

Cellular Inclusion Bodies: Structure and Mechanisms

by

Dalia Naser

A thesis

presented to the University of Waterloo

in fulfilment of the

thesis requirement for the degree of

Doctor of Philosophy

in

Chemistry

Waterloo, Ontario, Canada, 2022

©Dalia Naser 2022

Examining Committee Membership

The following served on the Examining Committee for this thesis. The decision of the Examining Committee is by majority vote.

External Examiner: Jan Rainey
 Professor, Department of Chemistry, University of Dalhousie

Supervisor: Elizabeth Meiering
 Professor, Department of Chemistry, University of Waterloo

Internal Member(s): Thorsten Dieckmann
 Associate Professor, Department of Chemistry, University of Waterloo
 Andrew Doxey
 Associate Professor, Department of Biology, University of Waterloo

Internal-External Marc Aucoin
Member: Professor, Department of Chemical Engineering, University of Waterloo

Author's Declaration

This thesis consists of material all of which I authored or co-authored: see Statement of Contributions included in the thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Statement of Contributions

This thesis consists in part of three manuscripts, all of which have been accepted for publication at the time of the submission of this thesis. Exceptions to sole authorship of material are included with each chapter. The manuscripts in this thesis are as follows:

A. Schaefer⁺, D. Naser⁺, B. Siebeneichler⁺, M. Tarasca, E. Meiering, *J. Biol. Chem.* **2022**, *In Press*

M. V Tarasca⁺, D. Naser⁺, A. Schaefer, T. G. Soule, E. M. Meiering, *Anal. Biochem.* **2022**

D. Naser⁺, M. V Tarasca⁺, B. Siebeneichler, A. Schaefer, H. K. Deol, T. G. Soule, J. Almey, S. Kelso, G. G. Mishra, H. Simon, E. M. Meiering, *Angew. Chemie - Int. Ed.* **2022**.

⁺ indicates both authors contributed equally to this work

Abstract

Inclusion bodies (IBs) are insoluble aggregate structures that commonly form upon overexpression of proteins in heterologous hosts with relevance to biomedicine and biotechnology. These structures are of broad interest in many industries: a near-pure source of protein, they are commonly used in protein purification. Additionally, their high mechanical and thermal stability, low toxicity, and resistance to proteases have made IBs an attractive candidate for being adapted as a vaccine-type molecule. Despite their applicability, the molecular mechanisms governing IB formation and structures remain obscure. The inhomogeneity, insolubility, and large size of protein aggregates makes traditional methods of structural determination such as nuclear magnetic resonance (NMR) or X-ray crystallography impossible or limited to low resolution, obstructing research into aggregation mechanisms and structures.

To address this issue, we undertook a multi-mutant study of the aggregation of Cu, Zn-Superoxide dismutase (SOD1) with the goal of determining if IB formation depends on protein primary sequence and stability, establish how IB structure can be modulated by varying cell growth conditions, and ultimately assemble a model of IB structure and formation.

Initial studies described in Chapter 2 involved measuring the aggregation propensity of eighteen SOD1 mutants under six different growth conditions (3 different temperatures, with and without metal cofactors to simulate maturation of the protein) using SDS-PAGE experiments combined with gel densitometry. Compelling correlations arise from these data: aggregation propensity is strongly correlated with the melting temperature of immature SOD1 ($R = -0.77$), as well as with the Gibb's free energy of dimer dissociation ($R = -0.68$) and monomer unfolding ($R = -0.58$), supporting previous findings that ALS mutations form aggregation-prone monomers, while also suggesting the contribution of dimer dissociation to aggregation. Correlations between these data and other biophysical parameters such as folding rate and disease duration were not observed, nor were correlations between aggregation propensity and normalized aggregation scores from several aggregation predictors. These results highlight how multiple variables can be involved in aggregation propensity.

To analyze aggregate structure, we used quenched hydrogen-deuterium exchange NMR (qHDX) measurements, a methodology that has been used extensively on purified proteins but whose application to protein aggregation in a cellular context has been little explored. The optimization of this method is discussed in detail in Chapter 3. It is then applied to the aggregates of 9 SOD1 mutant proteins under a single growth condition with predominant population of the most immature form of

the protein. The results point to overarching similarities in aggregate structure for all mutants. Extensive protection against exchange is seen throughout SOD1 regardless of a region's native hydrogen bonding or hydrophobicity. Closer examination of specific residues shows stretches of protection in regions previously shown to form protein-protein interfaces in both native and non-native SOD1 assemblies. Taken together, these data point to an ensemble of aggregation pathways in equilibrium and a heterogeneous aggregate.

Lastly, we aimed to expand the method to other protein systems, both to see if the procedure developed could be broadly applied, as well as to learn more about IB structure (Chapter 5). The first application was to Adnectins, an engineered family of proteins based on the 10th fibronectin type III domain. Adnectins have a beta-sheet sandwich structure with variable loops. In applying the qHDX method, we found that all conditions established for SOD1 readily applied to Adnectins, including 3D assignment strategies. We further found that, like SOD1, Adnectins show protection of many amides throughout the entire protein, beyond the limited regions that are protected by native structure or predicted to aggregate/form amyloid in the sequence. Combined with previous work showing that Adnectins may form native-like assemblies, these preliminary results similarly point to an ensemble model of IB formation. In contrast to the Adnectins and SOD1, apomyoglobin is a helical protein that has been seen to readily form IBs in *Escherichia coli*. In this case, rather than studying the effects of point mutations, we are evaluating different C-terminal truncation mutations to examine the impact of this change on the resulting aggregate structure. Preliminary results show that *in vivo* IB structures mimic previously studied *in vitro* aggregates, and that the protein becomes more native with increasing chain length. Overall, this thesis shows the utility of the qHDX method and provides a framework for understanding protein aggregation mechanisms and the resulting aggregate structures.

Acknowledgements

There are many times in my tenure in the Meiering lab where we jokingly uttered the phrase “it takes a village” at each other. Getting me through a PhD took that village.

Colleen Doyle and Dr. Duncan Mackenzie, you guided me through the start of graduate school and taught me to stand on my own two feet. You taught me how to be a confident learner and leader, and to cultivate wacky ideas. I hope I’ve made you proud and continued your legacy of science and mentorship.

Harmeen Deol, I owe you my sanity. Thank you for pushing me to never stop thinking and to never forget my love of asking and answering questions. Thank you for never letting me give up.

Dr. Jennifer Aguiar, whose friendship I’ve been lucky to have since my first undergraduate lab in Waterloo. Thank you for being my biggest cheerleader. You are always front row and center, smiling supportively in my life.

Ryan Moreira, you have always challenged me to do more and be better. Thanks for never doubting that I could do this.

To all the hands that have touched this project – Hilary Simon, Susan Kelso, Michael Tarasca, Tyler Soule, Bruna Siebeneichler and Anna Schaefer- thank you for giving it your all. Thank you for being my small team. It was an honor.

Thank you to my supervisor Dr Elizabeth Meiering and my committee - past and present - Dr Thorsten Dieckmann, Dr Andrew Doxey, Dr Marc Aucoin, and Dr Michael Palmer for their support throughout the course of this project.

To everyone else in the Meiering lab whose shoulder I’ve cried on -which is everyone- thank you.

We did it.

Dedication

To my parents, who always wanted me to be a doctor. I hope this teaches you to be more specific.

And, of course, to Betsy and Francis.

Table of Contents

List of Figures	xii
List of Tables	xiv
List of Abbreviations	xv
Chapter 1: Introduction	1
1.0 Preface	1
1.1 Protein Folding, Misfolding, and Aggregation	1
1.1.1 Theories on Protein Folding	1
1.1.2 Protein Misfolding	3
1.2 Methods to determine protein aggregate structure	4
1.2.1 Overview	4
1.2.2. Conformation specific antibodies: moving beyond detection to structural information	7
1.2.3 FTIR: Secondary structure at a glance	9
1.2.4 Solid-state NMR: a whole-structure view	14
1.2.5 qHDX NMR - Residue specific view	18
1.2.6 qHDX MS: a segment-specific view	20
1.3 Conclusions	23
Chapter 2: Quantifying levels of IB formation in a SOD1 model system	26
2.0 Preface	26
2.1 Introduction	26
2.1.1 Studying Inclusion Bodies (IBs)	26
2.1.2 Cu, Zn Superoxide Dismutase	27
2.2 Materials and Methods	30
2.2.1 Aggregation propensity prediction	30
2.2.2 Mutant SOD1 Expression System for IB Formation Studies	31
2.2.3 Expression, Sample Preparation, and Solubility Analysis	31
2.3 Results and discussion	33
2.3.1 Rationale for selected mutations and predicted behaviour	33
2.3.2 fALS SOD1 mutants aggregate more with increasing temperature	39
2.3.3 Addition of zinc rescues a portion of SOD1 from aggregation	48
2.3.4 Caveats on the use of gel densitometry	55
2.4 Conclusions	58
Chapter 3: Optimization of quenched hydrogen-deuterium amide exchange for impure IBs	60

3.0 Preface	60
3.1 Chapter Abstract	61
3.2 Introduction	62
3.2.1 Method Overview	64
3.3 Materials and Methods	66
3.3.1. Materials	66
3.3.2. Protein Growth and Expression	66
3.3.3. Preparation of IB Samples	67
3.3.4 Fourier transform infrared spectroscopy (FTIR)	67
3.3.5 Preparation of NMR Samples	68
3.3.6. NMR Experiments	68
3.3.7. Data Analysis	71
3.4 Results and Discussion	72
3.4.1 FTIR analysis of experimental conditions	72
3.4.2 Control of IB Sample Conditions	75
3.4.3 Spectral Analysis	83
3.5 Conclusions	89
Chapter 4: Quenched hydrogen-deuterium amide exchange reveals native-like structure in SOD1 IBs	90
4.0 Preface	90
4.1 Chapter Abstract	91
4.2 Introduction	91
4.3 Materials and Methods	92
4.3.1 SOD1 50 mL Inclusion Body Expression	92
4.3.2 Iodoacetamide SDS- PAGE	92
4.3.3 Preparation of IB Samples for NMR	93
4.3.4 Aggregation and Secondary Structure Prediction	93
4.3.5 Preparation and Purification of Recombinant SOD1	93
4.3.6 FTIR Spectroscopy	94
4.3.7 FTIR Spectral Analysis	94
4.3.8 Congo Red Binding	95
4.3.9 Urea denaturation of IBs	95
4.4 Results and discussion	96
4.4.1 SOD1 aggregation correlates to its biophysical parameters	96

4.4.2 qHDX reveals native structural features	97
4.4.3 SOD1 IBs exhibit multi-exponential amide exchange	101
4.4.4 CR and FTIR analysis of qHDX samples corroborates similar IB structures between mutants	104
4.4.5 Urea denaturation relates stability of SOD1 IBs to pure protein stability	107
4.4.6 CR, FTIR and urea denaturation of other condition IBs support similar structure but differing stability	111
4.4.7 Ensemble model describes SOD1 IB structure	115
4.5 Conclusions	116
Chapter 5: Adaptation of qHDX method to other systems	118
5.0 Preface	118
5.1 Introduction	118
5.1.1 Low aggregation systems and early aggregates	118
5.1.2 Adnectins: adapting qHDX to another protein system	119
5.1.3 Apomyoglobin: examining chain length dependence by qHDX	120
5.1.4 Significance of adapting qHDX to new systems	122
5.2 Materials and Methods	123
5.2.1 Limited isotopic labelling of IB (media switch) of SOD1	123
5.2.2 Protein expression of apoMb fragments	123
5.2.3 Protein expression of Adn variants	124
5.2.4 Detergent washing of apoMb fragments	124
5.3 Results and Discussion	125
5.3.1 Strategies to study earlier SOD1 aggregates	125
5.3.2 Assigning strategies: similarities, differences, and new insights with Adn and apoMB	130
5.3.3 Adnectin IBs display a native-like ensemble	140
5.3.4 Apomyoglobin appears to aggregate similarly <i>in vivo</i> and <i>in vitro</i>	142
5.4 Conclusions	147
Chapter 6: Conclusions and Future work	149
6.1 Conclusions	149
6.2 Logical Extensions	149
6.3 New directions	150
References	152
Appendix	168

List of Figures

Figure 1. 1 Theories of Protein Folding	2
Figure 1. 2 Technical considerations for the structural characterization of protein aggregates.	7
Figure 1. 3 Data processing workflow for FTIR analysis of protein aggregates.	13
Figure 1. 4 Sample production for solid-state NMR analysis of protein aggregates.	16
Figure 1. 5 Overview of a qHDX-MS workflow.	21
Figure 2.1 The Structure of mature Holo Cu, Zn SOD1	28
Figure 2.2. Mechanism for SOD1	28
Figure 2.3 Maturation States of SOD1	30
Figure 2.4 WT Sequence Aggregation Propensity	35
Figure 2.5 ZipperDB Predictions of SOD1 variants relative to WT.	36
Figure 2.6 ConSurf Prediction for WT SOD1	37
Figure 2.7 Logo for WT SOD1 amino acid sequence.	38
Figure 2.8 Impact of temperature of the aggregation of SOD1 variants.	39
Figure 2.9 Correlation between SOD1 aggregation and ConSurf sequence conservation score	46
Figure 2.10 Correlations between SOD1 aggregation propensity and SOD1 biophysical parameters	48
Figure 2.11 Impact of Zinc on SOD1 aggregation	49
Figure 2.12 Comparison of average SOD1 aggregation in LB to minimal media	51
Figure 2.13 Comparison of zinc rescue of SOD1 aggregation relative to mutant position	53
Figure 2.14 Proximity of A4 and L106 to dimer interface	54
Figure 2.15 Gel of SOD1 standards	56
Figure 2.16 Relationship between band density and mass of protein	57
Figure 3. 1 <i>qHDX method overview</i>	61
Figure 3. 2 Quenched hydrogen-deuterium amide exchange on protein aggregate	63
Figure 3. 3 Flowchart of the IB qHDX method.	65
Figure 3. 4 ^1H - ^{15}N HSQC spectra for measuring relative protection.	71
Figure 3. 5 Comparison of freeze-thaw and emulsifying on IB structure and quantity	74
Figure 3. 6 Impacts of dilution and incubation time on IB structure	75
Figure 3. 7 Measured intrinsic rate constants of amide exchange for different amino acids.	77
Figure 3. 8 DCA methyl ^1H peak as an internal indicator of sample pH.	79
Figure 3. 9 ^1H - ^{15}N HSQC spectral overlay of a qHDX sample in DMSO with 0% (v/v) and 5% H_2O added	80
Figure 3. 10 Comparison of water content in qHDX samples calculated by two different methods	82
Figure 3. 11 ^1H - ^{15}N HSQC spectrum of A4V SOD1 unfolded in 95% DMSO / 5% H_2O at pH 5.5 with backbone residue assignments	84
Figure 3. 12 Peak movements in SOD1 point mutations show impact of sequence on peak location	87
Figure 3. 13 Comparisons of replicate qHDX sample results.	89
Figure 4. 1 SOD1 structure, stability, and IB formation.	97
Figure 4. 2 Structure of mutant SOD1 IBs from qHDX analysis	100
Figure 4. 3 ^1H - ^{15}N HSQC spectra for aliquots of G85R SOD1 IBs at different times of exchange in D_2O .	102
Figure 4. 4 IB amide protons show fast and slow exchanging components in D_2O	103
Figure 4. 5 FTIR spectra of mutant SOD1 IBs exhibit similar absorbance and second derivative features	105
Figure 4. 6 Integrated area of the component bands from curve-fitting of mutant SOD1 FTIR spectra	105

Figure 4. 7. FTIR spectra of SOD1 IBs exhibit native and amyloid features	106
Figure 4. 8 CR spectra of SOD1 IBs exhibit native and amyloid features.	107
Figure 4. 9 Urea denaturation of SOD1 IBs grown at 37°C with no metals	109
Figure 4. 10 Correlations of urea denaturation to biophysical parameters of SOD1.	110
Figure 4. 11 CR spectra of SOD1 variants grown under different conditions.	111
Figure 4. 12 FTIR spectra of SOD1 variants grown at 25°C overnight.	112
Figure 4. 13 FTIR spectra for SOD1 variants grown at 37 °C with zinc	113
Figure 4. 14 Comparing urea denaturation curves for SOD1 variants grown under different conditions	115
Figure 4. 15 SOD1 IB formation involves multiple pathways that include native and non-native interactions	116
Figure 5. 1 Example of lower aggregation conditions of interest of SOD1.	119
Figure 5. 2 Natively folded Adnectin structure	120
Figure 5. 3 Natively folded full length apomyoglobin structure	121
Figure 5. 4 Overview of different expansions of qHDX method	122
Figure 5. 5 Comparison of SOD1 expression at different time points using a media switch growth	125
Figure 5. 6 Protonated spectra relative to average SOD1	126
Figure 5. 7 Fraction amide protection of T ₁ and T ₄ media switch of A4V SOD1 compared to average SOD1 IB protection	128
Figure 5. 8 Change in relative amide protection of a residue over expression time compared to average IB residue protection	129
Figure 5. 9 Comparison of average chemical shift of each nucleus type per amino acid between SOD1 and Adn in DMSO	132
Figure 5. 10 ¹ H- ¹⁵ N HSQC spectra of apoMb ₁₋₁₁₉ under different preparation conditions in 95% DMSO/5% H ₂ O	136
Figure 5. 11 Protonated ¹ H- ¹⁵ N HSQC spectra of apoMb fragments in 95%DMSO/5% H ₂ O	137
Figure 5. 12 Preliminary resonance assignments of apoMb ₁₋₇₇ in 95% DMSO/5% H ₂ O at pH 5.4	139
Figure 5. 13 Comparison of denatured apoMb ₁₋₇₇ backbone assignments under different solution conditions	140
Figure 5. 14 Spectral overlays of qHDX samples with their respective protonated spectra	140
Figure 5. 15 Measured intrinsic exchange rate constant of amide exchange for different amino acids	141
Figure 5. 16 Structure of Adnectin IBs from qHDX analysis	142
Figure 5. 17 Fraction amide protection of apoMb ₁₋₇₇ and apoMb ₁₋₁₁₉ for residues 1-77 compared to native and calculated parameters	145
Figure 5. 18 Fraction amide protection of apoMb ₁₋₁₁₉ for residues 1-119 compared to native structure	147
Figure A.2. 1 Example SDS-PAGE used to assess aggregation propensity	167
Figure A.3. 1 Amide exchange decay fits in CCPNMR software	169
Figure A.3. 2 Effects of pH on ¹ H- ¹⁵ N HSQC cross peak intensity and chemical shift	170
Figure A.3. 3 Effects of varying IB storage	171
Figure A.3. 4 ¹ H 1D NMR spectra overlaid to highlight differences in water content across samples.	172
Figure A.3. 5 Overlay of ¹ H- ¹⁵ N HSQC spectra at different acquisition temperatures.	173
Figure A.3. 6 Impact of resuspension in D ₂ O on qHDX experiments	174

List of Tables

Table 1. 1 FTIR spectroscopy on cellular protein aggregates reveals characteristic spectral features of different β-structures	11
Table 2. 1 Composition of Buffer and Solutions Used	32
Table 2. 2 fALS point mutation locations and biophysical parameters	33
Table 2. 3 Measured aggregation propensity of SOD1 variants measured at 25 °C	40
Table 2. 4 Measured aggregation propensity of SOD1 variants measured at 30 °C	42
Table 2. 5 Correlation of normalized aggregation propensity scores to cellular aggregation propensity assessed by SDS-PAGE	46
Table 2. 6 Measured distance of each mutation location to metal binding sites in mature SOD1	52
Table 2. 7 Standards loaded in SOD1 gel of standards and corresponding band densities	56
Table 2. 8 Expected and actual densitometry ratios of gel of standards	58
Table 2. 9 Expected and actual densitometry ratios of gel of standards compared to empty lanes	58
Table 3. 1 Experimental parameters for 2D and 3D NMR experiments used in resonance assignment	68
Table 3. 2 FTIR peak locations for different sample incubations	75
Table 3. 3 Comparison of SOD1 chemical shifts to random coil (RC)	85
Table 5. 1 Average deviation of different nucleus types for denatured proteins	131
Table 5. 2 Primary sequence effects on ^{15}N chemical shifts of amino acids	133
Table A.2. 1 p values for all correlations given in thesis	175
Table A.3. 1 Primary sequence effects on ^{15}N chemical shifts of amino acid	179
Table A.4. 1 Full chemical shift assignments of SOD1 in 95% DMSO/5% H_2O pH 5.4	180
Table A.4. 2 Comparison of qHDX protection to native structural features	190
Table A.4. 3 k_{ex} of SOD1 residues fit to both a single or double exponential decay	197
Table A.4. 4 Fit values of urea denaturation	202
Table A.4. 5 Fit peak locations for FTIR of SOD1 IB samples grown with zinc	204
Table A.4. 6 Fit peak locations for FTIR of SOD1 IB samples grown at lower temperatures	204

List of Abbreviations

Abs	conformation specific antibodies
AD	Alzheimer's Disease
Adn	Adnectin protein
AFM	atomic force microscopy
ALS	amyotrophic lateral sclerosis
ANS	1-anilino-8-naphthalene sulfonate
apo	Metal-free
apoMb _x	Apo myoglobin fragment of residues X
apoSH	Immature SOD1 with reduced disulfide and no metal
ATR-FTIR	Attenuated total reflection FTIR
CCS	Copper chaperone for SOD1
CD	Circular dichroism
C _{mid}	Concentration at which half the protein is denatured
cold shock protein A (CspA)	
CR	Congo red
DCA	Dichloroacetic acid
DMSO	Dimethyl sulfoxide
DNP	dynamic nuclear polarization
DTT	dithiothreitol
E coli	Escherichia coli
E, ZnSH	SOD1 with reduced sulfide and one zinc bound per monomer
EDTA	ethylenediaminetetraacetic acid
EM	electron microscopy
FTIR	Fourier transform infrared spectroscopy
HD	Huntington's Disease
HDX	Hydrogen deuterium exchange
HET-s 218-289	HET-s fungal prion
Holo	With all bound metals
holoSS	Mature SOD1 with disulfide intact and bound metals
HSQC	heteronuclear single quantum correlation
IB	Inclusion body
IPTG	isopropyl β-D-1-thiogalactopyranoside
k _{ex}	observed amide proton exchange rate
k _{int}	intrinsic amide proton exchange rate
LB	Luria broth
MAS	magic-angle-spinning
NMR	Nuclear magnetic resonance
NT	the N-terminal domain of major ampullate spidroin 1,
OD _{350nm}	Optical density at 350 nm
OD _{600nm}	Optical density at 350 nm
OMP F	Outer membrane protein F
PD	Parkinson's Disease
PrP	prion protein
PrPsc	PrP mutant aggregates
pWT	Pseudo wild-type
qHDX	Quenched hydrogen-deuterium exchange

qHDX MS	Quenched hydrogen deuterium exchange mass spectrometry
raSOD1	Immature SOD1 with reduced disulfide and no metal
S/N	signal-to-noise
SOD1	Cu,Zn-superoxide dismutase
ssNMR	Solid state nmr
ThT	thioflavin T
T _m	Melting temperature
Tris	tris(hydroxymethyl)aminomethane
TTR	transthyretin
WT	Wild type
xH	fraction of H ₂ O in the quenching buffer
ΔG_{dimer}	Free energy change of dimer dissociation
$\Delta G_{\text{monomer}}$	Free energy change of monomer unfolding
ΔG_{U}	Free energy change for the folding/unfolding reaction

Chapter 1: Introduction

1.0 Preface

This introductory chapter will cover the background of protein folding and misfolding before giving an overview of a range of low to high resolution methods that are implemented to study aggregate structure.

The latter portion of this chapter (structural methods) was previously published in an invited review in *The Journal of Biological Chemistry* that was co-authored by Dalia Naser, Anna Schaefer, and Bruna Siebenechler under the supervision of Elizabeth Meiering. Dalia Naser conceived the review and wrote sections on conformational specific antibodies, quenched hydrogen deuterium exchange by NMR, and quenched hydrogen deuterium exchange by mass spectrometry. Bruna Siebenechler wrote sections on solid state NMR and contributed to researching FTIR. Anna Schaefer wrote sections on FTIR as well as introductory and concluding statements. Consequently, the ssNMR and FTIR sections included in this introduction are greatly shortened to those in the article. Additionally, portions of discussion on quenched hydrogen deuterium exchange by NMR were moved to more relevant chapters (Chapter 4).

Citation: A. Schaefer⁺, D. Naser⁺, B. Siebeneichler⁺, M. Tarasca, E. Meiering, *J. Biol. Chem.* **2022**, *In Press*

1.1 Protein Folding, Misfolding, and Aggregation

1.1.1 Theories on Protein Folding

Proteins are large macromolecules essential for life. Comprised of a linear sequence of amino acids linked by peptide bonds, a protein has a long journey before reaching its final folded form. The completely unfolded state, or polypeptide chain, is known as the primary structure. The sequence of the polypeptide chain is a key determinant of the folded protein^[1]. The folding process, however, is complex, and there is no consensus on how proteins fold or why they fold a certain way^[1]. Prior to the advent of experimental methods that allow its observation, protein folding was thought to occur by following an identical series of steps and going through defined intermediates (Figure 1.1A). The classical theory began to change after several key studies. In 1961, Anfinsen established through his work on ribonuclease that proteins folded both spontaneously and quickly^[2]. Working in parallel was Levinthal, who posed the Levinthal paradox. Levinthal reasoned that if a protein were to sample every possible conformation to reach its global energy minimum, the time required would be longer than the age of the universe^[1]. As it was known from Anfinsen that folding occurred fairly quickly, Levinthal proposed that simple interactions between amino acids guide the folding process^[3]. This idea was soon

followed by the development of Anfinsen's dogma in 1973^[3]. The dogma states that, at optimal folding conditions, a unique and stable native state is both kinetically available and minimal in energy.

As the theory on protein folding began to shift, models of the process were developed^[4]. One major model is the framework model^[5]. The framework model suggests a stepwise process – the unfolded peptide forms a secondary structure, and this structure is used as a scaffold for the tertiary structure. Two possibilities are postulated for how this process occurs. The diffusion-collision process involves the formation of secondary structures which are free to diffuse and collide with one another, forming tertiary structures in the process^[6]. The nucleation process is similar, and proposes the slow formation of a native-like core structure which then samples tertiary conformations^[7]. A second model is the hydrophobic collapse model, in which the peptide initially collapses due to the hydrophobic effect to form a hydrophobic core around which packing interactions occur to form the final folded structure^[8]. Combined, these ideas led to a change in the theory of protein folding. Regardless of the nature of the interactions occurring, protein folding is now thought of in terms of an “energy funnel” (Figure 1.1B). This implies that there are multiple potential intermediates and folding pathways, but that the protein becomes more energetically stable as it moves down the energy funnel towards the final native conformation^[1].

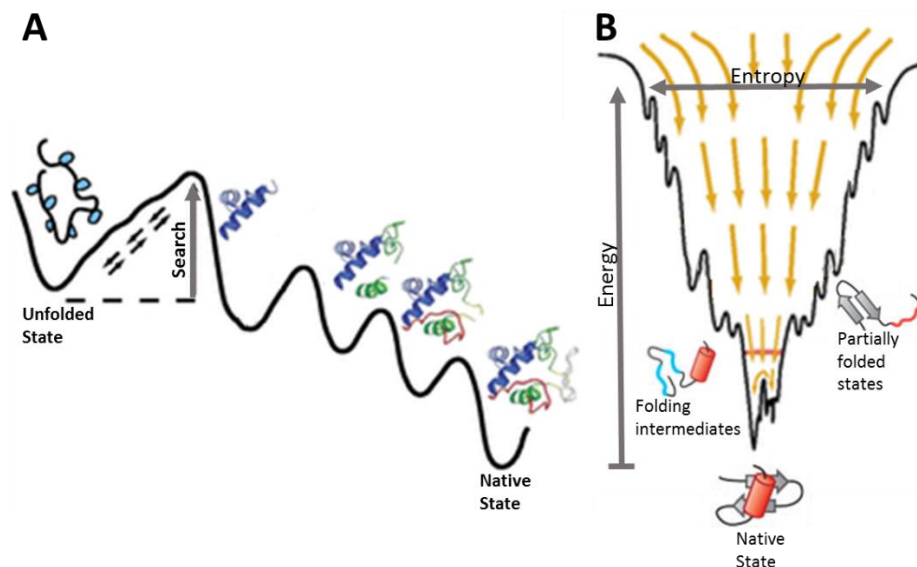


Figure 1. 1 Theories of Protein Folding

(A) The classical theory of protein folding, showing one defined pathway with consistent intermediates. (B) The energy landscape funnel, in which a protein may sample various conformations and intermediate states that gradually become lower in energy until it reaches the final state. Adapted from Englander & Mayne, 2014 ^[1] and Hartl & Hayer-Hartl, 2004 ^[9].

1.1.2 Protein Misfolding

Given the complexity of the folding process, it is not surprising that protein misfolding can and does occur^[10]. As partially folded intermediates are formed, it is possible for them to interact with each other incorrectly and form non-native complexes. Should improper interactions happen repeatedly, an aggregate is formed. It should be noted that aggregate formation is not limited to partially folded states, and that an assortment of protein states can lead to aggregation^[11]. This includes native protein, degraded, partially unfolded, and fully unfolded proteins. A convenient way to classify aggregates is by the types of bonds^[11]. Non-covalent aggregates are held together by various weak non-covalent interactions such as van der Waals, hydrogen bonding, and electrostatic forces. A common non-covalent aggregation pathway involves hydrophobic residues: if a protein adopts a conformation in which a hydrophobic patch is solvent-exposed, the thermodynamic drive to bury the region can result in improper protein-protein interactions. On the other hand, aggregates may also contain covalent bonds. These aggregates may be bound, for example, by disulphide linkages between cysteine residues or through other crosslinking mechanisms^[12,13].

As with protein folding, various models of protein misfolding mechanisms also exist. One such model is based on nucleation or seeding, and is called the nucleation-propagation polymerization model^[14]. As the name suggests, the idea is that an altered species will create a thermodynamic nucleus upon which aggregates may form in a growth phase^[14,15]. The model assumes that nucleus formation is thermodynamically unfavourable, and that the post-nucleation aggregation steps result in a lower energy state. A similar model is the downhill polymerization model, which is used to describe the aggregation of transthyretin (TTR)^[15]. In this model, the formation of the nucleating species is energetically favourable, and each subsequent growth step results in a lower energy state. A third model of interest is the “prepared model”^[16]. In this model both the protein’s sequence and its surroundings determine the likelihood of aggregation. Under certain conditions, the protein is “prepared” and the ensemble of protein species in solution will contain aggregation-prone species. Thus, a mutation can “prepare” the protein by altering the accessibility of an aggregation-prone state. This model suggests that small, local changes to the protein conformation can be as damning as unfolding to a protein^[16].

1.2 Methods to determine protein aggregate structure

1.2.1 Overview

Protein aggregation is a widespread phenomenon that is of great importance in disease and in biotechnology. Numerous pathological conditions such as Alzheimer's Disease (AD), Parkinson's Disease (PD), Huntington's Disease (HD), amyotrophic lateral sclerosis (ALS), diabetes, cataracts, and cancer are associated with the deposition of amyloid fibrils and other protein aggregates^[17,18]. Moreover, aggregation may occur at many points during the expression, purification, and use of recombinant proteins for medical or biotechnological purposes. Due to their broad significance, great effort has been put into determining the structural and biochemical properties of protein aggregates under native conditions^[19]. Many of these studies have been performed with hope of selectively preventing or controlling aggregation, while others investigate how protein folding, maturation and environment can influence toxic aggregates. However, limitations such as structural heterogeneity, sample impurity and size constraints of structure determination techniques have obstructed a high-resolution understanding of protein aggregates.

Historically, protein aggregates were classified into two structural categories: amyloid or amorphous aggregates. By definition, amyloid describes fibrillar aggregates with ordered cross β -sheet structure as identified using x-ray diffraction and electron microscopy^[20]. Mature amyloid fibers are often found in the characteristic plaques of protein deposition diseases and thus have been widely studied in terms of their structure, formation, and toxicity. Commonly, amyloid is formed via a nucleation-dependent polymerization; however, many models exist for the exact molecular mechanism^[21]. In contrast, large, insoluble deposits known as inclusion bodies (IBs) were categorized as amorphous when first discovered^[22,23]. IBs commonly form when recombinant proteins are overexpressed in bacteria, but they have also been compared with inclusions found in neurodegenerative diseases such as ALS and HD. For a long time, IBs were considered deposits of misfolded and inactive proteins that posed an obstacle in recombinant protein production. However, recent studies have revealed that this is not strictly the case and IBs have gained much attention in the fields of biotechnology and medicine as self-immobilized catalysts and drug delivery systems^[24]. A more thorough discussion of IBs can be found in Chapter 2.

Classical techniques for obtaining structural information about protein aggregates typically use low-resolution data on whether a protein is aggregating coupled with other methods to obtain a more detailed view. One of the simplest approaches to detect the presence of amyloid is UV-visible absorbance or fluorescence spectroscopy to monitor optical changes of the sample upon binding of dye

by the aggregate. Commonly used dyes include Congo red (CR), thioflavin T (ThT) and 8-anilidonaphthalene-1-sulfonic acid (ANS). Upon binding amyloid aggregates, CR exhibits characteristic green–yellow birefringence under polarized light and a shift of the absorption maximum from ~490 nm to ~540 nm, whereas ThT and ANS exhibit increases in fluorescence when they bind to amyloid species [25,26]. While these dyes have proved useful as diagnostic tools, they are non-specific and so can bind to other cellular components; for example, binding of CR to certain lipids and ThT to other protein hydrophobic pockets has been described [27,28]. ANS is even more limited because its binding to amyloid structure can induce protein conformational changes [26]. Another low-resolution technique is the binding of conformation specific antibodies (Abs), which can qualitatively detect the presence of aggregates and their location. Such Abs have been developed for both sequence dependent and independent (ie. fibril binding) structure recognition [29,30]. These Abs provide only a slightly more detailed view of structural features or surface exposed residues because they can bind a range of conformations with varying affinities [31–33].

Other characterization techniques, including protease digestion, atomic force microscopy (AFM) and electron microscopy (EM), report on global structure or morphology of aggregates. The degree of digestion with proteases can provide insight on the compactness of the aggregate structure [34,35] and, taking this one step further, fragment sizes can be obtained using subsequent detection by a variety of methods, each providing a different resolution. Approximate fragment sizes can be identified using SDS-PAGE or Western blotting while mass spectrometry (MS) provides much more specific information on what regions of the protein are exposed in the aggregate [36–38]. Similarly, AFM and EM can help visualize macroscopic properties such as length, width, rigidity, and the degree of branching. AFM provides the advantage that it can characterize aggregates at single molecule resolution. An adaptation of transmission EM known as cryo-EM is at the forefront of structural biology; much more detailed three-dimensional structures, with resolutions comparable to x-ray crystallography, can be obtained with the benefits that crystallization is not required, and molecules are not destroyed by the high energy electron beam. These methods will not be considered here as they have been covered in excellent recent reviews [39–42].

On a microscopic level, secondary structure determination methods such as circular dichroism (CD) spectroscopy and Fourier transform infrared (FTIR) spectroscopy can report on aggregates in many physical states [43]. While both are able to detect increased β -sheet content, a common feature of aggregating proteins, FTIR provides more detailed observations due to its higher extinction coefficient

for β -structures and better fitness for turbid samples. CD has high relative absorbance for α -helical structure, while FTIR provides a more uniform signal for different secondary structures allowing for easier visual assessment of structural changes ^[44,45]. At the atomic level, methods such as quenched hydrogen deuterium exchange (qHDX) and solid-state nuclear magnetic resonance (ssNMR) provide high resolution information on which residues are participating in the aggregate structure, allowing for a mechanistic interpretation of aggregate formation ^[46–51]. These specific methods will be emphasized later in this review as they provide a wealth of complementary and high-resolution information.

There is clearly an abundance of techniques to study protein aggregation. Unfortunately, due to limitations caused by sample heterogeneity or impurity, many investigations have largely been limited to studying aggregation *in vitro* - under conditions that may be far different from physiological or pathological conditions. Furthermore, most of the methods applied *in vitro* for the biophysical and structural characterization of these species usually involve purification processes that can perturb the aggregation process and ultimately may change the aggregate structure. Thus, it is crucial that structure determination methods be adapted to study cell-derived aggregates both *ex vivo* and in whole cells or tissue. This chapter will discuss the historical basis and recent advances in methods that can be used to study the structure of aggregates, with a specific focus on applications to cell-derived or in-cell and thus impure aggregates; cell-derived or cellular will refer to aggregates isolated from human or bacterial cells with minimal processing, whereas purified aggregates will describe purified protein induced to aggregate *in vitro* either by seeding or altered solution conditions. Determining aggregate structures is of great importance to elucidate critical details of aggregation-prone species and association mechanisms as they apply to disease and biotechnological applications. Methods reviewed include conformation specific antibodies, ATR-FTIR and FTIR microscopy, qHDX with detection by MS or NMR, and ssNMR. An overview of these methods is represented in Figure 1.2

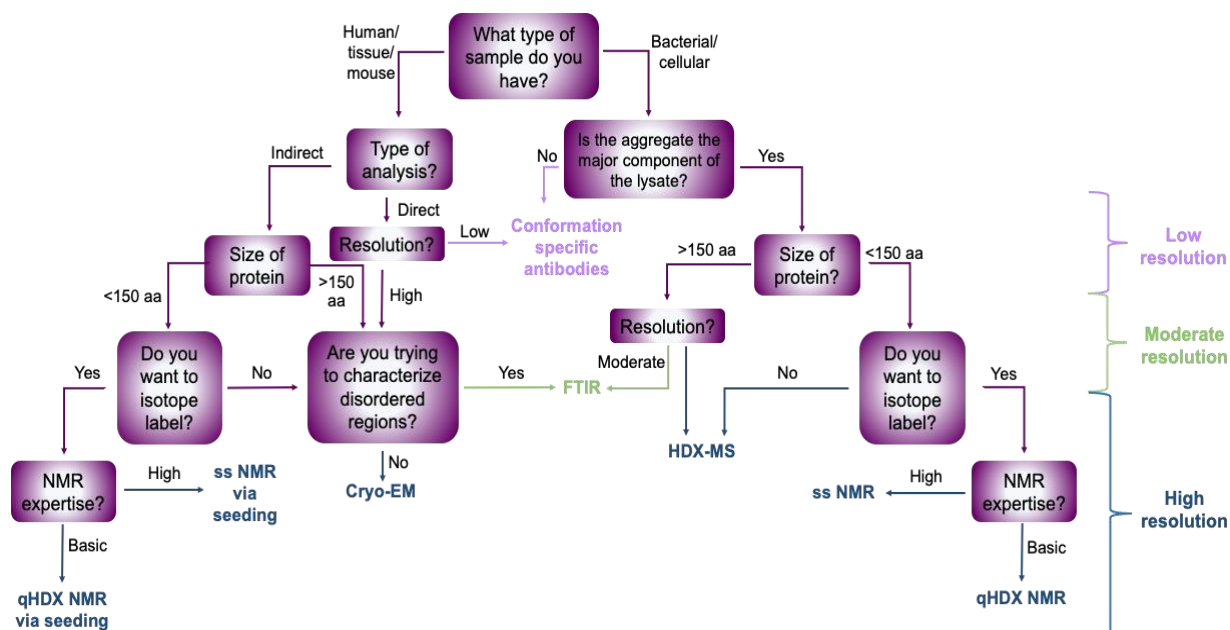


Figure 1. 2 Technical considerations for the structural characterization of protein aggregates.

Structural characterization methods organized based on resolution, sample limitations such as size and impurity, considerations of time, effort and cost, and level of expertise.

1.2.2. Conformation specific antibodies: moving beyond detection to structural information

Antibodies are often used in microbiology methods such as Western blotting, immunohistochemistry and dot blotting. While these methods do not all necessarily provide structural information, they can serve as a means of determining aggregate composition and a probe for detecting the protein of interest and its localization. The appeal of these methods is twofold: 1) they do not require a large amount of sample; and 2) the sample does not need to be pure (Figure 1.2). Thus, they are ideal for samples such as lysates (cell or animal) or even patient tissue ^[52,53] – to quantify a protein of interest ^[54] or diagnose a medical condition ^[55].

Charles Glabe's group conducted seminal work in using Abs for aggregate structure determination, developing sequence-independent antibodies that recognize amyloid fibrils and oligomers ^[29,56,57]. These helpful tools allow discrimination of different types of aggregates and helped broaden the definition of "amyloid" by detecting specific aggregated structures ^[58–61]. These antibodies are now commercially available and have become standard tools to study aggregates. However, further development of antibodies has also resulted in protein-, sequence- and conformation-specific antibodies (Abs) against many disease-associated proteins such as Tau, huntingtin, RNA-binding protein FUS, and Cu, Zn superoxide dismutase (SOD1).

For SOD1, Abs have been developed to recognize epitopes throughout the protein structure or detect specific conformations and maturation forms of the protein ^[62]. These Abs can be particularly useful in aggregate characterization. A simple example compared the binding of the B8H10 SOD1 Ab (specific for misfolded SOD1 with exposed loop 4) and C4F6 (specific for oxidized protein) to examine rat spinal cords with mutated SOD1 ^[30]. Pickles and colleagues detected lower binding of C4F6 relative to B8H10, indicating less oxidized protein and increased misfolded SOD1 in the aggregate. They also tested the B8H10 antibody on ALS patient lymphoblasts and saw binding, validating, through indirect structural information, that the mouse model aggregates are structurally like human ones. Conformation-specific and peptide recognizing antibodies have also identified distinct conformational strains of aggregates in mutant SOD1 ALS mice models ^[63–65]. Similar work has been done with Abs against Tau protein, in which antibodies have been shown to bind oligomeric rather than monomeric protein in aggregates, and have identified a key aggregation-prone region in the proteins' N-terminal domain ^[66,67], and on expanded huntingtin polyQ domains, wherein Ab studies discerned multiple toxic conformations ^[68].

An interesting recent example of the power of Abs to contribute aggregate structural information was provided by Semmler and colleagues ^[69]. Through immunoprecipitation with two SOD1 Abs analyzed by mass spectrometry, they identified a particular ubiquitin ligase binding partner associated with mitochondrial misfolded SOD1. This indicated that a specific SOD1 loop is exposed for this interaction to occur. Investigation of other SOD1 mutants via Ab-binding ascertained that not all variants could interact with the ligase, suggesting the existence of different misfolded structures and misfolding pathways. While this work does not directly investigate the aggregate structure, Semmler *et al* demonstrate, through discovery of the binding sites of associated proteins, how moderate structural resolution information about exposed segments in the *in vivo* misfolded SOD1 species can be obtained.

Another interesting study by Pickles *et al.* tracked SOD1 aggregation through disease progression by monitoring Ab binding to different rat tissues over time ^[32]. Through use of Abs that bind throughout the structure, the group established that there were two distinct classes of aggregates: one structure that exists primarily in the motor neurons and neutrophils, and a second that exists as a fibrillar network in neurons, axons, and dendrites. Additionally, dot blot experiments on pure SOD1 in different maturation states revealed that the Abs preferentially bind unmetalated (apo) and disulphide reduced protein over more mature folded protein. Other studies on SOD1 Ab affinities have been conducted and have also found preference for specific aggregate structures or different ALS mutant SOD1s ^[33]. Interestingly, it

was shown that not all the Abs were reactive in a Western blot, providing a note of caution when interpreting Western blots.

While none of the aforementioned is high resolution structural information, each piece is a structural clue, revealing that distinct structures exist, and potentially the order in which they are formed in a structural maturation process. These types of results highlight the applicability of Abs for in-cell studies, which is especially important as protein folding, maturation and environmental factors can influence the toxicity of aggregates and these conditions may be hard to mimic *in vitro*. Thus, many researchers have utilized Abs to investigate toxic species in diseased tissues ^[31,32,56,62,70–72]. One example of this strategy used hypersensitive Abs reporting on two known conformations of A β aggregates. By time-dependent staining of plaques from mouse tissue, Nyström and colleagues identified a unique structural transition of the plaque core at old age that they proposed could be the cause of a shift in toxicity ^[72]. Similarly, when examining brain tissue, conformation specific Abs identified amyloids formed by huntingtin exon-1 with looser loops and turns are more toxic than rigid, extended fibrils ^[71].

In summary, the use of conformation specific Abs has gained traction, with creative researchers applying them in new and exciting ways to obtain structural information on impure and even patient aggregates. Unfortunately, the results are limited by relatively low resolution and because antibodies typically bind with varying affinity to different conformations of a protein, limiting precise interpretation. Nevertheless, increasing studies performed with Abs are obtaining disease relevant information including localization and the identification of toxic species in diseased tissue. It is expected that as more conformation-specific Abs continue to become commercially available, that there will be increasing applications of these valuable tools in future studies as diagnostic tools and therapeutics for protein aggregation diseases ^[73]. The work done on systems such as SOD1 wherein many Abs have been tested for binding specific structures and sequence illustrates the value of such expansion.

1.2.3 FTIR: Secondary structure at a glance

FTIR is a rapid and moderate sensitivity means of monitoring protein secondary structure and performing compositional analysis for biological samples. FTIR spectra report on covalent bonds which absorb in the infrared region and result in a characteristic fingerprint for a given molecular mixture. Scientists studying protein conformational changes in the context of protein dynamics, misfolding and aggregation rely specifically on the carbonyl (C=O) stretching vibration of amide bonds which appears in the 1600-1700 cm⁻¹ range, termed the amide I region. This range has characteristic frequency ranges for

α -helical, β -strands, random coil, loop and turn structures due to distinctive hydrogen-bonding patterns of the protein backbone, and signal is largely independent of primary sequence allowing facile structure comparison between proteins ^[43,74]. As extensive FTIR studies have been performed on protein aggregates, a wealth of previous knowledge is available about their spectral characteristics which can assist current researchers with their investigations (Table 1). Additional signals, termed amide II and III, are observed at 1500-1600 cm^{-1} and 1175-1310 cm^{-1} , respectively. These regions are infrequently used for structure analysis due to their lower intensity and overlap with absorbance bands of amino acid side chains in the case of amide II. However, recent work has highlighted the potential of amide III to be very sensitive to secondary structure and assignments of secondary structure to specific wavenumbers were completed, similar to what has been done for amide I ^[75].

Table 1. 1 FTIR spectroscopy on cellular protein aggregates reveals characteristic spectral features of different β -structures

Bands (cm ⁻¹)	Assignment	Sample	Technique	Reference
Cell-derived IBs				
1622	Amorphous aggregates	Asparaginase (induction at 37-42°C)	ATR-FTIR	[35]
1628-1629	Amyloid aggregates	Asparaginase (induction at 20-30°C)	ATR-FTIR	
1632	Coiled-coil	ZapB	ATR-FTIR	[76]
1651	α-helix		ATR-FTIR	
1620-1630	Inter β-sheet			
1650	α-helix	Lipase	FTIR	[77]
1680	β-turn			
1623	Aggregates			
1634	Native β-sheet	Phospholipase A2	ATR-FTIR	[78]
1650	α-helix			
1625	Amyloid	VP1-GFP in <i>Pichia pastoris</i>	ATR-FTIR	[79]
1680	Antiparallel β-sheet			
In vitro-derived aggregates				
1622	Inter β-sheet	SOD1 fibers (60°C+shaking)	FTIR	[80]
1618	Inter β-sheet	Human Serum Albumin (heated)	ATR-FTIR	[81]
1630/1690	Amyloid (parallel β-sheet)	Aβ42 fibril	ATR-FTIR	[82]
1630/1690	Oligomers (Antiparallel β-sheet)	Aβ42 oligomer	ATR-FTIR	

Details of recent FTIR analyses on aggregates are compiled.

Protein aggregates have been extensively studied with FTIR as this method circumvents the typical limitations of structural analysis techniques: it has short sample measurement times allowing the observation of time dependent changes, it requires relatively low sample amount (<100 ng), and is applicable to insoluble and impure samples [74]. Although both transmission and attenuated total reflectance (ATR) FTIR have been used to study aggregates, ATR-FTIR has become the preferred method

as it is more tolerant of insoluble sample components and less sample preparation is required. FTIR spectra of proteins associated with neurodegenerative diseases such as A β (Alzheimer's), α -synuclein (Parkinson's), and many others have been acquired for fibrils produced *in vitro* ^[83]. These studies revealed a key spectral shift in the β band towards lower wavenumber (1611 cm⁻¹ to 1630 cm⁻¹) in fibrils, allowing differentiation of native and amyloid-like intermolecular β -sheets ^[84]. Recent studies show a broad spectral range for intermolecular β -sheets; an FTIR study on synthetic amyloid beta revealed the size of oligomers correlated with the FTIR absorbance, with larger oligomers having lower wavenumbers ^[85]. Others have distinguished between parallel and antiparallel- β -sheet shift assignments based on the appearance of a higher order β peak above 1685 cm⁻¹ in antiparallel aggregates ^[86–88]. All of the preceding information cannot be measured by other secondary structure determination techniques such as circular dichroism nor by high-resolution techniques such as qHDX.

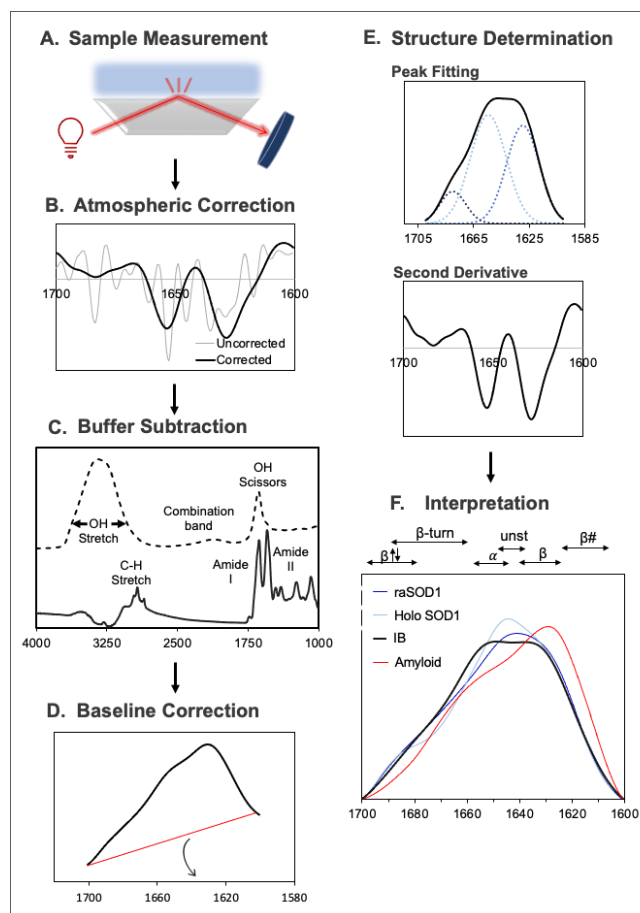


Figure 1.3 Data processing workflow for FTIR analysis of protein aggregates.

Following sample measurement by ATR-FTIR, the data are transformed to give an absorbance spectrum. This absorbance spectrum is sensitive to water and carbon dioxide in the atmosphere and therefore can be corrected using patented algorithms from acquisition and analysis software such as Bruker Opus. Following atmospheric correction, a post-buffer scan must be subtracted from the sample to remove contributions from water. A flat baseline is applied to the amide I region (1600-1700 cm^{-1}) and then the resulting spectrum can be analyzed using structure determination methods such as multivariate peak fitting, second derivative analysis, Fourier self-deconvolution and more [89]. Interpretation is often aided by comparison to FTIR spectra of purified proteins. Data are for SOD1, where panel F is modified from Naser et al. [49].

The above findings from studies of pure protein aggregates allowed enhanced interpretation of *in vivo* aggregation. Many cell-derived aggregates, including bacterial IBs, show increased β character [90,91] and in some cases, amyloid-like features have been identified in cell-derived IBs using FTIR [90,92]. More recently the coexistence of fibril-like intermolecular β -sheets with significant native structure [35,49,76,93] has further elucidated the heterogeneity of IBs. As the amount of residual native structure in IBs often correlates with the yield of enzymatically active protein, FTIR can be an essential screening tool for researchers trying to understand and optimize functional IBs [93–95]. Similar screening has been

performed to investigate the structure of tagged IBs where the tag of choice can facilitate the assembly of the protein of interest into active aggregates ^[96]. Moreover, the effect of expression temperature and culture pH has also been investigated in the context of functional IBs ^[35,97–99]. However, conflicting conclusions have been reported for the temperature dependence of IB structure. Studies on lipase from the psychotropic bacterium *Pseudomonas fragi* (PFL) indicate increased native-like structure by FTIR and increased biological activity at lower temperatures ^[97]. Conversely, asparaginase II IBs exhibited increased native-structure and higher enzymatic activity at higher temperatures ^[35]. It should be noted that this contradiction may be due to other biochemical properties of the enzymes, such as thermostability, affecting the structural ensemble in the cell and altering which conformations are involved in aggregation. In such studies, for accurate interpretation, the protein of interest should be the main component and not be obscured by cellular background.

In summary, traditional FTIR has yielded valuable knowledge of aggregate secondary structure by providing information such as the orientation of β strands and the presence of intermolecular β -sheets. FTIR can be applied to samples without isotopic labelling or advanced sample preparation, facilitating its application to cell-derived samples and providing moderate resolution structural data (Figure 1.2). Further, the development of different types of FTIR microscopy has allowed the study of whole cells and tissue samples where simultaneous cellular changes induced by misfolded protein or aggregates can be identified ^[100,101]. However, with all types of FTIR great care should be taken during the data-processing and analysis stages to avoid misinterpretation. Early popularity of FTIR for analyzing aggregate structure abated in part due to problems with interpretation but has increased again with more in depth structural knowledge obtained for a wide variety of aggregates providing a firm base for further investigations ^[44]. It is likely that the active development of hardware improvements will allow the examination of tissues with protein concentrations as low as 0.25 mg/mL, and thus FTIR (micro)spectroscopy may also become a critical tool in early clinical diagnosis of medical conditions, including Alzheimer's and cancer ^[102–104] and advance many studies of cellular protein aggregation.

1.2.4 Solid-state NMR: a whole-structure view

Solid-state nuclear magnetic resonance (ssNMR) is a powerful tool for high-resolution structure determination of cellular protein aggregates. The more conventional solution-state NMR is at best challenging and often impossible for large assemblies such as protein aggregates; even if aggregates can be suspended in solution, their large size causes them to tumble slowly, giving short transverse relaxation times (T_2) and the broadening (or even complete disappearance) of their signal ^[105]. ssNMR

has emerged as an alternative tool that does not require protein solubility or crystallinity. Thus, it has been used to discern the atomic-level structure of proteins in high molecular-weight complexes, both *in vivo* and *in vitro* ^[106]. In the field of protein aggregation, ssNMR has been used to elucidate molecular mechanisms and determine high resolution structures of disease fibrils, IBs and functional aggregates ^[50,107,108]. This section will summarize key advances, methods and prospects of ssNMR for characterizing cellular aggregates, including disease amyloid proteins, prions, IBs, functional amyloids, and secretory granules.

The implementation of fast magic-angle-spinning (MAS) and dynamic nuclear polarization (DNP) techniques were major advances for the application of ssNMR to study biomolecules in a native-like environment ^[109]. MAS was an early breakthrough for ssNMR to deal with direction-dependent interactions that occur in a solid sample. Fast sample rotation during data acquisition serves to average out direction dependent or anisotropic interactions as would naturally occur via molecular tumbling in solution NMR. This spinning decreases line broadening which, while having the potential to provide information on protein dynamics, generally obscures important structural information. However, the small polarization of ¹³C and ¹⁵N in MAS ssNMR was still a challenge causing low sensitivity. The more recently developed technique, DNP, provides a means to improve detection sensitivity by transferring the high polarization of unpaired electrons to the nuclei of interest using matched microwave irradiation ^[105]. DNP and MAS ssNMR can be used together and combined with other advanced NMR techniques, such as higher dimensional experiments, assays for structural restraints, proton (¹H) detection schemes, and non-uniform sampling methods ^[110].

ssNMR has been applied extensively to amyloid proteins associated with diseases, including AD, PD, HD, systemic amyloidosis, and prion diseases ^[107,111]. However, these generally have been synthetic or *in vitro* studies ^[105]. Unfortunately, ssNMR experiments on tissue samples are much less common, because measurements require milligram-scale quantities of isotopically labelled material (Figure 1.2). To overcome this limitation, amyloid aggregates isolated from patient tissue (*ex vivo*) with a disease pathology such as AD, have been used as a source for seeding additional labelled aggregates (Figure 1.4) ^[50]. Tycko and coworkers have performed intensive research in A β fibrils, illustrating the success of this seeding method ^[112–114]. Their ssNMR studies on A β 40 fibrils seeded from various brain regions of Alzheimer's patients revealed that each patient developed a single predominant fibril structure. Moreover, they determined a molecular-level structural model for patient A β 40 fibrils identified

Overall, these findings are crucial for the correlation of fibril structures with disease characteristics and hence may assist the future development of AD therapies and diagnostics. Many other amyloids and cellular aggregates formed by proteins such as TAR DNA-binding protein 43 (TDP-43), immunoglobulin light chain, and prion protein (PrP) have also been amplified for ssNMR studies using seeding ^[116–120]. These studies have been successful in determining residue-specific information on the aggregate structure and describe how broadening of cross-peaks or lack thereof can provide detail on conformational heterogeneity. Technique-wise, they are exemplary of the results that can be obtained from seeding combined with high-resolution characterization by ssNMR. Seeding is a valuable tool to obtain sufficient protein for structural studies while retaining structural characteristics of *in vivo* aggregates, allowing more concrete predictions of protein behavior in disease. However, it is important to consider that less abundant or “unseedable” yet still pathologically relevant aggregate species may be missed during the amplification process.

As IBs are often produced during protein overexpression in recombinant bacteria, the sample amount required for ssNMR studies is more readily achieved. Comparative studies of IBs and fibrils using ssNMR have proven useful in characterizing the composition of IBs for proteins such as the prion-forming domain of HET-s fungal prion (HET-s 218-289), A β peptides, and β 2-microglobulin ^[121–123]. The analysis of HET-s (218–289) revealed that both IBs (raw and partially purified) spectra reproduce all the peaks visible for HET-s (218–289) fibrils assembled *in vitro* ^[121]. In contrast, A β and β 2-microglobulin IBs demonstrated fewer similarities indicating more heterogeneous ensembles with more contributions from non- β -sheet conformations ^[122,123]. While the increased protein content of IB samples is advantageous, the heterogeneity of these aggregates can present a challenge. Often IBs exhibit additional resonances arising from phospholipids from the *E. coli* membrane, other proteins, or RNA. Therefore, it is common practice to perform a partial or total purification of IBs using detergent or buffer washes which may decrease but not eliminate these additional resonances ^[122]. Another approach to reduce the complexity of ssNMR spectra is selective isotopic labelling. An example strategy is the use of specifically ¹³C- and ¹⁵N-labelled amino acids corresponding to unique residue pairs. This strategy has been employed to obtain residue-specific secondary structure information for aggregates formed by a subunit of the influenza virus hemagglutinin protein (FHA2) and the GP41 protein of human immunodeficiency virus (HIV) in whole cells ^[124,125]. Upon comparison to known patterns in chemical shifts for amyloid fibrils and their respective soluble forms, both studies concluded that the protein of interest retained large amounts of native structure in IBs.

Different labelling schemes that help decrease background signals and isolate individual protein resonances can also be applied to seeded or purified protein (Figure 1.4) to resolve ambiguities for residues of interest ^[111,126]. Whereas uniformly labelled samples are required for sequence-specific resonance assignments, specifically placed labelled residues give well resolved spectra that can be used to determine long-range intermolecular distance restraints ^[127,128]. An example is the ssNMR study of the curli protein, a functional amyloid found in the extracellular matrix of *E. coli* involved in biofilm formation, host adhesion and invasion ^[129]. Schubeis *et al.* employed segmental labelling of repeat sequences in curli to obtain residue-specific assignments and information about the arrangement of the repeats. The resultant spectra of native curli extracted from biofilm and the lab-expressed protein had high similarity, revealing a well-defined β -solenoid structure. Interestingly, this structure has also been identified by ssNMR in human functional amyloids (secretory granules) of the endocrine system ^[51]. Residue-specific labelling has also been applied to the human functional amyloids of the glucagon hormone. Using sets of scattered labels, Gelenter and coworkers demonstrated an unexpected antiparallel hydrogen bonding pattern throughout glucagon fibrils ^[120]. All in all, non-uniform isotopic labelling methods for larger complexes can aid in the determination of previously undefined structures and interactions within the complex.

In conclusion, there are currently many ways to acquire valuable information about cellular aggregates using ssNMR, either from a bacterial expression system ^[130], whole cells ^[124,125] or from diseased tissue ^[112]. Although more laborious sample preparation such as uniform or selective isotopic labelling (Figure 1.2), there are many benefits of ssNMR studies. Unlike other methods reviewed here, ssNMR can provide information on amide acid side chains, conformational heterogeneity through cross-peak linewidths, and the fibril core can be identified at a residue-specific level, giving essential information about aggregation mechanisms and crucial relationships between toxicity and structure. To date, relatively few ssNMR studies have been reported for cell-derived aggregates, thus this is a likely fruitful area for further research.

1.2.5 qHDX NMR - Residue specific view

Quenched hydrogen deuterium exchange (qHDX) with solution NMR readout is a powerful but also little-used method to obtain high resolution structural information on large cellular complexes or aggregates. Traditional HDX provides data on native protein structure and dynamics by measuring the extent of protection for individual protein amides (NHs) in a protein structure against exchange with

solvent (D₂O) over time. However, this method cannot be performed on large assemblies directly as they are not amenable to NMR measurements; instead, protection may be measured using an adapted qHDX method.

Upon suspending aggregate samples in D₂O, NHs that are inaccessible to solvent are protected against exchange, while solvent accessible amides are converted to ND. Exchange is then quenched by flash freezing and the assembly is dissolved in dimethyl sulfoxide (DMSO), a solvent that slows exchange rates by ~four orders of magnitude compared to aqueous solutions ^[131]. DMSO also effectively denatures the protein by lowering the free energy of the unfolded state relative to the native state ^[132]. Thus, the originally too-large for traditional ¹⁵N-based solution NMR assembly becomes an amenable size; the majority of residues may be analyzable for protein subunits smaller than ~15 kDa. As deuterium is not observed in a ¹H-¹⁵N heteronuclear single quantum correlation (HSQC) spectrum, only NHs that were buried or H-bond protected in the assembly, and therefore did not exchange with D₂O, will be observed. Analysis of HSQC peaks reveals which specific residues were involved in protective structure and which were exposed, providing a high-resolution analysis of the protein assembly. The observed protection pattern reports on the average structure of all populated protein conformations in the aggregate. Extensive background on qHDX was described by Hoshino *et al.* ^[133] and methodological details for samples of pure proteins aggregated *in vitro* were described by Alexandrescu ^[134].

This method was developed independently by the Alexandrescu ^[47] and Goto ^[46] groups in the early 2000s, and it was initially used to determine structural details of *in vitro* amyloid. Alexandrescu applied qHDX to cold shock protein A (CspA) fibrils that were made through acid-denaturing the purified protein. His work elegantly showed that the entirety of the protein was structured in the fibrils, despite their unfolded origins. Soon after, the Goto group reported mapping the core of β 2-microglobulin amyloid fibrils. Similar to CspA, all amides, even those natively found in loop regions, showed strong protection against exchange, indicating the whole protein was involved in the fibril structure. In addition, the protection was compared to that of purified native β 2-microglobulin and found to be remarkably similar in the core, but varied in the loops and termini, suggesting an extensive hydrogen bonding network underlying the rigidity and structural stability of the fibrils. Further application of qHDX to amyloid- β fibrils made from purified protein segments showed, notably, that increasing the length of the aggregating fragment by as little as two residues drastically changed the fibril structure ^[135–137].

Some years later, the application of qHDX to cellular IBs was reported by Wang *et al.* in 2008 ^[48]. The cellular IBs were washed several times during their preparation to make largely pure samples. In a

remarkable contrast to the preceding studies of *in vitro* aggregates, Wang and coworkers observed protection of only a small proportion of amides in IBs formed by several structurally diverse proteins. Based on the qHDX and complementary experiments, they inferred that the proteins formed small stretches of amyloid and all other residues were unprotected. This result - the first qHDX for an *in vivo* aggregate - showed that cellular aggregation may be very different from *in vitro* aggregation. This was furthered by their 2010 study, where one protein was aggregated in different ways (heat, concentration, acid, amyloid fibrils, and IB) and its structure was analyzed ^[138]. The IB did not resemble the others, further highlighting the importance of *in vivo* studies. Strikingly, the IB resembled the soluble protein more closely than the other aggregates. qHDX studies on washed amyloid- β IBs again showed a new view by revealing the presence of a new, α -helical, aggregation-prone species ^[122].

qHDX has not yet been applied to visualize aggregation in whole cells and doing so would need to overcome challenges in detecting, identifying and assigning signals from the aggregated protein of interest. Additionally, parsing the contribution of soluble species of the same protein in the whole cells presents another challenge. On the other hand, the use of qHDX to investigate cell-derived aggregates has many advantages. Because aggregation-prone proteins tend to accumulate in large amounts in cells, relatively small quantities of cell culture can easily be used to generate many NMR samples ^[139]. Combined with the lack of purification minimizing protein loss, this greatly decreases the time and cost for these experiments. Furthermore, the simplicity of the protein preparation and the use of a standard HSQC with minimal setup means that little expertise or specialized equipment is required to obtain significant information on the extent of protection. For more detailed analyses, backbone amide resonance assignment strategies have been documented in detail for easy application, with Tarasca *et al.* demonstrating clearly that point mutants show nearly identical unfolded spectra, and that any peaks shifting due to water, pH, or mutation can be easily tracked ^[139]. The low barrier to entry is a significant advantage of qHDX, while residue-specific analysis provides a high-resolution view of cellular aggregation that provides a valuable complementary view to results obtained with other methods (Figure 1.2).

1.2.6 qHDX MS: a segment-specific view

An alternative readout to NMR for qHDX experiments is mass spectrometry (MS), which provides sensitive and precise information on the mass of proteins and peptides. We note that HDX readout by MS or NMR generally provide complementary information, as MS resolves *populations* that have

different net amide protection for a given protein or peptide, while NMR measures the *average* total protection of individual amides for the entire population ^[141]. In qHDX-MS, exchange in D₂O is often quenched by decreasing pH, followed by pepsin digestion and MS data acquisition. The extent of deuterium uptake with time can be monitored by MS based on how the mass of each fragment varies; this information can be mapped onto the associated segments of the protein and used to deduce aggregate structure (Figure 1.5). This method has been applied extensively to investigate the structure and dynamics of purified proteins; thus there are many existing reviews on its uses ^[142] and well developed recommendations on its implementation ^[143–145].

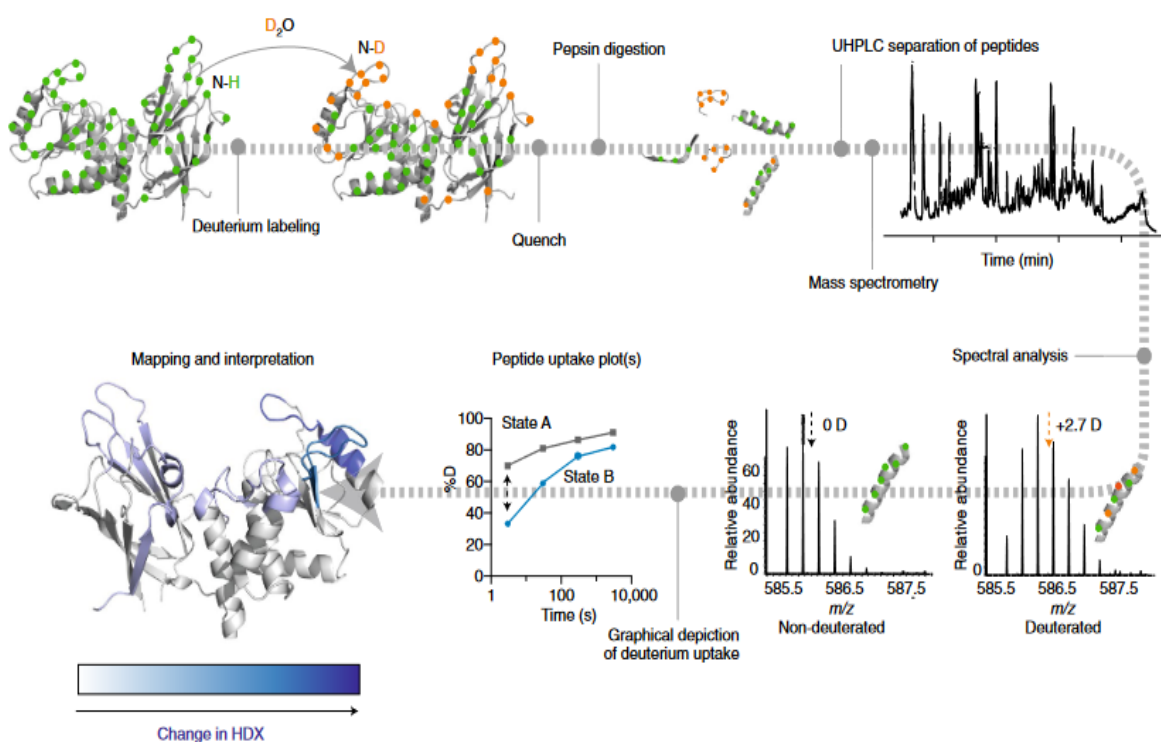


Figure 1.5 Overview of a qHDX-MS workflow.

Figure reproduced from Masson et al. with permission ^[143].

Although qHDX-MS is well established, to date it has been little applied to protein aggregates. Furthermore, the studies that have been completed often use purified aggregates to limit the number of additional species in the spectrum that may cause difficulties with assigning fragments of the protein of interest. Interestingly, a recent group of studies on impure PrP structures has provided evidence linking protein structure, infectivity, and function. First, a study on mouse derived PrP mutant aggregates (PrP^{Sc}) revealed evidence for strains - that is, prion proteins with distinct disease phenotypes and

distinct structures ^[146]. qHDX-MS showed region-specific differences in deuterium incorporation between three mutants, discerning their different aggregate structures. Additional comparison to amyloid formed by pure PrP also showed clear differences, indicating yet another structure and perhaps an explanation for different infectivities of the aggregates. Further studies on synthetic prion proteins showed differences in deuterium incorporation, distinctive in some helical portions of the proteins ^[147]. These studies also revealed a potential drawback of qHDX-MS when a fragment of the C-terminus exchanged very rapidly and then plateaued. This indicates that part of the segment is completely unordered/unprotected, whereas the rest is highly protected - however, it is impossible with this method to tell which residues show which behavior. Thus, while region- specific structural information can be obtained by qHDX-MS, residue-specific interpretation may be limited.

Similar to backbone amides, the imidazole protons of histidine residues are susceptible to exchange with solvent. Therefore, an avenue to obtain residue-specific information is through histidine HDX-MS. By exploiting the differing pH dependence of amide and imidazole exchange reactions, one can distinguish between their protection in an aggregate structure. A study on the human PrP used the combined results of qHDX of imidazole proteins and backbone amides to evaluate the structural differences between recombinant protein and protein seeded with sporadic Creutzfeldt-Jakob disease (sCJD) prions ^[148]. These results provided a good basis for using seeded protein as a more accurate model of disease conditions and highlighted possible differences in *in vivo* aggregation mechanisms.

More recently, Kaldmae and coworkers have attempted to expand the scope of qHDX-MS with the novel implementation of a cellular approach ^[149]. In this process, cells overexpressing the protein of interest (the N-terminal domain of major ampullate spidroin 1, NT) were incubated directly in a deuterated buffer before quenching, lysing, and performing MS. Deuteration for different amounts of time showed a clear time-dependent mass increase with time which could be localized to distinct segments of the protein through electron-transfer dissociation sequencing. These results demonstrated that despite extensive aggregation-prone regions in NT, the protein under cytosolic conditions adopts a very tight fold rather than amyloid. This approach was then taken one step further, to test the conformational stability of cellular NT. In-cell deuteration of an A β -NT fusion followed by MS detection revealed fibrillar aggregates. However, these fibrils largely did not incorporate NT, indicating it remains in a stable fold under cellular conditions. This phenomenal study clearly highlights the importance of studying protein folding and aggregation *in vivo*, as purified protein did not behave the same way as the in-cell aggregates. Furthermore, in this scenario qHDX-MS was able to provide crucial structural and

mechanistic information that would have otherwise remained obscure. While this is just the first example of an in-cell qHDX-MS study of aggregating protein, future developments to ion separation techniques may enable the separation of species based not only on mass and charge but also on shape. Further applications of these exciting qHDX-MS methods are a promising avenue for resolving specific aggregate conformations in living cells ^[142].

Overall, detection of hydrogen exchange via MS provides valuable ways to obtain high-resolution information on aggregate structures. Recent advances have allowed for a more in-depth study of in-cell aggregation mechanisms with the advantage that samples do not need to be labelled or purified. Additionally, the amount of sample required for MS is minuscule and the experiment time is short, allowing for technical and biological replicates to be performed more easily than for NMR methods ^[143]. Since *in vivo* conditions are less controlled than *in vitro* experiments, the ability to perform replicates is crucial. However, it should be noted that the presence of increased cellular contaminants will result in a more complex mass spectrum. Therefore, it is important that the protein of interest is present in a reasonably high concentration and/or a purified protein is available as a comparison during fragment assignment (Figure 1.2).

