Structure of Human Cu, Zn Superoxide Dismutase-1 Inclusion Body Aggregates using Quenched Hydrogen-Deuterium Amide Exchange

by

Michael V. Tarasca

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Science
in
Chemistry

Waterloo, Ontario, Canada, 2021

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 Contribution

This thesis consists in part of two manuscripts submitted for publication. Both manuscripts have not yet been accepted at the time of final submission of this thesis. Exceptions to sole authorship of material are as follows:

Research presented in Chapter 2:

This research was conducted at the University of Waterloo by Michael Tarasca in collaboration with Dalia Naser under the supervision of Dr. Elizabeth Meiering. Michael Tarasca and Dalia Naser contributed equally to both experimental design and execution, with both authors listed as co-first authors for publication. Anna Schaefer provided results for the protein Adnectin and provided intellectual input on the manuscript drafts. Tyler Soule aided in experimental execution during his time as an undergraduate research assistant.

Citation:

*__Tarasca, M. V__, *Naser, D., Schaefer, A., Soule, T. G.B., and Meiering, E. M. High-resolution structure of protein inclusion body aggregates by quenched hydrogen-deuterium amide exchange. *Anal. Biochem. (Under Review).* 2021

* Indicates these authors contributed equally to the work.

Research presented in Chapter 3:

This research was conducted at the University of Waterloo by Michael Tarasca in collaboration with Dalia Naser, Anna Schaefer, and Bruna Siebeneichler under the supervision of Dr. Elizabeth Meiering. Michael Tarasca and Dalia Naser contributed equally to both experimental design and execution, with both authors listed as co-first authors for publication. Michael Tarasca contributed to NMR spectral data acquisition, analysis, and interpretation, Dalia Naser contributed to NMR data acquisition, analysis including normalization, Congo red spectral assays, and FTIR acquisition. Bruna Siebeneichler and Anna Schaefer contributed FTIR data acquisition and analysis and to manuscript preparation. Harmeen Deol contributed to data interpretation. Tyler Soule, John Almey, Gyan Mishra, and Hilary Simon contributed preliminary or minor experimental work.

Citation:

*Naser D, *__Tarasca M V__, Siebeneichler B, Schaefer A, Deol HK, Soule T GB, Almey J, Kelso S, Mishra GG, Simon H, and Meiering EM. High Resolution NMR H / D Exchange of Human Cu, Zn Superoxide Dismutase Inclusion Bodies Reveals Significant Native Features Despite Structural Heterogeneity. *Angew. Chemie – Int. Ed.* (*Under Review).* 2021.

* Indicates these authors contributed equally to the work.

Abstract

Inclusion bodies (IBs) are insoluble protein aggregates that can be formed within recombinant bacteria. These aggregates, once thought to be an obstacle in recombinant protein expression and purification, have recently gained attention to their potential applications in biotechnology and medicine. However, despite their attractive qualities, there have been little investigations into their high-resolution structure. Presented is the method quenched hydrogen-deuterium amide exchange (qHDX), a method in which solvent-exposed amide protons of IB aggregates are exchanged with deuterium, lyophilized, then dissolved in DMSO and observed by two-dimension nuclear magnetic resonance (NMR). This method allows one to probe the map the solvent-exposed and solvent-protected backbone amides of the aggregated protein.

qHDX was optimized on IBs of Cu, Zn superoxide dismutase-1 (SOD1), a homodimeric metalloenzyme that is associated with the development of neurodegenerative diseases, including amyotrophic lateral sclerosis (ALS). The IB structure of SOD1 was probed using various mutants across three different expression conditions. Remarkably, the mutants all adopt similar, extensive native-like structure with some amyloid features, despite large differences in protein stabilities and mutation types. These results hold true across all measured expression conditions. These findings are of broad significance for understanding cellular protein aggregation. The structural characteristics of SOD1 IBs and effects of mutations may be valuable for advancing rational design of IBs. qHDX is a powerful tool that has potential to pave the way for advancements in the field of protein engineering as an effective method to measure the high-resolution structure of *in vivo* protein aggregates.

Acknowledgements

I would like to express gratitude to my supervisor, Dr. Elizabeth Meiering, for providing the opportunity to work in her group and to grow as a scientist through both my undergraduate and graduate studies.

I would also like to thank members of the Meiering lab, both past and present, for their love and support throughout my degree. I would not have made it this far without their help and input. I would specifically like to acknowledge Dalia Naser and Harmeen Deol for being there alongside me every step of the way, through both the ups and downs on my journey. I would also like to thank Fabio Cuzzucoli, Javan Buratynski, Khalid Shah, Khue Nguyen, Gyan Mishra, and Dr. Kyle Piccolo for their friendship and support throughout our time spent together in Waterloo.

Finally, I would like to thank my family and loved ones, especially my partner Samantha Siegel, and my parents Vince and Michelle Tarasca for being constant pillars of support in my life during my studies. Words cannot begin to express my gratitude to you.

Dedication

This one’s for the boys.

Table of Contents

Author's Declaration ii
Statement of Contribution iii
Abstract iv
Acknowledgements v
Dedication vi
List of Figures x
List of Tables xii
List of Abbreviations xiii
Chapter 1 1
1.1 Protein Folding, Misfolding, and Aggregation 2
1.2 Neurodegenerative Disease 7
1.2.1 Amyotrophic Lateral Sclerosis 7
1.2.2 Prion Strain Behaviour in Amyotrophic Lateral Sclerosis 8
1.3 Human Cu, Zn Superoxide Dismutase-1 10
1.3.1 Structure and Function 10
1.3.2 SOD1 Maturation 11
1.4 Nuclear Magnetic Resonance 13
1.4.1 NMR Background 13
1.4.2 Quenched Hydrogen-Deuterium Amide Exchange 14
1.5 Research Objectives 17
Chapter 2 18
2.1 Quenched HDX Workflow 19
2.2 Materials and Methods 21
2.2.1 Materials 21
2.2.2 Cellular Growth and Protein Expression 21
2.2.3 Preparation of IB Samples 22
2.2.4 Preparation of NMR Samples 23
2.2.5 NMR Experiments 24
2.2.6 Quenched HDX and k_{int} Control NMR Experiments 24
2.2.7 Fully Protonated Control NMR Experiments 25
2.2.8 Sequence Specific Resonance Assignments 26
2.2.9 Data Analysis 27

2.3 Control of IB Sample Conditions 27
2.3.1 Controlling the Intrinsic Exchange Rates 29
2.3.2 Sample pH 32
2.3.3 Sample H_2O Fraction, χ_H 35
2.3.4 Temperature 37
2.3.5 Time of IB Exchange in D_2O 38
2.3.6 Storage of Lyophilized Samples 39
2.4 Spectral Analysis 39
2.4.1 Resonance Assignments on an Unfolded Protein in DMSO 40
2.4.2 Quantitation of Amide Protection 44
2.4.3 Replicate Samples 46
2.5 Conclusions 47
Chapter 3 49
3.1 Introduction 50
3.1.1 SOD1 Inclusion Bodies Extent of Formation 50
3.1.2 SOD1 Aggregation Prediction 53
3.2 Materials and Methods 56
3.2.1 Protein Expression, IB Preparation, and NMR Data Acquisition 56
3.2.2 Aggregation Prediction 57
3.3 SOD1 qHDX Results 58
3.3.1 Reduced apo SOD1 IBs expressed at 37 °C for 4 hours 58
3.3.2 Reduced apo SOD1 IBs expressed at 25 °C overnight 61
3.3.3 Zn-bound SOD1 IBs expressed at 37 °C for 4 hours 64
3.4 SOD1 qHDX Discussion 68
3.5 qHDX applied to other proteins 78
3.6 Conclusions 80
Chapter 4 81
4.1 Conclusions 82
4.2 Future Work 83
References 84
Appendix 94
Appendix-Chapter 2 94

Appendix-Chapter 3 100

List of Figures

Figure 1.1: Classical and Energy Landscape Funnel Theories of Protein Folding. 3
Figure 1.2: Energy Landscape of Protein Folding and Aggregation. 5
Figure 1.3: Composition of IB Aggregate. 7
Figure 1.4: Model of Prion-Like Strains. 8
Figure 1.5: Ribbon Structure of Human Cu, Zn Superoxide Dismutase 1. 10
Figure 1.6: SOD1 Catalytic Cycle. 11
Figure 1.7: Maturation Scheme of SOD1. 12
Figure 1.8: Quenched hydrogen-deuterium amide exchange on protein aggregates. 16

Figure 2.1: Flowchart of the IB quenched HDX method. 19
Figure 2.2: Different ^{1}H-^{15}N HSQC spectra acquired for quenched HDX experiments. 26
Figure 2.3: Measured intrinsic rate constants of amide exchange for different amino acids. 32
Figure 2.4: DCA methyl ^{1}H peak as an internal indicator of sample pH. 34
Figure 2.5: ^{1}H-^{15}N HSQC spectral overlay of a quenched HDX sample in DMSO with 0% (v/v) water added (blue) and 5% water added (red). 35
Figure 2.6: Comparisons of replicate qHDX sample results. 47

Figure 3.1: SOD1 Structure, Stability, and IB Formation. 51
Figure 3.2: Aggregation and Secondary Structure Prediction for SOD1. 55
Figure 3.3: Structure of Mutant SOD1 IBs from qHDX Analysis. 60
Figure 3.4: qHDX amide protection of mutant SOD1s expressed at 25 °C overnight. 62
Figure 3.5: Comparison of average fraction amide protection between expression temperatures. 64
Figure 3.6: qHDX amide protection of mutant SOD1s expressed at 37 °C for 4 hours with Zn. 65
Figure 3.7: Average qHDX protection of mutant SOD1 expressed with and without Zn. 67
Figure 3.8: Comparison of fraction amide protection between expression conditions. 72
Figure 3.9: FTIR and CR Spectra of IBs exhibit Native and Amyloid Features. 74
Figure 3.10: SOD1 IB Formation Involves Multiple Pathways that Include Native and Non-Native Interactions. 77
Figure 3.11: Structure of Adnectin IBs from qHDX analysis. 79

Figure A2.1: Amide exchange decay fits in CCPTNMR software. 94
Figure A2.2: Effects of pH on ^{1}H-^{15}N HSQC cross peak intensity and chemical shift. 95
Figure A2.3: ^{1}H 1D NMR spectra overlaid to highlight differences in water content across samples. 96
Figure A2.4: Overlay of ^{1}H-^{15}N spectra at different acquisition temperatures. 97
Figure A2.5: Effects of varying IB storage. 98

Figure A2.6: ^{1}H-^{15}N spectrum of A4V SOD1 unfolded in 95% DMSO / 5% H_2O at pH 5.5 with backbone residue assignments. 99

Figure A3.1: SDS-PAGE of mutant SOD1 IB formation and disulphide status. 100
Figure A3.2: The % of SOD1 that forms insoluble IBs increases at higher growth temperature. 101
Figure A3.3: Normalized Delta Predictor Scores. 103
Figure A3.4: Protection ratios - rApo expressed at 37° C for 4 hours. 105
Figure A3.5: Protection ratios - rApo expressed at 25 °C overnight. 106
Figure A3.6: Protection ratios - Zn expressed at 37 °C for 4 hours. 107
Figure A3.7: Biphasic exchange of IB amide protons in D_2O. 108
Figure A3.8: ^{1}H-^{15}N HSQC spectra for aliquots of G85R SOD1 IBs at different times of exchange in D_2O. 109
Figure A3.9: Comparison of amide protection between different expression conditions. . 110

List of Tables

Table 2.1: Random coil chemical shifts of SOD1 in DMSO. 41
Table 2.2: Primary sequence effects on ^{15}N chemical shifts of amino acids. 43

Table 3.1: Summary of Aggregation Propensity and Biophysical Parameters of SOD1 52
Table 3.2: Normalized Aggregation Propensity Predictions. 54

Table A3.1: ^{1}H-^{15}N backbone assignments of A4V SOD1 in 95% DMSO / 5% H_2O at pH 5.5. 115

