x and y gradients). For encoding and decoding both x and y gradients were used at the same percentage of their maximum. Eight scans per FID were recorded, yielding a total experimental time of 15 minutes, and strong detuning of the probe was used to prevent radiation damping effects, thereby removing the need for presaturating or otherwise suppressing the H_2O signal. These 1D PFG NMR diffusion spectra were recorded at hydrostatic pressures of 1, 500, 1000, 1500, 2000, 2500, and 3000 bar. Data were processed with NMRPipe (Delaglio et al. 1995) and analyzed using in-house Python and NmrGlue (Helmus et al. 2013) scripts to automatically extract the attenuation of intensities resulting from translational diffusion (I_{grad}/I_0), where I_0 refers to the intensity in the spectrum with the weakest encoding/decoding gradients and I_{grad} to the intensities in the spectra with higher gradient strengths. Errors were obtained from the variance-covariance matrix derived from a fit of $\ln(I_{\text{grad}}/I_0)$ versus gradient strength squared. The time domain data were apodized with a 3 Hz exponential line broadening function, zero-filled to four times the number of data points, and then baseline corrected. Representative NMR resonances of the tracer molecules at the maximum encode/decode gradient strength for each pressure are shown in Fig. 1.

For 1D PFG NMR diffusion measurements on VA2-ubiquitin, WT ubiquitin, and αS a similar procedure was performed, except for usage of a 7 Hz exponential line broadening function. The range of encode/decode gradient strengths were adjusted to enable accurate measurement of the diffusion coefficients of both protein and dioxane molecules, and using only the data attenuated by less than 10-fold for fitting the diffusion decay constants. To this end, 11 spectra were recorded

with gradient strengths varied between 4 and 50% (WT, VA2-ubiquitin) or 58% (α S) of the maximum value. Typically, 80 (α S) or 160 (WT- and VA2-ubiquitin) scans per FID were recorded with a recycle delay of 1.5 s, yielding a total experimental time of 34 (α S) or 68 minutes (WT-, VA2-ubiquitin) per set of 11 spectra. Errors were obtained as above. Diffusion delays of 375 and 400 ms were used for ubiquitin and α S, respectively. For samples where the unfolded state of VA2-ubiquitin was lowly populated (e.g. 1000, 1200 bar), 320 scans per FID were collected for an experimental acquisition time of 136 minutes. 1D PFG NMR diffusion spectra were recorded for each protein at pressures ranging from 1 bar to 3000 bar in 500 bar increments, with finer pressure increments used in the pressure range where both folded and unfolded VA2-ubiquitin can be observed. For VA2-ubiquitin, diffusion data were also collected at urea concentrations of 0, 1, 2, 3, 3.25, 3.5, 3.75, 4, 5, 7, and 8 M.

2D PFG NMR diffusion data were collected as pseudo-3D spectra, comprising eight interleaved datasets of $40^* \times 1024^*$ points (where N^* indicates N complex data points) with acquisition times of 26 and 113 ms (t_1 , t_2) and a recycle delay of 1 s. Depending on the population of the unfolded state, either 8, 16, or 64 scans were acquired per FID, yielding a total experimental duration of 2.15 h (8 scans), 4.3 h (16 scans), or 17.2 h (64 scans). The diffusion delay (375 ms), maximum x/y gradient strengths, and gradient durations (2×1.5 ms) were set identically to those in the 1D PFG NMR datasets, to facilitate quantitative comparison of the results (Fig. 11). To decouple ^{15}N spins during t_2 acquisition, WALTZ-16 decoupling with a 1.39 kHz field was employed. Decreasing the decoupling field to 0.69 kHz had no impact on the measured diffusion coefficient, indicating that potential sample heating from decoupling pulses did not affect the analysis (Fig. 11). 2D data were processed with NMRPipe (Delaglio et al. 1995) with the time domain of the indirect dimension extended by 50% using the SMILE program (Ying et al. 2017), followed by apodization with a cosine-squared window. Spectra were visualized with NMRFAM-Sparky (Lee et al. 2015), and peak shapes were fit with FuDA (Vallurupalli et al. 2008). Intensities

from resonances that arose from the folded or unfolded states were grouped together and fit to a single exponential decay, with errors in the fitted diffusion coefficient estimated from a standard jackknife procedure when using different combinations of peaks to represent the folded or unfolded protein.

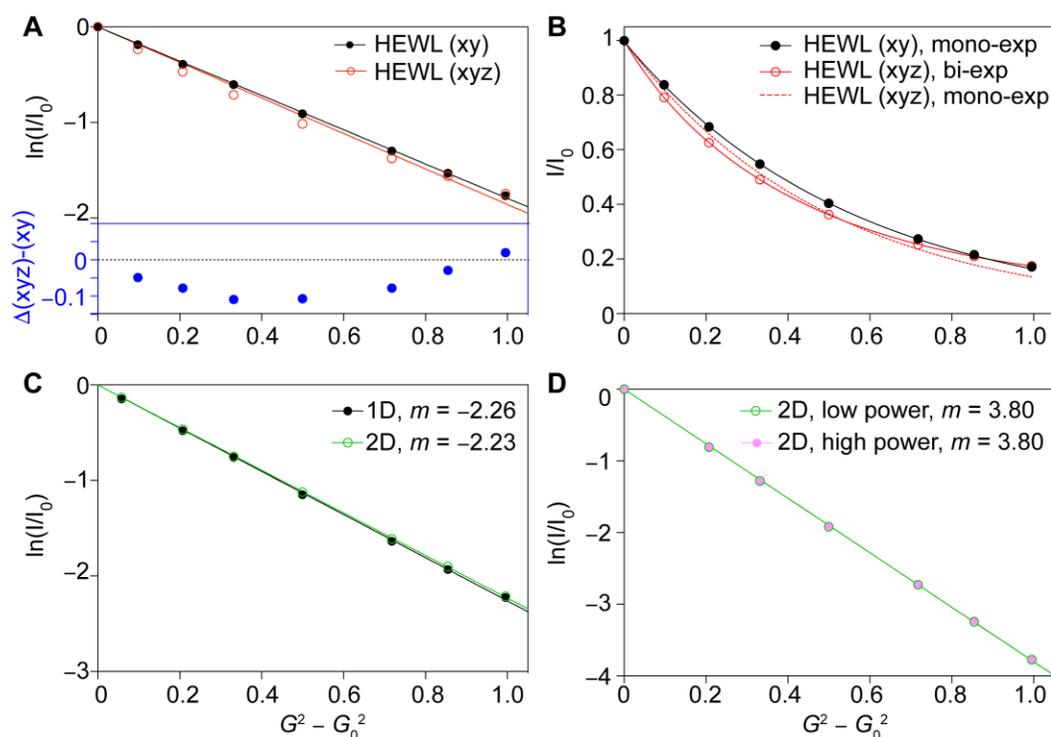


Figure 11 | Calibration of 1D and 2D PFG diffusion experiments. (A) The translational diffusion of hen egg white lysozyme (HEWL, 2.5 mM concentration, 30 mM sodium phosphate, pH 7, 288 K) was measured using the BPP-LED pulse sequence. A final concentration of 0.1% dioxane (v/v) was included for reference. The diffusion delay was set to 200 ms and the duration of the encoding/decoding gradients was 1.5 ms. The x-, y-, and z-components of the encoding and decoding gradients were set to either $\pm 27\%$ (x/y/z) or $\pm 50\%$ (x/y). In the experiment that utilized the z-gradient, significant deviations from mono-exponential signal decay were observed (red points), whereas x/y gradients produced mono-exponential behavior (blue points). Presumably this manifestation is due to convection, the effect of which is exacerbated by z-gradients (Kiraly et al. 2016, Swan et al. 2015). All translational diffusion experiments henceforth utilized x/y gradients only. (B) Data from (A) fit to exponential functions, either bi- or mono-exponential. (C) $\ln(I/I_0)$ values determined for 2VA-ubq using either a 2D (x-axis) or 1D (y-axis) PFG NMR diffusion experiment with the diffusion time set to 280 ms and the gradient duration set to 1.5 ms. The near identical signal attenuation enables direct comparisons to be made between 2D and 1D diffusion experiments. (D) Lowering the power level by a factor of four (1.39 kHz to 0.69 kHz, WALTZ-16 decoupling scheme) in the 2D PFG NMR diffusion experiment had no impact on the measured signal attenuation. Thus, a negligible amount of sample heating from ^{15}N decoupling occurs during t_2 acquisition, with no impact on translational diffusion. In (C) and (D), m refers to the fitted slope.

6.6 References

- Alderson TR, Charlier C, Torchia DA, Anfinrud P, Bax A. 2017. Monitoring Hydrogen Exchange During Protein Folding by Fast Pressure Jump NMR Spectroscopy, *J. Am. Chem. Soc.* 139: 11036-11039.
- Altieri AS, Hinton DP, Byrd RA. 1995. Association of Biomolecular Systems Via Pulsed-Field Gradient Nmr Self-Diffusion Measurements, *J. Am. Chem. Soc.* 117: 7566-7567.
- Baldwin AJ, Hilton GR, Lioe H, Bagn ris C, Benesch JLP, Kay LE. (2011) Quaternary Dynamics of α B-Crystallin as a Direct Consequence of Localised Tertiary Fluctuations in the C-Terminus, *J. Mol. Biol.* 413: 310-320.
- Baldwin AJ, Anthony-Cahill SJ, Knowles TJP, Lippens G, Christodoulou J, Barker PD, Dobson CM. (2008) Measurement of Amyloid Fibril Length Distributions by Inclusion of Rotational Motion in Solution NMR Diffusion Measurements, *Angew. Chem.* 47: 3385-3387.
- Beck M, Koehler J, Moeser B, Horinek D, Kremer W, Kalbitzer HR. (2014) Relationship between Nonlinear Pressure-Induced Chemical Shift Changes and Thermodynamic Parameters, *J. Phys. Chem. B* 118: 5681-5690.
- Benedek GB, Purcell EM. 1954. Nuclear magnetic resonance in liquids under high pressure, *J. Chem. Phys.* 22: 2003-2012.
- Bernad  P,  kerud T, Garc a de la Torre J., Akke M, Pons M. 2003. Combined Use of NMR Relaxation Measurements and Hydrodynamic Calculations To Study Protein Association. Evidence for Tetramers of Low Molecular Weight Protein Tyrosine Phosphatase in Solution, *J. Am. Chem. Soc.* 125: 916-923.
- Bertoncini CW, Jung YS, Fernandez CO, Hoyer W, Griesinger C, Jovin TM, Zweckstetter M. 2005. Release of long-range tertiary interactions potentiates aggregation of natively unstructured alpha-synuclein, *Proc. Natl. Acad. Sci. U. S. A.* 102: 1430-1435.
- Best RB, Zheng W, Borgia A, Buholzer K, Borgia MB, Hofmann H, Soranno A, Nettels D, Gast D, Grishaev A, Schuler B. 2018. Comment on "Innovative scattering analysis shows that hydrophobic disordered proteins are expanded in water", *Science* 361: eaar7101.
- Borgia A, Zheng W, Buholzer K, Borgia MB, Sch ler A, Hofmann H, Soranno A, Nettels D, Gast K, Grishaev A, Best RB, Schuler B. 2016. Consistent View of Polypeptide Chain Expansion in Chemical Denaturants from Multiple Experimental Methods, *J. Am. Chem. Soc.* 138: 11714-11726.
- Brady JP, Farber PJ, Sekhar A, Lin Y-H, Huang R, Bah A, Nott TJ, Chan HS, Baldwin AJ, Forman-Kay JD, Kay LE. 2017. Structural and hydrodynamic properties of an intrinsically disordered region of a germ cell-specific protein on phase separation, *Proc. Natl. Acad. Sci. U S A* 114: E8194-E8203.
- Buevich AV, Baum J. 2002. Residue-specific real-time NMR diffusion experiments define the association states of proteins during folding, *J. Am. Chem. Soc.* 124: 7156-7162.
- Caro JA, Wand AJ. 2018. Practical aspects of high-pressure NMR spectroscopy and its applications in protein biophysics and structural biology, *Methods* 148: 67-80.
- Charlier C, Alderson TR, Courtney JM, Ying J, Anfinrud P, Bax A. 2018a. Study of protein folding under native conditions by rapidly switching the hydrostatic pressure inside an NMR sample cell, *Proc. Natl. Acad. Sci. U S A* 115: 201803642.