1.3 Conclusions

In conclusion, the determination of *in vivo* aggregate structures and mechanisms of aggregation is of central importance to biotechnology and medicine. Amyloid and many other aggregate structures are present in the protein deposits of many maladies, such as Alzheimer's and Parkinson's disease, ALS, and cancer. Amyloid fibrils have also been identified in functional aggregates in the *E. coli* extracellular matrix involved in biofilm formation and secretory granules of the human endocrine system that work as long-term deposits of hormone units ^[24,127]. Furthermore, it has been shown that bacterial IBs are composed of functional native and native-like protein structures along with amyloid or aggregated forms of the same protein. This molecular architecture allows for retention of biological activity as well as high mechanical stability, showing promise for their use as functional materials ^[24,49]. Consequently, IBs are being explored for applications as drug delivery systems and self-immobilized catalysts. IBs formed by enzymes, such as reductases, kinases, lipases and synthases, have the advantage of catalyzing reactions while being reusable and recyclable, resistant to proteases, and relatively easy to purify ^[24,150–152]. In all these occurrences of aggregation, structural and mechanistic knowledge would be an asset,

contributing to the development of disease treatment, biotherapeutics and catalysts in industrial processes, for example ^[153–155].

As specifically highlighted in each section, the techniques reviewed here have all advanced to provide crucial structural information on cell-derived and in-cell protein aggregates. Although each has its advantages as a standalone technique, it is common practice to employ multiple techniques to gain an in-depth characterization of protein aggregates. Some methods, such as conformation specific Abs, capture the surface properties of the aggregate, while others, like FTIR, acquire signals from all components of the aggregate. Some techniques, such as cryoEM and ssNMR, report mainly on ordered portions of the structure. However, unlike cryoEM, ssNMR can provide indirect information on residues exhibiting structural polymorphisms through peak-broadening. Finally, qHDX with detection by either NMR or MS provides measurement of the aggregate ensemble at atomic resolution. To date, only a small number of investigations of cellular aggregate structures have used methods such as qHDX-MS and NMR. However, recent advances in these techniques have provided exciting prospects for gaining additional groundbreaking insights on cell-derived and *in vivo* aggregates. Similar to qHDX, various complementary MS and NMR methods to study aggregated protein *in vitro* may be adaptable to study cellular or cell-derived aggregates, in a variety of cellular contexts, such as the cytoplasm, cellular compartments or on membranes, and are promising areas for future research ^[141,156–160]. Our critical analysis of methodologies with consideration of sample characteristics such as purity, feasibility of isotopic labelling, concentration or amount, protein size and level of expertise may serve as a valuable resource for further advancing high resolution knowledge of cellular aggregate structures.

While the aforementioned methods have enabled much progress, overcoming previous limitations such as aggregate size, sample heterogeneity and insolubility, there are still obstacles to further advances. For the goal of determining aggregation mechanisms and structures in living cells, there is the complication of selectively observing the protein of interest. This means signals from contaminating proteins must be removed or otherwise taken into account. This may be accomplished by analyzing cell-derived aggregates *ex vivo* and can be further aided by sample processing such as washing with buffer or selective solubilization using detergents. Unfortunately, the effects on aggregate structures of such processing and other sample preparation requirements are variable and often unknown. It is essential to consider the consequences of such experimental variables in the application and further development of the structure analysis tools described herein. The current powerful suite of structure analysis techniques, and likely continued developments thereof, are applicable to a great variety of systems.

These methods provide a strong foundation for exciting further progress in understanding and ultimately controlling protein aggregation in medical and biotechnology settings.

Chapter 2: Quantifying levels of IB formation in a SOD1 model system

2.0 Preface

The work in the following chapter aims to set the stage for studying aggregate structure by first quantifying aggregate formation in a model system (SOD1) by SDS-PAGE. By correlating this to aggregation predictors and other biophysical data known for SOD1, the goal is to establish a framework for the aggregation-prone regions of SOD1 and to pinpoint a smaller subset of mutants for structural studies.

Experiments were primarily performed by Dalia Naser, with a subset being performed by Susan Kelso and Hilary Simon. Data analysis and interpretation was performed by Dalia Naser. The raw aggregation propensity values reported in this Chapter, as well as the correlations to melting temperature and free energy change of dimer dissociation, were previously published in the publication used for Chapter 4. The rest of the data and interpretations are unpublished.

2.1 Introduction

2.1.1 Studying Inclusion Bodies (IBs)

Inclusion bodies (IBs) are insoluble intracellular aggregates formed upon overexpression of heterologous protein in a foreign host^[161–163]. Due to their large size and insolubility, their mechanism of formation and structure are not well known. However, interest in IBs has been expanding over recent years. Historically, IBs were seen as a roadblock in protein production^[164,165], and earliest studies on IB formation were aimed at predicting IB structure/formation to *avoid* IB formation^[162,163]. As understanding of IBs grew, there became an interest in them as reservoirs of pure protein, with many studies suggesting that the aggregate contains active protein, and therefore natively folded, protein^[166–170]. Concurrently, other work showed that “soluble” heterologous protein was not synonymous with “well folded” or “functional” protein, but that it may contain soluble aggregates with varying levels of function^[169,171–173]. This shifted the mindset around IBs from a waste product to be avoided to a potentially useful cellular product that may be exploited for protein production and purification, as a functional biocatalyst with minimal purification/refolding^[174]. Consequently, interest in understanding IB structure and formation mechanisms grew.

Apart from biotechnological functions, IBs have also been viewed as a useful model to study aggregation as a cellular stress response^[175]. Many conditions such as heat shock, aging, mutation, or rapid protein production can cause a cell to be stressed which in turn can trigger cellular quality control responses.

Evidence has shown that stress response can occur simultaneously with protein aggregation^[175]. Tyedmers et al (2010) described that protein aggregation is a conserved process across different cell types, suggesting a universal response across bacterial, yeast, and mammalian cells^[176]. This observation indicates that knowledge gained from studying mechanisms responsible for bacterial IB formation can be applied to other, more complicated cell types.

Many complex human diseases are associated with protein aggregation. These diseases include Huntington's and Alzheimer's diseases and amyotrophic lateral sclerosis (ALS). There are several examples of bacterial IBs have provided insights towards the pathology of these diseases, many of which were discussed in Chapter 1 ^[175]. In this way, bacterial IBs provide a tractable model to study proteins that aggregate in disease.

2.1.2 Cu, Zn Superoxide Dismutase

SOD1 Structure and Function

The model system chosen for this thesis is Cu, Zn Superoxide Dismutase (SOD1). SOD1 is a homodimeric metalloenzyme found both in the cytoplasm and the mitochondria of cells.^[177] As an extremely well-studied aggregating protein with tangible maturation states, multiple disease-associated mutations, and the ability to be expressed in multiple cell types (for example, yeast and bacteria), SOD1 presents itself as an ideal model system for studying aggregation. In healthy cells, the enzyme exists as a homodimer with each monomer forming a Greek key barrel. Each monomer comprises 153 amino acid residues (Figure 2.1) with a conserved intramolecular disulphide bond between C57 and C146. Two metal ions are bound per monomer: a zinc ion and a copper ion, whose major roles are structural and catalytic, respectively. The zinc ion is coordinated in a structural loop known as the zinc-binding loop by four residues: H63, H71, H80 and D83. The copper ion shares the His63 ligand, and is also ligated by H46, H48, and H120. Residues 124-139 form a second functional loop known as the electrostatic loop. This loop aids in guiding superoxide molecules towards the catalytic copper site.

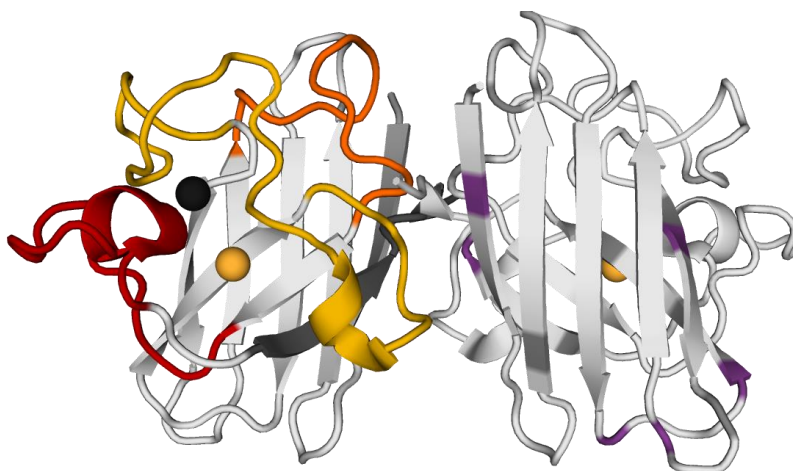


Figure 2.1 The Structure of mature Holo Cu, Zn SOD1

The protein is shown with bound copper (orange) and zinc (black) ions. Major structural features are shown on the left subunit as follows: N-terminus is in light grey, metal-binding loop is shown in red, loop 6 is shown in orange, electrostatic loop is shown in yellow, and the C-terminus is shown in dark grey. Mutations studied in this chapter are shown on the right subunit in dark grey or purple. Mutations shown in purple are used in studies in later chapters in this thesis. PDB 1HL5.

SOD1 is a ubiquitous enzyme which scavenges superoxide radicals ($O_2^{\cdot-}$)^[177,178]. Superoxide is prevalent in the cytoplasm, as it is a common by-product of standard cellular processes. However, high concentrations of superoxide are extremely toxic to cells^[177]. SOD1 catalyzes the dismutation of superoxide to produce molecular oxygen and hydrogen peroxide in a two-step “ping-pong” mechanism (Figure 2.2)^[178]. In the first step, the cupric ion in the active site is reduced by a superoxide molecule, resulting in dioxygen and a cuprous ion. This cuprous ion is reoxidized by a second superoxide molecule, to produce hydrogen peroxide and return the copper to its original oxidation state. Thus, each cycle of the enzyme removes two superoxide radicals^[178].

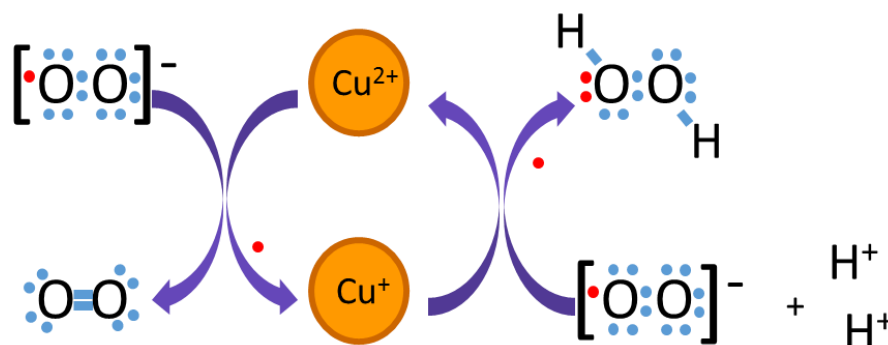


Figure 2.2. Mechanism for SOD1

The copper ion “ping-pongs” between the two possible dismutase reactions. The first reaction is shown on the left, with the Cu^{2+} ion being reduced by a superoxide molecule, producing dioxygen. The Cu^+ ion is then re-oxidized by another superoxide molecule, producing hydrogen peroxide (reaction on the right). Since the copper has been oxidized, it may now repeat the cycle. Adapted from Chakrabartty & Rakhit, 2006.^[178]

SOD1, Mutations, and ALS

Interest in SOD1 has been strongly motivated by its association with amyotrophic lateral sclerosis (ALS). More specifically, there are over 180 mutations in SOD1 at the protein-sequence level that have been associated with familial ALS (fALS). Disease-associated mutations are found throughout the entire protein structure^[177]. They are located in regular secondary structure elements, as well as in functionally important sites (such as the metal binding sites or dimer interface). The mutations encompass a myriad of changes to the protein, with residues changing in hydrophobicity, charge, size, and side-chain orientation.^[179,180] Despite this variation in mutation type, the mutants tend to be found as aggregates in SOD1 fALS patients.^[181] Interestingly, the stability of mutant protein or its propensity to aggregate does not necessarily correlate to disease duration. A marked example of this is the V148I mutant, associated with a heartbreakingly short disease duration of ~1.7 years – yet V148I is more stable than WT, with a similar aggregation propensity^[182].

SOD1 Maturation and Maturation States

The binding of metals, formation of an intrasubunit disulphide bond, and dimerization of SOD1 are all post-translational modifications that must occur for the protein to achieve its mature native state (Figure 2.3)^[180]. The most immature form of SOD1 is the reduced apo form (apoSH) which has no metals bound and a reduced disulphide^[180]. The reduction of the disulphide bond greatly increases dimer dissociation,^[183] meaning that this form of the protein tends to be monomeric. When using melting temperature (T_m , the temperature at which half the protein is unfolded) as a measure of stability, the apoSH form is only marginally stable, with differential scanning calorimetry (DSC) studies on WT apoSH showing a large decrease in T_m relative to that observed in the fully mature form (the oxidized holo form, or holoSS)- the apoSH form exhibits T_m of 47.6 °C, whereas the holoSS form displays a significantly higher T_m of 92.7°C.^[184,185] This difference in relative stabilities of mature and immature SOD1 is seen for mutant SOD1s as well. The high stability of mature SOD1 is owing to the metal binding, disulphide formation and dimerization. Thus, an understanding of the order of the post-translational modifications and the mechanisms by which they occur is a key factor in understanding SOD1 stability. While there is still some ambiguity in the details of the post-translational maturation in-cell, it has been said that the steps occur sequentially^[186]. The addition of copper is the most well studied step in the process and has helped shed light on the order that the maturation steps occur.

SOD1 maturation has been studied extensively, with several models proposed. Figure 2.3 shows one possible model of the maturation steps, although it should be noted that different studies propose that

the steps can occur concurrently. It is thought that the addition of the zinc ion occurs first, as zinc is highly stabilizing to the SOD1 monomer^[187,188]. However, the mechanism by which SOD1 obtains zinc has not yet been determined. As the intracellular free zinc concentration is generally low and tightly regulated, it has been proposed that a metallothionein is involved in zinc delivery^[178]. It is postulated that the addition of copper by the copper chaperone for SOD1 (CCS) follows, following the formation of the disulphide bond (which the CCS facilitates). The various possible and preferred maturation states of SOD1 are indicated in Figure 2.3 below.

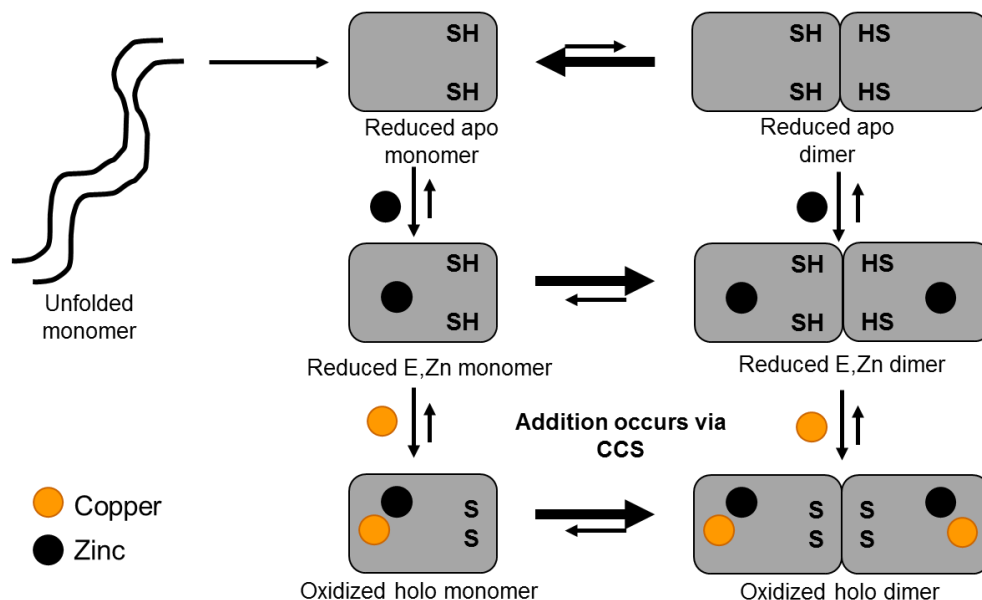


Figure 2.3 Maturation States of SOD1

Following the unfolded state, the reduced apo monomer (apoSH, top left) is the most immature form of folded protein. The addition of zinc is thought to occur first and stabilize the monomers enough for dimerization to occur (resulting in the reduced zinc-bound dimer, E,ZnSH). Copper binding and oxidation are thought to follow via interaction with the CCS, resulting in the holo dimeric form of the protein. Figure courtesy of Harmeen K Deol (Meiering group).

2.2 Materials and Methods

2.2.1 Aggregation propensity prediction

The servers TANGO (sequence-based)^[189], ZipperDB (sequence-based)^[190], AmylPred2, Kyte-Doolittle hydropathy index, and CamSol (sequence- and structure-based)^[191] were used to predict aggregation propensities for the WT SOD1 sequence. The sequence-based servers require only a primary amino acid sequence input, while CamSol uses a PDB file (PDB: 2GBU Chain A) for 3D structure-based prediction. Each residue in the sequence was given a binary ranking (1 is aggregation-prone, 0 if not) and these were summed for the sum binary aggregation score. For Kyte-Doolittle hydropathy index, a residue was scored as “aggregation-prone” if it was predicted to be hydrophobic, whereas other predictors were scored as “aggregation-prone” based on cutoffs given by the respective predictors. Additionally, each

mutant SOD1 sequence was run through the ZipperDB server to determine Rosetta energies. The WT sequence was also analyzed through the ConSurf server, which is a tool for determining the evolutionary conservation of an amino acid. Lastly, a multiple sequence alignment was performed on the WT SOD1 sequence using PredictProtein to determine preferred amino acids at each position.

2.2.2 Mutant SOD1 Expression System for IB Formation Studies

The protocol for measuring mutant SOD1 IB formation was adapted from *Trainor et al.*^[192,193] In this case, SOD1 mutants are in the WT background (e.g. no mutations at Cys-6 and Cys-111). SOD1 variants were expressed in BL21 DE(3) *Escherichia coli* (E. coli) cells with a pLys plasmid to encode for the production of T7 polymerase (an inhibitor of RNA polymerase) and chloramphenicol (Chlor) resistance, and a pET21 vector to code for β -lactamase and provide ampicillin (Amp) resistance. The composition of various buffers for the experiments are given in Table 2.1 below.

2.2.3 Expression, Sample Preparation, and Solubility Analysis

Glycerol stocks of BL21 (DE3) pLysS cells for expressing SOD1 were used to inoculate 10 mL of LB medium supplemented with 100 μ g/mL Amp and 30 μ g/mL Chlor. Cultures were grown overnight at 37 °C with shaking (200rpm). After overnight growth, two 1mL-aliquots of each overnight LB culture were removed and centrifuged at 8000 *g* at room temperature (RT) for 1 minute, followed by removal of 750 μ L of the supernatant. The pellet was resuspended in the remaining 250 μ L to make the inoculum. This inoculum was added to 50 mL of sterile 1xM9 media in 125 mL flasks following the addition of M9 supplements and antibiotics. For zinc-supplemented experiments, ZnSO₄ was added at this step to a final concentration of 1 mM. Inoculated the flasks were then placed in the shaker at 37 °C, 200 rpm, until they reached an optical density at 600 nm (OD₆₀₀) of ~0.5-0.8. Protein production was then induced by the addition of isopropyl β -D-1-thiogalactopyranoside (IPTG, 0.1 mM for 37 °C trials, 0.5 mM for the 25 °C and 30 °C trials). Incubation continued at the desired temperature with shaking at 200 rpm. Triplicate 1-mL samples were taking at 2-, 4-, and 6-hours post-induction (T₂, T₄, and T₆). One sample was used to read the OD₆₀₀, while the others were prepared as follows.

The 1-mL samples were centrifuged at 8000 *g* for 1 minute at RT. The supernatant was removed, and the pellet was resuspended in 30 μ L of TEN buffer. Of this, 10 μ L was removed as a whole cell fraction. An additional 10 μ L of TEN buffer was then added to each sample, and the samples were stored at -20 °C until time of analysis, at which point they were thawed to room temperature before undergoing cell lysis. The cell lysis process occurred as follows: samples were flash frozen in liquid N₂ and thawed for 5

minutes in a 37 °C water bath (a freeze-thaw cycle). A 40-minute incubation at 4 °C with 100 µg/mL DNase1 followed. Lysis was completed with three more freeze-thaw cycles. The soluble and insoluble fractions were separated by centrifugation (12000 g, 10 minutes, RT) and analyzed with SDS-PAGE (15% resolving gel). Gel densitometry was performed on the SOD1 bands in the pellet and supernatant fractions using the Totallab software. The %insoluble SOD1 protein determined was the ratio of the pellet band over the sum of both the pellet and supernatant band using the following equation:

$$\% \text{ insoluble} = \frac{\text{insoluble}}{\text{soluble} + \text{insoluble}} \times 100\% \quad \text{Eq. 2.1}$$

Experiments were conducted once at 37 °C (+/- zinc) and in triplicate at both 25 °C and 30 °C (also +/- zinc).

Table 2. 1 Composition of Buffer and Solutions Used

Media	Composition
2X Loading Buffer	0.125M Tris-HCl pH 6.8, 4% SDS, 20% Glycerol, 10% 2-mercaptoethanol, 0.03 mM bromophenol blue
Coomassie Brilliant Blue Stain	62.5%, 30% methanol, 7.5% acetic acid with 1g Coomassie R per liter
DNase I	2 mg/mL in storage buffer (10 mM Tris pH 7.5, 10 mM CaCl ₂ , 10 mM MgCl ₂ , 50% glycerol)
Luria Broth	20 g of LB (Lennox) per liter, where LB is 20 g bacto-tryptone, 5g bacto yeast extract, 10g NaCl
M9 supplements	250 µL 100 mg/mL NH ₄ Cl, 100 µL 1M MgSO ₄ , 1 mL 20% glucose, 50 µL 0.05% (w/v) thiamine, 50 µL 0.1 M CaCl ₂ , 50 µL 100 mg/mL Amp and 50 µL 30 mg/mL Chlor
10X M9 salts	67.8 g Na ₂ HPO ₄ [or 128g Na ₂ HPO ₄ ·7H ₂ O], 30 g KH ₂ PO ₄ , 5g NaCl per liter
SDS-PAGE destain	62.5%, 30% methanol, 7.5% acetic acid
TEN Buffer	20 mM Tris, 1 mM EDTA, 100 mM NaCl, pH 8.0

2.3 Results and discussion

2.3.1 Rationale for selected mutations and predicted behaviour

In this chapter, 18 fALS mutants are characterized in the wild-type SOD1 background. These mutants are dispersed throughout the SOD1 primary sequence and are involved in different secondary structure. Two of the selected mutants (H46R and G85R) are known to disrupt metal binding in different ways ^[177], while four others (A4V, I113T, V148G and V148I) are known to impact the dimerization of SOD1. Additionally, the mutations also change biophysical parameters such as charge and hydrophobicity and have been shown to have different structural stabilities. Lastly, they have varied fALS disease characteristics. Thus, these diverse mutants encompass a variety of features in SOD1 and can be used to decipher some of the determinants of SOD1's aggregation propensity. A summary of the selected variants and their characteristics is shown in Table 2.2.

Table 2. 2 fALS point mutation locations and biophysical parameters

Mutant	Location	Structural impact	ΔG (dimer) kcal/mol ^[194]	ΔG (monomer) kcal/mol ^[194]	apoSH T_m (°C) ^[184]	apoSS T_m (°C) ^[195]	HoloSS T_m (°C) ^[185]
WT	N/A	N/A	10.3	2.8	47.6	59.1	92.7
A4V	$\beta 1$	Dimer interface	6.4	2.3	36.3	50.7	86.7
G16S	$\beta 2$	---	---	---	---	---	---
G37R	Loop3	β -barrel plug	7.6	3.2	33.5	50.1	
G41D	Loop3	β -barrel plug	---	---	---	45.2	86
G41S	Loop3	β -barrel plug	---	---	---		84.4
H43R	$\beta 4$	---	7.5	1.7	35.4	48.1	86.3
H46R	$\beta 4$	Blocks copper site	8.4	5.3	---	---	---
L84V	$\beta 5$	---		---	52.6	55.6	---
G85R	$\beta 5$	Blocks zinc site	---	---	40.7	54.7	77.5
D90A	Loop5	β -barrel plug	---	---	---		77.5
G93A	Loop5	β -barrel plug	7.2	1.7		47.9	87.7
G93D	Loop5	β -barrel plug	---	---	---	45.6	85.1
E100G	$\beta 6$	---	8.0	1.6	33.2	51.2	86.2
L106V	Loop6	---	---	---	---	---	---
I113T	Loop6	Dimer interface	7.1	1.4		47.1	88.2
L144F	Loop7	---	---	---	---	---	---
V148G	$\beta 8$	Dimer interface	5.9	2.2	34	49.3	86.9
V148I	$\beta 8$	Dimer interface	8.9	5.7	51	60.5	92.7

Due to the prevalence of protein aggregation in disease and biotechnology, much work has been devoted to aggregation predictors. These predictors range from simple algorithms that use only primary sequence information to estimate a biophysical characteristic of the sequence (for example, the Chiti-Dobson method of calculating the weighted sum change of hydrophobicity, net charge, and secondary structure propensity upon mutation of a residue), sliding window algorithms based on the properties of a stretch of sequence (for example, ZipperDB assesses stretches of 6 amino acid residues for their amyloid-forming propensity), or more complex, structure-based algorithms that utilize a crystal or NMR structure for 3D structural information in addition to sequence information^[191].

The WT SOD1 sequence was run through 5 common aggregation predictors to identify aggregation hotspots. These predictors were: CamSol^[191], Tango^[189], ZipperDB^[190], AmylPred2^[196], and the Kyte-Doolittle hydrophathy index^[197]. CamSol predicts a residue's intrinsic solubility based on physiochemical properties of the amino acids with removal of bias towards amyloid. Tango predicts aggregation nucleating regions based on physiochemical principles of secondary structure propensity extended by the assumption that core regions in an aggregate are fully buried. ZipperDB compares a sequence to an amyloid-forming hexapeptide and evaluates its subsequent fibril-forming propensity, as well as comparing the quality of the resulting steric zipper to other amyloid peptides. AmylPred2 is a consensus of different methods developed specifically to predict amyloid formation. Lastly, the Kyte-Doolittle hydrophathy index is a simple calculation of the hydrophobicity (and therefore likelihood of burial) of each amino acid in a sequence. A binary (yes/no predicted aggregation) approach was taken to combine the output of the different algorithms. The sum is plotted in Figure 2.4.

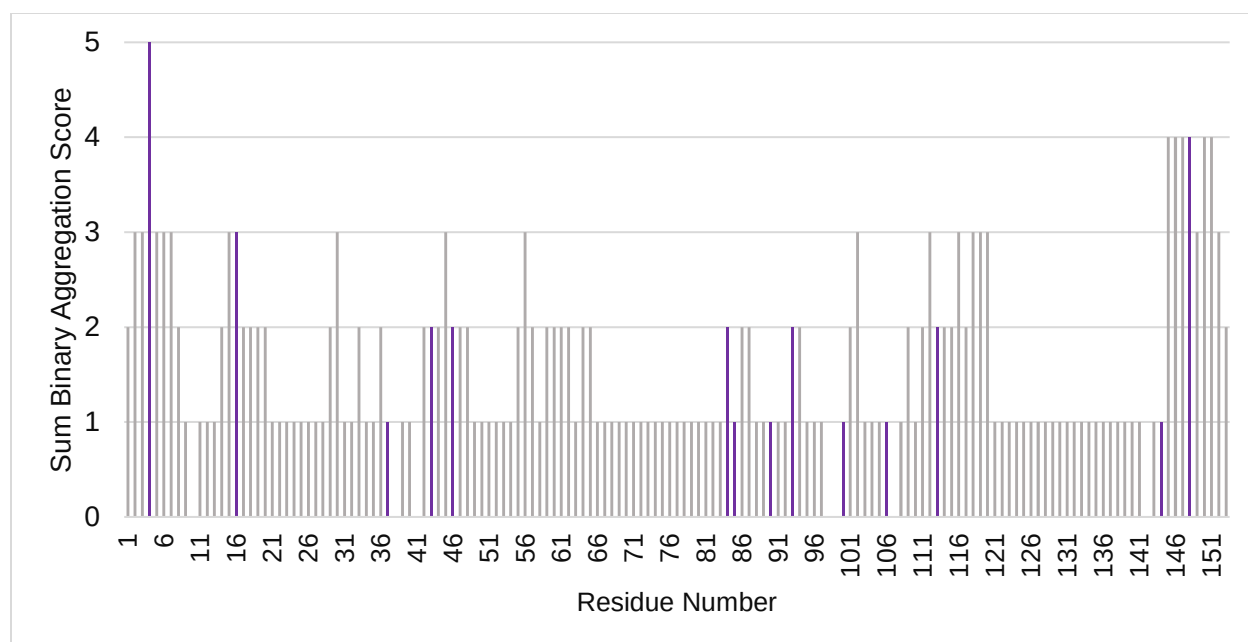


Figure 2.4 WT Sequence Aggregation Propensity

The WT SOD1 sequence was analyzed by Tango, Camsol, Kyte-Doolittle hydropathy index, AmelPred2, and ZipperDB and each residue was given a binary score of 1 or 0 depending on whether it was predicted to be aggregation-prone. Amino acids that are mutated and explored in this chapter are colored in purple.

The results of the aggregation predictors for WT SOD1 describe which areas in the protein may be particularly aggregation prone. Mutations in these areas may prevent or promote aggregation further. Based on these simple sequence predictor methods, not all the selected mutants are in aggregation-prone regions in the WT protein. For example, residue 41 does not appear to be aggregation-prone, and the selected residues flanking it (G37, H43, and H46) do not appear particularly aggregation prone either. The N- and C- termini appear to be the most important for aggregation and contain A4 and V148 mutations that are included in this study.

These aggregation predictors were also applied to mutant SOD1 sequences. Application of ZipperDB to the mutants reveals several regions which become more aggregation prone upon mutation (Figure 2.5). ZipperDB results were particularly noted as IBs are commonly thought to contain amyloid structure, and ZipperDB specifically predicts the likelihood of amyloid formation. The A4V mutation increases aggregation in the N-terminus, G16S greatly increases aggregation in Loop 1, and G41S makes the previously innocuous stretch of surrounding sequence incredibly aggregation-prone. The L84V, G93A, E100G, L106V, L144F and V148I mutations follow suit. Interestingly, the remaining mutants are predicted to decrease aggregation propensity relative to the WT variant.

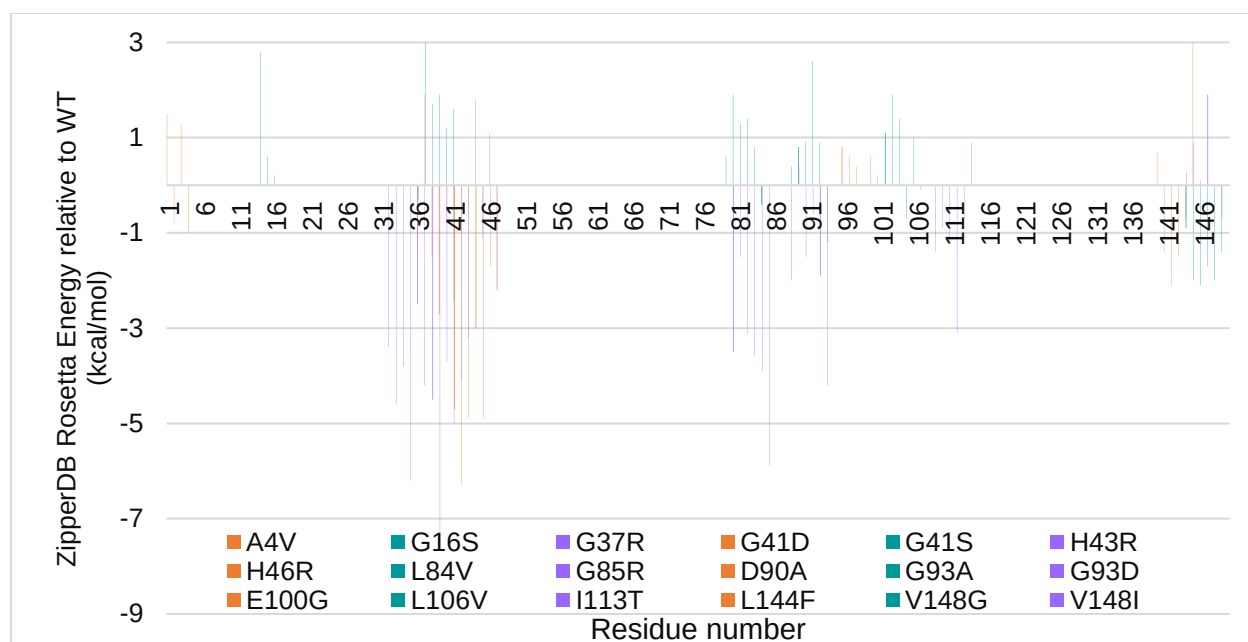


Figure 2.5 ZipperDB Predictions of SOD1 variants relative to WT.

Rosetta energies were calculated for each variant and subtracted from WT. As the energies are negative, a positive value on the plot indicates that a residue is more amyloid-prone than WT and a negative value indicates it is less amyloid-prone than WT. ZipperDB is a sliding window predictor so only residues adjacent to the mutation will be impacted for each variant. Consequently, colors repeat every 3 sequential variants for ease of visualization.

Evolutionary conservation is another metric that may be calculated and used to identify aggregation hot spots. This metric evaluates how often the WT residue is conserved at each position. Using the ConSurf server, multiple sequence alignments are performed on a sequence and its homologs, and the evolutionary conservation is predicted^[198]. As structurally or functionally important regions are more likely to be conserved, mutations in these sequences are more likely to perturb protein folding. The ConSurf prediction for SOD1 is shown in Figure 2.6, with the cut-off for 'conserved' shown in a black line (ConSurf score ≥ 7).

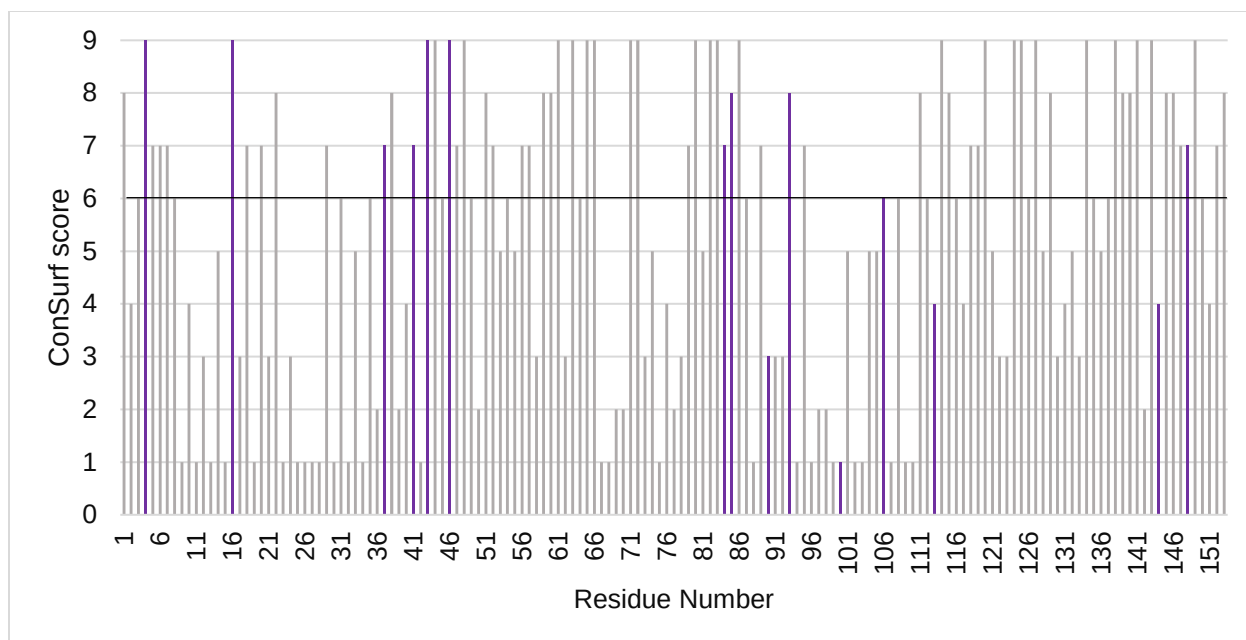


Figure 2.6 ConSurf Prediction for WT SOD1

A score of 1-9 is given to each amino acid based on its normalized conservation score, where 1 means the residue is not conserved, and 9 indicates the residue is very highly conserved. Scores equal to or greater than 7 are considered conserved, and this cut-off is represented by a black horizontal line at a score of 6. Residues examined in this study are colored purple.

ConSurf shows that the C-terminus is highly conserved, as are large portions of the metal binding loop (residues 51-84) and electrostatic loop (residues 122-143). Of the 18 mutations studied, 12 are in highly conserved residues and are expected to perturb the structure greatly. Previous ConSurf analyses have shown that 59% of mutant SOD1 variants (in a set of 233) occurred in highly conserved positions^[199], so the mutants in this study are representative of SOD1 mutations more generally and will inform on whether ConSurf can be used to predict aggregation. A multiple sequence alignment was also performed to determine the sequence preference at each position in SOD1. Figure 2.7 shows the preferred residues at each position in SOD1.

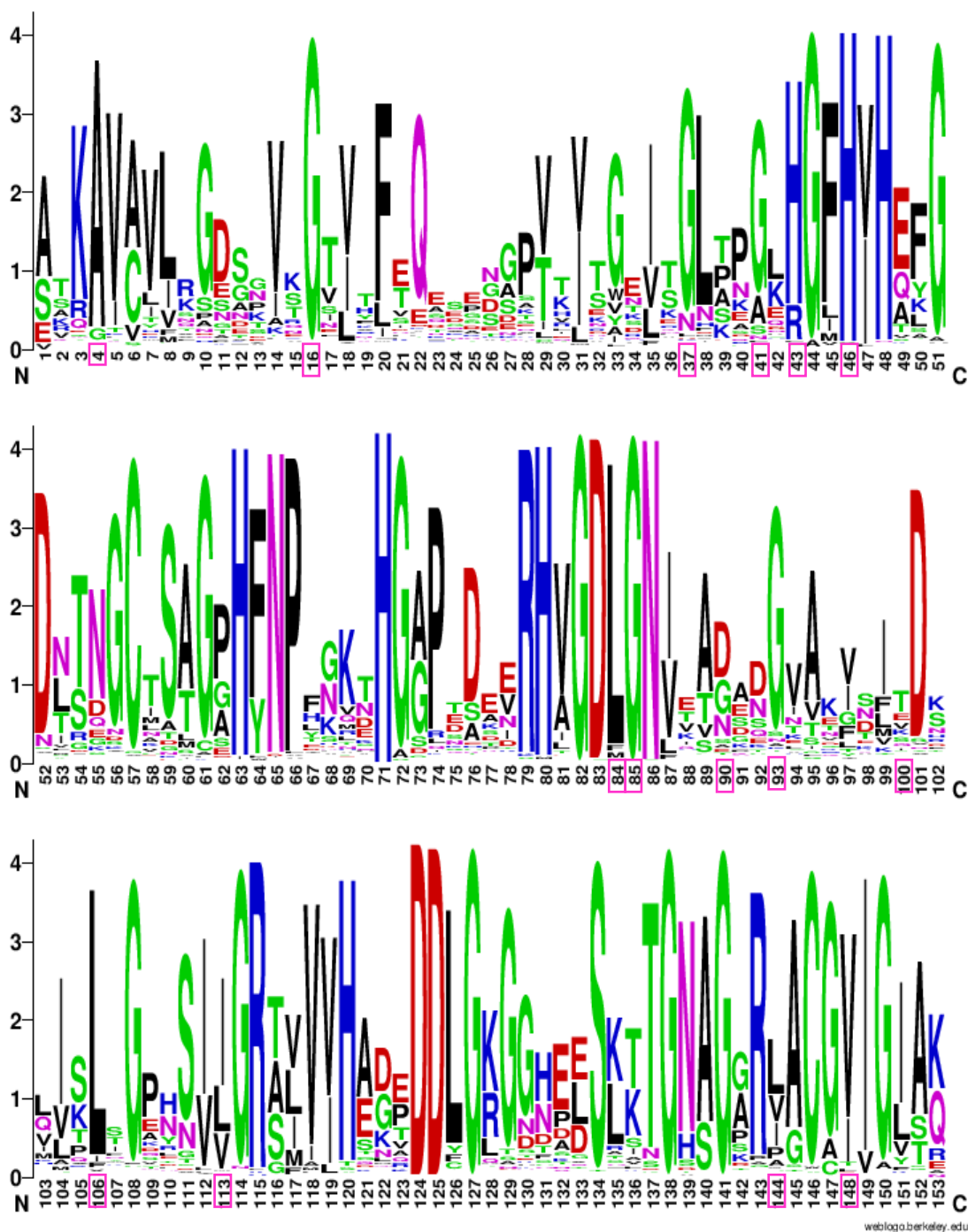


Figure 2.7 Logo for WT SOD1 amino acid sequence.

Logo is constructed based on a multiple sequence alignment of 1750 sequences from PredictProtein^[200]. The height of each letter corresponds to the frequency of the corresponding amino acid. Image created using WebLogo^[201].

2.3.2 fALS SOD1 mutants aggregate more with increasing temperature

The aggregation propensity of the 18 fALS mutants discussed above was measured at 25 °C, 30 °C and 37 °C in minimal media with no metals to mimic the most immature form of the protein, reduced apo SOD1 (apoSH) (Figure 2.8, example SDS-PAGE in Figure A.2.1). The melting temperature (T_m) for apoSH is known for 9 of the variants in this study and ranged from 33.2°C to 52.6°C, indicating that at the highest temperature studied (37 °C), several of the mutants include populations of unfolded protein. Indeed, it was found that all variants aggregated heavily at 37 °C, with even variants thought previously to be less aggregation prone (H46R, V148I) aggregating greater than 50%. Upon conducting the experiment at the lower temperatures, a clear trend emerged: all variants aggregated less. This observation implicates less well-folded species in aggregation. Raw aggregation values can be found in Tables 2.3 and 2.4.

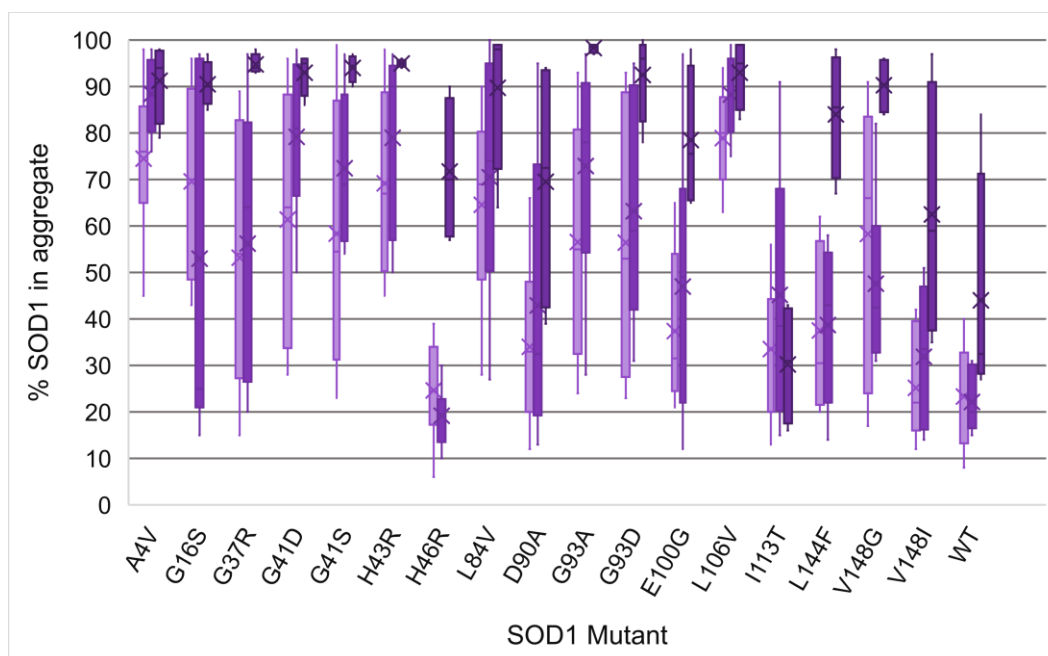


Figure 2.8 Impact of temperature of the aggregation of SOD1 variants.

Aggregation propensity was measured at 25 °C, 30 °C and 37 °C (shown as light to dark purple, respectively) at 4 and 6 hours post induction in triplicate where possible (at 25 °C and 30 °C) and only once at 37 °C. X is the mean; outliers are not shown.

Table 2. 3 Measured aggregation propensity of SOD1 variants measured at 25 °C

	T2		T4		T6	
	apo	zinc	apo	zinc	apo	zinc
WT	25.7 +/- 17.0	20.7 +/- 15.2	21.8 +/- 8.71	19.7 +/- 13.3	24.7 +/- 10.8	25.5 +/- 13.8
A4V	90.0 +/- 0.72	42.2 +/- 16.8	86.3 +/- 9.78	61.3 +/- 12.7	88.8 +/- 8.59	61.9 +/- 14.7
G16S	64.1 +/- 12.1	47.2 +/- 20.6	87.4 +/- 3.49	50.6 +/- 4.68	91.7 +/- 4.28	46.6 +/- 10.5
G37R	52.5 +/- 11.6	28.8 +/- 1.45	70.5 +/- 9.63	26.1 +/- 6.32	85.6 +/- 2.89	30.0 +/- 14.3
G41D	66.5 +/- 11.5	31.3 +/- 5.46	80.8 +/- 7.29	34.9 +/- 3.33	91.0 +/- 5.11	39.8 +/- 14.8
G41S	65.0 +/- 14.2	43.7 +/- 12.3	77.9 +/- 11.9	29.0 +/- 6.03	87.6 +/- 12.8	37.6 +/- 7.07
H43R	76.7 +/- 4.85	44.0 +/- 10.9	80.1 +/- 12.6	51.4 +/- 6.63	88.9 +/- 13.1	55.5 +/- 10.6
H46R	31.4 +/- 10.0	26.2 +/- 5.93	17.8 +/- 16.8	23.9 +/- 9.15	28.3 +/- 12.8	27.4 +/- 10.5
L84V	56.5 +/- 18.3	34.6 +/- 12.5	72.0 +/- 2.48	34.0 +/- 9.27	82.7 +/- 0.25	47.2 +/- 25.0
G85R	---	---	77.2 +/- 23.7	81.3 +/- 8.76	84.9 +/- 8.93	83.4 +/- 13.3
D90A	35.1 +/- 9.57	25.0 +/- 15.0	36.2 +/- 1.58	24.1 +/- 9.68	55.1 +/- 9.80	21.6 +/- 9.79
G93A	71.1 +/- 7.51	29.2 +/- 1.10	73.1 +/- 6.80	32.9 +/- 7.37	87.4 +/- 6.28	32.6 +/- 9.97
G93D	77.3 +/- 12.8	26.6 +/- 2.43	80.4 +/- 12.9	27.9 +/- 7.37	85.8 +/- 11.0	31.4 +/- 7.08
E100G	63.0 +/- 6.73	30.2 +/- 13.3	46.9 +/- 14.0	26.1 +/- 3.55	58.8 +/- 8.61	28.0 +/- 6.56
L106V	80.2 +/- 5.29	41.6 +/- 14.2	79.9 +/- 14.1	58.6 +/- 22.0	86.3 +/- 10.6	76.6 +/- 13.3
I113T	59.5 +/- 2.14	13.3 +/- 9.08	44.7 +/- 3.49	23.7 +/- 9.09	47.9 +/- 7.26	28.3 +/- 10.0
L144F	31.4 +/- 11.1	16.2 +/- 7.92	55.3 +/- 4.81	22.4 +/- 2.37	58.9 +/- 3.89	26.1 +/- 8.83

V148G	66.7 +/- 20.5	26.4 +/- 11.1	82.5 +/- 6.30	31.0 +/- 16.5	88.7 +/- 6.56	30.1 +/- 14.0
V148I	29.0 +/- 6.22	14.1 +/- 3.29	30.7 +/- 16.1	19.1 +/- 5.44	33.2 +/- 11.9	16.6 +/- 3.92

Variants are all measured without and with zinc at 2-, 4- and 6- hours of expression at 25 °C. Values are the average of 3 independent measurements (growth and SDS-PAGE) with corresponding standard deviations given.

Table 2. 4 Measured aggregation propensity of SOD1 variants measured at 30 °C

	T2		T4		T6	
	apo	zinc	apo	zinc	apo	zinc
WT	20.3 +/- 11.5	7.96 +/-	19.9 +/- 4.65	2.40 +/-	30.1 +/- 0.49	15.4 +/-
A4V	90.5 +/- 6.00	59.7 +/- 14.9	93.6 +/- 2.71	79.2 +/- 5.05	96.2 +/- 2.32	85.5 +/- 5.81
G16S	82.5 +/- 6.39	44.3 +/- 46.0	94.4 +/- 3.63	18.2 +/- 8.00	96.1 +/- 0.85	21.1 +/- 2.22
G37R	61.0 +/- 3.51	30.1 +/- 2.47	76.4 +/- 13.9	16.3 +/- 0.56	79.2 +/- 25.6	28.9 +/- 2.10
G41D	57.1 +/- 23.8	45.8 +/- 6.23	80.7 +/- 4.13	69.4 +/- 15.6	97.5 +/- 0.21	81.3 +/- 5.83
G41S	75.3 +/- 6.97	39.4 +/- 18.9	82.7 +/- 6.55	56.9 +/- 5.11	94.7 +/- 3.29	65.7 +/- 15.8
H43R	67.7 +/- 8.74	56.9 +/- 27.3	92.2 +/- 4.53	72.7 +/- 20.5	94.9 +/- 2.20	82.5 +/- 13.2
H46R	25.9 +/- 6.75	11.6 +/- 2.30	21.8 +/- 4.15	15.7 +/- 3.32	15.9 +/- 9.07	20.9 +/- 7.65
L84V	31.2 +/- 32.7	47.3 +/- 15.1	49.8 +/- 20.0	43.7 +/- 26.1	77.4 +/- 29.3	56.4 +/- 26.9
G85R	---	---	79.2 +/- 10.7	89.5 +/- 4.17	77.8 +/- 15.7	78.5 +/- 7.99
D90A	24.4 +/- 1.14	10.3 +/- 2.77	37.4 +/- 27.2	21.7 +/- 8.34	73.6 +/- 15.7	20.9 +/- 10.2
G93A	44.6 +/- 35.0	29.0 +/- 12.9	81.8 +/- 7.26	47.5 +/- 18.1	94.6 +/- 3.12	60.1 +/- 17.8
G93D	39.4 +/- 7.02	34.5 +/- 21.1	64.3 +/- 14.6	33.6 +/- 8.36	91.9 +/- 2.41	42.4 +/- 10.0
E100G	36.5 +/- 11.9	23.4 +/- 6.14	35.5 +/- 21.5	28.8 +/- 16.9	35.8 +/- 26.5	24.7 +/- 17.3
L106V	80.5 +/- 12.5	52.3 +/- 17.7	92.8 +/- 5.21	78.6 +/- 2.03	97.5 +/- 2.43	88.4 +/- 4.76
I113T	45.1 +/- 16.4	25.7 +/- 12.8	90.1 +/- 1.73	26.0 +/- 12.9	58.4 +/- 4.09	25.5 +/- 14.4
L144F	44.5 +/- 3.35	17.3 +/- 6.64	48.8 +/- 6.19	17.1 +/- 5.56	52.3 +/- 7.20	23.3 +/- 7.64
V148G	43.5 +/- 38.5	47.3 +/- 9.38	29.6 +/- 27.4	39.9 +/- 9.23	48.0 +/- 42.7	41.4 +/- 16.6
V148I	39.8 +/- 27.5	19.8 +/- 2.40	41.7 +/- 10.3	16.4 +/- 4.35	45.5 +/- 7.01	25.7 +/- 7.30

Variants are all measured without and with zinc at 2-, 4- and 6- hours of expression at 30 °C. Values are the average of 3 independent measurements (growth and SDS-PAGE) with corresponding standard deviations given.

Interestingly, at 25°C, different aggregation behaviours can be observed. Unsurprisingly, A4V – a mutant in a location that is highlighted as a hotspot by all predictors and that is predicted to increase the aggregation propensity of the N-terminus- aggregates heavily even at lower temperatures. The G16S and L106V mutants show a similar behaviour, a result that corroborates previous observations that they are in fibril forming regions and that the mutations do not inhibit aggregate formation but may rather enhance it^[202]. The increased propensity for fibril formation may outweigh any stabilizing effects imparted by the lower experimental temperature.

Aggregation predictors fail when 3D structure is important

H46R and V148I aggregate notably less than other variants, and behave more similarly to WT. However, both mutations were predicted to make the protein *more* aggregation prone than WT, with V148I being in a particularly aggregation prone region. This behaviour can somewhat be explained by the specific mutation positions in the 3D structure, something that is not considered by sequence-based predictors. H46R, located in β 4, is notable for hindering copper binding in mature SOD1 because H46, along with H43, H63, and H120, coordinate the copper ion, with the sidechain of H46 pointing into the Cu-binding pocket. Upon mutation to Arg, a Cu^{2+} can no longer be coordinated in the site due to the size of the side chain. However, the positively charged Arg residue may mimic the metal atom in the binding pocket^[203]. In this way, apoSH H46R is significantly more well-folded and more stable than other variants because it is mimicking a more mature form of the protein. Thus, the measured aggregation propensity for H46R more likely represents a “zinc bound” species rather than apoSH SOD1.

The V148I mutation is centrally located in β 8, an aggregation hot-spot and crucial strand for the formation of the SOD1 dimer interface. Previous characterizations of V148I show that it is more stable than pWT^[204], and it appears to fold properly^[205]. V148 is in a stretch of 6 residues that contribute to dimerization. The mutation is not likely to hinder the formation of a β -strand due to the similarity of Val and Ile in both size and hydrophobicity- Ile has previously been shown have a higher β -sheet propensity than Val ^[206]. The side chain of V148/I148 is also solvent exposed and may be stabilized by a larger, hydrophobic residue (V to I mutation). Additionally, while most of the surrounding residues are involved in H-bonding (144-147, 150—152), V148 is not involved in an H-bond, so the mutation likely does not disrupt or eliminate any stabilizing interactions. In fact, V148I has been shown to dimerize readily from the one zinc bound reduced apo^[207], E,Zn-SS ^[208], and oxidized metalated forms^[209]. Consequently, the variant behaves very similarly to pWT. The compilation of these data does not, however, explain why the mutant is so severe in ALS, or why it does not appear to agree with all predictions. A recent

observation on the aggregation for apoSH V148I may provide a clue – under aggregation-forming conditions, V148I was seen to form small, soluble, dimer-like aggregates^[205]. These aggregates would not be detected by SDS-PAGE experiments, as they would appear to be “soluble” protein. Additionally, in appearing so native-like, they are less likely to be detected and disposed of by the cell. This is in agreement with previous observations that soluble protein can consist of misfolded or aggregated species^[210–213]. While large, insoluble aggregates have been thought to be protective, soluble aggregates have been implicated as toxic in several diseases ^[214–217]. These observations offer insight on the observed behaviour of V148I and a limitation on the use of SDS-PAGE to measure aggregation of proteins.

Another mutant with unexpected behaviour is G37R. Based on the various aggregation predictors, G37 is not located in an aggregation prone region, nor is the mutation predicted to increase aggregation propensity (rather, it is predicted to make the area less likely to aggregate). Examination of the structure of SOD1 provides an explanation. G37 is located on Loop 3, which is spatially near Loop 5. A cluster of residues involving G37 form what is known as the β -barrel plug (res 37-42 and 90-95). Many substitution mutations in this region are significantly destabilized relative to the WT protein, suggesting that mutation in the plug may result in the region being more solvent exposed and likely to self-assemble^[218]. Other work on the β -barrel plug suggests that mutations in this region may be introducing steric clashes and slowing maturation, increasing the likelihood of aggregation ^[219]. Previous modelling studies have also shown that destabilizing the β -barrel plug leads to a deprotection of edge β -strands, creating an aggregation face that allows for fibril growth^[220]. Taken together, G37R again highlights the importance of 3D structure to aggregation prediction: despite the mutation to a charged residue appearing innocuous (Figure 2.5), it destabilizes a major structural element in SOD1, leading to aggregation.

Structural importance of glycine in protein folding and aggregation

Approximately 50% of the mutations in this study involve a Gly residue. Glycines have previously been observed to be evolutionarily conserved for their ability to inhibit aggregation^[221]. The WT SOD1 sequence is glycine rich, with 25 of 153 residues being glycines. Of these, 17 (68%) are considered highly conserved, and 16 (64%) are associated with fALS mutations (interestingly, all Glycines with no fALS mutations are found in loops). All mutated glycines in this study (G16, G37, G41, G85 and G93) are strongly conserved and aggregated heavily, even at lower temperatures. Intriguingly, most of these residues are in β -strands or on the edges of loops. While Gly is usually not preferred in β -structures due to its flexibility, glycine kinks have been shown to relieve strain in some β -structures^[222]. As a certain amount

of flexibility is needed in SOD1 to mature, the presence of these Gly residues may be important to this process, and the bulkiness of other residues upon mutation likely limits this flexibility. No clear correlation could be found with sequence position or conservation for Gly residues for which there is no known fALS mutation, indicating that there are other factors at play for the impact of Gly on a protein structure.

The curious behaviour of V148 is also seen in the V148G mutation, which is predicted to aggregate *less* than WT but aggregates much more. Other studies have shown that V148G is greatly destabilized, with a much lower T_m than WT (34 °C compared to WT's 47.6 °C) and greatly impaired dimerization^[184,207]. Considering the location of V148 in a β -strand and Gly's greatly decreased propensity for β -structure, the destabilized and aggregation-prone outcome should not be surprising. The true surprise lies in the oversight of the impact of the mutation by aggregation propensity predictors. This may be attributed to an overinterpretation of the predictors – whether or not a specific region is more aggregation prone may not correlate with overall extent of aggregation if the newly disrupted region, while less aggregation-prone in itself, prevents folding or the formation of a stabilizing interaction. However, the contrast between V148I and V148G, a conserved vs non-conserved mutation, in expected and observed behaviours makes them an interesting case study for further aggregation studies.

Sequence conservation captures some impacts of mutations

Surprisingly, ConSurf scores showed a correlation with aggregation propensity of SOD1 variants, but only if the H46R and V148I mutations are not considered ($R^2 = 0.53$, $p < 0.00001$ (note: all p values can be found in Table A.2.1)). With the inclusion of the two additional mutations, the correlation is weak ($R^2 = 0.17$, $p = 0.0923$). This is again due to the specific nature of the H46R mutation having a *stabilizing* effect on the structure that would not occur with other amino acid substitutions at this location, and V148I not being disruptive enough to the β_8 strand. Indeed, despite the conservation of a Val in position 148, multiple sequence alignments show that Ile is the second most common residue in this position, indicating it can be tolerated (see Figure 2.7). This is not true for H46R: while the Arg residue stabilizes the folded structure, it cannot bind the catalytic Cu^{2+} atom, and thus would not be a functional form of SOD1^[203,223]. All other variants show a better correlation with sequence conservation than to any other predicted parameters, indicating it may be a useful metric for aggregation prediction (Figure 2.9).

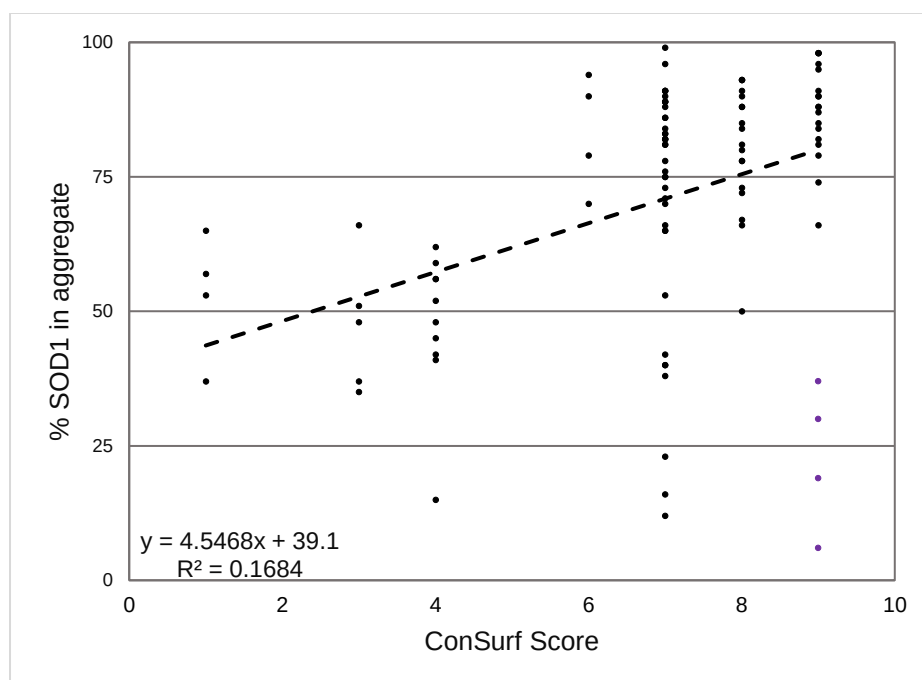


Figure 2.9 Correlation between SOD1 aggregation and ConSurf sequence conservation score

Aggregation data measured at 25°C and 30°C at 4 and 6 hours post induction with no metals added. Correlation shown ($R^2 = 0.17$, $p = 0.0923$) represents all mutations in the study.

Aggregation predictors do not correlate to aggregation propensity

V148I and H46R both demonstrate a downfall of most aggregation-propensity predictors: without structural information, predictions are extremely limited. This was highlighted when correlations were attempted between aggregation propensity data and normalized aggregation scores for each variant: no convincing correlations were found (Table 2.5). While there are predictors that take account a structure, this is limited by availability of structures, computational power required, and differences in structure that may (and do) occur between mutants.

Table 2. 5 Correlation of normalized aggregation propensity scores to cellular aggregation propensity assessed by SDS-PAGE

	Chiti-Dobson	Wang-Agar	Zyggre-gator	Fold-Ztox	Profile-Amyloid	PASTA	TANGO	Waltz	3D
Apo 25 °C:	0.1044	0.0942	0.0703	0.0252	0.3189	0.0404	0.0011	0.0775	0.0016
Zn 25 °C	0.0692	0.0585	0.0026	0.0475	0.2849	0.0285	0.1898	0.0817	0.0019
Apo 30 °C	0.0355	0.0549	0.0603	0.0211	0.0861	0.0105	0.1099	0.1404	0.004
Zn 30 °C:	0.0966	0.1015	0.0485	0.1907	0.173	0.0341	0.0602	0.2326	0.0153

Normalized aggregation scores are courtesy of Heather Primmer^[224]. No correlations are significant in a 90% confidence interval (Table A.2.1).

Aggregation propensity correlates to biophysical parameters

Measured aggregation propensity was then compared to known/measured biophysical parameters of SOD1 mutants. A summary of previously measured T_m and $\Delta G_{\text{dimer dissociation}}$ can be found in Naser *et al* [49]. The values are partial due to limitations in mutant availability in the pseudo WT background (a requirement for calorimetry measurements) and due to sample aggregation for unstable variants. Note that for the following analyses, 37 °C data were not utilized as significant differences in aggregation propensity were not observed. Additionally, the two-hour timepoint was not utilized, as the amount of protein present was generally low, making the method more error prone.

Aggregation propensity was found to correlate to apoSH T_m (Figure 2.10, top) with an $R^2 = 0.60$ ($p < 0.00001$). This was the strongest correlation found to the aggregation propensity data and implicates immature species in SOD1 aggregation. A correlation was also seen with $\Delta G_{\text{dimer dissociation}}$ with an $R^2 = 0.4675$ ($p < 0.00001$), again pointing to a less folded, immature species as a culprit in aggregation (Figure 2.10, bottom). This correlation is to be expected, as apoSH T_m and $\Delta G_{\text{dimer dissociation}}$ relate to each other ($R^2 = 0.56$, $p = 0.12$, plot not shown) and describe similar things about the SOD1 structure – variants that are less stable and do not stay well folded have more difficulty forming a dimer. However, it makes it hard to pinpoint whether a poorly folded monomer or an unfolded species is more responsible for aggregation. Less aggregation prone variants like H46R and V148I are better represented by these data because T_m takes structure into account. This supports structure being imperative for aggregation prediction. Unfortunately, these data are limited because of how destabilized many of the mutants are in the apoSH state: attempts to obtain more T_m 's were often marred by aggregation of the protein. This limitation means that only a portion of the data could be compared to these two biophysical parameters. Attempts at correlations with other biophysical parameters (eg folding rate, K_d zinc) were unsuccessful. This should not be taken to mean that no other parameters impact protein aggregation – they may not be properly reflected under IB producing conditions in a cell. Additionally, cellular parameters which cannot be measured may be impacting aggregation: rate of protein translation, cellular crowding, interactions with general chaperones, and protein degradation machinery may vary across the mutants and may impact the amount of aggregate formed. These variables may account for the amount of spread seen in the data within one variant/condition.

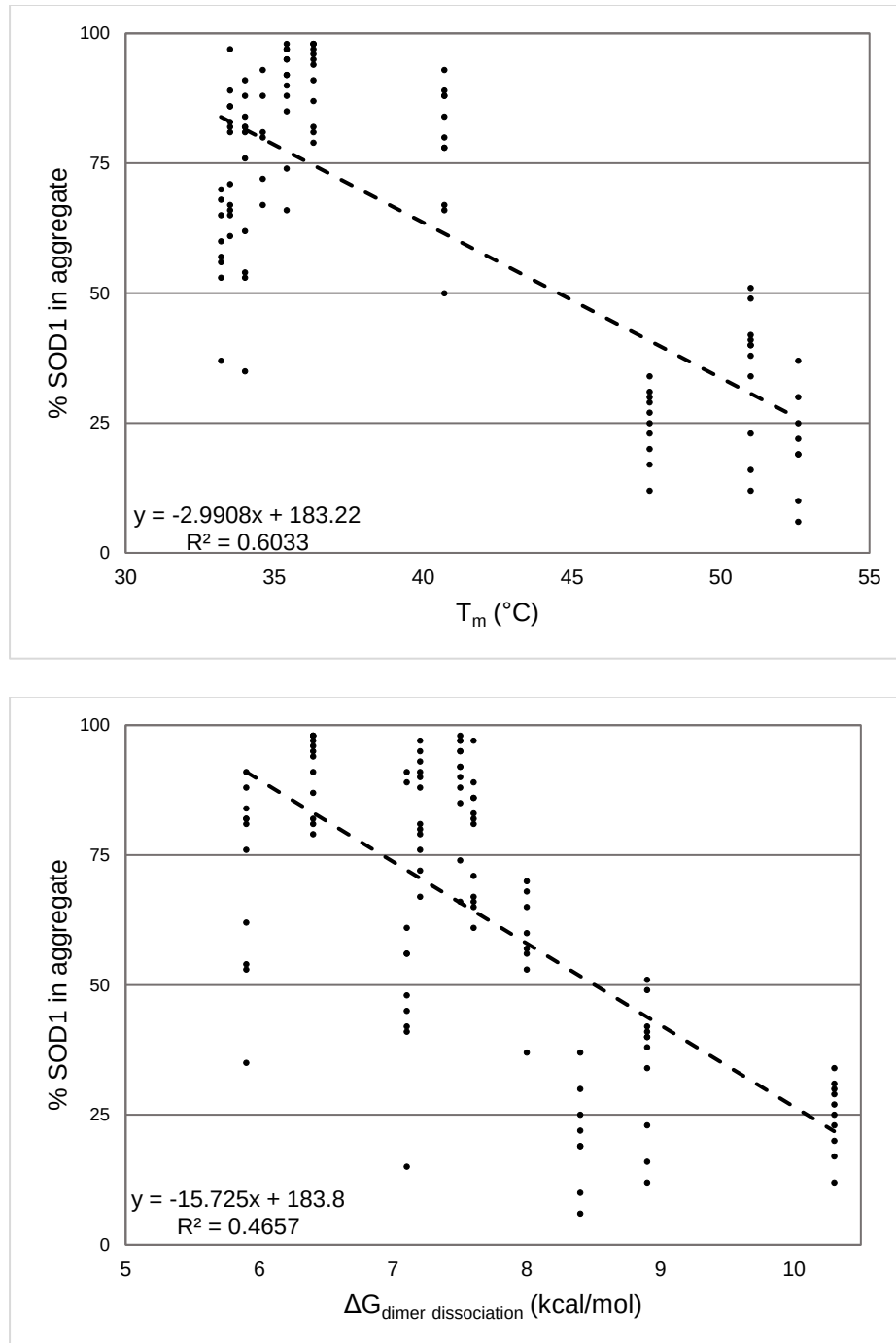


Figure 2.10 Correlations between SOD1 aggregation propensity and SOD1 biophysical parameters

Correlation to measured reduced apo T_m (top) and ΔG dimer dissociation (bottom) is compared to %SOD1 in aggregate at 25°C and 30°C at 4 and 6 hours post induction in duplicate or triplicate where possible.

2.3.3 Addition of zinc rescues a portion of SOD1 from aggregation

The aggregation propensity of the fALS mutants discussed was measured again at 25 °C, 30 °C and 37 °C in minimal media with zinc supplementation to mimic a more mature form of the protein, reduced SOD1

with one or two zincs bound (referred to as E, Zn SH or Zn₂SH, respectively). The T_m for oxidized apo SOD1 has previously been measured for 13 variants from this study and ranged from 45.6 °C to 60.5 °C, indicating that maturation has a notable effect on stability and that at the highest temperature studied (37 °C), no mutants will be significantly unfolded. Note that the T_m for E,Zn SH SOD1 has also been measured for few variants, and brings T_m to a similar or higher range^[207]. It was found that almost all variants show at least minor amounts of rescue from aggregation upon the addition of zinc, and that this trend is visible at all temperatures measured. This result supports the hypothesis that immature and destabilized SOD1 is responsible for initiating SOD1 aggregation. No major correlations to biophysical parameters emerged from the zinc studies, therefore the plots/correlation values are not included. Additionally, all conditions were compared to previous work in rich media (which contains metals and would allow the protein to mature further) and no significant correlations were found. This indicates that maturation can have substantial structural implications to the protein that impact the mutations in different ways. A summary of the data can be found in Figure 2.11, and raw values across the conditions are listed in Tables 2.3 and 2.4. Correlations to rich media can be found in Figure 2.12.

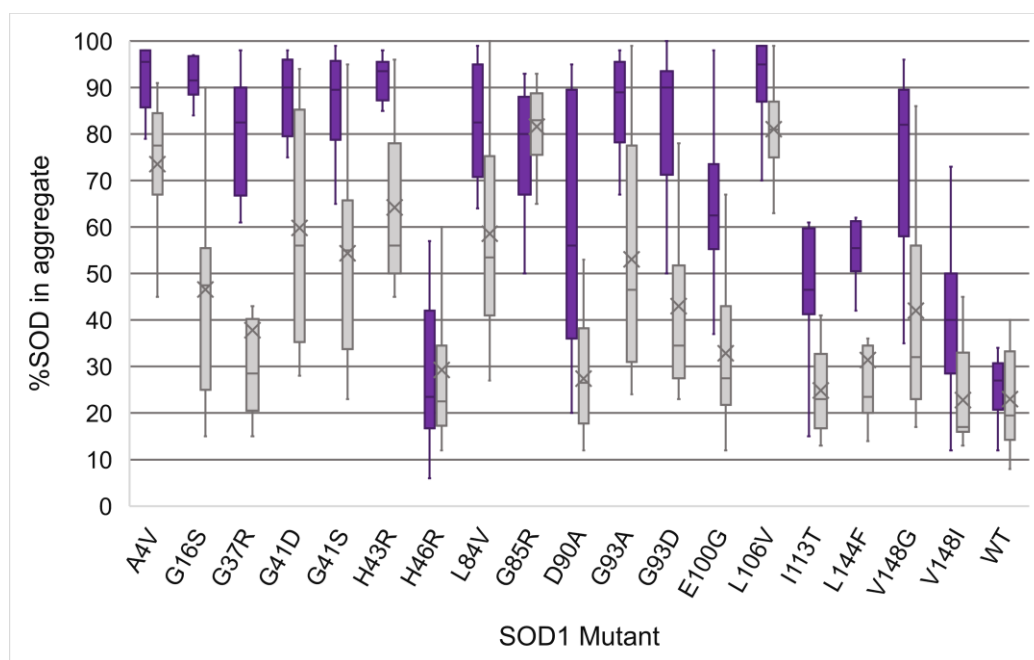
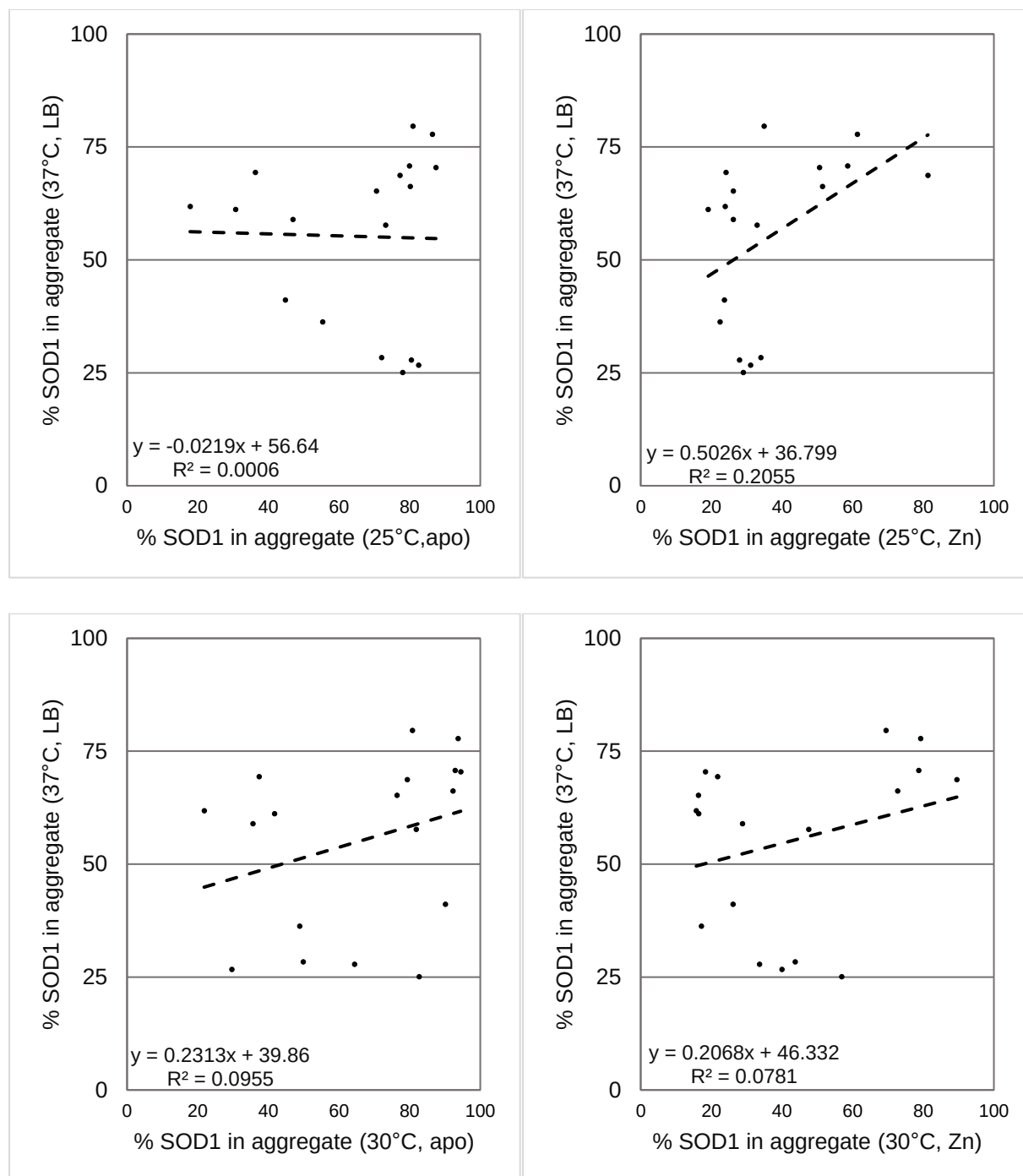


Figure 2.11 Impact of Zinc on SOD1 aggregation

SOD1 aggregation is shown for SOD1 grown in apo (purple) and with zinc (grey) at 25°C and 30°C at 4 and 6 hours post induction in duplicate or triplicate.