List of Abbreviations

1D	One-dimensional
2D	Two-dimensional
3D	Three-dimensional
ALS	Amyotrophic lateral sclerosis
CCS	Copper chaperone protein for superoxide dismutase 1
CD	Circular dichroism
CR	Congo red
Cu,Zn-SOD1^{SS}	holo SOD1
D1	Native dimer
D2	Non-native symmetric dimer
D3	Non-native asymmetric dimer
DCA	dichloroacetic acid
ΔG_d	Gibbs free energy of dimer dissociation
DMSO	Dimethyl sulfoxide
DSC	Differential scanning calorimetry
DTT	dithiothreitol
E,E-SOD1^{SH}	reduced apo SOD1
E,Zn-SOD1^{SH}	reduced Zn-bound SOD1
EDTA	ethylenediaminetetraacetic acid
fALS	Familial ALS
FTIR	Fourier transform infrared
HSQC	heteronuclear single quantum correlation
IB	Inclusion body
IPTG	isopropyl b-D-1-thiogalactopyranoside
k_{ex}	Observed exchange rate
k_{int}	Intrinsic rate of amide exchange
LB	Luria broth
T_m	Melting point
NMR	Nuclear magnetic resonance
polyQ	Repeated glutamine
PrP	Prion protein
pWT	pseudo wild-type
qHDX	Quenched hydrogen-deuterium amide exchange
rApo	reduced apo
S/N	Signal-to-noise
sALS	Sporadic ALS
SOD1	Human Cu, Zn superoxide dismutase 1
ssNMR	solid-state NMR
TDP43	TAR DNA binding protein 43
TEN	Tris, EDTA, NaCl
Tris	tris(hydroxymethyl)aminomethane
WT	Wild-type
χH	Residual water within qHDX sample
YFP	Yellow fluorescent protein

Chapter 1

Introduction: Protein Misfolding and Aggregation, SOD1, Quenched Hydrogen Deuterium Amide Exchange, and Nuclear Magnetic Resonance

1.1 Protein Folding, Misfolding, and Aggregation

Proteins are large biomacromolecules that exist in all life on Earth. They are composed of amino acid monomers covalently linked through peptide bonds to form a linear sequence known as the protein's primary structure.[1] These linear sequences fold to form higher order structures.[1] There is debate on how or why exactly proteins fold the way they do.[2,3] The classical model of protein folding (**Figure 1.1A**)[2] postulates that a protein folds along a distinct pathway, forming clearly defined intermediate structures before it ultimately achieves the final native fold.[2,4] This classical model provides a solution to Levinthal's paradox. Levinthal proposed that there is no undirected folding process for proteins, as the time it would take a protein to randomly sample every potential form before finding its native structure is longer than the current age of the universe.[2] Working at the same time as Levinthal, Anfinsen observed through his work on ribonuclease that protein folding was a quick process.[5,6] Because of this, the idea was proposed that a protein folds via intermediates, rather than randomly sampling available conformations.[2] This led to the development of Anfinsen's dogma.[5]

According to Anfinsen's dogma, also known as the thermodynamic hypothesis, the primary amino acid sequence ultimately determines the native three-dimensional (3D) folded structure of the protein.[2,5] Formation of the native structure requires three conditions to be fulfilled: a unique structure at a global free energy minimum, a stable structure that is unaffected by minor changes in the surrounding system, and a structure obtainable through a kinetic pathway.[2,5] From this hypothesis, several new models of protein folding began to develop. These new theories led to the development of the energy landscape funnel theory (**Figure 1.1B**).[2,7] This theory states that there are multiple routes that a protein may follow throughout the folding process.[2,8] Consequently, there are many folding intermediates that the protein may sample before it reaches its native state.

However, rather than the random sampling of folding pathways, in most cases each folding intermediate will generally be more energetically favourable than the preceding one, to not escape Levinthal's paradox.[2] Both the classical model and energy landscape funnel are illustrated in **Figure 1.1**.[2]

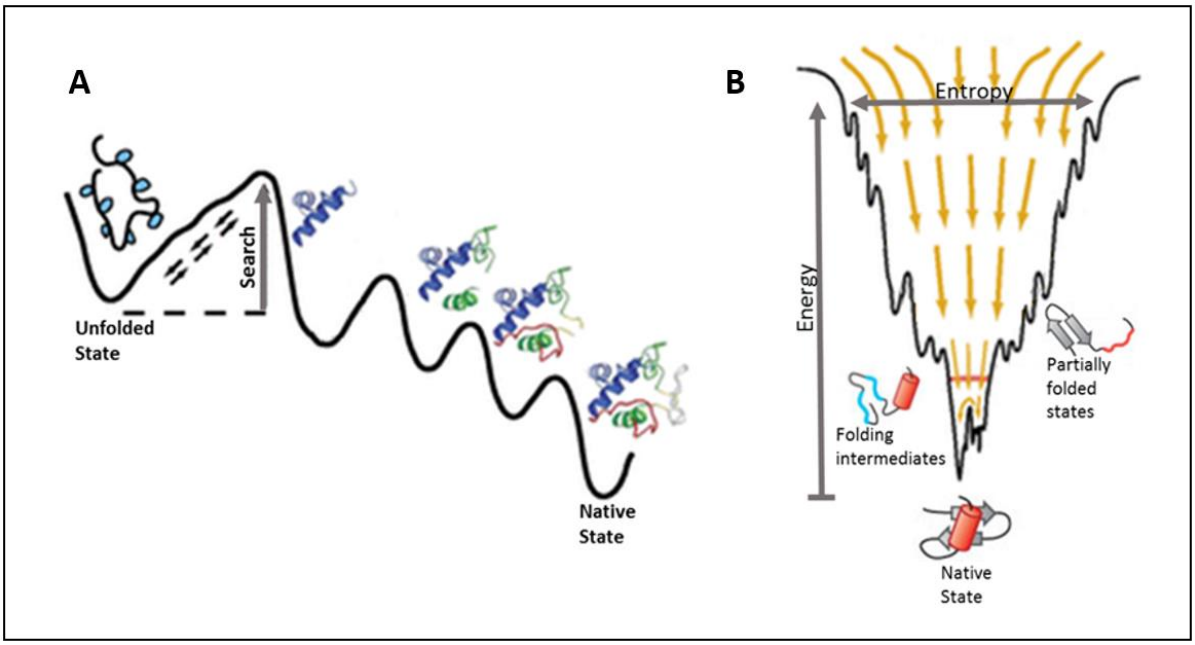


Figure 1.1: Classical and Energy Landscape Funnel Theories of Protein Folding.
The classical model, A, displays clearly defined folding intermediates that are formed prior to reaching the final native structure. The energy landscape funnel, B, illustrates that proteins can sample conformations that are decreasing in energy, until an energy minimum is reached. Both models A and B are shown to follow the conditions established in Anfinsen's dogma, while also not violating Levinthal's paradox. Slightly modified from (A,B) Englander and Mayne, 2014[2] and (B) Hartl & Hayer-Hartl, 2009[3].

Protein misfolding describes errors in the folding process.[2] Such errors may occur when a protein is sampling conformations and forms an off-pathway folding intermediate.[2,3] There are several aspects to a protein's structure that can increase the probability of misfolding. These factors may include long range interactions in a native fold that may be difficult to form, and folding intermediates containing interactions that do not occur in the native state.[3] Misfolded or partially folded proteins often contain solvent-exposed hydrophobic residues. This is in contrast to native

conformations, where these residues are generally buried in the core of the protein.[3] This observation is consistent with the hydrophobic collapse model of protein folding, according to which a primary sequence collapses into itself due to the hydrophobic effect, and followed by further folding steps.[9] Surface level hydrophobic interactions between two or more misfolded proteins can thermodynamically encourage the assembly of higher order structures, known as aggregates.[3,8]

Protein aggregates can be formed through both covalent and non-covalent interactions.[8] Covalent linkages can include disulfide bonds through free cysteine residues, and non-covalent linkages can originate through van der Waals forces, dipole interactions, and hydrogen bonding.[1,3] The formation of these thermodynamically stable structures is often modeled to form through the assembly of two or more misfolded proteins, followed by formation of small, soluble aggregates, then further assembly into larger insoluble structures.[3,8] Proteins involved in non-native interactions have been shown to be able to destabilize natively folded proteins and recruit them to the aggregate structure.[3] This can be modelled into an energy landscape funnel, as illustrated in **Figure 1.2**.[3] The stability of a protein directly correlates to the difference in free energy between folded and unfolded states at equilibrium.[10] The thermodynamic stability of the protein describes the population of different states at equilibrium, and the kinetic stability is correlated to the rate of unfolding of the native state.[8,10] The structure and stability of these aggregates can vary greatly, from highly ordered amyloid fibrils, to amorphous oligomers.[3,8] These aggregates can form within neuronal cells of humans, where they are associated with the development of neurodegenerative diseases, including Alzheimer's disease, Huntington's disease, and amyotrophic lateral sclerosis (ALS).[3,8,11] Due to the association of aggregate structures with these fatal diseases, it is imperative to have detailed structural characterization for these aggregates. There has been great recent

progress in defining the high resolution structure of aggregated proteins found in disease, notably amyloid fibrils, which are long unbranched aggregates that arrange in a cross-β sheet structure.[12]

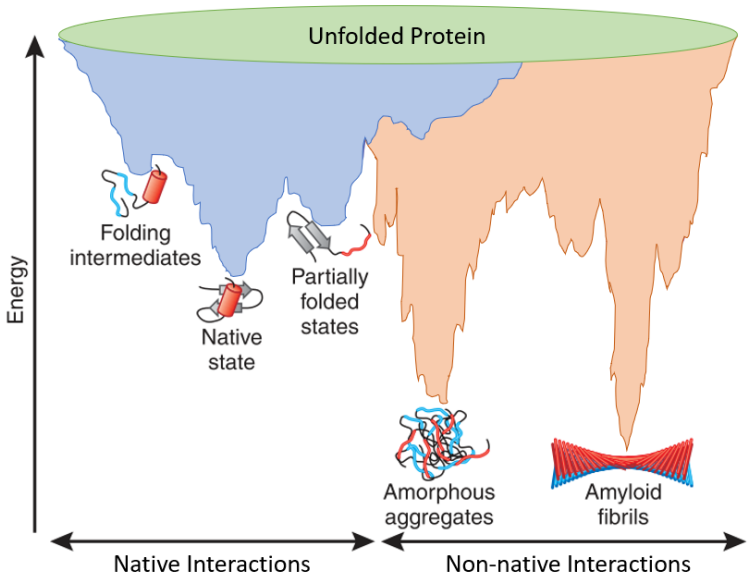


Figure 1.2: Energy Landscape of Protein Folding and Aggregation.
The blue surface shows unfolded protein folding into its native state. The orange surface shows proteins assembling into higher order aggregate structures. These surfaces are overlapped, illustrating the ability of the native states of the protein to transition into aggregate structures following destabilization. Molecular chaperones within the cell act to prevent the conversion of natively folded proteins to higher ordered structures. Slightly modified from Hartl and Hayer-Hartl, 2009.[3]

The overexpression of proteins in bacteria frequently leads to their assembly into large, insoluble aggregates known as inclusion bodies (IBs)[13,14]. While IBs have historically often been considered an obstacle in the preparation of protein using *E. coli* expression systems[15,16], recent investigations have turned to the useful functional characteristics of IBs[15]. IB applications have grown to include their use as functional nanomaterials[15,17,18], industrial large-scale protein purification[16,19], engineering of catalytically active protein aggregates[20], and medically as potential vaccines[21]. Central to developing such applications is a deeper understanding of the structures of proteins in IBs[15]. The current view is that IBs can contain a variety of conformations of

recombinant protein, including native structure, partially folded structures, and amyloid fibrils[13,14,16] (**Figure 1.3**). Despite the promise of IBs, there have been very few systematic studies characterizing their structures at high resolution. Such structural characterization may provide information key for understanding the mechanisms governing IB structure and advance the design of more robust and advanced recombinant protein expression.