- Charlier C, Courtney JM, Alderson TR, Anfinrud P, Bax A. 2018b. Monitoring ^{15}N Chemical Shifts During Protein Folding by Pressure-Jump NMR, *J. Am. Chem. Soc.* 140: 8096-8099.
- Chou JJ, Baber JL, Bax A. 2004. Characterization of Phospholipid Mixed Micelles by Translational Diffusion, *J. Biomol. NMR* 29: 299-308.
- Choy W-Y, Mulder FAA, Crowhurst KA, Muhandiram DR, Millett IS, Doniach S, Forman-Kay JD, Kay LE. 2002. Distribution of molecular size within an unfolded state ensemble using small-angle X-ray scattering and pulse field gradient NMR techniques, *J. Mol. Biol.* 316: 101-112.
- Defries T, Jonas J. 1977. Pressure-dependence of NMR proton spin-lattice relaxation-times and shear viscosity in liquid water in temperature range -15-10 degrees C, *J. Chem. Phys.* 66: 896-901.
- Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, A. Bax. 1995. NMRpipe - a multidimensional spectral processing system based on Unix pipes, *J. Biomol. NMR* 6: 277-293.
- Deniz AA, Laurence TA, Beligere GS, Dahan M, Martin AB, Chemla DS, Dawson PE, Schultz PG, Weiss S. 2000. Single-molecule protein folding: Diffusion fluorescence resonance energy transfer studies of the denaturation of chymotrypsin inhibitor 2, *Proc. Natl. Acad. Sci. U. S. A.* 97: 5179-5184.
- Didenko T, Boelens R, Rudiger SGD. 2011. 3D DOSY-TROSY to determine the translational diffusion coefficient of large protein complexes, *Prot. Eng. Des. Sel.* 24: 99-103.
- Dingley A, Mackay J, Chapman B, Morris M, Kuchel P, Hambly B, King G. 1995. Measuring protein self-association using pulsed-field-gradient NMR spectroscopy: Application to myosin light chain 2, *J. Biomol. NMR* 6: 321-328.
- Farrow NA, Zhang OW, Forman-Kay JD, Kay LE. 1994. A Heteronuclear Correlation Experiment For Simultaneous Determination of N-15 Longitudinal Decay and Chemical-Exchange Rates of Systems in Slow Equilibrium, *J. Biomol. NMR* 4: 727-734.
- Fuertes G, Banterle N, Ruff KM, Chowdhury A, Pappu RV, Svergun DI, Lemke EA. 2018. Comment on "Innovative scattering analysis shows that hydrophobic disordered proteins are expanded in water", *Science* 361: eaau8230.
- Gavish B, Gratton E, Hardy CJ. 1983. Adiabatic compressibility of globular proteins, *Proc. Natl. Acad. Sci. U S A* 80: 750-754.
- Helmus JJ, Jaroniec CP. 2013. Nmrplug: an open source Python package for the analysis of multidimensional NMR data, *J. Biomol. NMR* 55: 355-367.
- Herberhold H, Winter R. 2002. Temperature- and pressure-induced unfolding and refolding of ubiquitin: A static and kinetic Fourier transform infrared spectroscopy study, *Biochem.* 41: 2396-2401.
- Hofmann H, Soranno A, Borgia A, Gast K, Nettels D, Schuler B. 2012. Polymer scaling laws of unfolded and intrinsically disordered proteins quantified with single-molecule spectroscopy, *Proc. Natl. Acad. Sci. U. S. A.* 109: 16155-16160.
- Holthausen LMF, Rosgen J, Bolen DW. 2010. Hydrogen Bonding Progressively Strengthens upon Transfer of the Protein Urea-Denatured State to Water and Protecting Osmolytes, *Biochemistry* 49: 1310-1318.
- Horst R, Horwich AL, Wüthrich K. 2011. Translational diffusion of macromolecular assemblies measured using transverse-relaxation-optimized pulsed field gradient NMR., *J. Am. Chem. Soc.* 133: 16354-16357.

- Huang R, Ripstein ZA, Augustyniak R, Lazniewski M, Ginalski K, Kay LE, Rubinstein JL. 2016. Unfolding the mechanism of the AAA+ unfoldase VAT by a combined cryo-EM, solution NMR study, *Proc. Natl. Acad. Sci. U S A* 113: E4190-E4199.
- Huang R, Brady JP, Sekhar A, Yuwen T, Kay LE. 2017. An enhanced sensitivity methyl ¹H triple-quantum pulse scheme for measuring diffusion constants of macromolecules, *J. Biomol. NMR* 68: 249-255.
- Jansma AL, Kirkpatrick JP, Hsu AR, Handel TM, Nietlispach D. 2010. NMR analysis of the structure, dynamics, and unique oligomerization properties of the chemokine CCL27, *J. Biol. Chem.* 285: 14424-14437.
- Johnson, CS. 1999. Diffusion ordered nuclear magnetic resonance spectroscopy: principles and applications, *Prog. Nucl. Magn. Reson. Spectrosc.* 34: 203-256.
- Jones JA, Wilkins DK, Smith LJ, Dobson CM. 1997. Characterisation of protein unfolding by NMR diffusion measurements, *J. Biomol. NMR* 10: 199-203.
- Kalbitzer HR, Rosnizeck IC, Munte CE, Narayanan SP, Kropf V, Spoerner M. 2013. Intrinsic Allosteric Inhibition of Signaling Proteins by Targeting Rare Interaction States Detected by High-Pressure NMR Spectroscopy, *Angew. Chem.* 52: 14242-14246.
- Kamatari YO, Yamada H, Akasaka K, Jones JA, Dobson CM, Smith LJ. 2001. Response of native and denatured hen lysozyme to high pressure studied by ¹⁵N-¹H NMR spectroscopy, *Eur. J. Biochem.* 268: 1782-1793.
- Kay LE, Torchia DA, Bax A. 1989. Backbone dynamics of proteins as studied by ¹⁵N inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease., *Biochemistry* 28: 8972-8979.
- Kiraly P, Swan I, Nilsson M, Morris GA. 2016. Improving accuracy in DOSY and diffusion measurements using triaxial field gradients, *J. Magn. Reson.* 270: 24-30.
- Kitahara R, Yokoyama S, Akasaka K. 2005. NMR snapshots of a fluctuating protein structure: Ubiquitin at 30 bar-3 kbar, *J. Mol. Biol.* 347: 277-285.
- Kitahara R, Hata K, Li H, Williamson MP, Akasaka K. 2013. Pressure-induced chemical shifts as probes for conformational fluctuations in proteins, *Prog. Nucl. Magn. Reson. Spectrosc.* 71: 35-58.
- Lee D, Hilty C, Wider G, Wuthrich K. 2006. Effective rotational correlation times of proteins from NMR relaxation interference, *J. Magn. Reson.* 178: 72-76.
- Lee W, Tonelli M, Markley JL. 2015. NMRFAM-SPARKY: enhanced software for biomolecular NMR spectroscopy., *Bioinformatics* 31: 1325-1327.
- Leung RLC, Robinson MDM, Ajabali AAA, Karunanithy G, Lyons B, Raj R, Raoufmoghaddam S, Mohammed S, Claridge TDW, Baldwin AJ, Davis BG. 2017. Monitoring the Disassembly of Virus-like Particles by ¹⁹F-NMR, *J. Am. Chem. Soc.* 139: 5277-5280.
- Maeno A, Sindhikara D, Hirata F, Otten R, Dahlquist FW, Yokoyama S, Akasaka K, Mulder FAA, Kitahara R. 2015. Cavity as a Source of Conformational Fluctuation and High-Energy State: High-Pressure NMR Study of a Cavity-Enlarged Mutant of T4Lysozyme, *Biophys. J.* 108: 133-145.
- Michalet X, Weiss S, Jager M. 2006. Single-molecule fluorescence studies of protein folding and conformational dynamics, *Chem. Rev.* 106: 1785-1813.

- Moglich A, Joder K, Kiefhaber T. 2006. End-to-end distance distributions and intrachain diffusion constants in unfolded polypeptide chains indicate intramolecular hydrogen bond formation, *Proc. Natl. Acad. Sci. U. S. A.* 103: 12394-12399.
- Mok Y-K, Kay CM, Kay LE, Forman-Kay JD. 1999. NOE data demonstrating a compact unfolded state for an SH3 domain under non-denaturing conditions, *J. Mol. Biol.* 289 (1999) 619-638.
- Nilsson M, Duarte IF, Almeida C, Delgadillo I, Goodfellow BJ, Ana A, Gil M, Morris GA. 2004. High-Resolution NMR and Diffusion-Ordered Spectroscopy of Port Wine, *J. Agric. Food Chem.* 52: 3736-3743.
- Noppert A, Gast K, MullerFrohne M, Zirwer D, Damaschun G. 1996. Reduced-denatured ribonuclease A is not in a compact state, *FEBS Lett.* 380: 179-182.
- Nucci NV, Fuglestad B, Athanasoula EA, Wand AJ. 2014. Role of cavities and hydration in the pressure unfolding of T4 lysozyme, *Proc. Natl. Acad. Sci. U S A* 111: 13846-13851.
- Pelta MD, Morris GA, Stchedroff MJ, Hammond SJ. 2002. A one-shot sequence for high-resolution diffusion-ordered spectroscopy, *Magn. Reson. Chem.* 40: S147-S152.
- Peterson RW, Wand AJ. 2005. Self-contained high-pressure cell, apparatus, and procedure for the preparation of encapsulated proteins dissolved in low viscosity fluids for nuclear magnetic resonance spectroscopy, *Rev. Sci. Instrum.* 76.
- Price, WS. 1997. Pulsed-field gradient nuclear magnetic resonance as a tool for studying translational diffusion 1. Basic theory, *Concepts Magn. Reson.* 9: 299-336.
- Price, WS. 1998. Pulsed-field gradient nuclear magnetic resonance as a tool for studying translational diffusion: Part II. Experimental aspects, *Concepts Magn. Resonance* 10: 197-237.
- Price WS, Tsuchiya F, Arata Y. 1999. Lysozyme aggregation and solution properties studied using PGSE NMR diffusion measurements, *J. Am. Chem. Soc.* 121: 11503-11512.
- Price, WS. 2009. NMR studies of translational motion: Principles and applications, Cambridge University Press (England).
- Riback JA, Bowman MA, Zmyslowski AM, Knoverek CR, Jumper JM, Hinshaw JR, Kaye, EB Freed KF, Clark PL, Sosnick TR. 2017. Innovative scattering analysis shows that hydrophobic disordered proteins are expanded in water, *Science* 358: 238-241.
- Riback JA, Bowman MA, Zmyslowski AM, Knoverek CR, Jumper JM, Hinshaw JR, Kaye, EB Freed KF, Clark PL, Sosnick TR. 2018. Response to Comment on "Innovative scattering analysis shows that hydrophobic disordered proteins are expanded in water", *Science* 361: eaar7949.
- Roche J, Caro JA, Norberto DR, Barthe P, Roumestand C, Schlessman JL, Garcia AE, Garcia-Moreno B, Royer CA. 2012. Cavities determine the pressure unfolding of proteins, *Proc. Natl. Acad. Sci. U. S. A.* 109: 6945-6950.
- Roche J, Louis JM, Bax A, Best RB, 2015. Pressure-induced structural transition of mature HIV-1 protease from a combined NMR/MD simulation approach, *Proteins* 83: 2117-2123.
- Roche J, Royer CA. Roumestand C. 2017. Monitoring protein folding through high pressure NMR spectroscopy, *Prog. Nucl. Magn. Reson. Spectrosc.* 102-103: 15-31.
- Royer CA, Hinck AP, Loh SN, Prehoda KE, Peng XD, Jonas J, Markley JL. 1993. Effects of Amino-Acid Substitutions on the Pressure Denaturation of Staphylococcal Nuclease as Monitored by Fluorescence and Nuclear-Magnetic-Resonance Spectroscopy, *Biochemistry* 32: 5222-5232.

- Schuler B, Lipman EA, Eaton WA. 2002. Probing the free-energy surface for protein folding with single-molecule fluorescence spectroscopy, *Nature* 419: 743-747.
- Silva JL, Weber G. 1993. Pressure stability of proteins, *Annu. Rev. Phys. Chem.* 44: 89-113.
- Soper AK, Ricci MA. 2000. Structures of high-density and low-density water, *Phys. Rev. Lett.* 84: 2881-2884.
- Stejskal EO, Tanner JE. 1965. Spin Diffusion Measurements: Spin Echoes in the Presence of a Time-Dependent Field Gradient, *J. Chem. Phys.* 42: 288-292.
- Stilbs P. 1987. Fourier transform pulsed-gradient spin-echo studies of molecular diffusion, *Prog. Nucl. Magn. Reson. Spectrosc.* 19: 1-45.
- Tanner JE. 1970. Use of the Stimulated Echo in NMR Diffusion Studies, *J. Chem. Phys.* 52: 2523-2526.
- Vajpai N, Nisius L, Wiktor M, Grzesiek S. 2013. High-pressure NMR reveals close similarity between cold and alcohol protein denaturation in ubiquitin, *Proc. Natl. Acad. Sci. U. S. A.* 110: E368-E376.
- Vallurupalli P, Hansen DF, Kay LE. 2008. Structures of invisible, excited protein states by relaxation dispersion NMR spectroscopy, *Proc. Natl. Acad. Sci. U. S. A.* 105: 11766-11771.
- Watkins HM, Simon AJ, Sosnick TR, Lipman EA, Hjelm RP, Plaxco KW. 2015. Random coil negative control reproduces the discrepancy between scattering and FRET measurements of denatured protein dimensions, *Proc. Natl. Acad. Sci. U. S. A.* 112: 6631-6636.
- Weljie AM, Yamniuk AP, Yoshino H, Izumi Y, Vogel HJ. 2003. Protein conformational changes studied by diffusion NMR spectroscopy: application to helix-loop-helix calcium binding proteins., *Prot. Sci.* 12: 228-236.
- Wider G, Dotsch V, Wuthrich K. 1994. Self-Compensating Pulsed Magnetic-Field Gradients for Short Recovery Times, *J. Mag. Res.* 108: 255-258.
- Wilkins DK, Grimshaw SB, Receveur V, Dobson CM, Jones JA, Smith LJ. 1999. Hydrodynamic radii of native and denatured proteins measured by pulse field gradient NMR techniques, *Biochemistry* 38: 16424-16431.
- Wu DH, Chen AD, Johnson CS. 1995. An improved diffusion-ordered spectroscopy experiment incorporating bipolar-gradient pulses, *J. Mag. Reson.* 115: 260-264.
- Ying J, Delaglio F, Torchia DA, Bax A. 2017. Sparse multidimensional iterative lineshape-enhanced (SMILE) reconstruction of both non-uniformly sampled and conventional NMR data, *J. Biomol. NMR* 68 101-118.
- Yoo TY, Meisburger SP, Hinshaw J, Pollack L, Haran G, Sosnick TR, Plaxco K. (2012) Small-Angle X-ray Scattering and Single-Molecule FRET Spectroscopy Produce Highly Divergent Views of the Low-Denaturant Unfolded State, *J. Mol. Biol.* 418: 226-236.