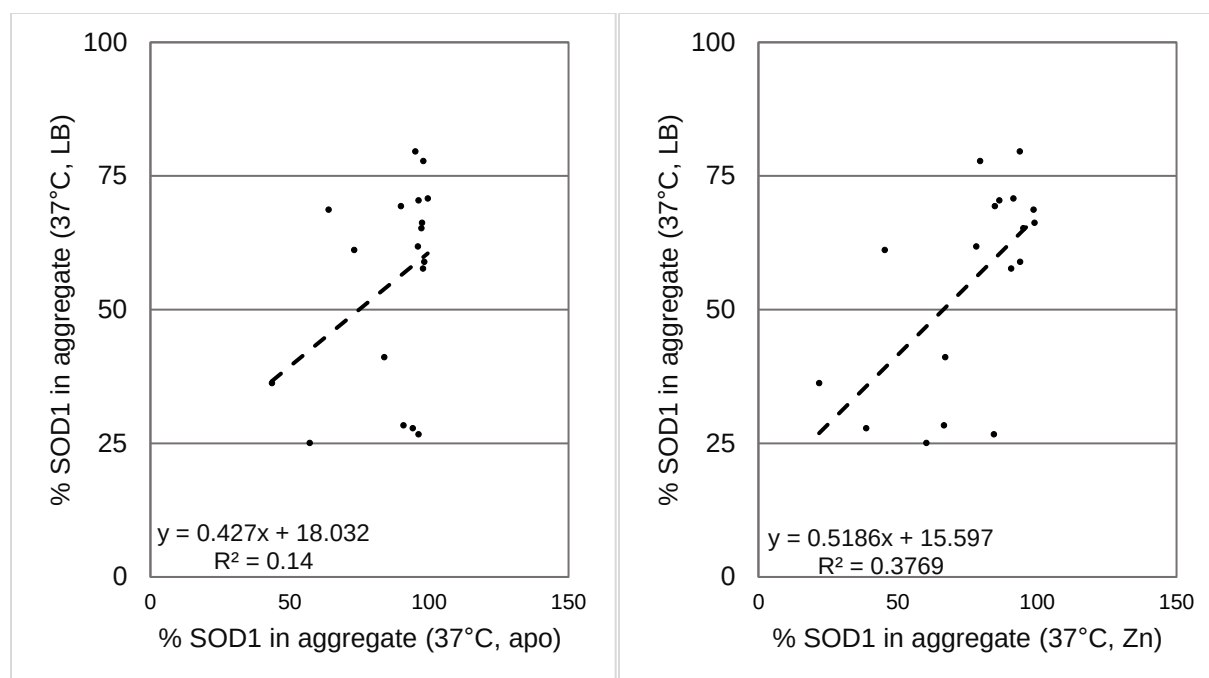


Figure 2.12 Comparison of average SOD1 aggregation in LB to minimal media

Comparison of %SOD1 in aggregate at 4 hours post induction for each mutation under each of the 6 conditions considered (25°C, 30°C and 37°C with and without metal) to %SOD1 in aggregate grown in LB at 37°C, 4 hours post induction. Rich media aggregation propensity was previously measured by John Almey^[193]. p values can be found in Table A.2.1.

Metal-binding mutants model the impact of zinc on aggregation

Two mutants that impact metal binding, H46R and G85R, were included in this study. These variants impact metal binding in different ways: in H46R, the arginine side chain both impairs metal coordination and mimics a metal ion, stabilizing the structure. Conversely, the arginine in G85R hinders metal binding without stabilizing the structure. Both mutants show an identical result upon the addition of zinc to the growth medium: there is no change in aggregation propensity. However, in the case of H46R, this is because aggregation was minimal regardless of metal content: the aggregation propensity of the other variants upon the addition of zinc comes to resemble that of H46R, confirming that the Arg side chain is mimicking a metal and stabilizing the protein. On the other hand, G85R has a high aggregation propensity, and sees no change upon zinc addition. This confirms that the mutation hinders zinc binding. Additionally, it shows that residual zinc binding in the copper-site is not imparting significant structural stability, and that the effects of zinc on the other mutants likely reflects only zinc binding to the correct binding site^[203].

Proximity to Zn site contributes to zinc rescue

While all mutants (except H46R and G85R) show rescue upon addition of zinc, the amount of rescue is variable. A4V, the most prominent aggregator in the metal-free experiments, experiences extremely

minimal rescue from aggregation, and once again, L106V mimics its behavior. Conversely, G16S, G37R, and both G93 mutants show considerable rescue from aggregation, despite G37R being more destabilized than A4V.

To attempt to explain this range of behaviours, the distance from each mutation to the zinc and copper binding sites was measured (on mature SOD1, PDB 1HL5) with the initial hypothesis centering on the idea that residues near the metal-binding site are more likely to be impacted by the stabilizing structural changes imparted by zinc binding. Distances can be found in Table 2.6. These distances were compared to the extent of rescue seen for each variant (Δ %SOD1 in aggregate = %SOD1 in aggregate_{apo} - %SOD1 in aggregate_{zn}) using the average aggregation propensity for a condition per mutant. While an initial comparison between Δ %SOD1 in aggregate at 25 °C and distance to the zinc atom seemed promising (R^2 = 0.3225, p = 0.19, Figure 2.13), including the 30 °C dataset lowers the correlation (R^2 = 0.2584, p = 0.12, plot not shown) and inclusion of distance to the copper binding site (which was included because the zinc atom has been shown to bind both sites) also did not appear promising (R^2 = 0.137, p = 0.43, R^2 to average distance of both metals = 0.2088, p = 0.22, plots not shown).

Table 2. 6 Measured distance of each mutation location to metal binding sites in mature SOD1

	Distance to Zn ²⁺ (Å)	Distance to Cu ²⁺ (Å)	Average distance (Å)
A4	19.6	20	19.8
G16	21.1	15.7	18.4
G37	23.2	17.8	20.5
G41	20.3	17	18.65
H43	14	10.9	12.45
H46	7.9	4.8	6.35
L84	6.8	8.1	7.45
G85	7.8	8.7	8.25
D90	22.6	19	20.8
G93	25	20.5	22.75
E100	16.5	17.2	16.85
L106	17.6	18.5	18.05
I113	16.2	15.6	15.9
L144	14.9	8.6	11.75
V148	17.6	13.2	15.4

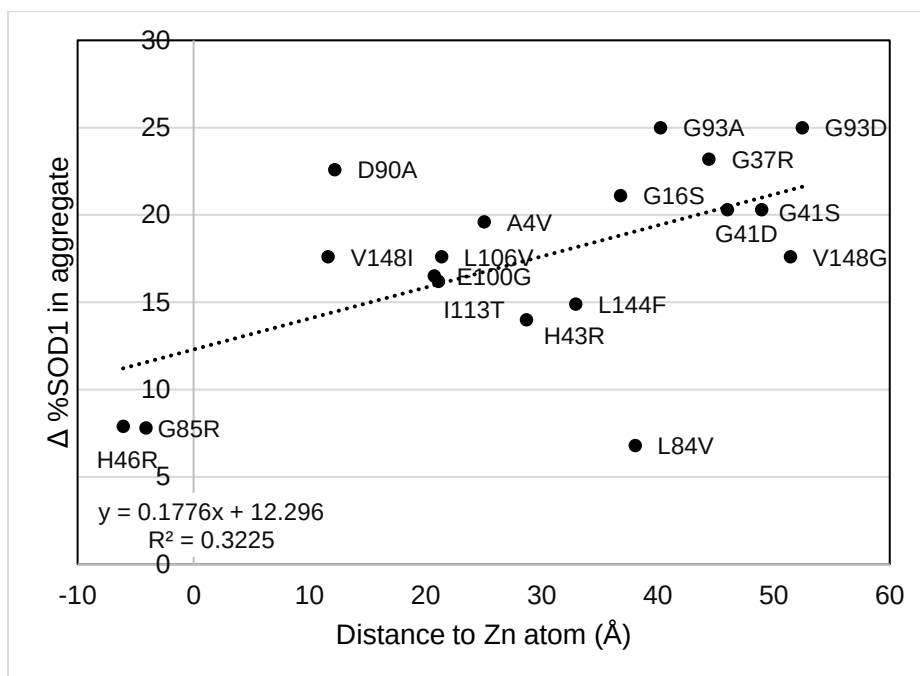


Figure 2.13 Comparison of zinc rescue of SOD1 aggregation relative to mutant position

Average %SOD1 aggregation at 25°C 4 hours post induction (with zinc) subtracted from the same growth condition (without zinc) compared to mutation backbone distance to zinc ion. Distances were measured on PDB 1HL5 using PyMol.

Loop 6 emerges as for a key contributor to aggregation

Previous work has shown that loop 6 (res 102 – 114) is crucial for SOD1 aggregation. More specifically, residues 101-107 were implicated in the formation of steric zippers through face-to-back packing, and calculated Rosetta energies for mutations in the region found that almost all mutations in the region maintain or increase the aggregation propensity of the region, particularly the L106V mutation. Interestingly, L106 is also the most highly conserved residue in the loop. While the loop has no specific function, it is a more “structured” loop, due to numerous (9) H-bonds formed both within the loop and to surrounding structures, with L106 H-bonding with res 22 (β 2) and G114 H-bonding with res 149 (β 8/C-term). Additionally, the latter half of loop 6 (res 113-115) is involved in the native dimer interface, with the addition of I104, L106, S111 and I112 if non-native dimer interfaces are concerned.

The three mutations in this study involving Loop 6 (E100G, which flanks the loop, as well as L106V, and I113T) model three different aggregation behaviours. E100G is technically not in the loop, but is greatly destabilized relative to WT. As Gly mutations decrease the β propensity of a residue, this mutation possibly extends loop 6 to include a portion of β 6. This may interfere with the β 5- β 6 cleft and reveal an aggregation hot-spot. Interestingly, E100 is the least conserved residue in this region – position 100 contains a Thr residue more often than a Glu (30% compared to 11%). E100G models the typical

behaviour seen by SOD1 mutants, aggregating a great deal under metal-free conditions, but showing rescue from aggregation.

On the other hand, L106V behaves very similarly to G85R (high aggregation and no metal rescue) despite not being near the zinc ion or having a known impact on metal binding. A comparable behaviour is observed for A4V: extremely high aggregation propensity was observed with minimal metal rescue, despite having no known impact on metal binding. On the contrary, A4V has been shown to readily bind zinc ions in both of its binding sites and form a thermally stable Zn-Zn species. However, A4V has also been shown to inhibit dimerization. Previous NMR experiment using chemical exchange saturation transfer (CEST) have found that the A4V mutation disrupts the dimer interface and favours a monomer species^[225]. A4V has also been shown to have difficulty binding the CCS, and to be copper deficient *in vivo* despite being able to bind copper^[186]. More recent calorimetry results support that A4V has a weakened dimer interface and have postulated that it forms a modified monomeric species that is stable but cannot dimerize^[207]. Such a species would not show rescue from aggregation in these studies, as native dimerization appears to be crucial in inhibiting aggregation, as modelled by G85R. If these results are taken together with L106's proximity to the dimer interface (and to A4, Figure 2.14), one may hypothesize that L106V may similarly be making a modified monomeric species.

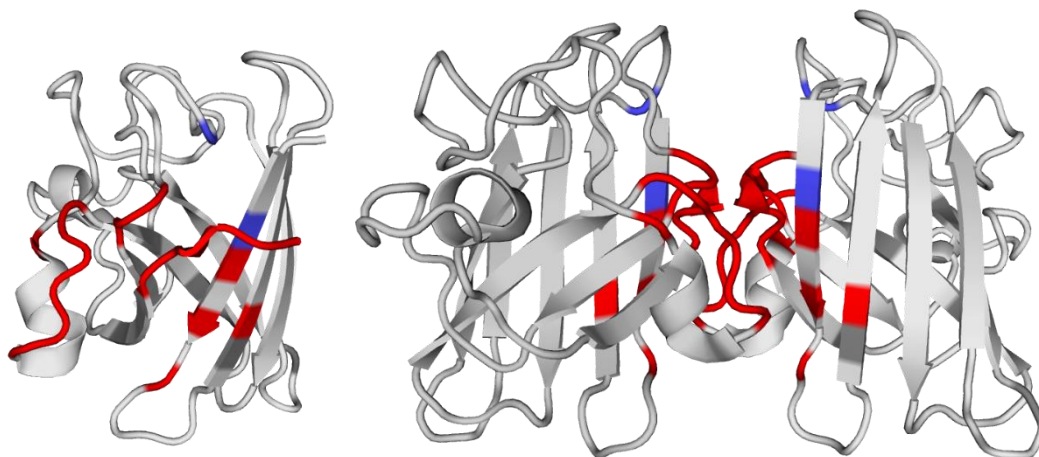


Figure 2.14 Proximity of A4 and L106 to dimer interface

Mutations are colored in light purple and residues involved in the native dimer interface are colored in red. Left: looking into the dimer interface on an SOD1 monomer. Right: front view of a SOD1 dimer with the interface intact. Both images made with PDB 1HL5 in PyMol.

The last loop 6 mutation examined is I113T. I113 is also involved in native dimerization. However, previous work has shown that I113T does not disrupt the interface as drastically as other mutations.

I113T is capable of both dimerization and metal binding, and previous *in vivo* studies have shown it to be a low aggregator. Additionally, RMSF and B-factor analyses have shown that the flexibility of the structure remains WT-like upon mutation^[199]. Consequently, we found that I113T has a low aggregation propensity, and shows some rescue upon zinc addition. Taken together, these three mutations show the sensitivity of loop 6 to changes, despite not being a functional loop. Small changes across the loop can drastically change the aggregation behaviour of the SOD1 protein. This example illustrates some of the difficulty in studying SOD1 aggregation: with many potential variables upon protein maturation, and many key structural contacts in the protein, changes can often have seemingly unpredictable impacts on the protein. While it is possible in hindsight to make sense of the observed data, it is much harder to anticipate or predict it.

V148G counters rescue trends

Unfortunately, the V148G variant contradicts the hypothesis that dimerization is the key maturation step that occurs upon zinc binding to slow aggregation. V148G is a dimer interface mutant that has been shown to not dimerize in the Zn-bound form (forming almost 100% monomeric, zinc-bound protein). Calorimetry measurements have shown that it has the weakest dimer interface by far^[207]. However, it is rescued from aggregation upon the addition of zinc, indicating one of two things: V148G may have a different pathway to aggregation, or a different pathway to rescue – or both. If V148G is aggregating through different structural sites than the other mutants, it is possible that the stable monomer is enough to inhibit this pathway and a dimer is not required. This is supported by structural predictions that the C-terminus as an aggregation prone region, and that the V148G mutation makes it less so. Combined with previous work that the C-terminus may nucleate aggregation, the aggregation of V148G likely stems from a different source than the C-terminus, and thus its rescue is not correlated to the rescue path seen for other mutants. However, it is not possible to determine this from the SDS-PAGE experiments.

2.3.4 Caveats on the use of gel densitometry

The data in this chapter are all based on a simple experiment: small-scale growth of mutant SOD1, a freeze-thaw lysis, an SDS-PAGE gel, and quantification of relative protein amounts through gel densitometry. While valiant attempts to minimize experimental error were made (use of the same glycerol cell stocks, use of the same incubator with the same settings, etc), a large spread in the data is still seen. To assess the extent that the variation may be attributed to the gel densitometry method, two controls were undertaken: firstly, as multiple students were working on the project, two different

individuals analyzed each gel by densitometry to ensure that personal bias in picking band locations or boundaries were not influencing the result. The output was consistent (data not shown), so this practice was eventually discarded. Secondly, a gel of standard, purified protein amounts was run and analyzed by gel densitometry to gauge the upper and lower limits of quantitation (gel is Figure 2.15, volume loaded, mass of protein, and band density are shown in Table 2.7). The relationship between band density and mass of protein was found to be linear in the range used (Figure 2.16 $R^2 = 0.99$ $p < 0.00001$), which was generally representative of protein amounts seen during the experiment. Notably, lanes 8 and 9 have the same amount of protein loaded but display different densitometry values.



Figure 2.15 SDS PAGE of SOD1 standards used to validate gel densitometry

15% SDS-PAGE with ten lanes. Each lane is labelled above the protein. Mass of SOD1 labelled can be found in Table 2.7.

Table 2. 7 Standards loaded in SOD1 gel of standards and corresponding band densities

Lane	Volume of Std		Band density
	(μ L)	Mass SOD1 (μ g)	
1	17.1	16.0	6.55E+06
2	15.0	14.0	5.52E+06
3	12.8	12.0	4.77E+06
4	10.7	10.0	4.01E+06
5	8.6	8.0	3.38E+06
6	6.4	6.0	2.42E+06
7	4.3	4.0	1.87E+06
8	2.1	2.0	9.09E+05
9	2.1	2.0	1.02E+06
10	0.0	0.0	3.36E+05

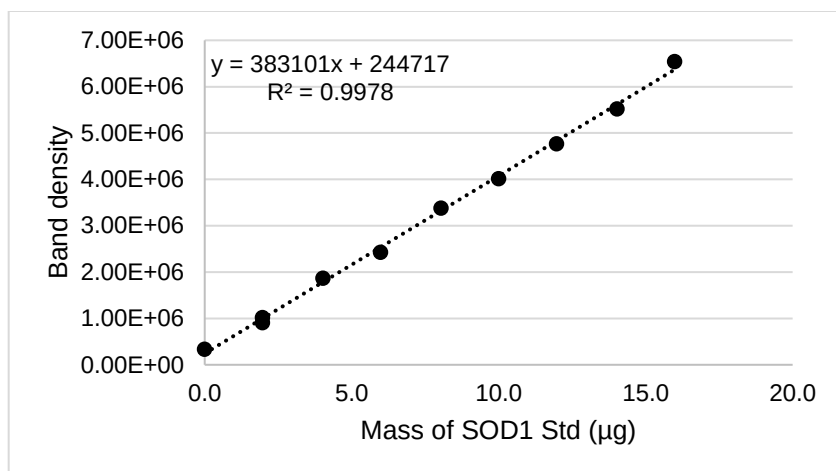


Figure 2.16 Relationship between band density and mass of protein

To assess how the small variations from linearity may impact the results, various ratios of the band densities were done and compared to the expected (Table 2.8). It was found that when bands of similar density were compared (expected ratios of ~30% to ~80%), the error was quite small. However, when bands with a larger disparity were compared, the error is quite significant (~10% variation from the expected), and if an empty lane is utilized, the error increases even more (25%). This means that variants that were lower expressing, or variants wherein the protein was almost completely soluble (H46R at 25 °C) or insoluble (A4V and others at 37 °C), are likely to have a greater variation between their replicate results. This result was verified by comparing all lanes relative to the empty lane (Table 2.9). There was always moderate error associated with this, but the error increased with the band intensities. Ultimately, these errors do not account for all the spread in the SOD1 aggregation propensity data. However, they should be considered for future densitometry work.

Table 2. 8 Expected and actual densitometry ratios of gel of standards

Ratio of protein (μg)	Expected result	Actual result	% error
2/(2+0)	100%	75%	25%
16/(16+2)	89%	87%	3%
16/16+4	80%	78%	3%
14/14+6	70%	69%	1%
12/12+8	60%	58%	2%
8/8+12	40%	42%	3%
4/4+6	40%	43%	8%
6/14+6	30%	31%	2%
2/(8+2)	20%	21%	8%
4/(4+16)	20%	22%	10%
2/(16+2)	11%	13%	23%

Table 2. 9 Expected and actual densitometry ratios of gel of standards compared to empty lanes

Band compared	Expected ratio	Actual ratio	%error
1	1	0.951	5%
2	1	0.943	6%
3	1	0.934	7%
4	1	0.923	8%
5	1	0.910	9%
6	1	0.878	12%
7	1	0.847	15%
9	1	0.729	27%
9	1	0.752	25%

2.4 Conclusions

SOD1 is a rich system for examining cellular aggregation propensity by SDS-PAGE. Examination of a set of mutants throughout SOD1 under different conditions reveals details of protein aggregation. All SOD1 variants including WT showed increased aggregation at higher expression temperatures, a condition under which the variants are known to be less stable and connecting aggregation with protein stability. However, aggregation did not correlate with folding rates, affinity for metal binding, or aggregation in different growth conditions. The addition of zinc, which is known to stabilize SOD1, decreased the aggregation of all variants, again connecting aggregation with stability. The extent to which the protein

is rescued from aggregation varies for each mutant depending on how much the zinc stabilizes the normal protein, and how the mutation impacts zinc binding.

Aggregation predictors frequently did not capture the aggregation behaviour of SOD1 – an analysis of where this was true revealed that 3D structure, and long-range contacts are crucial determinants of SOD1 aggregation and are not captured by aggregation predictors. This is specifically evident with H46R and V148I, which are predicted to make the protein more aggregation prone than WT protein, but in fact make it less. Examination of sequence conservation did appear to correlate with aggregation: strongly conserved residues, particularly Gly, were very detrimental protein solubility. It is hypothesized that this is due to glycine's role in providing flexibility and relieving structural strain.

Lastly, these studies revealed hot spots for aggregation. The dimer interface is clearly involved in both native structure and IBs, as are Loop6 and the β -barrel plug. Accordingly, mutations in these regions were chosen as a focus for further structural examination of IBs. More specifically, based on the trends described in this chapter, the A4V and V148 mutant pair appear to be involved in different aggregation processes. As identifying aggregate structures and mechanisms of IB formation is the overarching goal of this thesis, the work laid out in this chapter will be the basis of further work. Subsequent chapters in this thesis will examine the aggregates of SOD1 mutants identified herein as of potential interest further with both low- and high-resolution methods.

Chapter 3: Optimization of quenched hydrogen-deuterium amide exchange for impure IBs

3.0 Preface

Continuing on from Chapter 2, Chapter 3 describes the optimization of a method to study aggregate structure at high-resolution. As described in Chapter 1, quenched hydrogen-deuterium exchange via NMR has been little applied to the impure cellular aggregates that were quantified in Chapter 2. Chapter 3 will walk the reader through the multiple steps of the method that need to be controlled for success of the experiment and validate the method fully before its application to the SOD1 system (Chapter 4).

This research was conducted in collaboration with Michael V Tarasca and supervised by Elizabeth Meiering. Dalia Naser and Michael Tarasca contributed equally to experimental design and research and are listed as co-first authors on the publication. Dalia Naser focused on optimizing early conditions, minimizing water content in samples, 3D resonance assignment strategies and execution, and data normalization. Michael Tarasca focused on optimizing early conditions, minimizing intrinsic exchange rates, spectral analysis, and measuring/fitting intrinsic exchange rates. Manuscript writing was done primarily by Michael Tarasca with input from Dalia Naser; figures were made primarily by Dalia Naser. In addition to these published results, this chapter included unpublished FTIR analysis of experimental conditions, analysis of incubation on IB structure, and comparison of water content in IB samples, performed and analyzed by Dalia Naser. Tyler Soule contributed to some data analysis and experimental execution working as an undergraduate in the Meiering group. Anna Schaefer contributed Adnectin data to the publication, which are not included in this chapter.

Citation: M. V Tarasca⁺, D. Naser⁺, A. Schaefer, T. G. Soule, E. M. Meiering, *Anal. Biochem.* **2022**

3.1 Chapter Abstract

Inclusion bodies (IBs) are large, insoluble aggregates that often form during the overexpression of proteins in bacteria. These aggregates are of broad fundamental and practical significance, for recombinant protein preparation and due to their relevance to aggregation-related medical conditions and their recent emergence as promising functional nanomaterials. Despite their significance, high resolution knowledge of IB structure remains very limited. Such knowledge will advance understanding and control of IB formation and properties in myriad practical applications. Here, we report a detailed quenched hydrogen-deuterium amide exchange (qHDX) method with NMR readout to define the structure of IBs at the level of individual residues throughout the protein. Applying proper control of experimental conditions, such as sample pH, water content, temperature, and intrinsic rate of amide exchange, yields in depth results for these cellular protein aggregates.

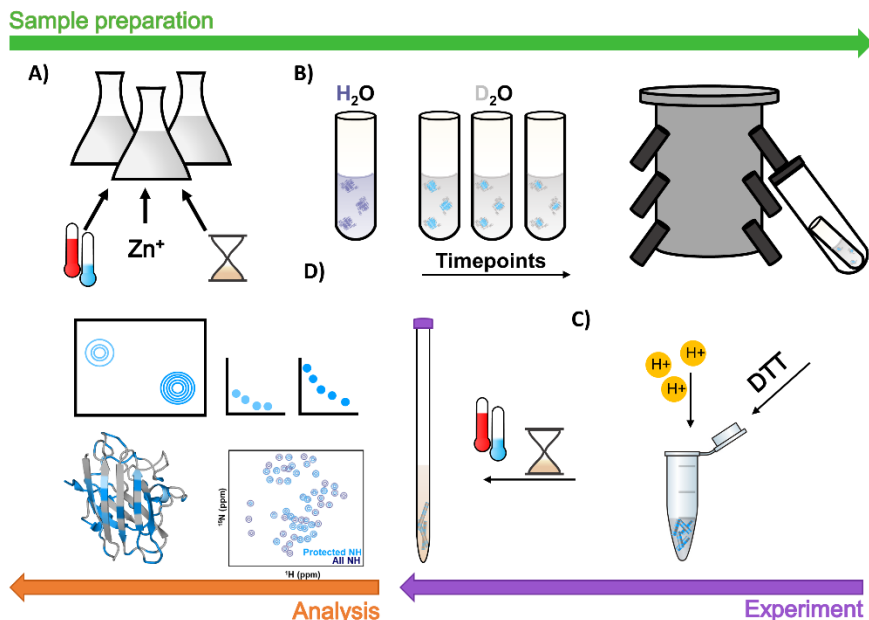


Figure 3. 1 qHDX method overview

The qHDX method has 3 major sections: sample preparation, experiment, and analysis. (A) The first step of qHDX is IB expression in isotopically labelled minimal media. Growth time, temperature, and presence of co-factors can be modified. (B) Crude IB samples are incubated in either H_2O or D_2O for control or exchange experiments. Incubation is lysed by freezing and lyophilization. (C) Samples are dissolved in DMSO buffer containing acid and reducing agent. Quenched amide exchange is measured by NMR. (D) Peak intensities are measured and analysed for aggregate protection.

3.2 Introduction

The overexpression of proteins in bacteria frequently leads to their assembly into large, insoluble aggregates known as inclusion bodies (IBs) ^[226,227]. Recent investigations have focused on the useful functional characteristics of IBs ^[228], with their applications including use as functional nanomaterials ^[228–230], industrial large-scale protein purification ^[231,232], engineering of catalytically active protein aggregates ^[233], and as potential vaccines ^[234]. Central to developing such applications is a deeper understanding of the structures of proteins in IBs ^[228]. Although IBs were originally considered amorphous assemblies, many studies have found evidence for defined structure, which may range from functional and native-like conformations through to unfolded and amyloid. Yet, there is a paucity of atomistic high-resolution structural information on IBs, as their large size, heterogeneity, and impurity have mostly limited structural characterization to low and medium resolution techniques such as dye binding and optical spectroscopies ^[174,228,229,235,236]. These spectroscopic methods may provide general structural information such as the presence of amyloid β structure or differences between two samples but cannot define molecular features in detail.

Quenched hydrogen-deuterium amide exchange (qHDX) measurements have been reported as a means to characterize IB structure at the level of individual amino acid residues ^[237,238]. In qHDX monitored by NMR, aggregated protein is allowed to exchange with solvent deuterium oxide (D_2O), then exchange is quenched, and the aggregate is disassembled, unfolded and analyzed by ^{15}N -based solution NMR (Figure 3.2) ^[237,239,240]. Backbone amides may be protected against exchange by the aggregate structure (hydrogen bonding) and the extent of protection quantitated. Attractive features of qHDX are the relatively low cost, quick preparation of samples with minimal processing, and detailed structural information.

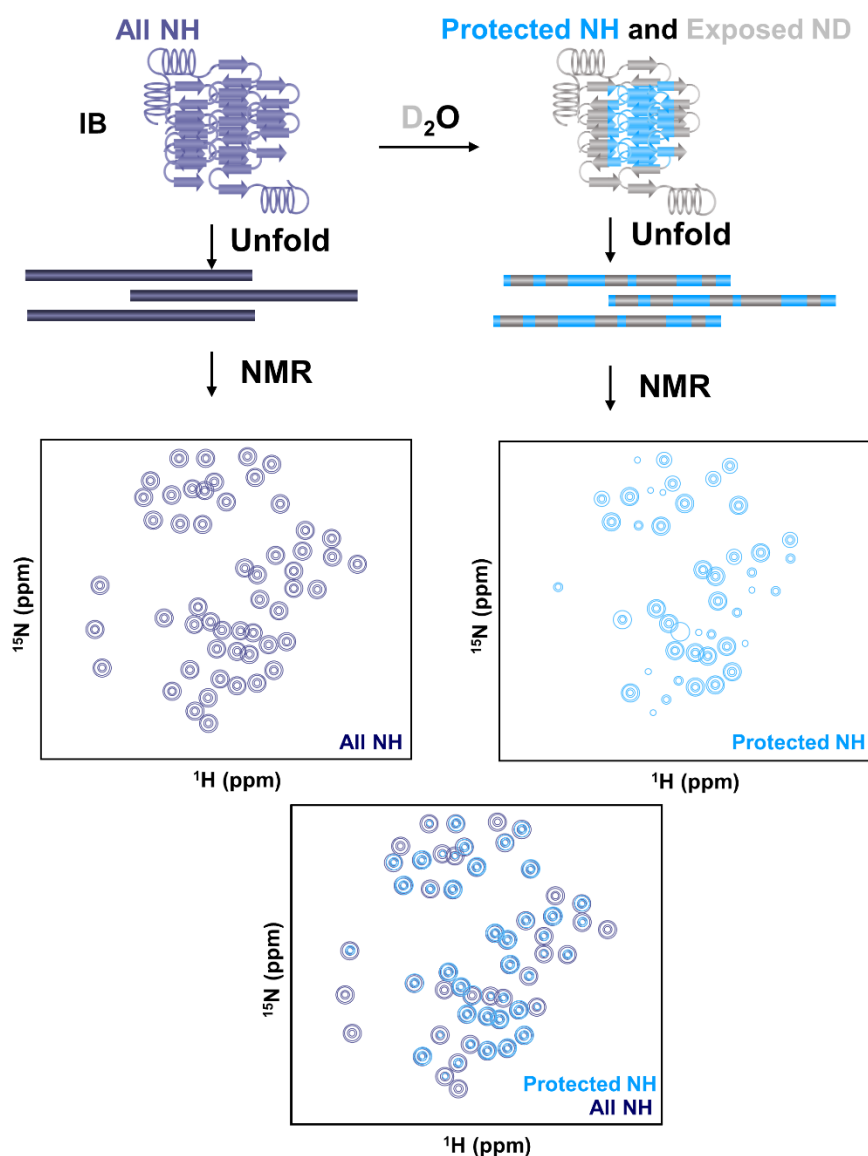


Figure 3. 2 Quenched hydrogen-deuterium amide exchange on protein aggregate

The solvent exposed amino acid amide protons are shown to exchange with deuterons when exposed to D_2O , resulting in a loss of signal when measured by ^1H - ^{15}N HSQC NMR. When this exchanged sample is compared to a spectrum of unfolded protein that has not undergone qHDX, differences in signal intensities are observed. Signal intensities for amino acid residues that are closer to the signal of the fully protonated spectrum are highly protected from the solvent, and residue signals that are lower in intensity are more solvent exposed [239–241].

This methodology was initially applied to *in vitro* aggregates formed by a variety of purified proteins, natural and disease-associated [46,242]. A range of extents and patterns of protection was observed, often with protection of many amides distributed through the protein. For example, protection was observed for many amides in β -microglobulin [46], which may be consistent with cross- β amyloid as well as native-like structure. Subsequently, qHDX was adapted to study a small number of proteins in IB aggregates

^[237,238]. Strikingly, protection in these IBs was observed only for limited regions, indicating the prominent formation of amyloid steric zippers, while most amides were unprotected. Recently, high resolution cryo-EM and solid-state NMR have highlighted that the structures of *in vivo* and *in vitro* aggregates of a given protein often differ ^[243,244], highlighting a need for further structural studies of cellular aggregates.

Seeking to apply qHDX to IBs of additional proteins, we identified previously unrecognized obstacles to obtaining reliable and consistent results; these technical difficulties may account in part for the limited studies in this area. We therefore report here a detailed qHDX NMR methodology, demonstrating which experimental aspects must be closely controlled and which may be varied for successful high-resolution structural analysis of cellular protein aggregates. We further document the impacts of incorrect or unoptimized experimental parameters to aid other researchers in recognizing these previously overlooked errors. Lastly, through comparison to existing works, we provide benchmarks for assessing the success of crucial parts of qHDX experiments, such as backbone resonance assignment, sequence effects, and expected ranges of exchange rates. Our results are of broad significance for studies of both *in vitro* and cellular protein aggregation.

3.2.1 Method Overview

The workflow of qHDX experiments on IBs is summarized in Figure 3.3. The experiments consist of 3 stages: sample preparation (green), NMR experiments (purple), and data analysis (orange). These stages are outlined here and described in detail in the following sections.

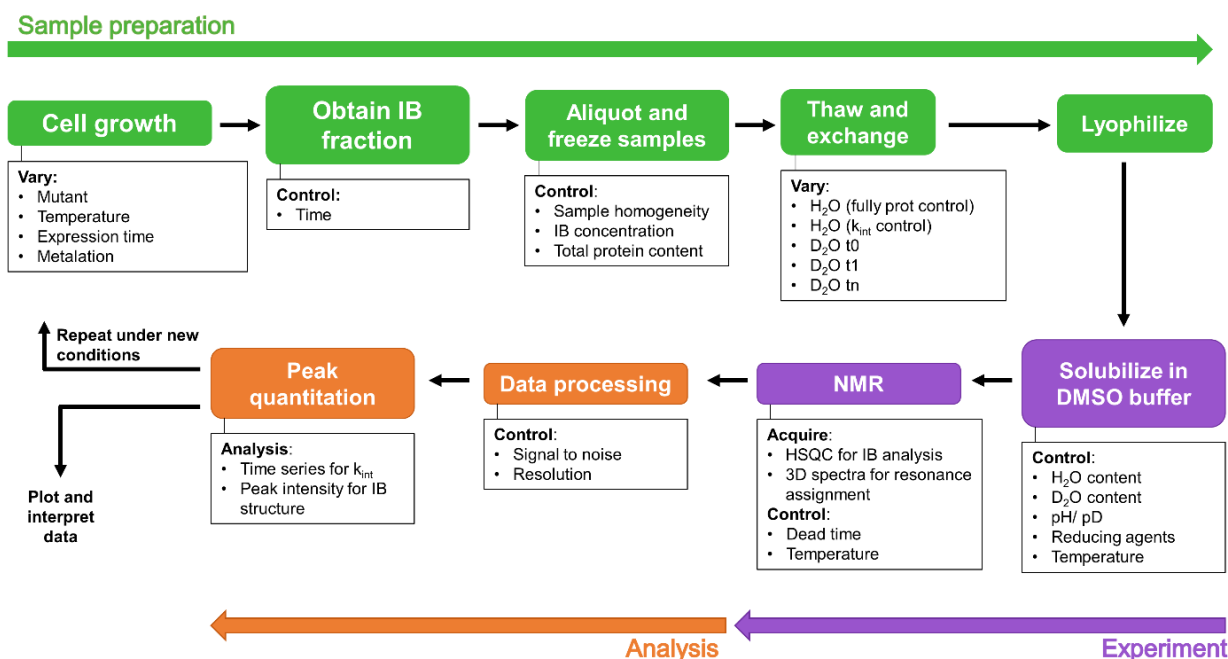


Figure 3. 3 Flowchart of the IB qHDX method.

The workflow can be divided into three stages: sample preparation (green), an experiment stage (purple), and data analysis (orange). The sample preparation requires careful control of sample conditions to avoid errors in amide protection measurements and unintended changes to IB structures. With such control in mind, growth and IB preparation conditions can be intentionally varied and the effects on the structure of IBs can be observed. The experimental stage involves the acquisition of NMR spectra, with ¹H-¹⁵N HSQC spectra acquired for IB structural analysis, and 3D spectra acquired for sequence specific resonance assignment. The third stage, data analysis, involves the quantification of cross peak signal intensities of ¹H-¹⁵N HSQC spectra to provide high-resolution structural information of the IB aggregate.

In the sample preparation stage, there are multiple experimental parameters that may be varied. Cell culture conditions can be adjusted to determine their effects on the structure of the IBs. Conditions of interest may include expressing different variants of the protein of interest, growth temperature, time of protein expression, and addition of protein cofactors such as metals. In contrast, throughout the procedure, there are key steps where other variables must be carefully controlled. For example, during the preparation of the IB sample from the insoluble fraction of the cells, it is important to control the sample homogeneity, IB concentration, and total protein concentration to avoid introducing unintended structural changes that interfere with NMR signal intensities.

The final step in the sample preparation stage is exchange, in which the IB preparation is resuspended in either H₂O or D₂O. Resuspension in H₂O is used for two types of experiments: 1) a fully protonated control experiment and 2) a k_{int} control experiment. The fully protonated control experiment is the acquisition of an NMR spectrum of the fully protonated unfolded protein where intensities are maximal. A k_{int} control experiment measures the intrinsic rate of amide exchange in DMSO for each amino acid in the protein. IBs are resuspended in D₂O for qHDX experiments. Varying the time of resuspension in D₂O

allows different extents of exchange to occur. This measurable rate is known as the observed exchange rate in the aggregate (k_{ex})^[239]. The exchange is stopped, or quenched, by flash freezing the sample in liquid nitrogen followed by lyophilization. To minimize changes to the IB structure as well as signal scrambling, lyophilized samples are stored at -80 °C for less than two weeks.

In the second stage, NMR experiments include acquiring two-dimensional (2D) ^1H - ^{15}N HSQC experiments for IB structural analysis, and three-dimensional (3D) experiments to obtain sequence specific resonance assignments. We note that general information about the extent of protection may be obtained using only HSQCs, but sequence specific resonance assignments of the HSQC cross peaks are a prerequisite for detailed structural interpretation of qHDX data. All experiments involve the solubilization of the IB in a denaturing DMSO buffer. The experiments described here are based on 10 minutes dead time from dissolving a qHDX sample and a 20 minute ^1H - ^{15}N HSQC spectrum acquisition time.

The third stage of the workflow is the data analysis. Acquired NMR data are processed, and the cross-peak signal intensities are quantified to measure both the extent of protection of the aggregate and the k_{int} values of the sample. The workflow may be repeated to examine new experimental conditions. We illustrate all the steps in the stages and representative results in the sections below.

3.3 Materials and Methods

3.3.1. Materials

Solutions were prepared using Milli-Q ultrapure water and chemicals were analytical grade unless otherwise stated. TEN buffer, 10X M9 salts, and M9 supplements are as described in Chapter 2, Table 2.1. Bottles of D₂O were sealed with Parafilm and stored at 4 °C. Ultrapure 99.9% DMSO-d₆ in single use 1 mL glass ampules (Sigma-Aldrich) were used for NMR sample preparation. Stock solutions of 1 M dithiothreitol (DTT) and 5% (v/v) dichloroacetic acid (DCA) were made in DMSO-d₆ and stored at -80 °C. Note that the DTT was frozen in one-use aliquots, as it would otherwise alter the pH of the final solution.

3.3.2. Protein Growth and Expression

Growth protocol is modified from Section 2.2.2 for larger sample volumes and isotopic labeling. Our model protein, human Cu, Zn superoxide dismutase 1 (SOD1), was expressed in BL21(DE3) cells containing a pET 21 vector and grown overnight at 37 °C with shaking at 200 RPM in 50 mL of Luria broth (LB) containing 100 µg mL⁻¹ ampicillin and 30 µg mL⁻¹ chloramphenicol. 30 mL of the overnight culture was centrifuged at 4000g at 4 °C for 10 minutes, and 20 mL of the supernatant was removed. The pellet was resuspended in the remaining 10 mL of the supernatant and used to inoculate 1 L of M9 minimal

media containing the same antibiotics, then incubated at 37 °C with shaking at 200 RPM to an A_{600} of 0.6-0.8. Overexpression was induced using 0.1 mM isopropyl β -D-1-thiogalactopyranoside (IPTG) for expression at 37 °C, or 0.5 mM IPTG for expression at 25 °C. Samples at 37 °C were grown for 4 h post-induction, and samples at 25 °C were grown for 24 h post-induction. The cells were pelleted by centrifugation at 4000g at 4 °C for 20 minutes, flash frozen in liquid nitrogen, and stored at -80 °C. For NMR experiments, duplicate flasks (isotopically labelled and not labelled) are always grown together for each sample type (see 3.3.5).

3.3.3. Preparation of IB Samples

Cell pellets were resuspended in 30 mL of TEN buffer, then flash-frozen in liquid nitrogen and rapidly thawed in a 37 °C water bath for 15 minutes, with gentle swirling every 5 minutes (a freeze-thaw cycle). 3 mg of DNase I (Grade II, Sigma-Aldrich) was then added to each sample and mixed by inversion. The sample was incubated for 45 minutes at 4 °C, then subjected to three more freeze-thaw cycles. Freeze-thaw cycles were chosen as opposed to other lysis methods, such as sonication, to avoid unwanted structural changes to the aggregate ^[245]. The soluble and insoluble fractions were separated by centrifugation at 20,000g at 4 °C for 20 minutes and the supernatants were removed. The pellets were resuspended in 8 mL of TEN buffer to create a homogenous slurry and separated into eight 1 mL aliquots, each containing roughly 4 mg of insoluble SOD1 (confirmed by SDS-PAGE, typically performed for a parallel unlabelled growth). The samples were flash frozen and stored at -80 °C.

3.3.4 Fourier transform infrared spectroscopy (FTIR)

Fourier transform infrared spectroscopy (FTIR) was used as a measure of structural changes that may occur during sample preparation. FTIR experiment and data processing was performed as in Chapter 4. Samples were compared to the average IB FTIR spectra shown in Chapter 4 (Sections 4.3.6 and 4.3.7).

To examine lysis, a cell pellet suspended in TEN buffer (as in 3.3.2) was divided in half. Half of the sample was processed as in 3.3.2 while the other half was lysed with 3 passes through an emulsifier (Emulsiflex C5, pressure changes from 20,000 psi to 8,000 psi) followed by centrifugation at 20,000g at 4 °C for 20 minutes and the supernatants were removed. The pellets were resuspended in TEN buffer and analyzed by FTIR and SDS-PAGE.

To examine the impacts of dilution on IB samples (as is performed for qHDX, see 3.3.4), FTIR was performed on an IB sample diluted in 5 mL of H₂O and portions of the sample were incubated at room temperature (RT) or at 4°C for 1 hour and 2 weeks (the standard incubation times in literature).

3.3.5 Preparation of NMR Samples

IB sample aliquots were thawed at room temperature, centrifuged at 21,000g at 4 °C for 20 minutes and the supernatant was removed. The pellet was resuspended in 5 mL of either H₂O for a protonated control experiment or an intrinsic exchange rate (k_{int}) control experiment, or D₂O for a qHDX experiment. The time of incubation in D₂O for qHDX samples was varied to observe different extents of exchange. The sample was then flash frozen in liquid nitrogen and extensively lyophilized (48 hours). The lyophilized sample may be stored at -80 °C for a short period. We recommend no longer than two weeks, to avoid changes in aggregate structure and signal scrambling.

To ensure the dead time prior to NMR data acquisition is minimized for a qHDX experiment, sample conditions were first determined for a replicate, non-isotopically labelled sample. This replicate will be referred to as the test sample. We find that separate aliquots of a given cell growth pellet provide consistent results in qHDX experiments, therefore one can determine optimal sample conditions on a single aliquot and use these conditions on replicates from the same cell growth. The sample to be analyzed for the experiment will be referred to as the qHDX sample. The test sample was dissolved in 500 µL of solution containing 470 µL DMSO-d₆, 30 µL of D₂O, and 25 mM DTT (from 1 M DTT in DMSO-d₆). The pD of the test sample ($pD = pH_{read} + 0.4$) was then adjusted to 5.5 using 5% (v/v) DCA in DMSO-d₆. The same volume of 5% DCA required for the test sample was used to prepare the qHDX sample. The test sample volume was made up to 600 µL with DMSO-d₆. This test sample was further used to set up NMR experimental conditions. A premixed 600 µL solution for the qHDX sample was prepared containing DMSO-d₆, 30 µL of D₂O, 25 mM DTT, and the volume of 5% (v/v) DCA determined using the test sample. This premixed solution was used to dissolve the lyophilized qHDX sample.

3.3.6. NMR Experiments

qHDX and k_{int} control NMR experiments follow almost the same procedure as for the test sample. The key difference is that the k_{int} control sample does not undergo amide exchange and is lyophilized directly out of H₂O. A fully protonated control experiment was used as a reference for exchange in D₂O samples; this control contains H₂O and so did not undergo amide exchange; thus, the protonated control spectra can be acquired without consideration of dead time. All NMR experiments were performed at a temperature of 19 °C.

Sequence Specific Resonance Assignments

To obtain NMR resonance assignments, protein was grown and expressed as described (see 3.3.2), using 0.2% ¹³C glucose and 0.5 mg L⁻¹ ¹⁵NH₄Cl to make a double-labelled sample. Cell culture was grown at 37

°C for 6 hours post induction and prepared as in 3.3.3 and 3.3.5. Two 1 mL aliquots of double labelled SOD1 were combined to make an NMR sample. For making assignments, five spectra were acquired: HNCO, HN(CA)CO, HNCACB, CBCA(CO)NH, and HN(CA)NNH [246–251]. Spectra were acquired based on standard Bruker parameter sets with small modifications (Table 3.1). Computer aided resonance assignment (CARA)^[252] was used for making assignments. Assignments were readily transferred to other mutants based on the characteristic denatured chemical shifts for different amino acids. Detailed information on strategies for resonance assignment is given in 3.5.3.

Table 3. 2 Experimental parameters for 2D and 3D NMR experiments used in resonance assignment

Experiment <i>Pulse sequence</i>	# dummy scans	# scans	Dimensions	# complex points	Sweep width (ppm)	Carrier frequency (ppm)
¹ H- ¹⁵ N HSQC <i>hsqcetf3gpsi</i>	16	8	¹ H ¹⁵ N	2048 128	13.97 30.00	3.828 113
HNCO <i>hncogp3d</i>	32	16	¹ H ¹⁵ N ¹³ C	2048 48 128	13.97 30.00 9.00	3.828 113 41
HN(CA)CO <i>hncacogp3d</i>	16	24	¹ H ¹⁵ N ¹³ C	2048 44 128	13.97 30.00 9.00	3.828 113 41
HNCACB <i>hncacbgp3d</i>	16	24	¹ H ¹⁵ N ¹³ C	2048 48 128	13.97 30.00 60.00	3.828 113 41
CBCA(CO)NH <i>cbcaconhgp3d</i>	16	16	¹ H ¹⁵ N ¹³ C	2048 48 120	13.97 30.00 60.00	3.828 113 41
HN(CA)NNH <i>hncannhgp3d</i>	16	8	¹ H ¹⁵ N ¹⁵ N	2048 64 64	13.97 30.00 30.00	3.828 113 113

Note: the HN(CA)NNH has an additional delay that must be set. d22 should be set to 12.5 ms to show N_i-1 and N_i+1, or to 6.25 ms to see N_i. For these experiments it was set to 12.5 ms. Some older pulse programs or tutorials call d22 “d27”.

Quenched HDX NMR Experiments

The NMR spectrometer setup was determined using the unlabelled test sample. For the test sample, the spectrometer reference signal was locked to DMSO-d₆, the shims were adjusted, spectral parameters were determined, and a ¹H 1D NMR experiment was performed. Determined parameters would be used for the qHDX sample.

The lyophilized qHDX sample was dissolved through mixing by micropipette in the premixed solution. The spectrometer was matched and tuned to the sample, and the shims readjusted. We have found this

readjustment is generally fast, as the test sample is very similar to the qHDX sample. A ^1H 1D NMR experiment was then acquired for the qHDX sample.

A series of nine 20 minute ^1H - ^{15}N heteronuclear single quantum correlation (HSQC) experiments were initiated after a 10 minute dead time/setup time. After acquiring these spectra, the shim files were saved for acquiring the final ^1H - ^{15}N HSQC after the sample reached H-D exchange equilibrium. The qHDX sample was stored at room temperature for one week, after which three additional ^1H - ^{15}N HSQC spectra were acquired. The entire series of spectra was used to determine the k_{int} of each amino acid residue in the sample, and the first ^1H - ^{15}N HSQC spectrum was used to measure the relative IB protection.

While k_{int} can be determined for each qHDX sample signal decay for the 12 acquired ^1H - ^{15}N HSQC spectra, it is important also to conduct an experiment in which fully protonated protein (rather than exchanged protein) is dissolved in DMSO with 5% D_2O . This k_{int} control experiment is valuable as the sample has higher starting signal and allows for more accurate measurement of k_{int} values, while also permitting the identification of peaks with such high k_{int} that they fully exchange within the dead time of acquiring the first ^1H - ^{15}N HSQC.

Fully Protonated Control NMR Experiments

Fully protonated control samples contain 5% (v/v) H_2O and could be analyzed directly. These control samples were prepared as described in 3.3.5, with H_2O used in the place of D_2O . A ^1H 1D NMR experiment followed by a ^1H - ^{15}N HSQC experiment were performed. The ^1H 1D NMR spectrum is acquired for normalization for sample concentration. This ^1H - ^{15}N HSQC spectrum is used alongside the qHDX ^1H - ^{15}N HSQC spectrum to determine the relative protection of the aggregate (Figure 3.4).

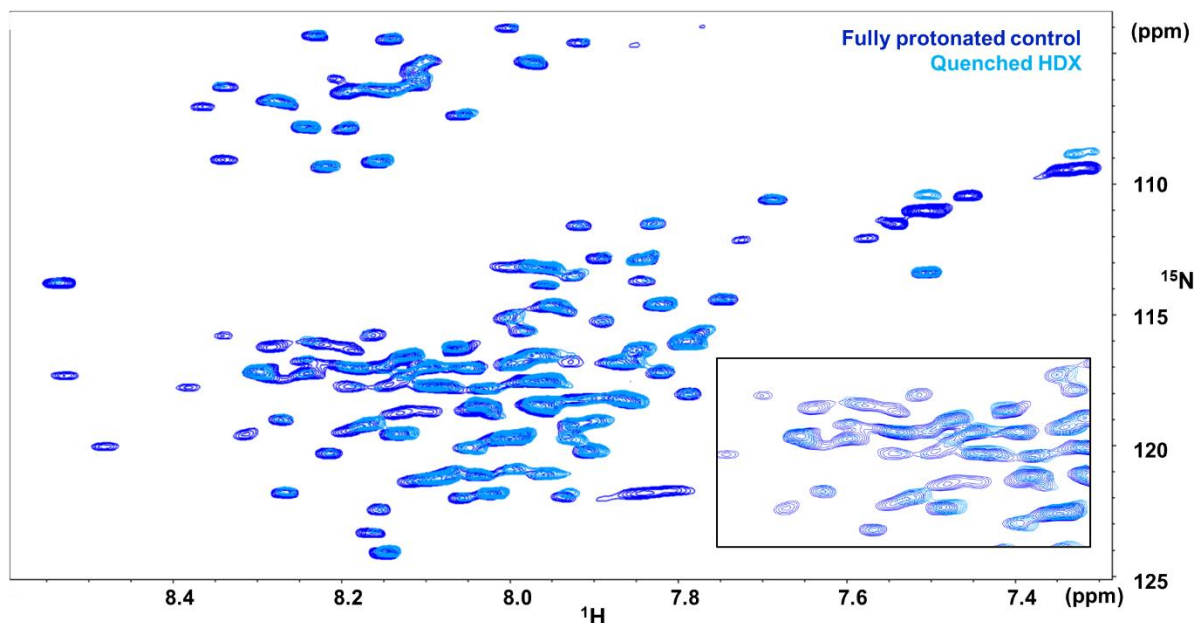


Figure 3.4 ^1H - ^{15}N HSQC spectra for measuring relative protection.

Shown in dark blue is a fully protonated control sample, exhibiting the maximal cross-peak intensities. Shown in light blue is the first spectrum of a qHDX sample that has exchanged with D_2O for one hour, prior to quenching. A ratio of the signal intensities in the qHDX spectrum is calculated relative to the fully protonated control, providing a measurement of the relative protection of the aggregate. Some peaks are not observed in the qHDX spectrum due to H-D exchange. The inset shows an expanded view of a region containing highly protected residues and solvent exposed residues, illustrated by peaks with higher and lower signals in the qHDX spectra, respectively.

3.3.7. Data Analysis

All ^1H - ^{15}N HSQC spectra were processed using Bruker Topspin software with a 90° shifted sine bell squared window function. Processed spectra were analyzed using CcpNMR Analysis Software ^[253], which allows for automatic picking of the cross peaks in a series of spectra, and fitting of these peak intensities as a function of time to a single exponential decay with an offset, to obtain the intrinsic rate constant for exchange, k_{int} , for each amide.

The first spectrum the decay series corresponds to an exchange time in 95% DMSO / 5% D_2O of 20 minutes (10-minute dead time plus 10 minutes corresponding to half of the time to acquire the ^1H - ^{15}N HSQC spectrum). Under suitable conditions, the ratio of the signal from the first spectrum to the signal of a fully protonated control spectrum can provide a quick measure of the extent of protection against exchange, and hence, residue-specific information on the aggregate structure. Further details on the quantitation of amide protection are given in 3.4.3.

3.4 Results and Discussion

3.4.1 FTIR analysis of experimental conditions

Gentle and efficient cell lysis is critical for qHDX experiments on cellular IBs. Some methods have been shown to induce structural changes in proteins^[245]. Freeze-thaw lysis and emulsification are both methods considered to be gentle and therefore were compared for use in qHDX. FTIR analysis of the resulting pellets showed no drastic changes to the IB structure compared to the average of that variant (A4V, average of 6 samples) (Figure 3.5 A, B). However, SDS-PAGE analysis of these samples showed significantly less protein in the pellet – both SOD1 and contaminant proteins (Figure 3.5 C). In the case of SOD1, overexpression is substantial enough such that contaminant proteins are negligent in comparison to SOD1 (see Chapter 2). Thus freeze-thaw is preferred because more sample is obtained from a single preparation and high concentration is a requirement for NMR. However, emulsifying does not appear to interrupt the IB structure, and may be used if preferable for other protein systems.

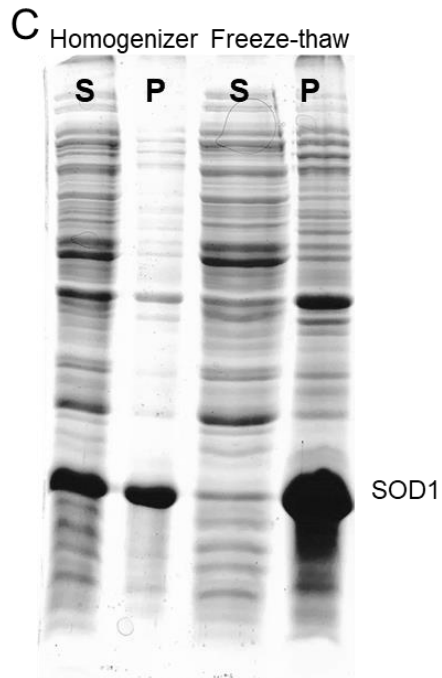
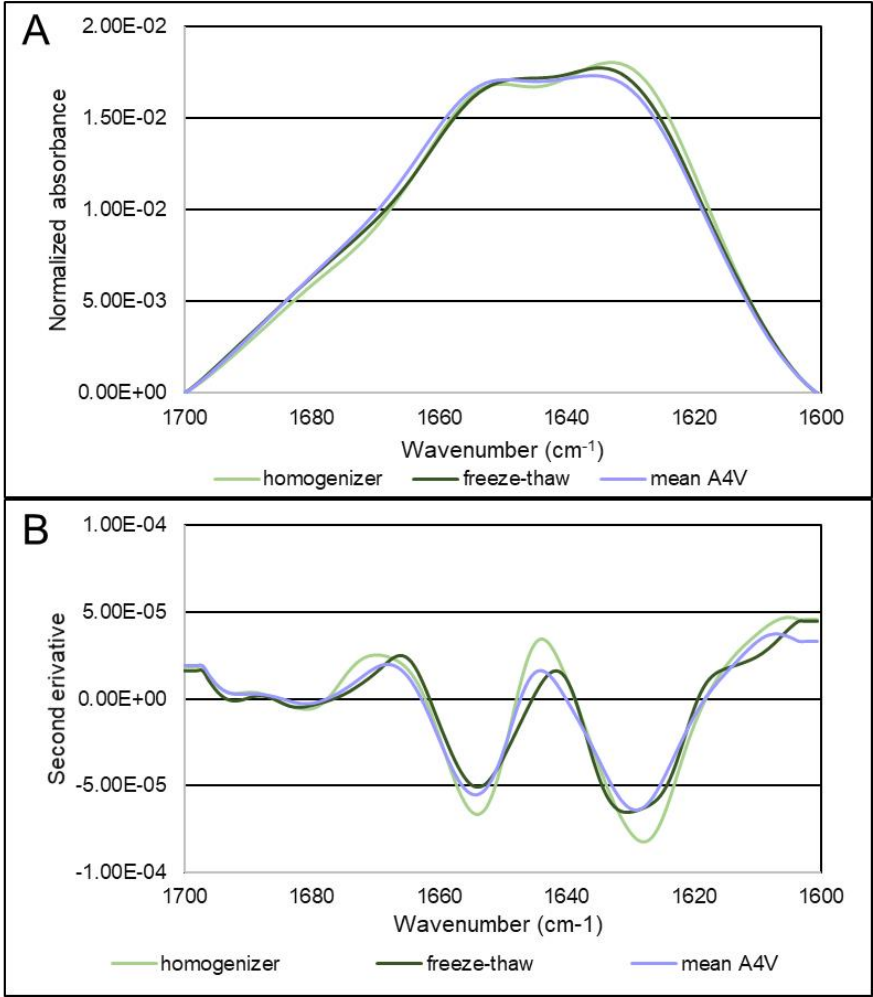


Figure 3. 5 Comparison of freeze-thaw and emulsifying on IB structure and quantity

Absorbance (A) FTIR spectra of cells from a single growth of A4V SOD1 lysed with homogenizer and freeze thaw compared to average of 6 A4V IB samples (lysed by freeze-thaw). Second derivative is shown in (B). No major structural changes are seen. (C) SDS-PAGE comparison of lysis conditions shows significantly more protein in the pellet fraction of the freeze-thaw lysis than the homogenizer lysis. Note that no loading control was used, and that the homogenizer sample was likely diluted more than the freeze-thaw sample due to the volume requirements of the homogenizer. However, relative protein amounts within each method can still be compared.

Previous qHDX studies incubate IBs for extended periods of time to calculate protection factors of each residue in the aggregate. It is possible that the IB structure may change during the long sample incubation, even though some previous work claims that this does not impact aggregate structures^[237]. FTIR analysis of a sample incubated in H₂O for different times or under different conditions can detect whether SOD1 IB samples are stable during incubation. It was found that all conditions examined (1 hour RT, 1 hour 4 °C, 2 weeks RT, 2 weeks 4°C) differed from unincubated IB samples (see Chapter 4; average is of 9 mutants in 3x technical replicate x2 biological replicate). As anticipated, the 2-week RT incubation appears to have the largest change, with an increase in signal at $\sim 1655\text{ cm}^{-1}$, a region associated with unfolded or disordered protein (Figure 3.6A). As previous work suggests that the outside of an aggregate is a less ordered “shell”^[237], this would indicate that these loose proteins are coming off of the aggregate into solution, and a different “shell” is formed. Examination of the second derivative of the spectra reveals small shifts in the two primary peaks in the amide 1 region (Figure 3.6B, Table 3.2). It is notable that the RT incubations deviate more from the average IB but appear stable over time, indicating a structural change on a short time scale (perhaps from a change in equilibrium occurring upon dilution), whereas incubation at 4°C appears to be changing more over time, albeit slowly. It should be noted that amide exchange for SOD1 is slowed 5.75 times at 4°C compared to room temperature^[254], so a sample would need to be incubated ~ 6 longer to achieve a comparable incubation time. Thus, the smaller changes in structure over time seen at 4°C may not prove advantageous once the longer incubation is done.

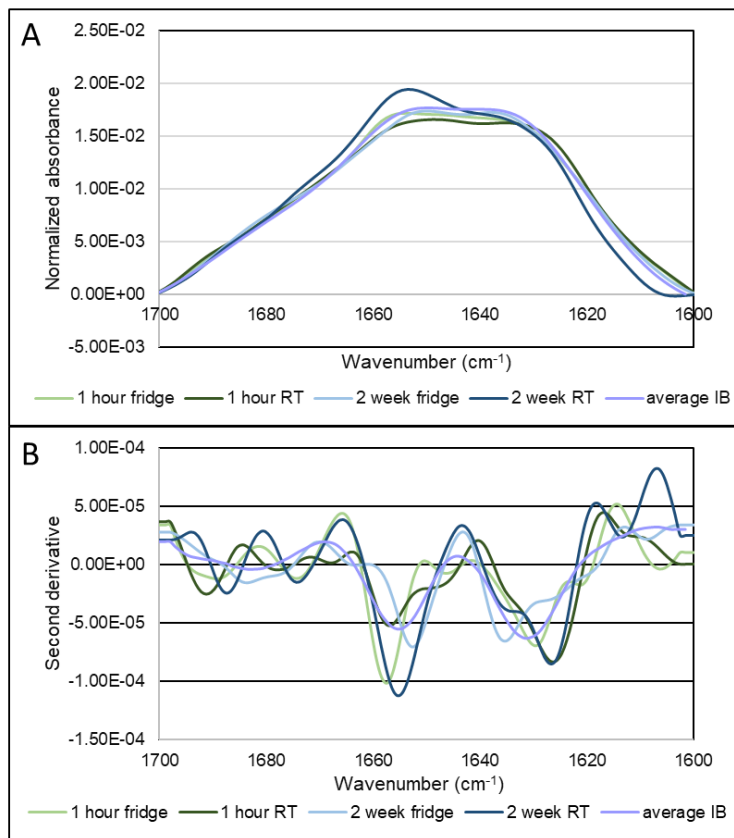


Figure 3. 6 Impacts of dilution and incubation time on IB structure

Absorbance (A) and second derivative (B) FTIR spectra for 4 different incubations compared to undiluted, unincubated average IB show small structural changes to the IB structure, particularly with long, RT incubations. Note that signal-to-noise is decreased in these spectra due to the sample dilution.

Table 3. 2 FTIR peak locations for different sample incubations

	Literature Value, (range)	Average IB	1 hour fridge	1 hour RT	2-week fridge	2-week RT
β region (cm^{-1})	1633 (1623 – 1641)	1631.47	1626.65	1629.54	1626.65	1635.33
Unfolded region (cm^{-1})	1654 (1642 – 1657)	1655.58	1656.54	1657.51	1655.58	1652.69

3.4.2 Control of IB Sample Conditions

The sample conditions throughout qHDX experiments must be carefully controlled to ensure that the signal observed by NMR is representative of the IB structure and to prevent unintended changes to the structure of the IBs, allowing the experiments to be as reproducible as possible. This ensures consistent peak positions and k_{int} between samples. The data acquired from a ^1H - ^{15}N HSQC are typically mapped onto a protein structure to visualize protected or solvent-exposed amino acid residues [239,241]. Provided

pH, water content, and temperature conditions are maintained between different samples, we find both chemical shifts and spectral intensities are directly comparable between mutants.

Measuring Intrinsic Exchange Rates

Control of k_{int} is critical for quantitative peak analysis and comparisons between qHDX experiments. As soon as the lyophilized sample is dissolved and unfolded in 95% DMSO / 5% D₂O, the exposed amides start exchanging further with available deuterons according to their k_{int} values, meaning that the observed signal is lower than its true value. To account for exchange, it is necessary to know the k_{int} for the individual amino acid residues. Provided the value of k_{int} is low enough, the observed NMR signals after the dead time will be close to the signals at the time of dissolution of the IB. Previous results from the Goto group recommended sufficiently low k_{int} values as $9.0 \times 10^{-3} \text{ min}^{-1}$ or lower [239]. For k_{int} of $9.0 \times 10^{-3} \text{ min}^{-1}$, the half-life for exchange is $\sim 77 \text{ min}$; therefore, with a deadtime of 10 min and ¹H-¹⁵N HSQC acquisition time of 20 min, signal intensity at the midpoint of sample acquisition will decrease to $\sim 85\%$ of the original intensity, allowing for a reasonable measure without a correction of the signal at time of dissolution [239]. The value of k_{int} depends on multiple factors: the sample pH, the water content of the sample, the sample temperature, and the amino acid type; thus, it is crucial that these conditions be carefully controlled [239–241].

Values of k_{int} can be measured by acquiring a series of ¹H-¹⁵N HSQC spectra. Here, a sample suspended in H₂O is lyophilized and then dissolved in 95% DMSO buffer with 5% D₂O (Appendix Figure A.3.1.). The extent of amide exchange is monitored by measuring ¹H-¹⁵N HSQC spectra as a function of time. The cross-peak intensities for each residue are measured and fit to a single exponential decay with an offset, where k_{int} is the rate constant of the decay. An equation with an offset is used because the peaks do not decay to zero but rather to a minimal signal based on the water content of the sample. k_{int} values can be used to calculate the signal at the time of dissolution (correcting for exchange in the dead time) using Eq. 3.1:

$$\text{Corrected signal intensity} = A \times e^{(-k_{int}^{av} \times t)} + C \quad \text{Eq. 3.1}$$

where A is the amplitude of the fit of the 12 spectra decay, k_{int}^{av} is the average k_{int} value for a specific residue, t is time sample has been unfolded in DMSO, and C is the offset value of the fit.

Residues with values of k_{int} greater than $9.0 \times 10^{-3} \text{ min}^{-1}$ decay too rapidly for quantitation without correction, which can be made by measuring k_{int} and correcting for signal loss in the experimental dead time, or by taking a ratio of the reference spectrum to the exchanged spectrum [239].

We compared k_{int} values obtained here for IBs of our model protein human SOD1 with values obtained previously for *in vitro* fibrils of pure prion protein (PrP) under similar solution conditions of DMSO, pH, and temperature ^[255] (Figure 3.7). Overall, the k_{int} values are similar between the proteins and between different amino acids, within experimental variation. Some amino acids may have inherently faster or slower exchange, for example, Gly and Asp versus Val and other hydrophobic residues, respectively. The variation in k_{int} tends to be higher for amino acids with higher exchange rates, which likely reflects higher uncertainties in the k_{int} determination due to faster signal decay, and for N-terminal residues, which we find tend to have large fluctuations in signal intensity and consequently k_{int} . Such residues should be excluded from analysis.

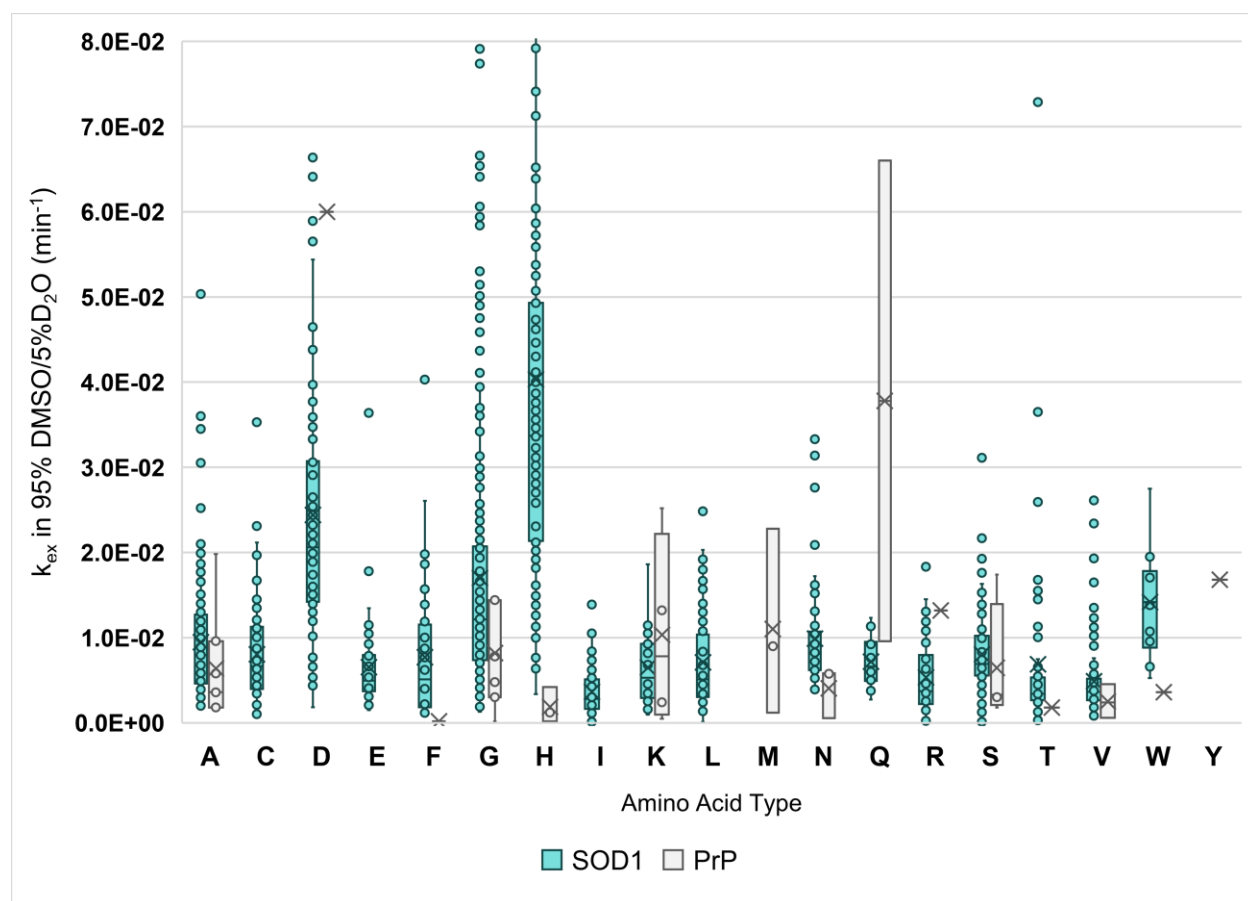


Figure 3. 7 Measured intrinsic rate constants of amide exchange for different amino acids.