In addition, IBs have been proposed as a useful model to study principles of aggregation in cells.[14,19] This is supported by the concept of protein aggregation being a conserved cell response.[19] While aggregation will differ greatly between human cells and *E. coli*, bacterial cells do provide a strong model to study the fundamentals of protein aggregation. There are noteworthy examples of IB research in disease from literature. A 2012 study by Invernizzi *et al.* investigated the aggregates of proteins with repeated glutamine (polyQ) motifs, which also occur in proteins that form aggregates in Huntington's Disease.[22] They found that the aggregates formed protected the cell from smaller soluble cytotoxic oligomers, displaying aggregation as a conservative cell response.[19,22] Another study by Sabate *et al.* in 2009 reported that the rate of transmission by IBs containing prions is dependent on the local cellular environment during aggregation.[23] Since IB formation and structure provide a tractable model to study aggregation-prone disease proteins, it naturally follows that the development of a robust method to study IB structure at high resolution would provide valuable information for aggregation associated neurodegenerative diseases. This information will also advance understanding the fundamental principles of protein aggregation in cells.[24]

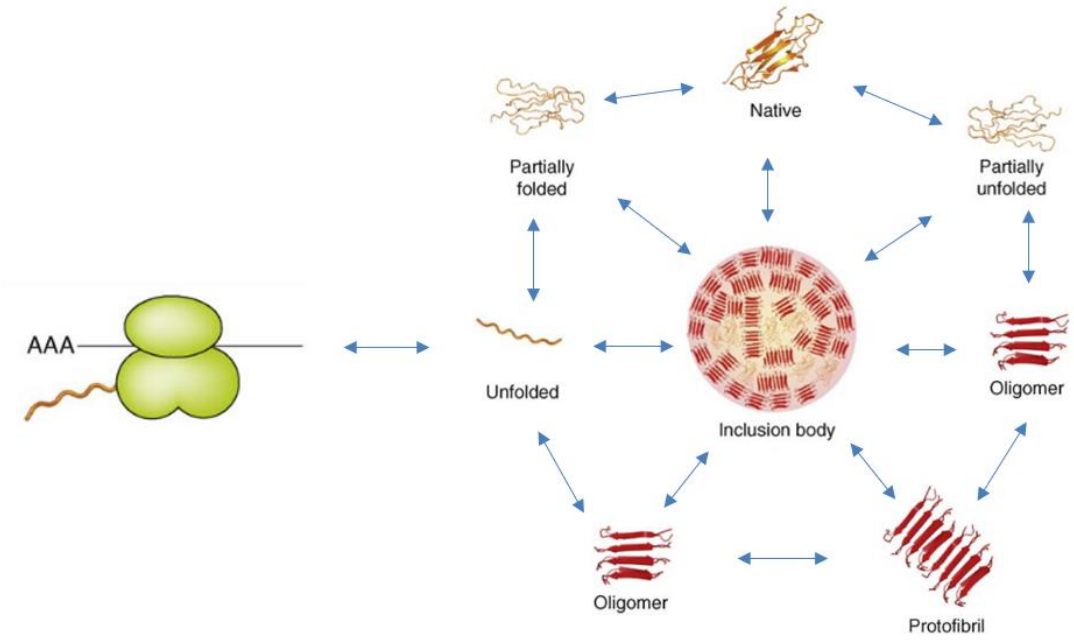


Figure 1.3: Composition of IB Aggregate.
Upon synthesis of recombinant protein in cells by the ribosome, the protein may aggregate into higher order structures. These large, dense aggregates may be composed of various forms of a recombinantly expressed protein, including native, unfolded, and partially folded structures as well as various assemblies thereof. Adapted from de Groot *et al.*, 2009.[14]

1.2 Neurodegenerative Disease

1.2.1 Amyotrophic Lateral Sclerosis

Amyotrophic Lateral Sclerosis (ALS) is a devastating neurodegenerative disease, characterized by the death of motor neurons in the spinal cord, brainstem, and motor cortex.[11,10] Patients lose control of vital muscles, often resulting in death due to respiratory failure. Familial (fALS) is the inherited form of ALS and accounts for roughly 10% of all ALS cases, and sporadic (sALS), is the uninherited form which is responsible for about 90% of cases.[11] There are many aggregation-prone proteins that have been linked to fALS, including Cu, Zn superoxide dismutase 1 (SOD1).[11] Approximately 15-20% of all fALS cases are found to be associated with SOD1 mutations.[10,11,25] Different fALS-associated SOD1 mutants result in different disease durations in ALS, however there is still considerable heterogeneity in duration even for a given mutation.[26]

Despite very extensive investigations, a clear relationship between biophysical properties of SOD1 mutants and ALS disease characteristics has not been found.[10,11]

1.2.2 Prion Strain Behaviour in Amyotrophic Lateral Sclerosis

Prions are infectious misfolded proteins that promote the conversion of natively folded protein to a pathogenic state.[27] Misfolded prion-like proteins form various non-native conformations and recruit other native proteins to enter these non-native conformations as well.[27] Different conformations, or strains, can aggregate to both smaller soluble oligomers and larger structures that can both potentially lead to neurotoxicity (**Figure 1.4**).[27] There are emerging studies of prion-like spreading of misfolded and aggregated proteins in numerous neurodegenerative diseases.[27–29] Different aggregate structures of ALS-associated proteins, including SOD1 and TAR DNA binding protein 43 (TDP43), have been observed in patients and may potentially contribute to varying degrees of neurodegeneration.[28,30]

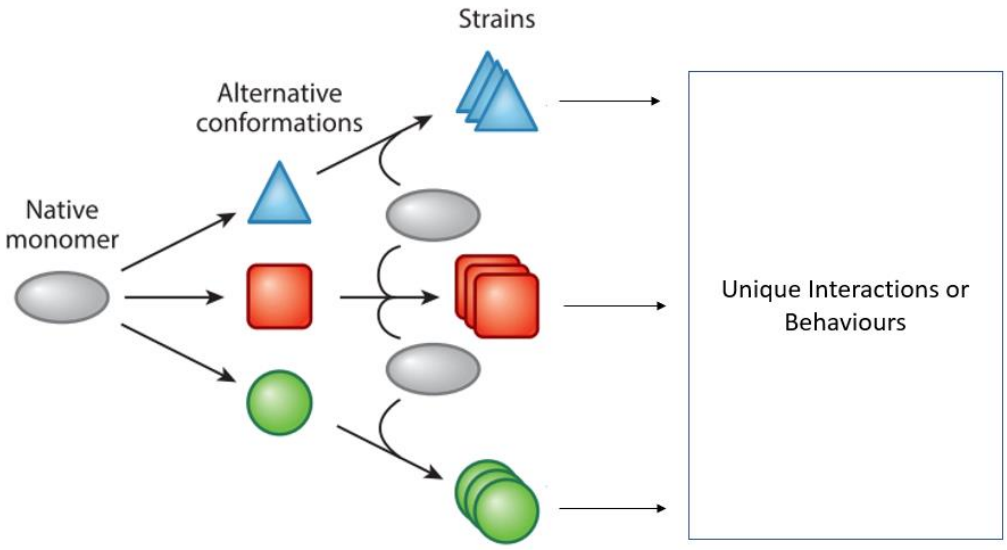


Figure 1.4: Model of Prion-Like Strains.
Natively folded monomers convert to alternate misfolded conformations which can behave in a prion like way, self associating with native monomers to further aggregate into larger complexes.

The various misfolded species may contribute to neurotoxicity. Adapted from Vaquer-Alicea and Diamond, 2019[27].

These remarkable findings support the hypothesis that differently mutated or covalently modified SOD1 proteins may form distinct prion-like assemblies with different toxic effects. This idea is supported by the yet-unexplained varying levels of neurodegeneration and disease characteristics amongst SOD1 mutants.[26] For example, some mutants of SOD1 have average disease durations of less than one year, such as A4V, and others have average disease durations of over ten years, such as H46R.[26] One potential origin of this variation in duration is prion-like propagation of these aggregates between neuronal cells, with different aggregate strains propagating in different ways.[31] As noted above, there is no clear relationship between known quantifiable properties of SOD1 aggregates and disease duration.[26]

In recent years, evidence for the occurrence of prion-like protein aggregation in neurodegenerative diseases, including SOD1 in ALS, has been accumulating.[30–33] A 2016 study by Ayers *et al*. examined mice that carry the G85R mutation of SOD1, which exhibit ALS symptoms that do not present until roughly 20 months of age. This G85R SOD1 was tagged with yellow fluorescent protein (YFP) to visualize SOD1 localization and assembly.[31,32] These newborn mice were injected with spinal homogenates from paralyzed mice that carry the G93A mutation of SOD1, a mutant which causes symptoms to develop much earlier, at six months of age. This injection produced symptoms of neurodegenerative disease within roughly 3 months post-injection.[32] These transgenic mouse studies demonstrate that different mutant aggregates of SOD1 result in different disease characteristics, and these characteristics can be replicated upon injecting ALS symptomless mice.[30–32] The studies of SOD1 IB formation presented herein further examine the hypothesis of mutant dependent strain behaviour in aggregated SOD1.

1.3 Human Cu, Zn Superoxide Dismutase-1

1.3.1 Structure and Function

The structural features and function of native SOD1 provide context for understanding modes of misfolding and aggregation. SOD1 (**Figure 1.5**) is a homodimeric metalloenzyme that helps prevent build-up of the toxic oxygen species superoxide (O_2^-) through its reduction to hydrogen peroxide and oxygen (**Figure 1.6**).[11,10,34] The mature, or holo, form of SOD1 is a homodimer, and contains one intramolecular disulphide bond between residues Cys57 and Cys146, one zinc ion, and one copper ion per monomer.[11,34] SOD1 also contains two long functional loops, the electrostatic loop, L7, and the zinc binding loop, L4. L7 helps guide O_2^- to the active site, which contains a catalytic copper cofactor that is coordinated by residues His46, His48, His63, and His120.[34] The other loop, L4, contains amino acids responsible for the coordination of a zinc structural cofactor, specifically His63, His71, His80, and Asp83.[11,34]

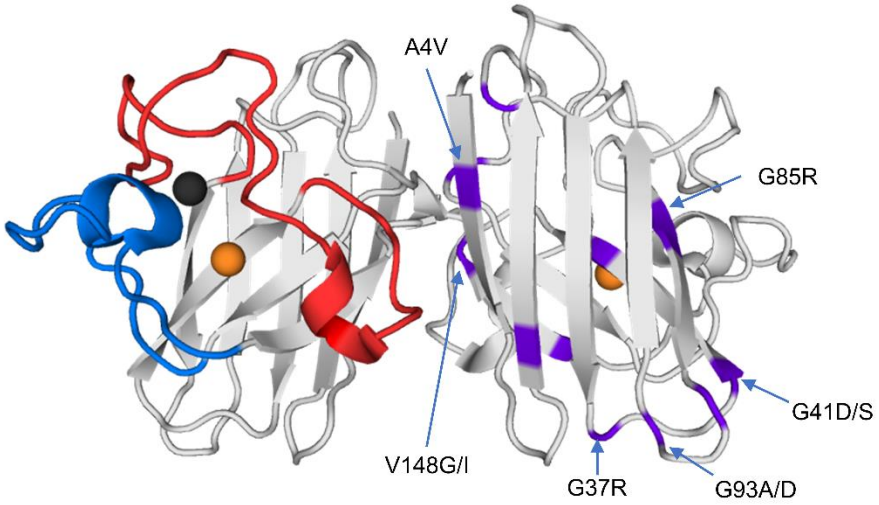


Figure 1.5: Ribbon Structure of Human Cu, Zn Superoxide Dismutase 1.
This SOD1 ribbon has its copper and zinc cofactors illustrated in orange and black, respectively. Loop L4, containing zinc binding residues His63, His71, His80, and Asp83, is shown in red, and loop L7, containing copper coordinating residues His46, His48, His63, and His120, is highlighted in blue. Sites of mutations of potential interest in this study are highlighted on the right monomer in purple. Figure created in Pymol. PDB: 1HL5

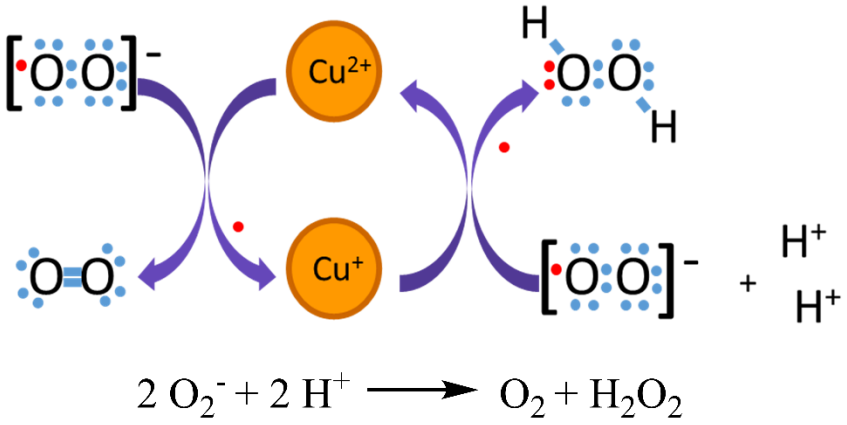


Figure 1.6: SOD1 Catalytic Cycle.
SOD1 utilizes a 'Ping-Pong' type reaction to catalyze the reduction of superoxide into hydrogen peroxide and molecular oxygen to reduce oxidative stress in cells. At left, Cu(II) is reduced to Cu(I) by superoxide, and then at right oxidized to Cu(II) by a second equivalent of superoxide. In the first step, Cu(II) is reduced by a superoxide molecule, generating dioxygen (O_2) and Cu(I). The Cu(I) is then re-oxidized by a second equivalent of superoxide, resulting in Cu(II) and hydrogen peroxide.[35] The enzyme can then undergo repeated catalytic cycles. Figure provided by Dalia Naser.