7

Measurement of *cis*-proline in disordered and unfolded proteins

7.1 Abstract

In unfolded proteins, peptide bonds involving Pro residues exist in an equilibrium between the minor *cis* and major *trans* conformations. Folded proteins predominantly contain *trans*-Pro bonds, and slow *cis-trans* Pro isomerization in the unfolded state is often found to be a rate limiting step in protein folding. Moreover, kinases and phosphatases that act upon Ser/Thr-Pro motifs exhibit preferential recognition of either the *cis*- or *trans*-Pro conformer. Here, I determined the fraction of *cis*-Pro in four proteins that are intrinsically disordered or pressure-denatured. At both atmospheric and elevated pressures, the population of *cis*-Pro falls well below previous estimates, an effect attributed to the use of short peptides with charged termini in most prior model studies. For the intrinsically disordered protein α -synuclein, *cis*-Pro populations at all five of its X-Pro bonds are less than 5%, with only modest ionic strength dependence and no detectable effect of the previously demonstrated interaction between the N- and C-terminal halves of the protein. Comparison to small peptides with the same amino acid sequence indicates that peptides, particularly those with unblocked, oppositely charged amino and carboxyl end groups, strongly overestimate the amount of *cis*-Pro. Across 15 different X-Pro bonds measured in four proteins, the fractions of *cis*-Pro are 10% or less. ^{15}N spin relaxation experiments further indicate that *cis*- and *trans*-Pro conformers exhibit similar backbone dynamics. My data reveal that *cis*-Pro formation is an inherent property of unfolded polypeptide chains, with limited structural or dynamical differences between the *trans*- or *cis*-Pro-containing regions.

7.2 Introduction

Within proteins, the vast majority (>99.5%) of peptide bonds not involving proline exist in the *trans* conformation, in which the dihedral angle ω is 180° . The lowly populated *cis* conformer requires a 180° rotation about the planar CO(*i*-1)–N(*i*) peptide bond ($\omega = 0^\circ$), but such a rotation induces a steric clash between the C $^\alpha$ (*i*-1) and C $^\alpha$ (*i*) atoms. This creates a free energy difference between the *trans* and *cis* conformations of *ca.* 2-6 kcal/mol in non-Pro peptide bonds, and a high-energy barrier to rotation of the partial double bond (*ca.* 20 kcal/mol) overwhelmingly favors the *trans* state (Dugave & Demange 2003, Wedemeyer et al. 2002, Weiss et al. 1998). However, in peptide bonds between any amino acid (X) and proline (X-Pro), the *trans* and *cis* conformers have a substantially lower energy difference, owing to the cyclic nature of the proline side-chain. Thus, *cis*-peptidyl-prolyl (*cis*-Pro) conformations in unfolded polypeptide chains are populated to significantly higher levels, with values that range from 5 to 80% in model peptides (Deber et al. 1972, Dyson et al. 1988, Grathwohl & Wüthrich 1976, 1981; Osváth & Gruebele 2003, Raleigh et al. 1992, Reimer et al. 1998, Sarkar et al. 1984, Torchia 1972, Yao et al. 1994), depending on the precise amino acid composition, with little dependence on temperature (Ahuja et al. 2016, Osváth & Gruebele 2003, Raleigh et al. 1992, Reimer et al. 1998). In folded proteins, local interactions around X-Pro bonds typically induce 100% population of either the *cis* or *trans* conformation (Capaldi et al. 1999, Lu et al. 2007, Sarkar et al. 2007, Walport et al. 2015).

Cis-Pro bonds and their slow isomerization to the *trans* state, *ca.* 10^{-3} to 10^{-2} s $^{-1}$ at room temperature, depending on the types of adjacent residues (Reimer et al. 1998), can be the rate-limiting step in protein folding (Wedemeyer et al. 2002), as most non-native *cis*-Pro bonds in the unfolded protein require isomerization to the native *trans* conformations in order for folding to proceed. Indeed, a class of molecular chaperones has evolved to catalyze *cis-trans* proline isomerization in nascent polypeptides (Schmid 1993), and *in vitro* refolding studies have demonstrated that such peptidyl-prolyl isomerases, including the ribosome-localized trigger factor (Hesterkamp et al. 1996), enhance the rate of protein folding (Lang et al. 1987).

The human proteome contains 6.3% proline residues (Morgan et al. 2013), and intrinsically disordered proteins (IDPs) generally contain primary sequences that are enriched in proline by nearly 2-fold over folded proteins (Theillet et al. 2014). Biologically relevant post-translational modifications occur at numerous Ser-/Thr-Pro motifs, with some kinases and phosphatases specifically recognizing either the *cis*- or *trans*-proline conformation (Werner-Allen et al. 2011). Moreover, in chemically denatured proteins used for *in vitro* folding studies, native *trans*-Pro bonds will isomerize to the *cis* state upon denaturation and equilibration. For proteins with multiple Pro residues, knowledge of the fraction of *cis*-Pro at each X-Pro bond is critical for the resultant analysis of refolding, which typically contains multiple phases. However, the quantification of *cis*-Pro propensity in intact, denatured proteins or intrinsically disordered regions has remained limited (Ahuja et al. 2016, Alderson et al. 2017, Follis et al. 2015, Gibbs et al. 2017, Sarkar et al. 2007), with most studies instead opting to quantify *cis*-Pro populations in small, model peptides (Deber et al. 1972, Dyson et al. 1988, Grathwohl & Wüthrich 1976, 1981; Osváth & Gruebele 2003, Raleigh et al. 1992, Reimer et al. 1998, Sarkar et al. 1984, Torchia 1972, Yao et al. 1994).

Despite the often low abundance of *cis*-Pro conformers, NMR is well suited to characterize these states, as *cis-trans* Pro isomerization is in the slow exchange regime on the NMR timescale ($\gg 1 \text{ s}^{-1}$) (e.g. see Chapter 2). Therefore, separate resonances for the minor *cis* states are observable, with the respective peak intensities reporting on the concentrations of each conformer. Multi-dimensional NMR provides a spectroscopic probe at every backbone amide moiety, thereby enabling accurate, quantitative analysis of *cis*-Pro populations from multiple residues affected by the same Pro, as well as the structural impact of this isomerization. Here, I have employed two- and three-dimensional solution-state NMR to accurately quantify the fractions of *cis*-Pro at 15 different X-Pro bonds in three unfolded proteins, an IDP (α -synuclein) and two pressure-denatured proteins, the α -crystallin domain of HSP27 (cHSP27) and ubiquitin V17A/V26A (VA2-ubq). Comparison of these fractions to those measured for small peptides of identical sequence reveal

that peptides tend to overestimate the fraction of *cis*-Pro, in particular when the peptides contain oppositely charged N-terminal amino and C-terminal carboxyl groups. The relatively uniform impact of *cis*-Pro on the chemical shifts of nearby residues will facilitate identification of *cis*-Pro conformers in future NMR studies of IDPs and intrinsically disordered regions (IDRs) of proteins.

7.3 Results

7.3.1 Measurement of *cis*-proline fractions in the IDP α -synuclein

IDPs typically yield a single set of intense cross-peaks in the two-dimensional (2D) ^1H - ^{15}N HSQC spectrum. The HSQC spectrum of α -synuclein (αS) exemplifies such behavior, but weak signals from alternate conformers are observable at low contour levels (Fig. 1, Fig. 2). There are five Pro residues in the acidic C-terminal region of αS , and the minor signals corresponding to *cis*-Pro bonds cluster near resonances of residues that are close in sequence to Pro. To assign the chemical shifts from the *cis*-Pro conformers in αS , triple resonance NMR spectra were acquired on a 0.9 mM [U - ^{13}C , ^{15}N]-labeled sample. High signal-to-noise ratios in these spectra afforded unambiguous assignment of the residues in the vicinity of all *cis*-Pro conformers (P108, P117, P120, P128, P138), despite the low population of these states (3-4.5%, *i.e.*, 27-40 μM effective sample concentration).

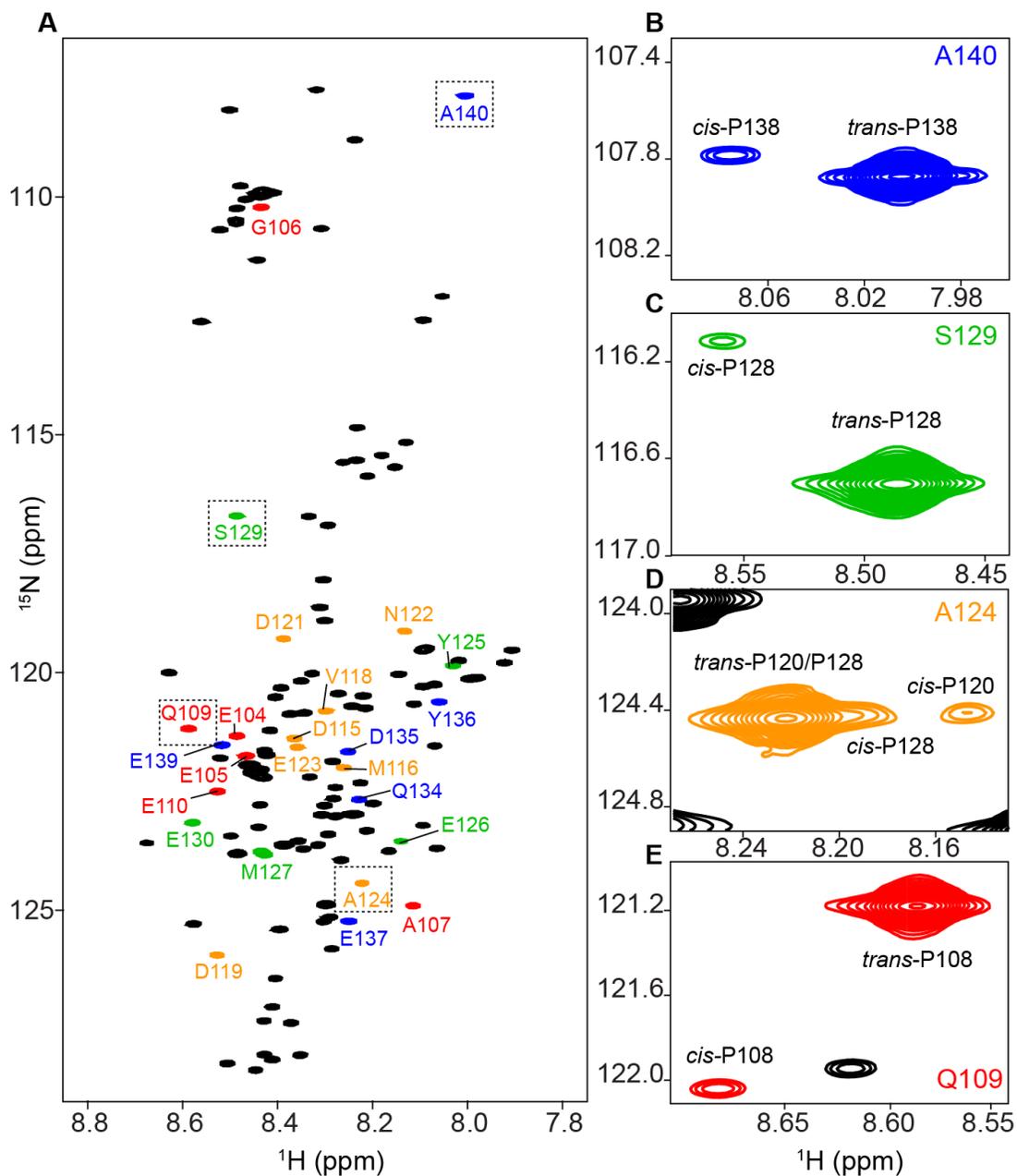


Figure 1. Identification of *cis*-Pro bonds in α S. **(A)** 2D ^1H - ^{15}N HSQC of 0.9 mM ^{13}C , ^{15}N - α S at pH 6, 288 K. Resonances impacted by the *cis*/*trans* state of nearby Pro residues are colored (P108, *red*; P117, P120, *orange*; P128, *green*; P138, *blue*). **(B-E)** Zoomed-in regions corresponding to the boxed areas from **(A)** are shown at a lower contour level. The major *trans*- and minor *cis*-Pro peaks are indicated with their assignments. See Figure 2 for the full, low-contour spectrum. Reproduced with permission from John Wiley and Sons.

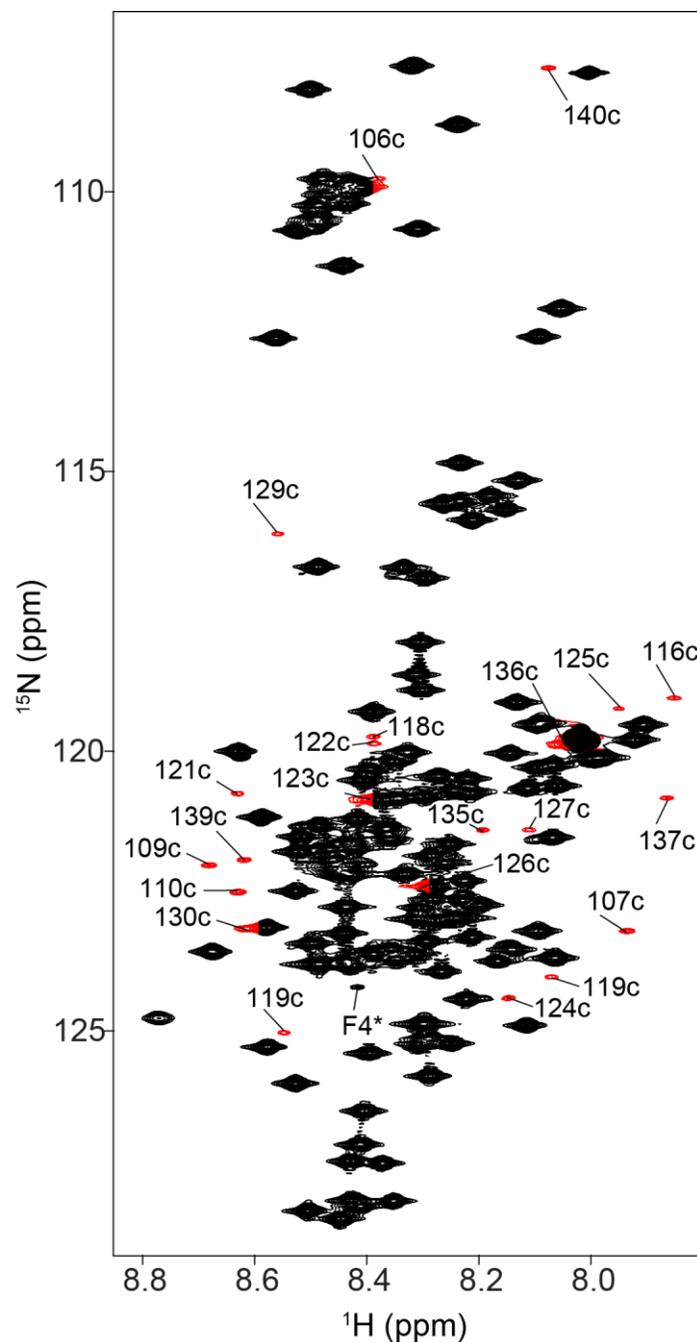


Figure 2. 2D ^1H - ^{15}N HSQC NMR spectrum of αS , with resonance assignments marked for residues impacted by *cis*-Pro formation. The sample contained 0.9 mM αS in 20 mM NaH_2PO_4 at pH 6, 2 mM EDTA, 2 mM benzamidine, 288 K. Assignments for resonances affected by *cis*-Pro formation are colored *red* whereas resonances corresponding to *trans*-Pro are shown in *black*. For each of the five Pro residues, the fraction of *cis* was less than 5%, resulting in an effective NMR-observable concentration of ≤ 40 μM protein. A total of 27 out of 140 residues are visibly impacted by *cis*-Pro formation (including the Pro residues, which are not observable in the spectrum above). The peak labeled F4* corresponds to a $\sim 3.5\%$ fraction of protein that has been oxidized at M5. Reproduced with permission from John Wiley and Sons.