Values are plotted for SOD1 (teal) IB samples together with literature values for pure PrP aggregates (light grey) ^[255]. Values include results for 18 SOD1 samples (two biological replicates for each of nine different mutants) and a single mutant (P101L) for PrP ^[255]. Data are shown as box plots, with horizontal line at the median values and 'x' at the mean values. Boxes correspond to the interquartile ranges. Individual data points are shown as circles. Box plots were selected to visualize the distributions and ranges of rates for each amino acid type. For residues with one data point the mean and median are coincident. For residues with two data points, edges of the box correspond to the two data points. Note that the N-terminal Q residue is included in the plot for PrP; we have found N-terminal residues to exhibit highly variable signals, which may account for the large range in the values in this case.

Sample pH

The pH of the sample influences both the values of k_{int} and the observed chemical shifts (Figure A.3.2). As the sample in qHDX experiments contain D₂O rather than H₂O, the acidity of the sample is measured as pD ($pD = pH_{read} + 0.4$)^[241]. Previous studies showed that the k_{int} of amide protons in 95% DMSO:5% D₂O is minimized at a pD of ~5.5^[241]. The pKa of dichloroacetic acid (DCA) in 95% DMSO:5% D₂O is 5.72, which makes it suitable for buffering at the target pD of 5.5 in order to minimize k_{int} ^[241]. Trifluoroacetic acid can be considered if DCA does not sufficiently solubilize the sample^[237,256], however due to its pKa of 3.45 in 95% DMSO^[257], it does not sufficiently buffer at pD 5.5. After endpoint ¹H-¹⁵N HSQC spectra are obtained at equilibrium, the pD of the sample can be accurately measured using a glass pH probe with a ceramic reference junction. It is necessary to use this specific type of pH probe, as DMSO will damage many other probes, particularly standard epoxy. The measurements are acquired after the final spectrum to prevent the addition of salt and water through the reference junction of the pH probe, which experiments on non-isotopically labelled samples have shown can readily ruin the sample.

The pD of the sample can also be assessed by using DCA as an internal reporter. A ¹H 1D NMR spectrum of DCA contains a strong lone singlet peak at approximately 6.5 ppm in 95% DMSO at pD 5.5, corresponding to the acetyl group proton of DCA. A linear relationship is observed between the chemical shift of this peak and the measured pD of the sample (Figure 3.8A). This can be used to assess the approximate pD of the sample before acquiring the ¹H-¹⁵N HSQC series (Figure 3.8B, C), eliminating the need to use the pH probe and keeping the dead time between sample dissolution and NMR data acquisition at a minimum. As k_{int} is dependent on sample pD, it is key that this variable is controlled in these experiments.

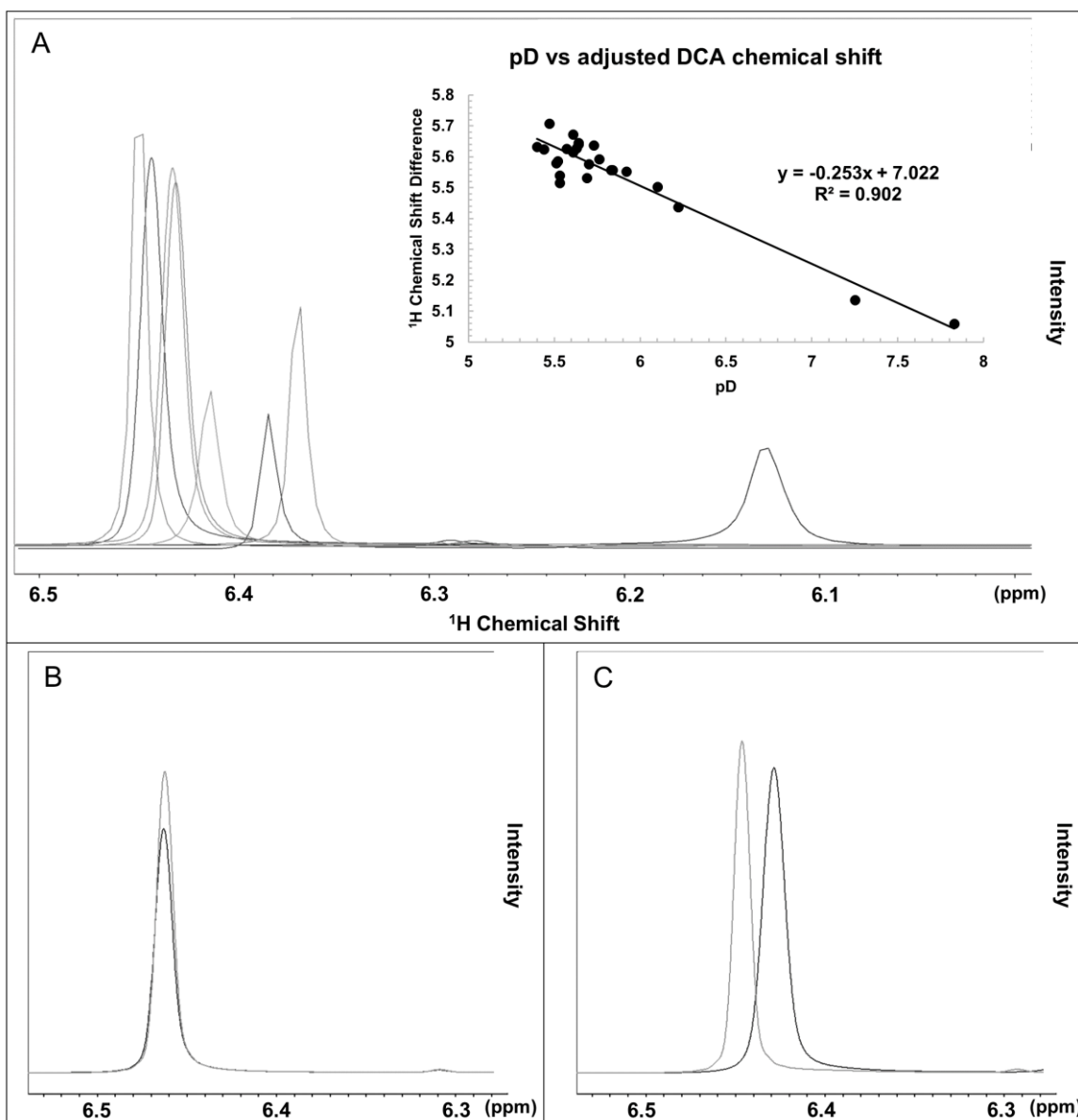


Figure 3. 8 DCA methyl ^1H peak as an internal indicator of sample pH.

A) ^1H 1D NMR spectral overlay between qHDX samples at different pD values, zoomed in to encompass the peak corresponding to the DCA addition. Its adjusted chemical shift, being the difference in chemical shift between the DCA peak and the most upfield peak in the methyl region, is plotted against measured pD and a linear relationship is observed ($R^2 = 0.902$, $p < 0.00001$). B) An ideal overlay between the setup test sample (light grey) and the ^{15}N labelled qHDX sample (dark grey). Overlap between the two peaks indicate near identical pD values. C) A non-ideal overlay between the setup test sample (light grey) and the ^{15}N labelled qHDX sample (dark grey). The lack of overlap indicates that the sample pD is different than that of the setup test sample.

Sample H_2O Fraction, χ_{H}

The fraction of residual water ($\chi_{\text{H}} = [\text{H}_2\text{O}] / \{[\text{H}_2\text{O}] + [\text{D}_2\text{O}]\}$) in the sample must be controlled for two key reasons. The first is the dependence of protein chemical shifts on χ_{H} , and the second is the back

exchange of sample amide deuterons with water protons. As illustrated in Figure 3.9, there are clear differences in chemical shift between two spectra of samples that differ only in their water fraction.

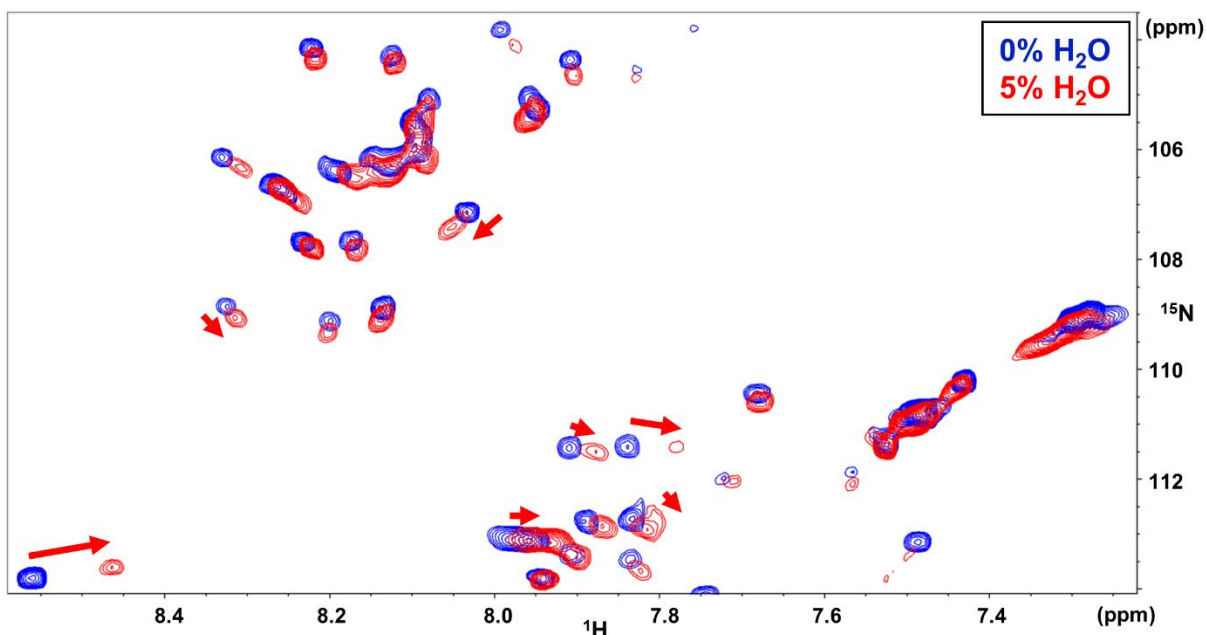


Figure 3.9 ^1H - ^{15}N HSQC spectral overlay of a qHDX sample in DMSO with 0% (v/v) and 5% H_2O added

0% water shown in blue and 5% water shown in red. Differences in chemical shifts are observed, therefore water content must be consistent between samples for the purposes of resonance transfer.

Back exchange will occur if χ_H is larger than the fraction amide protected. As back exchange occurs, signal intensities will increase in acquired ^1H - ^{15}N HSQC spectra and could result in data misinterpretation. To achieve sufficiently low χ_H , small volumes of IB aggregates are suspended in large volumes of D_2O for exchange. For example, the 1 mL aliquots of IB aggregates once centrifuged yield a pellet volume of 300 μL . This is suspended in 5 mL of D_2O , resulting in a maximum χ_H value of <0.06 ($\chi_H = 0.300/\{5.0 + 0.300\}$)^[240]. Additionally, to prevent back exchange, 5% D_2O (v/v) is included in the NMR sample to ensure dominant forward exchange. Previous studies have illustrated an χ_H value less than 0.30, combined with appropriate k_{int} and the addition of D_2O , will result in negligible back and forward exchange, providing high quality data without the need for additional corrections^[239,240].

The χ_H can also be minimized through extensive lyophilization. We recommend using a constant time for lyophilization of all samples of ~ 48 hours; this extended period ensures that as much moisture is removed from the sample as possible. Note that lyophilized samples should be used directly - we have observed that the spectra of replicate IB samples analyzed immediately after lyophilization, stored at room temperature, and stored at -80°C all differ from one another, whereas biological and technical

replicates that have been analyzed immediately after lyophilization display high consistency and reproducibility (Figure A.3.3). Variations may be caused by IB structural changes or signal scrambling due to residual water ^[258].

A final key step to ensure the minimization of χ_H is the utilization of a set-up sample to measure the amount of DCA that must be added to the sample. Use of a set-up sample removes the need to check the pD of the sample during sample preparation, which prevents χ_H from increasing dramatically. Combining the steps described above allows χ_H to be suitably low and reproducible to allow for comparisons between experiments.

An initial rough measure of the residual H₂O in the sample can be obtained by acquiring a ¹H 1D NMR spectrum on the qHDX sample. Residual water in the NMR sample will appear as a peak at a chemical shift of about ~3.68 ppm (Figure A.3.4). In exchanged samples where the height of the water peak is greater than those of the DMSO or methyl peaks, the water content is too high (evidence by peak shifts in a ¹H-¹⁵N HSQC spectrum and inconsistent sample signal compared to replicates).

Water content can be measured for samples in several ways once the qHDX sample is at equilibrium (i.e. fully exchanged to endpoint levels). One may take a ratio of the amide region (8.1 ppm – 8.5 ppm) from the ¹H spectrum of the equilibrated HDX sample to the same region from the corresponding fully protonated spectrum, scaled by a ratio of the methyl peaks (0.5 ppm – 0.95 ppm) in each sample as a measure of protein concentration. The ratio of methyl peaks was generally found to be close to 1, indicating comparable protein concentrations in the samples. A summary of water content in SOD1 variants calculated using this method can be found in Figure 3.10. A second method to check water content is to take a ratio of the sum of the peak heights from the endpoint sample to the sum of the peak heights from the corresponding protonated sample. These values are also found on Figure 3.10. An advantage of this method is that it only considers amides, whereas the 1D comparison includes any compounds with chemical shifts in the 8.1 ppm – 8.5 ppm region – for example, some protons of nucleic acids have chemical shifts in the range of 6.5 ppm – 8.5 ppm range, and as these are impure samples, nucleic acids are very likely present. This may explain why the water content from the 1D method appears to be larger than the 2D method in almost all cases.

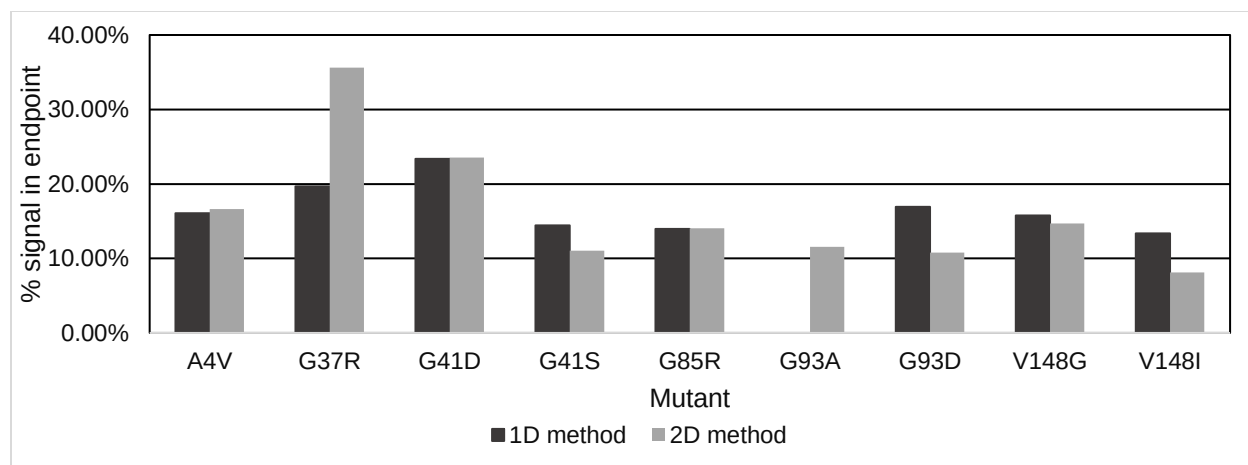


Figure 3. 10 Comparison of water content in qHDX samples calculated by two different methods

Water content in qHDX samples from Chapter 4 calculated using the two methods described above shows general consistency between the methods, with the 1D method trending slightly higher than the 2D method. G93A samples did not have associated 1D spectra.

Temperature

Another factor that affects k_{int} is the temperature of the sample ^[239,241]. Since k_{int} decreases with temperature, it is best to lower the temperature as much as possible, while still maintaining suitable NMR spectral quality in terms of signal to noise and peak resolution ^[239,241]. We recommend a temperature for NMR data acquisition of ~19 °C, the melting point of DMSO, which provides spectra with a reasonable signal to noise and resolution. We further note that peak position is not influenced by temperature since there is no structure, so resonance assigning should be done at a higher temperature for better signal-to-noise (S/N) and resolution (Figure A.3.5). An additional suggestion is to choose a running temperature approximately equal to the ambient temperature of the room containing the NMR spectrometer. This results in less time required for sample thermal equilibration, therefore reducing the dead time prior to acquisition. It additionally ensures that exchange during dissolution and during experimental set-up is occurring at a consistent rate.

Time of IB Exchange in D₂O

Amide exchange of the IB occurs more rapidly with solvent exposed amides, and slower with solvent protected amides. Upon very short incubation, only the solvent exposed amides will have the time to exchange with deuterons, but the solvent protected amides will remain protonated. A single D₂O incubation timepoint provides a measurement of the extent of H-D exchange of the aggregate during the D₂O at for that time. The extent of protection is defined as the ratio between the signal intensities of a qHDX ¹H-¹⁵N HSQC spectrum and a fully protonated spectrum (Figure 3.2, 3.4).

The time of incubation in D₂O can be varied to observe exchange of solvent protected residues (Figure 3.3). By measuring the extent of exchange and fitting to a function of time, the observed rate of hydrogen-deuterium exchange of the aggregate (k_{ex}) can be determined. Measured k_{ex} values are roughly 100-1000 faster than measured k_{int} values, as this exchange happens in aqueous conditions as opposed to DMSO [239,241]. The ratio between k_{ex} and k_{int} is called the protection factor, P [239]. The protection factor differs from the previously mentioned extent of protection, as it is a measurement across a series of timepoints of D₂O incubations. Typical protection factor values observed can range from 10^1 to 10^4 or higher [239,242]. Through varying the duration of the D₂O incubation, protection factors can be determined for each amino acid residue, providing high resolution structural data on the solvent exposed and protected residues of the observed IB aggregate.

3.4.3 Spectral Analysis

qHDX experiments require optimization of spectral processing. Firstly, as the IBs are not purified from a cellular lysate, other proteins and cellular components will also be visible. Secondly, as the protein is unfolded, there is a high degree of overlap in the peaks, which can make some undistinguishable. Thus, the correct balance of S/N and resolution must be found when processing the spectra. qHDX samples are also continuously exchanging their NMR-observable amide protons for unobservable deuterons, therefore the signal diminishes over time. Spectral processing must be performed in a way to consider both S/N as well as signal resolution (see 3.3.6).

Resonance Assignment on Unfolded Protein in DMSO

¹H, ¹⁵N and ¹³C backbone and C β sequence specific resonance assignments for unfolded SOD1 were obtained to 80% completion using a combination of standard 3D experiments for uniformly ¹⁵N, ¹³C - labelled protein (Figure 3.11). It is of note that a large stretch of electrostatic residues could not be assigned despite quality spectra, perhaps hinting at transient or residual association in DMSO. Several spectral qualities were observed throughout the process that may aid in future denatured IB assignment and will be discussed in the following. Along with having only one cross peak per amide, the HNCO spectrum also had the best S/N of the set, and consequently was the most convenient for finding the locations of amide peaks in the crowded regions of the ¹H-¹⁵N HSQC. We found that C α and C β chemical shifts were very similar for a given amino acid for the DMSO-unfolded proteins, and that sequence did not influence these shifts. Ala, Cys, Phe, Gly, Pro, Ser, Thr, Ile, and Val each have characteristic C α /C β chemical shifts which can be used to easily identify amino acid type. Carbonyl shifts were less characteristic for amino acid type but were frequently essential for distinguishing assignments for

Table 3. 3 Comparison of SOD1 chemical shifts to random coil (RC)

Residue	¹³ C=O		¹³ Cα		¹³ Cβ		¹⁵ NH		HN	
	SOD1	RC	SOD1	RC	SOD1	RC	SOD1	RC	SOD1	RC
Ala	173.16	177.8	48.95	52.5	17.92	19.1	122.50	123.8	8.09	8.24
Arg	172.01	176.3	52.80	56	28.86	30.9	119.30	120.5	8.06	8.23
Asn	171.46	175.2	49.69	53.1	37.24	38.9	118.27	118.7	8.21	8.4
Asp	171.27	176.3	50.03	54.2	36.25	41.1	116.73	120.4	8.18	8.34
Cys	170.41	174.6	55.50	58.2	26.33	28	116.68	118.8	8.03	8.32
Gln	172.94	176	52.41	55.7	27.61	29.4	118.26	119.8	8.05	8.32
Glu	171.90	176.6	52.64	56.6	27.24	29.9	118.49	120.2	8.04	8.42
Gly	169.85	174.9	42.25	45.1			106.46	108.8	8.14	8.33
His	170.56	174.1	52.17	55	27.31	29	117.35	118.2	8.28	8.42
Ile	171.78	176.4	57.82	61.1	36.51	38.8	118.63	119.9	7.91	8
Leu	172.89	177.6	51.81	55.1	40.54	42.4	119.40	121.8	7.96	8.16
Lys	172.19	176.6	53.00	56.2	31.24	33.1	119.16	120.4	8.01	8.29
Met		176.3		55.4		32.9		119.6		8.28
Phe	171.55	175.8	54.60	57.7	37.50	39.6	117.59	120.3	8.09	8.3
Pro		177.3	60.02	63.3	29.14	32.1		-		-
Ser	170.82	174.6	55.47	58.3	61.81	63.8	113.93	115.7	7.89	8.31
Thr	170.63	174.7	59.10	61.8	66.82	69.8	112.37	113.6	7.91	8.15
Trp	172.47	176.1	53.86	57.5	27.91	29.6	121.34	121.3	8.09	8.25
Tyr		175.9		57.9		38.8		120.3		8.12
Val	171.59	176.3	58.46	62.2	30.59	32.9	116.12	119.2	7.86	8.03
Rsq		0.721		0.990		0.996		0.934		0.542

Data shown are for uniformly ¹⁵N, ¹³C labelled A4V SOD1 in 95% DMSO / 5% H₂O, pH 5.50, wherein ¹³C=O refers to carbonyl chemical shift, ¹³Cα refers to alpha carbon chemical shift, ¹³Cβ refers to beta carbon chemical shift, ¹⁵NH refers to backbone amide nitrogen chemical shift, and HN refers to backbone amide proton chemical shift. All spectra were acquired at 25 °C. All chemical shifts are displayed in parts per million (ppm). Different amino acids are shown to have characteristic Cα, Cβ, and N chemical shifts, allowing for assignment of the unfolded protein in DMSO. Random coil (RC) chemical shifts are as described by Wishart and coworkers, referenced to internal DSS at 25 °C, pH ~5.0^[259]. Pearson correlations (rsq) between the two data sets are given for each nucleus types.

A valuable spectrum for making assignments was the HN(CA)NNH^[251], which provides connectivity of amide *i* to the ¹⁵N chemical shifts for the *i* + 1 and *i* - 1 amides (no directionality). In unfolded polypeptides, amino acid types tend to cluster by type in the HSQC, with several (Ala, Gly, Ser/Thr)

occurring in distinct regions of the ^1H - ^{15}N HSQC and resulting in (albeit limited) dispersion in the nitrogen dimension (Figure 3.11). This makes the HN(CA)NNH a powerful tool in resonance assignment, frequently leading to a specific sequential amide.

Protein primary sequence is a major determinant on the feasibility of resonance assignments of the unfolded protein, and the utility of the listed set of spectra. In addition to unique $\text{C}\alpha/\text{C}\beta$ or NH chemical shifts for specific amino acids, there are distinct primary sequence effects on amide nitrogen. Residues preceded by Ala resonated upfield in nitrogen relative to others of the same amino acid type, whereas residues preceded by Ile or Val shifted downfield. This trend was checked for other proteins in DMSO and found to hold (Table A.3.1). Thus, proteins rich in these residues, particularly if they occur in the sequence before more “difficult” (i.e. overlapped) amino acid types (Asn, Gln, Glu, Asp, Lys), can be much easier to assign and have the added benefit of having more dispersion/isolated peaks, making final analysis easier as well.

A general difficulty in assigning unfolded protein is sequence repeats. While pairs of amino acids that repeat in the sequence can be discerned in the 3D spectra used herein, longer stretches of repeated sequences may well require additional multi-dimensional experiments to be resolved.

In further comparisons among mutant proteins, we found that mutations in residue i affected the chemical shifts of residues $i+1$ and $i+2$, with a larger effect on the chemical shift of $i+1$. While no consistent trend was found for residues preceded by Gly, we did find that mutation of residue i to Gly resulted in a clear upfield shift of residue $i+1$, a trend that is useful for transferring assignments to mutants (Figure 3.12). We recommend that if residues succeeding the mutation are in a crowded area of the spectrum, residues in that region should not be used for quantitation unless the shifts can be clearly identified, as peaks may be overlapped in unexpected areas (Figure 3.11). A useful check is to compare the intensity of each peak in a protonated spectrum: we found that these are consistent, and outliers in a mutant spectrum may be indicative of shifts in overlapped peaks. In addition to the peak intensities, k_{int} values may also be checked to confirm peak identity, as k_{int} values tend to be consistent per residue between samples.

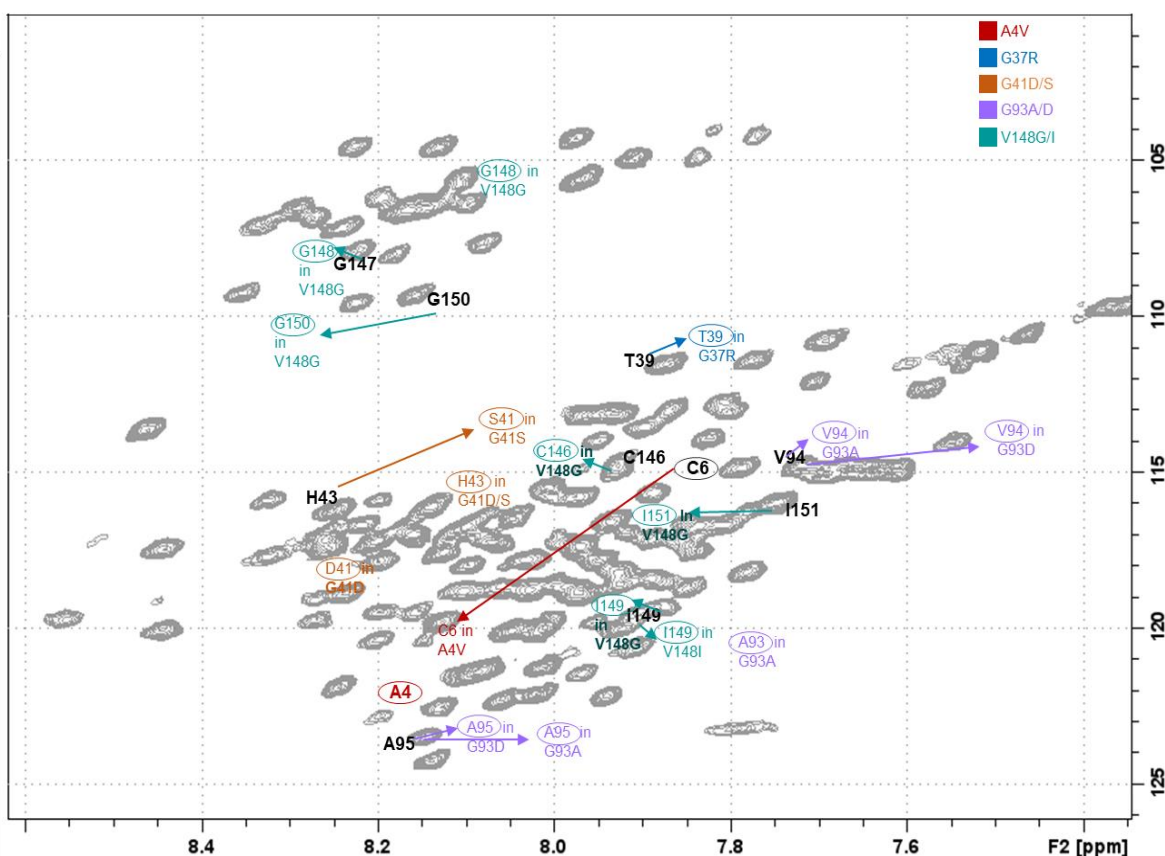


Figure 3. 12 Peak movements in SOD1 point mutations show impact of sequence on peak location

Quantitation of Amide Protection

The signal intensities of peaks in the ^1H - ^{15}N HSQC spectrum report on the structure of the inclusion body sample at the time of dissolution in DMSO. A facile measure of the extent of protection against exchange, and so structure in the IB, is the ratio of the signal from this spectrum and the signal intensities of a fully protonated control spectrum. This ratio will be referred to as the relative protection. Amino acid residues with low relative protection are solvent-exposed in the aggregate, and residues with high levels of relative protection are solvent-protected (Figure 3.4).

Taking the concentration of the protein of interest in NMR samples into account must be done carefully due to uncontrollable variations in overall protein concentration in the IB. Accordingly, we report the fraction of amide i protected in a D_2O sample relative to fully protonated H_2O sample ($\text{FractionProtection}_i$) using the following procedure. The smoothed intensity for each amide in the first ^1H - ^{15}N HSQC spectrum was scaled for protein concentration by dividing by the average intensity of all analyzed amides in the sample.

$$qHDX\ Intensity_i = \frac{Smoothed\ Intensity_i}{Average\ Intensity_{All}} \quad \text{Eq. 3.2}$$

Each amide in every protonated sample is similarly scaled for protein concentration using average intensity of amides in the sample before than being averaged across all samples (Average Protonated Intensity_i). Since the average amide D₂O intensity for a sample is determined by protein concentration and the non-exchanged signal, the extent of exchange must be accounted for or else the fraction protected may be erroneously high ^[239]. Accordingly, the extent of exchange for the qHDX samples was calculated as the ratio of the integrated area of the ¹H 1D amide region (7.5-8.5 ppm) of the D₂O sample divided by the intensity of the corresponding H₂O sample (from the same growth), where these ¹H 1D spectra were scaled for protein concentration using the most upfield peak in the methyl region (~0.81 ppm); the average ratio, *R*, for all SOD1 samples was ~0.5. Thus, fraction amide protected ^[255] was calculated as follows:

$$FractionProtection_i = R \times \frac{qHDX\ Intensity_i}{Average\ Protonated\ Intensity_i} \quad \text{Eq. 3.3}$$

Replicate Samples

Protein aggregation is well established to be variable among both in vitro and in vivo replicate experiments ^[9,226], thus analyses of IB structure by qHDX require averaging the results from repeated experiments. The variability in replicate experiments is illustrated in Figure 3.13. Provided identical sample conditions are maintained, there is outstanding consistency in the observed scaled signal intensities between biological replicates of SOD1 mutants, in both fully protonated and qHDX samples (Figure 3.13A, B). We had additionally found that extent of resuspension in D₂O, or how vigorously the resuspension occurs, does not impact the amide exchange of the aggregate (Figure A.3.6). This is advantageous in that replicates may be performed by different people and still achieve consistent results. We also find that the standard deviation of scaled, fully protonated signal intensity between biological replicate samples is lower than that of all SOD1 mutants analyzed (Figure 3.13C). This means that there is less error between biological replicates compared to other samples. Accordingly, we advise generally to use the fully protonated signal intensities of the same mutant when calculating sample protection, rather than the averaged fully protonated signal intensities of all mutants (Eq. 3.3).

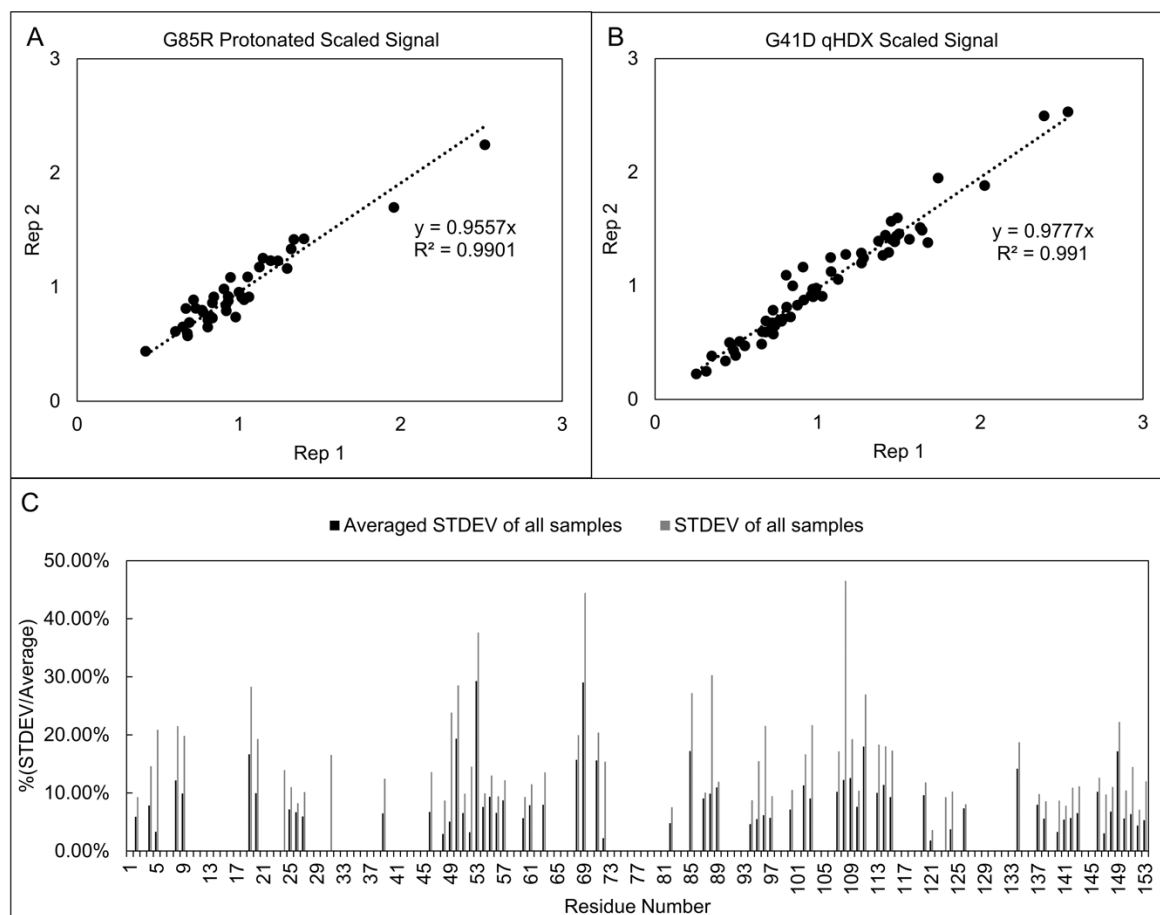


Figure 3.13 Comparisons of replicate qHDX sample results.

A) Biological replicates of G85R SOD1 fully protonated samples. $R^2 = 0.99$, $p < 0.00001$. B) Biological replicates of mutant SOD1 qHDX samples. $R^2 = 0.99$, $p < 0.00001$. C) Comparison of the averaged standard deviation of all biological replicates and the standard deviation of all samples. The standard deviation between each pair of samples in one mutant were measured, then averaged across all mutants. The standard deviation across all samples were also measured. Each value was then divided by the average scaled signal across each mutant and represented as a percentage. We found that the averaged standard deviation values are considerably lower across the entire sequence, indicating that biological replicates have lower deviations from each other than they do to other mutants.

3.5 Conclusions

We report here a detailed and optimized qHDX NMR method for the residue-level structural analysis of IBs, aggregates formed in recombinant bacteria that have many potential practical uses. Quantitative qHDX experiments require very careful control of experimental conditions, including sample pH, water content, temperature, and k_{int} . The results demonstrated here will provide a valuable guide to apply qHDX to explore different recombinant expression systems and aggregates formed by diverse proteins under different conditions. Ultimately, high resolution qHDX structural analysis of IBs promises to illuminate molecular determinants of IB formation and provide valuable information to tailor IB formation in a host of practical applications.

Chapter 4: Quenched hydrogen-deuterium amide exchange reveals native-like structure in SOD1 IBs

4.0 Preface

Chapter 4 applies the validated qHDX method established in Chapter 3 to determine the structure and aggregation mechanisms of SOD1 aggregates for mutations and conditions examined in Chapter 2. In addition, the SOD1 IBs are further characterized by lower resolution methods. The data are then taken together to establish a model of SOD1 aggregation.

This research was conducted in collaboration with Michael V Tarasca under the supervision of Elizabeth Meiering. Dalia Naser and Michael Tarasca contributed equally to experimental design and research and are listed as co-first authors on the publication. Dalia Naser had oversight of this research. For qHDX, Dalia Naser focused on sample preparation, NMR execution, resonance assigning, data normalization, and data interpretation., whereas Michael Tarasca focused on sample preparation, NMR execution, spectral analysis, rate fitting, and data interpretation. Dalia Naser additionally conducted Congo red binding experiments, iodoacetamide SDS-PAGE, and optimized FTIR conditions. Michael Tarasca additionally conducted aggregation and secondary structure prediction and performed growth and sample preparation of IBs under different conditions (data not shown in this thesis). Manuscript writing was done primarily by Dalia Naser with input from Michael Tarasca; NMR figures were made primarily by Dalia Naser. FTIR was performed by Bruna Siebenechler, and fitting/associated figures were done by Bruna Siebenechler aided by Anna Schaefer. Harmeen Deol contributed to data interpretation. Other non-corresponding authors all contributed preliminary or minor experimental work.

Citation: D. Naser⁺, M. V Tarasca⁺, B. Siebeneichler, A. Schaefer, H. K. Deol, T. G. Soule, J. Almey, S. Kelso, G. G. Mishra, H. Simon, E. M. Meiering, *Angew. Chemie - Int. Ed.* **2022**.

A section of discussion on biphasic exchange in IBs was previously published in a review article in the *Journal of Biological Chemistry* co-authored by Dalia Naser, Anna Schaefer, and Bruna Siebenechler. The included section was written by Dalia Naser.

Citation: A. Schaefer⁺, D. Naser⁺, B. Siebeneichler⁺, M. Tarasca, E. Meiering, *J. Biol. Chem.* **2022**, *In press*

Chapter 4 contains additional unpublished research performed by Dalia Naser: urea denaturation experiments (Figures 4.9, 4.10), Congo red of SOD1 grown under different conditions (Figure 4.11), FTIR

of SOD1 grown under different conditions (Figures 4.12, 4.13) and urea denaturation for SOD1 grown under different conditions (Figure 4.14) and associated discussion and conclusions.

4.1 Chapter Abstract

Protein aggregation is central to aging, disease and biotechnology. While there has been recent progress in defining structural features of cellular protein aggregates, many aspects remain unclear due to heterogeneity of aggregates presenting obstacles to characterization. Here we report high-resolution analysis of cellular inclusion bodies (IBs) of immature human superoxide dismutase (SOD1) mutants using NMR quenched amide hydrogen/deuterium exchange (qHDX), FTIR and Congo red binding. The extent of aggregation is correlated with mutant global stability and, notably, the free energy of native dimer dissociation, indicating contributions of native-like monomer associations to IB formation. This is further manifested by a common pattern of extensive protection against H/D exchange throughout nine mutant SOD1s despite their diverse characteristics. These results reveal multiple aggregation-prone regions in SOD1 and illuminate how aggregation may occur via an ensemble of pathways.

4.2 Introduction

Inclusion bodies (IBs) are cellular aggregates that commonly form upon overexpression of proteins in heterologous hosts^[260–262]; these aggregates are of broad fundamental and practical interest^[263,264] as they are important in protein preparation^[265,266], as functional materials^[267–269], and are medically relevant to toxic protein aggregation in disease^[269,270]. While IBs have been characterized using various low-resolution methods, their insolubility and heterogeneity^[270] has left a dearth of high resolution structural data, limiting understanding of the molecular details of their aggregation mechanisms and structures *in vitro* and yet more *in vivo*.

As has been shown in Chapter 3, the obstacle of aggregate heterogeneity can be overcome using quenched amide hydrogen-deuterium exchange (qHDX) monitored by nuclear magnetic resonance (NMR), which has provided structural information at the level of individual amino acid residues for many purified and some in cell globular proteins^[46,47,237]. For IBs, protection against exchange was observed for a small proportion of residues, clustered together in the protein primary sequence and near regions predicted to form amyloid steric zipper assemblies. To date, the effects of mutations in globular proteins on IB structure at high resolution have not been determined.

We focus here on Cu, Zn-Superoxide dismutase (SOD1, Fig. 1A). SOD1 is an attractive system for elucidating molecular mechanisms contributing to IB formation as many mutants associated with familial amyotrophic lateral sclerosis (fALS) promote its aggregation in disease (<http://alsod.iop.kcl.ac.uk/Als>) and their in vivo and biophysical properties have been extensively characterized^[180,271] (see Chapter 2). We analyze a diverse set of fALS mutant SOD1 IBs, from freshly and gently lysed cells, in solution using qHDX and complementary Fourier transform infrared (FTIR) and Congo red (CR) spectroscopies. Remarkably, the mutants all show similar, extensively native-like structure with some amyloid features, despite large differences in protein stabilities and mutation types.

4.3 Materials and Methods

4.3.1 SOD1 50 mL Inclusion Body Expression

Human Cu,Zn superoxide dismutase 1 (SOD1) was expressed from pET21 in BL21(DE3) cells as described in Chapter 2 (2.2.2). The same protocol was repeated for a subset of mutants at 25 °C in minimal media, and at 37 °C in minimal media with the addition of 1 mM ZnSO₄.

Aggregation Propensity of SOD1 in Cells

Aggregation propensity was as described in Chapter 2 (2.2.2).

4.3.2 Iodoacetamide SDS- PAGE

To analyze the thiol status of the SOD1 cysteines in the IBs, iodoacetamide modification was performed^[272]. 1 mL aliquots were removed from each cell culture just prior to harvesting, centrifuged at 8000 g (room temperature) for 1 minute and the pellets resuspended in 30 µL of 0.125 M iodoacetamide (IA) in TEN buffer (20 mM tris(hydroxymethyl)aminomethane (Tris), 1 mM ethylenediaminetetraacetic acid (EDTA), and 100 mM NaCl at pH 8.1). This sample was subjected to one freeze-thaw cycle (freezing in liquid N₂ followed by thawing in a 37 °C water bath), then incubated with 100 µg/mL DNase at 4 °C for 45 minutes. Following incubation, the sample was subjected to three more freeze-thaw cycles. The soluble and insoluble fractions were separated by centrifugation at 21,000 g for 20 minutes and the supernatants were removed. 30 µL of 0.25 M IA dissolved in TEN buffer was added to both the soluble and insoluble fractions. The samples were then analyzed by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE; 15% resolving gel) and stained with Coomassie Blue. The gels were imaged with a BIS303P gel documentation system (DNR Bio-Imaging Systems).

4.3.3 Preparation of IB Samples for NMR

Sample preparation is as described in Chapter 3 (3.3.5).

Amide Exchange Samples

Amide exchange is as described in Chapter 3 (3.3.5).

qHDX NMR Experiments and Analysis

qHDX NMR experiments and analysis are as described in Chapter 3 (3.3.6)

Error may be propagated per residue for each mutant using the standard equation

$$\frac{\Delta z}{z} = \sqrt{\left(\frac{\Delta x}{x}\right)^2 + \left(\frac{\Delta y}{y}\right)^2} \quad \text{Eq. 4.1}$$

where x is intensity of Residue i amide in D_2O sample, y is the intensity of Residue i amide in H_2O sample, z is the ratio of x/y , and Δx , Δy , and Δz are the errors of each component. When data for a residue are averaged across mutants for comparison, the propagated errors are then also averaged:

Average Propogated Error_{Residue i}

$$= \text{Propogated Error}_{\text{Residue } i} \times \text{Averaged propogated errors of all mutants} \quad \text{Eq. 4.2}$$

Resonance Assignments

Resonance assignments are as described in Chapter 3 (3.3.6).

4.3.4 Aggregation and Secondary Structure Prediction

The servers TANGO (sequence-based)^[189], ZipperDB (sequence-based)^[190], and CamSol (sequence- and structure-based)^[191] were used to predict aggregation propensities for each mutant SOD1. The sequence-based servers require only a primary sequence input, while CamSol uses a PDB file (PDB: 2GBU Chain A) for 3D structure-based prediction.

4.3.5 Preparation and Purification of Recombinant SOD1

SOD1 was expressed and purified as described previously^[273,274], modified to include purification using Poros HP2 hydrophobic interaction chromatography. The pure proteins have a pseudo wild-type (pWT) background in which the free cysteines are mutated (C6A, C111A) to prevent their aberrant oxidation

during protein purification and sample preparation; pWT has been extensively characterized and closely resembles the WT in structure, function, stability, and aggregation^[204]. The remaining cysteines (C57, C146) which form a conserved disulfide bond were reduced as described previously^[204].

4.3.6 FTIR Spectroscopy

Attenuated total reflectance fourier transform infrared spectroscopy (ATR-FTIR) was performed following established recommendations^[89,275] using a Tensor 37 FTIR spectrometer (Bruker Optics, Germany) continuously purged with N₂ gas, equipped with a LN-MCT (mercury-cadmium-telluride) detector cooled with liquid N₂. The sample cell is a Bio-ATR II cell with a ZnSe crystal. For all measurements, four spectra were acquired consecutively with 256 scans each with spectral range of 4000-1200 cm⁻¹ and 4 cm⁻¹. For each protein sample, the following series was acquired. A measurement of the background (air) and buffer (TEN) were collected. Next, the SOD1 IB sample (prepared as above) was measured. The crystal was then flushed thoroughly with buffer before collecting post-sample buffer spectra. As recommended, due to clear evidence of sample adsorbed on the crystal despite the extensive flushing, the post-sample buffer was used for liquid water subtraction from the sample spectrum. For each mutant, six independent spectra were measured, comprising two biological replicates with three technical replicates each. Measurements were made at 25°C controlled with a Thermo HAAKDC water bath (Haake).

4.3.7 FTIR Spectral Analysis

All data processing and analyses for SOD1 IB and pure protein samples were performed using the fourth spectrum to ensure thermal equilibrium. Initial processing and analyses of FTIR spectra used the OPUS 6.5 software (Bruker) to perform atmospheric compensation and buffer subtraction. Next, spectra were baseline-corrected by subtracting a straight line fit between 1700 cm⁻¹ and 1600 cm⁻¹, followed by normalizing the area of the amide I band to 1 by using the absolute area of the amide I peak range (1700 cm⁻¹ and 1600 cm⁻¹), using OriginPro 2018. The final average spectrum of each SOD1 mutant was obtained by averaging the six replicates.

Using OriginPro 2018, the second derivative of the final average spectrum of each SOD1 mutant was taken with Savitzky-Golay smoothing (second order with a 7-point window). Local minima were found by searching the second derivative in ranges where secondary structures are known to be found based on literature (1623 cm⁻¹ – 1641 cm⁻¹ for the β region and 1642 cm⁻¹ – 1657 cm⁻¹ for the unfolded region) for minima.

Component peaks in the absorbance spectra were determined by multiple peak fitting in OriginPro 2018. Peaks were fit to a convolution of Lorentzian and Gaussian curve profiles. For a given spectrum, initial peak position estimates for the fitting were obtained based on the corresponding second derivative spectrum. All accepted fits had an R^2 greater than 0.99.

4.3.8 Congo Red Binding

A 300 μ M stock solution of Congo red (CR) was prepared in TEN buffer with 10% ethanol and filtered 3 times through a 0.2 μ m Acrodisc filter. CR-IB assays were conducted using a protocol modified from Klunk *et al.*^[276]. All measurements were made using a Cary 4000 spectrophotometer (300 nm - 700 nm) blanked with TEN buffer. For each sample, a dilution of the IB was made to obtain an absorbance spectrum below 1, bringing IB concentrations to an average of $\sim$ 2 mg/mL. As controls, 20 μ M CR and diluted IB were measured separately before each binding experiment. Each diluted IB was combined with 300 μ M CR for a final concentration of 20 μ M CR and was incubated for 10 mins before being scanned 3 times. Absorbance was unchanged when remeasured after 15 minutes. Differences plotted correspond to the IB+CR solution with diluted IB and blank CR subtracted.

$$\text{CR bound protein} = \text{IB+CR} - (\text{diluted IB}) - (\text{CR}) \quad \text{Eq. 4.3}$$

4.3.9 Urea denaturation of IBs

Mutant SOD1 IBs (20 μ L) were diluted with 20 μ L of assay buffer and then solubilized in 160 μ L of 20 mM Tris-HCl buffer (pH 8) containing different concentrations of urea (from 0 to 6M) for a final volume of 200 μ L in a 96-well plate. Samples were incubated at room temperature for 3 hours with regular shaking for solubilization. Optical density at 350 nm ($\text{OD}_{350 \text{ nm}}$) was measured every 20 seconds for the each sample. The final 20 minutes of readings were determined to have reached equilibrium across the samples, and were averaged for the $\text{OD}_{350 \text{ nm}}$. To account for differing concentrations for each mutant, data were normalized by the turbidity of the control sample with no urea added. Each mutant was run twice on each plate to account for variability associated with pipetting small amounts of IB and potential heterogeneity of IB samples. Urea solubility profiles for each typ of IB were obtained by plotting the average $\text{OD}_{350 \text{ nm}}$ against urea concentration. To estimate relative stability of IBs, data points were fit to a sigmoidal function Eq 4.6 (where c_{mid} is the concentration of urea at which the half the sample is denatured and n is the Hill coefficient) using OriginPro 2018.

$$y = \frac{1}{1 + \left(\frac{c_{\text{mid}}}{x}\right)^n} \quad \text{Eq. 4.4}$$

4.4 Results and discussion

4.4.1 SOD1 aggregation correlates to its biophysical parameters

First, we measured IB formation for 18 chemically diverse fALS point mutations (Chapter 2; Table 2.2) that are distributed throughout the protein structure (Figure 4.1A, *right*) in *E. coli* cells grown in metal-free media at 25 °C, 30 °C and 37 °C. Under these conditions, the IBs are formed by immature SOD1, i.e., with reduced cysteines and unmetalled (apo) (Chapter 2 Results and Discussion, Figure 4.1B). Native reduced apo SOD1 is a marginally stable predominantly monomeric 8-stranded β -barrel with dynamic loops^[204,272,277]. Similar to some proteins^[278,279] but not others^[261,266], for SOD1 there is a correlation between increased IB formation and decreased stability, measured by the apparent midpoint temperature of its thermal unfolding transition (T_m) (Chapter 2; Figure 2.10)^[204]. The T_m values report on global stability but not on accompanying changes in local stability. For example, local perturbations decrease the Gibbs free energy of dimer dissociation (ΔG_d) in mutant apo SOD1 with oxidized disulfide bond^[194]. Notably, increased formation of mutant monomers with disrupted interfaces is also correlated with IB formation (Chapter 2; Figure 2.10). A more thorough discussion of aggregation propensity can be found in Chapter 2 of this thesis.

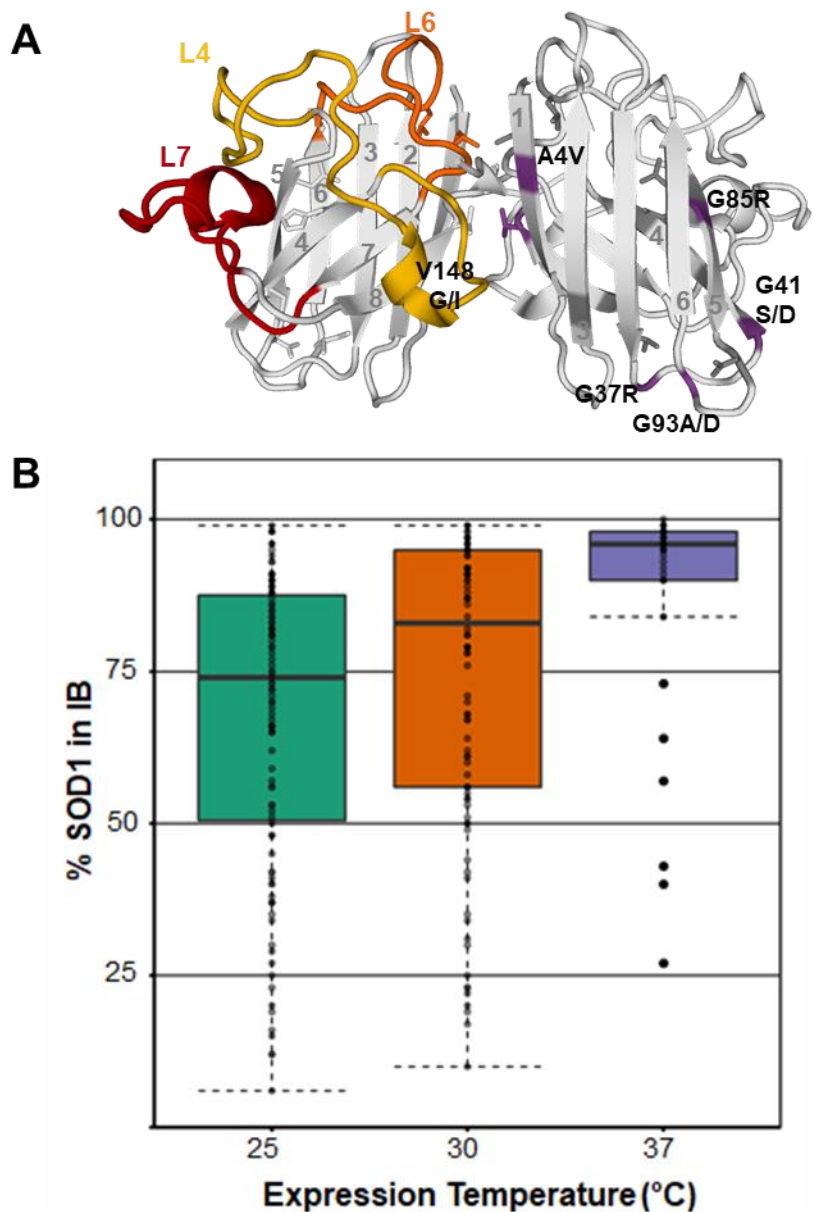


Figure 4.1 SOD1 structure, stability, and IB formation.

(A) Native homodimer, with left subunit highlighting major structural features: b-strands 1-8 are numbered ($\beta 1$ - $\beta 8$) and loops that make intermolecular interactions are coloured and numbered; PDB entry 1HL5. Right subunit shows mutations examined herein: grey/purple analyzed for % SOD1 in IB, purple also analyzed by qHDX, FTIR, and CR. (B) % SOD1 in IB is calculated as described in Chapter 2. Data are shown for samples taken at 4 and 6 hours post induction at 25 °C, 30 °C (both typically in triplicate) and 37 °C (single measurements).

4.4.2 qHDX reveals native structural features

To define the IB structures, we performed qHDX experiments for nine fALS SOD1s (Figure 4.1A, right) containing point mutations that alter residue charge, hydrophobicity, secondary structure propensity, and increase or decrease aggregation propensity; the mutant T_m s range from ~51 °C (for V148I) to ~33

°C (for G37R) (Chapter 2; Table 2.1). The qHDX experiments require a high level of IB formation, thus we measured qHDX for high aggregating mutants at 37 and 25 °C, using resonance assignments for resolved amides (Table A.4.1).

Although IB formation is decreased at lower temperature, in all cases protection is observed throughout the mutant proteins (Figure 4.2A). This extensive protection (Figure 4.2A) is in sharp contrast to the much more limited localized protection reported for IBs of other proteins^[138,237]. Moreover, it is startling that, despite their varying chemical nature and stabilities, the SOD1 mutants have very similar HDX protection.

To elucidate the molecular basis for this protection, we examined experimental and predicted biophysical properties of SOD1 (Figure 4.2B, C). The primary sequence distribution of protected amides in IBs generally resembles that in native reduced apo and holo SOD1, where protection is primarily conferred by intramolecular hydrogen bonding^[280]. These forms of SOD1 exhibit protection throughout the 8 strands of the β -barrel (β 1- β 8) (Fig. 4.2A, B). Protection in the loops connecting the strands is much more pronounced for reduced apo in IBs than expected for the soluble monomer^[277,281]. The loop protection in the IBs resembles that in dimeric holo SOD1, which includes intermolecular hydrogen bonding. Notably, the long loops, 4 and 7, are more extensively protected in the IBs than in the native dimer and despite their strongly hydrophilic character^[197] (Fig. 4.2C).

To further discern the cause of the extensive protection in the IBs, regions in SOD1 with propensity to promote intermolecular interactions and aggregation were analyzed using predictors of structure and aggregation propensity (including TANGO^[189] and ZipperDB^[190]) (Fig. 4.2D). The C-terminal strand, β 8, is generally predicted as the most highly prone to aggregate, followed by β 1. These strands as well as β 2, β 3 and L6 contain short stretches of residues with high predicted propensity to form steric zippers, assemblies consisting of hydrogen bonded antiparallel or parallel b-strands. Previous experimental studies observed zipper formation by short peptides corresponding to β 2, β 3, β 8, and L6, whereas a β 2- β 3 peptide formed a non-amyloid corkscrew oligomer^[282,283]. It is important to note that native structure can modulate or prevent these short sequences from being available to assemble into zippers; also, the steric zipper regions in SOD1 comprise only a fraction of the many protected residues in the IBs.

We therefore considered additional contributions of native-like structure in IBs. A model based on the β 2- β 3 peptide oligomer proposed that much of the rest of the SOD1 sequence could adopt native-like structure. Moreover, multiple experimental studies show full-length SOD1 can form various native-like

assemblies. “Amyloid-like” filaments of near-native apo SOD1 dimers in crystal structures involve intersubunit packing and hydrogen bonding between Loop 7 and $\beta 5$ - $\beta 6$ ^[284–286], regions that also show HDX protection in IBs (Fig. 4.2E). Furthermore, NMR found that reduced apo monomers of wild-type and fALS mutant SOD1s can form the native dimer (D1) and 2 non-native dimers (a symmetric one, D2, and an asymmetric one, D3) in solution^[225,277] (Fig. 4.2F). Importantly, the interfaces for these dimers include intermolecular interactions of loops Loop 4, Loop 6 and Loop 7, which are also protected against exchange in IBs (Fig. 4.2F, G).

The altered HDX protection in V148G SOD1 (Fig. 4.2A, H) further supports contributions of D1, D2 and D3 to IB structure. V148G is situated in $\beta 8$, packing in the center of the native D1 dimer interface (Fig. 4.1A), which is much weakened by the V148G mutation^[194]. Increased dissociation of native D1 would decrease protection in $\beta 8$ and Loop 4, and the non-native asymmetric D3 interface is also likely to be disrupted by V148G thereby decreasing protection of Loop 7^[277]. Thus, the relative protection of L6 in symmetric D2 may be increased, as observed in V148G.

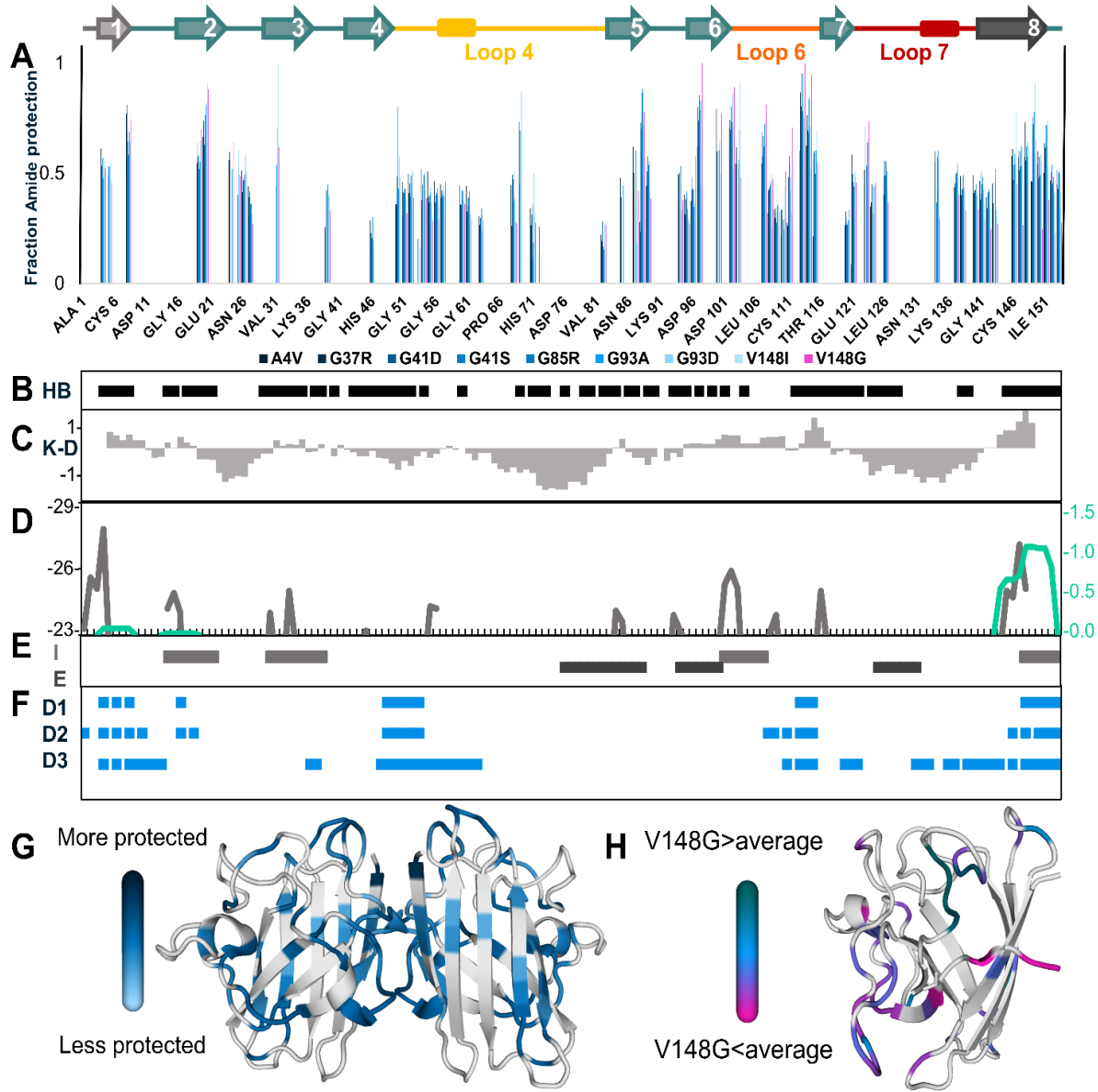


Figure 4. 2 Structure of mutant SOD1 IBs from qHDX analysis

(A) Fraction amide protected for 9 SOD1 mutants. Each bar is an average of two biological replicates. Gaps indicate areas of no data. (B) WT holo SOD1 traditional amide HDX protection^[280]. (C) Kyte-Doolittle hydropathy index of WT SOD1^[197]. (D) Tango^[189] (green) and ZipperDB^[190] (grey) predicted aggregation regions in WT SOD1. More negative values correspond to higher predicted aggregation propensity. V148G was predicted to have the largest effect of the mutations studied, markedly decreasing aggregation propensity near the C-terminus. (E) Gain of interaction regions observed for steric zipper peptides by Ivanova *et al.*^[282] (light grey, top) and filaments of native dimers by Elam *et al.*^[284] (dark grey, bottom). (F) Residues that engage in native dimerization (D1) and non-native dimerization (D2, D3)^[277]. (G) Average qHDX protection of all mutants except V148G shown on the holo SOD1 dimer (PDB 1HL5). Protection from 0.5 to 1.0 is colored as light to dark blue, respectively. Grey indicates no data. (H) Ratio of V148G protection relative to the average qHDX protection. Ratio is colored from 0.5 to 1.5 in pink to teal, respectively (PDB 2GBU).

4.4.3 SOD1 IBs exhibit multi-exponential amide exchange

In addition to identifying amide protection patterns, qHDX experiments can be used to study exchange kinetics by measuring protection for different D₂O exchange times. This allows one to calculate a protection factor, and gauge the stability of the aggregate structure ^[46]. It might be expected that this would result in a single exponential decay, as is usually observed for purified proteins. Instead, the exchange time course has fast and slow exchanging components, similar to previous studies and indicating conformational heterogeneity (Figure 4.3, 4.4)^[46,237,238]. Figure 4.4 depicts the decay of a set of residues showing the general trend of an initial fast decrease in intensity followed by a slower decrease in intensity. The data are fit to a biphasic decay to guide the eye, but this is an assumption and may not accurately describe the IB. A reliable fit to a double exponential decay would require considerably more time points. Additionally, while the data points do appear to have plateaued, the samples were shown to not be fully exchanged upon dissolution in DMSO. This indicates that there may be an additional, even slower phase that is not being measured in the time points of this experiment. As shown in Chapter 3, incubations in D₂O for extended periods of time may result in structural changes, and results from longer time points may not be reliable. Consequently, longer incubations were not conducted, and calculated exchange rate may not be reliable. For reference, they are included in Table A.4.3.

While a single exponential decay is observed in some cases of qHDX on aggregates ^[47,287], a surprising number of the studies show a biphasic or double exponential decay. Yamaguchi *et al.* originally reported that the majority of residues in the β 2-microglobulin fibrils could not be well fit using a single exponential function, but instead were characterized by a period of fast initial exchange followed by a slower phase or a plateau ^[288]. This biphasic behavior was explained by a model proposing that the entire fibril was ordered, with the outer edges of the fibrils exchanging quickly. The same residue would also be present in an inner face of the fibril, protected from solvent, resulting in a slower exchange. The combination of these two environments results in biphasic exchange. Wang *et al.* reported similar biphasic behavior for each of several different protein IBs ^[48]. However, they hypothesized that the fast-exchanging population corresponded to amorphous structure coexisting with slow exchanging, homogeneous, amyloid-like structure. Interestingly, similar biphasic behavior was also observed in qHDX NMR studies of the HypF protein ^[138], despite the various aggregates in the study having very different structures, suggesting that despite varying aggregation mechanism, in general, aggregated proteins may retain a disordered, amorphous, or perhaps misfolded “shell”. Similarly, fast and slow exchanging

populations were also observed in the SOD1 IB studies, suggesting different parts of structures were exposed in a heterogeneous aggregate. Unfortunately, qHDX has revealed only the presence and not the structure of this fast-exchanging population, as it is very difficult to monitor and characterize. Furthermore, the multiexponential exchange is an important piece of information that sheds light on why there may be disagreement between results obtained with different structural methods across different aggregate structure studies.

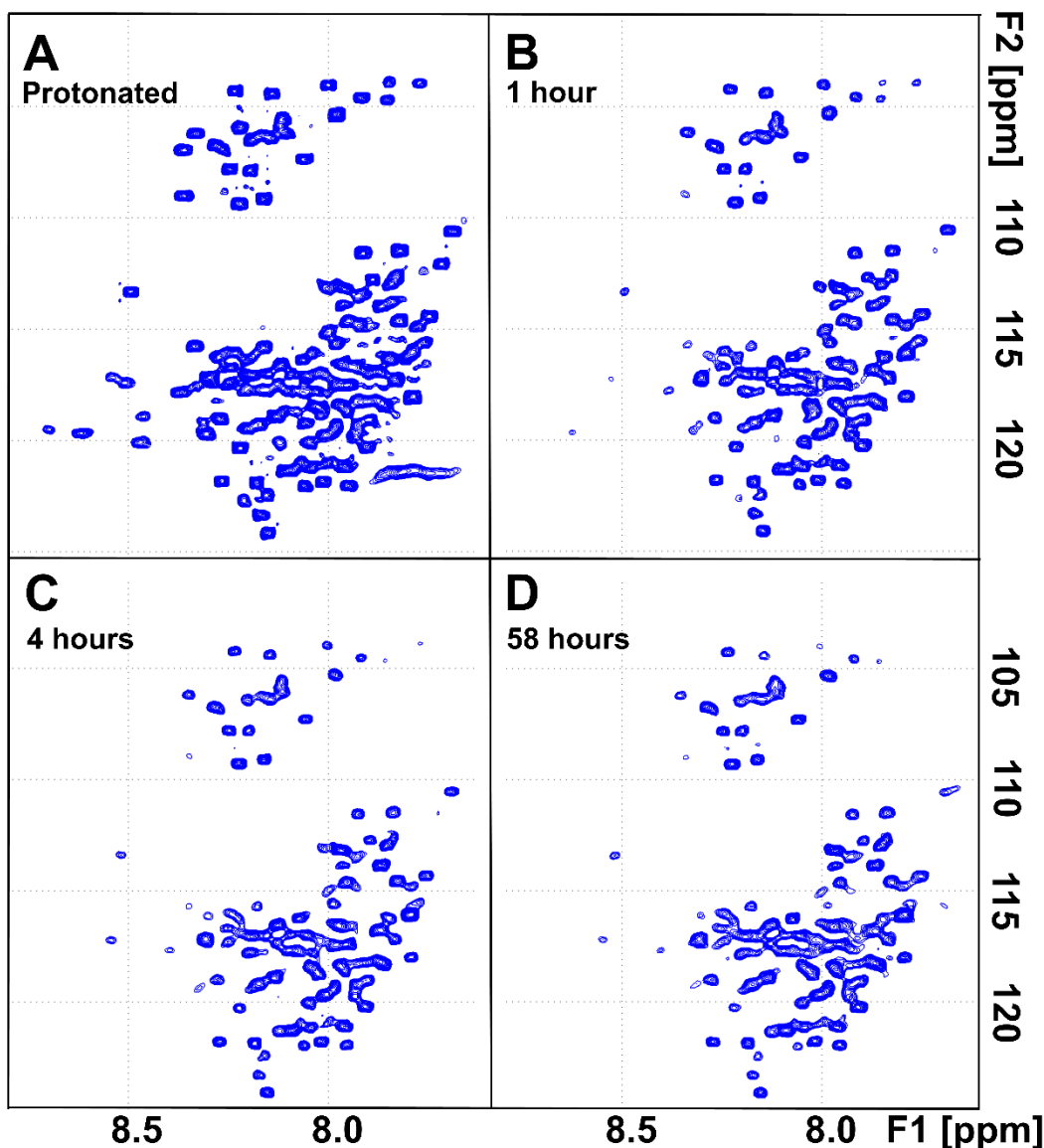


Figure 4.3 ^1H - ^{15}N HSQC spectra for aliquots of G85R SOD1 IBs at different times of exchange in D_2O .

A: Fully protonated, not incubated in D_2O (0 minutes); B: Exchanged for 1 hour (60 minutes) in D_2O at room temperature; C: Exchanged for 4 hours (240 minutes) in D_2O at room temperature; D: Exchanged for 2 weeks at 4°C in D_2O (converted to 58 hrs, i.e. 3480 min at 20°C using intrinsic exchange rates calculated through SPHERE (A Server Program for Hydrogen Exchange Rate Estimation)^[289,290], a server which calculates aqueous exchange rate for each residue in a sequence under a set of temperature/pH conditions))

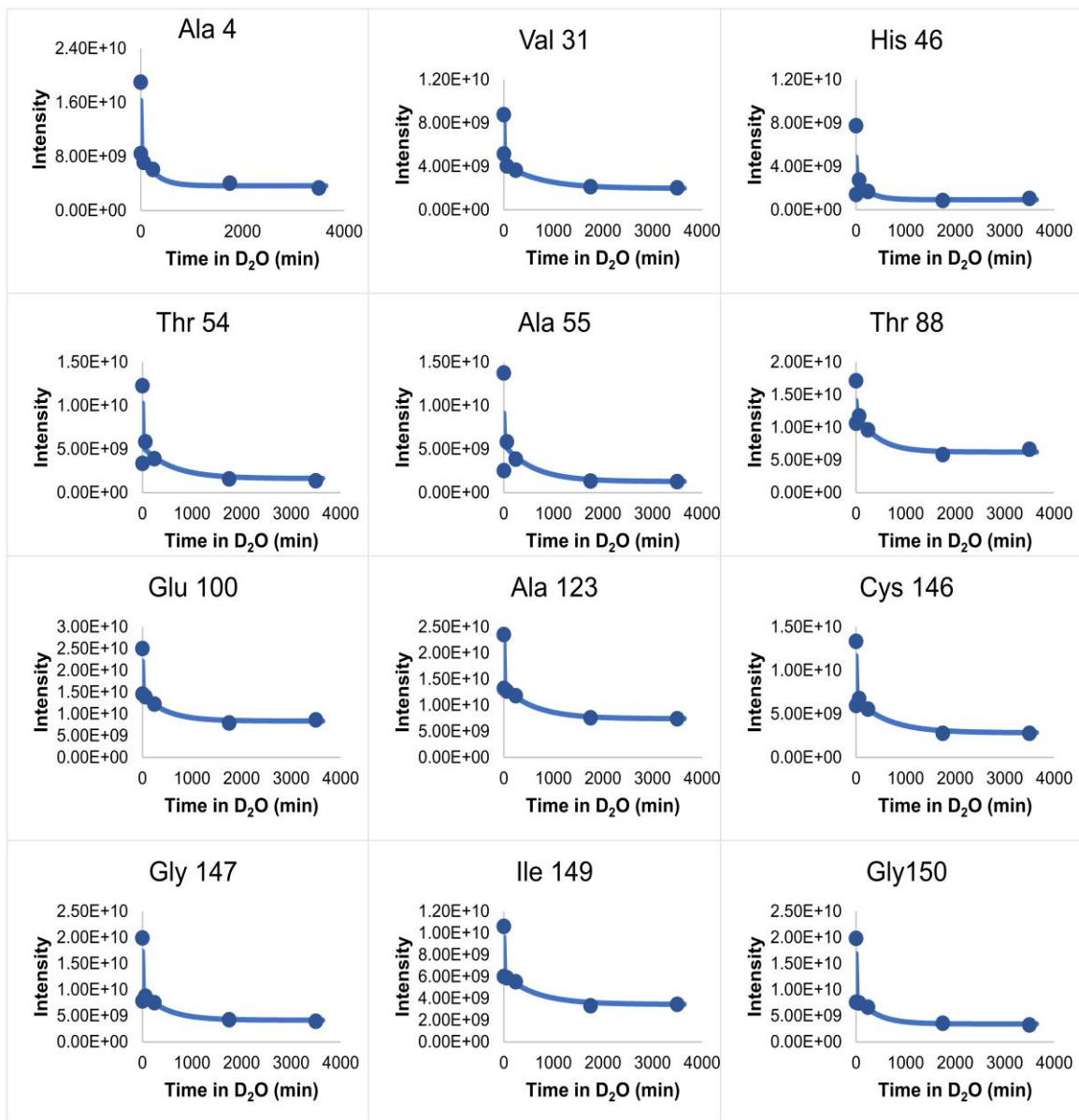


Figure 4. 4 IB amide protons show fast and slow exchanging components in D₂O

Aliquots of IBs from the same culture expressing G85R SOD1 grown at 37 °C and harvested at 4 hours post-induction were incubated in D₂O for 0 (protonated), 2, 60, or 240 minutes at room temperature, and for 1 week or 2 weeks at 4 °C (converted to 1750 and 3480 minutes, respectively at 20 °C using SPHERE^[289,290]) (Fig. S5). To guide the eye, solid lines show fits of the biphasic time course of exchange in D₂O to a double exponential decay with an offset, Amide Peak Intensity = $A_1e^{(-k_1 \cdot t)} + A_2e^{(-k_2 \cdot t)} + \text{offset}$, using OriginPro 7. The initial intensity decrease is attributed to fast exchange indicating unstable and/or unstructured species, while the slower exchange is attributed to more stable and hydrogen bonded structured species.