1.3.2 SOD1 Maturation

SOD1 undergoes maturation to reach its mature holo form, and extensive studies show that immature forms of SOD1 have increased propensity to aggregate.[10,11] The most immature form of this protein, reduced apo (E,E-SOD1^{SH}) has no bound metals, has not yet formed its internal disulfide bond, and is predominantly monomeric.[34] This protein matures through several maturation steps, as shown in **Figure 1.7**.[36] A structurally stabilizing Zn ion is added through a currently unknown mechanism to form a singly metallated species (E,Zn-SOD1^{SH}).[34] Additional modifications include the introduction of a copper ion, which is responsible for redox reaction catalysis, by the copper chaperone protein for superoxide dismutase 1 (CCS).[34,37] The internal disulfide bond between Cys57 and Cys146 is formed oxidatively, catalyzed by CCS during its interaction with SOD1.[34,37] The final, mature state of SOD1 is denoted as Cu,Zn-SOD1^{SS}.

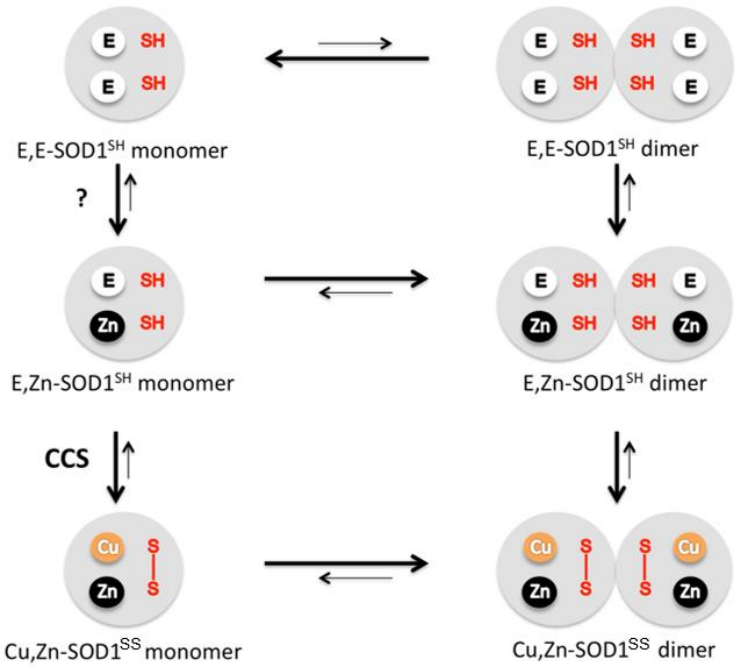


Figure 1.7: Maturation Scheme of SOD1.
E,E-SOD1^{SH}, which is predominantly monomeric[38], obtains a stabilizing zinc ion cofactor to form E,Zn-SOD1^{SH}. This form then interacts with CCS to acquire a catalytic copper ion and to oxidize the disulfide bond between Cys57 and Cys146. This forms the mature protein, Cu,Zn-SOD1^{SS}. Figure provided by Harmeen Deol.[36]

As SOD1 matures it becomes increasingly thermodynamically and kinetically stable.[10] Differential scanning calorimetry (DSC) experiments have been performed to quantify the global stability of SOD1 for different maturation states.[10,36] Thermal stability is measured by increasing the temperature within the instrument and allowing the protein to unfold, with a higher melting point (T_m) indicating a more thermally stable structure.[36] DSC studies on SOD1 show an increase in T_m as the protein matures. Wild type E,E-SOD1^{SH} displays the lowest thermal stability, with a T_m of 47.6°C, while the most mature form of the protein, Cu,Zn-SOD1^{SS}, has a much higher T_m of 92.7°C.[39,40] Recent DSC studies by Harmeen Deol of the Meiering group have shown the monomeric E,Zn-SOD1^{SH} to have a T_m of about 57°C, indicating a more thermally stable structure than the metal-free variant, but less thermally stable than the mature form.[36]

Extensive studies have shown the immature E,E-SOD1^{SH} is particularly prone to aggregation.[10] Notably, under certain conditions, small amounts of aggregated E,E-SOD1^{SH} can seed the aggregation of native-like stable structures of SOD1.[10,41,42] This form of the protein has displays a tendency to aggregate, with different mutants forming aggregates of different sizes.[10] These findings suggest the possibility that partially mature forms of SOD1, notably zinc bound variants, might aggregate, given that they are not as thermodynamically stable as the mature form.[37] This is supported by the observation that mature SOD1 can form aggregates upon loss of metals, which is promoted through mutation or impaired maturation.[43] Investigating the extent of formation of these aggregates will prove valuable for studying both SOD1 maturation as well as the fundamentals of protein aggregation.

1.4 Nuclear Magnetic Resonance

1.4.1 NMR Background

Nuclear magnetic resonance (NMR) is the primary method of analysis for this project. NMR spectroscopy is used to analyze the structures of a variety of biomolecules, notably proteins.[44,45] NMR observes the nuclei of atoms with a non-zero nuclear spin, meaning that they must have an odd atomic mass number.[44,45] The most commonly studied biologically relevant nuclei observed through NMR are hydrogen, carbon, and nitrogen.[44,45] Carbon and nitrogen are both not naturally abundant in their ^{13}C and ^{15}N isotopes, respectively.[44,45] Therefore, to facilitate high resolution NMR spectroscopy on a protein, it can be isotopically labelled with ^{13}C or ^{15}N during cell growth and protein expression.[44,45] This will allow for the protein to have its ^{14}N and ^{12}C atoms uniformly replaced with ^{15}N and ^{13}C atoms, respectively.[44,45]

The nuclei observed will resonate at a specific frequency within a magnetic field, and this frequency is dependent on the strength of the magnetic field as well as the local environment of the nucleus.[44,45] By applying a uniform magnetic field to a molecule, multiple different signals will be observed based on the environment of the nucleus.[44,45] These signals, called chemical shifts, can then be used to gain insight on the structure of a complex molecule, such as a protein.[44,45] The common NMR experiment which will be used to analyze samples of our protein of interest, SOD1, is the ^{1}H-^{15}N heteronuclear single quantum correlation (HSQC) experiment. An HSQC is a two-dimensional (2D) experiment that produces a 2D spectrum with a visible cross-peak for each detected proton-bound nitrogen atom.[44,45] Consequently, a peak will be present for each amino acid amide backbone, except for that of proline, as well as –NH containing side chains in the protein.[44,45]

1.4.2 Quenched Hydrogen-Deuterium Amide Exchange

A powerful high-resolution technique that can be used for IB structural analysis is quenched hydrogen-deuterium amide exchange (qHDX). qHDX is an NMR approach that can be used to measure the high-resolution structure of insoluble aggregates that are a size unfit for traditional aqueous NMR experiments. A key early qHDX study by the Roder group in 1995[46] helped pave the way for future studies in *in-vitro* protein aggregation. qHDX was developed for the purposes of peptide conformational studies.[46] Zhang and coworkers dissolved the small peptide melittin in D_2O and measured the extent of hydrogen-deuterium amide exchange in the aggregate through unfolding in DMSO and measurement using NMR.[46] This resulted in the observation of both solvent-exposed and solvent-protected amino acid residues of the peptide.[46] Interestingly, the same methodology has been further developed and applied to aggregates of purified protein.[47,48]

In these studies, amyloid fibrils formed by various purified proteins were subjected to qHDX, providing structural characterization at individual amino acid residue level.[49,50]

The sequence of steps in a typical qHDX NMR experiment as applied to an aggregate structure is shown in **Figure 1.8**.[49–51] There are many benefits to the use of quenched HDX, including the lower material costs on account of the high expression of IBs, the time saved and lack of loss of sample due to no purification, and it can provide accurate depictions of aggregates in the cell. Traditionally, structural characterization of large aggregates through solution NMR techniques or X-ray crystallography is met with significant difficulties due to the size limitations of each method, as protein aggregates are too large for observation using these tools.[50] However, by unfolding the protein aggregate after it has undergone amide exchange, qHDX reveals structural data about the large protein aggregate while simultaneously reducing the structure to a size suitable to traditional NMR experiments.

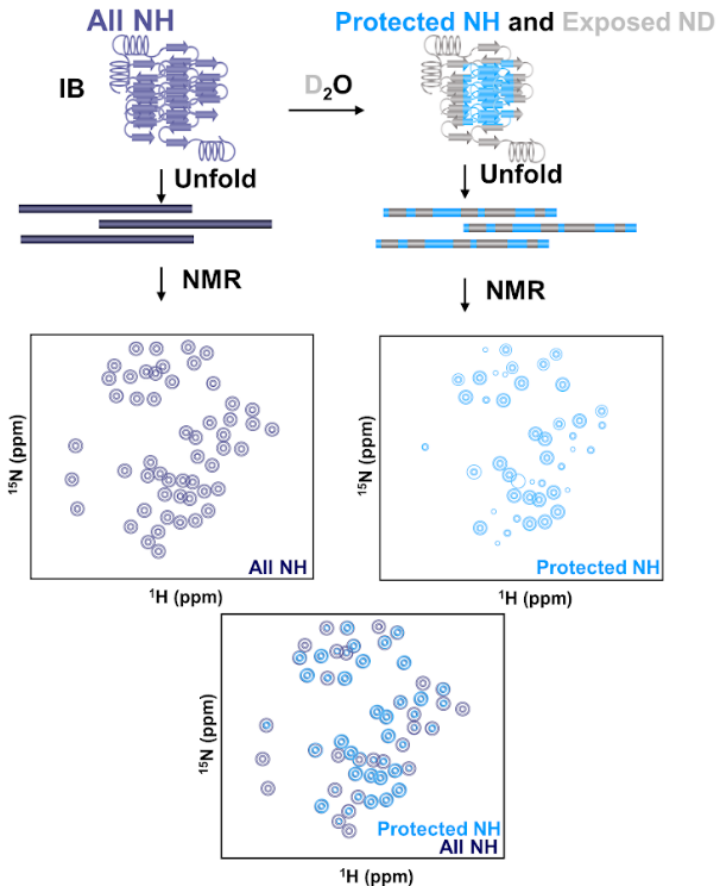


Figure 1.8: Quenched hydrogen-deuterium amide exchange on protein aggregates.
The solvent exposed amino acid amide protons are shown to exchange with deuterons when exposed to D_2O, resulting in a loss of signal when observed through a ^{1}H-^{15}N HSQC NMR experiment. When this exchanged sample is compared to a spectrum of unfolded protein that has not undergone qHDX, differences in signal intensities are observed. Signal intensities for amino acid residues that more closely resemble the fully protonated spectrum are considered to be more highly protected from the solvent, and residue signals that are considerably lower in intensity than the fully protonated spectrum are considered to be more solvent exposed.[46,49,50]

qHDX has been applied previously as a method to characterize IB structures [51]. Lei Wang *et al.* have used this method to display evidence that IBs, which were previously thought to be amorphous aggregates, contained organized cross-β amyloid structures [51]. Additional investigations, including solid-state NMR spectroscopy (ssNMR), alongside low-resolution methods that include Fourier transform infrared (FTIR) spectroscopy, circular dichroism (CD), and X-ray diffraction have also suggested amyloid properties within IBs [17,52]. Surprisingly, after these investigations, no further studies were reported after the initial intriguing results. These

results showed that different proteins exhibited very different protection in IBs ranging from just a short stretch of a loop, to most of the protein sequence [51].

As we discovered, obtaining reproducible results requires careful control of many experimental aspects, described extensively in Chapter 2. In this thesis, qHDX is optimized to further probe the high-resolution structures of IBs, and applied to investigate the effects of mutations, protein expression conditions, and post-translational modifications on the structure of IBs.