In the *cis*-Pro state, the $i-1$ residue with respect to Pro shows significant upfield changes in chemical shifts for $^1\text{H}^{\text{N}}$ (ca. 0.2-0.4 ppm) and ^{15}N (ca. 2-4 ppm), with a smaller and usually upfield shift for $^{13}\text{C}^{\alpha}$ (Fig. 3). The small, downfield shift of E137 $^{13}\text{C}^{\alpha}$, in close proximity to the C-terminus, is an exception to this rule. Another diagnostic chemical shift change of the *cis* state is seen for the $i-2$ $^{13}\text{C}'$ resonance, which exhibits a substantial upfield change (ca. 0.4-1 ppm) that is similar in sign and magnitude to the $^{13}\text{C}'$ chemical shift change of the isomerizing Pro (Fig. 3). Residues in the $i+1$ and $i+2$ positions show downfield $^1\text{H}^{\text{N}}$ chemical shift changes in the *cis* state (up to 0.2 ppm), but variable ^{15}N chemical shift differences. The chemical shift differences typically become vanishingly small for residues in the $i\pm 4$ positions and beyond.

Using the peak intensities from resonances impacted by the *cis* and *trans* states, the *cis*-Pro population was calculated as $[\textit{cis}]/([\textit{cis}] + [\textit{trans}])$. In αS , the fraction of *cis*-Pro ranged from 3% to 4.5%, significantly lower than values generally reported in the literature (e.g. 10-20%) (Osváth & Gruebele 2003). The fraction of *cis*-Pro is highly consistent among the various impacted residues near a given Pro, and the signal-to-noise ratio for each *cis*-Pro-impacted cross peak was near ca. 80:1, indicating the high precision of these measurements. For *de novo* assignment purposes, peak intensities from the minor *cis* resonances can be used to alleviate potential ambiguities, in the absence of unambiguous triple resonance data.

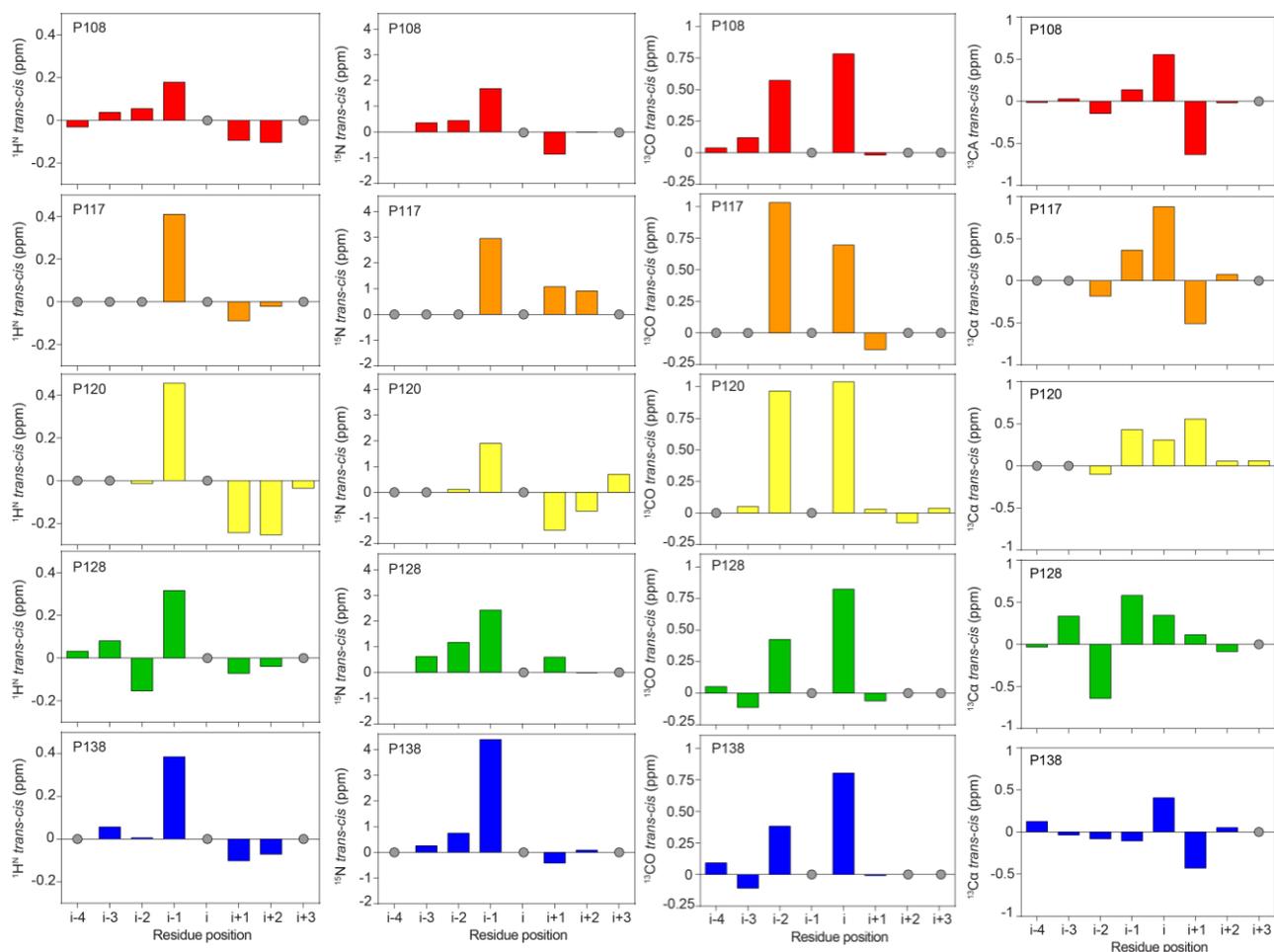


Figure 3. Chemical shift differences ($\omega_{trans} - \omega_{cis}$) between the *trans*- and *cis*-Pro-affected resonances in α S. Shown here are the differences for (1st column) $^1\text{H}^N$, (2nd column) ^{15}N , (3rd column) $^{13}\text{C}^\alpha$, and (4th column) $^{13}\text{C}^\alpha$. Diagnostic chemical shift changes include $^1\text{H}^N$ and ^{15}N for the *i*-1 residue with respect to the isomerizing Pro, which show large positive values. Since these resonances are readily observed in 2D ^1H - ^{15}N HSQC spectra, together with the constraint that the *cis/trans* intensity ratio must be the same for amides affected by a given Pro isomerization, can often be used for assigning the *cis*-isomer resonances to their nearby *trans* counterparts. $^{13}\text{C}^\alpha$ chemical shifts of *i* and *i*-2 residues show large upfield shifts for the *cis* isomer, and the same applies for the Pro $^{13}\text{C}^\alpha$. Reproduced with permission from John Wiley and Sons.

7.3.2 Small peptides from α -synuclein overestimate the fraction of *cis*-proline

To compare the *cis*-Pro values in α S to values of *cis*-Pro in small peptides of identical sequence, natural abundance NMR spectra were acquired on a set of tetra-peptides (Ac-XXPX-NH₂) that were N-terminally acetylated and C-terminally amidated. All of the blocked, tetra-peptides, except that corresponding to P120, displayed a significantly higher population of *cis*-Pro than full-length

α S (Fig. 4), indicating that longer-range interactions, presumably steric clashing or repulsive electrostatic interactions of the chains, are unfavorable to the formation of the *cis* state at P108, P117, P128, and P138. The highly negatively charged nature of the C-terminal tail of α S (net charge of -14 for the last 40 residues) is expected to enhance the electrostatic contribution. Indeed, at very high ionic strength (1.1 M NaCl), increased *cis* fractions are observed, albeit remaining below those of the blocked tetra-peptides (Fig. 4). NMR spectra were collected at such high salt concentrations by using a thick-walled tube with a smaller inner diameter (Methods).

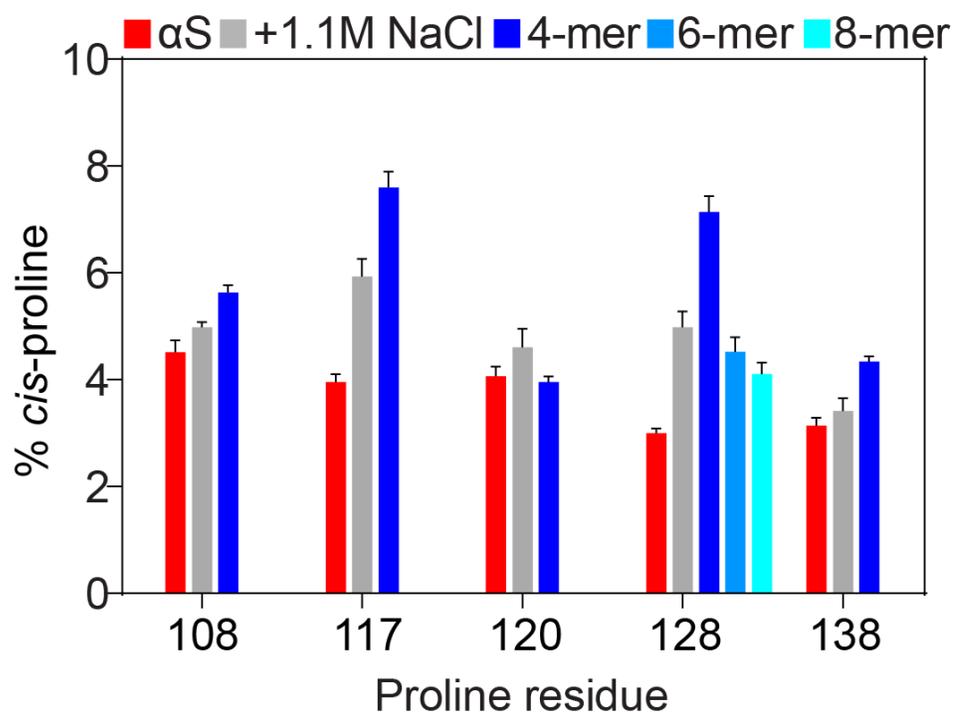


Figure 4. Fractions of *cis*-Pro bonds in α S and comparison to oligo-peptides. All samples were at pH 6, 288 K. The fraction of *cis*-Pro is shown for each Pro residue in α S (red), with error bars representing one standard deviation from the mean. Upon the addition of 1160 mM NaCl, the fraction of *cis*-Pro increases (grey), indicative of decreased electrostatic repulsion in the acidic C-terminal region of α S. Blocked, tetra-peptides (blue) in general contain elevated levels of *cis*-Pro, whereas hexa- (teal) and octa-peptides (cyan) approach the values seen in full length α S. Reproduced with permission from John Wiley and Sons.

High fractions of *cis*-Pro previously have been reported for peptides with an aromatic residue in the *i*-1 position (Thomas et al. 2006); however, these peptides mostly were studied with

unblocked, charged terminal residues. For example, a high *cis* fraction was previously reported for the FPA tri-peptide (Grathwohl & Wüthrich 1981). However, charge neutralization by N-terminal acetylation and C-terminal amidation reduces the *cis*-Pro fraction by more than one-fifth (Fig. 5). Analogously, the addition of 1.1 M NaCl to the unblocked peptide decreases *cis*-Pro by nearly the same fraction (Fig. 5). In the tri-peptide MPS, I observe even larger reductions in the *cis* fraction upon the addition of blocking groups at the termini or 1.1 M NaCl, which decrease the *cis*-Pro fraction by about a third and a quarter, respectively (Fig. 5). Similar results were obtained for the peptide APQ, which displayed fractional changes in *cis* content of 43% and 17% upon terminal blockage or the addition of 1.1 M NaCl (Fig. 5). Therefore, electrostatic attraction between oppositely charged termini enhances the *cis*-Pro fraction in oligo-peptides.

7.3.3 Long-range interactions not involving inter-domain contacts influence cis-proline propensity in α -synuclein

As P128 showed the largest discrepancy between values obtained for full-length α S and its corresponding blocked tetra-peptide (Fig. 4), the impact of longer range interactions on *cis*-Pro was assessed by comparing the fraction of *cis*-Pro at P128 as a function of peptide length. Upon increasing the length of the peptide from four to six to eight residues (Fig. 4), a progressive decrease in the fraction of *cis*-Pro was observed. However, even though the blocked octa-peptide exhibited a *cis*-Pro fraction that approached that of full-length α S, the remaining significant difference of 1.5% must result from weak interactions that extend beyond the ± 4 residue range.