4.4.4 CR and FTIR analysis of qHDX samples corroborates similar IB structures between mutants

Next, we analyzed the IBs using FTIR and CR spectroscopies to measure their secondary structure and amyloid content, respectively. Consistent with the qHDX results, the FTIR spectra of the SOD1 mutant IBs coincide closely (Figure 4.5A, 4.6), demonstrating extensive structural similarities among the mutant IBs. The spectra and secondary structure analyses of IBs are akin to those of native reduced apo and holo SOD1 in solution (Fig. 4.7A, B), in agreement with native-like features in qHDX. In contrast, there are larger differences between the average IB spectrum compared to heat denatured SOD1 and sonicated SOD1, which has characteristic features of amyloid^[291]. Amyloid typically exhibits absorbance at low wavenumbers arising from short intermolecular hydrogen bonds ($<1630\text{ cm}^{-1}$), which is prominent only for sonicated SOD1. Significant absorbance at high wavenumbers ($>1680\text{ cm}^{-1}$) for IBs, native SOD1, and aggregates of reduced apo SOD1 in rats overexpressing human G93A SOD1^[292] is characteristic of antiparallel β -sheet and turns. Conversely, parallel β -amyloid lacks absorbance in this region^[74,293,294], as is also apparent for sonicated SOD1. Curve fitting of secondary structure components agrees with the preceding observations (Figure 4.6).

The FTIR data for mutant SOD1 IBs are consistent with qHDX results showing that the IB structures are largely similar and consistent between replicate samples. We note that detailed interpretation of FTIR spectra is difficult, owing to overlapping bands^[74,293]; however, comparisons of differences between spectra provides a conservative approach that can provide useful structural insights. Second derivatives of absorbance spectra (Figure 4.5B) are widely used to identify different bands present in the absorbance spectra^[89,295]; in IBs, the non- β band at 1654 cm^{-1} may correspond to unstructured/helical structure, while the peak at 1630 cm^{-1} and the broad region between $1670\text{--}1700\text{ cm}^{-1}$ (including signal for loops and a band for turns at 1685 cm^{-1}) are indicative of antiparallel β -structure^[74]. The shift of the unstructured/helix band in native SOD1 at 1646 cm^{-1} to 1654 cm^{-1} with concomitant broadening in IBs may be caused by various altered loop conformations due to packing interactions in IBs. Overall, the FTIR data indicate a lack of extensive amyloid structure, some disorder, and substantial presence of native-like conformations in the IBs.

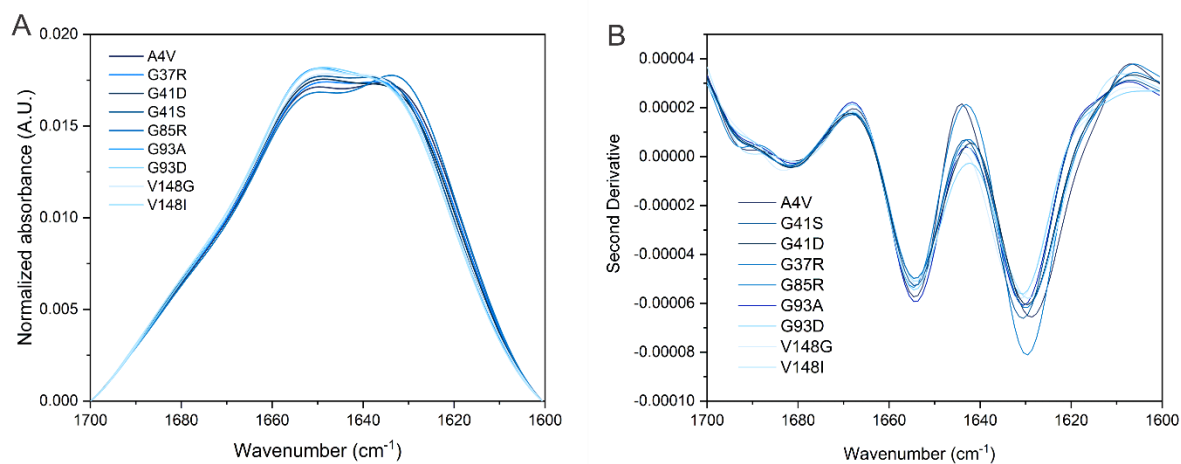


Figure 4. 5 FTIR spectra of mutant SOD1 IBs exhibit similar absorbance and second derivative features

Each spectrum is the average of 3 technical replicates for 2 separate culture growths. Each absorbance spectrum is normalized to an area of 1 (A), before calculating the second derivative (B).

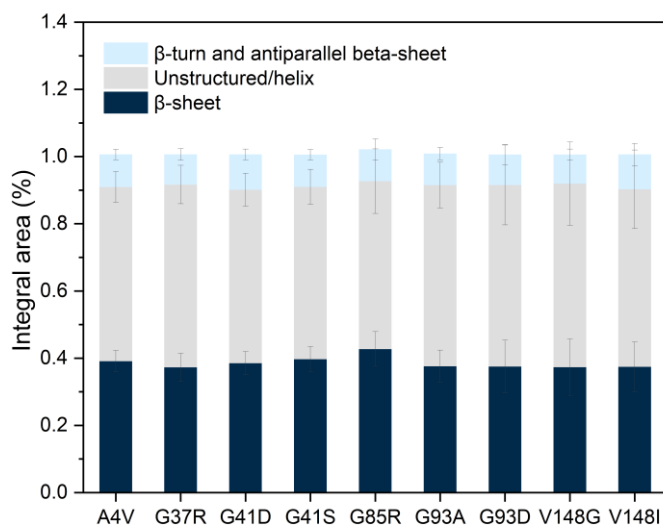


Figure 4. 6 Integrated area of the component bands from curve-fitting of mutant SOD1 FTIR spectra

The stacked bars correspond to the total normalized FTIR absorbance area of amide I band upon fitting of three components around 1630 cm^{-1} (β -sheet), 1654 cm^{-1} (unstructured/helix) and 1680 cm^{-1} (β -turn and antiparallel β -sheet). These peak positions were based on the peaks from the second derivative spectra of the IBs

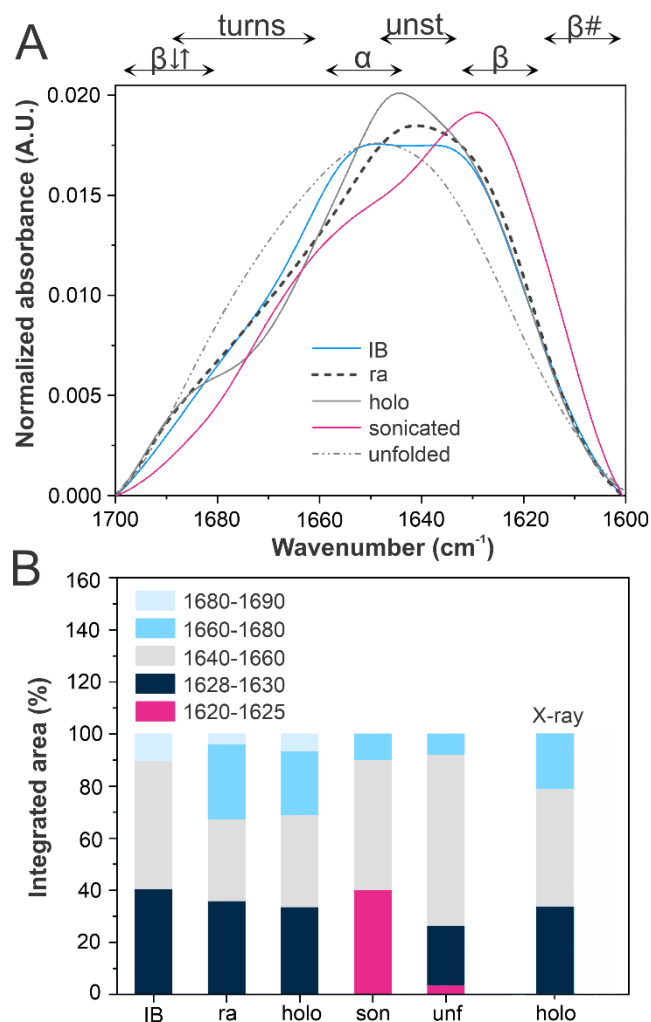


Figure 4. 7. FTIR spectra of SOD1 IBs exhibit native and amyloid features

(A) ATR-FTIR spectra in the amide I region for average IB, pure native reduced apo (ra) and holo pWT, ra A4V sonicated, and heat unfolded at 55 °C. Spectral regions for antiparallel β -sheet ($\beta\downarrow\uparrow$), β -turn, α -helix (α), unstructured (unst), β -sheet (β), and β -amyloid with strong intermolecular hydrogen bonds ($\beta\#$) are indicated. (B) Integrated area of secondary structure bands from curve-fitting analysis and the holo SOD1 x-ray structure^[296].

The binding of CR to amyloid causes a red-shift in the CR absorbance spectrum, with maximal difference at ~541 nm^[276]. Sonicated SOD1 exhibits typical CR spectral features corresponding to amyloid^[291]. The spectra of mutant IBs are muted in comparison (Fig. 3C), and more closely resemble the CR binding of uninduced cells. Importantly, ALS is not an amyloid disease and aggregates of SOD1 in ALS do not have amyloid characteristics^[297]. CR spectroscopy is therefore also consistent with limited amyloid structure in the SOD1 IBs.

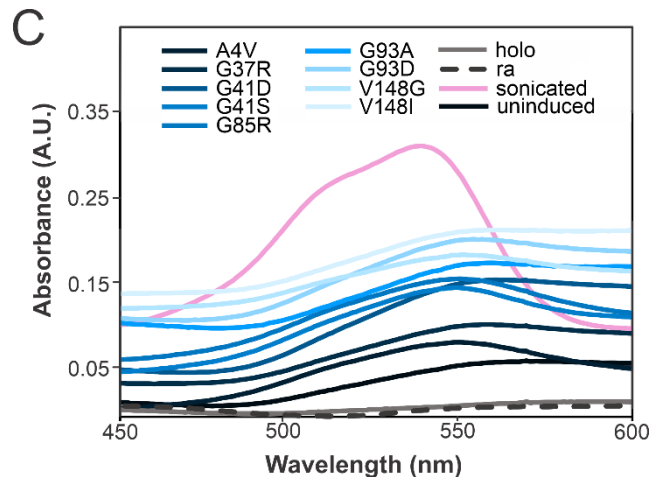


Figure 4.8 CR spectra of SOD1 IBs exhibit native and amyloid features.

CR difference absorbance spectra for mutant IBs, uninduced cells, and pure holo, ra and sonicated amyloid SOD1. Note that because these are “difference” spectra, a flat line or lack of signal indicates no CR binding.

4.4.5 Urea denaturation relates stability of SOD1 IBs to pure protein stability

Urea denaturation is traditionally used on folded protein to assess protein stability. Generally, measurements using a physical technique that can distinguish between the folded and unfolded states are made with different urea concentrations. If measurements are taken at equilibrium, reversibility is possible, and sufficient points are obtained in the pre- and post- transition baseline, the resulting denaturation curve will resemble a negative sigmoid and can be fit to obtain the free energy change, ΔG_U , for the folding/unfolding reaction^[298].

Adapting this method for IBs poses some caveats. IBs are impure and contain varying amounts of other proteins. This drastically limits the detection method for unfolding by urea – traditional methods such as monitoring Trp or Tyr fluorescence signal cannot be applied due to the number of proteins in the sample that may possess these residues. Additionally, the aggregate may contain multiple species in different states of folding, which will omit different extents of fluorescence. Furthermore, multiple processes are occurring in the urea denaturation of an aggregate – the aggregate is coming apart into protein subunits, and the subunits are unfolding. Depending on the structure of these subunits (and the homogeneity of the structure), these may be separate or overlapping events that cannot be deconvoluted. Measuring reversibility is also greatly limited in IB samples.

As a measure of IB stability, sample turbidity measurements at 350 nm were used, with a drop in turbidity interpreted as denaturation of the aggregate^[37]. The reaction is measured over time until

equilibrium is reached, and only the endpoint is used. These measurements were used to determine the relative compactness of packing of the different IBs.

Four different solubilization profiles were seen for the 8 mutants in this study. Two of the profiles (Figure 4.9A, B) show an immediate drop in signal even at the lowest urea concentration used with no pre-transition baseline and different steepness, whereas the other two (Figure 4.9C, D) show a small amount of pre-transition baseline with different steepness in the transition. Data were fit to a negative sigmoid function to help guide the eye and to estimate relative differences between the variants. Fit values can be found in Table A.4.4. SDS-PAGE analysis of the samples revealed that background proteins such as outer membrane protein F (OMP F) do not appear to be denaturing, as was seen in previous studies^[299].

A convenient way to group the different profiles is to take a ratio of c_{mid}/n and group mutants together according to this value. The mutants with the least compact/strongly held together aggregate are G41D and G93D (Figure 4.9A), which have a similarly steep transition, c_{mid} values of approximately 2.1M, and seemingly no pre-transition baseline. The second grouping contains G41S and V148I, with both mutants showing a shallower transition and higher c_{mid} values (Figure 4.9B). The third grouping of A4V, G93A and V148G shows curve that appear to have a very small pre-transition baseline (Figure 4.9C), and the last grouping of G37R and G85R are the variants that most resemble a urea denaturation curve, with a slight pre-transition baseline, a sharp transition, and a clear post-transition baseline (Figure 4.9D). Some of the observed behaviors appear to correlate with ZipperDB scores: G41D and G93D both stabilize their respective regions to a similar extent, G37R and G85R are both the most stabilizing according to ZipperDB, and G41S and V148I similarly destabilize their respective regions. It is possible that the energy changes in these regions correspond to how tightly packed or held together the aggregate may be. However, there is no such trend between A4V, G93A and V148G.

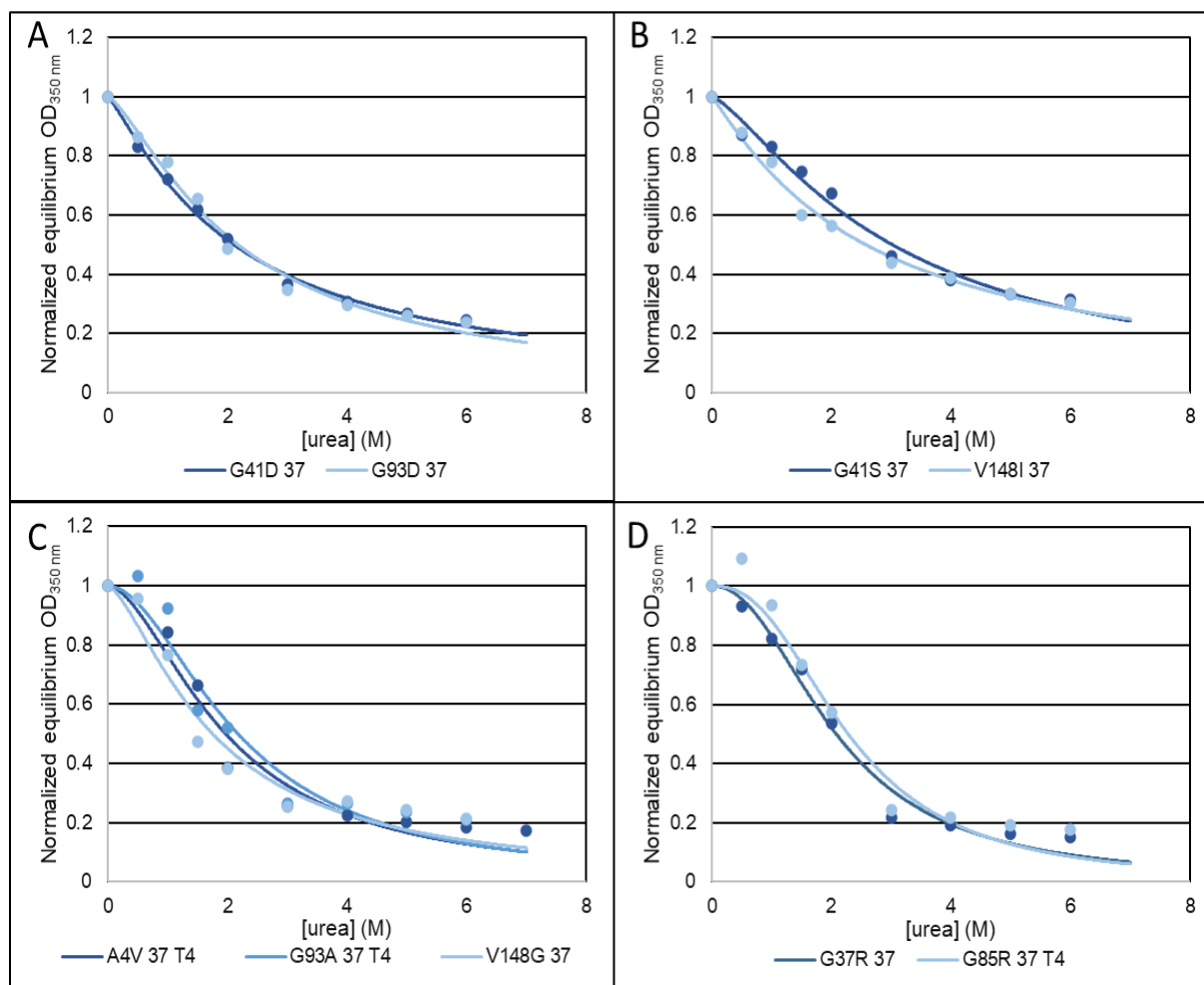


Figure 4.9 Urea denaturation of SOD1 IBs grown at 37°C with no metals

Urea denaturation curves are grouped by c_{mid}/n values, wherein (A) is ~ 1.6 , (B) is ~ 2.2 , (C) is ~ 1.1 and (D) is ~ 1 . Points are the average of 2 technical replicates, line is the fit of the average.

An interesting trend emerges upon comparing the fit c_{mid} values to the T_m of SOD1 variants (Figure 4.10 *top*). When comparing to apoSH T_m , the form of the protein implicated in aggregation, a very strong correlation is seen ($R^2 = 0.8118$, $p = 0.0952$). Unfortunately, this correlation is only defined by a handful of data points due to the difficulty in obtaining apoSH T_m 's. However, the correlation is still present if apoSS T_m 's are used – while it is lower and less significant ($R^2 = 0.4035$, $p = 0.32$); apoSS T_m 's do not scale up perfectly from apoSH T_m 's. It appears that the IBs have a comparable relative stability to their mutant monomers, but we cannot distinctly say that this is because of monomer unfolding, as this is not directly being measured – turbidity is not a measure of protein unfolding but rather change in solubility (ie a monomer may be disaggregated but still folded). This is additionally seen when comparing c_{mid} values to

$\Delta G_{\text{monomer unfolding}}$ and $\Delta G_{\text{dimer dissociation}}$: an extremely strong correlation is seen with $\Delta G_{\text{dimer dissociation}}$, but a robust (albeit less significant) correlation is also seen for $\Delta G_{\text{monomer unfolding}}$ (Figure 4.10B). Since these parameters are related (more strongly and better folded monomers form stronger dimers), they cannot be deconvoluted here. However, the presence of the relationships supports a theory of native/non-native SOD1 dimers being present in the IB.

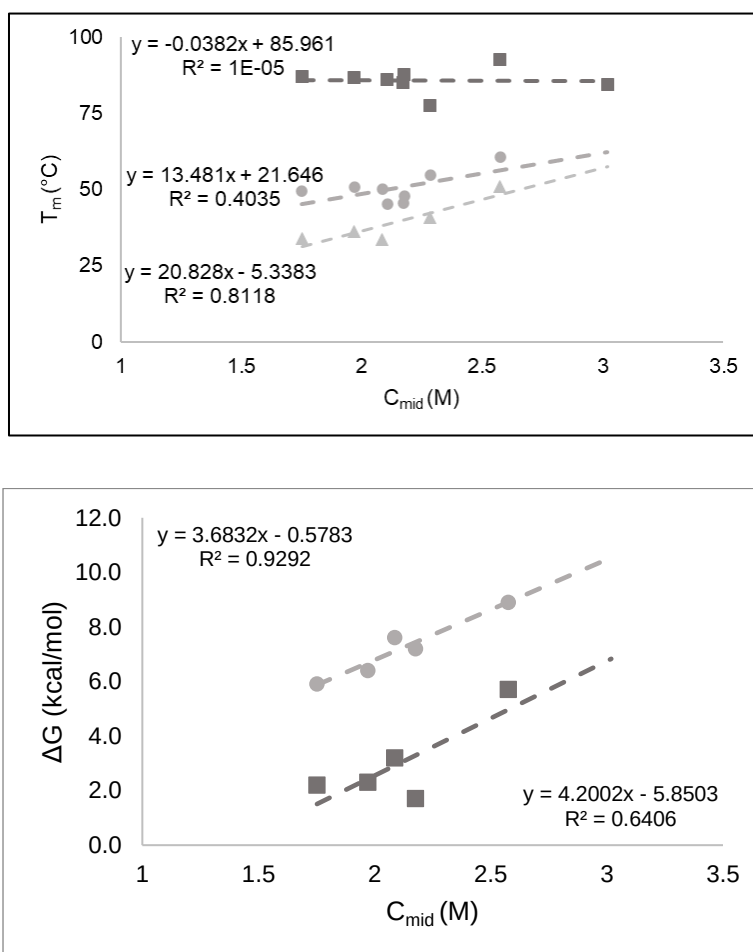


Figure 4. 10 Correlations of urea denaturation to biophysical parameters of SOD1.

Top: Correlations to reduced apo (Δ), oxidized apo ($\bullet$) and holo ($\square$) melting temperatures to fit C_{mid} values. A strong correlation exists for reduced apo, but the correlation is gone by the time the protein has fully matured. As holoSS SOD1 is extremely stabilized by bound metals, one would not expect a correlation with metal-free aggregates. *Bottom* correlation to ΔG_{dimer} ($\bullet$) and $\Delta G_{\text{monomer unfolding}}$ ($\square$) with fit C_{mid} values. A stronger correlation exists with ΔG_{dimer} , supporting that the IB consists of native and non-native dimers. p values can be found in Table A.2.1.

4.4.6 CR, FTIR and urea denaturation of other condition IBs support similar structure but differing stability

A subset of mutants was studied under other growth conditions (37 °C with grown with zinc for 4 hours and 25 °C grown overnight) using qHDX, CR, FTIR and urea denaturation. qHDX was conducted by Michael Tarasca and detailed results may be found in his thesis. Overall, no dramatic differences from average IB at 37 °C were found by qHDX. Similarly, CR binding spectra under these growth conditions show a small amount of CR binding to the protein, indicating a small amount of amyloid present (Figure 4.11). FTIR spectra for mutants grown at 25 °C also look almost identical to the average 37 °C IB FTIR spectrum, corroborating a similar structure (Figure 4.12, Table A.4.5).

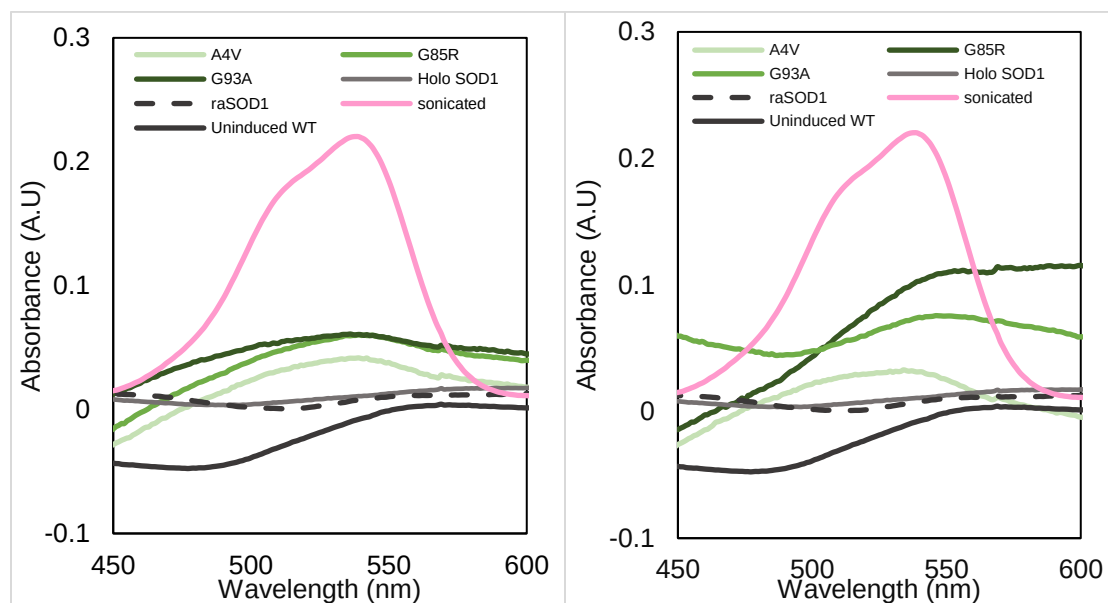


Figure 4. 11 CR spectra of SOD1 variants grown under different conditions.

CR spectra of samples grown at 37 °C with zinc (*left*) and at 25 °C without zinc (*right*). Uninduced WT IB as well as reduced apo SOD1 (raSOD1), holo SOD1 and sonicated reduced apo SOD1 (representing pure amyloid) are shown on both plots for comparison. All IB samples show some CR binding.

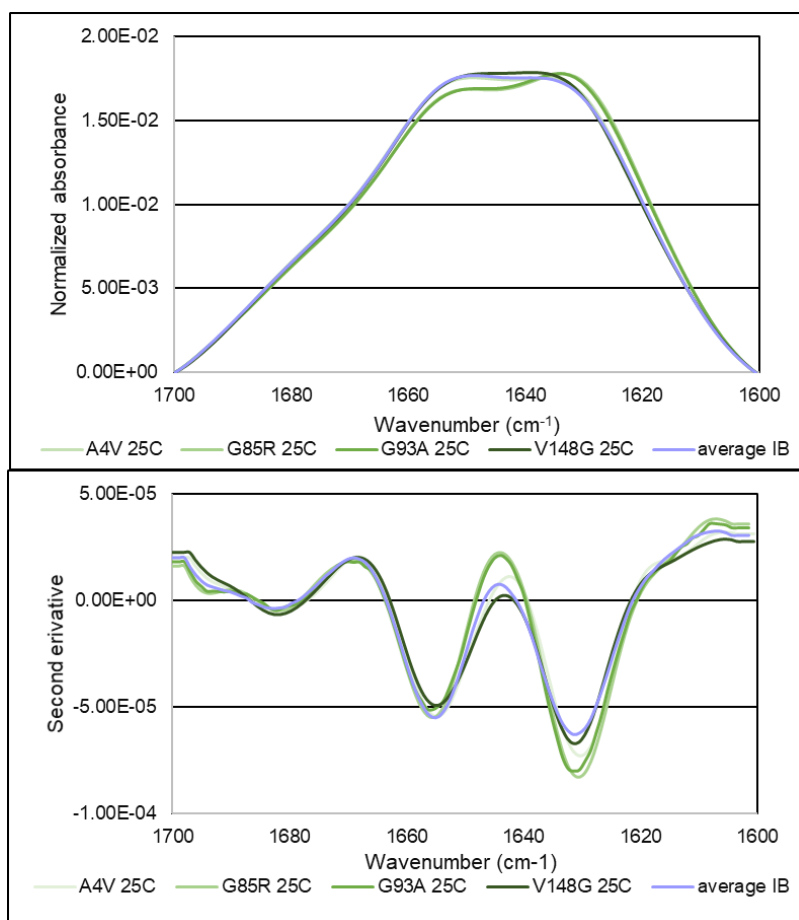


Figure 4.12 FTIR spectra of SOD1 variants grown at 25°C overnight.

Absorbance (*top*) and second derivative (*bottom*) spectra for SOD1 variants grown at a lower temperature.

Interestingly, FTIR spectra for mutants grown at 37 °C with zinc do appear different from the average 37 °C IB FTIR spectrum: while peak *locations* do not appear different, relative heights between the main two peaks (at 1650 cm⁻¹ and 1630 cm⁻¹) appear different with the 37 °C zinc mutants (Figure 4.13, Table A.4.6). More specifically, there appears to be a smaller amount of the 1650 cm⁻¹ peak, which corresponds to disordered, and a greater amount of the 1630 cm⁻¹ peak, which corresponds to β -structure. While this may initially suggest a different IB structure, technical replicates revealed variability in the height of the disorder peak, and examination of what protein remained adsorbed to the FTIR crystal shows primarily disordered protein. This suggests that the IB grown under this condition may actually have a similar structure to the others (as shown by qHDX) but different packing: if the aggregate has a more tightly packed core and less tightly packed disorder “shell” than the others, with variability in FTIR results arising from differential disorder adsorption to the crystal.

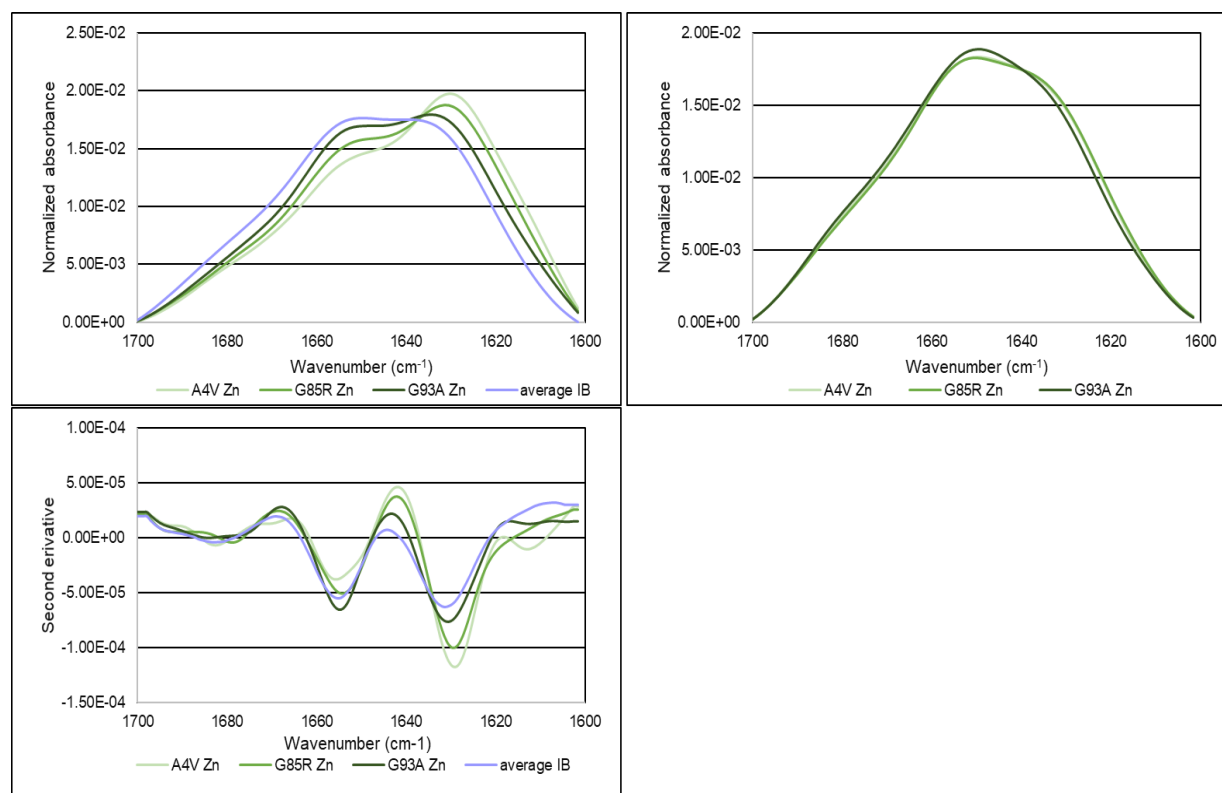


Figure 4. 13 FTIR spectra for SOD1 variants grown at 37 °C with zinc

Absorbance (*top, left*) and second derivative (*bottom*) spectra for SOD1 variants grown at a lower temperature. *Top right* displays absorbance spectra of protein that stuck to the ATR crystal, showing that the apparent difference in IB structure seen at 37 °C with zinc from average IB may be due to different proteins sticking to the crystal during the experiment.

Urea denaturation of mutants grown at 25 °C and for mutants grown at 37 °C with zinc corroborate this hypothesis. Where 25 °C denaturation curves are nearly identical to the 37 °C of the corresponding mutant, in all cases, 37 °C with zinc variants show a right-shifted denaturation curve, indicating that more urea is required to disaggregate/solubilize the aggregate (Figure 4.14). This strengthens the hypothesis that IBs grown under different conditions may have different tightness of packing in the IB.

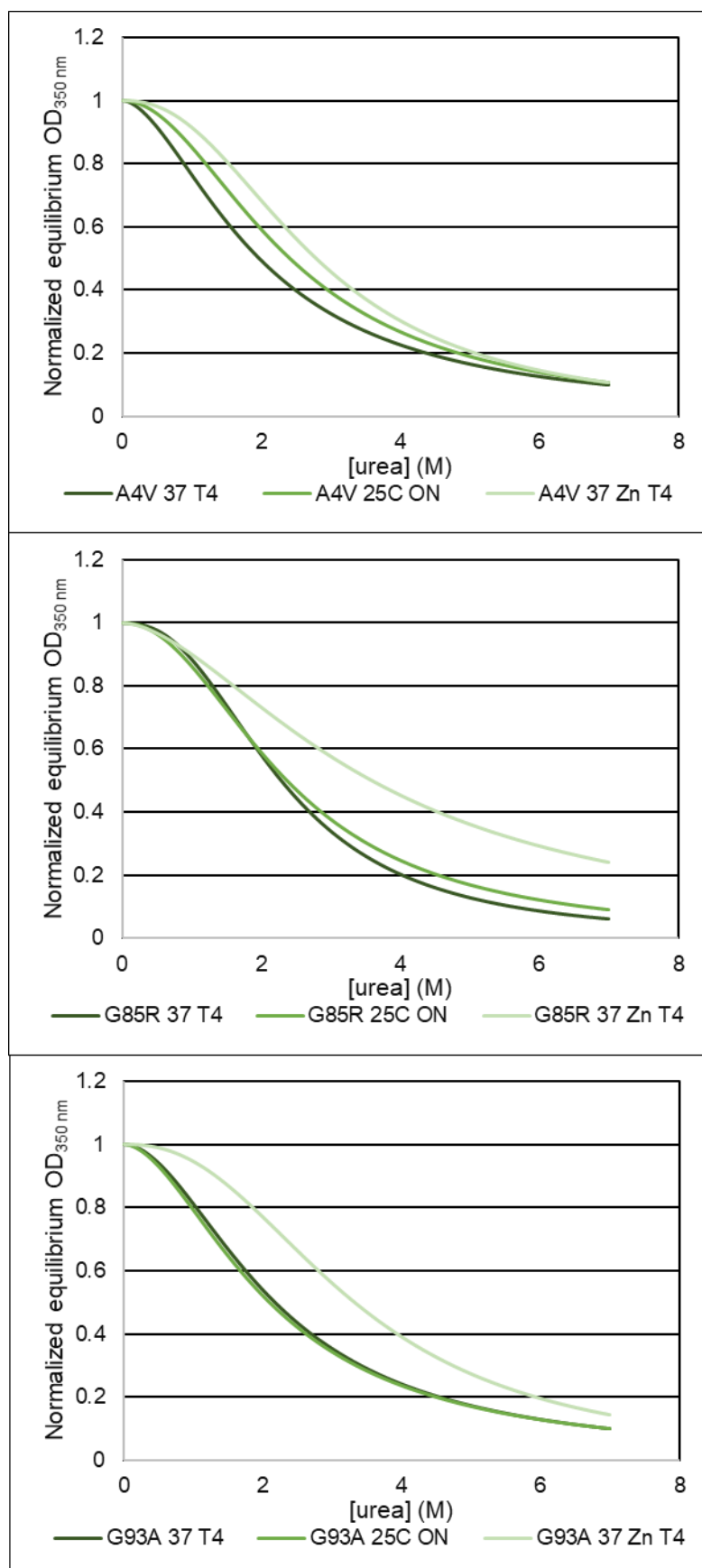


Figure 4. 14 Comparing urea denaturation curves for SOD1 variants grown under different conditions

Comparison of urea denaturation of A4V (*top*), G85R (*middle*) and G93A (*bottom*) mutants grown under different conditions show that IBs grown with zinc have a different relative stability.

4.4.7 Ensemble model describes SOD1 IB structure

Taken together the results presented herein provide compelling evidence for substantial native-like structure in SOD1 IBs, based on: 1) correlation of increased IB formation with decreased fALS mutant ΔG_d and consequently increased population of folded, aggregation prone monomers^[194] 2) native-like pattern of qHDX protection in IBs; 3) similar secondary structure by FTIR in IBs and native SOD1. In addition, associations of native-like and partially folded SOD1 conformations similar to ones previously reported^[225,277,284–286] will confer protection of residues in loops, which are unstructured in native reduced apo SOD1^[277,281]. In-cell NMR of SOD1 in *E. coli* exhibits only native signal for W32, a reporter of global folding, as well as disordered signals indicating multiple SOD1 conformations^[225,300]. In addition, amyloid structures are not prominent in the IBs, based on CR and FTIR spectral features and the distribution of qHDX protection beyond potential steric zippers. Collectively, these findings provide strong support for a model of SOD1 IB formation comprising an ensemble of pathways, including a range of prominent native-like as well as some amyloid-like conformations (Figure 4.15).

The ensemble model explains why SOD1 point mutations generally have very little effect on IB structure: no single mutation substantially alters all pathways and self-association interfaces. The outlier that proves the rule is V148G: due to its central location in multiple interfaces, this mutation significantly shifts the IB aggregation ensemble. Interestingly, decreasing SOD1 stability by mutation, or increasing temperature, tends to increase the proportion of protein in IBs (Fig. 4.1B, Chapter 2); why and how this occurs is unclear and requires further investigation.

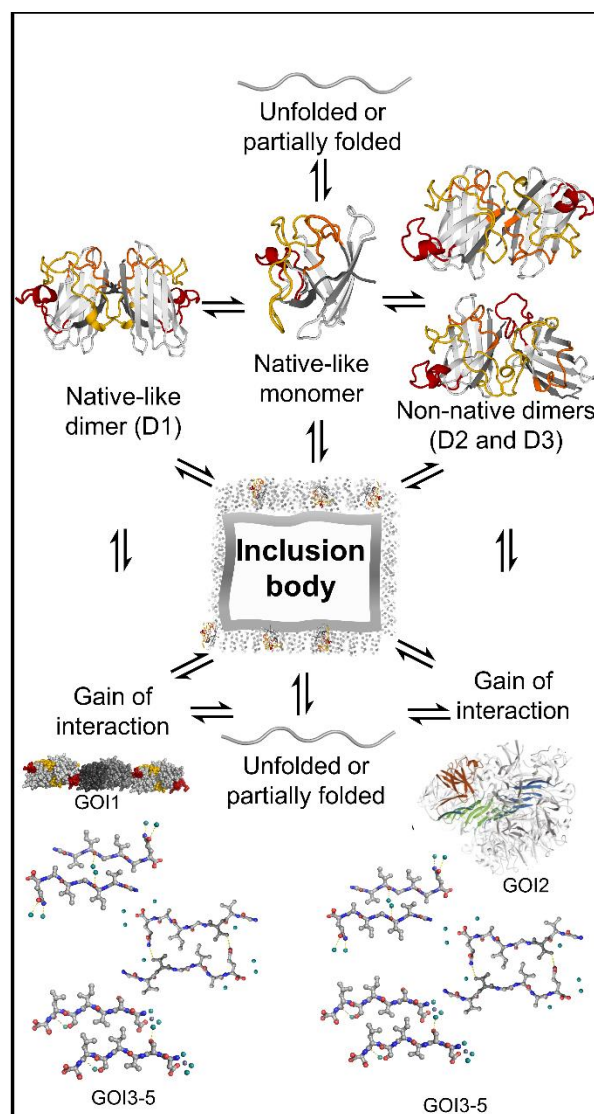


Figure 4. 15 SOD1 IB formation involves multiple pathways that include native and non-native interactions

Loops are coloured as in Fig.2. D1, D2, D3 form in solution^[225,277]. Gain of interactions in crystal structures include: GOI1, an “amyloid-like” filament of near-native dimers where L7 (red) packs against the $\beta 5$ - $\beta 6$ edge strands of the next dimer including intermolecular hydrogen bonding (PDB entry 1OZU)^[284,285]; GOI2 is a corkscrew oligomer model (5IIW)^[283]; GOI3-5 are peptide steric zippers (4NIP, 4NIO, 4NIN) that may form from D1-D3 and contribute to assembly of more or less native-like species resembling GOI1 or GOI2^[282].

4.5 Conclusions

Our findings for SOD1 IBs are of broad significance for understanding and, ultimately, controlling cellular protein aggregation. SOD1 IBs may resemble high quality, native-like IBs^[301], which are desirable for the production of native proteins and as functional protein particles^[267,268,302]. The structural characteristics of SOD1 IBs and effects of mutations may be valuable for advancing rational design of IBs, currently an

inefficient endeavour with often unpredictable outcomes. In addition, comparisons between the fALS mutant SOD1 IBs and disease aggregates highlight key similarities and differences in the assemblies of different proteins. SOD1 IBs and aggregates in ALS models^[272,292] and patients^[297] exhibit similar characteristics. The common antiparallel- β structure of the SOD1 aggregates contrasts with the prominent parallel β -amyloid in mature inclusions of other unrelated proteins in amyloid diseases, such as immunoglobulins, prions and A β ^[294,303], although some early stage aggregates of these proteins also contain significant antiparallel structure^[304,305]. The cellular structures and impacts of aggregates can be affected by many factors, such as the stage of assembly or age of the aggregate, biophysical characteristics of the aggregating protein, interactions of the protein with a variety of cellular species, and the specific procedures used for aggregate isolation and characterization^[113,264,269,303,306]. For example, different cellular aggregate structures determined at moderate resolution by mass spectrometry qHDX for prion proteins exhibit widely varying toxicity that varies with the method of aggregate preparation^[147,307]. It is therefore crucial to be mindful of such factors when, for example, seeking to develop therapeutics to target protein aggregation. The method we present here has the advantage of minimal sample processing and associated changes to the protein structures formed in cells.

In conclusion, our study presents a coherent basis for understanding many previously unconnected observations of SOD1 self-association as an ensemble process, which may be similar to the aggregation of many other proteins. NMR qHDX is demonstrated as an accessible and valuable tool that may be applied to determining the impacts of biological and processing variables on IB structures at high resolution and so advance fundamental knowledge of the molecular basis of protein aggregation in cells as well as the rational modification of such aggregation for myriad practical purposes. Further applications and variations of qHDX NMR will continue to be explored in Chapter 5.

Chapter 5: Adaptation of qHDX method to other systems

5.0 Preface

Moving beyond SOD1, Chapter 5 discusses limitations of the qHDX method established in this thesis. Using SOD1 and two other protein systems, Adnectins and apomyoglobin, this chapter will show how to successfully adapt qHDX to lower-aggregation systems. Successful strategies will be highlighted, and necessary modifications will be considered and validated.

This research was conducted by Dalia Naser in collaboration with Anna Schaefer and Bruna Siebenechler. Dalia Naser had oversight over the experimental design and interpretation. SOD1 experiments and data analysis and interpretation were performed by Dalia Naser. Adnectin experiments and structural analysis was carried out by Anna Schaefer. This portion of the Chapter was previously published.

Citation: M. V Tarasca⁺, D. Naser⁺, A. Schaefer, T. G. Soule, E. M. Meiering, *Anal. Biochem.* **2022**

Comparison of resonance assignments and peak movement trends of Adnectin was conducted by Dalia Naser. Apomyoglobin growth optimization and qHDX experiments were carried out by Bruna Siebenechler. Preliminary resonance assignments and data analysis on apomyoglobin was done by Dalia Naser.

5.1 Introduction

5.1.1 Low aggregation systems and early aggregates

A requirement of the qHDX method applied to IBs as discussed in this thesis is the overexpression of the protein of interest to higher levels than other proteins in the sample. To achieve this, the protein must be extremely aggregation prone. It must generally also be expressed for a long enough time to accumulate the protein. This is limiting for several reasons. Many SOD1 variants discussed in Chapter 2 display interesting aggregation data but do not overexpress significantly enough compared to other proteins, meaning they cannot be studied successfully by qHDX. Studying growth conditions that lower aggregation (25°C or with metals) is limited despite these conditions having potential to show interesting differences (Figure 5.1). Additionally, there is evidence that earlier oligomers are key toxic species and may contain crucial structural information^[211,308–312]. In ensemble systems, earlier aggregates may show if a specific pathway is predominant and then a conformational change occurs. Consequently, adapting the qHDX method to study earlier aggregates is a goal for expanding the method beyond the current studies.

There are two primary approaches to tackling this issue. One approach is by sample clean up through detergent or buffer washing to remove contaminants and increase the relative amount of protein of interest in the sample. This approach is being studied by others in the Meiering lab for its feasibility and whether the additional sample processing changes the structures of the aggregate. A second approach is selective or limited isotope labelling during protein growth and expression. Cells can be grown without isotope labelling until the induction of protein expression. This would not increase the amount of protein of interest relative to others, but the amount of protein visible by a ^1H - ^{15}N HSQC. This approach will be examined in this chapter of the thesis.

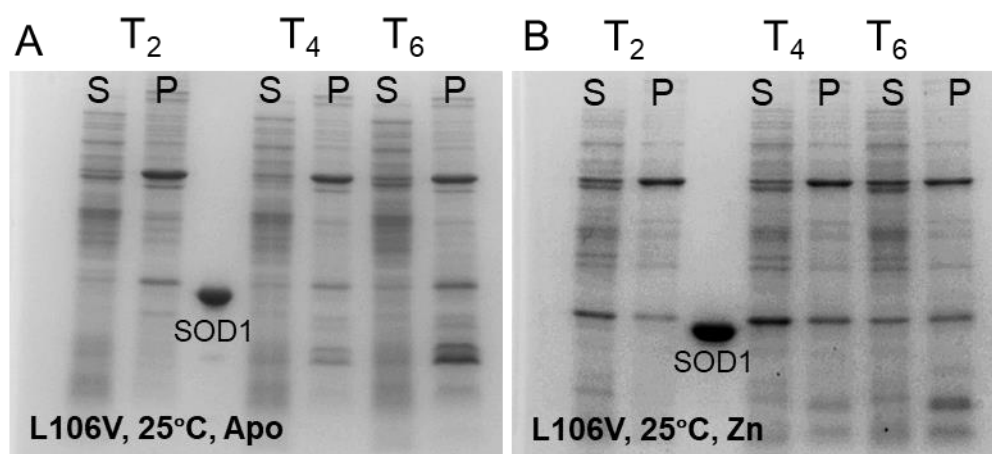


Figure 5.1 Example of lower aggregation conditions of interest of SOD1.

SDS-PAGE of soluble (S) and insoluble (P) L106V mutant of SOD1, which is highlighted in Chapter 2 as of interest to study further, is shown expressed at a lower temperature (*left*) and at a lower temperature growth with zinc (*right*) at 2, 4 and 6 hours of expression (T_2 , T_4 , and T_6 , respectively). Standards of pure SOD1 protein are loaded on each SDS-PAGE and labelled as "SOD1".

5.1.2 Adnectins: adapting qHDX to another protein system

Adnectin (Adn), an engineered protein based on the 10th human fibronectin type III domain, is a protein used to model aggregation in the Meiering group. Adns were designed to specifically bind to therapeutically relevant targets. They have a β -sheet sandwich immunoglobulin-like fold with 3 diverse loops (Figure 5.2). This allows them to bind with high affinity and specificity to a wide variety of targets making them attractive for a wide variety of applications [313–315]. Adn also serve as a powerful system for unravelling specific molecular determinants of IB formation owing to the existing knowledge of their aggregation behaviour and propensity [316]. For these reasons, they were used as the first system to adapt qHDX to. Because of the range of solubilities, they are also a candidate for adaptation of qHDX to lower expression systems or conditions.

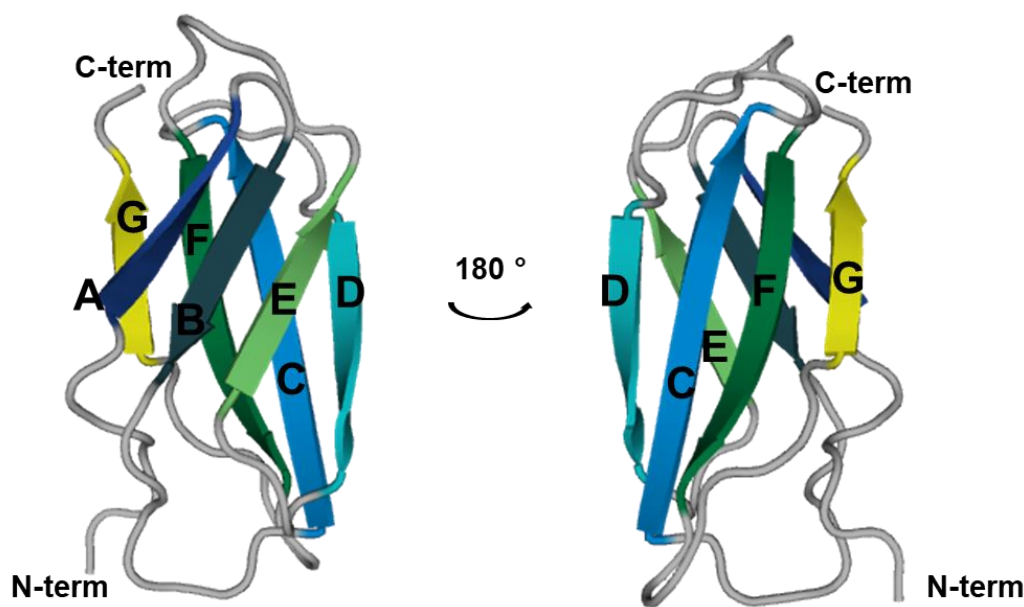


Figure 5.2 Natively folded Adnectin structure

The A and G β -strands (corresponding to the N- and C- terminal regions) of Adn are colored in dark blue and yellow, respectively, with intermittent strands in a blue-green-yellow gradient. Shown on the *right* is a 180° rotation of the structure. Mutations are found in the FG loop. PDB is a homology model made by Julia Steckner of the Meiering group based on 1FNF (the 10 fibronectin 3 domain).

5.1.3 Apomyoglobin: examining chain length dependence by qHDX

Proteins are synthesized from their N-terminal ends, and it has been suggested that folding or misfolding can begin while the protein is still being synthesized^[317]. Consequently, the impact of chain length on protein folding and structure has been heavily studied. Initial studies by Anfinsen and Taniuchi on N-terminal fragments of staphylococcal nuclease (SNase, 149 residues) revealed no native-like conformation could be formed until more than 85% of residues had been synthesized^[318]. Later studies on large N-terminal fragments of SNase (SNase₁₋₁₀₃, SNase₁₋₁₁₂, SNase₁₋₁₂₈, SNase₁₋₁₃₆) purified and refolded from IBs were soluble and did not aggregate and displayed significant secondary structure with increases in tertiary structure as the chain elongates^[319,320]. The longest two fragments also display significant to full catalytic activity. Taken together, this information signifies that most of the protein is required for folding, but that final C-terminal residues contribute minimally.

Apomyoglobin (apoMb, 153 residues) is a helical protein that is frequently used as a model system in protein folding investigations^[321]. Four fragments of apoMb have been examined for impacts of chain elongation: apoMb₁₋₃₆, apoMb₁₋₇₇, apoMb₁₋₁₁₉, apoMb₁₅₃ (full length)^[322]. Like SNase, apoMb gains more

structure as the chain elongates. apoMb₁₋₃₆ is extremely aggregation prone and has a non-native, predominantly β -sheet structure, likely due to the extremely hydrophobic residues in the N-terminus of the protein. As chain length increases, the corresponding peptide fragments contain increasing helical character, particularly as the protein reaches full length. The inclusion of the final residues also drastically reduces the aggregation propensity of the protein: shorter length fragments form high-order assemblies with amyloid characteristics, whereas apoMb₁₋₁₁₉ forms dimeric and trimeric species with more helical features. Full length apoMb achieves a nativelike fold and has minimal aggregation propensity.

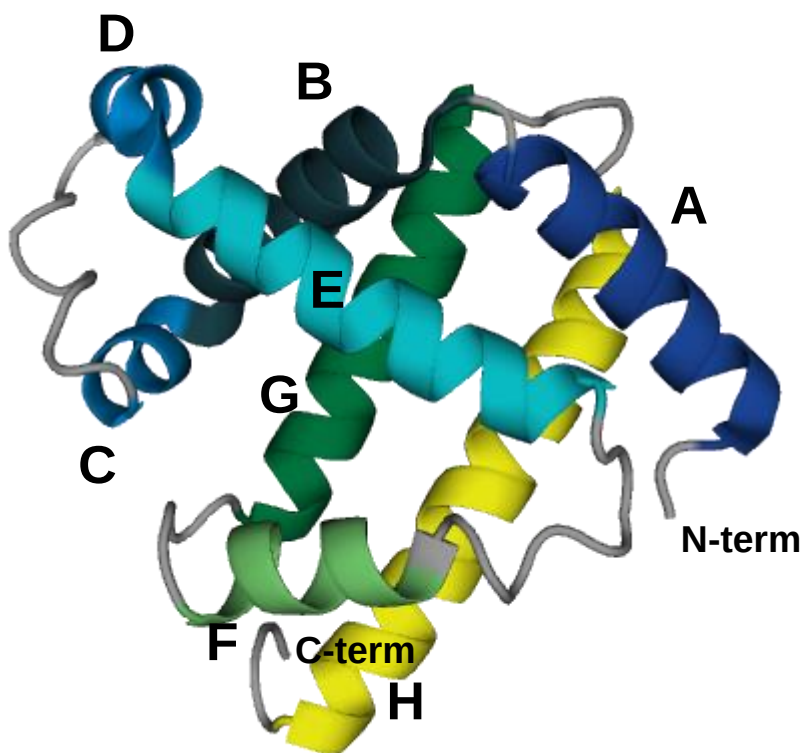


Figure 5. 3 Natively folded full length apomyoglobin structure

The N- and C terminus are labelled. Helices are colored in shades of blue, as in 5.17 and 5.18. Secondary structures shown should not be assumed for the fragments. PDB 1VXA ^[323].

Due to their aggregation propensity, apoMb fragments are frequently purified from IBs for study. Consequently, this provides a tractable model for studying the impacts of chain elongation on *in vivo* aggregates by qHDX. Unlike point mutations, chain elongation significantly and obviously interrupts the folding process of the polypeptide chain. Previous work on apoMb and its fragments indicates that fundamental properties of the amino acid chain are responsible for the folding and misfolding of this protein. Thus, one would expect *in vivo* aggregates to mimic *in vitro* aggregates in the case of the apoMb fragments.

5.1.4 Significance of adapting qHDX to new systems

This chapter gives an overview of several extensions of the qHDX method: expansion of the method to apply to lower aggregation systems, application of the method as-is to another protein (Adnectin) and expansion of a modified qHDX method to a system that studies a different type of protein modification (chain length, apomyoglobin). This is summarized in Figure 5.4. These three variations on qHDX all tie in to answer the crucial question posed in this thesis: can one determine the structure of the IB, and how did the protein fold or misfold to achieve this structure? It is the sincere hope of the author that the extensions discussed in this chapter show the qHDX method as adaptable and informative.

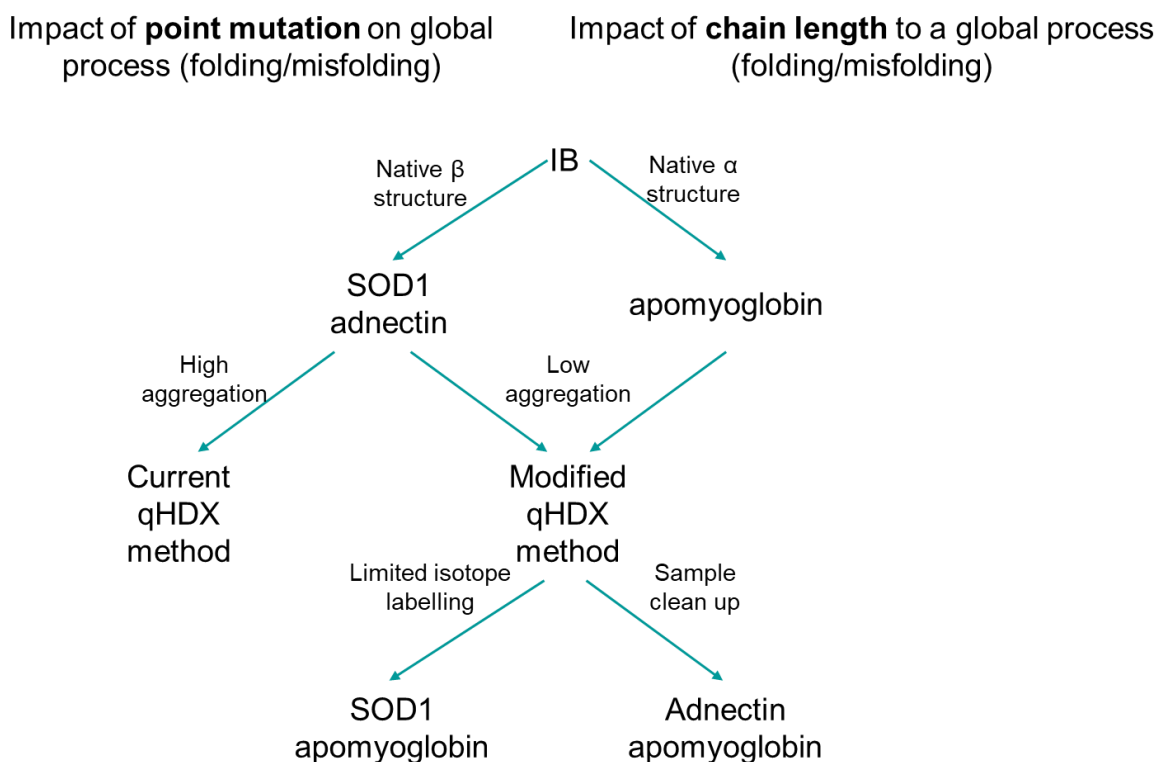


Figure 5. 4 Overview of different expansions of qHDX method

Different avenues that are connected to the overall goal of determining IB structure and misfolding pathways are shown.

5.2 Materials and Methods

5.2.1 Limited isotopic labelling of IB (media switch) of SOD1

Growth protocol is modified from Section 3.3.2. Overnight cultures were prepared as described. 15 mL of overnight culture was used to inoculate 1 L of LB broth supplemented with 100 $\mu\text{g mL}^{-1}$ ampicillin and 30 $\mu\text{g mL}^{-1}$ chloramphenicol then incubated at 37 °C with shaking at 200 RPM to an A_{600} of 0.6. Cultures were harvested by centrifugation at 4000g for 20 minutes at 4°C, and then resuspended in 1 L of pre warmed and aerated minimal media (1 X minimal media salts and all M9 supplements as described in Chapter 2, Table 2.1 using $^{15}\text{NH}_4\text{Cl}$ as the nitrogen source) and the same antibiotics. Culture was left to equilibrate at 37 °C with shaking at 200 RPM for 1 hour before induction of protein expression using 0.1 mM IPTG. A portion of the cells were harvested at 1-, 2-, 3- and 4-hours post-induction (T_1 , T_2 , T_3 , and T_4 , respectively) through centrifugation at 4000g for 20 minutes at 4°C and flash frozen in liquid nitrogen. Preparation of IB samples, NMR spectroscopy and all data analysis was all conducted as described in Chapter 3.

5.2.2 Protein expression of apoMb fragments

Plasmids expressing apoMb fragments were generously gifted by Silvia Cavagnero of the Cavagnero research group (University of Wisconsin). Optimization of apoMb growth and expression was led by Bruna Siebenechler of the Meiering group under my guidance. ApoMb fragments are expressed from a pET blue-1 plasmid with carbenicillin resistance in *E coli* Tuner (DE3)pLacI cells encoding chloramphenicol resistance. Overnight cultures were prepared as described for SOD1 in Section 3.3.2, using 50 $\mu\text{g/mL}$ carbenicillin rather than ampicillin. 30 mL of the overnight culture was centrifuged at 4000g at 4 °C for 10 minutes, and 20 mL of the supernatant was removed. The pellet was resuspended in the remaining 10 mL of the supernatant and used to inoculate 1 L of M9 minimal media containing the same antibiotics, then incubated at 37 °C with shaking at 200 RPM to an A_{600} of 0.6-0.8. Overexpression was induced using 1 mM IPTG. Due to lower expression than SOD1, protein was expressed for 4 hours. A limited isotopic labelling growth was also attempted on apoMb₁₋₁₁₉ as described in Section 5.2.1, expressing protein for 4 and 18 hours (overnight) post-induction. In all cases, the cells were pelleted by centrifugation at 4000g at 4 °C for 20 minutes, flash frozen in liquid nitrogen, and stored at -80 °C. Preparation of IB samples, NMR spectroscopy and all data analysis was all conducted as described in Chapter 3.

5.2.3 Protein expression of Adn variants

Protein expression of Adn variants was conducted primarily by Anna Schafer as described in Chapter 3 with the following modifications. Growth required kanamycin at 50 µg/mL and chloramphenicol at 34 µg/mL. Protein expression was induced with 1 mM IPTG, and variants were grown for 18 hours (V75R) or 6 hours (L78I). Preparation of IB samples, NMR spectroscopy and all data analysis was all conducted as described in Chapter 3.

5.2.4 Detergent washing of apoMb fragments

Due to limited protein expression and an abundance of contaminating protein, detergent washing was conducted on apoMb samples. The impacts of detergent washing on IB structures has been explored in more detail by Anna Schaefer on the Adnectin system and has generally shown to have minimal structural impact. Initial IB sample preparation is done as described in 3.3.3. 1 mL aliquots were thawed and centrifuged at 21 000 G for 20 minutes, and the supernatant was removed. Each aliquot was resuspended in 2.5 mL of 1% w/v N-octyl glucoside and immediately centrifuged for 5 minutes at 5000G. This was repeated a second time before resuspension in H₂O or D₂O for NMR, as described in Section 3.3.5 Due to lower protein expression and some protein loss due to the detergent washing, 2 aliquots were combined for apoMb NMR samples. Aliquots are combined at the commencement of the washing step. The resuspension volume of detergent is scaled up accordingly, but the final resuspension in H₂O or D₂O for NMR is not, as the washed pellets have considerably less protein.

5.3 Results and Discussion

5.3.1 Strategies to study earlier SOD1 aggregates

Media switch approach successful for SOD1 A4V

The media switch protocol was tested on A4V SOD1 to examine earlier time points and determine if earlier aggregate structures differ from those determined at 4 hours. As a particularly high expressing and aggregating mutant, it was deemed the most likely to have sufficient signal to noise under all conditions examined. Additionally, as A4V alters the SOD1 native dimer interface, and it was surprising that it behaved similarly to other SOD1 mutants under conditions investigated previously (Chapter 4), it is of interest to determine whether the earlier aggregates show structural differences to other SOD1 IBs. Consequently, it was postulated that A4V may show differences earlier on in the aggregation process. Expression of A4V at different times is shown in Figure 5.4. Based on the expression levels, under normal experimental conditions, T_1 would have been unusable, T_2 would have produced poor spectra, and T_3 and T_4 would have produced usable spectra.

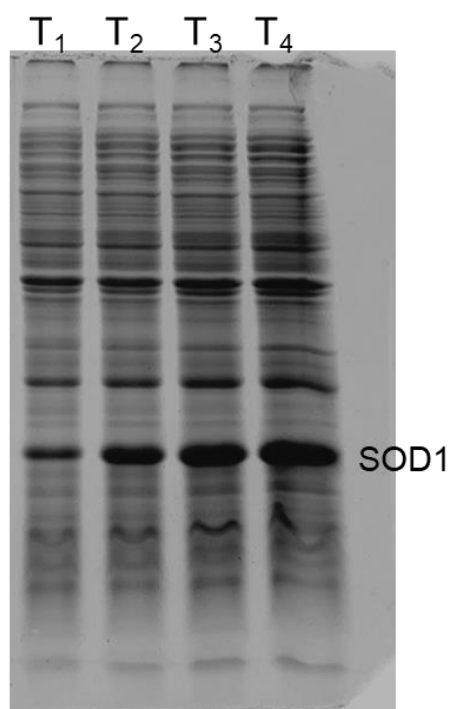


Figure 5. 5 Comparison of SOD1 expression at different time points using a media switch growth

SDS-PAGE of whole cell fraction of SOD1 media switch growth at 1, 2, 3 and 4 hours of expression (T_1 , T_2 , T_3 and T_4 respectively) expression at 37 °C with no metals.

To assess the quality of the spectra at each time point, protonated spectra were analyzed and compared to the average of protonated spectra across other SOD1 samples. Ratios of each spectrum to the

average signal of all protonated SOD1 spectra are shown in Figure 5.5. Ratios were largely found to be near 1, indicating that most of the peaks were of comparable quality to the original experiments. Some residues, such as residues 55 and 56, did appear to be much lower intensity in the T_1 spectrum. Conversely, residues 5 and 68 appear higher than the others (albeit at all time points) Residues with varied intensities in the T_1 spectrum, particularly if they are low, should be considered very critically, as they may have had very poor signal to noise. The identification of these residues shows the importance of checking each peak across time when attempting to examine lower intensity spectra, to ensure that results are not being skewed by inconsistencies caused by signal to noise.

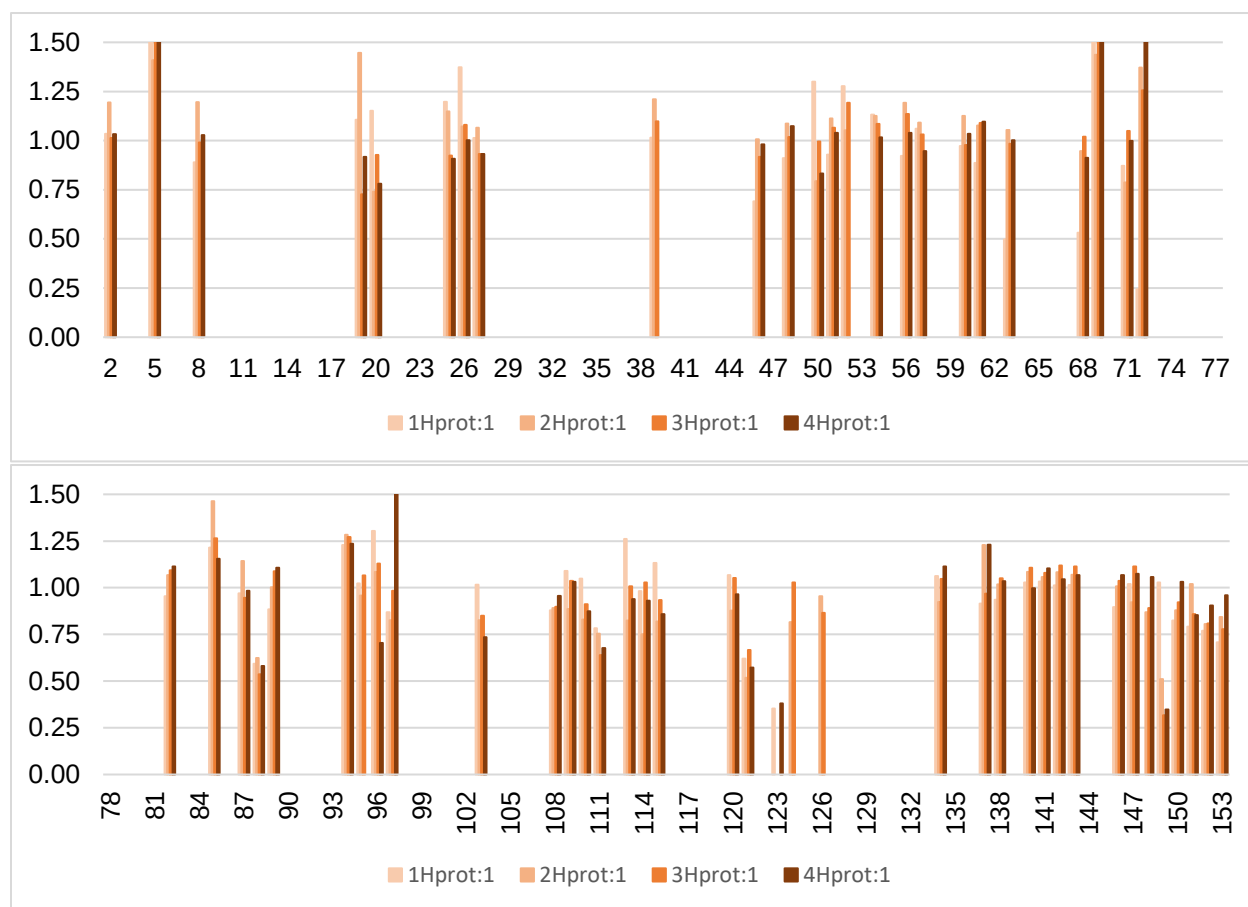
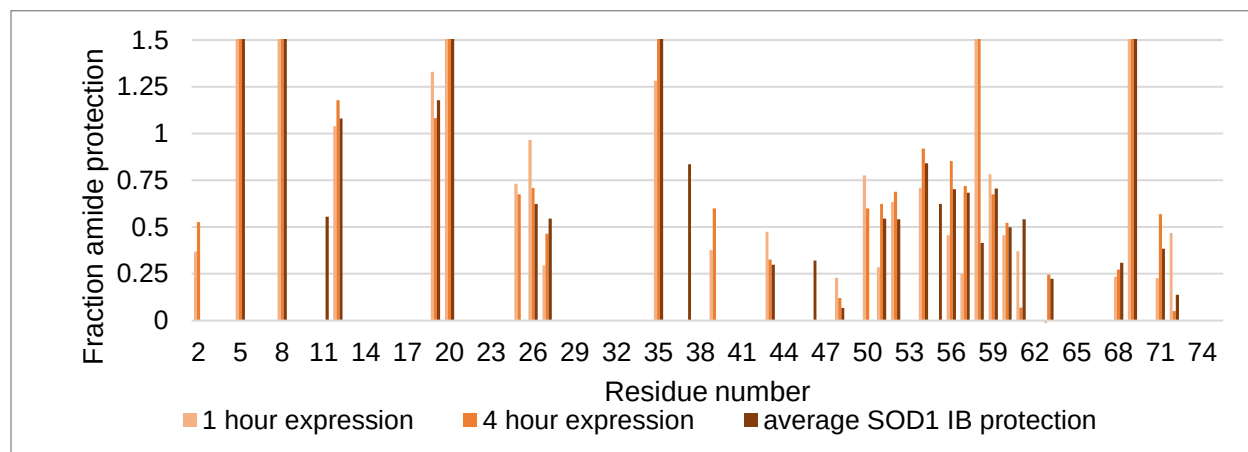


Figure 5. 6 Protonated spectra relative to average SOD1

Signal of each residue normalized to total signal of the sample relative to average protonated SOD1 shown used in analysis for Chapter 4 for samples expressed for 1, 2, 3 and 4 hours in minimal media at 37 °C with no metals. A ratio of 1 indicates that the values were identical, greater than 1 indicates that the residue had a higher signal in the media switch sample, and a lower value indicates that the residue had a lower signal in the media switch sample. Two plots are used (one showing residues 2-77, and the second showing residues 78—153) for ease of comparison of the bars.

Earlier A4V pathway displays a modified ensemble

Examining the protection pattern of A4V over time shows several stretches of residues that appear to be less protected at the earlier timepoints. The entire protection pattern for T₁ and T₄ compared to average SOD1 IB protection are shown in Figure 5.7, and protection of individual key residues is shown in Figure 5.8. Residues 56, 57, 60 and 61 (a small portion of loop 4) show decreased protection; it is however notable that residues 58 and 59 do not, and that residue 57 displays a lower-than-average protonated intensity. Residues 137-141 and 150, 152 and 153 also show lower protection. Both these sets of residues are involved in non-native dimer 3 (D3, Chapter 4, Figure 4.2). More specifically, the 56-61 and 137-141 sets are *only* involved in D3. These data suggest that earlier A4V IB ensembles may contain a lower contribution from D3 relative to the other structures. The progression of the intensity of these residues (Figure 5.8) shows that they do increase to average levels by T₄. A small potential compensation can be seen in residues 82-94, where there is a short stretch of residues with increased protection. These residues are protected by H-bonding in native SOD1. Residues 90-94 form the highly H-bonded β -barrel plug. Earlier aggregates may be smaller oligomers formed more predominantly of smaller, lower order species that more heavily consist of native dimer species rather than larger complexes of non-native dimers.