1.5 Research Objectives

The focus of this thesis is on the optimization of qHDX NMR to determine the high-resolution structure of IBs. SOD1 is used as the model system for optimizing the methodology. Through careful control experiments, the development of qHDX as a robust method that is quick, efficient, and cost-effective to measure, with accuracy, the structure of *in-vivo* aggregates was achieved. Consistency in results between both technical and biological replicates is key, as any observed differences must be a result of variability in IB structure. To achieve this reproducibility, careful control of experimental parameters is required, and the requisite controls are detailed in Chapter 2.

Using the optimized method, the effects of point mutation, expression temperature, expression time, and post-translational modifications on IB formation and structure were investigated using SOD1. Three different expression conditions will be investigated in this research: reduced apo (rApo) SOD1 expressed at 37 °C for 4 hours, rApo SOD1 expressed at 25 °C overnight (roughly 18 hours), and Zn-grown SOD1 expressed at 37 °C for 4 hours. The structures of these aggregates at each condition are reported in Chapter 3.

Chapter 2

Inclusion Body Quenched Hydrogen-Deuterium Amide Exchange Method Development

This chapter has been submitted for publication to *Analytical Biochemistry* and is currently under review.

Citation:

*__Tarasca, M. V__, *Naser, D., Schaefer, A., Soule, T. G., and Meiering, E. M. High-resolution structure of protein inclusion body aggregates by quenched hydrogen-deuterium amide exchange. *Anal. Biochem. (Under Review)* 2021

* Indicates these authors contributed equally to the work.

2.1 Quenched HDX Workflow

The workflow of hydrogen-deuterium amide exchange (qHDX) experiments on inclusion bodies (IBs) is summarized in **Figure 2.1**. The experiments consist of 3 stages: sample preparation (green) followed by NMR experiments (purple) and data analysis (orange). These stages are outlined here and described in detail in the following sections.

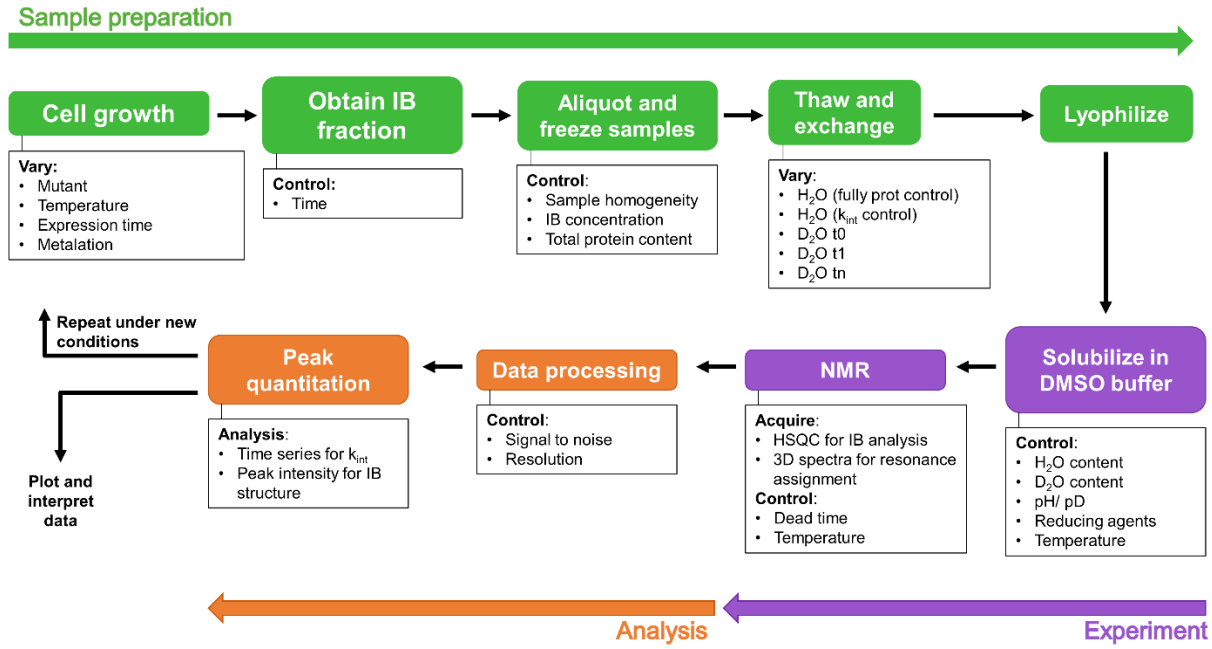


Figure 2.1: Flowchart of the IB quenched HDX method.
The workflow can be divided into three distinct stages: sample preparation (green), an experimental stage (purple), and data analysis (orange). The sample preparation requires very careful control of sample conditions, so as to avoid unintended variations in measured structures. With this control in mind, growth conditions can be varied and their effects on the structure of IBs can be observed by NMR. The experimental stage involves the acquisition of NMR spectra, with ^{1}H-^{15}N 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 crosspeak signal intensities of ^{1}H-^{15}N HSQC spectra to provide high-resolution structural information of the aggregate.

In the sample preparation stage, there are multiple experimental parameters that may be varied. Cell culture conditions can be varied to determine their effects on the structure of the IBs. Conditions of interest may include, for example, expressing different mutant forms of the protein

of interest, growth temperature, time of protein expression, and addition of protein cofactors such as metals. 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 IB preparation is initially resuspended in either H_2O or D_2O. Resuspension in H_2O 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 (see 2.3.1). IBs are resuspended in D_2O for qHDX experiments. Varying the time of resuspension in D_2O allows different extents of exchange to occur. This measurable rate is known as the observed exchange rate in the aggregate (k_{ex}) [50] (see 2.3.5). The exchange is stopped, or quenched, by flash freezing the sample in liquid nitrogen, lyophilized, then solubilized in DMSO buffer for NMR experiments. To minimize changes to the IB structure as well as signal scrambling, lyophilized samples are stored at -80 °C for a limited time (see 2.3.6).

In the second stage, NMR experiments include acquiring 2D ^{1}H-^{15}N HSQC experiments for IB structural analysis (see 2.4.2), and 3D experiments to obtain sequence specific resonance assignments (see 2.4.1). The experiments described here are based on 10 minutes dead time from dissolving a qHDX sample and a 20 min ^{1}H-^{15}N HSQC spectrum acquisition time. The third stage of the workflow is the data analysis. Acquired NMR data is 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 (see 2.2.9, 2.3.1, 2.4.2). The entire workflow may be repeated to examine

new experimental conditions. We illustrate all the steps in the stages and representative results in the sections below.

2.2 Materials and Methods

2.2.1 Materials

Solutions were prepared using Milli-Q ultrapure water and chemicals were analytical grade unless otherwise stated. TEN buffer, containing 20 mM tris(hydroxymethyl)aminomethane (Tris), 1 mM ethylenediaminetetraacetic acid (EDTA), and 100 mM NaCl at pH 8.1, was stored at 4 °C. Bottles of D_2O were sealed with Parafilm and stored at 4 °C. Using 10X M9 minimal media stocks, consisting of 500 mM Na_2HPO_4, 250 mM KH_2PO_4, and 5000 mg L^{-1} NaCl, 1 L of 1X media was prepared and incubated overnight at 37 °C with shaking at 200 RPM. At the time of inoculation, 0.4% Glucose, 100 μM $CaCl_2$, 2 mM $MgSO_4$, 500 mg L^{-1} ^{15}N-NH_4Cl (or ^{14}N-NH_4Cl for non-isotopically labelled samples), and 0.1% Thiamine-HCl, were added to the medium together with 100 μg mL^{-1} ampicillin and 30 μg mL^{-1} chloramphenicol. Ultrapure 99.9% DMSO-d_6 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-d_6 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.

2.2.2 Cellular Growth and Protein Expression

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 4000*g* 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 4000*g* at 4 °C for 20 min, 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 2.2.4).

In addition to SOD1, we also used qHDX to analyze mutants of Adnectins, a family of engineered proteins derived from the human 10^{th} fibronectin type 3 domain [53]. All qHDX experimental work and data analysis on Adnectins were performed by Anna Schaefer in our group. Adnectin variants were expressed in *E. coli* BL21(DE3) pLysS cells containing a pET-9d vector [24]. Cells were grown as described for SOD1 with the following changes: selection was achieved using 50 μg mL^{-1} Kanamycin and 34 μg mL^{-1} chloramphenicol, expression was induced using 1 mM IPTG and cells were grown at 37 °C following induction.

2.2.3 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. The soluble and insoluble fractions were separated by centrifugation at

20,000*g* 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.

2.2.4 Preparation of NMR Samples

IB sample aliquots were thawed at room temperature, centrifuged at 21,000*g* at 4 °C for 20 minutes and the supernatant was removed. The pellet was resuspended in 5 mL of either H_2O for a protonated control experiment or an intrinsic exchange rate (k_{int}) control experiment (see 2.3.1), or D_2O for a qHDX experiment. The time of incubation in D_2O 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 (see 2.3.6).

To ensure the dead time prior to NMR data acquisition is minimized for a qHDX experiment (see 2.3.1), 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-d_6, 30 µL of D_2O, and 25 mM DTT (from 1 M DTT in DMSO-d_6). The pD of the test sample (pD = pH_{read} + 0.4; see 2.3.2) was then adjusted to 5.5 using 5% (v/v) DCA in DMSO-d_6. A glass pH

probe with a ceramic reference junction was used for measurements to avoid unwanted reactions with DMSO. 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-d_6. This test sample was further used to set up NMR experimental conditions (see 2.2.6). A premixed 600 µL solution for the qHDX sample was prepared containing DMSO-d_6, 30 µL of D_2O, 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.

2.2.5 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 H_2O (see 2.3.1). A fully protonated control experiment was used as a reference for the amount of exchange in D_2O samples; this control contains H_2O 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 (see 2.3.4).

2.2.6 Quenched HDX and k_{int} Control NMR Experiments

The NMR spectrometer setup was determined using the unlabelled test sample (2.2.4). 500 µL of the test sample was slowly pipetted into an NMR tube (to avoid bubbles) and inserted into the spectrometer. For the test sample, the spectrometer reference signal was locked to DMSO-d_6, the shims were adjusted, and spectral parameters were determined. A ^{1}H 1D NMR experiment was performed, and the test sample was removed. The parameters determined were copied to experiment files for the qHDX sample.

The lyophilized qHDX sample was dissolved (~1 min of mixing by micropipette) in the premixed solution (2.2.4) and a timer was started. Next, the sample was centrifuged for 10 s, and 500 µL of the supernatant was slowly pipetted into an NMR tube. The qHDX sample was inserted into the spectrometer, reference signal locked to DMSO-d_6, 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 ^{1}H 1D NMR experiment was then acquired for the qHDX sample.

Once the ten-minute mark on the timer was reached, a series of nine 20 min ^{1}H-^{15}N heteronuclear single quantum correlation (HSQC) experiments were initiated. After acquiring these spectra, the shim files were saved for acquiring the final three spectra after the sample reached H-D exchange equilibrium (see 2.3.1). The qHDX sample was removed from the spectrometer and stored at room temperature for one week, after which three additional ^{1}H-^{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 (see 2.2.9 and 2.3.1), and the first ^{1}H-^{15}N HSQC spectrum was used to measure the relative IB protection (**Figure 2.2**) (see 2.2.9 and 2.4.2).

2.2.7 Fully Protonated Control NMR Experiments

Fully protonated control samples contain 5% (v/v) H_2O and could be analyzed directly (without the use of a test sample). These control samples were prepared as described in 2.2.4, with H_2O used in the place of D_2O. A ^{1}H 1D NMR experiment followed by a ^{1}H-^{15}N HSQC experiment were performed. The ^{1}H 1D NMR spectrum is acquired for normalization for sample concentration (3.2.2). This protonated ^{1}H-^{15}N HSQC spectrum is used alongside the qHDX ^{1}H-^{15}N HSQC spectrum to determine the relative protection of the aggregate (**Figure 2.2**) (see 2.2.9 and 2.4.2).

2.2.8 Sequence Specific Resonance Assignments

To obtain NMR resonance assignments, protein was grown and expressed as described (see 2.2.2), using 0.2% ^{13}C glucose and 0.5 mg L^{-1} ^{15}N-NH_4Cl to make a double-labelled sample. Cell culture was grown at 37 °C for 6 hours post induction and prepared as in 2.2.3 and 2.2.4. 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 [54–59]. The assignments for SOD1 unfolded in DMSO were determined by Dalia Naser in our group. Spectra were acquired based on standard Bruker parameter sets.