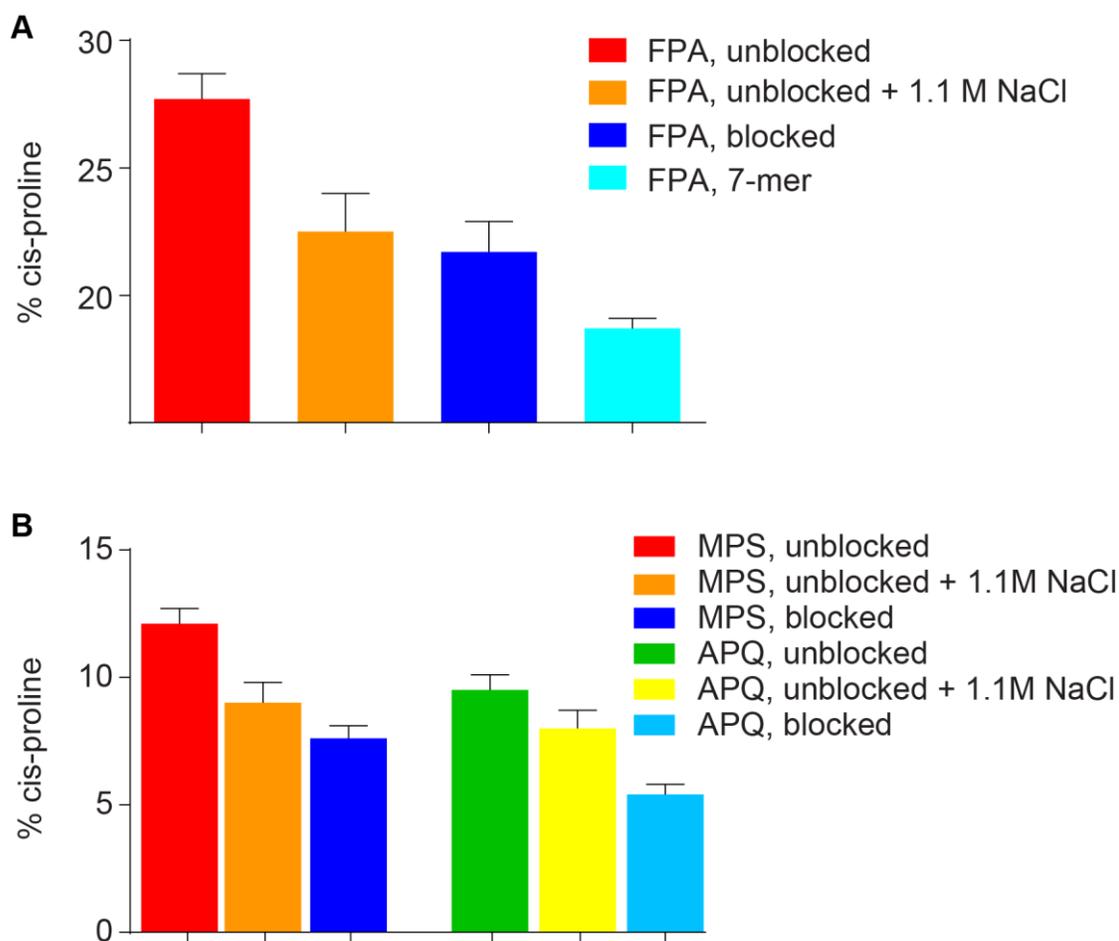


Figure 5. Impact of charged termini on *cis*-Pro formation. (A) Fraction of *cis*-Pro in a tri-peptide (FPA) without (*red*) or with (*blue*) acetylation and amidation of the N- and C-termini, respectively. Addition of 1.1 M NaCl to the unblocked FPA peptide (*orange*) significantly diminishes the fraction of *cis*-Pro, lowering it to levels essentially identical to the blocked peptide. Extending the peptide chain from a blocked tri-peptide to a blocked hepta-peptide (*cyan*) further decreases the fraction of *cis*-Pro. (B) Fraction of *cis*-Pro in α S tri-peptides MPS and APQ. The unblocked tri-peptides (*red*, *green*) show elevated *cis*-Pro fractions as compared to their blocked counterparts (*blue*, *cyan*). The addition of 1.1 M NaCl lowers the *cis*-Pro fraction of the unblocked MPS peptide to that of the blocked peptide. In both cases, the fractional change in *cis*-Pro upon blocking the termini and adding 1.1 M NaCl is *ca.* 40% and 20%, respectively. Reproduced with permission from John Wiley and Sons.

I therefore also evaluated the impact on *cis*-Pro formation in full-length α S from potential inter-domain contacts. Previous NMR measurements have indicated that the oppositely charged N- and C-terminal regions of α S transiently contact each other (Bertoncini et al. 2005), an effect held responsible for the compaction of the radius of hydration of this molecule relative to that of a

fully random coil (Uversky et al. 2001) and decreased amyloidogenicity relative to sequences that lack the C-terminal tail (Thomas et al. 2006, van der Wateren et al. 2018). Moreover, the N- and C-termini have been observed to make transiently intermolecular contacts (Wu & Baum 2010). Depending on the nature of such contacts, they could either restrict or favor *cis*-Pro formation. However, a construct comprising residues 87–140 of α S showed no detectable differences in the populations of *cis*-Pro between the full-length and truncated proteins (Fig. 6), indicating that the N-terminal region does not measurably impact *cis*-Pro propensities in the C-terminal region. Nevertheless, the small chemical shift differences between the full-length protein and this C-terminal 54-residue peptide (Fig. 6) are consistent with the previously noted interactions between the two domains, but are too weak to cause a detectable impact on the fraction of *cis*-Pro.

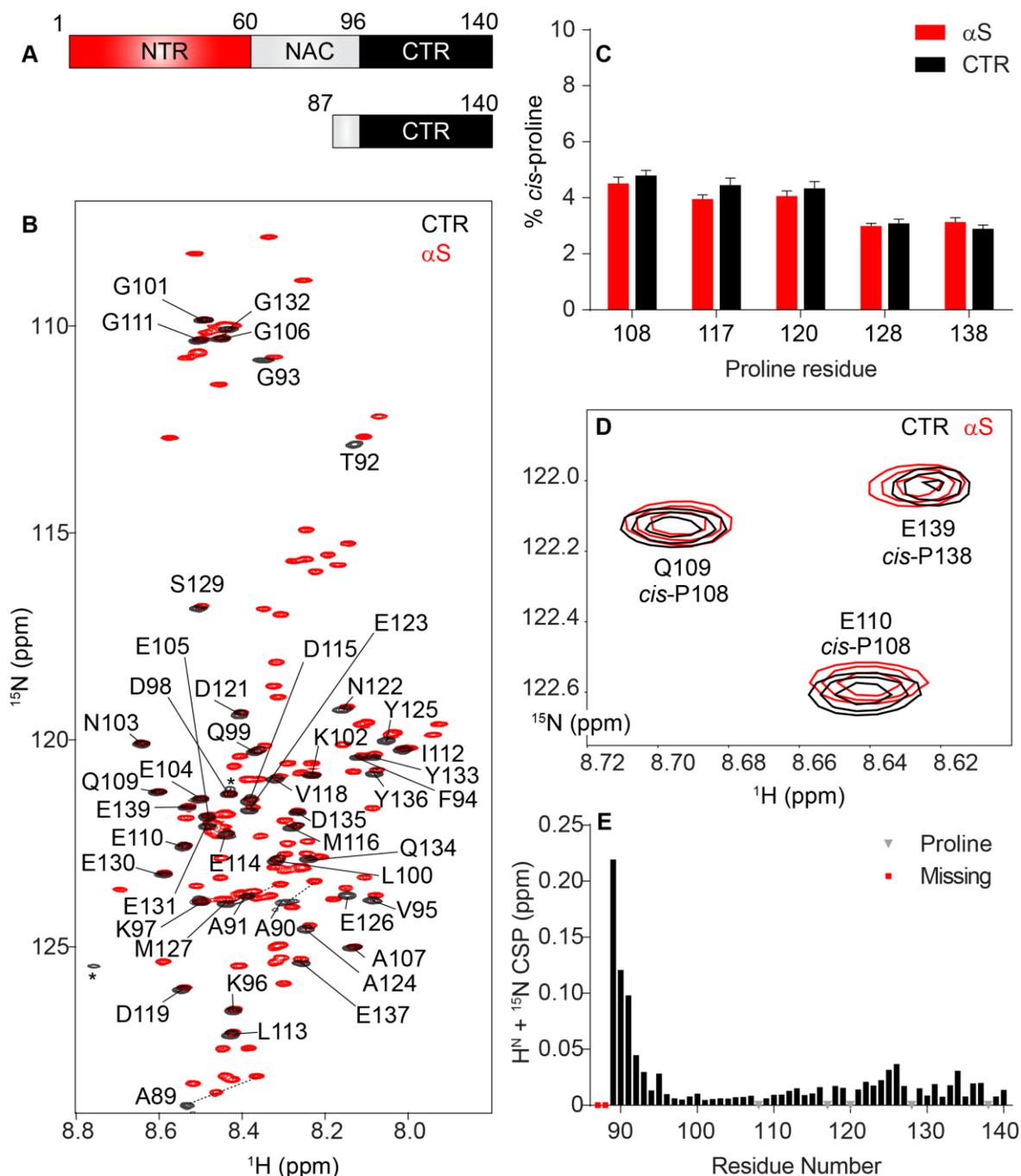


Figure 6. The isolated C-terminal region of α S retains its native *cis*-Pro fractions. **(A)** Depiction of α S and its three regions: the amphipathic N-terminal region (NTR), the hydrophobic NAC region, and the acidic C-terminal region (CTR). The isolated CTR (*bottom*) contains an S87C mutation. **(B)** 2D ^1H - ^{15}N HSQC spectra of full-length α S (*red*) and the isolated CTR (*black*). Buffer conditions were as in Figure 1, except that 5 mM BME was added to the isolated CTR to prevent disulfide formation. The peak indicated with an asterisk arises from an impurity. **(C)** Quantification of the *cis*-Pro fractions in both full-length α S and its isolated CTR. **(D)** Zoomed-in region from panel (B) showing resonances affected by *cis*-Pro. **(E)** Combined and weighted H^{N} and ^{15}N chemical shift perturbations (CSPs) between full-length α S and its isolated CTR. Reproduced with permission from John Wiley and Sons.

7.3.4 High-pressure NMR to measure *cis*-proline fractions in unfolded proteins

To compare the amount of *cis*-Pro in α S with other unfolded proteins, I used high-pressure NMR, which was previously shown to have no significant effect on the structure and dynamics of α S. To verify that hydrostatic pressure does not impact the *cis*-Pro content in an unfolded protein, I also measured the fraction of *cis*-Pro in α S at 2.5 kbar. Indeed, these values were found to be unchanged with respect to their 1 bar counterparts (Fig. 7).

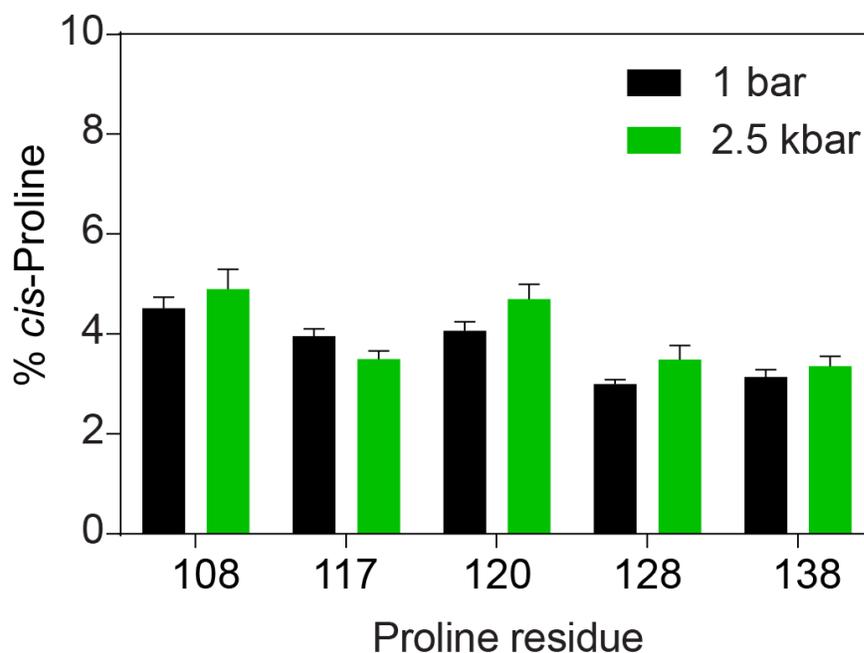


Figure 7. Impact of hydrostatic pressure on *cis*-Pro formation in α S. The fraction of *cis*-Pro was measured at 1 bar (*black*) and at 2.5 kbar (*green*). Agreement to within experimental uncertainty between the two datasets indicates that high pressures has no detectable impact on *cis*-Pro formation. Reproduced with permission from John Wiley and Sons.

Two proteins were denatured at 2.5 kbar and their NMR signals corresponding to residues impacted by *cis*-Pro were assigned. The α -crystallin domain of HSP27 (cHSP27), a 2x10 kDa non-covalent dimer, contains seven Pro residues per subunit that all adopt the *trans* conformation

in the native state. cHSP27 readily unfolds at high pressures (Fig. 8), and 3D NMR spectra were recorded at 2.5 kbar in order to assign the chemical shifts of the pressure-denatured state at pH 7 and 288 K. Previous data indicated that *cis*-Pro populations were independent of pH for all X-Pro bonds excluding His-Pro and Tyr-Pro, neither of which exist in cHSP27. At least 32 out of 86 ^1H - ^{15}N correlations in cHSP27 are visibly impacted by *cis*-Pro formation (Fig. 8). Interestingly, cHSP27 contains Pro clustering, with a P145-P146 pair close to P150, and P168 near P170. As a result, nearby residues are affected by *cis-trans* isomerization from multiple X-Pro bonds. For example, four signals are observed for G147: all *trans*, *cis*-P145, *cis*-P146, and *cis*-P150 (Fig. 8), with sequential assignments providing unambiguous identification of each X-Pro bond of interest. As compared to αS , the amount of *cis*-Pro in pressure-denatured cHSP27 varied widely, from 5% to 10% across the seven X-Pro bonds (Fig. 8).