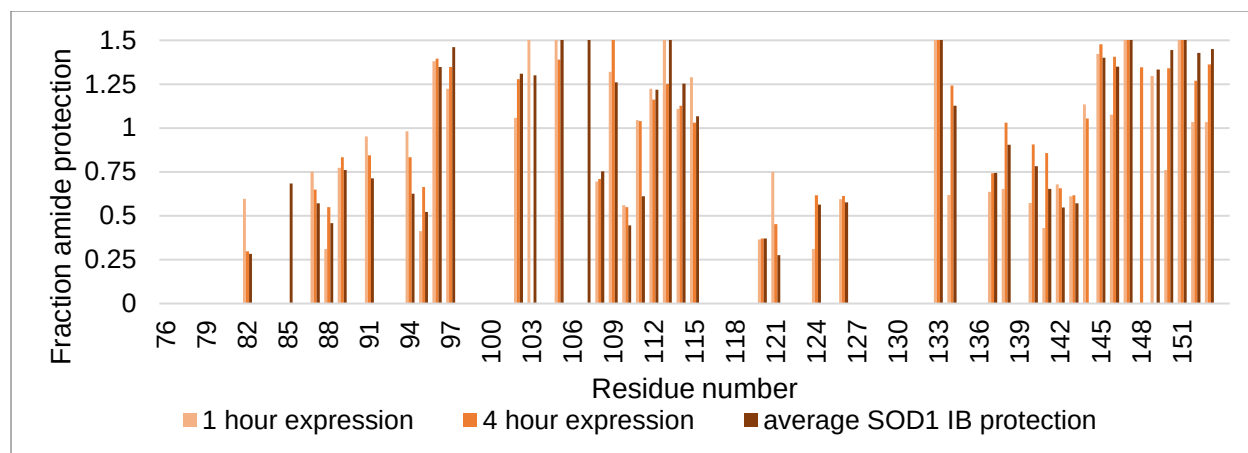
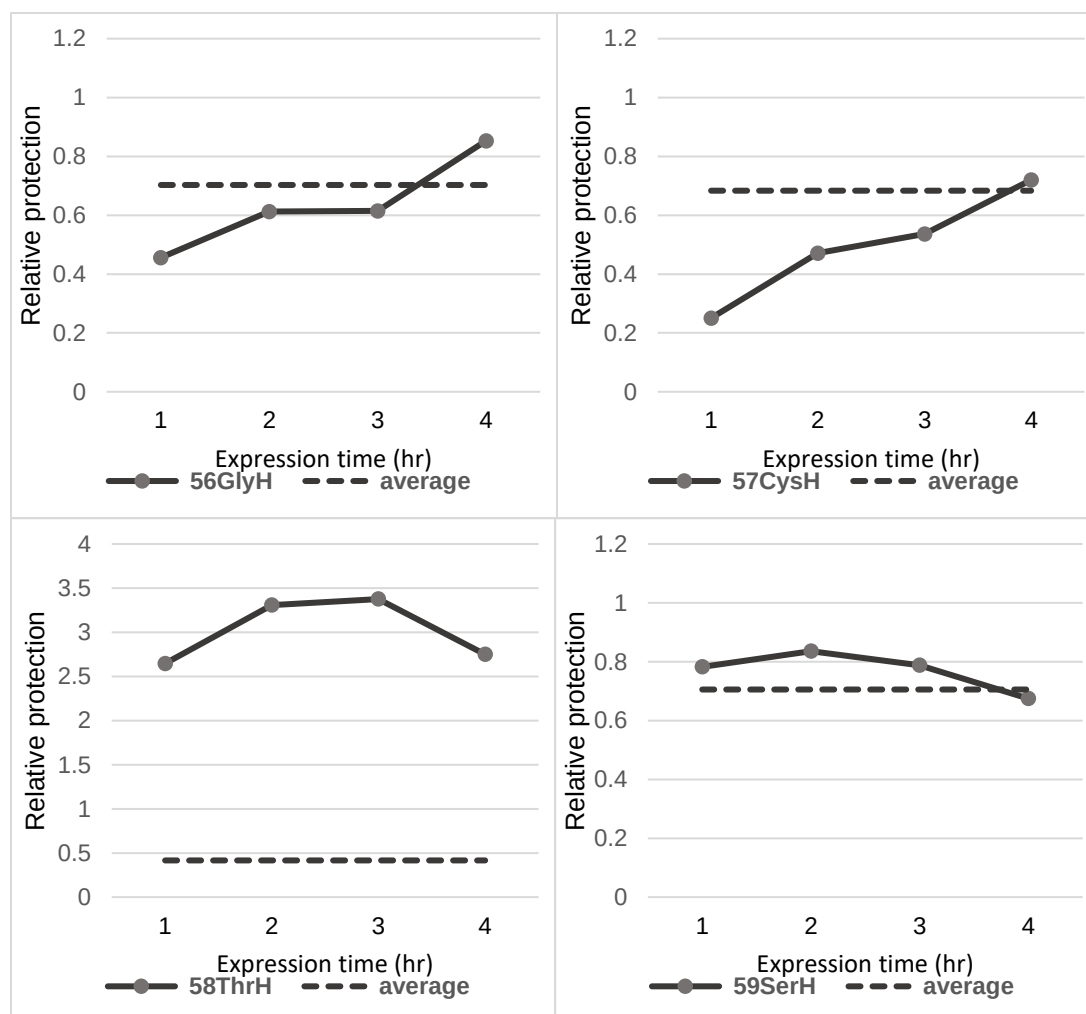


Figure 5.7 Fraction amide protection of T₁ and T₄ media switch of A4V SOD1 compared to average SOD1 IB protection
 Fraction amide protection is calculated as described in Chapter 3, using each sample's corresponding protonated spectrum. Samples are expressed for 1 and 4 hours post induction in minimal media at 37 °C with no metals. Two plots are used (one showing residues 2-75, and the second showing residues 76—153) for ease of comparison of the bars.



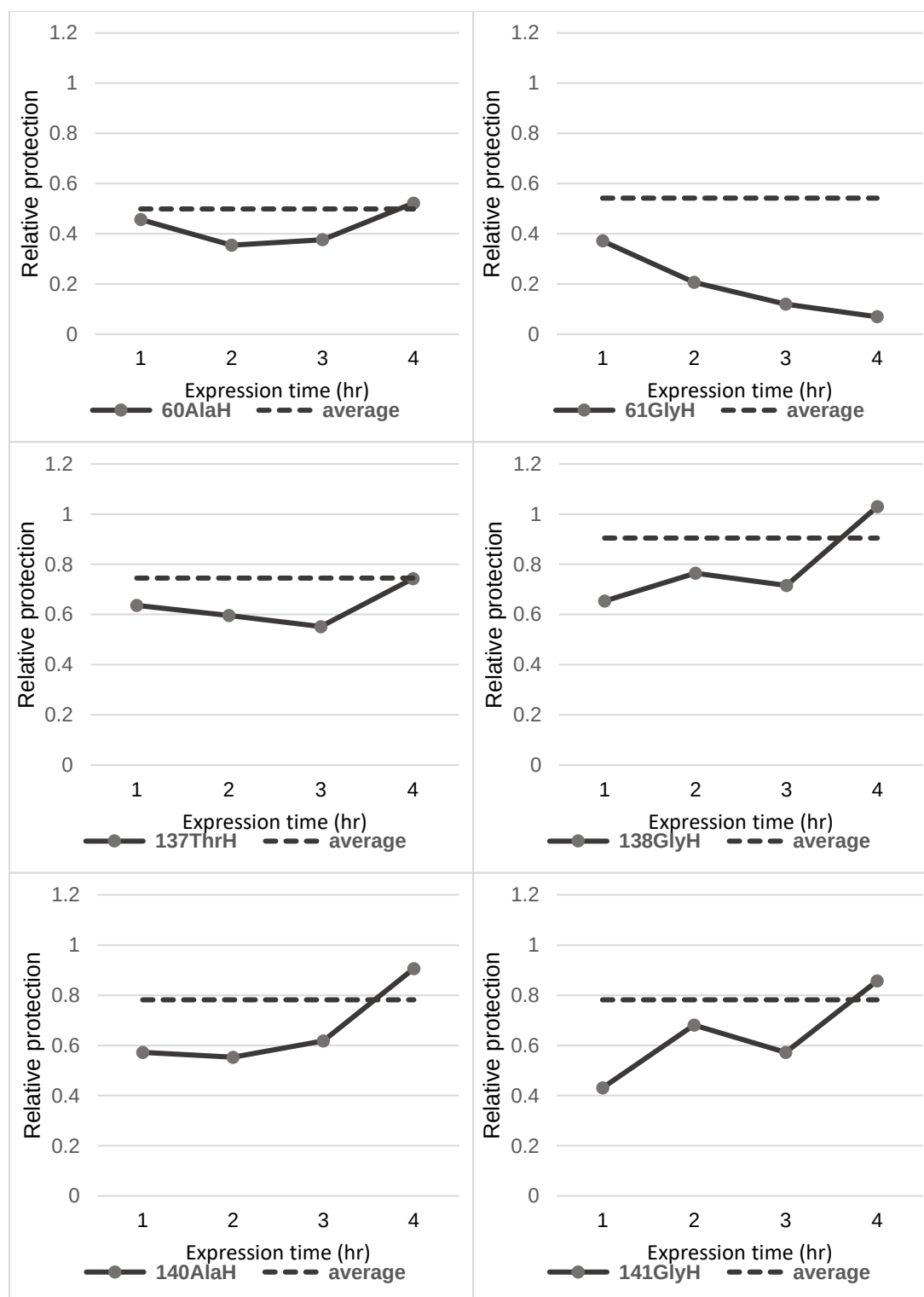


Figure 5. 8 Change in relative amide protection of a residue over expression time compared to average IB residue protection
 Fraction amide protection is calculated as described in Chapter 3, using each sample's corresponding protonated spectrum. Samples are expressed for 1, 2, 3, and 4 hours post induction in minimal media at 37 °C with no metals and are depicted as points connected by a solid line. The dashed line shows the protection of that residue in the average IB sample (Chapter 4) at 4 hours post induction in minimal media at 37 °C with no metals.

5.3.2 Assigning strategies: similarities, differences, and new insights with Adn and apoMB

Existing chemical shift assignments inform new assignments

Resonance assignments for Adn were obtained using a similar strategy to SOD1. The average chemical shift of each nucleus type within each amino acid was determined for Adnectin and compared to SOD1 (Figure 5.9). As expected, a strong correlation exists between ^{15}N chemical shifts, whereas a weak correlation exists between proton chemical shifts. Proton chemical shifts are extremely sensitive to environment: small changes in solution conditions such as pH, very small local structural changes, and interactions with solvent may influence the shift^[209]. Additionally, the range of proton random coil chemical shifts is very small (~ 1 ppm), so a seemingly small change in chemical shift is significant relative to the range of shifts. Further note that a similar correlation was seen when comparing SOD1 chemical shifts when denatured in DMSO to those of SOD1 denatured in urea: amide nitrogen chemical shifts show an excellent correlation ($R^2 = 0.9563$, plot not shown) whereas the relationship between proton chemical shifts was weaker ($R^2 = 0.4353$) and when comparing SOD1 chemical shifts to standard random coil ($R^2 = 0.934$ for ^{15}N and $R^2 = 0.542$ for proton, see Chapter 3). This indicates that the variability seen between Adn and SOD1 proton chemical shifts is not due to impacts of sequence but rather an inherent variation in proton chemical shifts. These data further suggest that if one is making preliminary resonance assignments or attempting resonance assignments with limited data, more weight should be placed on random coil location of nitrogen chemical shifts rather than proton chemical shifts.

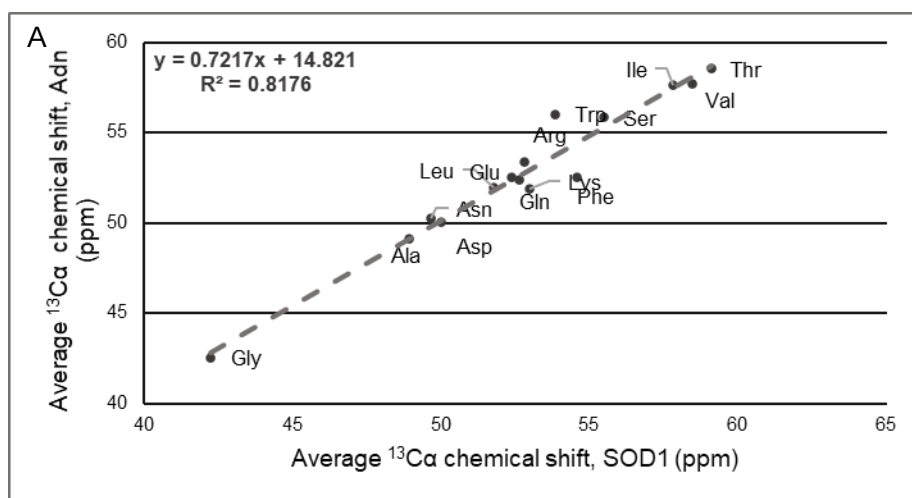
Alpha and beta carbon shifts were also compared between SOD1 and Adn and were both found to be strongly correlated. An interesting feature of chemical shifts is revealed when one examines the average deviation of different atom types across denatured proteins. Deviation in ^{15}N is large across the proteins, which is practical given sequence effects on ^{15}N chemical shifts (see Chapter 3, see also further in Chapter 5). However, $\text{C}\alpha$ and $\text{C}\beta$ show a higher average standard deviation for Adn when compared to SOD1, which was assigned under identical solution conditions. A closer examination of $\text{C}\alpha$ and $\text{C}\beta$ shifts per amino acid for Adn compared to SOD1 revealed that, while some residues were identical, a subset of residues deviated from the SOD1 values. These residues were in stretches together of the Adn sequence: residues 14, 16-19, 21, 23-27, 30, 63, 66, 78, 91, 94 and 96. These are electrostatic stretches in the Adn sequence. Further, the peaks were noted to be broader in shape, and a significantly lower intensity than others. Taken together, these data imply a residual structure in DMSO for these regions that is either forming after the protein denatures, or is residual from the IB structure, which is leading to the broadening. Interestingly, a large stretch of electrostatic residues in SOD1 were unassignable despite quality spectra. There were no candidate peaks for these residues, and at the time it was assumed that

they were overlapped. However, the broadening and lowered intensity demonstrated in Adn suggests that the peaks may simply have not been visible in the spectrum – if the stretch of sequences is similarly forming transient structures in DMSO, they may be greatly broadened/weakened and not visible in the HSQC. This information also informs assignments of new proteins: proteins with a high degree of electrostatic residues in stretches may be more difficult to assign, and if differences are being sought in electrostatic regions, they may be beyond the scope of the experiment.

Table 5. 1 Average deviation of different nucleus types for denatured proteins

	SOD1	FAS 1-4	Cyt C	Adn	acid denatured ApoMb
C	0.434157	0.3937	0.8367	N/A	0.47
C α	0.350894	0.3414	0.9213	0.4743	0.44
C β	0.205357	0.3263	1.2391	0.2488	0.27
H	0.113532	0.0910	0.2352	0.0930	0.57
N	1.374833	1.3183	1.6509	1.1964	1.37

The average and standard deviation of chemical shifts for each amino acid type were averaged for each nucleus. The standard deviation values were averaged for the average deviation. C α and C β for Adn show a larger average deviation than other proteins examined. p values are given in Table A.2.1.



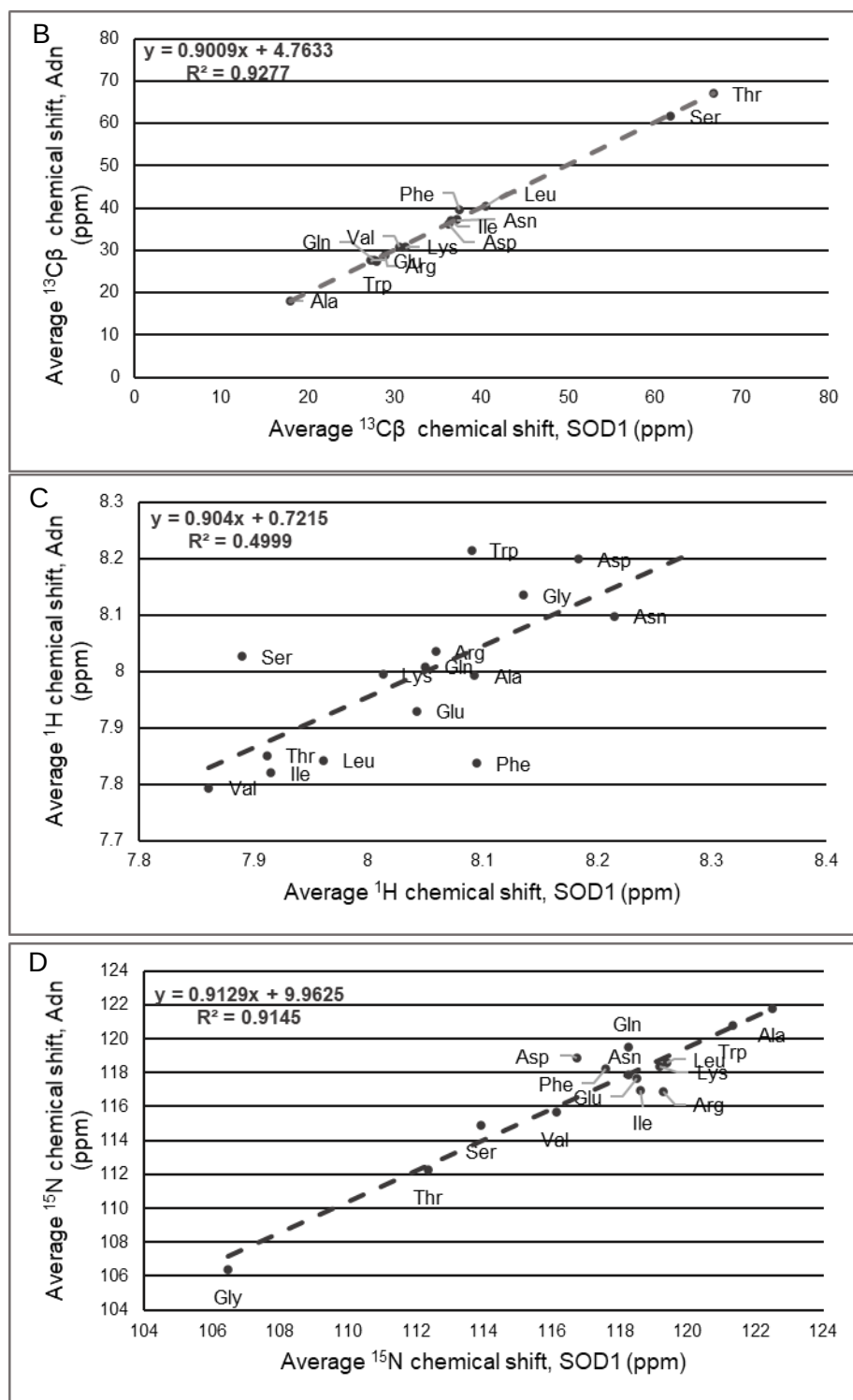


Figure 5.9 Comparison of average chemical shift of each nucleus type per amino acid between SOD1 and Adn in DMSO
Comparison of (A) $\text{C}\alpha$, (B) $\text{C}\beta$, (C) ^1H , and (D) ^{15}N chemical shifts in 95% DMSO/5% H_2O at pH 5.4 Each amino acid is labelled with its type. Pearson correlations are shown on each plot. p values are given in Table A.2.1.

Adn and apoMb (previously assigned in the acid denatured state ^[324]) denatured assignments were analyzed as in Chapter 3 for ¹⁵N sequence effects. ApoMB follows the trends seen in other proteins (Chapter 3) in that 75% of the residues preceded by Ala, Ile, and Val show a deviation from the average of other residues of that type by more than one standard deviation. The proportion of Adn residues that follow this trend, however, is lower at 50%. However, 3/20 residues examined happen to also be in stretches deemed to be impacted by residual structure, and these appear to be the residues not experiencing the expected sequence effects. As a comparison to structured proteins has not been conducted, it cannot be determined how the impact of structure specifically alters chemical shift. However, it can be said that there are more factors impacting the chemical shift which makes parsing trends more difficult.

Table 5. 2 Primary sequence effects on ¹⁵N chemical shifts of amino acids

Preceded by A			Preceded by I			Preceded by V		
	Residue	deviation from average (ppm)	stdev for that res type (ppm)		Residue	deviation from average (ppm)	stdev for that res type (ppm)	
Adnectin	A13	2.12	1.33	T35	-1.59	1.35	S2	-1.3
	T14	-0.73	1.35	S60	-1.04	1.43	R6	-0.63
	R25	0.74	0.92	T71	-2.77	1.35	V11	-2.05
	R30	0.85	0.92	S 85	-1.43	1.43	A12	-1.36
	T58	2.12	1.35	N87	-1.71	1.51	Q46	0
	V75	1.54	1.7	D93	-1.72	1.46	K52	0.97
							Y55	-1.03
							D67	-0.42
							Y73	-1.14
							T76	-0.3
apoMb	D20	4.65	1.35	L29	-1.32	1.93	L2	-3.32
	G23	1.76	0.86	R31	-1.75	1.34	L11	-0.52
	E54	2.24	1.33	L76	-1.92	1.93	K16	2.03
	S58	2.40	1.51	I101	-0.37	1.76	E18	-2.26
	L72	2.28	1.93	K102	-2.57	1.38	A22	-2.29
	I75	2.03	1.76	S108	-2.30	1.51	T67	-3.54
	E85	1.34	1.33	I112	-2.17	1.76	L69	-2.92
	Q91	0.30	1.73	H113	-2.75	1.42	L115	-2.02
	T95	1.36	2.05	A143	-1.49	1.33	L2	-3.32
	I111	2.13	1.76				L11	-0.52
	D126	1.72	1.35				K16	2.03
	Q128	1.40	1.72					

	M131	1.30	1.84						
	L135	3.88	1.93						
	A144	2.41	1.33						
	K145	3.13	1.38						
SOD1	T2	-1.21	1.22	I18	-1.35	1.48	V5	-2.73	1.39
	G56	1.92	1.50	N19	-1.65	1.37	C6	-3.06	2.22
	G61	2.25	1.50	K36	-2.13	1.39	L8	-2.67	1.60
	D90	0.69	0.50	E100	-1.34	0.93	Q15	-1.78	2.52
	D96	-0.20	0.50	S105	-2.46	1.47	H48	-2.68	1.80
	D124	0.81	0.50	I113	-1.87	1.48	G82	-2.73	1.50
	G141	1.95	1.50	G114	-3.02	1.50	T88	-1.43	1.22
	C146	2.13	2.22	G150	-2.83	1.50	A95	-0.92	1.00
	Q153	2.01	2.52	A152	-1.69	1.00	S98	-2.45	1.47
	I104	0.05	1.48				H120	-2.34	1.80

Primary sequence effects are calculated as in Chapter 3. Displayed are the effects on chemical shifts of residues preceded by A, I, and V. Adn and SOD1 chemical shifts are measured in DMSO, and apoMb is measured in aqueous solution at low pH (BMRB: 4676^[325]). SOD1 data are as in Chapter 3 and are included for reference. Depending on the preceding amino acid, the ¹⁵N chemical shift may move upfield or downfield relative to the average value for each amino acid type. The average value was determined per nucleus for each amino acid type.

Preliminary apoMb resonance assignments in DMSO can be informed from previous assignments

Initial experiments to obtain spectra of denatured apoMb and apoMb fragments posed some difficulties due to insufficient protein content relative to contaminant proteins. Four conditions were ultimately examined for feasibility: traditional qHDX sample preparation, a sample that had been detergent washed, and two samples with selective isotope labelling expressed for two different periods of time (Figure 5.10, A through D respectively). Two primary contaminants are present: OmpF, and an unidentified protein of approximately 10 kDa (determined by SDS-PAGE, not shown). OmpF can be identified by a peak at 8.3 ppm and 109 ppm (confirmed by increasing intensity in samples with less protein of interest and in NMR spectrum of uninduced cells, data not shown), and the second contaminant can be identified by a peak at approximately 7.9 ppm and 105.5 ppm. Both peaks are prevalent in Figure 5.10, a sample prepared as in Chapter 3. Figure 5.10B, a detergent washed sample, shows a drastic increase in peak quality and resolution, and a drop in the OmpF contaminant based on the change in intensity in the OmpF peak. Figures 5.10C and D show poor spectral quality and the presence of OmpF, and a decrease in peaks associated with the smaller contaminant. Concurrent work in the Meiering lab on detergent washing of Adns generally improves spectral quality regardless of which proteins are removed due to decreasing the viscosity of the sample, likely due to removal of other

cellular components such as nucleic acid fragments and cell wall fragments that are known to be present in IBs. However, the 10 kDa contaminant does not appear to solubilize differently upon detergent washing and remains in the sample. For preliminary studies, the washed sample processing was selected for optimal resolution.

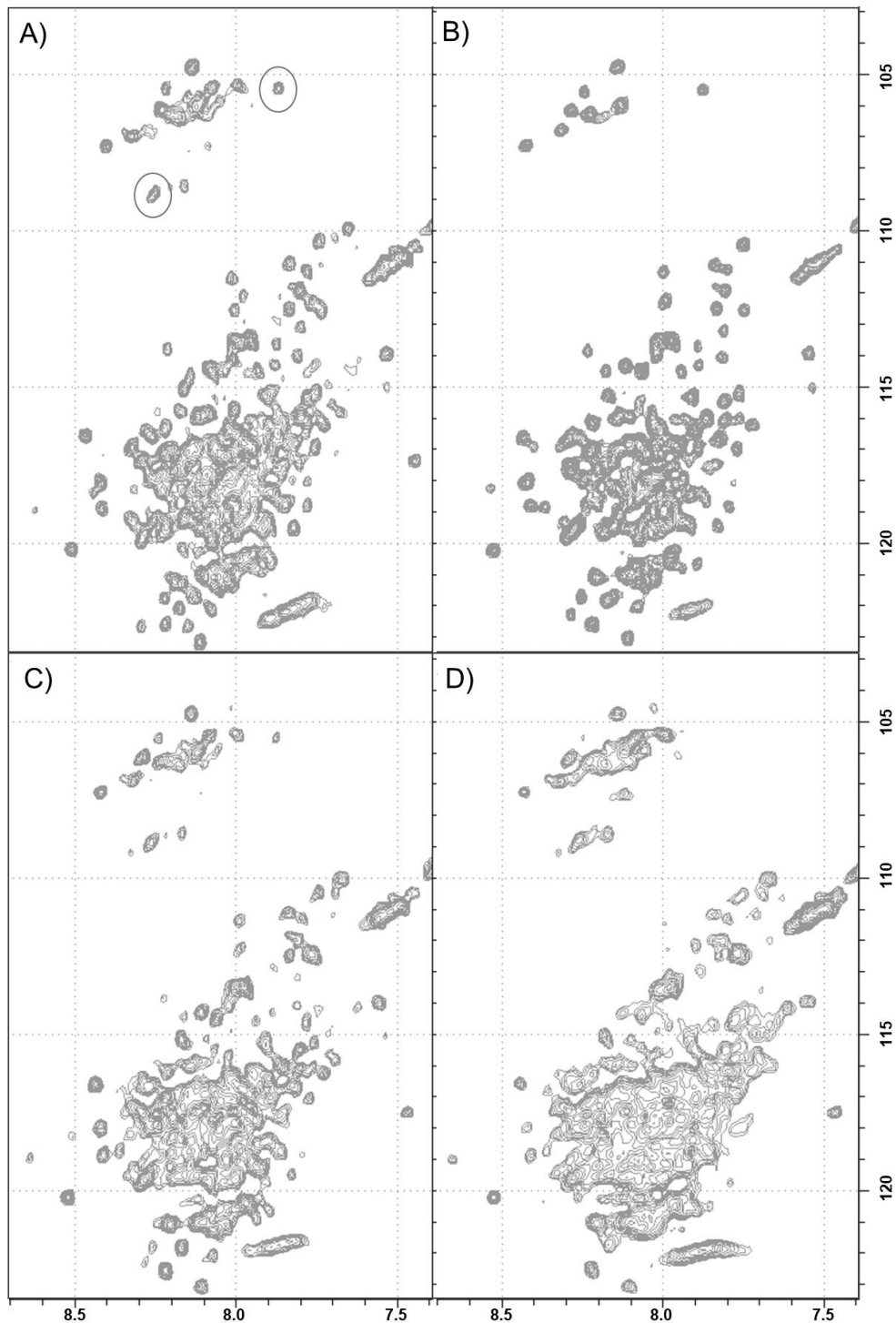


Figure 5. 10 ^1H - ^{15}N HSQC spectra of apoMb₁₋₁₁₉ under different preparation conditions in 95% DMSO/5% H₂O

Traditional (A), detergent washed (B), media switched, 4 hours of expression (C), and media switched, O/N expression (D). OmpF can be identified by a peak at 8.3 ppm and 109 ppm, and the second contaminant can be identified by a peak at approximately 7.9 ppm and 105.5 ppm. These peaks are circled in light grey circles in panel A.

The sample preparation was applied to the 3 apoMb fragment under examination (Figure 5.11). ApoMb₁₋₇₇ and apoMb₁₋₁₁₉ produced adequate spectra for preliminary analysis. Preparation of apoMb₁₋₁₅₃ has not yet been optimized and remains low quality for the time being. Interestingly, a spectral overlay of the three variants shows that the 10 kDa contaminant is most present in the smallest fragment, and least present in the full-length protein. Additionally, apoMb₁₋₇₇ makes the largest amount of IB, and full-length makes the least (which contributes to the poor sample quality but agrees to what is expected from *in vitro* work^[322]). A reasonable assumption for the identity of this protein is GroES, a 10 kDa protein whose primary function is to facilitate proper protein folding: since the shorter fragments are more aggregation prone, it stands to reason that *E. coli* may synthesize more GroES when they are being expressed. Unfortunately, existing DMSO denatured assignments of GroES do not overlap fully with suspected contaminant peaks (data not shown^[326]). Non-overlapped peaks can be analyzed regardless of the presence of the contaminant.

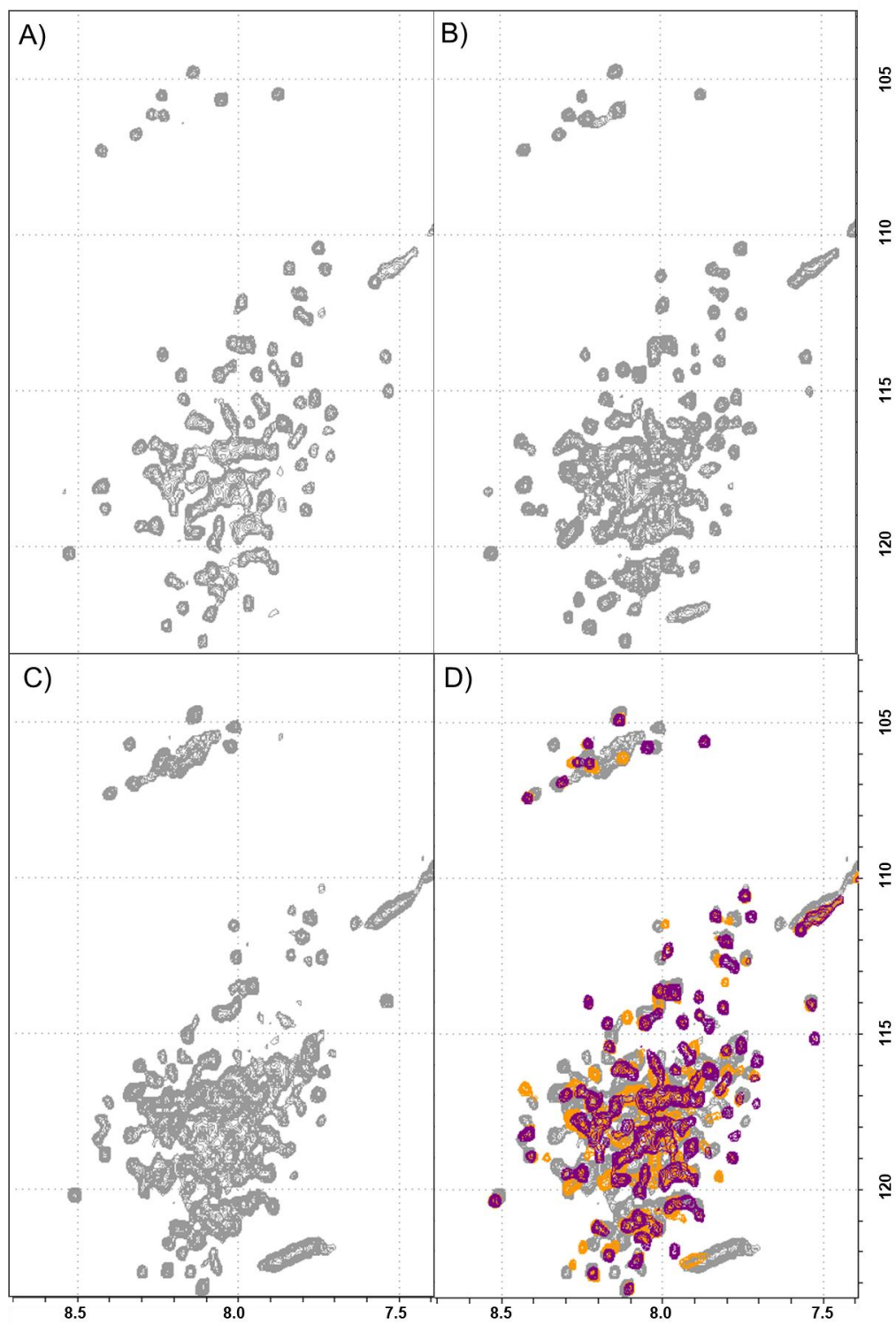


Figure 5. 11 Protonated ^1H - ^{15}N HSQC spectra of apoMb fragments in 95%DMSO/5% H_2O

Detergent washed apoMb₁₋₇₇ (A), apoMb₁₋₁₁₉ (B), full length (C), and all overlaid (D, apoMb₁₋₇₇ in purple, apoMb₁₋₁₁₉ in orange, and full length in grey) grown in minimal media at 37 °C for four hours.

Using previously existing acid-denatured spectra and calculated sequence effect trends, preliminary resonance assignment of ¹H-¹⁵N HSQCs of apoMb₁₋₇₇ and apoMb₁₋₁₁₉ was performed. Assignments are shown on an apoMb₁₋₇₇ spectrum in Figure 5.12; few additional resonance assignments were found for the longer fragment (8 residues) so a second spectrum is not shown. Assignment efforts focused on Gly, Thr, Ser, Ala and Leu residues, particularly those that are more isolated or influenced by sequence effects (Table 5.2). Assignments were compared to the same residues in acid-denatured apoMb and excellent correlations were found (Figure 5.13, $R^2 = 0.6919$ and $R^2 = 0.9431$ for proton and nitrogen, respectively) which strengthens confidence in assignments. Note that when comparing assignment transference in different buffers, R^2 represents the “goodness” of the assignment transfer, whereas the slope of the line represents the relative spread of the peaks (which may be different based on solution conditions). In total, 44% of residues in apoMb₁₋₇₇ were assigned with no additional experiments, speaking to the feasibility of preliminary work in qHDX experiments. Initial D₂O exchanges were also conducted on apoMb₁₋₇₇ and apoMb₁₋₁₁₉ fragments and determined that a reasonable number of peaks had enough signal to noise for analysis (Figure 5.14). Given the quality of the spectra, an initial analysis was undertaken and is discussed later in this chapter.

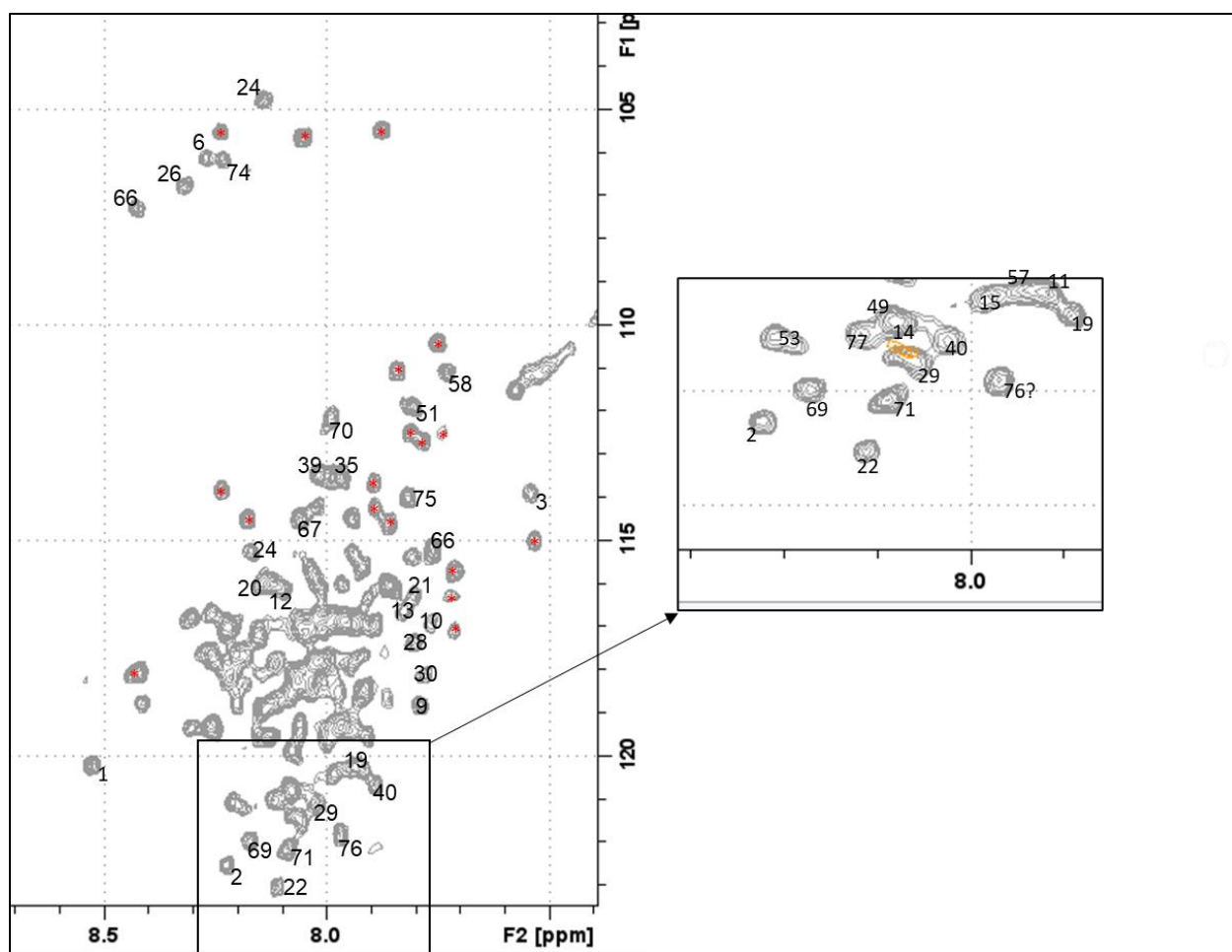


Figure 5. 12 Preliminary resonance assignments of apoMb₁₋₇₇ in 95% DMSO/5% H₂O at pH 5.4

Sequence specific resonance assignments made at 19 °C based on existing denatured apoMb assignments conducted at low pH. Peaks with a * are residues thought to be from contaminating proteins.

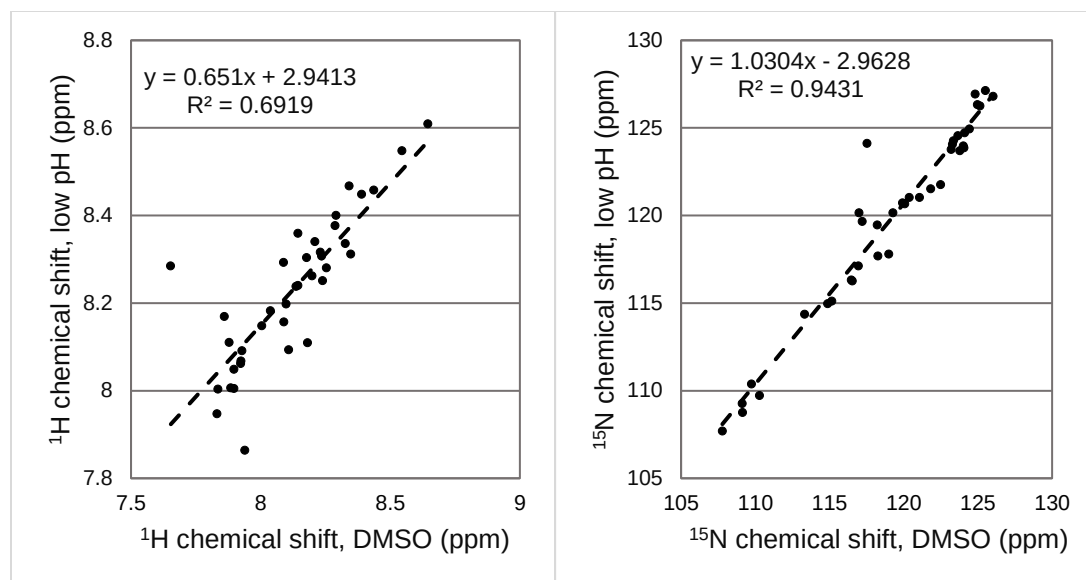


Figure 5.13 Comparison of denatured apoMb₁₋₇₇ backbone assignments under different solution conditions

Proton (*left*) and nitrogen (*right*) chemical shifts of apoMb₁₋₇₇ denatured in DMSO or at low pH. Agreement between the shifts values gives confidence to the preliminary assignments in DMSO.

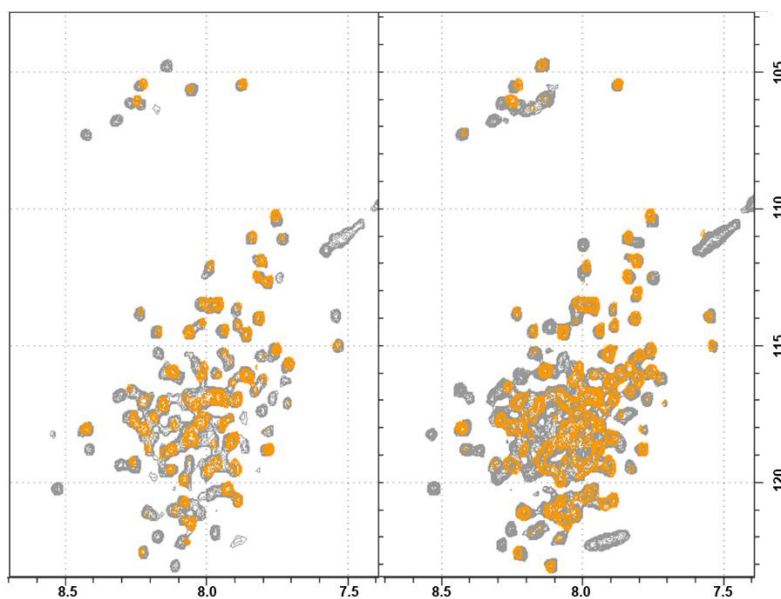


Figure 5.14 Spectral overlays of qHDX samples with their respective protonated spectra

qHDX spectra of apoMb₁₋₇₇ (left) and apoMb₁₋₁₁₉ (right) exchanged for 1 hour compared with their protonated spectra show differential protection of residues.

5.3.3 Adnectin IBs display a native-like ensemble

The qHDX method was applied to Adn by Anna Schaefer in the Meiering Group with my guidance and a more detailed analysis of the data can be found in her thesis. A brief comparison will be shown here for the benefit of discussion. Comparison of chemical shifts is undertaken earlier in this thesis chapter (Section 5.3.2). A comparison of intrinsic exchange rates in DMSO to SOD1 and PrP (as seen in Chapter 3)

with the addition of Adn is shown in Figure 5.15 for 6 Adn samples. Adn k_{int} rates are very much in the range of SOD1, indicating two things: the method conditions are directly influencing the rates in a reproducible way, and the rates are not significantly impacted by sequence. Adn exchange rates were found to be sufficiently slow for detailed qHDX analysis.

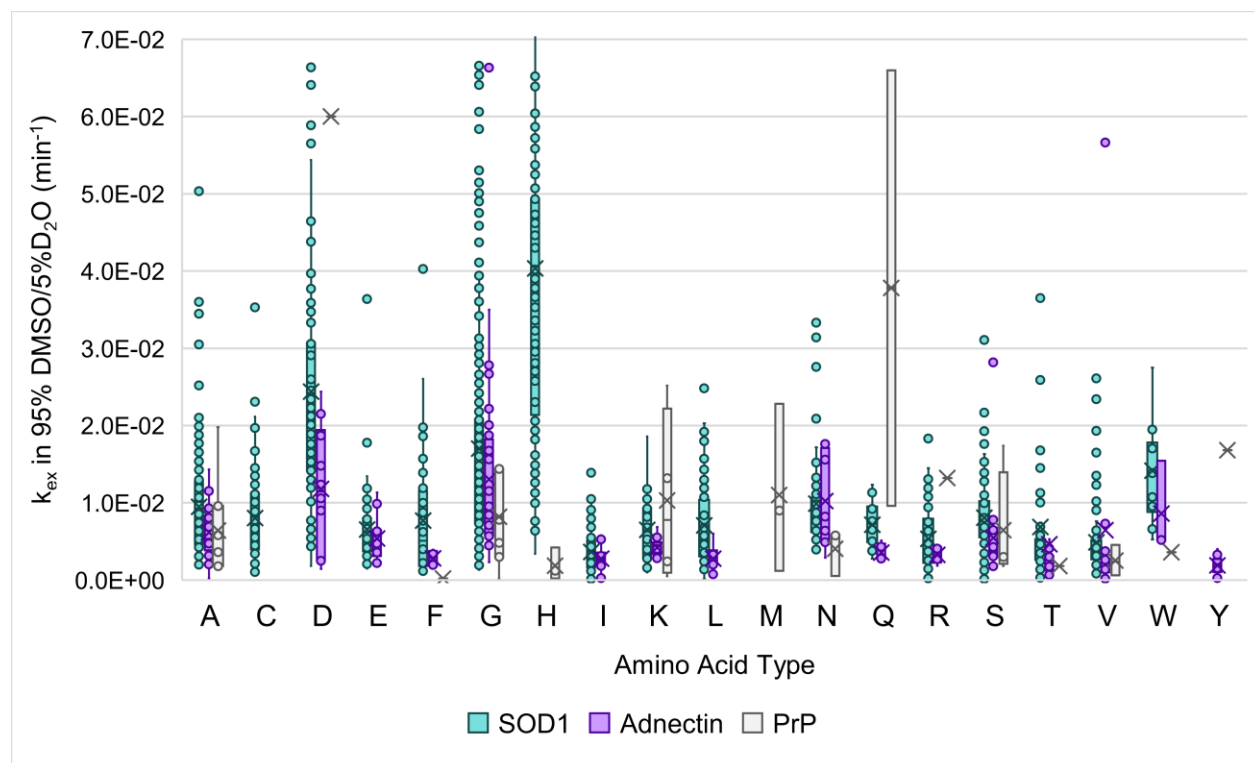


Figure 5. 15 Measured intrinsic exchange rate constant of amide exchange for different amino acids

Values are plotted for SOD1 (teal) and Adn (purple) compared to literature values for pure PrP aggregates (light grey). PrP and SOD1 are as in Figure 3.7 and are included for comparison. Values include results for 18 SOD1 samples (two biological replicates for each of nine different mutants), six Adnectin samples, and a single mutant (P101L) for PrP.

Studies conducted on two Adn variants (V75R and L78I) reveal extensive protection throughout the sequence (Figure 5.15), similar to results for SOD1 (Chapter 4). The general pattern is also maintained between the Adn mutants, despite their widely differing solubilities. The parent Adn sequence contains three amyloidegenic regions according to ZipperDB: residues 18-21, 55-60, 69-75. These regions overlap with those predicted by Tango to aggregate, as well as regions calculated to be hydrophobic. While these regions are protected in both Adn variants, residues not predicted to be in any aggregate are also protected. Natively folded Adn is a β -sheet sandwich that is very strongly H-bonded throughout. B-strands C and D, two strands not implicated by any aggregation predictor, are very strongly H-bonded in native Adn. These strands are also protected in the Adn IB, suggesting that Adn IBs contain native structure. Adn has also been shown to self-associate by domain swapping via its terminal β -strand; this

mechanism may be occurring in tandem to non-native interactions to result in an ensemble of aggregates in the IB structure, as the data available does not preclude contributions from cross- β amyloid or alternative protected conformations.

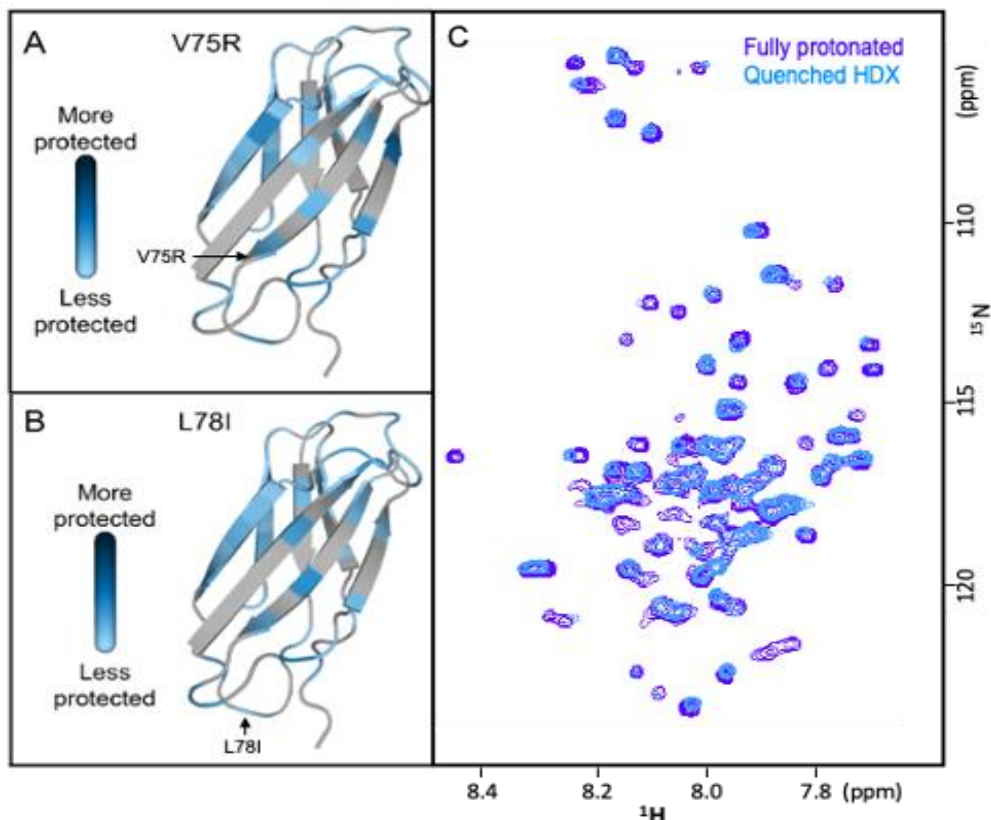


Figure 5.16 Structure of Adnectin IBs from qHDX analysis

A) Average qHDX protection for two biological replicates of Adnectin V75R (model based on PDB 1FNF made with software Swiss-Model^[327]). Protection fractions from 0.5 to 1.0 are coloured as light to dark blue, respectively. Grey indicates no data. B) Average qHDX protection of two biological replicates of Adnectin L78I. The two mutants have very similar protection, implying similar IB structure, as observed also observed for SOD1 mutants. (C) A spectral overlay of protonated (dark blue) and exchanged (light blue) V75R IB. Many peaks in the exchanged spectrum show strong intensity demonstrating protection throughout the protein structure (A).

5.3.4 Apomyoglobin appears to aggregate similarly *in vivo* and *in vitro*

In vitro aggregation studies have previously been performed on apomyoglobin fragments by the Cavagnero group^[322]. While the overall structure of the aggregates was not determined, a battery of lower resolution methods was applied to establish the structural features of the aggregates. It was found that apoMb₁₋₁₁₉ aggregates contained 30% α -helix, 10% β -strands, 40% β -strands + turns and 30% random coil (compared to the 16%, 29%, 51% and 33% found for apoMb₁₋₇₇, respectively). Furthermore, size exclusion chromatography of apoMb₁₋₁₁₉ showed primarily dimer or trimer species, whereas

apoMb₁₋₇₇ aggregates contained primarily higher-order structures such as hexamers. Additionally, thioflavin-T binding showed more amyloid character for shorter apoMb fragments. Taken together and in the absence of qHDX data on the apoMb aggregates, one can assume that apoMb would behave as an amyloid-like aggregate (like those seen by Wang *et al* ^[44]) whereas apoMb₁₋₁₁₉ will exhibit more native-like features.

Preliminary analyses of qHDX data on apoMb₁₋₇₇ and apoMb₁₋₁₁₉ suggests that *in vivo* IBs contain similar structures to those present in *in vitro* aggregates. Across a comparable residue range (1-77), apoMb₁₋₇₇ appears to have lowered protection throughout the protein, with small stretches of strong protection (eg residues 29-39). Conversely, apoMb₁₋₁₁₉ appears more strongly protected across the sequence. As the natively folded, full length apoMb is very strongly H-bonded and generally slow-exchanging by traditional HDX, more extensive protection supports a more native fold. Considering that apoMb is helical and has very little β -propensity, it is a reasonable assumption that the more extensive protection in apoMb₁₋₁₁₉ is due to more native folding rather than increased amyloid features.

For N-terminal fragments, sequence-based predictors may be more advantageous than in previous examples in this thesis, as the three-dimensional structure of the fragments will vary greatly from the full-length protein. TANGO and ZipperDB both suggested that short stretches of residues at around residue 10 and residue 27 were aggregation prone and differed in the latter part of the sequence: TANGO suggests that residues 67-77 are extremely aggregation prone, whereas ZipperDB highlights the stretch of residues 55—67 as the culprit in aggregation. Unfortunately, there is not a strong correlation between the most strongly protected residues in apoMb₁₋₇₇ (residues 29-39) and the predicted regions; this may of course be due to the preliminary nature of the backbone chemical shift assignments. Conversely, the structures of properly folded N-terminal fragments are unknown and may be drastically influencing the resulting aggregates.

Comparison to native physical features may also provide insight into the protection pattern observed. Kyte-Doolittle (K-D) hydrophathy index may in principle highlight aggregation-prone regions; unfortunately, in this case, these regions overlapped with regions predicted by TANGO to be aggregation-prone. Solvent-accessible surface area (SASA) is another parameter that may be examined. More specifically, the ratio of nonpolar solvent accessible surface area to total solvent accessible surface area is shown in Figure 5.17. It is known that β -sheets are more effective than α -helices at burying nonpolar surfaces^[328]. The highest proportion of nonpolar residues/SASA is in the N-terminal region of the protein. It has been suggested that hydrophobic collapse, a model for protein folding, can also be

used to model aggregation. This supports the more amyloid fold for shorter N-terminal fragments and a more native-like aggregate for longer chain fragments, as is observed herein.

A great deal of previous work has been conducted on the myoglobin folding pathway. It has been determined that refolding of apoMb contains a “burst phase” intermediate that forms in the dead time of folding experiments^[329]. Probes in these folding experiments can be characterized as “fast” folding, “slow” folding, or both (termed fast+slow, medium, or biphasic depending on the work referred to). These data provide two key pieces of information: first, they demonstrate a heterogeneous intermediate, and second, they highlight which residues should be protected if a similar intermediate is being formed. The fast and slow residues are shown in Figure 5.17. Note that these works are for full length apoMb. While the studies largely agree, they do differ at residues 67-77, where Jennings^[321] and Garcia^[330] allege slow folding, but the more recent Nishimura study signifies heterogeneous or fast folding. This region shows minimal protection in the apoMb₁₋₇₇ fragment, which would be supported by being slower folding residues. Overall, the majority of full-length structural data appears to not explain the apoMb₁₋₇₇ aggregate, whereas the more targeted studies on N-terminal fragments do. This supports the non-native aggregate but also highlights a difficulty in these studies in comparison to point-mutant studies in that more extensive work likely needs to be done on each fragment to understand the folding of the aggregate.

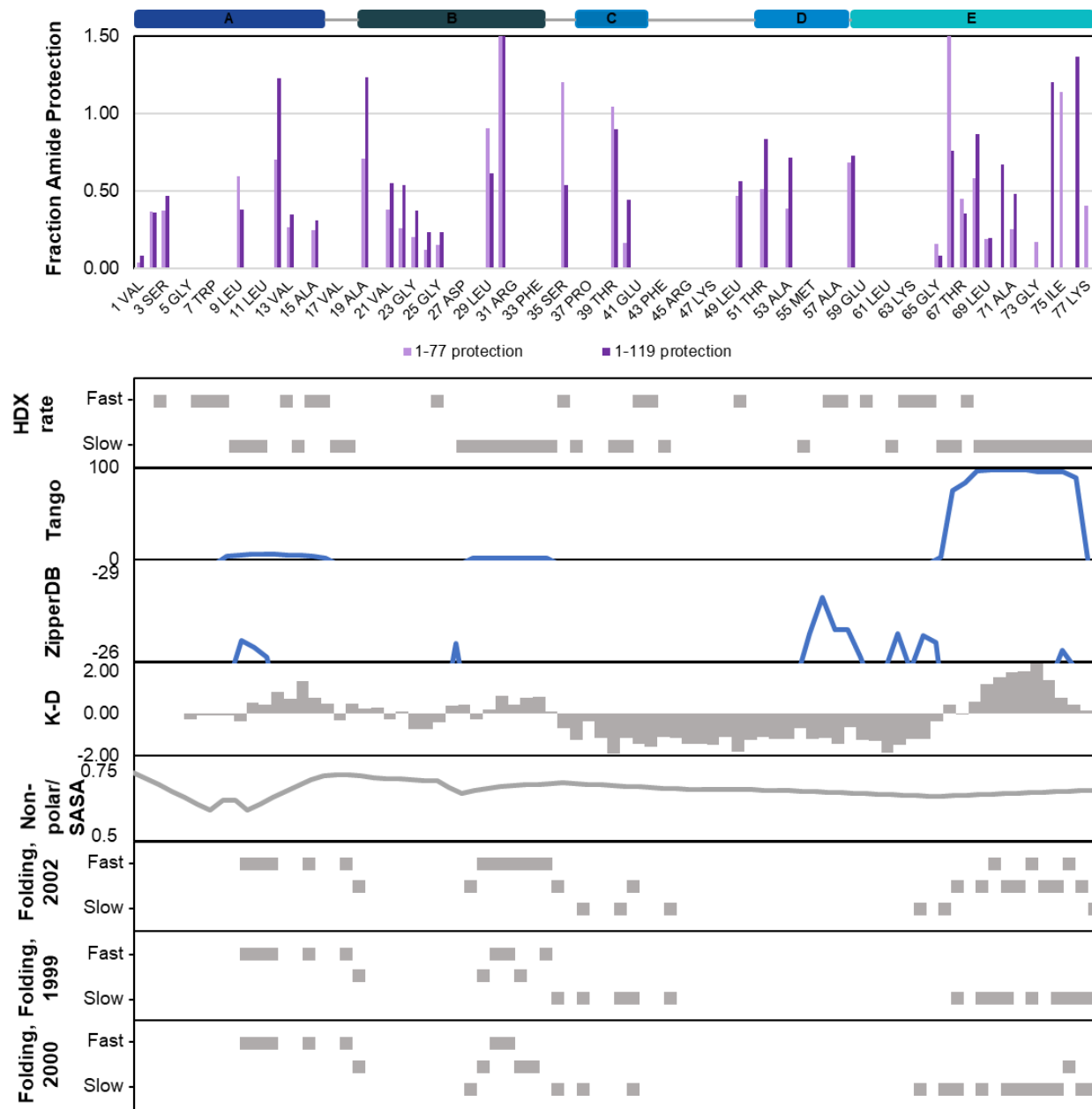


Figure 5. 17 Fraction amide protection of apoMb₁₋₇₇ and apoMb₁₋₁₁₉ for residues 1-77 compared to native and calculated parameters

Fraction amide protection is calculated as described in Chapter 2. HDX rate refers to exchange rate of folded, native, full-length structure as measured by X. Regions of slow exchange are protected in native structure. Tango, ZipperDB, and Kyte-Doolittle hydropathy (KD) predictions are not impacted by truncation as they both use a sliding window, so the same prediction is relevant to both fragments. Non-polar surface area/surface accessible surface area (non-polar/SASA) obtained from X calculated for full-length apoMb. Folding rates 2002, 2000 and 1999 are all for full-length protein. Residues that fold quickly are more likely to be protected in the native structure.

The protection for additional residues in apoMb₁₋₁₁₉ was also determined and is shown in Figure 5.18. Interestingly, almost all additional residues show very low protection, which appears to contradict the hypothesis that is making a native and well folded aggregate. However, the structure of this fragment

has been studied further^[331]. It was demonstrated that the C terminal portion of apoMb₁₋₁₁₉ was highly unstructured. Comparison of previously published C α secondary shift data (Figure 5.18, lower) between the fragment (purple) and the full-length protein (grey) show that the terminal residues on the fragment have very low values, which is consistent with being much less structured (albeit with residual helicity). Further, NMR spectra of soluble apoMb₁₋₁₁₉ aggregates show only a fraction of the expected residues, all of which are C-terminal and found in a very narrow chemical shift dispersion, again indicating that they are highly disordered. This previous work concluded that the N-terminal region of the fragment forms a folded aggregate core, whereas the C-terminal region of the fragment remains dynamic with residual helicity. The current protection pattern shown for apoMb₁₋₁₁₉ IB supports this model for the *in vivo* aggregate as well. Despite the preliminary nature of the chemical shift assignments, all peaks assigned to residues between 78-119 are peaks which do not appear in the apoMb₁₋₇₇ fragment. Thus, while their exact resonance assignment may be incorrect, they are all in the C-terminus of the fragment and can report on this portion of the protein.

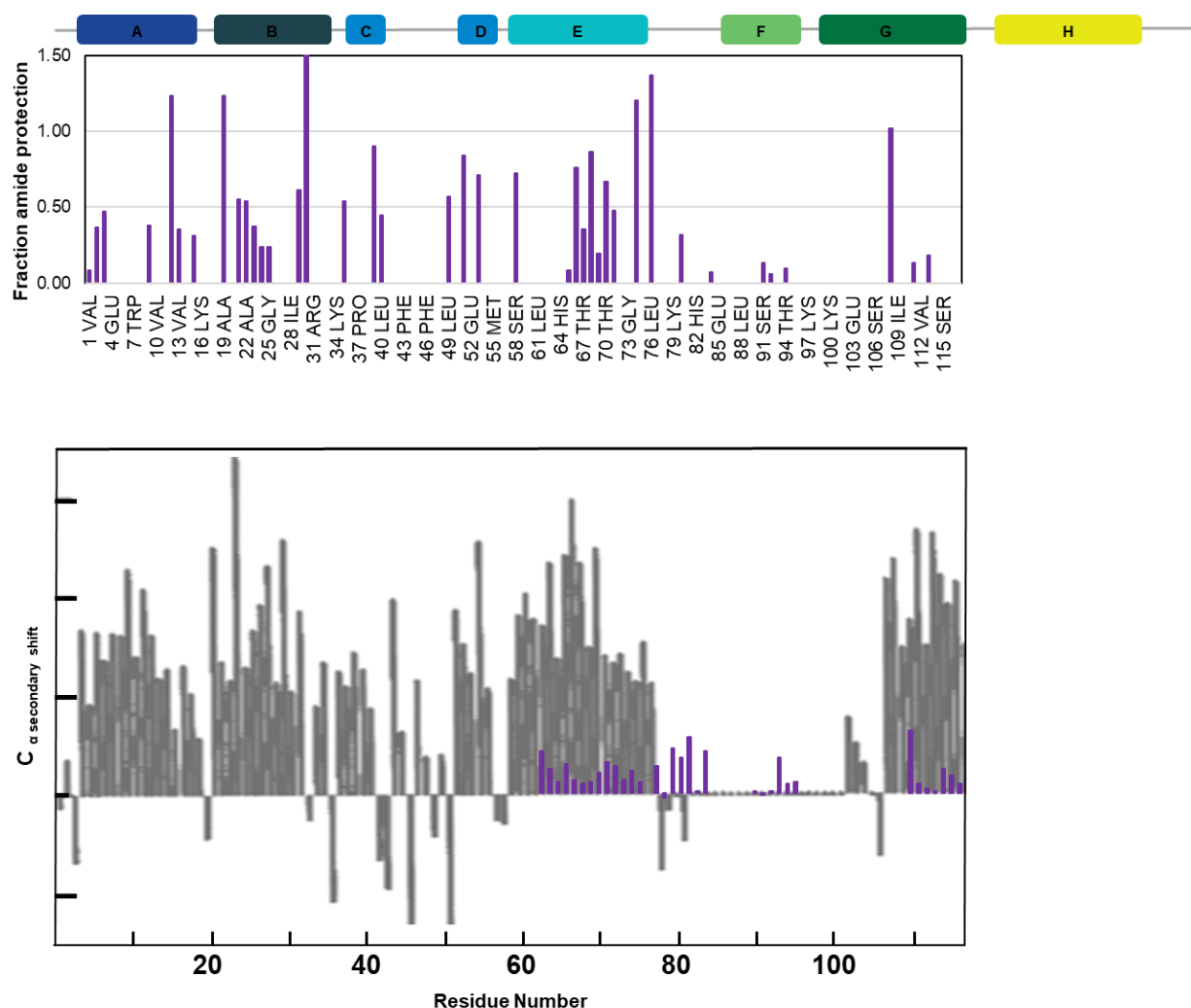


Figure 5.18 Fraction amide protection of apoMb₁₋₁₁₉ for residues 1-119 compared to native structure

Top Fraction amide protection of all residues preliminarily assigned in apoMb₁₋₁₁₉. *Bottom* Comparison of C_{α} secondary shift between the apoMb₁₋₁₁₉ (purple) and the apoMb₁₋₁₅₃ (grey) shows that the terminal region of the apoMb₁₋₁₁₉ fragment is much less structured than the full-length protein. Data is not available for residues 1-60 of apoMb₁₋₁₁₉. Data reproduced from^[331].

5.4 Conclusions

The protein requirement of the qHDX method can be limiting for many protein systems. Adapting the qHDX method to be robust on lower aggregation systems is the logical next step for understanding determinants of aggregates and of aggregate structures. One strategy employed in this chapter was limited isotopic labelling, which was validated on the SOD1 A4V mutant. Results showed that the aggregate at equal expression times was very similar, indicating that the modified protein expression did not greatly impact the aggregate structure. The method allows one to examine much earlier aggregates, which revealed differences in the aggregation structure: namely that the extend of non-native dimers

made may vary. These results corroborate earlier predictions that A4V may have a modified monomer with a different dimer interface.

Application of qHDX to Adn, another β -structured protein, further validated the qHDX method. The majority of the protocol was easily applied to Adn, and assignment strategies readily produced consistent results, giving confidence to the procedure developed. Use of the strategies established from the two systems then allowed for excellent preliminary data on apoMb, a lower-aggregation system with a trove of data on native structure available. apoMb results highlight that structural information on IBs can be identified without before investing a lot of time into obtaining the perfect sample or making financial commitments to double isotope labelling and 3D NMR experiments for resonance assignments. Overall, data in this Chapter show that the qHDX method developed in this thesis is versatile and has potential to be tested or applied to a myriad of protein systems.

Chapter 6: Conclusions and Future work

6.1 Conclusions

This thesis explored the aggregation of the model system SOD1 with the goal of answering the questions “how and why do proteins form IBs?”. Chapter Two explored relationships between aggregation propensity of point mutants to evolutionary parameters such as sequence conservation, biophysical parameters such as protein stability and dimer dissociation, and to predicted parameters of aggregation properties. This chapter ascertained the complicated relationships between proteins and their likelihood to aggregate and highlighted several SOD1 point mutants that were possibly participating in different aggregation processes. Chapter 3 went on to establish a method to study impure and heterogeneous aggregates at high-resolution with confidence, and Chapter 4 applied this method (qHDX by NMR) to 9 SOD1 variants. Combined with low-resolution structural methods, it was determined that SOD1 aggregated via an ensemble of different native and non-native structures, and that the ensemble could be shifted if primary aggregation hot spots were eliminated. The utility of qHDX was further shown in Chapter 5 as the method was modified for application to earlier SOD1 aggregates and other lower-aggregation systems (Adnectin, apomyoglobin). Taken together, the work in this thesis has developed a robust framework for understanding protein aggregation and has opened up many avenues for future work.

6.2 Logical Extensions

qHDX of earlier SOD1 aggregates

The limited isotope labelling demonstrated in Chapter 5 of this thesis showed a viable method for studying earlier aggregates. It additionally demonstrated that the earlier timepoint aggregate had different structural features which would have otherwise been overlooked. Since the approach has been validated, other SOD1 mutants can now be examined at the shorter timepoints. More specifically, G85R and V148G, variants that are more likely to impact the structure (G85R because of its impact on the zinc-binding pocket and V148G for its impact on the dimer interface) have potential to show interesting differences in their earlier structures.

qHDX of additional SOD1 point mutants

Several SOD1 point mutations were described in Chapter 2 of this thesis that are of high interest for further study, but whose study was limited by the protein requirement of qHDX. More specifically, the L106V mutation was suggested to form a modified monomer structure. As was shown in Chapter 4, Loop4 is a key portion of the protein for aggregation, and it is where L106 is located. Applying qHDX to

this variant (likely with the augmentation of limited isotope labelling or detergent washing to compensate for lower protein amounts) is an obvious addition to the data set.

Another variant of specific interest is H46R, which is a low-aggregation variant proposed to mimic a metal-bound species. As aggregates grown with zinc appeared similar apo aggregates by qHDX (but different by urea denaturation), this mimetic that is known to have a different apo monomer than others will provide valuable additional information on the seeming discrepancy.

Metal content of Zn growth condition

As aforementioned, IBs of SOD1 variants grown with zinc appear structurally similar but have differing stabilities. Determination of the metal content of the aggregates via ICP-AES, as has previously been done in the Meiering group on other IBs, may shed light on the contribution of the metal ion.

qHDX of SOD1 truncations

Work on apoMb N-terminal fragments has highlighted the importance of the C-terminus for proper protein folding. Conversely, work on SOD1 aggregate structures have shown that the C-terminus – a main component of the dimer interfaces- is a key component of SOD1 aggregates. An interesting extension of the SOD1 project is to examine SOD1 truncation mutants. A potential choice is G127X, a truncation mutant that is associated with fALS that would remove the entirety of residues involved in dimerization from the C-terminus. This would also leave Loop6 is the primary aggregation region remaining in SOD1, which in principle should show a different type of protection pattern wherein one region (Loop6) is strongly protected, and the rest of the protein is not.

6.3 New directions

Function of IB aggregates

This thesis has postulated that SOD1 IBs have native-like structure. Previous work on IBs that demonstrated native features showed that the IBs had residual enzymatic activity. SOD1 is an enzyme with an established activity assay. Function could not be checked on the IBs in this thesis because they were not grown with the catalytic metals. However, it would be possible to grow SOD1 IBs in LB media, examine their structure through low-resolution methods such as FTIR and CR to see if they resemble those done by qHDX, and then determine their activity via the pyrogallol assay. Another option would be to grow SOD1 IBs in minimal media with metals so that isotopic labelling is possible, and then examine the IBs by qHDX as well as its enzymatic activity.

Structure of soluble IBs: NMR and SEC

It has been demonstrated on all three systems discussed in this thesis that a portion of the protein expressed tends to remain soluble, particularly under certain growth conditions (ie low temperature). However, soluble protein should not be conflated with “properly folded” or “monomeric” protein – a protein can be misfolded or forming lower order oligomers (dimers, trimers) and still be soluble. Examination of the soluble cellular protein fraction may be rich with information on aggregate precursors. Data exists on how well-folded soluble protein appears by NMR for all systems in this thesis. The soluble fraction of an isotopically labelled sample may simply be examined by NMR to see if the protein appears like folded, monomeric protein, like folded dimeric protein (SOD1), like unfolded protein (by comparison to random coil or DMSO denatured samples) or like a mixture of species. Low resolution and peak broadness would be indicators of lower-order aggregated species in the mixture. The soluble fraction can also be analyzed by size exclusion chromatography (SEC). Monomeric, dimeric, and higher-order species would be separated by this method, and relative amounts of them can then be determined by SDS-PAGE. Different amounts of these species in different variants may account for aggregation behaviours that are otherwise not explained by the insoluble fraction.