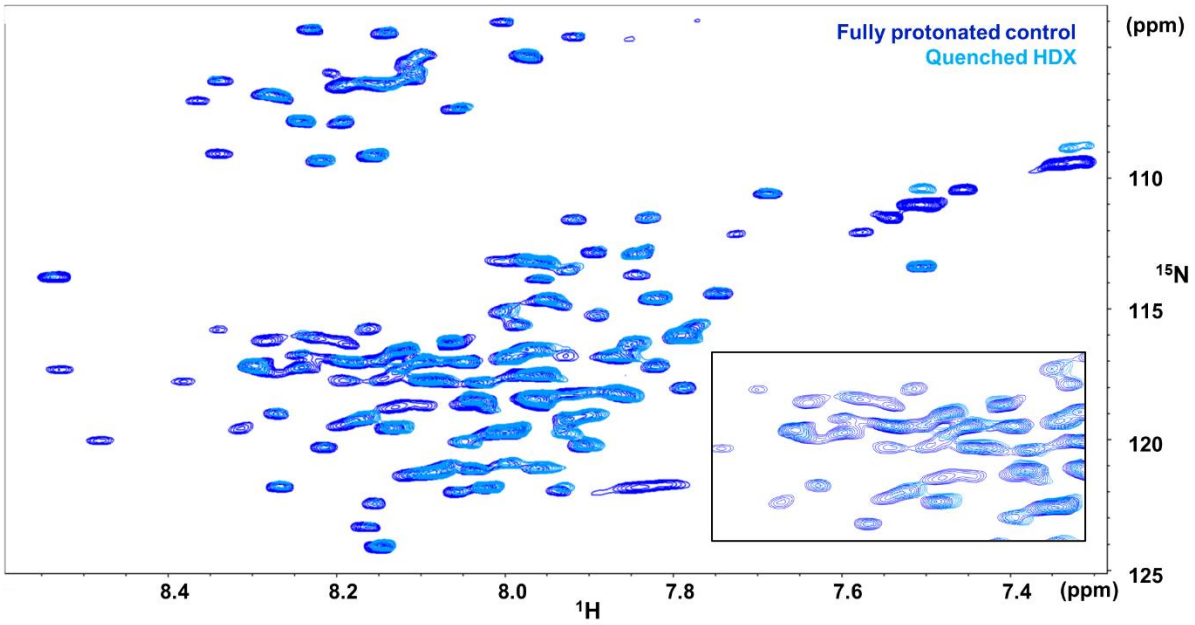


Figure 2.2: Different ^{1}H-^{15}N HSQC spectra acquired for quenched HDX experiments.
The signal intensities of each spectrum visibly decrease as the sample exchanges with D_2O. Shown in dark blue is a fully protonated control experiment. All crosspeak intensities are at their maximum. Shown in light blue is a quenched HDX experiment. A ratio of the signal intensities measured in this spectrum and the fully protonated control experiment is taken, providing a measurement of the relative protection of the aggregate. Relative to the fully protonated control spectrum, there are peaks missing due to H-D exchange. Highlighted is an area within the pair of spectra that possess both highly protected residues and solvent exposed residues, illustrated by peaks with higher and lower signals in the quenched HDX spectra, respectively.

2.2.9 Data Analysis

All ^{1}H-^{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 software [60], 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. While k_{int} can be determined for each qHDX sample signal decay for the 12 acquired ^{1}H-^{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 ^{1}H-^{15}N HSQC (see 2.3.1).

The first spectrum the decay series corresponds to an exchange time in 95% DMSO / 5% D_2O of 20 min (10 min dead time plus 10 min corresponding to half of the time to acquire the ^{1}H-^{15}N HSQC spectrum) (see 2.3.1). 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 2.4.2.

2.3 Control of IB Sample Conditions

The sample conditions throughout qHDX experiments must be carefully controlled. Reasons for this strict control include ensuring that the signal observed by NMR is representative

of the IB structure, preventing unintended changes to the structure of the IBs, and ensuring the samples are as reproducible as possible, so that peak positions and k_{int} do not change from sample to sample. The data acquired from a ^{1}H-^{15}N HSQC are typically mapped onto a protein structure to visualize protected or solvent-exposed amino acid residues [46,50]. However, for the signal intensities observed at the time of the first NMR experiment to be representative of the sample at the time of dissolution, the intrinsic amide exchange rate of protein unfolded in DMSO (k_{int}) must be sufficiently low (see 2.3.1).

Throughout cell culture growth, protein expression, and sample preparation it is important that uncontrolled factors minimally impact the structure of the IB. Examples of such factors include the time and temperature of protein expression and sample preparation (see 2.3.4), and vigour of resuspending the IB sample in buffer after cell lysis. In addition, the chemical shifts of each amino acid residue must be consistent between spectra to allow for comparisons between samples. Because the protein is unfolded during protection readout, the chemical shifts of a given amino acid residue between replicates of a given mutant will be nearly identical, with the variations being dependent on sample conditions. These conditions include pH, water content, and temperature (see 2.3.2, 2.3.3, 2.3.4). Provided conditions are maintained between different samples, we find both chemical shifts and spectral intensities are directly comparable between mutants. The chemical shifts of residues are unchanged except near the site of mutation; here, clear sequence dependent changes in chemical shift allow for ready transfer of assignments between mutants (see 2.4.1).

2.3.1 Controlling the 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% D_2O, the exposed amides start exchanging with deuterons according to their k_{int} values. This forward exchange, while much slower than in aqueous conditions [46,50], can lead to significant signal loss in the dead time preceding NMR data acquisition. The addition of D_2O is necessary, as without it the unfolded protein will undergo poorly controlled back exchange with H_2O (see 2.3.3) [49,50]. 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 9.0×10^{-3} min^{-1} or lower [50]. For k_{int} of 9.0×10^{-3} min^{-1}, the half-life for exchange is ~77 min; therefore, with a deadtime of 10 min and ^{1}H-^{15}N HSQC acquisition time of 20 min, signal intensity at the midpoint of sample acquisition will decrease to ~85% of the original intensity [50]. 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 [46,49,50].

Values of k_{int} can be measured by acquiring a series of ^{1}H-^{15}N HSQC spectra. Here, a sample suspended in H_2O is lyophilized and then dissolved in 95% DMSO buffer with 5% D_2O (**Appendix Figure A2.1**). The fully protonated amide groups then exchange with D_2O, and the extent of exchange is monitored by measuring ^{1}H-^{15}N HSQC spectra as a function of time. Note that the difference in signal between the first ^{1}H-^{15}N HSQC and a fully protonated sample in 95% DMSO / 5% H_2O corresponds to signal lost in the dead time, and residues that show very low or no signal in the first spectrum cannot be assessed by qHDX as they exchange too fast to measure.

Using five to ten half-lives of signal decay to define k_{int}, we have found that 7 days is sufficient for our slowest residues to equilibrate **(Figure A2.1**). We recommend acquiring ~12 spectra throughout the exchange, including three at the time of equilibrium, to accurately determine k_{int} values. The crosspeak 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. Residues with a k_{int} value smaller than the recommended value of 9.0×10^{-3} min^{-1} decay sufficiently slowly that their signal intensity is a reasonable measure without a correction of the signal at time of dissolution [50]. For samples with limited signal to noise, and hence increased experimental uncertainties and apparent variations in measurements of k_{int} for separate samples, using an average over multiple samples of k_{int} values can decrease experimental variability in the results. Also, k_{int} values can be used to calculate the signal at the time of dissolution (correcting for exchange in the dead time), as we did here for SOD1 using the averages of 18 values – two biological replicates of nine mutants (see 2.2.9) using Eq. 1:

Corrected signal intensity

$$= A * e^{(-k_{int\ av} * t)} + C \qquad \textbf{Eq. 1}$$

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 across 18 samples, t is zero for the back-calculation, and C is the offset value of the fit.

Residues with larger values of k_{int} decay too rapidly for quantitation without correction for forward exchange [50]. The results described here do not include corrections for forward exchange in the dead time, which can be neglected in some conditions, including if the fraction of residual H_2O in the sample (χ_H) is less than 0.3 (see 2.3.3) [49,50] and if k_{int} is less than 9.0×10^{-3} min^{-1}. If

these conditions cannot be met, then corrections for exchange can be made by measuring k_{int} and correcting for signal loss in the experimental dead time, as described by Hoshino *et al.*, 2007 [50], or by taking a ratio of the reference spectrum to the exchanged spectrum [50]. In situations where requisite conditions are satisfied, the measured ^{1}H-^{15}N HSQC signal intensities for each amino acid residue in the protein can be used directly to assess IB structure.

k_{int} values obtained for IBs SOD1 and Adnectins (measured by Anna Schaefer, included with permission) were compared with values obtained previously for *in vitro* fibrils of pure prion protein (PrP) under similar solution conditions of DMSO, pH, and temperature [61] (**Figure 2.3**). Overall, the k_{int} values are similar between the three 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 values for His standout as particularly high for SOD1, with average k_{int} for His in SOD1 of 4.24×10^{-2} min^{-1}, over four times higher than the recommended threshold value of 9.0×10^{-3} min^{-1} for k_{int}. The high k_{int} values for His residues preclude quantitative analysis of their protection in SOD1 IBs. 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. Accordingly, using average values of k_{int} can be beneficial to decrease experimental scatter in the protection analysis (see 2.4.2).

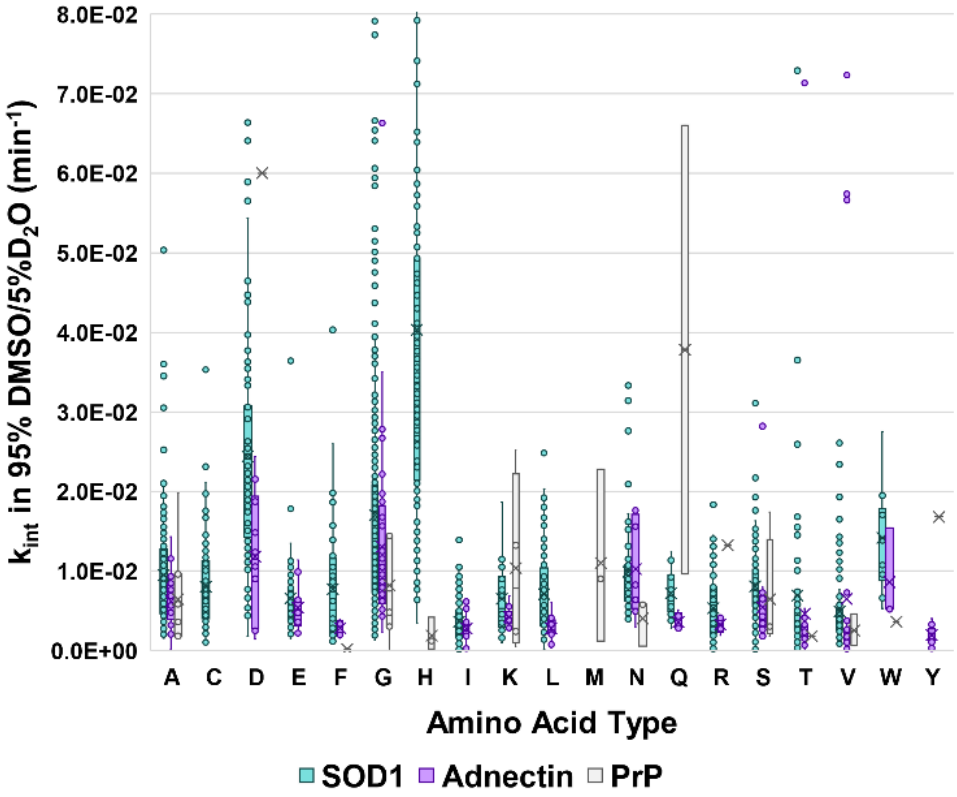


Figure 2.3: Measured intrinsic rate constants of amide exchange for different amino acids. Values are plotted for SOD1 (teal) and Adnectin (purple) IB samples together with literature values for pure PrP aggregates (light grey) [61]. Values include results for 18 SOD1 samples (two biological replicates for each of nine different mutants), six Adnectin samples and a single mutant (P101L) for PrP [61]. 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. 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.