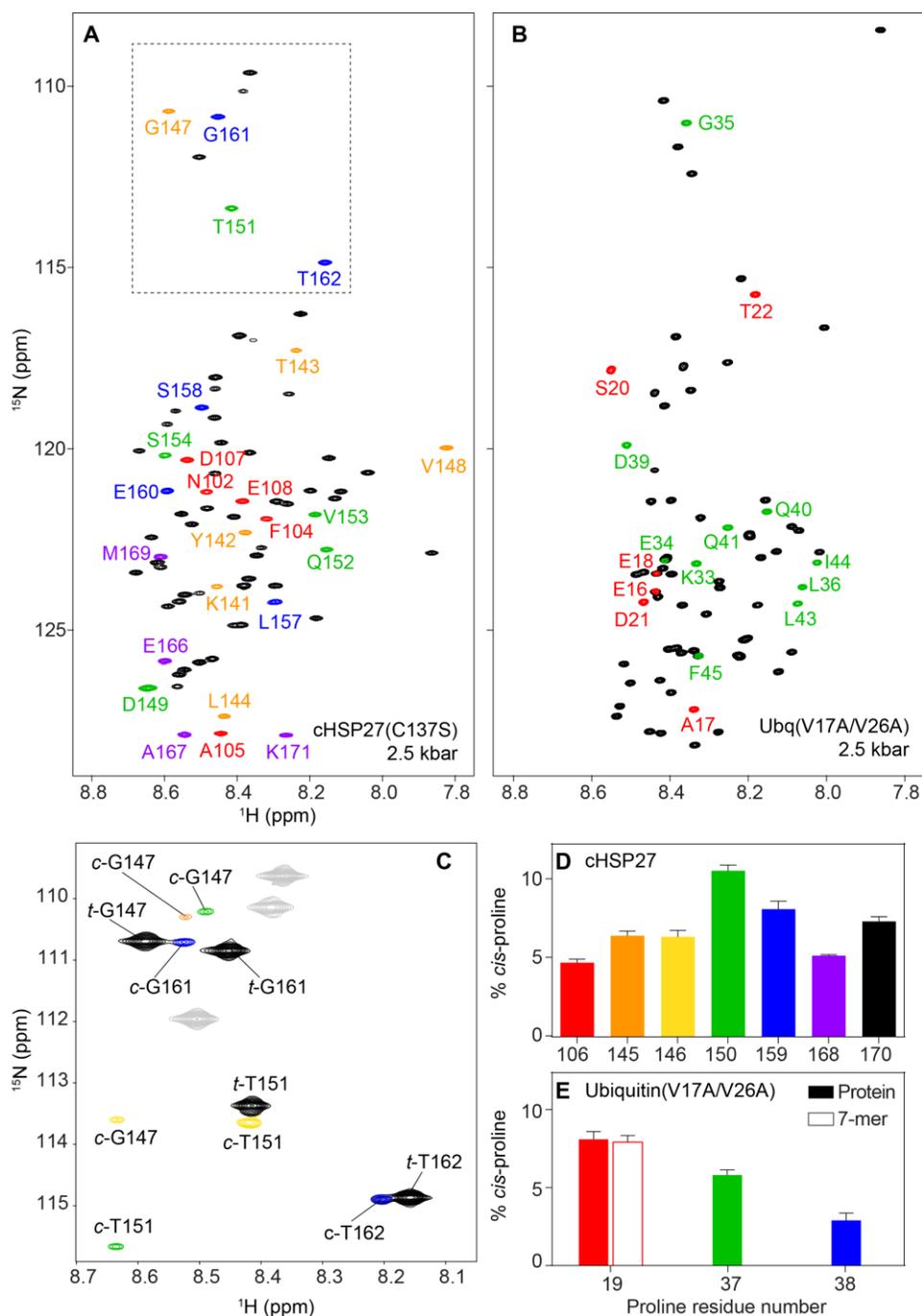


Figure 8. Quantification of *cis*-Pro bonds in two pressure-denatured proteins. **(A)** 1.6 mM ^{13}C , ^{15}N -labeled α -crystallin domain of HSP27 (cHSP27) at 2.5 kbar in 30 mM sodium phosphate, pH 7, 2 mM EDTA, 2 mM benzamidine, 288 K. **(B)** 0.75 mM ^2H , ^{13}C , ^{15}N -labeled ubiquitin-(V17A/V26A) at 2.5 kbar in 20 mM potassium phosphate buffer, pH 6.4, 2 mM benzamidine, 288 K. In both **(A)** and **(B)**, colored resonances have distinguishable assignments for the *trans* and *cis* conformations. **(C)** Boxed region of **(A)** at a lower contour level; *c* and *t* denote *cis* and *trans*, respectively. **(D,E)** Average *cis*-Pro fraction at each Pro residue for the samples in **(A)** and **(B)**. Reproduced with permission from John Wiley and Sons.

Ubiquitin contains three Pro residues, but the wild-type protein does not easily unfold at high pressure. Therefore, a cavity-containing double mutant (V17A/V26A) (Alderson et al. 2017b) was prepared in order to facilitate pressure-induced unfolding. These mutations lowered the midpoint of pressure denaturation from >5 kbar to ca. 1.4 kbar, enabling acquisition of a set of 3D NMR spectra to assign the minor signals corresponding to *cis*-Pro (Fig. 8). The *cis*-Pro fraction varied by a factor of three across the Pro residues, from as low as 2.5% at P38, which follows P37, to 8% at P19 (Fig. 8). A hepta-peptide of the same amino acid composition centered around P19 exhibited the same fraction of *cis*-Pro as full-length, pressure-unfolded ubiquitin (Fig. 8). The P37-P38 bond in ubiquitin-(V17A/V26A) displays the lowest *cis*-Pro value (2.5%) across all 15 X-Pro bonds. However, the chemically similar P145-P146 bond in cHSP27 contains ca. 6% *cis*-Pro, highlighting the importance of residues remote from the Pro-Pro bond.

7.3.5 Similar backbone dynamics in the *cis*- and *trans*-proline conformers

Finally, to characterize the potential impact of *cis*-Pro formation on the internal dynamics of the polypeptide chain, I recorded ^{15}N spin relaxation data that probe motions on the pico-nanosecond timescale (Jarymowycz & Stone 2006). These rapid motions reflect fluctuations of N–H bond vectors over time, providing insight into the conformational entropy and local dynamics of the involved residues. For this purpose, I studied the intrinsically disordered C-terminal region (CTR) of HSP27, which comprises the final ca. 30 residues and contains two Pro residues. While the average molecular mass of HSP27 oligomers is ca. 500 kDa (Fig. 9a), the unfolded CTR remains highly dynamic and tumbles independently of the large oligomers, thus giving rise to sharp NMR signals (Fig. 9b). Recently, a solution-state NMR publication attributed the observation of multiple sets of peaks in the CTR of αB -crystallin to an equilibrium between the bound and unbound forms (Mainz et al. 2015). However, the oligomer-bound CTR would tumble with a rotational correlation time that is similar to, if not equal to, that of the large αB -crystallin oligomers, and would thus be invisible in conventional ^1H - ^{15}N HSQC spectra.

The native mass spectrum of HSP27 at 25 μM monomer concentration reveals many overlapping signals in the region between *ca.* 5000-12000 m/z (Fig. 9a), diagnostic of the assembly of polydisperse oligomers. However, excluding the minor resonances caused by *cis*-Pro formation (see below), the 2D ^1H - ^{15}N HSQC spectrum of ^{15}N -labeled HSP27 revealed a single set of peaks, suggesting that the dynamic CTR remains unfolded in the heterogeneous ensemble of HSP27 oligomers. NMR signals from the disordered CTR were assigned by standard methods (Sattler et al. 1999), with multiple resonances observed for residues near P194 (Fig. 9b, Fig. 9c). The minor peaks were attributed to formation of a *cis*-Pro conformation, as evidenced by the $^{13}\text{C}\beta$ and $^{13}\text{C}\gamma$ chemical shifts from P194 (Fig. 9d). Formation of a *cis*-Pro conformation increases the difference between Pro $^{13}\text{C}\beta$ and $^{13}\text{C}\gamma$ chemical shifts to *ca.* 10 ppm, whereas this value is closer to 5 ppm in the *trans* conformation. I recorded a 3D C(CO)NH spectrum to measure the ^{13}C chemical shifts for Pro side-chains, and the anticipated *ca.* 10 ppm difference in $^{13}\text{C}\beta$ and $^{13}\text{C}\gamma$ chemical shifts was observed for the minor state (Fig. 9d).

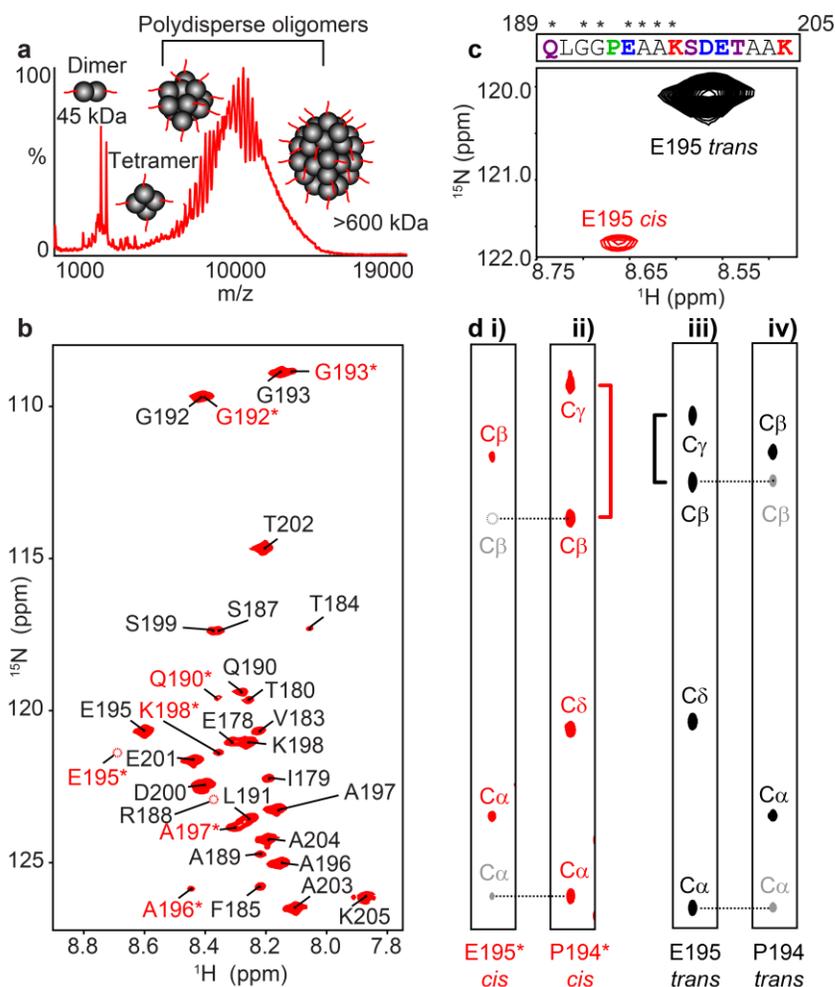


Figure 9. *Cis-trans* proline isomerization about the G193-P194 peptide bond in HSP27. (a) Native mass spectrum of 25 μM HSP27 in 200 mM ammonium acetate, pH 6.9. The y-axis depicts signal intensity and the x-axis displays the mass-to-charge ratio. (b) 2D ^1H - ^{15}N HSQC spectrum of 2 mM [U - ^{13}C , ^{15}N]-HSP27 at 298 K in 30 mM NaH_2PO_4 , 100 mM NaCl , 2 mM EDTA, 2 mM NaN_3 , pH 7. Each peak corresponds to an N-H bond and resonance assignments are listed next to each peak. Residues that are colored in red and labeled with an asterisk (*) indicate doubled peaks. The dashed circles indicate resonances that are weak and not observable at the plotted contour level. (c) Residues that exhibit two peaks are shown with asterisks (*) above their one-letter amino acid code. A zoomed-in region from (b), which shows the major (black) and minor (red) peaks that arise from E195. (d) 2D ^1H - ^{13}C strip plots taken from 3D HNCACB (i. and iv.) and C(CO)NH (ii. and iii.) spectra at the ^{15}N frequency corresponding to E195 or E195*. In the HNCACB spectrum, the $^{13}\text{C}\alpha$ and $^{13}\text{C}\beta$ nuclei from E195 (red/black) and P194 (grey) are indicated with dashed lines to reveal their corresponding positions in the C(CO)NH spectrum taken at the E195 ^{15}N frequency. The difference between $^{13}\text{C}\beta$ and $^{13}\text{C}\alpha$ chemical shifts in proline residues is diagnostic of *cis*- or *trans*-proline bonds. In the minor state the $^{13}\text{C}\beta$ and $^{13}\text{C}\alpha$ chemical shift difference is ~ 10 ppm (red), indicative of a *cis*-conformation, whereas in the major state this difference is only ~ 5 ppm (black), indicative of a *trans*-conformation. Reproduced freely under a Creative Commons CC BY license.

With resonance assignments for both the *cis*- and *trans*-Pro conformers in hand, I recorded ^{15}N spin relaxation experiments to probe backbone dynamics. Namely, I measured $\{^1\text{H}\}$ - ^{15}N Overhauser enhancement (hetNOE) value and longitudinal (R_1) and transverse (R_2) ^{15}N relaxation rates (Fig. 10). While these data collectively provide a description of the fast dynamics associated with each N–H bond vector, the hetNOE is most sensitive to picosecond dynamics (Kay et al. 1989). For folded proteins, hetNOE values typically range from ca. 0.65–0.85 at 600 MHz, while relatively flexible loops and regions with enhanced picosecond dynamics exhibit lower values. Consistent with this, the hetNOE measures I obtained for the *trans*-P194 state of the CTR at 600 MHz are near zero for residues 178–188, and decrease toward a minimum of -1.2 at the C-terminus (Fig. 10a). This enhanced flexibility with increased distance from the ACD is similar to observations made in αA - and αB -crystallin (Treweek et al. 2010). The *cis*-P194 hetNOE values are nearly identical to those for the *trans*-P194 state, except for residues E195 and K198, which are somewhat smaller. The overall resemblance of hetNOE values for both proline isomer states, however, demonstrates similar dynamic behavior, with picosecond-timescale motions throughout the CTR that increase toward the C-terminus for both conformations.