References

- [1] S. W. Englander, L. Mayne, *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 1–8.
- [2] C. B. Anfinsen, E. Haber, Sela, F. H. White, *Proc. Natl. Acad. Sci.* **1961**, *47*, 1309–1314.
- [3] C. B. Anfinsen, *Science (80-.)*. **1973**, *181*, 223–230.
- [4] V. Daggett, A. Fersht, *Nat. Rev. Mol. Cell Biol.* **2003**, *4*, 497–502.
- [5] P. S. Kim, R. L. Baldwin, *Annu. Rev. Biochem.* **1982**, *51*, 459–489.
- [6] M. Karplus, D. L. Weaver, *Prot. Sci.* **1994**, *3*, 650–668.
- [7] D. B. Wetlaufer, *Proc. Natl. Acad. Sci.* **1973**, *70*, 697–701.
- [8] O. B. Ptitsyn, *J. Protein Chem.* **1987**, *6*, 273–293.
- [9] F. U. Hartl, M. Hayer-Hartl, *Nat. Struct. Mol. Biol.* **2009**, *16*, 574–581.
- [10] C. M. Dobson, *Nature* **2003**, *426*, 884–890.
- [11] H. C. Mahler, W. Friess, U. Grauschopf, S. Kiese, *J. Pharm. Sci.* **2009**, *98*, 2909–2934.
- [12] M. E. M. Cromwell, E. Hilario, F. Jacobson, *AAPS J.* **2006**, *8*, E572–E579.
- [13] D. A. Malencik, S. R. Anderson, *Amino Acids* **2003**, *25*, 233–247.
- [14] F. F. Ferrone, in *Methods Enzymol.*, **1999**, pp. 256–274.
- [15] A. R. Hurshman, J. T. White, E. T. Powers, J. W. Kelly, *Biochemistry* **2004**, *43*, 7365–7381.
- [16] Z. Guo, D. Eisenberg, *J. Biol. Chem.* **2007**, *282*, 14968–14974.
- [17] M. G. Iadanza, M. P. Jackson, E. W. Hewitt, N. A. Ranson, S. E. Radford, *Nat. Rev. Mol. Cell Biol.* **2018**, *19*, 755–773.
- [18] F. Chiti, C. M. Dobson, *Annu. Rev. Biochem.* **2006**, *75*, 333–366.
- [19] F. Chiti, C. M. Dobson, *Nat. Chem. Biol.* **2009**, *5*, 15–22.
- [20] M. Sunde, L. C. Serpell, M. Bartlam, P. E. Fraser, M. B. Pepys, C. C. Blake, *J. Mol. Biol.* **1997**, *273*, 729–739.
- [21] E. Chatani, N. Yamamoto, *Biophys. Rev.* **2018**, *10*, 527–534.
- [22] G. A. Bowden, A. M. Paredes, G. Georgiou, *Nat. Biotechnol.* **1991**, *9*, 725–730.
- [23] A. L. Fink, *Fold. Des.* **1998**, *3*, R9–R23.
- [24] U. Rinas, E. Garcia-Fruitós, J. L. Corchero, E. Vázquez, J. Seras-Franzoso, A. Villaverde, *Trends Biochem. Sci.* **2017**, *42*, 726–737.
- [25] A. A. Reinke, J. E. Gestwicki, *Chem. Biol. Drug Des.* **2011**, *77*, 399–411.
- [26] M. I. Sulatsky, A. I. Sulatskaya, O. I. Povarova, I. A. Antifeeva, I. M. Kuznetsova, K. K. Turoverov, *Prion* **2020**, *14*, 67–75.

- [27] E. I. Yakupova, L. G. Bobyleva, I. M. Vikhlyantsev, A. G. Bobylev, *Biosci. Rep.* **2019**, *39*, 1–22.
- [28] M. Biancalana, S. Koide, *Biochim. Biophys. Acta - Proteins Proteomics* **2010**, *1804*, 1405–1412.
- [29] R. Kaye, E. Head, F. Sarsoza, T. Saing, C. W. Cotman, M. Necula, L. Margol, J. Wu, L. Breydo, J. L. Thompson, S. Rasool, T. Gurlo, P. Butler, C. G. Glabe, *Mol. Neurodegener.* **2007**, *2*, 18 (7 pages).
- [30] S. Pickles, L. Destroismaisons, S. L. Peyrard, S. Cadot, G. A. Rouleau, R. H. Brown, J.-P. Julien, N. Arbour, C. Vande Velde, *Hum. Mol. Genet.* **2013**, *22*, 3947–3959.
- [31] R. Rakhit, J. Robertson, C. Vande Velde, P. Horne, D. M. Ruth, J. Griffin, D. W. Cleveland, N. R. Cashman, A. Chakrabarty, *Nat. Med.* **2007**, *13*, 754–759.
- [32] S. Pickles, S. Semmler, H. R. Broom, L. Destroismaisons, L. Legroux, N. Arbour, E. Meiering, N. R. Cashman, C. Vande Velde, *Acta Neuropathol. Commun.* **2016**, *4*, (14 pages).
- [33] R. S. Atlasi, R. Malik, C. I. Corrales, L. Tzeplaeff, J. P. Whitelegge, N. R. Cashman, G. Bitan, *ACS Chem. Biol.* **2018**, *13*, 2794–2807.
- [34] L. Masino, V. Musi, R. P. Menon, P. Fusi, G. Kelly, T. A. Frenkiel, Y. Trottier, A. Pastore, *FEBS Lett.* **2003**, *549*, 21–25.
- [35] A. Singh, V. Upadhyay, A. Singh, A. K. Panda, *Front. Microbiol.* **2020**, *11*, 1–10.
- [36] P. Polverino De Laureto, N. Taddei, E. Frare, C. Capanni, S. Costantini, J. Zurdo, F. Chiti, C. M. Dobson, A. Fontana, *J. Mol. Biol.* **2003**, *334*, 129–141.
- [37] A. K. Upadhyay, A. Murmu, A. Singh, A. K. Panda, *PLoS One* **2012**, *7*, e33951 (11 pages).
- [38] S. Yoshinaga, T. Yamanaka, H. Miyazaki, A. Okuzumi, A. Hiyama, S. Murayama, N. Nukina, *Biochem. Biophys. Res. Commun.* **2020**, *522*, 655–661.
- [39] F. S. Ruggeri, J. Habchi, A. Cerreta, G. Dietler, *Curr. Pharm. Des.* **2016**, *22*, 3950–3970.
- [40] F. S. Ruggeri, T. Šneirderis, M. Vendruscolo, T. P. Knowles, *Arch. Biochem. Biophys.* **2019**, *664*, 134–148.
- [41] R. Guerrero-Ferreira, L. Kovacik, D. Ni, H. Stahlberg, *Curr. Opin. Neurobiol.* **2020**, *61*, 89–95.
- [42] B. C. Creekmore, Y.-W. Chang, E. B. Lee, *J. Neuropathol. Exp. Neurol.* **2021**, *80*, 494–513.
- [43] J. T. Pelton, L. R. Mclean, *Anal. Biochem.* **2000**, *277*, 167–176.
- [44] R. Sarroukh, E. Goormaghtigh, J. M. Ruysschaert, V. Raussens, *Biochim. Biophys. Acta - Biomembr.* **2013**, *1828*, 2328–2338.
- [45] J. Milošević, R. Prodanović, N. Polović, *Molecules* **2021**, *26*, 970 (15 pages).
- [46] M. Hoshino, H. Katou, Y. Hagihara, K. Hasegawa, H. Naiki, Y. Goto, *Nat. Struct. Biol.* **2002**, *9*, 332–336.
- [47] A. Alexandrescu, *Pacific Symp. Biocomput.* **2001**, *78*, 67–78.
- [48] L. Wang, S. K. Maji, M. R. Sawaya, D. Eisenberg, R. Riek, *PLoS Biol.* **2008**, *6*, 1791–1801.
- [49] D. Naser, M. V Tarasca, B. Siebeneichler, A. Schaefer, H. K. Deol, T. G. Soule, J. Almey, S. Kelso, G. G. Mishra, H. Simon, E. M. Meiering, *Angew. Chemie - Int. Ed.* **2022**, *61*.

- [50] R. Tycko, *Cold Spring Harb. Perspect. Med.* **2016**, 6, a024083.
- [51] C. Seuring, J. Verasdonck, J. Gath, D. Ghosh, N. Nespovitaya, M. A. Wälti, S. K. Maji, R. Cadalbert, P. Güntert, B. H. Meier, R. Riek, *Nat. Struct. Mol. Biol.* **2020**, 27, 1178–1184.
- [52] P. A. Jonsson, K. Ernhill, P. M. Andersen, D. Bergemalm, T. Brännström, O. Gredal, P. Nilsson, S. L. Marklund, *Brain* **2004**, 127, 73–88.
- [53] K. Forsberg, P. A. Jonsson, P. M. Andersen, D. Bergemalm, K. S. Graffmo, M. Hultdin, J. Jacobsson, R. Rosquist, S. L. Marklund, T. Brännström, *PLoS One* **2010**, 5, e11552.
- [54] F. A. Aprile, P. Sormanni, M. Podpolny, S. Chhangur, L.-M. Needham, F. S. Ruggeri, M. Perni, R. Limbicker, G. T. Heller, T. Sneideris, T. Scheidt, B. Mannini, J. Habchi, S. F. Lee, P. C. Salinas, T. P. J. Knowles, C. M. Dobson, M. Vendruscolo, *Proc. Natl. Acad. Sci.* **2020**, 117, 13509–13518.
- [55] Y. Ma, M. Ueda, A. Ueda, S. Shinriki, A. Nagatoshi, A. Isoguchi, M. Okada, M. Tasaki, T. Nomura, Y. Inoue, T. Masuda, Y. Misumi, T. Yamashita, H. Matsui, Y. Ando, *J. Neurol. Sci.* **2020**, 415, 116931.
- [56] R. Kayed, E. Head, J. L. Thompson, T. M. McIntire, S. C. Milton, C. W. Cotman, C. G. Glabe, *Science* **2003**, 300, 486–4899.
- [57] R. Kayed, A. Pensalfini, L. Margol, Y. Sokolov, F. Sarsoza, E. Head, J. Hall, C. Glabe, *J. Biol. Chem.* **2009**, 284, 4230–4237.
- [58] C. G. Glabe, *J. Biol. Chem.* **2008**, 283, 29639–29643.
- [59] L. Breydo, D. Kurouski, S. Rasool, S. Milton, J. W. Wu, V. N. Uversky, I. K. Lednev, C. G. Glabe, *Biochem. Biophys. Res. Commun.* **2016**, 477, 700–705.
- [60] R. Kayed, I. Canto, L. Breydo, S. Rasool, T. Lukacsovich, J. Wu, R. Albay, A. Pensalfini, S. Yeung, E. Head, J. L. Marsh, C. Glabe, *Mol. Neurodegener.* **2010**, 5, 1–10.
- [61] A. Laganowsky, C. Liu, M. R. Sawaya, J. P. Whitelegge, J. Park, M. Zhao, A. Pensalfini, A. B. Soriaga, M. Landau, P. K. Teng, D. Cascio, C. Glabe, D. Eisenberg, *Science (80-.)*. **2012**, 335, 1228–1231.
- [62] S. Pickles, C. Vande Velde, C. Vande Velde, *Amyotroph. Lateral Scler.* **2012**, 13, 333–340.
- [63] E. Ekhtiari Bidhendi, J. Bergh, P. Zetterström, P. M. Andersen, S. L. Marklund, T. Brännström, *J. Clin. Invest.* **2016**, 126, 2249–2253.
- [64] J. I. Ayers, J. Diamond, A. Sari, S. Fromholt, A. Galaleldeen, L. W. Ostrow, J. D. Glass, P. J. Hart, D. R. Borchelt, *Acta Neuropathol.* **2016**, 132, 827–840.
- [65] J. Bergh, P. Zetterström, P. M. Andersen, T. Brännström, K. S. Graffmo, P. A. Jonsson, L. Lang, J. Danielsson, M. Oliveberg, S. L. Marklund, *Proc. Natl. Acad. Sci.* **2015**, 112, 4489–4494.
- [66] B. Combs, C. Hamel, N. M. Kanaan, *Neurobiol. Dis.* **2016**, 94, 18–31.
- [67] C. A. Lasagna-Reeves, D. L. Castillo-Carranza, U. Sengupta, J. Sarmiento, J. Troncoso, G. R. Jackson, R. Kayed, *FASEB J.* **2012**, 26, 1946–1959.
- [68] J. Legleiter, G. P. Lotz, J. Miller, J. Ko, C. Ng, G. L. Williams, S. Finkbeiner, P. H. Patterson, P. J. Muchowski, *J. Biol. Chem.* **2009**, 284, 21647–21658.
- [69] S. Semmler, M. Gagné, P. Garg, S. R. Pickles, C. Baudouin, E. Hamon-Keromen, L.

- Destroismaisons, Y. Khalfallah, M. Chaineau, E. Caron, A. N. Bayne, J. F. Trempe, N. R. Cashman, A. T. Star, A. S. Haqqani, T. M. Durcan, E. M. Meiering, J. Robertson, N. Grandvaux, S. S. Plotkin, H. M. McBride, C. Vande Velde, G. N. DeMartino, *J. Biol. Chem.* **2020**, *295*, 3808–3825.
- [70] T. Yokoyama, K. M. Kimura, Y. Ushiki, S. Yamada, A. Morooka, T. Nakashiba, T. Sassa, S. Itohara, *J. Biol. Chem.* **2001**, *276*, 11265–11271.
- [71] Y. Nekooki-Machida, M. Kurosawa, N. Nukina, K. Ito, T. Oda, M. Tanaka, *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 9679–9684.
- [72] S. Nystro, K. M. Psonka-Antonczyk, G. Ellingsen, L. B. G Johansson, N. Reitan, S. Handrick, S. Prokop, F. L. Heppner, B. M. Wegenast-Braun, M. Jucker, M. Lindgren, B. Torger Stokke, P. Hammarstro, K. R. Peter Nilsson, *ACS Chem. Biol.* **2013**, *8*, 1128–1133.
- [73] A. A. Desai, M. D. Smith, Y. Zhang, E. K. Makowski, J. E. Gerson, E. Ionescu, C. G. Starr, J. M. Zupancic, S. J. Moore, A. B. Sutter, M. I. Ivanova, G. G. Murphy, H. L. Paulson, P. M. Tessier, **2021**, DOI 10.1016/j.jbc.2021.100508.
- [74] A. Barth, *Biochim. Biophys. Acta - Bioenerg.* **2007**, *1767*, 1073–1101.
- [75] C. Stani, L. Vaccari, E. Mitri, G. Birarda, *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* **2020**, *229*, 118006–118012.
- [76] M. Gil-Garcia, S. Navarro, S. Ventura, *Microb. Cell Fact.* **2020**, *19*, 117.
- [77] F. Rashno, K. Khajeh, C. Capitini, R. H. Sajedi, M. M. Shokri, F. Chiti, *Biochim. Biophys. Acta - Proteins Proteomics* **2017**, *1865*, 652–663.
- [78] N. A. Valdez-Cruz, G. I. Reynoso-Cereceda, S. Pérez-Rodriguez, S. Restrepo-Pineda, J. González-Santana, A. Olvera, G. Zavala, A. Alagón, M. A. Trujillo-Roldán, *Microb. Cell Fact.* **2017**, *16*, 1–12.
- [79] F. Rueda, B. Gasser, A. Sánchez-Chardi, M. Roldán, S. Villegas, V. Puxbaum, N. Ferrer-Miralles, U. Unzueta, E. Vázquez, E. Garcia-Fruitós, D. Mattanovich, A. Villaverde, *Microb. Cell Fact.* **2016**, *15*, DOI 10.1186/s12934-016-0565-9.
- [80] G. Musteikyte, M. Ziaunys, V. Smirnovas, *PeerJ* **2020**, *8*, 1–17.
- [81] D. Usoltsev, V. Sitnikova, A. Kajava, M. Uspenskaya, *Biomolecules* **2019**, *9*, 1–17.
- [82] J. Waeytens, V. Van Hemelryck, A. Deniset-Besseau, J. M. Ruyschaert, A. Dazzi, V. Raussens, *Molecules* **2020**, *25*, DOI 10.3390/molecules25122899.
- [83] N. Wilkosz, M. Czaja, S. Seweryn, K. Skirlinska-Nosek, M. Szymonski, E. Lipiec, K. Sofinska, *Molecules* **2020**, *25*, 1–26.
- [84] G. Zandomeneghi, M. R. H. Krebs, M. G. McCammon, M. Fändrich, *Protein Sci.* **2004**, *13*, 3314–3321.
- [85] F. Vosough, A. Barth, *ACS Chem. Neurosci.* **2021**, *12*, 473–488.
- [86] E. Goormaghtigh, V. Cabiaux, J. M. Ruyschaert, in *Physicochem. Methods Study Biomembr. Subcell. Biochem.* (Eds.: H.J. Hilderson, G.B. Ralston), Springer, Boston, MA, **1994**, pp. 329–362.
- [87] E. Cerf, R. Sarroukh, S. Tamamizu-Kato, L. Breydo, S. Derclayes, Y. F. Dufrênes, V. Narayanaswami, E. Goormaghtigh, J. M. Ruyschaert, V. Raussens, *Biochem. J.* **2009**, *421*, 415–423.

- [88] L. Zhou, D. Kurouski, *Anal. Chem.* **2020**, 92, 6806–6810.
- [89] H. Yang, S. Yang, J. Kong, A. Dong, S. Yu, *Nat. Protoc.* **2015**, 10, 382–396.
- [90] B. Shivu, S. Seshadri, J. Li, K. A. Oberg, V. N. Uversky, A. L. Fink, *Biochemistry* **2013**, 52, 5176–5183.
- [91] L. M. Miller, M. Gragg, T. G. Kim, P. S. H. Park, *FEBS Lett.* **2015**, 589, 3119–3125.
- [92] M. Morell, R. Bravo, A. Espargaró, X. Sisquella, F. X. Avilés, X. Fernàndez-Busquets, S. Ventura, *Biochim. Biophys. Acta - Mol. Cell Res.* **2008**, 1783, 1815–1825.
- [93] O. Cano-Garrido, A. Sánchez-Chardi, S. Parés, I. Giró, W. I. Tatkiewicz, N. Ferrer-Miralles, I. Ratera, A. Natalello, R. Cubarsi, J. Veciana, À. Bach, A. Villaverde, A. Arís, E. Garcia-Fruitós, *Acta Biomater.* **2016**, 43, 230–239.
- [94] A. Vera, N. González-Montalbán, A. Arís, A. Villaverde, *Biotechnol. Bioeng* **2007**, 96, 1101–1106.
- [95] J. Fan, L. Huang, J. Sun, Y. Qiu, J. Zhou, Y. Shen, *J. Biotechnol.* **2015**, 209, 16–22.
- [96] W. Wu, L. Xing, B. Zhou, Z. Lin, *Microb. Cell Fact.* **2011**, 10, 1–8.
- [97] S. M. Doglia, D. Ami, A. Natalello, P. Gatti-Lafranconi, M. Lotti, *Biotechnol. J.* **2008**, 3, 193–201.
- [98] S. Jevševar, V. Gaberc-Porekar, I. Fonda, B. Podobnik, J. Grdadolnik, V. Menart, *Biotechnol. Prog.* **2005**, 21, 632–639.
- [99] C. Calcines-Cruz, A. Olvera, R. M. Castro-Acosta, G. Zavala, A. Alagón, M. A. Trujillo-Roldán, N. A. Valdez-Cruz, *Int. J. Biol. Macromol.* **2018**, 108, 826–836.
- [100] M. Bonda, V. Perrin, B. Vilenó, H. Runne, A. Kretlow, L. Forró, R. Luthi-Carter, L. M. Miller, S. Jeney, *Anal. Chem.* **2011**, 83, 7712–7720.
- [101] D. Ami, B. Sciandrone, P. Mereghetti, J. Falvo, T. Catelani, C. Visentin, P. Tortora, S. Ventura, A. Natalello, M. E. Regonesi, *Int. J. Mol. Sci.* **2021**, 22, 1–21.
- [102] D. Ami, F. Lavatelli, P. Rognoni, G. Palladini, S. Raimondi, S. Giorgetti, L. Monti, S. M. Doglia, A. Natalello, G. Merlini, *Sci. Rep.* **2016**, 6, 29096.
- [103] B. Byrne, J. W. Beattie, C. L. Song, S. G. Kazarian, *ATR-FTIR Spectroscopy and Spectroscopic Imaging of Proteins.*, **2020**.
- [104] F. Santos, S. Magalhaes, M. C. Henriques, M. Fardilha, A. Nunes, *Curr. Metabolomics* **2018**, 6, 103–111.
- [105] C. Li, J. Zhao, K. Cheng, Y. Ge, Q. Wu, Y. Ye, G. Xu, Z. Zhang, W. Zheng, X. Zhang, X. Zhou, G. Pielak, M. Liu, *Annu. Rev. Anal. Chem.* **2017**, 10, 157–182.
- [106] S. Reckel, J. J. Lopez, F. Löhr, C. Glaubitz, V. Dötsch, *ChemBioChem* **2012**, 13, 534–537.
- [107] Patrick C A van der Wel, *Solid State Nucl. Magn. Reson.* **2017**, 8, 1–14.
- [108] A. Daskalov, N. El Mammeri, A. Lends, J. Shenoy, G. Lamon, Y. Fichou, A. Saad, D. Martinez, E. Morvan, M. Berbon, A. Grélard, B. Kauffmann, M. Ferber, B. Bardiaux, B. Habenstein, S. J. Saupe, A. Loquet, *Front. Mol. Neurosci.* **2021**, 14, 1–18.

- [109] V. S. Mandala, M. Hong, *Curr. Opin. Struct. Biol.* **2019**, *58*, 183–190.
- [110] R. Rogawski, A. E. McDermott, *Arch. Biochem. Biophys.* **2017**, *628*, 102–113.
- [111] K. P. Scherpelz, S. Wang, P. Pytel, R. S. Madhurapantula, A. K. Srivastava, J. R. Sachleben, J. Orgel, Y. Ishii, S. C. Meredith, *Sci. Rep.* **2021**, *11*, 1–14.
- [112] J.-X. Lu, W. Qiang, W.-M. Yau, C. D. Schwieters, S. C. Meredith, R. Tycko, *Cell* **2013**, *154*, 1257–1268.
- [113] U. Ghosh, K. R. Thurber, W. M. Yau, R. Tycko, *Proc. Natl. Acad. Sci. U. S. A.* **2021**, *118*, DOI 10.1073/pnas.2023089118.
- [114] W. Qiang, W. M. Yau, J. X. Lu, J. Collinge, R. Tycko, *Nature* **2017**, *541*, 217–221.
- [115] D. Li, C. Liu, *Nat. Chem. Biol.* **2021**, *17*, 237–245.
- [116] Z. Chang, J. Deng, W. Zhao, J. Yang, *Biochem. Biophys. Res. Commun.* **2020**, *532*, 655–661.
- [117] T. Pradhan, K. Annamalai, R. Sarkar, S. Huhn, U. Hegenbart, S. Schönland, M. Fändrich, B. Reif, *J. Biol. Chem.* **2020**, *295*, 18474–18484.
- [118] B. R. Groveman, M. A. Dolan, L. M. Taubner, A. Kraus, R. B. Wickner, B. Caughey, *J. Biol. Chem.* **2014**, *289*, 24129–24142.
- [119] T. Theint, P. S. Nadaud, D. Aucoin, J. J. Helmus, S. P. Pondaven, K. Surewicz, W. K. Surewicz, C. P. Jaroniec, *Nat. Commun.* **2017**, *8*, DOI 10.1038/s41467-017-00794-z.
- [120] M. D. Gelenter, K. J. Smith, S. Y. Liao, V. S. Mandala, A. J. Dregni, M. S. Lamm, Y. Tian, W. Xu, D. J. Pochan, T. J. Tucker, Y. Su, M. Hong, *Nat. Struct. Mol. Biol.* **2019**, *26*, 592–598.
- [121] C. Wasmer, L. Benkemoun, R. Sabaté, M. O. Steinmetz, B. Coulary-Salin, L. Wang, R. Riek, S. J. Saupe, B. H. Meier, *Angew. Chemie - Int. Ed.* **2009**, *48*, 4858–4860.
- [122] M. Dasari, A. Espargaro, R. Sabate, J. M. Lopez Del Amo, U. Fink, G. Grelle, J. Bieschke, S. Ventura, B. Reif, *ChemBioChem* **2011**, *12*, 407–423.
- [123] G. F. Taylor, S. P. Wood, K. Mörs, C. Glaubitz, J. M. Werner, P. T. F. Williamson, *ChemBioChem* **2011**, *12*, 556–558.
- [124] J. Curtis-Fisk, R. M. Spencer, D. P. Weliky, *J. Am. Chem. Soc.* **2008**, *130*, 12568–12569.
- [125] E. P. Vogel, J. Curtis-Fisk, K. M. Young, D. P. Weliky, *Biochemistry* **2011**, *50*, 10013–10026.
- [126] K. K. Frederick, V. K. Michaelis, M. A. Caporini, L. B. Andreas, G. T. Debelouchina, R. G. Griffin, S. Lindquist, *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114*, 3642–3647.
- [127] T. Schubeis, T. Lühns, C. Ritter, *ChemBioChem* **2015**, *16*, 51–54.
- [128] K. Hun Lim, A. K. R Dasari, I. Hung, Z. Gan, J. W. Kelly, P. E. Wright, D. E. Wemmer, **2016**, DOI 10.1021/acs.biochem.6b00649.
- [129] T. Schubeis, P. Yuan, M. Ahmed, M. Nagaraj, B. J. Van Rossum, C. Ritter, *Angew. Chemie - Int. Ed.* **2015**, *54*, 14669–14672.
- [130] S. Narasimhan, C. Pinto, A. Lucini Paioni, J. van der Zwan, G. E. Folkers, M. Baldus, *Nat. Protoc.*

- 2021**, *16*, 893–918.
- [131] Y. Z. Zhang, Y. Paterson, H. Roder, *Protein Sci.* **1995**, *4*, 804.
 - [132] T. Arakawa, Y. Kita, S. N. Timasheff, *Biophys. Chem.* **2007**, *131*, 62–70.
 - [133] M. Hoshino, H. Katou, K. ichi Yamaguchi, Y. Goto, *Biochim. Biophys. Acta - Biomembr.* **2007**, *1768*, 1886–1899.
 - [134] A. T. Alexandrescu, *Methods Mol. Biol.* **2016**, *1345*, 211–222.
 - [135] J. H. Ippel, A. Olofsson, J. Rgen Schleucher, E. Lundgren, S. S. Wijmenga, *PNAS* **2002**, *99*, 2021.
 - [136] N. A. Whittemore, R. Mishra, I. Kheterpal, A. D. Williams, R. Wetzel, E. H. Serpersu, *Biochemistry* **2005**, *44*, 4434–4441.
 - [137] A. Olofsson, A. E. Sauer-Eriksson, A. Öhman, *J. Biol. Chem.* **2006**, *281*, 477–483.
 - [138] L. Wang, D. Schubert, M. R. Sawaya, D. Eisenberg, R. Riek, *Angew. Chemie - Int. Ed.* **2010**, *49*, 3904–3908.
 - [139] M. V Tarasca, D. Naser, A. Schaefer, T. G. Soule, E. M. Meiering, *Anal. Biochem.* **2022**, 652.
 - [140] M. V Tarasca, D. Naser, A. Schaefer, T. G. Soule, E. M. Meiering, *Quenched Hydrogen-Deuterium Amide Exchange Optimization for High-Resolution Structural Analysis of Cellular Protein Aggregates*, **2022**.
 - [141] N. Carulla, M. Zhou, E. Giralt, C. V Robinson, C. M. Dobson, *Acc. Chem. Res.* **2010**, *43*, 1072–1079.
 - [142] E. I. James, T. A. Murphree, C. Vorauer, J. R. Engen, M. Guttman, *Chem. Rev.* **2021**, [acs.chemrev.1c00279](https://doi.org/10.1021/acs.chemrev.1c00279).
 - [143] G. R. Masson, J. E. Burke, N. G. Ahn, G. S. Anand, C. Borchers, S. Brier, G. M. Bou-Assaf, J. R. Engen, S. W. Englander, J. Faber, R. Garlish, P. R. Griffin, M. L. Gross, M. Guttman, Y. Hamuro, A. J. R. Heck, D. Houde, R. E. Iacob, T. J. D. Jørgensen, I. A. Kaltashov, J. P. Klinman, L. Konermann, P. Man, L. Mayne, B. D. Pascal, D. Reichmann, M. Skehel, J. Snijder, T. S. Strutzenberg, E. S. Underbakke, C. Wagner, T. E. Wales, B. T. Walters, D. D. Weis, D. J. Wilson, P. L. Wintrode, Z. Zhang, J. Zheng, D. C. Schriemer, K. D. Rand, *Nat. Methods* **2019**, *16*, 595–602.
 - [144] D. L. Smith, Y. Deng, Z. Zhang, *J. Mass Spectrom.* **1997**, *32*, 135–146.
 - [145] J. R. Engen, T. E. Wales, *Annu. Rev. Anal. Chem.* **2015**, *8*, 127–148.
 - [146] V. Smirnovas, G. S. Baron, D. K. Offerdahl, G. J. Raymond, B. Caughey, W. K. Surewicz, *Nat. Struct. Mol. Biol.* **2011**, *18*, 504–506.
 - [147] Q. Li, F. Wang, X. Xiao, C. Kim, J. Bohon, J. Kiselar, J. G. Safar, J. Ma, W. K. Surewicz, *J. Biol. Chem.* **2018**, *293*, 18494–18503.
 - [148] C. Kim, X. Xiao, S. Chen, T. Haldiman, V. Smirnovas, D. Kofskey, M. Warren, K. Surewicz, N. R. Maurer, Q. Kong, W. Surewicz, J. G. Safar, *Nat. Commun.* **2018**, *9*, 1–11.
 - [149] M. Kaldmäe, A. Leppert, G. Chen, M. Sarr, C. Sahin, K. Nordling, N. Kronqvist, M. Gonzalvo-Ulla, N. Fritz, A. Abelein, S. Laín, H. Biverstål, H. Jörnvall, D. P. Lane, A. Rising, J. Johansson, M. Landreh, *FEBS J.* **2020**, *287*, 2823–2833.

- [150] G. Ölçücü, O. Klaus, K. E. Jaeger, T. Drepper, U. Krauss, *ACS Sustain. Chem. Eng.* **2021**, *9*, 8919–8945.
- [151] U. Krauss, V. D. Jäger, M. Diener, M. Pohl, K.-E. Jaeger, *J. Biotechnol.* **2017**, *258*, 136–147.
- [152] A. de Marco, N. Ferrer-Miralles, E. García-Fruitós, A. Mitraki, S. Peternel, U. Rinas, M. A. Trujillo-Roldán, N. A. Valdez-Cruz, E. Vázquez, A. Villaverde, *FEMS Microbiol. Rev.* **2019**, *43*, 53–72.
- [153] A. Espargaró, A. Medina, O. Di Pietro, D. Muñoz-Torrero, R. Sabate, **2016**, DOI 10.1038/srep23349.
- [154] M. V. Céspedes, Y. Fernández, U. Unzueta, R. Mendoza, J. Seras-Franzoso, A. Sánchez-Chardi, P. Álamo, V. Toledo-Rubio, N. Ferrer-Miralles, E. Vázquez, S. Schwartz, I. Abasolo, J. L. Corchero, R. Mangues, A. Villaverde, *Sci. Rep.* **2016**, *6*, 1–10.
- [155] P. Singhvi, A. Saneja, S. Srichandan, A. K. Panda, *Trends Biotechnol.* **2020**, *38*, 474–486.
- [156] W. M. Babinchak, W. K. Surewicz, *J. Mol. Biol.* **2020**, *432*, 1910–1925.
- [157] M. I. Ivanova, Y. Lin, Y. H. Lee, J. Zheng, A. Ramamoorthy, *Biophys. Chem.* **2021**, *269*, 106507.
- [158] D. Aucoin, Y. Xia, T. Theint, P. S. Nadaud, K. Surewicz, W. K. Surewicz, C. P. Jaroniec, *J. Struct. Biol.* **2019**, *206*, 36–42.
- [159] E. A. Hodge, M. A. Benhaim, K. K. Lee, *Protein Sci.* **2020**, *29*, 843–855.
- [160] M. Tang, D. Lam, *Solid State Nucl. Magn. Reson.* **2019**, *103*, 9–16.
- [161] D. C. Williams, R. M. Van Frank, W. L. Muth, J. P. Burnett, *Science (80-.)*. **1982**, *215*, 687–689.
- [162] R. Wetzel, L. Jeanne Perry, C. Veilleux, *Bio/Technology* **1991**, *9*, 731–737.
- [163] B. A. Chranyaksp, J. Evans\$, J. Lillquistol, P. Youngti, R. Wetzel, \$ li, *J. Biol. Chem.* **1993**, *268*, 18053–18061.
- [164] C. H. Schein, *Bio/Technology* 1989 711 **1989**, *7*, 1141–1149.
- [165] F. A. O Marston, *Biochem. J* **1986**, *240*, 1–12.
- [166] K. Oberg, B. A. Chrunk, R. Wetzel, A. L. Fink, *Biochemistry* **1994**, *33*, 2628–2634.
- [167] E. García-Fruitós, N. González-Montalbán, M. Morell, A. Vera, R. M. Ferraz, A. Arís, S. Ventura, A. Villaverde, *Microb. Cell Fact.* **2005**, *4*, 1–6.
- [168] K. Tokatlidis, P. Dhurjati, J. Millet, P. Béguin, J. P. Aubert, *FEBS Lett.* **1991**, *282*, 205–208.
- [169] E. García-Fruitós, A. Arís, A. Villaverde, *Appl. Environ. Microbiol.* **2007**, *73*, 289.
- [170] J. P. Arié, M. Miot, N. Sassoon, J. M. Betton, *Mol. Microbiol.* **2006**, *62*, 427–437.
- [171] E. García-Fruitós, M. M. Carrió, A. Arís, A. Villaverde, *Biotechnol. Bioeng.* **2005**, *90*, 869–875.
- [172] A. Vera, N. González-Montalbán, A. Arís, A. Villaverde, *Biotechnol. Bioeng.* **2007**, *96*, 1101–1106.
- [173] N. González-Montalbán, E. García-Fruitós, S. Ventura, A. Arís, A. Villaverde, *Microb. Cell Fact.* **2006**, *5*, 1–9.

- [174] E. García-Fruitós, *Microb. Cell Factories* 2010 91 **2010**, 9, 1–3.
- [175] A. Ramón, M. Señorale-Pose, M. Marín, *Front. Microbiol.* **2014**, 5, 56.
- [176] J. Tyedmers, A. Mogk, B. Bukau, *Nat. Rev. Mol. Cell Biol.* 2010 1111 **2010**, 11, 777–788.
- [177] J. S. Valentine, P. A. Doucette, S. Zittin Potter, *Annu. Rev. Biochem.* **2005**, 74, 563–593.
- [178] R. Rakhit, A. Chakrabartty, *Biochim. Biophys. Acta* **2006**, 1762, 1025–37.
- [179] C. M. Doyle, J. A. Rumfeldt, H. R. Broom, A. Broom, P. B. Stathopoulos, K. A. Vassall, J. J. Almey, E. M. Meiering, *Arch Biochem Biophys* **2013**, 531, 44–64.
- [180] H. R. Broom, J. A. O. Rumfeldt, E. M. Meiering, *Essays Biochem.* **2014**, 56, 149–65.
- [181] A. Kerman, H.-N. N. Liu, S. Croul, J. Bilbao, E. Rogaeva, L. Zinman, J. Robertson, A. Chakrabartty, *Acta Neuropathol.* **2010**, 119, 335–344.
- [182] Q. Wang, J. L. Johnson, N. Y. R. Agar, J. N. Agar, *PLoS Biol* **2008**, 6, e170.
- [183] M. J. Lindberg, J. Normark, A. Holmgren, M. Oliveberg, *Proc. Natl. Acad. Sci. U. S. A.* **2004**, 101, 15893–8.
- [184] K. a Vassall, H. R. Stubbs, H. a Primmer, M. S. Tong, S. M. Sullivan, R. Sobering, S. Srinivasan, L.-A. K. Briere, S. D. Dunn, W. Colón, E. M. Meiering, *Proc. Natl. Acad. Sci. U. S. A.* **2011**, 108, 2210–2215.
- [185] P. B. Stathopoulos, J. A. O. Rumfeldt, F. Karbassi, C. A. Siddall, J. R. Lepock, E. M. Meiering, *J. Biol. Chem.* **2006**, 281, 6184–6193.
- [186] L. Banci, I. Bertini, F. Cantini, T. Kozyreva, C. Massagni, P. Palumaa, J. T. Rubino, K. Zovo, *Proc. Natl. Acad. Sci. U. S. A.* **2012**, 109, 13555–60.
- [187] T. Endo, T. Fujii, K. Sato, N. Taniguchi, J. Fujii, *Biochem. Biophys. Res. Commun.* **2000**, 276, 999–1004.
- [188] J. J. Goto, H. Zhu, R. J. Sanchez, A. Nersissian, E. B. Gralla, J. S. Valentine, D. E. Cabelli, *J Biol Chem* **2000**, 275, 1007–1014.
- [189] A. M. Fernandez-Escamilla, F. Rousseau, J. Schymkowitz, L. Serrano, *Nat. Biotechnol.* **2004**, 22, 1302–1306.
- [190] M. J. Thompson, S. A. Sievers, J. Karanicolas, M. I. Ivanova, D. Baker, D. Eisenberg, *Proc. Natl. Acad. Sci. U. S. A.* **2006**, 103, 4074–4078.
- [191] P. Sormanni, F. A. Aprile, M. Vendruscolo, *J. Mol. Biol.* **2015**, 427, 478–490.
- [192] K. Trainor, Z. Gingras, C. Shillingford, H. Malakian, *J. Mole* **2016**, 428, 1365–137414.
- [193] J. Almey, Inclusion Body Formation of Familial Amyotrophic Lateral Sclerosis Associated Cu , Zn-Superoxide Dismutase Mutants in Escherichia Coli By, University of Waterloo, **2015**.
- [194] H. R. Broom, J. A. O. Rumfeldt, K. A. Vassall, E. M. Meiering, *Protein Sci.* **2015**, 24, 2081–2089.
- [195] K. A. Vassall, P. B. Stathopoulos, J. A. O. Rumfeldt, J. R. Lepock, E. M. Meiering, **2006**, 7366–7379.
- [196] A. C. Tsois, N. C. Papandreou, V. A. Iconomidou, S. J. Hamodrakas, *PLoS One* **2013**, 8, 1–6.

- [197] J. Kyte, R. F. Doolittle, *J. Mol. Biol.* **1982**, *157*, 105–132.
- [198] H. Ashkenazy, S. Abadi, E. Martz, O. Chay, I. Mayrose, T. Pupko, N. Ben-Tal, *Nucleic Acids Res.* **2016**, *44*, W344–W350.
- [199] G. Rodrigues, C. Pereira, B. Rbara De Azevedo, A. Vieira, J. F. De Mesquita Id, **2021**, DOI 10.1371/journal.pone.0247841.
- [200] M. Bernhofer, C. Dallago, T. Karl, V. Satagopam, M. Heinzinger, M. Littmann, T. Olenyi, J. Qiu, K. Schütze, G. Yachdav, H. Ashkenazy, N. Ben-Tal, Y. Bromberg, T. Goldberg, L. Kajan, S. O'Donoghue, C. Sander, A. Schafferhans, A. Schlessinger, G. Vriend, M. Mirdita, P. Gawron, W. Gu, Y. Jarosz, C. Trefois, M. Steinegger, R. Schneider, B. Rost, *bioRxiv* **2021**, 2021.02.23.432527.
- [201] G. E. Crooks, G. Hon, J.-M. Chandonia, S. E. Brenner, **n.d.**, DOI 10.1101/gr.849004.
- [202] M. I. Ivanova, S. A. Sievers, E. L. Guenther, L. M. Johnson, D. D. Winkler, A. Galaleldeen, M. R. Sawaya, P. J. Hart, D. S. Eisenberg, *Proc. Natl. Acad. Sci.* **2014**, *111*, 197–201.
- [203] L. Banci, L. Barbieri, I. Bertini, F. Cantini, E. Luchinat, *PLoS One* **2011**, *6*, 1–8.
- [204] K. A. Vassall, H. R. Stubbs, H. A. Primmer, M. S. Tong, S. M. Sullivan, R. Sobering, S. Srinivasan, L.-A. K. Briere, S. D. Dunn, W. Colón, E. M. Meiering, *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 2210–2215.
- [205] H. K. Deol, H. R. Broom, B. Siebenechler, B. Lee, Z. Leonenko, E. M. Meiering, *Biophys. Chem.* **2022**.
- [206] C. K. Smith, L. Regan, J. M. Withka, *Biochemistry* **2002**, *33*, 5510–5517.
- [207] H. K. Deol, **n.d.**
- [208] C. M. Karch, M. Prudencio, D. D. Winkler, P. J. Hart, D. R. Borchelt, *Proc. Natl. Acad. Sci.* **2009**, *106*, 7774–7779.
- [209] C. M. Doyle, J. A. Rumfeldt, H. R. Broom, A. Sekhar, L. E. Kay, E. M. Meiering, *Biochemistry* **2016**, *55*, 1346–1361.
- [210] P. O. Souillac, V. N. Uversky, A. L. Fink, *Biochemistry* **2003**, *42*, 8094–8104.
- [211] F. Bemporad, F. Chiti, *Chem. Biol.* **2012**, *19*, 315–327.
- [212] T. Yang, S. Li, H. Xu, D. M. Walsh, D. J. Selkoe, *J. Neurosci.* **2017**, *37*, 152.
- [213] C. Haass, D. J. Selkoe, *Nat. Rev. Mol. Cell Biol.* **2007**, *8*, 101–112.
- [214] S. De, D. C. Wirthensohn, P. Flagmeier, C. Hughes, F. A. Aprile, F. S. Ruggeri, D. R. Whiten, D. Emin, Z. Xia, J. A. Varela, P. Sormanni, F. Kundel, T. P. J. Knowles, C. M. Dobson, C. Bryant, M. Vendruscolo, D. Klenerman, *Nat. Commun.* **2019**, *10*, 1–11.
- [215] S. De, D. R. Whiten, F. S. Ruggeri, C. Hughes, M. Rodrigues, D. I. Sideris, C. G. Taylor, F. A. Aprile, S. Muyldermans, T. P. J. Knowles, M. Vendruscolo, C. Bryant, K. Blennow, I. Skoog, S. Kern, H. Zetterberg, D. Klenerman, *Acta Neuropathol. Commun.* **2019**, *7*, 120.
- [216] G. T. Corbett, Z. Wang, W. Hong, M. Colom-Cadena, J. Rose, M. Liao, A. Asfaw, T. C. Hall, L. Ding, A. DeSousa, M. P. Frosch, J. Collinge, D. A. Harris, M. S. Perikinton, T. L. Spires-Jones, T. L. Young-

- Pearse, A. Billinton, D. M. Walsh, *Acta Neuropathol.* **2020**, *139*, 503–526.
- [217] M. Stefani, *Methods Mol. Biol.* **2010**, *648*, 25–41.
- [218] J. A. Rodriguez, B. F. Shaw, A. Durazo, S. H. Sohn, P. A. Doucette, A. M. Nersissian, K. F. Faull, D. K. Eggers, A. Tiwari, L. J. Hayward, J. S. Valentine, *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 10516–10521.
- [219] A. M. Crown, B. L. Roberts, K. Crosby, H. Brown, J. I. Ayers, P. J. Hart, D. R. Borchelt, *G3 Genes/Genomes/Genetics* **2019**, *9*, 719.
- [220] S. D. Khare, N. V. Dokholyan, *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 3147–3152.
- [221] C. Parrini, N. Taddei, M. Ramazzotti, D. Degl’Innocenti, G. Ramponi, C. M. Dobson, F. Chiti, *Structure* **2005**, *13*, 1143–1151.
- [222] J. Dou, A. A. Vorobieva, W. Sheffler, L. A. Doyle, H. Park, M. J. Bick, B. Mao, G. W. Foight, M. Y. Lee, L. A. Gagnon, L. Carter, B. Sankaran, S. Ovchinnikov, E. Marcos, P. S. Huang, J. C. Vaughan, B. L. Stoddard, D. Baker, *Nat. 2018 5617724* **2018**, *561*, 485–491.
- [223] J. S. Elam, A. B. Taylor, R. Strange, S. Antonyuk, P. A. Doucette, J. A. Rodriguez, S. S. Hasnain, L. J. Hayward, J. S. Valentine, T. O. Yeates, P. J. Hart, *Nat. Struct. Mol. Biol.* **2003**, *10*, 461–467.
- [224] H. Primmer, *Predicting and Measuring Molecular Mechanisms of Protein Aggregation.*, University of Waterloo, **2011**.
- [225] A. Sekhar, J. A. O. Rumfeldt, H. R. Broom, C. M. Doyle, R. E. Sobering, E. M. Meiering, L. E. Kay, *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113*, E6939–E6945.
- [226] A. L. Fink, *Fold. Des.* **1998**, *3*, R9–R23.
- [227] N. S. de Groot, R. Sabate, S. Ventura, *Trends Biochem. Sci.* **2009**, *34*, 408–416.
- [228] E. García-Fruitós, E. Vázquez, C. Díez-Gil, J. L. Corchero, J. Seras-Franzoso, I. Ratera, J. Veciana, A. Villaverde, *Trends Biotechnol.* **2012**, *30*, 65–70.
- [229] U. Rinas, E. Garcia-Fruitós, J. L. Corchero, E. Vázquez, J. Seras-Franzoso, A. Villaverde, *Trends Biochem. Sci.* **2017**, *42*, 726–737.
- [230] O. Cano-Garrido, A. Sánchez-Chardi, S. Parés, I. Giró, W. I. Tatkiewicz, N. Ferrer-Miralles, I. Ratera, A. Natalello, R. Cubarsi, J. Veciana, À. Bach, A. Villaverde, A. Arís, E. Garcia-Fruitós, *Acta Biomater.* **2016**, *43*, 230–239.
- [231] A. Ramón, M. Señorale-Pose, M. Marín, *Front. Microbiol.* **2014**, *5*, 2010–2015.
- [232] P. Singhvi, A. Saneja, S. Srichandan, A. K. Panda, *Trends Biotechnol.* **2020**, *38*, 474–486.
- [233] V. D. Jäger, R. Kloss, A. Grünberger, S. Seide, D. Hahn, T. Karmainski, M. Piqueray, J. Embruch, S. Longerich, U. Mackfeld, K.-E. Jaeger, W. Wiechert, M. Pohl, U. Krauss, *Microb. Cell Fact.* **2019**, *18*, 33.
- [234] S. T. T. Schetters, W. S. P. Jong, L. J. W. Kruijssen, H. B. van den B. van Saparoea, S. Engels, W. W. J. Unger, D. Houben, J. M. M. den Haan, J. Luirink, Y. van Kooyk, *Cell. Mol. Immunol.* **2020**, *17*, 415–417.

- [235] L. Wang, <http://dx.doi.org/10.4161/pri.3.3.9922> **2009**, *3*, 139–145.
- [236] A. Schaefer, D. Naser, B. Siebeneichler, M. Tarasca, E. Meiering, *J. Biol. Chem.* **2022**, *Submitted*, Invited review.
- [237] L. Wang, S. K. Maji, M. R. Sawaya, D. Eisenberg, R. Riek, *PLoS Biol.* **2008**, *6*, 1791–1801.
- [238] L. Wang, D. Schubert, M. R. R. Sawaya, D. Eisenberg, R. Riek, *Angew. Chemie - Int. Ed.* **2010**, *49*, 3904–3908.
- [239] M. Hoshino, H. Katou, K. ichi Yamaguchi, Y. Goto, *Biochim. Biophys. Acta - Biomembr.* **2007**, *1768*, 1886–1899.
- [240] A. T. Alexandrescu, *Methods Mol. Biol.* **2016**, *1345*, 211–22.
- [241] Y. Zhang, H. Roder, Y. Paterson, *Protein Sci.* **1995**, *4*, 804–814.
- [242] A. T. Alexandrescu, *Pac. Symp. Biocomput.* **2001**, *78*, 67–78.
- [243] K. Annamalai, F. Liberta, M.-T. Vielberg, W. Close, H. Lilie, K.-H. Gührs, A. Schierhorn, R. Koehler, A. Schmidt, C. Haupt, U. Hegenbart, S. Schönland, M. Schmidt, M. Groll, M. Fändrich, *Angew. Chemie Int. Ed.* **2017**, *56*, 7510–7514.
- [244] V. Smirnovas, G. S. Baron, D. K. Offerdahl, G. J. Raymond, B. Caughey, W. K. Surewicz, *Nat. Struct. Mol. Biol.* **2011**, *18*, 504–506.
- [245] P. B. Stathopoulos, G. A. Scholz, Y.-M. Hwang, J. A. O. Rumfeldt, J. R. Lepock, E. M. Meiering, *Protein Sci.* **2008**, *13*, 3017–3027.
- [246] L. E. Kay, M. Ikura, R. Tschudin, A. Bax, *J. Magn. Reson.* **1990**, *89*, 496–514.
- [247] S. Grzesiek, A. Bax, *J. Am. Chem. Soc.* **1992**, *114*, 6291–6293.
- [248] D. R. Muhandiram, L. E. Kay, *J. Magn. Reson. Ser. B* **1994**, *103*, 203–216.
- [249] R. T. Clubb, V. Thanabal, G. Wagner, **n.d.**
- [250] M. Wittekind, L. Mueller, *J. Magn. Reson. Ser. B* **1993**, *101*, 201–205.
- [251] R. Weisemann, H. Rüterjans, W. Bermel, *J. Biomol. NMR* **1993**, *3*, 113–120.
- [252] R. L. J. Keller, **2004**.
- [253] S. P. Skinner, R. H. Fogh, W. Boucher, T. J. Ragan, L. G. Mureddu, G. W. Vuister, *J. Biomol. NMR* **2016**, *66*, 111–124.
- [254] Y. Bai, S. J. Milne, L. Mayne, S. W. Englander, *Proteins Struct. Funct. Genet.* **1993**, *17*, 75–86.
- [255] S. M. Damo, A. H. Phillips, A. L. Young, S. Li, V. L. Woods, D. E. Wemmer, D. SM, P. AH, Y. AL, L. S, W. VL, W. DE, *J. Biol. Chem.* **2010**, *285*, 32303–32311.
- [256] M. Vilar, L. Wang, R. Riek, *Structural Studies of Amyloids by Quenched Hydrogen-Deuterium Exchange by NMR*, **2012**.
- [257] F. G. Bordwell, *Acc. Chem. Res.* **1988**, *21*, 456–463.
- [258] B. Moorthy, L. Iyer, E. Topp, *Curr. Pharm. Des.* **2015**, *21*, 5845–5853.

- [259] D. S. Wishart, C. G. Bigam, A. Holm, R. S. Hodges, B. D. Sykes, *J. Biomol. NMR* **1995**, *51* **1995**, 5, 67–81.
- [260] D. Williams, R. Van Frank, W. Muth, J. Burnett, *Science* (80-.). **1982**, *215*, 687–689.
- [261] B. A. Chrnyk, J. Evans, J. Lillquist, P. Young, R. Wetzel, *J. Biol. Chem.* **1993**, *268*, 18053–61.
- [262] A. Mitraki, J. King, *Nat. Biotechnol.* **1989**, *7*, 690–697.
- [263] U. Rinas, E. Garcia-Fruitós, J. L. Corchero, E. Vázquez, J. Seras-Franzoso, A. Villaverde, *Trends Biochem. Sci.* **2017**, *42*, 726–737.
- [264] P. C. Ke, R. Zhou, L. C. Serpell, R. Riek, T. P. J. Knowles, H. A. Lashuel, E. Gazit, I. W. Hamley, T. P. Davis, M. Fändrich, D. E. Otzen, M. R. Chapman, C. M. Dobson, D. S. Eisenberg, R. Mezzenga, *Chem. Soc. Rev.* **2020**, *49*, 5473–5509.
- [265] P. Singhvi, A. Saneja, S. Srichandan, A. K. Panda, *Trends Biotechnol.* **2020**, *38*, 474–486.
- [266] K. Trainor, Z. Gingras, C. Shillingford, H. Malakian, M. Gosselin, D. Lipovsek, E. M. Meiering, *J. Mol. Biol.* **2016**, *428*, 1365–1374.
- [267] U. Krauss, V. D. Jäger, M. Diener, M. Pohl, K. E. Jaeger, *J. Biotechnol.* **2017**, *258*, 136–147.
- [268] M. V. Céspedes, O. Cano-Garrido, P. Álamo, R. Sala, A. Gallardo, N. Serna, A. Falgàs, E. Voltà-Durán, I. Casanova, A. Sánchez-Chardi, H. López-Laguna, L. Sánchez-García, J. M. Sánchez, U. Unzueta, E. Vázquez, R. Mangues, A. Villaverde, *Adv. Mater.* **2020**, *32*, 1907348(1)-1907348(9).
- [269] F. Chiti, C. M. Dobson, *Annu. Rev. Biochem.* **2017**, *86*, 27–68.
- [270] N. S. de Groot, R. Sabate, S. Ventura, *Trends Biochem. Sci.* **2009**, *34*, 408–416.
- [271] G. S. A. Wright, S. V. Antonyuk, S. S. Hasnain, *Q. Rev. Biophys.* **2019**, *52*, e12.
- [272] Y. Furukawa, T. V. O’Halloran, *J. Biol. Chem.* **2005**, *280*, 17266–17274.
- [273] E. D. Getzoff, D. E. Cabelli, C. L. Fisher, H. E. Parge, M. S. Viezzoli, L. Banci, R. A. Hallewell, *Nature* **1992**, *358*, 347–351.
- [274] P. B. Stathopoulos, J. A. O. Rumfeldt, G. A. Scholz, R. A. Irani, H. E. Frey, R. A. Hallewell, J. R. Lepock, E. M. Meiering, *Proc Natl Acad Sci U S A* **2003**, *100*, 7021–7026.
- [275] K. A. Oberg, A. L. Fink, *Anal. Biochem.* **1998**, *256*, 92–106.
- [276] W. E. Klunk, R. F. Jacob, R. P. Mason, *Methods Enzymol.* **1999**, *309*, 285–305.
- [277] A. Sekhar, J. A. O. Rumfeldt, H. R. Broom, C. M. Doyle, G. Bouvignies, E. M. Meiering, L. E. Kay, *Elife* **2015**, *4*, e07296.
- [278] G. Calloni, S. Zoffoli, M. Stefani, C. M. Dobson, F. Chiti, *J. Biol. Chem.* **2005**, *280*, 10607–13.
- [279] V. Castillo, A. Espargaró, V. Gordo, J. Vendrell, S. Ventura, *Proteomics* **2010**, *10*, 4172–4185.
- [280] A. K. Museth, A.-C. Brorsson, M. Lundqvist, L. A. E. Tibell, B.-H. Jonsson, *Biochemistry* **2009**, *48*, 8817–8829.
- [281] L. Banci, I. Bertini, F. Cramaro, R. Del Conte, M. S. Viezzoli, *Biochemistry* **2003**, *42*, 9543–9553.

- [282] M. I. Ivanova, S. A. Sievers, E. L. Guenther, L. M. Johnson, D. D. Winkler, A. Galaleldeen, M. R. Sawaya, P. J. Hart, D. S. Eisenberg, *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 197–201.
- [283] S. Sangwan, A. Zhao, K. L. Adams, C. K. Jayson, M. R. Sawaya, E. L. Guenther, A. C. Pan, J. Ngo, D. M. Moore, A. B. Soriaga, T. D. Do, L. Goldschmidt, R. Nelson, M. T. Bowers, C. M. Koehler, D. E. Shaw, B. G. Novitch, D. S. Eisenberg, *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114*, 8770–8775.
- [284] J. S. Elam, A. B. Taylor, R. Strange, S. Antonyuk, P. A. Doucette, J. A. Rodriguez, S. S. Hasnain, L. J. Hayward, J. S. Valentine, T. O. Yeates, P. J. Hart, *Nat. Struct. Biol.* **2003**, *10*, 461–467.
- [285] R. W. Strange, S. Antonyuk, M. A. Hough, P. A. Doucette, J. A. Rodriguez, P. J. Hart, L. J. Hayward, J. S. Valentine, S. S. Hasnain, *J. Mol. Biol.* **2003**, *328*, 877–891.
- [286] A. Nordlund, L. Leinartaitė, K. Saraboji, C. Aisenbrey, G. Gröbner, P. Zetterström, J. Danielsson, D. T. Logan, M. Oliveberg, *Proc. Natl. Acad. Sci.* **2009**, *106*, 9667–9672.
- [287] A. T. Alexandrescu, *PLoS One* **2013**, *8*, DOI 10.1371/journal.pone.0056467.
- [288] K. I. Yamaguchi, H. Katou, M. Hoshino, K. Hasegawa, H. Naiki, Y. Goto, *J. Mol. Biol.* **2004**, *338*, 559–571.
- [289] Y.-Z. Zhang, *Protein Pept. Struct. Interact. Stud. by Hydrog. Exch. NMR, PhD Thesis, Univ. Pennsylvania* **1995**.
- [290] Y. Bai, J. Milne, L. Mayne, S. Englander, *Proteins* **1993**, *17*, 75–86.
- [291] P. B. Stathopoulos, G. A. Scholz, Y.-M. Hwang, J. A. O. Rumfeldt, J. R. Lepock, E. M. Meiering, *Protein Sci.* **2004**, *13*, 3017–3027.
- [292] P. Andjus, S. Stamenković, T. Dučić, *Eur. Biophys. J.* **2019**, *48*, 475–484.
- [293] R. Sarroukh, E. Goormaghtigh, J. M. Ruysschaert, V. Raussens, *Biochim. Biophys. Acta - Biomembr.* **2013**, *1828*, 2328–2338.
- [294] K. Annamalai, F. Liberta, M. T. Vielberg, W. Close, H. Lilie, K. H. Gührs, A. Schierhorn, R. Koehler, A. Schmidt, C. Haupt, U. Hegenbart, S. Schönland, M. Schmidt, M. Groll, M. Fändrich, *Angew. Chemie - Int. Ed.* **2017**, *56*, 7510–7514.
- [295] K. A. Oberg, A. L. Fink, *Tech. Protein Chem.* **1995**, *6*, 475–484.
- [296] M. Levitt, J. Greer, *J. Mol. Biol.* **1977**, *114*, 181–239.
- [297] A. Kerman, H. N. Liu, S. Croul, J. Bilbao, E. Rogaeva, L. Zinman, J. Robertson, A. Chakrabarty, *Acta Neuropathol.* **2010**, *119*, 335–344.
- [298] B. A. Shirley, *Methods Mol. Biol.* **1995**, *40*, 177–190.
- [299] U. Rinas, F. Hoffmann, E. Betiku, D. Estapé, S. Marten, *J. Biotechnol.* **2007**, *127*, 244–257.
- [300] L. Banci, L. Barbieri, I. Bertini, F. Cantini, E. Luchinat, *PLoS One* **2011**, *6*, e23561(1)–e23561(8).
- [301] W. Wu, L. Xing, B. Zhou, Z. Lin, *Microb. Cell Fact.* **2011**, *10*:9, 1–8.
- [302] A. De Marco, N. Ferrer-Mirallès, E. García-Fruitós, A. Mitraki, S. Peternel, U. Rinas, M. A. Trujillo-Roldán, N. A. Valdez-Cruz, E. Vázquez, A. Villaverde, *FEMS Microbiol. Rev.* **2019**, *43*, 53–72.

- [303] M. Fändrich, M. Schmidt, *Biochem. Soc. Trans.* **2021**, *49*, 977–985.
- [304] F. S. Ruggeri, G. Longo, S. Faggiano, E. Lipiec, A. Pastore, G. Dietler, *Nat. Commun.* **2015**, *6*, DOI 10.1038/ncomms8831.
- [305] D. Röhr, B. D. C. Boon, M. Schuler, K. Kremer, J. J. M. Hoozemans, F. H. Bouwman, S. F. El-Mashtoly, A. Nabers, F. Großerueschkamp, A. J. M. Rozemuller, K. Gerwert, *Acta Neuropathol. Commun.* **2020**, *8*, DOI 10.1186/s40478-020-01091-5.
- [306] J. Bergh, P. Zetterström, P. M. Andersen, T. Brännström, K. S. Graffmo, P. A. Jonsson, L. Lang, J. Danielsson, M. Oliveberg, S. L. Marklund, *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112*, 4489–4494.
- [307] V. Smirnovas, G. S. Baron, D. K. Offerdahl, G. J. Raymond, B. Caughey, W. K. Surewicz, *Nat. Struct. Mol. Biol.* **2011**, *18*, 504–506.
- [308] B. Barz, O. O. Olubiyi, B. Strodel, *Chem. Commun.* **2014**, *50*, 5373–5375.
- [309] F. E. Cohen, J. W. Kelly, *Nat.* **2003** *426*, 905–909.
- [310] S. Lesné, T. K. Ming, L. Kotilinek, R. Kaye, C. G. Glabe, A. Yang, M. Gallagher, K. H. Ashe, *Nat.* **2006** *440*, 352–357.
- [311] D. M. Walsh, I. Klyubin, J. V. Fadeeva, W. K. Cullen, R. Anwyl, M. S. Wolfe, M. J. Rowan, D. J. Selkoe, *Nat.* **2002** *416*, 535–539.
- [312] G. Wei, N. Mousseau, P. Derreumaux, <http://dx.doi.org/10.4161/pri.1.1.3969> **2007**, *1*, 3–8.
- [313] D. J. Donnelly, R. Adam Smith, P. Morin, D. Lipovsek, J. Gokemeijer, D. Cohen, V. Lafont, T. Tran, E. L. Cole, M. Wright, J. Kim, A. Pena, D. Kukral, D. D. Dischino, P. Chow, J. Gan, O. Adelakun, X. T. Wang, K. Cao, D. Leung, S. J. Bonacorsi, W. Hayes, *Synthesis and Biologic Evaluation of a Novel 18 F-Labeled Adnectin as a PET Radioligand for Imaging PD-L1 Expression*, **2018**.
- [314] P. G. Chandler, A. M. Buckle, *Cells* **2020**, *9*, DOI 10.3390/cells9030610.
- [315] D. Lipovšek, *Protein Eng. Des. Sel.* **2011**, *24*, 3–9.
- [316] K. Trainor, Z. Gingras, C. Shillingford, H. Malakian, M. Gosselin, D. Lipovsek, E. M. Meiering, *J. Mol. Biol.* **2016**, *428*, 1–14.
- [317] R. E. Canfield, C. B. Anfinsen, *Biochemistry* **1963**, *2*, 1073–1078.
- [318] H. Taniuchi, C. B. Anfinsen, *J. Biol. Chem.* **1969**, *244*, 3864–3875.
- [319] D. Shortle, A. K. Meeker, *Biochemistry* **1989**, *28*, 936–944.
- [320] Y. Feng, D. Liu, J. Wang, *J. Mol. Biol.* **2003**, *330*, 821–837.
- [321] P. A. Jennings, P. E. Wright, *Science* **1993**, *262*, 892–896.
- [322] C. C. Chow, C. Chow, V. Raghunathan, T. J. Huppert, E. B. Kimball, S. Cavagnero, *Biochemistry* **2003**, *42*, 7090–7099.
- [323] F. Yang, G. N. Phillips, *J. Mol. Biol.* **1996**, *256*, 762–774.
- [324] D. V. Fedyukina, S. Rajagopalan, A. Sekhar, E. C. Fulmer, Y. J. Eun, S. Cavagnero, *Biophys. J.* **2010**, *99*, L37.

- [325] J. Yao, J. Chung, D. Eliezer, P. E. Wright, H. J. Dyson, *Biochemistry* **2001**, *40*, 3561–3571.
- [326] M. S. Chandak, T. Nakamura, K. Makabe, T. Takenaka, A. Mukaiyama, T. K. Chaudhuri, K. Kato, K. Kuwajima, *J. Mol. Biol.* **2013**, *425*, 2541–2560.
- [327] A. Waterhouse, M. Bertoni, S. Bienert, G. Studer, G. Tauriello, R. Gumienny, F. T. Heer, T. A. P. De Beer, C. Rempfer, L. Bordoli, R. Lepore, T. Schwede, *Nucleic Acids Res.* **2018**, *46*, W296–W303.
- [328] R. M. Murphy, A. M. Tsai, **n.d.**
- [329] C. Nishimura, H. Jane Dyson, P. E. Wright, *J. Mol. Biol.* **2002**, *322*, 483–489.
- [330] C. Garcia, C. Nishimura, S. Cavagnero, H. J. Dyson, P. E. Wright, *Biochemistry* **2000**, *39*, 11227–11237.
- [331] S. Rajagopalan, N. Kurt, S. Cavagnero, *Biophys. J.* **2011**, *100*, 747.
- [332] J. Stangroom, “P Value from Pearson (R) Calculator,” can be found under <https://www.socscistatistics.com/pvalues/pearsondistribution.aspx>, **n.d.**
- [333] A. I. Karsisiotis, O. M. Deacon, C. Macdonald, T. M. A. Blumenschein, G. R. Moore, J. A. R. Worrall, *Biomol. NMR Assign.* **2017**, *11*, 165–168.
- [334] N. V. Kulminskaya, Y. Yoshimura, K. Runager, C. S. Sørensen, M. Bjerring, M. Andreassen, D. E. Otzen, J. J. Enghild, N. C. Nielsen, F. A. A. Mulder, *Biomol. NMR Assign.* **2016**, *10*, 25–29.
- [335] M. S. Chandak, T. Nakamura, K. Makabe, T. Takenaka, A. Mukaiyama, T. K. Chaudhuri, K. Kato, K. Kuwajima, *J. Mol. Biol.* **2013**, *425*, 2541–2560.

Appendix

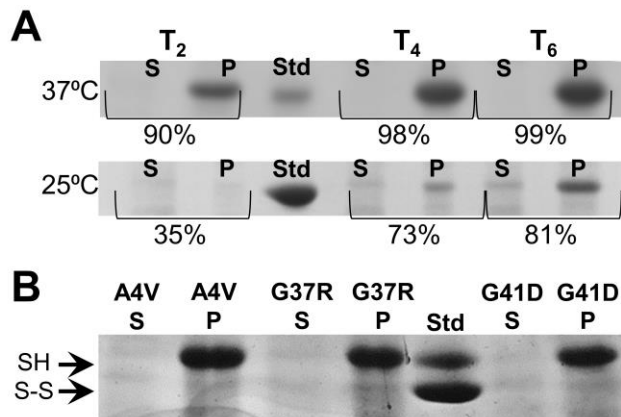


Figure A.2. 2 Example SDS-PAGE used to assess aggregation propensity

(A) Samples of cell culture taken at 2, 4, and 6 hours post induction at 37 °C (top) and 25 °C (bottom) were lysed and centrifuged, then soluble (S) and insoluble (P) fractions were analyzed by 15% SDS-PAGE. Gel densitometry was performed on the SOD1 band in each lane using TotalLab 100 software package (Non-linear Dynamics) with rolling ball background subtraction. %SOD1 in IB = $P/(S+P) \times 100\%$. (B) Samples taken at 4 hours post induction at 37 °C were treated with iodoacetamide to label free thiol groups in Cys residues, then analyzed by 15% non-reducing SDS-PAGE to determine the thiol status of the cysteine residues in the IBs. SOD1 with Cys reduced (SH) or oxidized (S-S) migrate differently in SDS-PAGE. The IB pellets contain primarily fully reduced SOD1

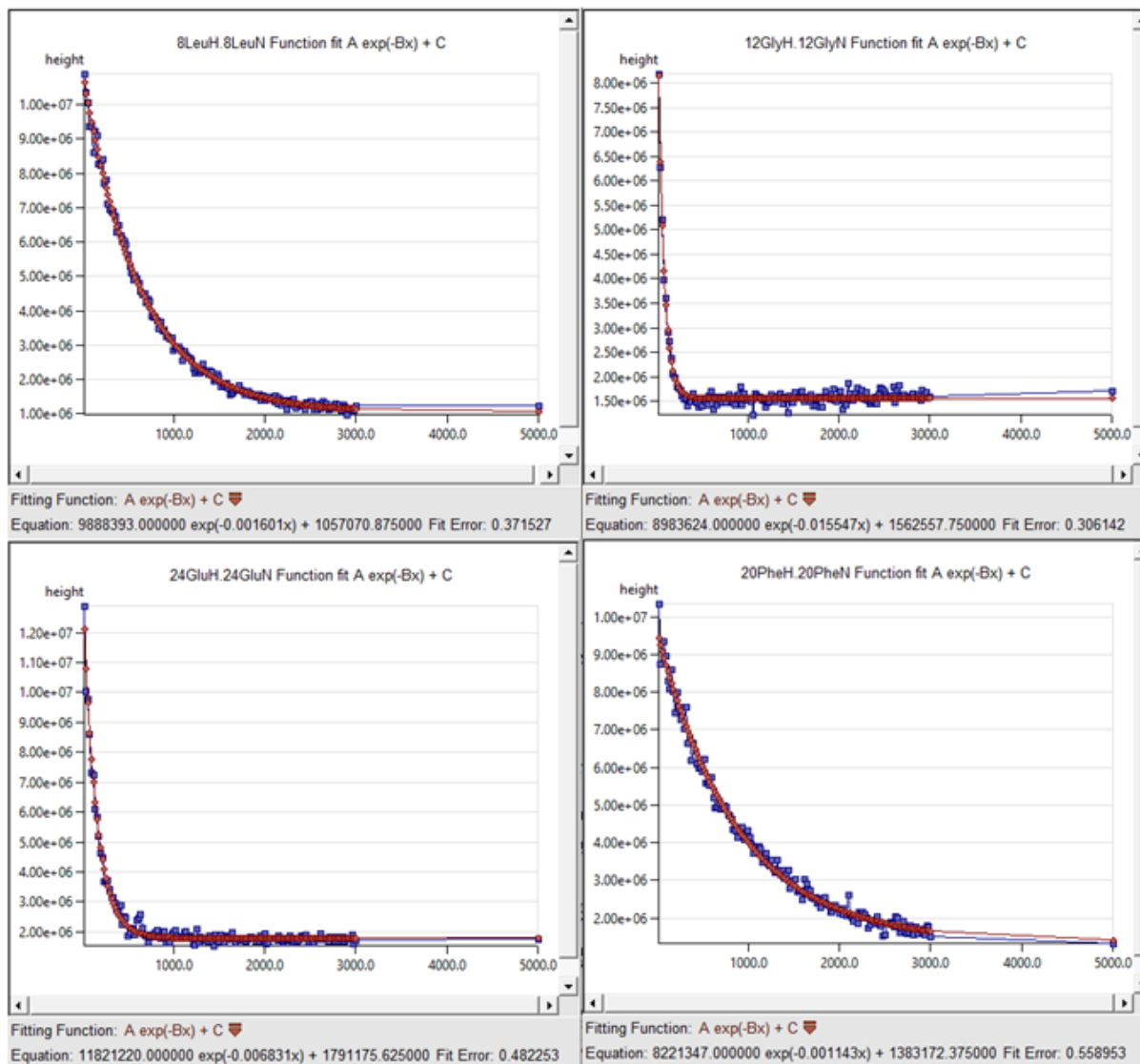


Figure A.3. 3 Amide exchange decay fits in CCPNMR software

CCPNMR produces an output of the intensities of each assigned cross peak in the series of ^1H - ^{15}N HSQC spectra fit as a function of time. Here, a series of 150 ^1H - ^{15}N HSQC spectra were obtained, assignments were transferred automatically throughout the series after manual transfer to one spectrum. The output produces a fit in the form of $(I = I_0 \cdot e^{(-k \cdot x)} + C)$, with I being the intensity at time x , I_0 being the initial intensity, $-k$ being the k_{int} , and C being the signal of the cross peak at HD exchange equilibrium.

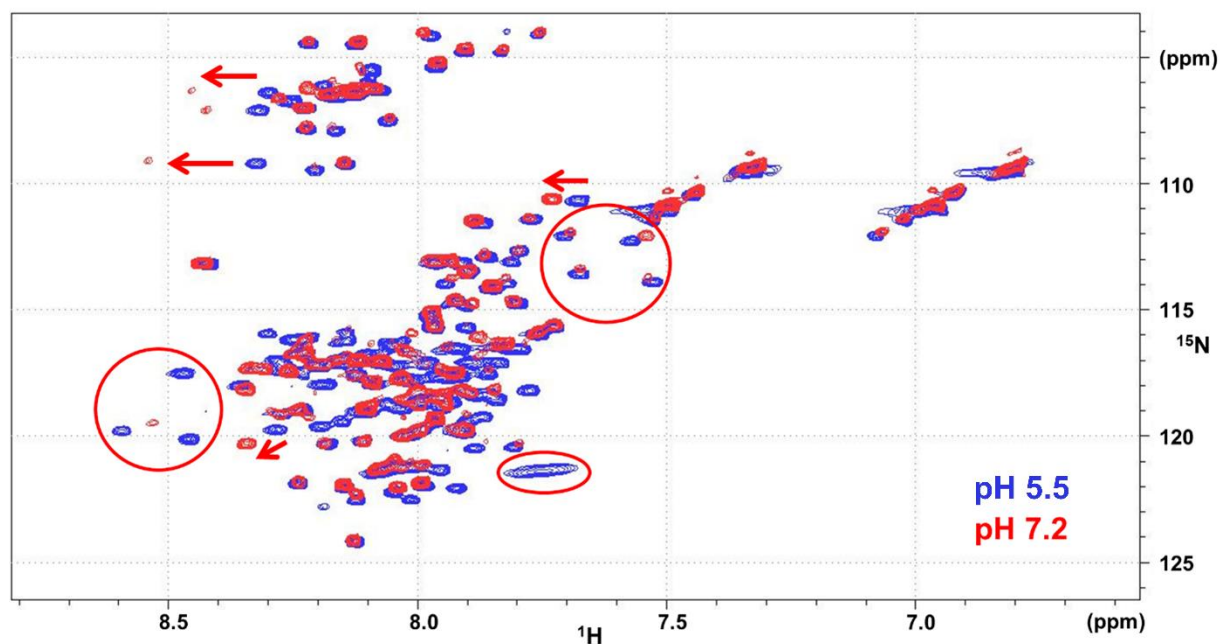


Figure A.3. 4 Effects of pH on ^1H - ^{15}N HSQC cross peak intensity and chemical shift

Replicate samples with different pH values. Areas where signal intensities differ greatly are circled, and areas where there are larger differences in chemical shifts are marked with arrows. The random coil region ($\sim 7.8 - 8.2$ ppm) also displays larger variability in peak intensity and chemical shift. Due to these variations the sample pH must be constant throughout experiments.