2.3.2 Sample pH

The pH of the sample influences both the values of k_{int} (see 2.3.1) and the observed chemical shifts (**Figure A2.2**). As the sample in qHDX experiments contain D_2O rather than H_2O the acidity of the sample is measured as pD (pD = pH_{read} + 0.4) [46]. Previous results from the Roder group showed that the k_{int} of amide protons in 95% DMSO / 5% D_2O is at a minimum at a pD of ~5.5 [46]. The pKa of dichloroacetic acid (DCA) in 95% DMSO / 5% D_2O is 5.72, which makes it suitable for buffering at the target pD of 5.5 in order to minimize k_{int} [46]. Trifluoroacetic acid has

also been used previously for some similar experiments [51,62]; however, its pKa in 95% DMSO is 3.45 [63] which does not result in appropriate buffering at the target pD of 5.5, therefore it is not used in the qHDX experiments reported here. Nevertheless, it can be considered if DCA does not sufficiently solubilize the sample [62]. After three endpoint ^{1}H-^{15}N HSQC spectra are obtained at equilibrium (2.3.1), 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, such as those made of standard epoxy. The sample is still buffered after this week-long period and should provide an accurate pD reading. The pD is measured after the final spectra are acquired, as opposed to during the sample preparation stage, to prevent the addition of salt and water through the reference junction of the pH probe (see 2.3.3 on minimizing H_2O content).

The pD of the sample can also be assessed by using DCA as a convenient internal reporter. A ^{1}H 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 2.4A**). This can be used to assess the approximate pD of the sample before acquiring the ^{1}H-^{15}N HSQC series (**Figure 2.4BC**): if the difference in chemical shift between the DCA peak and the most upfield methyl peak is greater than ~5.6 ppm it can be concluded that pD of the sample is higher than pD 5.5 and will consequently result in a high k_{int}. Additional benefits of using the DCA reporter peak are that there are no complications arising from the addition of water through the junction of a pH probe, and the dead time between sample dissolution and NMR data acquisition is kept at a minimum. By ensuring appropriate sample pD, the k_{int} will be minimized, therefore the

observed signal through NMR will be representative of the sample at the time of dissolution in DMSO.

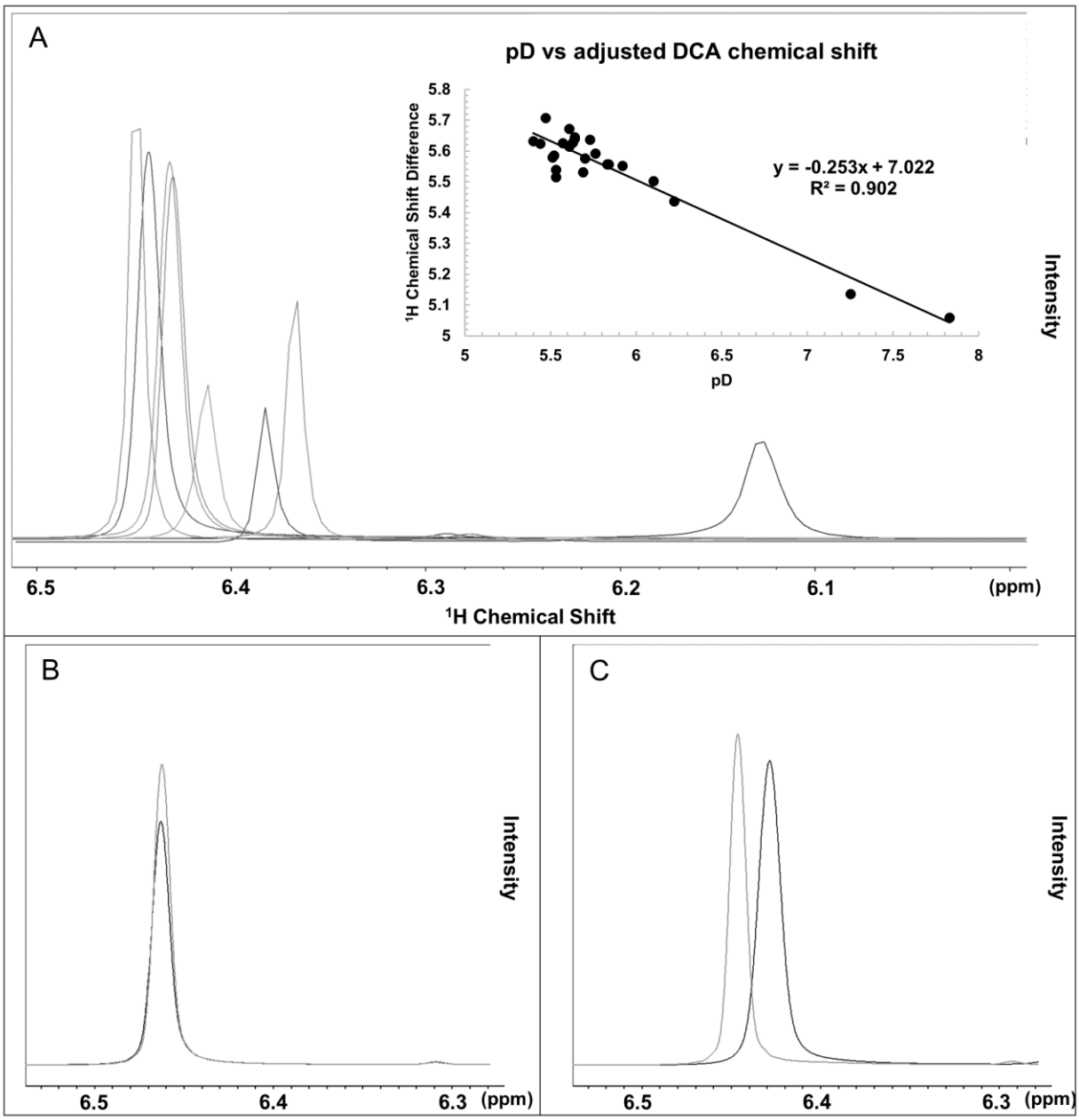


Figure 2.4: DCA methyl ^{1}H peak as an internal indicator of sample pH.
A) ^{1}H 1D NMR spectral overlay between qHDX samples at different pD values, zoomed in to encompass the peak corresponding to the DCA addition. Its adjusted chemical shift, being the difference in chemical shift between the DCA peak and the most upfield peak in the methyl region, is plotted against measured pD and a linear relationship is observed. 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.

2.3.3 Sample H_2O Fraction, χ_H

The fraction of residual water (χ_H = [H_2O] / {[H_2O] + [D_2O]}) 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 exchange of sample amide deuterons with water protons or deuterons (see 2.3.1). As illustrated in **Figure 2.5**, there are clear differences in chemical shift between two spectra of samples that differ only in their water fraction.

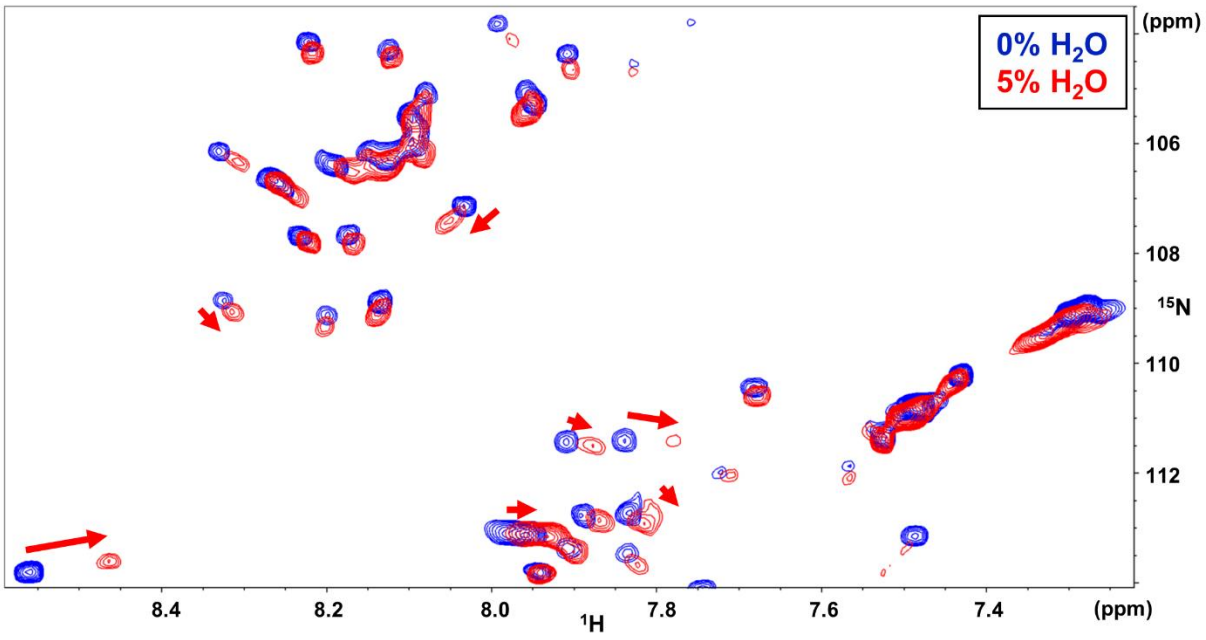


Figure 2.5: ^{1}H-^{15}N HSQC spectral overlay of a quenched HDX sample in DMSO with 0% (v/v) water added (blue) and 5% water added (red).
Differences in chemical shifts are observed, therefore water content must be consistent between samples for the purposes of resonance transfer.

Back exchange will occur if χ_H is larger than the fraction protected. As back exchange occurs, signal intensities will increase in acquired ^{1}H-^{15}N HSQC spectra. This could result in the misinterpretation of a solvent-exposed amide with a solvent-protected amide. 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 (2.2.4) 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 (χ_H = 0.300/{5.0 + 0.300}) [49]. 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} (see 2.3.1) and the addition of D_2O, will result in negligible back and forward exchange, providing high quality data without the need for additional corrections [49,50].

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 and helps to prevent continued exchange of the IB with residual D_2O upon removal of the sample from the lyophilizer. To further minimize contaminating H_2O in the sample one may use separate 1 mL DMSO-d_6 ampules for each sample. DMSO is highly hygroscopic and will absorb H_2O from the atmosphere, therefore it is necessary to minimize its exposure to air [64]. 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 (see 2.2.4). Experiments on non-isotopically labelled samples have shown that water added through the junction of a pH probe can readily ruin 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.

A rough measure of the residual H_2O 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 A2.3**). We have found that in samples with low

residual water, the largest peak in the spectrum is the solvent peak (DMSO, 2.511 ppm), followed by the height of the methyl peak (0.80 ppm), which is representative of protein concentration. 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 ^{1}H-^{15}N HSQC spectrum and inconsistent sample signal compared to replicates).

2.3.4 Temperature

Another factor that affects k_{int} is the temperature of the sample [46,50]. Since k_{int} decreases with temperature, a lower k_{int} will decrease forward exchange and allow for more accurate quantitation of qHDX. Therefore, 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 [46,50]. As the temperature is lowered to the freezing point of the DMSO, the molecular tumbling is slowed, broadening lines and decreasing peak intensity and resolution, while also reducing the rate of amide exchange. We recommend a temperature for NMR data acquisition of ~19 °C, the melting point of DMSO. We find that the NMR samples do not freeze at 19 °C, while providing spectra with a reasonable signal to noise and resolution (**Figure A2.4**). An additional recommendation 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. These combined factors contribute to the recommended NMR acquisition temperature of 19 °C.

2.3.5 Time of IB Exchange in D_2O

As the IB is suspended in D_2O its backbone amide protons are exchanging with deuterons (**Figure 1.8**). This exchange 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. From a single D_2O incubation timepoint, an extent of protection can be measured. This provides a measurement of the extent of H-D exchange of the aggregate during the D_2O at one timepoint. The extent of protection is defined as the ratio between the signal intensities of a qHDX ^{1}H-^{15}N HSQC spectrum and a fully protonated spectrum (**Figure 2.2**).

The time of incubation in D_2O can be varied to observe exchange of solvent protected residues (**Figure 2.1**). 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. As this exchange occurs in an aqueous solution, as opposed to k_{int} which is measured in DMSO, measured k_{ex} values are roughly 100-1000 faster than measured k_{int} values [46,50]. The ratio between this observed exchange rate for each residue, k_{ex}, and the intrinsic exchange rate for each residue if freely exposed, k_{int}, is called the protection factor [50]. The protection factor is defined as $P = k_{int}/k_{ex}$ [50]. The protection factor differs from the previously mentioned extent of protection, as it is a measurement across a series of timepoints of D_2O incubations. Typical protection factor values observed can range from 10^1 to 10^4 or higher [48,50]. Through varying the duration of the D_2O 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.