Both ^{15}N R_1 and ^{15}N R_2 are less sensitive to rapid internal dynamics than the hetNOE, yet provide information regarding motions on the nanosecond timescale that are faster than the rotational correlation time. The values obtained for the *cis*- and *trans*-P194 forms are very similar (Fig. 10b, Fig. 10c), with the average values of R_1 and R_2 of ca. 1.2 and 5 s^{-1} , respectively indicating that both *cis*- and *trans*-P194 conformational ensembles are also highly dynamic. In addition to these motions in the CTR of HSP27, it is possible, as in the case of αB -crystallin (Baldwin et al. 2011a), that there are contributions from conformational exchange to the observed R_2 values in the IPV motif. Taken together, my data have revealed that the CTR, including the *cis*-P194 conformer, is highly dynamic on the picosecond-nanosecond timescale, with the magnitude of these motions increasing upon moving away from the ACD towards the C-terminus.

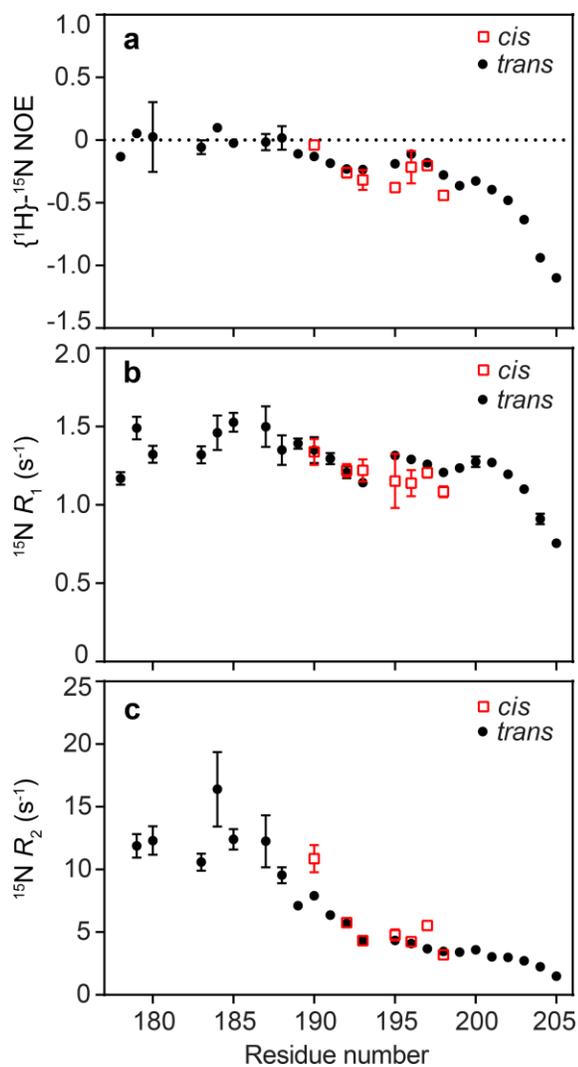


Figure 10. Comparison of backbone motions for the *cis*- and *trans*-Pro conformations in the C-terminal region of HSP27. (a-c) ^{15}N spin relaxation data obtained for both *cis*- (red) and *trans*-P194 (black) HSP27. In panels (a), the $\{^1\text{H}\}$ - ^{15}N nuclear Overhauser effect (NOE), (b) ^{15}N longitudinal relaxation rate (R_1), and (c) the ^{15}N transverse relaxation rate are shown. All data were recorded at 600 MHz and 298 K. Reproduced freely under a Creative Commons CC BY license.

7.4 Conclusions

My results indicate that populations of *cis*-Pro conformers in unfolded proteins are considerably lower than for corresponding short peptides that were previously used. The presence of *cis*-Pro is often considered a hurdle in protein folding, and with multiple Pro residues per chain, the fraction of all-*trans* X-Pro chains can be strongly impacted by even a modest change in the *cis*-Pro ratios. I attribute the lower *cis*-Pro fractions in the longer polypeptides to increased steric clashing between the segments of the chain preceding and following the *cis* peptide bond, which effectively reverses the chain direction and reduces conformational space available to the two segments. Electrostatic repulsion between these segments, as applies in α S, can enhance this effect. The utility of high-pressure NMR to accurately quantify *cis*-Pro populations in proteins is key to ongoing studies of protein folding by pressure-jump NMR, and the reported *cis/trans* chemical shift differences serve as convenient guides for identifying *cis*-Pro related resonances in NMR-based studies of IDPs and denatured proteins.

Antibody data have implicated elevated *cis*-Pro levels in toxic fibril formation of the intrinsically disordered Tau protein, associated with traumatic brain injury (Thomas et al. 2006), but NMR data refuted this conclusion (Ahuja et al. 2016). In context of α S and its formation of amyloid fibrils, mutation of its Pro residues to Ala accelerates fibril formation (Meuvis et al. 2010). Considering that the C-terminal region does not become embedded in the fibril core (Tuttle et al. 2016), this points to a possible inhibitory role of *cis*-Pro in fibril elongation, where a low *cis* fraction would suffice for such a role. NMR spectra of α S obtained in living bacterial and mammalian cells (Binolfi et al. 2012, 2016; Theillet et al. 2016) indicate that α S retains its predominantly *trans*-Pro conformation, implying that *in vivo* interactions do not drastically alter the inherent *cis*-Pro propensities of α S.

7.5 Methods

7.5.1 Proteins and peptides

Isotopically labelled proteins were expressed in *E. coli* and purified as described previously (Masuda et al. 2006, Chapter 3, Chapter 6). All NMR spectra were acquired on Bruker Avance III spectrometers (700 and 800 MHz) equipped with cryogenically cooled probes. The amino acid sequences of oligo-peptides and proteins used in this study are listed in Table 1 and Table 2, respectively.

Codon-optimized human α S (Masuda et al. 2006) but lacking the post-translational amino-terminal acetylation, was expressed and purified from *E. coli* BL21(DE3) using standard methods. Briefly, α S expression was induced with 1 mM IPTG when the OD of 1 L of $^{13}\text{C},^{15}\text{N}$ -M9 minimal medium at 37 °C reached 0.6-0.8 units. Protein expression proceeded for 3 hours, upon which cells were pelleted and stored at -80 °C until use. Cell pellets were resuspended in 15 mL of 20 mM Tris-HCl, 2 mM EDTA, 2 mM benzamidine, pH 7, and then placed in a water bath at 90 °C for 10 minutes. One protease inhibitor tablet (Roche) was included in the initial buffer, with only EDTA and benzamidine used in subsequent buffers. The resultant solution was centrifuged at 20,000 x g for 20 minutes, and the supernatant was loaded onto a HiTrap Q column equilibrated in the resuspension buffer. α S was eluted using a linear gradient of 1M NaCl from 7%-20%, with elution near 15% NaCl. All α S-containing fractions were pooled and then purified further by size exclusion chromatography and a Superdex S75 26/60 column (GE Healthcare) equilibrated in 20 mM NaH_2PO_4 , pH 6, 2 mM EDTA, and 2 mM benzamidine as a broad spectrum protease inhibitor.

The C-terminal fragment of human α S (87-140), including an S87C mutation (Jiang et al. 2017), was purified from *E. coli* BL21(DE3) as above, except that 5 mM β -mercaptoethanol (BME) was included in all buffers to keep the Cys residue reduced.

7.5.2 Preparation of peptide solutions

Oligo-peptides used are listed in Table 1 and were kindly synthesized by Galina Abdoulaeva at the Center for Biologics Evaluation and Research (CBER) of the Food and Drug Administration (FDA). The N- and C-termini of α S peptides were blocked with acetylation and amidation, and the tri-peptides FPA, MPS, and APQ were synthesized with and without blocked termini. All peptides were dissolved at a concentration of 40-50 mM in 20 mM NaH₂PO₄ and the pH of the solution was adjusted to 6 using concentrated NaOH. To match protein samples, the solution also contained 2 mM EDTA and 2 mM benzamidine.

Table 1. Peptides used to quantify *cis*-Pro fractions. Ac and NH₂ refer to acetylation and amidation, respectively. The bolded proline numbers correspond to those in α S and are included for clarity.

Peptide sequence ^a	% <i>cis</i> -Pro
FPA	27.7 ± 1.1%
FPA + 1.1 M NaCl	22.5 ± 1.5%
Ac-FPA-NH ₂	21.7 ± 1.2%
Ac-EGFPAEG-NH ₂	18.7 ± 0.4%
P108: Ac-GAPQ-NH ₂	5.6 ± 0.2%
GAPQ ^b	4.5 ± 0.2
Ac-GGPQ-NH ₂	7.7 ± 0.2%
APQ	9.5 ± 0.7%
APQ + 1.1 M NaCl	8.0 ± 0.8%
Ac-APQ-NH ₂	5.4 ± 0.4%
P117: Ac-DMPV-NH ₂	7.6 ± 0.2%
DMPV ^b	3.9 ± 0.2%
P120: Ac-VDPD-NH ₂	4.0 ± 0.1%
VDPD ^b	4.0 ± 0.2%
P128: Ac-EMPS-NH ₂	7.1 ± 0.1%
EMPS ^b	3.0 ± 0.1%
MPS	12.1 ± 0.6%
MPS + 1.1 M NaCl	9.0 ± 0.8%
Ac-MPS-NH ₂	7.6 ± 0.5%
Ac-YEMPSE-NH ₂	4.5 ± 0.2%
Ac-AYEMPSEE-NH ₂	4.1 ± 0.3%
P138: Ac-YEPE-NH ₂	4.3 ± 0.3%
YEPE ^b	3.1 ± 0.2%
Ac-EAEPST-NH ₂ ^c	7.9 ± 0.4%

- a. Unless otherwise indicated, all peptides were prepared in 20 mM sodium phosphate, 2 mM EDTA, 2 mM benzamidine, pH 6, and some contained 1.1M NaCl as noted. b Data obtained on full-length α S in the same buffer as above.
- b. 25 mM potassium phosphate, pH 6.4. This peptide corresponds to P18 in human ubiquitin bearing the V17A mutation to match the protein used in this study (V17A/V26A ubiquitin).

Table 2. Amino acid sequences for the four proteins used in this study.

Human α S (Uniprot: P37840)

MDVFMKGLSKAKEGVVAAAEEKTKQGVAEAAAGKTKEGVLYVGSKTKEGVVHGVAATVAEKTKEQVTNVGGAVVTGVTAVA
QKTVEGAGSIAAATGFVKKDQLGKNEEGA**P**QEGILEDM**P**V**D****P**DNEAYEM**P**SEEGYQDYE**P**EA

Human α S S87C-140

CIAAATGFVKKDQLGKNEEGA**P**QEGILEDM**P**V**D****P**DNEAYEM**P**SEEGYQDYE**P**EA

cHSP27(C137S), the α -crystallin domain of human HSP27 (Uniprot: P04792)

GVSEIRHTADRWVSLDVNHFA**P**DELTVKTKDGVVEITGKHEERQDEHGYISRSFTRKYTL**PP**GV**D****P**TQVSSSL**P**EG
TLTVEA**P****MP****K**

Human ubiquitin(V17A/V26A) (Uniprot: P0CG48)

MQIFVKTLTGKTITLEAE**P**SDTIENAKAKIQDKEGI**PP**DQQRLLIFAGKQLEDGRTLSDYNIQKESTLHLVLRLLGG

7.5.3 NMR spectroscopy

All samples for NMR spectroscopy included 7% D₂O and NMR spectra were recorded at 288 K. ¹³C,¹⁵N- α S was concentrated to 0.9 mM in 20 mM NaH₂PO₄, pH 6, 2 mM EDTA, 2 mM benzamidine, with 290 μ L loaded into a Shigemi micro NMR cell. A separate sample was prepared with 1.1 M NaCl, using a thick-walled, 2.2 mm ID NMR tube (New Era Enterprises, Inc.). Spectra were recorded on a Bruker Avance III 700 MHz spectrometer equipped with a cryogenically cooled probe.

¹³C,¹⁵N-ACD(C137S) was concentrated to 1.6 mM in 30 mM NaH₂PO₄, pH 7, 2 mM EDTA, 2 mM benzamidine. 280 μ L of the sample was loaded into a high-pressure tube (Daedalus Innovations) and pressurized to 2.5 kbar. ²H,¹³C,¹⁵N-ubiquitin(V17A/V26A) was concentrated to 0.7 mM in 25 mM potassium phosphate buffer at pH 6.4, loaded into a high-pressure tube, and pressurized to 2.5 kbar. Spectra were recorded on a Bruker Avance III 800 MHz spectrometer equipped with a cryogenically cooled probe.

3D HNCO and HNCA spectra were recorded to facilitate identification and assignment of the lowly populated *cis*-Pro conformations. When non-uniform sampling was used, the spectra

were reconstructed using the SMILE program (Ying et al. 2016) and all NMR spectra were processed and analyzed with NMRPipe (Delaglio et al. 1995) and Sparky (Lee et al. 2015). Finally, Chemical shift perturbations were calculated using the following equation: $\Delta\delta_{\text{eff}} = \sqrt{(\Delta\delta H)^2 + (\Delta\delta N/5)^2}$

7.6 Acknowledgements

The work outlined in this chapter has been published in *ChemBioChem* in 2018 under the title, “Propensity for *cis*-proline formation in unfolded proteins,” by the authors T. Reid Alderson, Jung Ho Lee, Cyril Charlier, Jinfa Ying, and Ad Bax with the DOI 10.1002/cbic.201700548. In addition, this chapter has been published in *Cell Stress & Chaperones* in 2017 under the title, “Proline isomerization in the C-terminal region of HSP27,” by the authors T. Reid Alderson, Justin L. P. Benesch, and Andrew J. Baldwin with the DOI 10.1007/s12192-017-0791-z.

The author expressed and purified indicated proteins in the study and collected and analyzed the data. J.H.L had initiated the project; C.C. provided the VA2-ubq sample; J.Y. provided assistance with data collection.