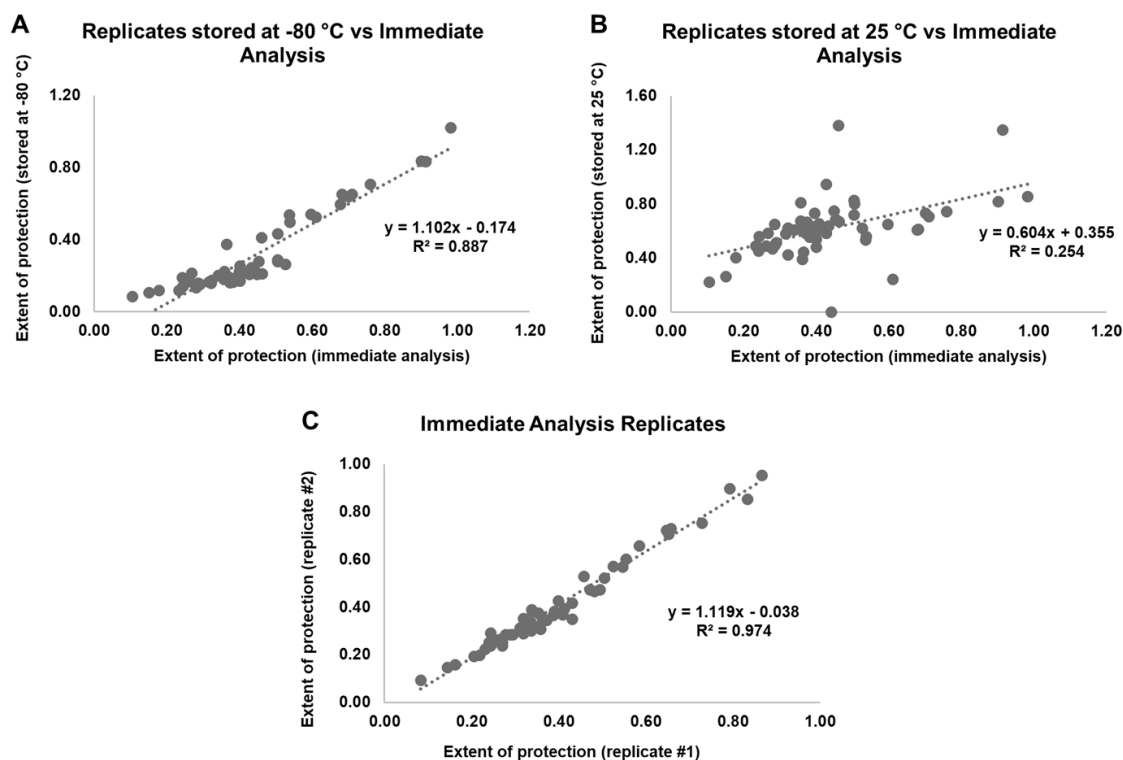


Figure A.3. 5 Effects of varying IB storage

A) Replicate samples, one analyzed immediately after lyophilization compared with one analyzed after storing the sample at -80 °C for three months. Some changes are observed, notably for residues with lower protection. B) Replicate samples analyzed immediately after lyophilization or are storage at 25 °C for two weeks. Extensive differences are observed, likely due to signal scrambling due to H-D exchange in the lyophilized sample. C) Replicate samples, both analyzed immediately after lyophilization. The high correlation shows that the most consistent results are achieved by analyzing samples immediately after lyophilization.

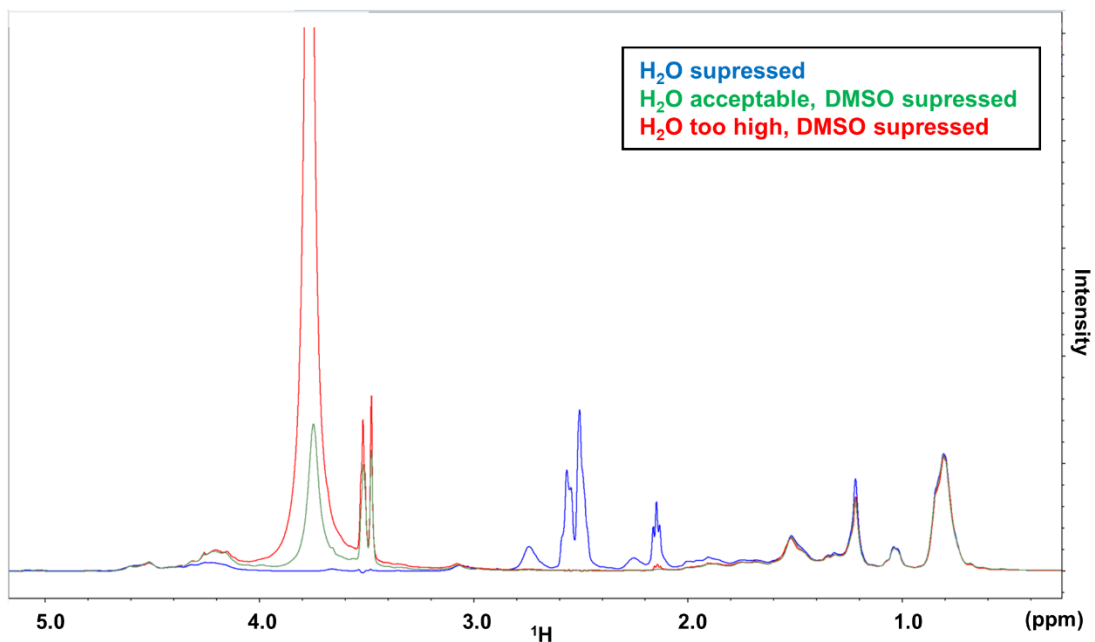


Figure A.3. $6\ ^1\text{H}$ 1D NMR spectra overlaid to highlight differences in water content across samples.

Here we have three spectra: Blue) a spectrum where the H_2O signal (3.815 ppm) is suppressed, allowing for the residual DMSO signal (2.512 ppm) to be observed; Green) a spectrum where the DMSO signal is suppressed and the H_2O signal is deemed acceptable; and Red) a spectrum where the DMSO signal is suppressed, and the H_2O signal is deemed too high.

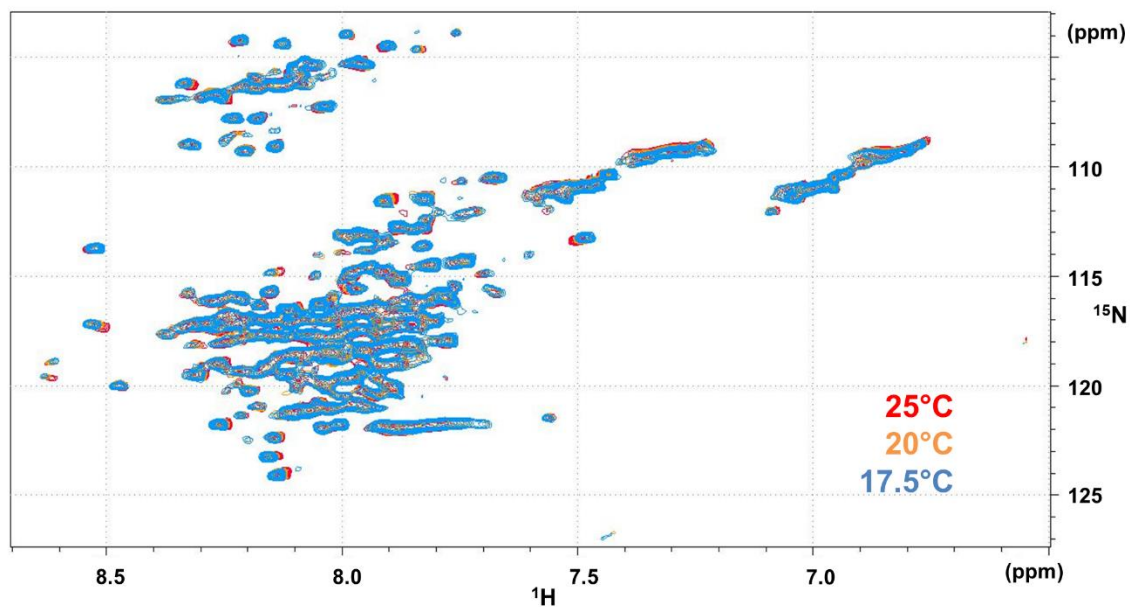


Figure A.3. 7 Overlay of ^1H - ^{15}N HSQC spectra at different acquisition temperatures.

A series of temperature controlled ^1H - ^{15}N HSQC spectra show that cross peaks move mostly uniformly in the unfolded protein, allowing for resonance assignments of a spectrum measured at 25 °C to be easily transferred to one measured at lower temperatures, being 19 °C for our experiments.

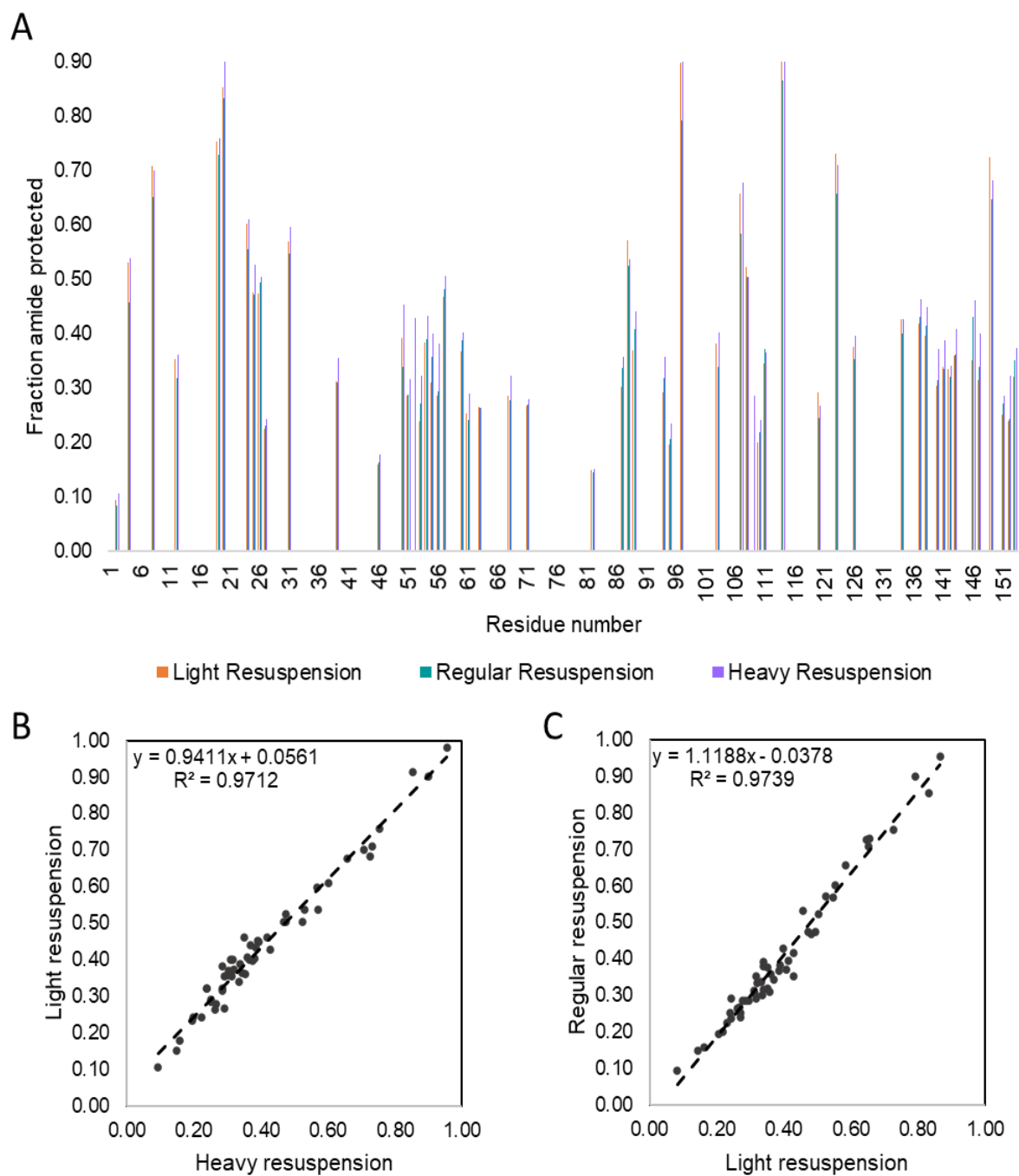


Figure A.3. 8 Impact of resuspension in D₂O on qHDX experiments

(A) Residue-specific comparison of fraction protected compared to how vigorously the sample was resuspended in D₂O. (B) Correlation of heavy resuspension to light resuspension (C) Correlation of light resuspension to regular resuspension.

Table A.2. 3 p values for all correlations given in thesis

Correlation	Figure #	R ² (Pearson)	n (number of samples)	p value	Notes
SOD1 and consurf	2.9	0.1684	101	0.0923	All variants
SOD1 and consurf	n/a	0.53	82	< .00001	H46 and V148 removed
Apo 25 °C agg propensity and normalized Chiti-Dobson score	Table 2.4	0.1044	16	0.7004	Average agg propensity score for each mutant used; G16S, G93D and L106V not included
Zn 25 °C agg propensity and normalized Chiti-Dobson score	Table 2.4	0.0692	16	0.7990	^
Apo 30 °C agg propensity and normalized Chiti-Dobson score	Table 2.4	0.0355	16	0.8961	^
Zn 30 °C agg propensity and normalized Chiti-Dobson score	Table 2.4	0.0966	16	0.7219	^
Apo 25 °C agg propensity and normalized Wang-Agar score	Table 2.4	0.0942	16	0.7286	^
Zn 25 °C agg propensity and normalized Wang-Agar score	Table 2.4	0.0585	16	0.830	^
Apo 30 °C agg propensity and normalized Wang-Agar score	Table 2.4	0.0549	16	0.8399	^
Zn 30 °C agg propensity and normalized Wang-Agar score	Table 2.4	0.1015	16	0.7083	^
Apo 25 °C agg propensity and normalized Zyggregator score	Table 2.4	0.0703	16	0.7959	^
Zn 25 °C agg propensity and normalized Zyggregator score	Table 2.4	0.0026	16	0.9924	^
Apo 30 °C agg propensity and normalized Zyggregator score	Table 2.4	0.0603	16	0.8244	^
Zn 30 °C agg propensity and normalized Zyggregator score	Table 2.4	0.0485	16	0.8584	^
Apo 25 °C agg propensity and normalized Ztox score	Table 2.4	0.0252	16	0.9262	^
Zn 25 °C agg propensity and normalized Ztox score	Table 2.4	0.0475	16	0.8613	^
Apo 30 °C agg propensity and normalized Ztox score	Table 2.4	0.0211	16	0.9382	^
Zn 30 °C agg propensity and normalized Ztox score	Table 2.4	0.1907	16	0.4793	^
Apo 25 °C agg propensity and normalized Fold Amyloid score	Table 2.4	0.3189	16	0.2286	^
Zn 25 °C agg propensity and	Table 2.4	0.2849	16	0.2848	^

normalized Fold Amyloid score					
Apo 30 °C agg propensity and normalized Fold Amyloid score	Table 2.4	0.0861	16	0.7512	^
Zn 30 °C agg propensity and normalized Fold Amyloid score	Table 2.4	0.173	16	0.5217	^
Apo 25 °C agg propensity and normalized PASTA score	Table 2.4	0.0404	16	0.8819	^
Zn 25 °C agg propensity and normalized PASTA score	Table 2.4	0.0285	16	0.9166	^
Apo 30 °C agg propensity and normalized PASTA score	Table 2.4	0.0105	16	0.9692	^
Zn 30 °C agg propensity and normalized PASTA score	Table 2.4	0.0341	16	0.9002	^
Apo 25 °C agg propensity and normalized TANGO score	Table 2.4	0.0011	16	0.9968	^
Zn 25 °C agg propensity and normalized TANGO score	Table 2.4	0.1898	16	0.4814	^
Apo 30 °C agg propensity and normalized TANGO score	Table 2.4	0.1099	16	0.6853	^
Zn 30 °C agg propensity and normalized TANGO score	Table 2.4	0.0602	16	0.8247	^
Apo 25 °C agg propensity and normalized Waltz score	Table 2.4	0.0775	16	0.7754	^
Zn 25 °C agg propensity and normalized Waltz score	Table 2.4	0.0817	16	0.7636	^
Apo 30 °C agg propensity and normalized Waltz score	Table 2.4	0.1404	16	0.6040	^
Zn 30 °C agg propensity and normalized Waltz score	Table 2.4	0.2326	16	0.3851	^
Apo 25 °C agg propensity and normalized Profile 3D score	Table 2.4	0.0016	16	0.9953	^
Zn 25 °C agg propensity and normalized Profile 3D score	Table 2.4	0.0019	16	0.9944	^
Apo 30 °C agg propensity and normalized Profile 3D score	Table 2.4	0.004	16	0.9883	^
Zn 30 °C agg propensity and normalized Profile 3D score	Table 2.4	0.0153	16	0.9551	^
Aggregation propensity at 25 °C and 30 °C and apoSH T _m	2.10 top	0.6933	102	< .00001	
Aggregation propensity at 25 °C and 30 °C and ΔG_{dimer} dissociation	2.10 bottom	0.4675	106	< .00001	
as apoSH T _m and ΔG_{dimer} dissociation to each other	n/a	0.56	9	0.1168	
Apo 25 °C aggregation propensity and LB aggregation propensity	2.12	0.0006	18	0.9981	Average agg propensity score for each

					mutant used
Zn 25 °C aggregation propensity and LB aggregation propensity	2.12	0.2055	18	0.4133	^
Apo 30 °C aggregation propensity and LB aggregation propensity	2.12	0.0955	18	0.7062	^
Zn 30 °C aggregation propensity and LB aggregation propensity	2.12	0.0781	18	0.7580	^
Apo 37 °C aggregation propensity and LB aggregation propensity	2.12	0.14	18	0.5795	^
Zn 37 °C aggregation propensity and LB aggregation propensity	2.12	0.3769	18	0.1231	^
Mutation distance to zinc atom vs extent of rescue at 25 °C	2.13	0.3225	18	0.1918	^
Mutation distance to zinc atom vs extent of rescue at 25 °C and 30 °C	N/A	0.2584	36	0.1281	
Mutation distance to copper atom vs extent of rescue at 25 °C and 30 °C	N/A	0.137	36	0.4255	
Average mutation distance to both metal ions vs extent of rescue at 25 °C and 30 °C	N/A	0.2088	36	0.2217	
Mass SOD1 std to band density	2.16	0.9978	10	< .00001	
pD of qHDX samples to chemical shift of DCA peak	3.8	0.902	21	< .00001	
Comparison of replicate qHDX samples – protonated G85R	3.13	0.9901	70	< .00001	
Comparison of replicate qHDX samples – exchanged G41D	3.13	0.991	71	< .00001	
apoSH T _m to C _{mid}	4.10	0.8118	5	0.0952	
apoSS T _m to C _{mid}	4.10	0.4035	8	0.3215	
Holo SOD1 T _m to C _{mid}	4.10	1.00E-05	8	0.9998	
ΔG _{dimer dissociation} to C _{mid}	4.10	0.9292	5	0.0224	
ΔG _{monomer unfolding} to C _{mid}	4.10	0.6404	5	0.2444	
Average chemical shift of Cα SOD1 vs Adn	5.9	0.8176	15	0.0002	
Average chemical shift of Cβ SOD1 vs Adn	5.9	0.9277	15	< .00001	
Average chemical shift of ¹ H SOD1 vs Adn	5.9	0.4999	15	0.0577	
Average chemical shift of ¹⁵ N SOD1 vs Adn	5.9	0.9145	15	< .00001	

Backbone ¹ H chemical shifts of apoMb unfolded in DMSO vs acid denatured (low pH)	5.1	0.6919	38	< .00001
Backbone ¹⁵ N chemical shifts of apoMb unfolded in DMSO vs acid denatured (low pH)	5.1	0.9431	38	< .00001
Analysis of sample storage: protection of sample stored at -80°C vs immediate analysis	A.3.3	0.867	70	< .00001
Analysis of sample storage: protection of sample stored at 25°C vs immediate analysis	A.3.3	0.254	70	0.0339
Analysis of technical reproducibility: two immediate analysis replicates	A.3.3	0.974	70	< .00001
Impact of vigor of resuspension on amide protection: heavy vs light resuspension	A.3.6	0.9712	70	< .00001
Impact of vigor of resuspension on amide protection: light vs regular resuspension	A.3.6	0.9739	70	< .00001
C=O SOD1 vs RC	Table 3.3	0.721	17	0.0011
Average chemical shift of Cα: SOD1 vs RC	Table 3.3	0.99	18	< .00001
Average chemical shift of Cβ: SOD1 vs RC	Table 3.3	0.996	17	< .00001
Average chemical shift of ¹⁵ N: SOD1 vs RC	Table 3.3	0.934	17	< .00001
Average chemical shift of ¹ H: SOD1 vs RC	Table 3.3	0.543	17	0.0243

p values were calculated from Pearson (R) value of each correlation with an online calculator^[332] that uses the following equation: $t = \frac{R\sqrt{n-2}}{\sqrt{1-R^2}}$ where the p value is 2 x P(T > t) where T follows a distribution with n-2 degrees of freedom, n is the number of observations (pairs) and R is the correlation coefficient. A p value was calculated in every instance that a Pearson correlation was given, however, note that the *lack* of correlation was frequently discussed and in this situation a p value does not provide any new information and is just included for posterity. Values that are significant to a 90% confidence interval are colored in blue, and those that are not significant are colored in red. A 90% cutoff was selected because some of the datasets were limited in datapoints (eg apoSH T_m); in these cases caveats to whether or not the correlation was significant was discussed in-text. “^” in the notes section implies an identical note to the preceding sample.

Table A.3. 1 Primary sequence effects on ^{15}N chemical shifts of amino acid

	Preceded by A			Preceded by I			Preceded by V		
	Residue	deviation from average (ppm)	stdev for that res type (ppm)	Residue	deviation from average (ppm)	stdev for that res type (ppm)	Residue	deviation from average (ppm)	stdev for that res type (ppm)
FAS1-4	G501	-0.07	1.55	Q523	-2.24	1.32	M506		
	A521	1.39	1.115	L569	-1.34	1.46	L509	-1.68	1.46
	I522	1.83	1.75	G575	-3.12	1.55	A520	-1.48	1.12
	G526	2.24	1.55	L578	-3.25	1.46	F540	-0.79	1.75
	L550	1.27	1.46	G584	-2.77	1.55	S580	-1.41	1.45
	K563	0.94	0.52	T629	-2.47	1.35	R588	-2.17	1.12
	N567	1.11	1.37	F564	-1.78	1.75	S600	-1.73	1.45
	L586	1.64	1.46				V606	-1.80	1.35
	E615	0.44	0.65				S607	-1.85	1.45
	T621	1.88	1.35				N609	-2.42	1.37
	N637	2.56	1.38				A614	-1.78	1.12
	D648	0.82	0.68				V625	-3.45	1.35
	L651	2.79	1.46				H626	-1.61	2.27
							I628	-3.63	1.75
							L632	-0.61	1.46
Cyt C	A51	2.00	1.20	F10	-1.25	2.05	E4	-1.63	1.94
	N52	1.47	1.37	I58	-2.67	1.33	E21	-1.21	1.94
	D93	1.31	1.18	F82	-1.12	2.05	G84	-2.08	1.31
	Y97			K86	-3.17	1.40			
	T102	1.80	0.98	A96	-1.06	1.20			
GroES	G23	2.07	1.48	V12	-2.60	2.41	I11	-2.82	2.45
	A33	2.18	1.37	V26	-2.46	2.41	L27	-0.27	0.37
	K34	1.44	2.04	V65	-1.82	2.41	L41	0.27	0.37
	V43	3.41	2.41	D79	-0.65	1.63	G44	-1.83	1.48
	I94	1.51	2.45	V95	-0.19	2.41	G62	-1.82	1.48
							D84	-2.45	1.63
SOD1	T2	-1.21	1.22	I18	-1.35	1.48	V5	-2.73	1.39
	G56	1.92	1.50	N19	-1.65	1.37	C6	-3.06	2.22
	G61	2.25	1.50	K36	-2.13	1.39	L8	-2.67	1.60
	D90	0.69	0.50	E100	-1.34	0.93	Q15	-1.78	2.52
	D96	-0.20	0.50	S105	-2.46	1.47	H48	-2.68	1.80
	D124	0.81	0.50	I113	-1.87	1.48	G82	-2.73	1.50
	G141	1.95	1.50	G114	-3.02	1.50	T88	-1.43	1.22
	C146	2.13	2.22	G150	-2.83	1.50	A95	-0.92	1.00
	Q153	2.01	2.52	A152	-1.69	1.00	S98	-2.45	1.47
	I104	0.05	1.48				H120	-2.34	1.80
							I149	-0.80	1.48

Clear primary sequence effects are observed on the chemical shift of amide cross peaks. Displayed are the effects on chemical shifts of residues preceded by A, I, and V. All chemical shift deviations are measured in DMSO, and we have measured sequence effects on SOD1, Cyt C (BMRB: 26973) ^[333], Fas1-4 (BMRB: 25425) ^[334], and GroES (BMRB: 18949) ^[335]. Depending on the preceding amino acid, the ^{15}N chemical shift may move upfield or downfield relative to the average value for each amino acid type. The average value was determined per nucleus for each amino acid type.

Table A.4. 1 Full chemical shift assignments of SOD1 in 95% DMSO/5% H₂O pH 5.4

Residue		Nucleus	Chemical shift (ppm)
2	THR	C	169.931
2	THR	CA	58.964
2	THR	CB	66.977
2	THR	H	8.456
2	THR	N	113.582
3	LYS	C	171.935
3	LYS	CA	52.775
3	LYS	CB	31.554
3	LYS	H	8.019
3	LYS	N	120.049
4	ALA	C	171.599
4	ALA	CA	58.458
4	ALA	CB	30.56
4	ALA	H	7.969
4	ALA	N	117.258
5	VAL	C	171.553
5	VAL	CA	58.339
5	VAL	H	7.969
5	VAL	N	118.85
6	ALA	C	170.173
6	ALA	CA	55.293
6	ALA	CB	26.272
6	ALA	H	8.124
6	ALA	N	119.741
7	VAL	C	171.366
7	VAL	CA	58.428

7	VAL	CB	30.674
7	VAL	H	7.951
7	VAL	N	117.813
8	LEU	C	172.569
8	LEU	CA	51.378
8	LEU	CB	40.565
8	LEU	H	8.016
8	LEU	N	122.063
9	LYS	C	172.363
9	LYS	CA	52.808
9	LYS	CB	31.305
9	LYS	H	7.974
9	LYS	N	118.641
10	GLY	C	169.381
10	GLY	CA	42.291
10	GLY	H	8.162
10	GLY	N	106.526
11	ASP	C	171.338
11	ASP	H	8.214
11	ASP	N	116.688
12	GLY	C	169.391
12	GLY	CA	41.917
12	GLY	H	7.965
12	GLY	N	105.526
14	VAL	C	171.776
14	VAL	CA	58.836
14	VAL	CB	30.311
14	VAL	H	7.991
14	VAL	N	115.677

15	GLN	C	172.224
15	GLN	CA	52.864
15	GLN	CB	27.938
15	GLN	H	8.053
15	GLN	N	120.037
16	GLY	CA	42.305
16	GLY	H	8.14
16	GLY	N	106.426
17	ILE	C	171.879
17	ILE	CA	57.717
17	ILE	CB	36.712
17	ILE	H	7.838
17	ILE	N	116.662
18	ILE	C	171.562
18	ILE	CA	58.401
18	ILE	CB	36.526
18	ILE	H	7.921
18	ILE	N	119.977
19	ASN	C	171.665
19	ASN	CA	50.212
19	ASN	CB	37.228
19	ASN	H	8.046
19	ASN	N	119.924
20	PHE	C	171.171
20	PHE	CA	55.214
20	PHE	CB	37.147
20	PHE	H	8.099
20	PHE	N	117.846
24	GLU	CA	52.457

24	GLU	CB	27.311
24	GLU	H	8.034
24	GLU	N	117.775
25	SER	C	170.351
25	SER	CA	55.293
25	SER	CB	61.943
25	SER	H	7.899
25	SER	N	113.509
26	ASN	C	171.618
26	ASN	CA	50.135
26	ASN	CB	37.333
26	ASN	H	8.234
26	ASN	N	118.834
27	GLY	C	176.38
27	GLY	CA	41.932
27	GLY	H	7.909
27	GLY	N	104.944
31	VAL	C	171.562
31	VAL	CA	58.09
31	VAL	CB	30.995
31	VAL	H	7.769
31	VAL	N	116.087
32	TRP	C	172.471
32	TRP	CA	53.864
32	TRP	CB	27.908
32	TRP	H	8.09
32	TRP	N	121.339
33	GLY	C	169.428
33	GLY	CA	42.429

33	GLY	H	8.232
33	GLY	N	107.032
34	SER	C	171.031
34	SER	CA	55.149
34	SER	CB	61.999
34	SER	H	7.942
34	SER	N	113.039
35	ILE	C	171.944
35	ILE	CA	57.896
35	ILE	CB	36.313
35	ILE	H	8.033
35	ILE	N	118.614
36	LYS	CA	53.88
36	LYS	CB	30.932
36	LYS	H	8.086
36	LYS	N	121.288
37	GLY	C	169.456
37	GLY	CA	42.367
37	GLY	H	8.1
37	GLY	N	106.308
38	LEU	C	172.988
38	LEU	CA	51.875
38	LEU	CB	40.751
38	LEU	H	7.929
38	LEU	N	118.557
39	THR	C	170.854
39	THR	CA	58.836
39	THR	CB	66.852
39	THR	H	7.868

39	THR	N	111.438
40	GLU	C	172.196
40	GLU	CA	52.808
40	GLU	CB	27.39
40	GLU	H	7.974
40	GLU	N	118.887
41	GLY	C	169.577
41	GLY	CA	42.305
41	GLY	H	8.17
41	GLY	N	106.535
42	LEU	C	172.904
42	LEU	CA	51.938
42	LEU	CB	40.441
42	LEU	H	8
42	LEU	N	118.614
43	HIS	CA	52.186
43	HIS	CB	27.328
43	HIS	H	8.245
43	HIS	N	116.155
44	GLY	CA	42.305
44	GLY	H	8.16
44	GLY	N	106.306
46	HIS	C	170.63
46	HIS	CA	52.124
46	HIS	CB	27.39
46	HIS	H	8.439
46	HIS	N	117.447
47	VAL	C	171.655
47	VAL	CA	58.572

47	VAL	CB	30.569
47	VAL	H	7.933
47	VAL	N	116.649
48	HIS	C	170.49
48	HIS	CA	52.124
48	HIS	CB	27.266
48	HIS	H	8.441
48	HIS	N	120.023
49	GLU	C	171.59
49	GLU	CA	52.471
49	GLU	CB	27.396
49	GLU	H	8.065
49	GLU	N	118.63
50	PHE	C	171.935
50	PHE	CA	54.486
50	PHE	CB	37.582
50	PHE	H	8.198
50	PHE	N	117.719
51	GLY	C	169.465
51	GLY	CA	42.305
51	GLY	H	8.283
51	GLY	N	106.444
52	ASP	C	171.329
52	ASP	CA	49.762
52	ASP	CB	36.463
52	ASP	H	8.213
52	ASP	N	116.759
53	ASN	C	171.655
53	ASN	CA	50.073

53	ASN	H	8.251
53	ASN	N	117.245
54	THR	C	170.695
54	THR	CA	58.898
54	THR	CB	66.604
54	THR	H	7.687
54	THR	N	110.695
55	ALA	C	173.482
55	ALA	CA	49.327
55	ALA	CB	17.758
55	ALA	H	8.056
55	ALA	N	122.232
56	GLY	C	169.685
56	GLY	CA	42.429
56	GLY	H	8.128
56	GLY	N	104.54
57	CYS	C	170.649
57	CYS	CA	55.604
57	CYS	CB	26.396
57	CYS	H	7.972
57	CYS	N	115.759
58	THR	C	170.49
58	THR	CA	58.711
58	THR	CB	66.915
58	THR	H	7.968
58	THR	N	113.076
59	SER	C	170.36
59	SER	CA	55.48
59	SER	CB	61.819

59	SER	H	7.93
59	SER	N	114.91
60	ALA	C	173.146
60	ALA	CA	48.892
60	ALA	CB	18.131
60	ALA	H	8.127
60	ALA	N	122.493
61	GLY	C	167.769
61	GLY	CA	41.748
61	GLY	H	7.972
61	GLY	N	104.21
63	HIS	C	170.173
63	HIS	CA	52.124
63	HIS	CB	27.142
63	HIS	H	8.318
63	HIS	N	115.836
64	PHE	CA	54.113
64	PHE	CB	37.768
64	PHE	H	7.986
64	PHE	N	117.201
65	ASN	C	170.863
65	ASN	CA	47.898
65	ASN	CB	37.147
65	ASN	H	8.553
65	ASN	N	119.681
68	SER	C	171.012
68	SER	CA	55.728
68	SER	CB	61.632
68	SER	H	7.575

68	SER	N	112.254
69	ARG	CA	52.808
69	ARG	CB	28.695
69	ARG	H	7.959
69	ARG	N	119.72
71	HIS	C	171.003
71	HIS	CA	52.248
71	HIS	CB	27.514
71	HIS	H	8.151
71	HIS	N	116.364
72	GLY	C	169.866
72	GLY	CA	42.305
72	GLY	H	8.329
72	GLY	N	106.95
73	GLY	C	167.853
73	GLY	CA	41.725
73	GLY	H	8.197
73	GLY	N	106.187
82	GLY	C	169.484
82	GLY	CA	42.23
82	GLY	H	8.348
82	GLY	N	109.185
83	ASP	C	171.357
83	ASP	CA	49.877
83	ASP	CB	36.313
83	ASP	H	8.205
83	ASP	N	117.011
84	LEU	C	173.146
84	LEU	CA	52.062

84	LEU	CB	40.565
84	LEU	H	8.029
84	LEU	N	118.818
85	GLY	H	8.105
85	GLY	N	105.476
86	ASN	CA	50.111
86	ASN	CB	37.137
86	ASN	H	8.083
86	ASN	N	116.949
87	VAL	C	171.944
87	VAL	CA	58.898
87	VAL	CB	30.311
87	VAL	H	7.888
87	VAL	N	115.608
88	THR	C	170.556
88	THR	CA	60.451
88	THR	CB	66.728
88	THR	H	7.825
88	THR	N	113.803
89	ALA	C	173.193
89	ALA	CA	49.017
89	ALA	CB	17.882
89	ALA	H	7.938
89	ALA	N	122.14
90	ASP	CA	50.135
90	ASP	CB	35.966
90	ASP	H	8.137
90	ASP	N	116.045
91	LYS	C	172.075

91	LYS	CA	52.994
91	LYS	CB	30.87
91	LYS	H	7.84
91	LYS	N	117.456
92	ASP	C	171.503
92	ASP	CA	50.179
92	ASP	CB	36.135
92	ASP	H	8.167
92	ASP	N	116.654
93	GLY	C	169.391
93	GLY	CA	42.59
93	GLY	H	7.973
93	GLY	N	105.583
94	VAL	C	171.301
94	VAL	CA	58.214
94	VAL	CB	30.684
94	VAL	H	7.714
94	VAL	N	114.719
95	ALA	C	173.062
95	ALA	CA	48.582
95	ALA	CB	18.255
95	ALA	H	8.142
95	ALA	N	123.418
96	ASP	C	171.031
96	ASP	CA	49.935
96	ASP	CB	36.401
96	ASP	H	8.255
96	ASP	N	116.926
97	VAL	C	171.469

97	VAL	CA	57.951
97	VAL	CB	31.1
97	VAL	H	7.545
97	VAL	N	113.981
98	SER	C	171.012
98	SER	CA	55.604
98	SER	CB	61.694
98	SER	H	8.08
98	SER	N	116.383
99	ILE	C	171.888
99	ILE	CA	57.99
99	ILE	CB	36.461
99	ILE	H	7.879
99	ILE	N	118.47
100	GLU	CA	52.745
100	GLU	CB	27.079
100	GLU	H	8.006
100	GLU	N	119.834
101	ASP	C	171.171
101	ASP	CA	50.193
101	ASP	CB	36.087
101	ASP	H	8.122
101	ASP	N	117.551
102	SER	C	170.733
102	SER	CA	55.604
102	SER	CB	61.757
102	SER	H	7.802
102	SER	N	112.945
103	VAL	C	171.758

103	VAL	CA	58.678
103	VAL	CB	30.373
103	VAL	H	7.805
103	VAL	N	116.621
104	ILE	C	171.785
104	ILE	CA	57.779
104	ILE	CB	36.526
104	ILE	H	7.847
104	ILE	N	118.578
105	SER	C	170.872
105	SER	CA	55.497
105	SER	CB	61.77
105	SER	H	7.932
105	SER	N	116.388
106	LEU	CA	51.852
106	LEU	CB	40.316
106	LEU	H	8.074
106	LEU	N	121.326
107	SER	H	7.924
107	SER	N	113.075
108	GLY	C	169.754
108	GLY	CA	42.616
108	GLY	H	8.176
108	GLY	N	107.924
109	ASP	C	171.171
109	ASP	CA	50.197
109	ASP	H	8.158
109	ASP	N	117.022
110	HIS	C	170.369

110	HIS	CA	52.248
110	HIS	CB	27.204
110	HIS	H	8.132
110	HIS	N	115.912
111	SER	C	170.285
111	SER	CA	55.603
111	SER	CB	26.272
111	SER	H	8.096
111	SER	N	116.677
112	ILE	CA	57.841
112	ILE	CB	36.339
112	ILE	H	8.153
112	ILE	N	119.446
113	ILE	C	171.991
113	ILE	CA	57.679
113	ILE	CB	36.563
113	ILE	H	7.906
113	ILE	N	120.499
114	GLY	C	169.344
114	GLY	CA	42.338
114	GLY	H	8.219
114	GLY	N	109.476
115	ARG	CA	52.482
115	ARG	CB	29.259
115	ARG	H	8.022
115	ARG	N	117.869
119	VAL	CA	58.836
119	VAL	CB	30.249
119	VAL	H	8.007

119	VAL	N	115.458
120	HIS	C	170.705
120	HIS	CA	52.124
120	HIS	CB	27.328
120	HIS	H	8.266
120	HIS	N	119.686
121	GLU	C	171.814
121	GLU	CA	52.58
121	GLU	H	8.106
121	GLU	N	118.661
122	LYS	C	171.925
122	LYS	CA	52.994
122	LYS	CB	31.119
122	LYS	H	8.194
122	LYS	N	119.434
126	LEU	C	173.081
126	LEU	CA	51.875
126	LEU	CB	40.565
126	LEU	H	7.78
126	LEU	N	118.112
127	GLY	C	169.614
127	GLY	CA	42.367
127	GLY	H	8.117
127	GLY	N	105.802
128	LYS	C	172.662
128	LYS	CA	52.745
128	LYS	CB	31.368
128	LYS	H	7.966
128	LYS	N	117.531

129	GLY	C	169.987
129	GLY	CA	42.367
129	GLY	H	8.268
129	GLY	N	106.725
133	GLU	C	172.019
133	GLU	CA	52.797
133	GLU	CB	27.037
133	GLU	H	8.067
133	GLU	N	117.148
134	SER	C	170.966
134	SER	CA	55.418
134	SER	CB	61.757
134	SER	H	7.949
134	SER	N	113.843
135	THR	C	170.733
135	THR	CA	58.846
135	THR	CB	66.753
135	THR	H	7.807
135	THR	N	112.704
136	LYS	CA	52.808
136	LYS	CB	31.554
136	LYS	H	8.011
136	LYS	N	119.726
137	THR	C	171.18
137	THR	CA	58.96
137	THR	CB	66.915
137	THR	H	7.771
137	THR	N	111.318
138	GLY	C	169.475

138	GLY	CA	42.491
138	GLY	H	8.08
138	GLY	N	107.567
139	ASN	C	171.506
139	ASN	CA	50.135
139	ASN	CB	37.333
139	ASN	H	8.121
139	ASN	N	117.008
140	ALA	C	173.519
140	ALA	CA	49.452
140	ALA	CB	17.696
140	ALA	H	8.24
140	ALA	N	121.802
141	GLY	C	169.698
141	GLY	CA	42.491
141	GLY	H	8.223
141	GLY	N	104.51
142	SER	C	171.059
142	SER	CA	55.48
142	SER	CB	61.943
142	SER	H	7.868
142	SER	N	112.942
143	ARG	C	172.009
143	ARG	CA	53.118
143	ARG	CB	28.633
143	ARG	H	8.196
143	ARG	N	120.303
144	LEU	C	172.662
144	LEU	CA	51.658

144	LEU	CB	40.562
144	LEU	H	7.897
144	LEU	N	118.293
145	ALA	C	172.913
145	ALA	CA	48.83
145	ALA	CB	17.696
145	ALA	H	8.011
145	ALA	N	121.22
146	CYS	C	170.537
146	CYS	CA	55.48
146	CYS	CB	26.398
146	CYS	H	7.927
146	CYS	N	114.554
147	GLY	C	169.214
147	GLY	CA	42.429
147	GLY	H	8.217
147	GLY	N	107.808
148	VAL	C	171.497
148	VAL	CA	58.214
148	VAL	CB	30.684
148	VAL	H	7.789
148	VAL	N	114.773
149	ILE	C	171.916

149	ILE	CA	57.756
149	ILE	CB	36.664
149	ILE	H	7.905
149	ILE	N	119.43
150	GLY	C	169.251
150	GLY	CA	42.291
150	GLY	H	8.156
150	GLY	N	109.287
151	ILE	C	171.236
151	ILE	CA	57.344
151	ILE	H	7.747
151	ILE	N	115.978
152	ALA	C	172.774
152	ALA	CA	48.563
152	ALA	CB	18.006
152	ALA	H	8.135
152	ALA	N	124.191
153	GLN	C	173.659
153	GLN	CA	51.952
153	GLN	CB	27.289
153	GLN	H	8.046
153	GLN	N	116.48

Resonance assignments for A4V inclusion body (IB) sample grown at 37°C and harvested 6 hours after induction. Experiments for making assignments were acquired at 25 °C for IB sample at pH 5.5 in 95% DMSO/5% H₂O/25mM DTT. CO chemical shifts were obtained from HNCO/HNCαCO experiments, and Cα and Cβ shifts were obtained from HNCαCβ/CβCα(CO)NH experiments.

Table A.4. 2 Comparison of qHDX protection to native structural features

SS	Res #	AA	Ave qHDX ^a	V148G qHDX	Holo SOD1 H bonding [306]	raSOD1 D1 dimer [225,277]	raSOD1 D2 dimer [225,277]	raSOD1 D3 dimer [225,277]
N	1	ALA	---	---			●	●
N	2	THR	---	---			●	
β1	3	LYS	---	---				
β1	4	ALA	0.55 +/- 0.09	0.52 +/- 0.12	●			
β1	5	VAL	0.52 +/- 0.1	0.44 +/- 0.09	●	●	●	●
β1	6	CYS	---	---	●			
β1	7	VAL	---	---	●	●	●	●
β1	8	LEU	0.70 +/- 0.24	0.74 +/- 0.20	●			
L1	9	LYS	---	---		●	●	●
L1	10	GLY	---	---				●
L1	11	ASP	---	---			●	●
L1	12	GLY	---	---				●
L1	13	PRO	---	---				●
β2	14	VAL	---	---	●			●
β2	15	GLN	---	---	●			
β2	16	GLY	---	---				
β2	17	ILE	---	---	●	●	●	
β2	18	ILE	---	---	●			
β2	19	ASN	0.59 +/- 0.18	0.69 +/- 0.21	●		●	
β2	20	PHE	0.75 +/- 0.24	0.88 +/- 0.17	●			
β2	21	GLU	---	---	●			

β2	22	GLN	---	---		
L2	23	LYS	---	---		
L2	24	GLU	---	---		
L2	25	SER	0.51 +/- 0.09	0.48 +/- 0.09		
L2	26	ASN	0.49 +/- 0.12	0.50 +/- 0.10		
L2	27	GLY	0.36 +/- 0.1	0.26 +/- 0.06		
L2	28	PRO	---	---		
β3	29	VAL	---	---	●	
β3	30	LYS	---	---	●	
β3	31	VAL	0.75 +/- 0.13	0.61 +/- 0.11	●	
β3	32	TRP	---	---	●	
β3	33	GLY	---	---	●	
β3	34	SER	---	---	●	
β3	35	ILE	---	---	●	
β3	36	LYS	---	---		
L3	37	GLY	---	---	●	●
L3	38	LEU	---	---	●	●
L3	39	THR	0.40 +/- 0.09	0.33 +/- 0.06		
L3	40	GLU	---	---	●	
β4	41	GLY	---	---		
β4	42	LEU	---	---		
β4	43	HIS	---	---	●	
β4	44	GLY	---	---	●	
β4	45	PHE	---	---	●	
β4	46	HIS	---	---	●	

β4	47	VAL	---	---	●				
β4	48	HIS	---	---	●				●
β4	49	GLU	---	---	●	●	●	●	●
L4	50	PHE	0.49 +/- 0.2	---	●	●	●	●	●
L4	51	GLY	0.40 +/- 0.11	0.31 +/- 0.03	●	●	●	●	●
L4	52	ASP	0.46 +/- 0.13	0.38 +/- 0.08	●	●	●	●	●
L4	53	ASN	---			●	●	●	●
L4	54	THR	0.47 +/- 0.11	0.38 +/- 0.04	●	●	●	●	●
L4	55	ALA	0.40 +/- 0.13	0.28 +/- 0.06					●
L4	56	GLY	0.40 +/- 0.11	0.28 +/- 0.02					●
L4	57	CYS	0.42 +/- 0.11	0.46 +/- 0.06					●
L4	58	THR	---	---					●
L4	59	SER	---	---					●
L4	60	ALA	0.40 +/- 0.11	0.35 +/- 0.07	●				●
L4	61	GLY	0.37 +/- 0.1	0.29 +/- 0.03					●
L4	62	PRO	---	---					●
L4	63	HIS	---	---					●
L4	64	PHE	---	---					
L4	65	ASN	---	---					
L4	66	PRO	---	---					
L4	67	LEU	---	---					
L4	68	SER	0.42 +/- 0.15	0.27 +/- 0.06					
L4	69	ARG	0.58 +/- 0.34	---	●				
L4	70	LYS	---	---					
L4	71	HIS	0.30 +/- 0.11	0.27 +/- 0.06	●				

L4	72	GLY	---	---	●
L4	73	GLY	---	---	●
L4	74	PRO	---	---	
L4	75	LYS	---	---	
L4	76	ASP	---	---	●
L4	77	GLU	---	---	
L4	78	GLU	---	---	
L4	79	ARG	---	---	●
L4	80	HIS	---	---	●
L4	81	VAL	---	---	
L4	82	GLY	0.22 +/- 0.06	0.26 +/- 0.03	●
β5	83	ASP	---	---	●
β5	84	LEU	---	---	●
β5	85	GLY	---	---	
β5	86	ASN	---	---	●
β5	87	VAL	0.45 +/- 0.2	0.42 +/- 0.14	●
β5	88	THR	0.65 +/- 0.23	0.77 +/- 0.25	
β5	89	ALA	0.51 +/- 0.14	0.38 +/- 0.06	●
L5	90	ASP	---	---	●
L5	91	LYS	---	---	
L5	92	ASP	---	---	
L5	93	GLY	---	---	●
β6	94	VAL	0.44 +/- 0.12	0.37 +/- 0.08	●
β6	95	ALA	0.36 +/- 0.12	0.27 +/- 0.04	●
β6	96	ASP	0.41 +/- 0.19	0.57 +/- 0.22	

β6	97	VAL	0.73 +/- 0.16	1.16 +/- 0.20	●				
β6	98	SER	---	---					
β6	99	ILE	---	---	●				
β6	100	GLU	0.63 +/- 0.22	0.77 +/- 0.09					
β6	101	ASP	---	---	●				
L6	102	SER	0.77 +/- 0.48	0.89 +/- 0.58					
L6	103	VAL	0.65 +/- 0.28	0.47 +/- 0.12					
L6	104	ILE	---	---	●				
L6	105	SER	---	---					
L6	106	LEU	---	---					
L6	107	SER	0.63 +/- 0.24	0.81 +/- 0.29					
L6	108	GLY	0.41 +/- 0.21	0.47 +/- 0.22				●	
L6	109	ASP	0.33 +/- 0.1	0.32 +/- 0.06				●	
L6	110	HIS	0.31 +/- 0.08	0.50 +/- 0.11					
L6	111	CYS	0.42 +/- 0.14	0.70 +/- 0.21				●	●
L6	112	ILE	---	---	●				
L6	113	ILE	0.83 +/- 0.24	1.18 +/- 0.25	●	●	●	●	
L6	114	GLY	0.71 +/- 0.22	0.94 +/- 0.32	●	●	●	●	
β7	115	ARG	0.54 +/- 0.23	0.54 +/- 0.21	●	●	●	●	
β7	116	THR	---	---	●				
β7	117	LEU	---	---	●				
β7	118	VAL	---	---	●				
β7	119	VAL	---	---	●				
β7	120	HIS	0.29 +/- 0.07	0.33 +/- 0.04	●				●
L7	121	GLU	0.44 +/- 0.04	0.45 +/- 0.01	●				●

L7	122	LYS	---	---	●	●
L7	123	ALA	0.60 +/- 0.14	0.73 +/- 0.11		
L7	124	ASP	0.40 +/- 0.2	0.45 +/- 0.20	●	
L7	125	ASP	---	---	●	
L7	126	LEU	0.49 +/- 0.11	0.36 +/- 0.06	●	
L7	127	GLY	---	---	●	
L7	128	LYS	---	---	●	
L7	129	GLY	---	---		
L7	130	GLY	---	---		
L7	131	ASN	---	---		●
L7	132	GLU	---	---		●
L7	133	GLU	---	---		●
L7	134	SER	0.53 +/- 0.18	0.29 +/- 0.11		
L7	135	THR	---	---		
L7	136	LYS	---	---		●
L7	137	THR	0.49 +/- 0.12	0.40 +/- 0.08		●
L7	138	GLY	0.45 +/- 0.12	0.32 +/- 0.04	●	
L7	139	ASN	---	---	●	●
L7	140	ALA	0.43 +/- 0.12	0.33 +/- 0.04		●
L7	141	GLY	0.43 +/- 0.12	0.29 +/- 0.05		●
L7	142	SER	0.37 +/- 0.1	0.24 +/- 0.03		●
β8	143	ARG	0.41 +/- 0.12	0.26 +/- 0.04		●
β8	144	LEU	---	---		●
β8	145	ALA	---	---	●	
β8	146	CYS	0.58 +/- 0.15	0.45 +/- 0.09	●	● ●

β8	147	GLY	0.58 +/- 0.13	0.26 +/- 0.03	●			
β8	148	VAL	0.54 +/- 0.15	---	●	●	●	●
β8	149	ILE	0.67 +/- 0.22	0.64 +/- 0.17	●	●		●
β8	150	GLY	0.53 +/- 0.13	0.24 +/- 0.03	●	●	●	●
β8	151	ILE	0.63 +/- 0.15	0.37 +/- 0.28	●	●	●	●
C	152	ALA	0.47 +/- 0.1	0.22 +/- 0.02	●	●	●	●
C	153	GLN	0.46 +/- 0.13	0.27 +/- 0.04	●	●	●	●

SS – Secondary structure, Res# - residue number, AA– amino acid. Table information corresponds to that in Figure 4.2. Colors on table columns correspond with colors for secondary structure elements in Figures 4.1 and 4.2, with loop 4 in yellow, loop 6 in orange, loop 7 in red, N- and C-termini in grey, and all other tructures in teal. ^aAve qHDX – average protection per residue across all mutants/replicates (shown in Figure 3), excluding V148G. Error estimates are the average of propagated errors for each mutant (see Materials and Methods Eq. 4.1 and Eq. 4.2). V148G qHDX – protection observed per residue across the two V148G replicates. Error estimates are the propagated errors for V148G (Eq. 4.1). H-bonding measured by traditional NMR native amide HDX ^[306] and participation in D1, D2 or D3 interfaces measured by solution NMR^[225,277] denoted by ●.

Table A.4. 3 k_{ex} of SOD1 residues fit to both a single or double exponential decay

	y ₀	y ₀ error	A ₁	A ₁ error	t ₁	t ₁ error	A ₂	A ₂ error	t ₂	t ₂ error
2ThrH	2.05E+09	1.03E+10	1.54E+10	2.30E+10	0.01192	8.37E+68				
2ThrH	2.83E+08	7.72E+09	6.56E+09	9.14E+10	0.04777	2.30E+16	6.45E+09	8.77E+10	79.0643	1453.352
4ValH	5.15E+09	8.79E+08	1.39E+10	1.97E+09	1.38808	0.56649				
4ValH	3.66E+09	1.21E+10	7.33E+09	3.62E+10	0.01294	5.57E+63	5.31E+09	3.33E+10	261.2689	3601.522
5ValH	4.09E+09	2.65E+09	7.30E+09	3.87E+09	82.95246	146.9094				
5ValH	3.64E+09	1.91E+09	6.38E+09	4.95E+09	0.01956	3.79E+40	4.27E+09	4.38E+09	394.9296	1180.861
8LeuH	4.33E+09	1.40E+09	6.05E+09	1.92E+09	169.9642	163.27834				
8LeuH	4.03E+09	6.44E+08	5.14E+09	9.96E+08	1.27073	0.66398	3.59E+09	7.89E+08	706.1808	704.6282
19AsnH	2.71E+09	7.90E+08	3.50E+09	1.08E+09	194.4874	179.26479				
19AsnH	2.62E+09	1.17E+08	2.63E+09	2.64E+08	0.54397	0.57027	2.38E+09	2.22E+08	493.2668	164.7017
20PheH	4.45E+09	8.00E+08	4.43E+09	1.09E+09	192.3072	142.06526				
20PheH	4.40E+09	1.65E+08	2.64E+09	4.42E+08	0.06974	1.15E+09	3.26E+09	3.97E+08	362.3356	119.5134
25SerH	2.04E+09	1.16E+09	2.40E+09	1.65E+09	94.80851	217.5845				
25SerH	1.97E+09	1.23E+09	1.93E+09	3.42E+09	0.04725	4.71E+15	1.38E+09	3.11E+09	326.7352	1842.592
26AsnH	2.20E+09	1.54E+09	2.99E+09	2.22E+09	90.26862	223.41997				
26AsnH	2.11E+09	2.08E+09	1.83E+09	6.07E+09	0.04368	2.39E+17	1.62E+09	5.56E+09	280.9227	2207.201
27GlyH	1.97E+09	1.85E+09	4.07E+09	2.86E+09	55.92187	131.01445				
27GlyH	1.65E+09	2.11E+09	3.40E+09	6.01E+09	0.03765	1.45E+20	2.36E+09	5.49E+09	304.2828	1692.975
31ValH	2.58E+09	9.44E+08	4.52E+09	1.47E+09	51.84843	56.78081				
31ValH	1.98E+09	7.57E+07	4.51E+09	1.14E+08	1.24095	0.08669	2.28E+09	9.09E+07	719.0645	130.6362
39ThrH	2.40E+09	2.10E+09	4.63E+09	3.29E+09	50.65832	120.99563				
39ThrH	2.01E+09	2.44E+09	3.88E+09	7.06E+09	0.02152	9.37E+36	2.67E+09	6.46E+09	289.5331	1628.309
46HisH	1.16E+09	1.59E+09	3.56E+09	2.40E+09	66.13533	148.08272				
46HisH	9.41E+08	2.32E+09	1.98E+09	6.98E+09	0.01878	1.06E+43	1.94E+09	6.43E+09	253.962	1821.005
50PheH	2.18E+09	3.30E+09	3.04E+09	3.20E+09	1941.943	4618.4565				
50Phe	6.41E+	6.70E+	1.03E+	1.49E+	71857.	1.12E+1	3.96E+	8.18E+	25086.	1.29E+

H	11	16	12	16	45	0	11	16	83	09
51GlyH	1.70E+ 09	2.37E+ 09	5.27E+ 09	3.47E+ 09	82.120 66	180.517 74				
51GlyH	1.48E+ 09	2.42E+ 09	4.33E+ 09	6.90E+ 09	0.0623 6	2.59E+1 1	3.11E+ 09	6.30E+ 09	303.16 93	1464.1 18
52Asp H	2.49E+ 09	5.40E+ 09	8.22E+ 09	1.21E+ 10	0.0136 6	5.29E+5 9				
52Asp H	1.74E+ 09	2.69E+ 09	3.46E+ 09	5.17E+ 09	0.0025 2	--	2.62E+ 09	4.31E+ 09	210.30 95	933.91 72
54ThrH	1.51E+ 09	2.57E+ 09	6.26E+ 09	3.49E+ 09	213.61 14	357.119 71				
54ThrH	1.63E+ 09	2.70E+ 09	5.00E+ 09	5.09E+ 09	0.0332 5	7.47E+2 2	3.66E+ 09	4.07E+ 09	595.97 46	2776.1 5
55AlaH	1.37E+ 09	3.22E+ 09	6.72E+ 09	4.39E+ 09	200.26 35	391.948 71				
55AlaH	1.28E+ 09	5.11E+ 09	3.56E+ 09	9.68E+ 09	0.0489 8	1.74E+1 5	4.27E+ 09	7.75E+ 09	594.09 18	4508.4 25
56GlyH	1.97E+ 09	3.00E+ 09	7.48E+ 09	4.56E+ 09	65.099 56	131.709 28				
56GlyH	1.40E+ 09	1.19E+ 10	6.54E+ 09	3.22E+ 10	0.0144 3	8.09E+5 6	4.38E+ 09	2.90E+ 10	351.99 9	6178.8 23
57CysH	1.21E+ 09	2.13E+ 09	5.19E+ 09	2.92E+ 09	150.05 48	260.390 99				
57CysH	1.14E+ 09	1.79E+ 09	4.38E+ 09	4.00E+ 09	0.0505 7	1.73E+1 4	3.09E+ 09	3.36E+ 09	500.02 86	1965.7 23
60AlaH	1.49E+ 09	1.77E+ 09	4.25E+ 09	2.68E+ 09	66.341 88	138.733 91				
60AlaH	1.20E+ 09	1.93E+ 09	3.28E+ 09	5.37E+ 09	0.046 6	1.30E+1 6	2.48E+ 09	4.88E+ 09	325.00 08	1597.9 2
61GlyH	1.16E+ 09	2.55E+ 09	5.60E+ 09	3.59E+ 09	105.01 76	221.498 74				
61GlyH	9.41E+ 08	8.98E+ 09	4.73E+ 09	2.32E+ 10	0.0075 6	9.55E+1 10	3.29E+ 09	2.05E+ 10	396.57 65	7229.6 86
63HisH	9.86E+ 08	4.90E+ 08	3.80E+ 09	1.10E+ 09	0.0426 9	5.83E+1 6				
63HisH	4.80E+ 08	3.73E+ 09	1.86E+ 09	1.11E+ 10	0.0115 9	3.64E+7 1	1.64E+ 09	1.02E+ 10	260.58 26	3574.7 37
68SerH	9.29E+ 08	1.94E+ 09	4.15E+ 09	2.69E+ 09	123.64 52	256.677 57				
68SerH	6.77E+ 08	1.46E+ 09	6.27E+ 09	1.55E+ 09	0.0031 6	3.44E+1 7	1.56E+ 09	1.48E+ 09	849.43 34	3363.9 55
71HisH	2.10E+ 09	6.63E+ 08	6.19E+ 09	1.48E+ 09	0.0266 7	3.18E+2 8				
71HisH	1.45E+ 09	2.55E+ 09	3.50E+ 09	8.09E+ 09	0.0352 4	6.30E+2 1	2.36E+ 09	7.47E+ 09	222.00 84	1436.7 99
82GlyH	1.10E+ 09	6.17E+ 08	5.20E+ 09	1.38E+ 09	0.0500 4	7.59E+1 3				
82GlyH	6.84E+ 08	2.64E+ 09	2.43E+ 09	9.70E+ 09	0.0420 4	1.58E+1 8	2.03E+ 09	8.96E+ 09	166.74 47	1385.6 79
87ValH	1.69E+ 09	1.48E+ 09	3.17E+ 09	2.15E+ 09	86.775 92	196.088 07				
87ValH	1.53E+ 09	1.77E+ 09	2.39E+ 09	4.78E+ 09	0.0210 9	8.33E+9 9	1.88E+ 09	4.30E+ 09	355.96 9	2.17E+ 9

	09	70	09	70	4	8	09	70	7	64
88ThrH	6.26E+09	1.90E+09	7.52E+09	2.57E+09	263.2714	277.21747				
88ThrH	6.22E+09	2.72E+09	1.65E+09	6.87E+09	0.01223	3.20E+67	6.24E+09	6.03E+09	414.0123	1213.192
89AlaH	2.99E+09	2.46E+09	4.94E+09	3.51E+09	94.88983	224.38259				
89AlaH	2.59E+09	1.68E+09	7.38E+09	2.23E+09	0.00103	--	1.99E+09	1.89E+09	679.9355	2843.223
94ValH	2.60E+09	2.62E+09	4.80E+09	3.62E+09	129.753	310.83226				
94ValH	2.51E+09	3.47E+09	2.95E+09	9.15E+09	0.04958	1.24E+15	2.65E+09	8.16E+09	378.1637	3270.411
95AlaH	3.19E+09	1.17E+09	1.08E+10	2.62E+09	0.06669	5.70E+09				
95AlaH	1.97E+09	4.04E+09	6.47E+09	1.36E+10	0.03304	2.19E+23	4.42E+09	1.26E+10	192.6431	1072.495
96AspH	4.42E+09	4.18E+09	1.11E+10	6.60E+09	45.63048	93.32718				
96AspH	3.48E+09	1.74E+10	9.83E+09	5.09E+10	0.01482	2.19E+55	6.38E+09	4.67E+10	275.5648	4552.074
97ValH	5.18E+09	1.41E+09	4.82E+09	1.79E+09	428.7594	598.76338				
97ValH	5.19E+09	1.83E+09	2.08E+09	2.84E+09	0.04187	6.66E+17	3.89E+09	2.25E+09	703.1041	1850.044
100GluH	8.70E+09	3.07E+09	1.08E+10	4.25E+09	129.9247	161.50301				
100GluH	8.38E+09	2.37E+09	7.41E+09	5.54E+09	0.03615	5.18E+20	6.33E+09	4.71E+09	472.9041	1208.399
103ValH	2.59E+09	1.83E+09	3.37E+09	2.50E+09	175.9806	394.29765				
103ValH	2.59E+09	2.77E+09	2.80E+09	7.89E+09	237.5869	1407.9716	9.21E+08	8.56E+09	0.03802	4.63E+20
107SerH	1.25E+10	1.54E+09	1.82E+10	3.43E+09	1.0319	0.68442				
107SerH	9.98E+09	3.65E+09	1.01E+10	5.60E+09	0.00167	--	6.96E+09	4.58E+09	562.0062	1552.145
108GlyH	3.05E+09	3.54E+08	9.78E+09	7.93E+08	0.63469	0.3786				
108GlyH	2.41E+09	2.96E+09	3.77E+09	1.03E+10	0.0115	1.43E+75	3.01E+09	9.50E+09	182.0689	1105.924
109AspH	5.44E+09	2.27E+09	1.85E+10	5.08E+09	0.02662	4.13E+28				
109AspH	4.31E+09	4.20E+09	2.62E+10	2.34E+09	0.00151	--	9.72E+08	3.99E+09	1417.635	17111.35
110HisH	1.90E+09	3.33E+08	6.34E+09	7.44E+08	0.05777	2.12E+11				
110HisH	1.30E+09	2.48E+09	1.97E+09	7.94E+09	0.04375	2.69E+17	1.93E+09	7.33E+09	215.6266	1654.478
111CysH	8.02E+09	1.33E+09	2.05E+10	2.98E+09	0.80556	0.55815				
111Cys	5.76E+09	3.85E+09	1.19E+09	1.01E+09	0.0400	3.37E+1	7.08E+09	8.97E+09	384.14	1383.2

H	09	09	10	10	3	8	09	09	61	57
113Ile	4.93E+	9.70E+	3.46E+	1.18E+	498.78	687.549				
H	09	08	09	09	29	46				
113Ile	5.08E+	1.05E+	1.63E+	1.79E+	0.0508	1.69E+1	2.67E+	1.42E+	650.04	1524.9
H	09	09	09	09	6	4	09	09	55	71
114Gly	4.30E+	1.14E+	4.04E+	1.52E+	317.13	387.475				
H	09	09	09	09	98	52				
114Gly	4.31E+	4.38E+	2.61E+	5.50E+	0.0199	1.56E+4	2.56E+	4.61E+	834.00	6788.9
H	09	09	09	09	3	0	09	09	93	19
115Arg	2.88E+	3.91E+	7.17E+	8.73E+	0.9422	0.44417				
H	09	08	09	08	4					
115Arg	2.24E+	1.16E+	4.03E+	3.74E+	0.0476	1.68E+1	2.52E+	3.45E+	213.36	589.42
H	09	09	09	09	9	5	09	09	49	12
120His	1.23E+	2.80E+	6.25E+	6.26E+	1.4190	0.40501				
H	09	08	09	08	4					
120His	8.16E+	5.29E+	3.21E+	2.45E+	0.0170	7.60E+4	2.30E+	2.26E+	128.62	2238.8
H	08	09	09	10	8	7	09	10	23	76
123Ala	7.85E+	3.12E+	1.01E+	4.28E+	155.15	201.069				
H	09	09	10	09	09	53				
123Ala	7.41E+	5.20E+	9.78E+	9.32E+	0.0359	8.75E+1	6.13E+	7.41E+	624.06	325.35
H	09	08	09	08	6	9	09	08	31	71
126Leu	2.41E+	2.20E+	4.14E+	3.02E+	143.05	324.895				
H	09	09	09	09	05	98				
126Leu	2.37E+	4.45E+	1.53E+	1.35E+	0.0085	4.41E+9	2.95E+	1.24E+	250.79	2270.6
H	09	09	09	10	9	7	09	10	14	58
134Ser	3.03E+	2.23E+	4.86E+	3.05E+	182.18	343.557				
H	09	09	09	09	29	1				
134Ser	3.00E+	3.30E+	1.49E+	9.15E+	0.0404	1.51E+1	3.66E+	8.30E+	328.25	1873.0
H	09	09	09	09	3	9	09	09	76	91
137Thr	2.72E+	2.27E+	4.28E+	3.20E+	104.03	255.996				
H	09	09	09	09		47				
137Thr	2.64E+	5.34E+	2.38E+	1.57E+	0.0073	2.33E+1	2.31E+	1.44E+	274.03	3848.5
H	09	09	09	10	5	14	09	10	37	49
138Gly	3.67E+	2.82E+	6.08E+	4.35E+	58.933	139.826				
H	09	09	09	09	15	68				
138Gly	3.17E+	3.99E+	3.85E+	1.12E+	0.0380	1.46E+2	3.44E+	1.02E+	312.70	2266.7
H	09	09	09	10	2	0	09	10	17	54
140Ala	3.65E+	3.34E+	7.10E+	5.14E+	58.906	141.426				
H	09	09	09	09	75	06				
140Ala	3.15E+	5.05E+	3.97E+	1.52E+	0.0701	2.25E+1	3.87E+	1.40E+	254.38	1996.0
H	09	09	09	10	3	0	09	10	17	42
141Gly	3.51E+	3.11E+	6.82E+	4.87E+	49.993	120.409				
H	09	09	09	09	33	93				
141Gly	2.90E+	4.42E+	4.38E+	8.31E+	0.0047	--	3.84E+	6.94E+	277.14	1431.3
H	09	09	09	09	5		09	09	87	42
142Ser	3.67E+	1.36E+	1.01E+	3.04E+	0.0489	1.95E+1				
H	09	09	10	09	8	4				
142Ser	2.30E+	4.51E+	5.46E+	1.33E+	0.0515	2.23E+1	4.45E+	1.23E+	269.00	1651.5
H	09	09	09	10	7	4	09	10	62	48
143Arg	1.89E+	2.42E+	6.21E+	3.68E+	65.229	128.199				
H	09	09	09	09	27	82				
143Arg	1.51E+	2.17E+	5.36E+	6.27E+	0.0265	2.40E+2	3.62E+	5.74E+	290.56	1071.9

H	09	09	09	09		9	09	09	32	61
146Cys	2.87E+	2.16E+	6.58E+	2.94E+	200.21	268.379				
H	09	09	09	09	72	36				
146Cys	2.84E+	1.81E+	4.90E+	3.22E+	0.0569	1.89E+1	3.95E+	2.55E+	630.03	1764.9
H	09	09	09	09	4	2	09	09	56	47
147Gly	4.76E+	3.33E+	9.10E+	4.76E+	94.814	164.932				
H	09	09	09	09	15	53				
147Gly	4.16E+	2.49E+	7.81E+	5.67E+	0.0631	8.33E+1	5.36E+	4.78E+	488.25	1542.2
H	09	09	09	09		0	09	09	94	68
148Val	5.24E+	9.45E+	7.53E+	2.11E+	1.0671	1.02147				
H	09	08	09	09	4					
148Val	5.24E+	1.64E+	3.85E+	--	0.9165	--	3.67E+	--	1.2277	--
H	09	09	09		1		09		9	
149Ile	3.48E+	1.41E+	4.60E+	1.92E+	199.66	249.967				
H	09	09	09	09	11	52				
149Ile	3.45E+	1.03E+	3.41E+	1.80E+	0.0455	6.59E+1	2.77E+	1.42E+	638.56	1431.9
H	09	09	09	09	1	5	09	09	44	63
150Gly	4.60E+	2.95E+	9.60E+	4.69E+	38.446	69.7251				
H	09	09	09	09	79	9				
150Gly	3.44E+	2.44E+	7.96E+	6.42E+	0.0167	1.00E+4	5.50E+	5.72E+	380.45	1116.7
H	09	09	09	09	2	8	09	09	93	24
151Ile	6.56E+	2.27E+	7.99E+	3.07E+	250.22	292.819				
H	09	09	09	09	46	79				
151Ile	6.42E+	1.10E+	6.21E+	1.62E+	0.0362	1.49E+2	5.30E+	1.29E+	736.20	820.84
H	09	09	09	09	8	0	09	09	12	72
152Ala	4.42E+	3.56E+	1.11E+	5.52E+	55.351	91.5326				
H	09	09	10	09	45	2				
152Ala	3.53E+	2.58E+	9.48E+	7.31E+	0.0494	3.50E+1	6.41E+	6.67E+	307.24	769.67
H	09	09	09	09	2	4	09	09	14	3
153Gln	4.50E+	4.79E+	1.24E+	6.70E+	112.27	197.091				
H	09	09	10	09	14	7				
153Gln	4.26E+	3.52E+	1.06E+	9.55E+	0.0441	3.97E+1	7.28E+	8.61E+	351.16	1097.7
H	09	09	10	09	8	6	09	09	98	53

Table A.4. 4 Fit values of urea denaturation

Sample	C _{mid}	error	n	error
uninduced	4.33	2.95	-1.09	1.11
A4V 37 T4	1.97	0.89	-1.73	1.36
G37R 37 T4	2.09	0.78	-2.18	1.66
G41D 37 T4	2.11	1.15	-1.19	0.98
G41S 37 T4	3.02	1.47	-1.36	1.15
G85R 37 T4	2.28	0.8	-2.45	1.9
G93A 37 T4	2.18	0.9	-1.88	1.45

G93D 37 T4	2.17	1.08	-1.36	1.08
V148G 37 T4	1.75	0.87	-1.48	1.16
V148I 37 T4	2.57	1.45	-1.11	0.97
A4V 37 Zn T4	2.79	0.98	-2.31	1.79
G85R 37 Zn T4	1.49	1.52	-1.72	1.49
G93A 37 Zn T4	3.33	1.12	-2.39	1.96
A4V 25C ON	2.41	0.94	-1.98	1.51
G85R 25C ON	2.37	0.89	-2.12	1.6
G93A 25C ON	2.1	0.88	-1.81	1.37
V148G 25C ON	1.81	1.11	-1.07	0.91

Table A.4. 5 Fit peak locations for FTIR of SOD1 IB samples grown with zinc

	Literature Value, (range)	Average IB	A4V Zn	G85R Zn	G93A Zn
β region (cm^{-1})	1633 (1623 – 1641)	1631.47	1629.54	1629.54	1630.51
Unfolded region (cm^{-1})	1654 (1642 – 1657)	1655.58	1655.58	1654.61	1654.61

Table A.4. 6 Fit peak locations for FTIR of SOD1 IB samples grown at lower temperatures

	Literature Value, (range)	Average IB	A4V 25 ON	G85R 25 ON	G93A 25 ON	V148G 25 ON
β region (cm^{-1})	1633 (1623 – 1641)	1631.47	1630.51	1629.54	1630.51	1631.47
Unfolded region (cm^{-1})	1654 (1642 – 1657)	1655.58	1654.61	1654.61	1654.61	1654.61