I thank Drs. J.L. Baber, Y. Shen, and D. Garrett for technical support; Drs. J. Roche, Robert Best, and D.A. Torchia for insightful discussions; Drs. Zhiping Jiang and Jennifer C. Lee (NHLBI) for the plasmid coding for the C-terminal region of α S; Prof. Justin Benesch, Prof. Andrew Baldwin, and Dr. Heidi Gastall (University of Oxford) for the plasmid encoding cHSP27; and Galina Abdoulaeva (CBER/FDA) for peptide synthesis.

7.7 References

- Ahuja P, Cantrelle F-X, Huvent I, Hanouille X, Lopez J, et al. 2016. Proline Conformation in a Functional Tau Fragment. *J. Mol. Biol.* 428(1):79–91
- Alderson TR, Benesch JLP, Baldwin AJ. 2017a. Proline isomerization in the C-terminal region of HSP27. *Cell Stress Chaperones.* 22(4):
- Alderson TR, Charlier C, Torchia DA, Anfinrud P, Bax A. 2017b. Monitoring Hydrogen Exchange During Protein Folding by Fast Pressure Jump NMR Spectroscopy. *J. Am. Chem. Soc.* 139(32):11036–39
- Bertoncini CW, Jung Y-S, Fernandez CO, Hoyer W, Griesinger C, et al. 2005. From The Cover: Release of long-range tertiary interactions potentiates aggregation of natively unstructured τ -synuclein. *Proc. Natl. Acad. Sci.* 102(5):1430–35
- Binolfi A, Limatola A, Verzini S, Kosten J, Theillet F-X, et al. 2016. Intracellular repair of oxidation-damaged α -synuclein fails to target C-terminal modification sites. *Nat. Commun.* 7:10251
- Binolfi A, Theillet F-X, Selenko P. 2012. Bacterial in-cell NMR of human α -synuclein: a disordered monomer by nature?: Figure 1. *Biochem. Soc. Trans.* 40(5):950–54
- Capaldi AP, Ferguson SJ, Radford SE. 1999. The Greek key protein apo-pseudoazurin folds through an obligate on-pathway intermediate. *J. Mol. Biol.* 286(5):1621–32
- Deber CM, Torchia DA, Wong SC, Blout ER. 1972. Conformational interconversions of the cyclic hexapeptide cyclo(Pro-Gly) 3. *Proc. Natl. Acad. Sci. U. S. A.* 69(7):1825–29
- Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A. 1995. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR.* 6(3):277–93
- Dugave C, Demange L. 2003. Cis-trans isomerization of organic molecules and biomolecules: Implications and applications. *Chem. Rev.* 103(7):2475–2532
- Dyson HJ, Rance M, Houghten RA, Lerner RA, Wright PE. 1988. Folding of immunogenic peptide fragments of proteins in water solution. I. Sequence requirements for the formation of a reverse turn. *J. Mol. Biol.* 201(1):161–200
- Follis AV, Llambi F, Merritt P, Chipuk JE, Green DR, Kriwacki RW. 2015. Pin1-Induced Proline Isomerization in Cytosolic p53 Mediates BAX Activation and Apoptosis. *Mol. Cell.* 59(4):677–84
- Gibbs EB, Lu F, Portz B, Fisher MJ, Medellin BP, et al. 2017. Phosphorylation induces sequence-specific conformational switches in the RNA polymerase II C-terminal domain. *Nat. Commun.* 8:15233
- Grathwohl C, Wüthrich K. 1976. Nmr studies of the molecular conformations in the linear oligopeptides H-(L-Ala)_n-L-Pro-OH. *Biopolymers.* 15(10):2043–57
- Grathwohl C, Wüthrich K. 1981. Nmr studies of the rates of proline cis–trans isomerization in oligopeptides. *Biopolymers.* 20(12):2623–33
- Hesterkamp T, Hauser S, Lütcke H, Bukau B. 1996. Escherichia coli trigger factor is a prolyl isomerase that associates with nascent polypeptide chains. *Proc. Natl. Acad. Sci. U. S. A.* 93(9):4437–41
- Hochberg GK a., Ecroyd H, Liu C, Cox D, Cascio D, et al. 2014. The structured core domain of β -crystallin can prevent amyloid fibrillation and associated toxicity. *Proc. Natl. Acad. Sci.* 111(16):E1562–70
- Jarymowycz VA, Stone MJ. 2006. Fast time scale dynamics of protein backbones: NMR relaxation methods, applications, and functional consequences. *Chem. Rev.* 106(5):1624–71

- Jiang Z, Heinrich F, McGlinchey RP, Gruschus JM, Lee JC. 2017. Segmental Deuteration of α -Synuclein for Neutron Reflectometry on Tethered Bilayers. *J. Phys. Chem. Lett.* 8(1):29–34
- Lang K, Schmid FX, Fischer G. 1987. Catalysis of protein folding by prolyl isomerase. *Nature.* 329(6136):268–70
- Lee W, Tonelli M, Markley JL. 2015. NMRFAM-SPARKY: enhanced software for biomolecular NMR spectroscopy. *Bioinformatics.* 31(8):1325–27
- Lu KP, Finn G, Lee TH, Nicholson LK. 2007. Prolyl cis-trans isomerization as a molecular timer. *Nat. Chem. Biol.* 3(10):619–29
- Mainz A, Peschek J, Stavropoulou M, Back KC, Bardiaux B, et al. 2015. The chaperone α B-crystallin uses different interfaces to capture an amorphous and an amyloid client. *Nat. Struct. Mol. Biol.* 22(11):898–905
- Masuda M, Dohmae N, Nonaka T, Oikawa T, Hisanaga S, et al. 2006. Cysteine misincorporation in bacterially expressed human α -synuclein. *FEBS Lett.* 580(7):1775–79
- Meuvis J, Gerard M, Desender L, Baekelandt V, Engelborghs Y. 2010. The conformation and the aggregation kinetics of α -synuclein depend on the proline residues in its C-terminal region. *Biochemistry.* 49(43):9345–52
- Morgan AA, Rubenstein E, Hosack D, Yang J, Gao W. 2013. Proline: The Distribution, Frequency, Positioning, and Common Functional Roles of Proline and Polyproline Sequences in the Human Proteome. *PLoS One.* 8(1):e53785
- Osváth S, Gruebele M. 2003. Proline can have opposite effects on fast and slow protein folding phases. *Biophys. J.* 85(2):1215–22
- Raleigh DP, Evans PA, Pitkeathly M, Dobson CM. 1992. A peptide model for proline isomerism in the unfolded state of staphylococcal nuclease. *J. Mol. Biol.* 228(2):338–42
- Reimer U, Scherer G, Drewello M, Kruber S, Schutkowski M, Fischer G. 1998. Side-chain effects on peptidyl-prolyl cis/trans isomerisation. *J. Mol. Biol.* 279(2):449–60
- Sarkar P, Reichman C, Saleh T, Birge RB, Kalodimos CG. 2007. Proline cis-trans isomerization controls autoinhibition of a signaling protein. *Mol. Cell.* 25(3):413–26
- Sarkar SK, Young PE, Sullivan CE, Torchia DA. 1984. Detection of cis and trans X-Pro peptide bonds in proteins by ^{13}C NMR: application to collagen. *Proc. Natl. Acad. Sci. U. S. A.* 81(15):4800–4803
- Sattler M, Schleucher J, Griesinger C. 1999. Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. *Prog. Nucl. Magn. Reson. Spectrosc.* 34(2):93–158
- Schmid FX. 1993. Prolyl Isomerase: Enzymatic Catalysis of Slow Protein-Folding Reactions. *Annu. Rev. Biophys. Biomol. Struct.* 22(1):123–43
- Theillet F-X, Binolfi A, Bekei B, Martorana A, Rose HM, et al. 2016. Structural disorder of monomeric α -synuclein persists in mammalian cells. *Nature.* 530(7588):45–50
- Theillet F-X, Kalmar L, Tompa P, Han K-H, Selenko P, et al. 2014. The alphabet of intrinsic disorder. *Intrinsically Disord. Proteins.* 1(1):e24360
- Thomas KM, Naduthambi D, Zondlo NJ. 2006. Electronic control of amide cis-trans isomerism via the aromatic-prolyl interaction. *J. Am. Chem. Soc.* 128(7):2216–17

- Torchia DA. 1972. Evidence for cis peptide bonds in copolypeptides of glycine and proline. *Biochemistry*. 11(8):1462–68
- Tuttle MD, Comellas G, Nieuwkoop AJ, Covell DJ, Berthold DA, et al. 2016. Solid-state NMR structure of a pathogenic fibril of full-length human α -synuclein. *Nat. Struct. Mol. Biol.* 23(5):409–15
- Uversky VN, Li J, Fink AL. 2001. Evidence for a partially folded intermediate in alpha-synuclein fibril formation. *J. Biol. Chem.* 276(14):10737–44
- van der Wateren IM, Knowles TPJ, Buell AK, Dobson CM, Galvagnion C. 2018. C-terminal truncation of α -synuclein promotes amyloid fibril amplification at physiological pH. *Chem. Sci.* 9(25):5506–16
- Walport LJ, Morra R, Mancini EJ, Redfield C. 2015. ¹H, ¹³C, and ¹⁵N resonance assignments for the tandem PHD finger motifs of human CHD4. *Biomol. NMR Assign.* 9(2):239–42
- Wedemeyer WJ, Welker E, Scheraga HA. 2002. Proline Cis–Trans Isomerization and Protein Folding. *Biochemistry*. 41(50):14637–44
- Weiss MS, Jabs A, Hilgenfeld R. 1998. Peptide bonds revisited. *Nat. Struct. Mol. Biol.* 5(8):676–676
- Werner-Allen JW, Lee C-J, Liu P, Nicely NI, Wang S, et al. 2011. cis -Proline-mediated Ser(P)⁵ Dephosphorylation by the RNA Polymerase II C-terminal Domain Phosphatase Ssu72. *J. Biol. Chem.* 286(7):5717–26
- Wu K-P, Baum J. 2010. Detection of Transient Interchain Interactions in the Intrinsically Disordered Protein α -Synuclein by NMR Paramagnetic Relaxation Enhancement. *J. Am. Chem. Soc.* 132(16):5546–47
- Yao J, Feher VA, Espejo BF, Reymond MT, Wright PE, Dyson HJ. 1994. Stabilization of a type VI turn in a family of linear peptides in water solution. *J. Mol. Biol.* 243(4):736–53
- Ying J, Delaglio F, Torchia DA, Bax A. 2016. Sparse multidimensional iterative lineshape-enhanced (SMILE) reconstruction of both non-uniformly sampled and conventional NMR data. *J. Biomol. NMR*

8

Conclusions

8.1 Concluding remarks

In this thesis, I have outlined new nuclear magnetic resonance (NMR) methodology that has provided novel insight into protein folding. In addition, I have described applications of NMR that have offered new biological insight into the mechanism and regulation of the molecular chaperone HSP27.

In Chapter 1, I provided a brief overview of protein folding and the importance of structural modeling in biochemical research. Protein folding plays a critical role in human health and disease, as the misfolding of proteins contributes to numerous so-called protein misfolding diseases, including Alzheimer's and Parkinson's diseases. I present a short description of molecular chaperones, namely the small heat shock proteins that were the focus of a large portion of this thesis.

In Chapter 2, I introduced NMR and its utility in biochemical research. I provide a brief theoretical overview before focusing on aspects of NMR that were particularly relevant to this thesis: nuclear spin relaxation and chemical exchange. Results from Carr-Purcell-Meiboom-Gill relaxation dispersion experiment were analyzed in detail by simulating and fitting data.

In Chapter 3, I presented an NMR-driven study of the small heat shock protein HSP27. I found that the stable, structured HSP27 dimer partially unfolds upon monomerization. This order-to-partial-disorder transition had not been previously known. I found that the partially disordered HSP27 monomer is a more effective molecular chaperone than its structured dimeric counterpart, suggesting that structural plasticity in HSP27 is important for its chaperone function. However, the partially disordered monomer was found to be unstable and prone to aggregation, thus highlighting the delicate balance between protein folding and function/dysfunction.

In Chapter 4, I used NMR spectroscopy to investigate the impact of a mutation in HSP27 that is implicated in the onset of a motor neuropathy, Charcot-Marie-Tooth disease. A highly

conserved region of HSP27, the IxI/V motif, binds to the structured core domain. A single amino acid variation in the IxI/V motif (P182L) causes Charcot-Marie-Tooth disease. I used Carr-Purcell-Meiboom-Gill relaxation dispersion to compare the binding kinetics of peptides from the WT and disease mutant forms, and found that the disease mutation causes a drastic attenuation in the association rate. The mutant form of the protein is chaperone-inactive and forms very large, soluble oligomers that are *ca.* 30-fold larger than the WT form. These results highlight the importance of the IxI/V motif in HSP27's chaperone function and molecular organization.

In Chapter 5, I outlined a new NMR method that provides insight into protein folding via hydrogen exchange coupled to fast pressure-jump NMR spectroscopy. The method is not presently generally applicable, as it requires a protein that can be quickly and reversibly unfolded at pressures below 3 kbar. Nevertheless, I show in this proof-of-principle study that amide-specific time constants for hydrogen-bond formation can be measured as the protein folds. A variant of the model system ubiquitin was used for this study.

In Chapter 6, I utilized translational diffusion NMR measurements to monitor hydration radii of proteins as the fraction of unfolded protein is progressively lowered by removing denaturant. I used both high pressures and urea to facilitate denaturation. In both cases, I observed that the polypeptide chain only marginally collapses under the lowest denaturant conditions. In addition, I outlined the necessary corrections for usage of dioxane in high-pressure diffusion NMR studies.

In Chapter 7, I reported on measurements of *cis*-proline fractions in unfolded and disordered proteins. I first demonstrated that the intrinsically disordered protein α -synuclein contains values of *cis*-proline in the 3.5-5% range, and that elevated pressures have no impact on these fractions. I then used high-pressure to denature two other proteins, and found that the 10 Xaa-Pro bonds in these proteins all contains values of *cis*-proline ranging from 2-10%. I determined that small oligo-peptides tend to overestimate the fraction of *cis*-proline, and so these results on intact proteins provide more accurate values.