2.3.6 Storage of Lyophilized Samples

The structure of lyophilized IB samples can change markedly over time, depending on their storage conditions. We have observed that the spectra acquired from replicate samples analyzed immediately after lyophilization, stored at room temperature, and stored at -80 °C all differ from one another (**Figure A2.5**). Proteins, even when lyophilized and stored at -80 °C, have been shown to be susceptible to changes in structure [65]. Another factor to consider is differences in spectra as a result of residual moisture within the sample, either H_2O, D_2O, or HOD (water with one proton and one deuteron), exchanging with the lyophilized IB. Over time, especially at room temperature but also at -80 °C, this results in both the scrambling as well as lowering of observed signal [66]. This scrambling leads to spectra with differences in measured signal intensity, and therefore influences the measured qHDX data. Consequently, when performing structural analysis of these aggregates it is recommended to observe the samples through NMR as soon as possible, as opposed to storing them at either room temperature or -80 °C.

2.4 Spectral Analysis

The spectra obtained from qHDX experiments can be analyzed to determine the relative protection of each amino acid residue in the IB. Provided the value of k_{int} is sufficiently low, a residue can be analyzed without correction (see 2.3.1) [50]. The residues with k_{int} values that are deemed to be too high are not included in analysis, as their signal intensities are no longer representative of the signal at the time of dissolution. Published work by the Goto group includes mathematical corrections that can accommodate for high k_{int} values, however provided residues are below the 9.0×10^{-3} min^{-1} threshold as described in 2.3.1, specific structural information can still be obtained.

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 signal-to-noise (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 signal to noise as well as signal resolution (see 2.2.9).

2.4.1 Resonance Assignments on an Unfolded Protein in DMSO

Sequence specific resonance assignments for unfolded SOD1 were obtained using a combination of standard 3D experiments for uniformly ^{15}N, ^{13}C -labelled protein (**Figure A2.6**). 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 ^{1}H-^{15}N HSQC. We found that Cα and Cβ 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 Cα/Cβ 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 sequence repeats, where Cα/Cβ shifts are identical. These trends do agree with the general behaviour of random coil peptides [67]. **Table 2.1** shows the average random coil chemical shifts for SOD1 unfolded in DMSO.

	C	Cα	Cβ	H	N
ALA	173.156	48.952	17.918	8.093	122.499
ARG	172.009	52.803	28.862	8.059	119.297
ASN	171.461	49.686	37.236	8.215	118.274
ASP	171.271	50.028	36.252	8.184	116.731
CYS	170.411	55.495	26.335	8.030	116.683
GLN	172.942	52.408	27.614	8.050	118.259
GLU	171.905	52.643	27.243	8.042	118.489
GLY	169.851	42.245		8.136	106.457
HIS	170.562	52.168	27.310	8.285	117.346
ILE	171.775	57.823	36.513	7.914	118.628
LEU	172.892	51.805	40.538	7.961	119.398
LYS	172.192	53.001	31.243	8.013	119.161
PHE	171.553	54.604	37.499	8.094	117.589
SER	170.822	55.473	61.813	7.890	113.929
THR	170.634	59.095	66.821	7.912	112.374
TRP	172.471	53.864	27.908	8.090	121.339
VAL	171.589	58.460	30.592	7.861	116.125
PRO		60.016	29.142		

Table 2.1: Random coil chemical shifts of SOD1 in DMSO.
Data shown are for uniformly ^{15}N, ^{13}C labelled A4V SOD1 in 95% DMSO / 5% H_2O, pH 5.50. 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.

A very valuable spectrum for making assignments was the HN(CA)NNH [59], which provides connectivities of amide *i* to the ^{15}N 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 ^{1}H, ^{15}N HSQC and resulting in, albeit limited, dispersion in the nitrogen dimension (**Figure A2.6**). 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 Cα/Cβ 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 2.2**). 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. Note that due to the larger chemical shift effects of Ala, Val, and Ile, which can sometimes extend through multiple adjacent residues, peak assignment transferring of neighbor residues should be done with extreme care if they are mutated in a variant, and sequence effects should be taken greatly into consideration.

	Preceded by A			Preceded by I			Preceded by V		
FAS1-4	Residue	deviation from average	stdev for that res type	Residue	deviation from average	stdev for that res type	Residue	deviation from average	stdev for that res type
	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
Adnectin	A15	2.12	1.33	T37	-1.60	1.35	S4	-0.90	1.79
	T16	-0.73	1.35	S62	-0.64	1.79	V13	-1.86	1.69
	R27	0.74	0.92	T73	-2.77	1.35	A14	-1.36	1.33
	R32	0.85	0.92	S87	-1.03	1.79	Q48	-0.86	1.21
	T60	2.12	1.35	N89	-1.43	2.02	D69	-0.33	1.66
	V77	1.73	1.69	D95	-1.63	1.66	T78	-0.30	1.35
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

Table 2.2: Primary sequence effects on ^{15}N chemical shifts of amino acids.
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 our proteins, Adnectin and

SOD1, as well as Cyt C (BMRB: 26973) [68], Fas1-4 (BMRB: 25425) [69], and GroES (BMRB: 18949) [70]. 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.

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. 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 A2.6**). 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 different samples.

2.4.2 Quantitation of Amide Protection

The signal intensities of peaks in the ^{1}H-^{15}N HSQC spectrum report on the structure of the IB 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 2.2**).

Taking the concentration of the protein of interest in NMR samples into account must be done carefully due to variations in overall sample concentration in the IB. Accordingly, we report the fraction of amide protected in a qHDX sample relative to fully protonated H_2O sample using the following procedure. The smoothed intensity for each residue in the first ^{1}H-^{15}N HSQC spectrum was scaled for protein concentration by dividing by the average amide signal in the sample.

Intensity of Residue *i* amide in D_2O sample

$$= \text{Smoothed Intensity}_{\text{Residue } i} / (\text{AverageIntensity}_{\text{All}})_{\text{residues}} \quad \textbf{Eq. 2}$$

Since the average amide D_2O signal for a sample is a combination of protein concentration and the extent of signal lost to exchange, the extent of exchange must be accounted for or else the fraction protected may be erroneously high. Accordingly, the extent of exchange for the qHDX samples was calculated as the ratio of the intensity of the ^{1}H 1D amide region (7.5-8.5 ppm) of the qHDX sample divided by the intensity of the corresponding H_2O sample (from the same growth), where these ^{1}H 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 was calculated as:

Fraction of Residue *i* amide protected

$$= R \times (\text{Scaled Intensity}_{\text{Residue } i \text{ amide in qHDX sample}} / \text{Average Scaled Intensity}_{\text{Residue } i \text{ amide in H2O}} \quad \textbf{Eq. 3}$$ [61]

2.4.3 Replicate Samples

Protein aggregation is well established to be variable among both in vitro and in vivo replicate experiments [3,13], thus analyses of IB structure by qHDX require averaging the results from repeated experiments. The variability in replicate experiments is illustrated in **Figure 2.6**. Provided identical sample conditions are maintained, there is consistency in the observed scaled signal intensities between biological replicates of SOD1 mutants (**Figure 2.6AB**). We find this trend in both fully protonated and qHDX samples. 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 2.6C**). 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**).

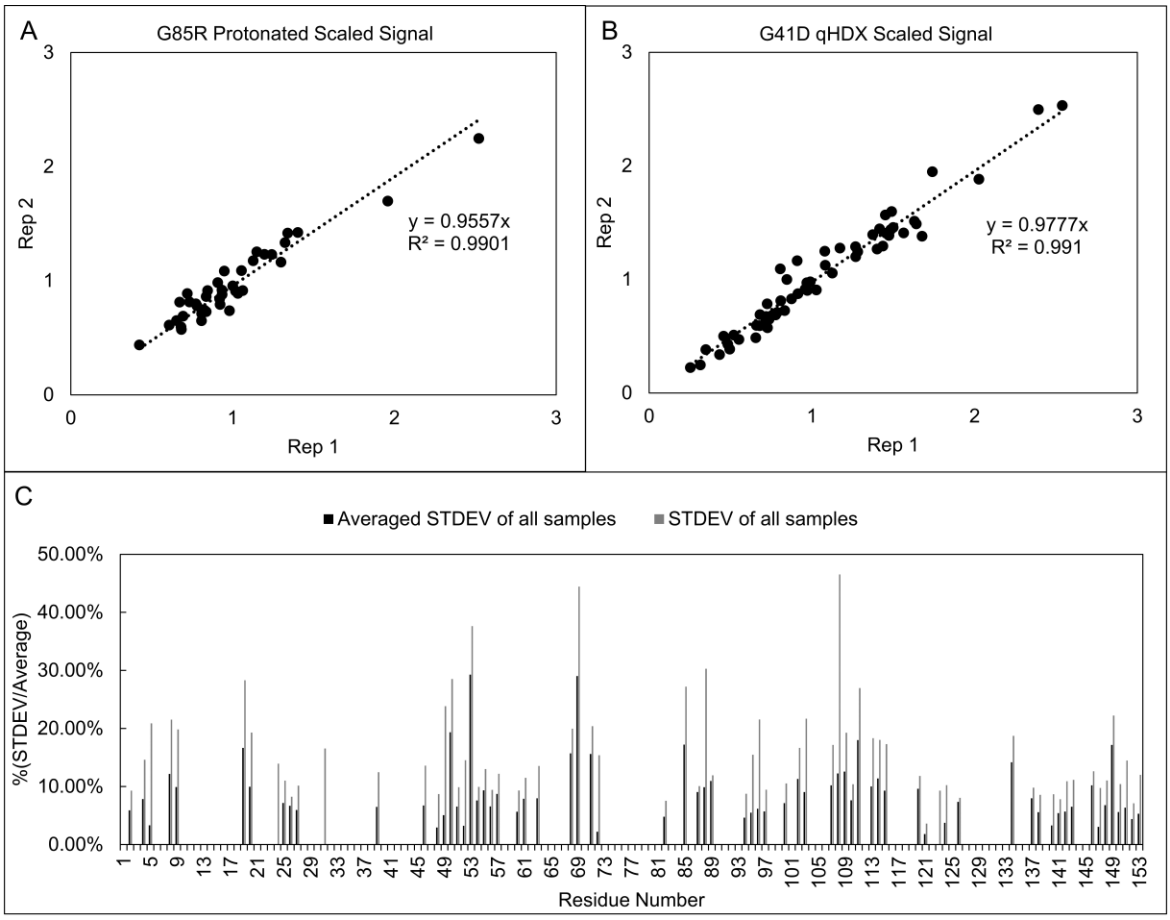


Figure 2.6: Comparisons of replicate qHDX sample results.
A) Biological replicates of G85R SOD1 fully protonated samples. Signal has been scaled using Eq. 2. B) Biological replicates of G41D SOD1 qHDX samples. Signal has been scaled using Eq. 2. C) Comparison of the averaged standard deviation of all 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.

2.5 Conclusions

We report here a detailed and optimized qHDX NMR method for the structural analysis of IBs, aggregates formed in recombinant bacteria that have many potential practical uses. High-

resolution structural data on these aggregates provides valuable information for further developing practical applications of IBs. Quantitative qHDX experiments require very careful control of experimental conditions, including sample pH, water content, temperature, and k_{int}. Detailed qHDX data on SOD1 are presented in Chapter 3. 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 3

Characterizing the structure of SOD1 IB aggregates using quenched HDX

This chapter has been submitted for publication to *Angewandte Chemie – International Edition* and is currently under review.

Citation:

*Naser D, ***Tarasca M V**, Siebeneichler B, Schaefer A, Deol HK, Soule T GB, Almey J, Kelso S, Mishra GG, Simon H, and Meiering EM. High Resolution NMR H / D Exchange of Human Cu, Zn Superoxide Dismutase Inclusion Bodies Reveals Significant Native Features Despite Structural Heterogeneity. *Angew. Chemie – Int. Ed*. (*Under Review).* 2021.

* Indicates these authors contributed equally to the work.

The material presented has been rearranged with respect to the above article. The following sections have not been submitted for publication and contain material not included in the above article:

- 3.3.2 Reduced apo SOD1 IBs expressed at 25 °C overnight
- 3.3.3 Zn-bound SOD1 IBs expressed at 37 °C for 4 hours
- 3.4 SOD1 qHDX Discussion (in part)
- 3.5 qHDX applied to other proteins
- 3.6 Conclusions

3.1 Introduction

3.1.1 SOD1 Inclusion Bodies Extent of Formation

Inclusion bodies (IBs) are cellular aggregates that commonly form upon overexpression of proteins in heterologous hosts[71–73]; these aggregates are of broad fundamental and practical interest[74,75] as they are important in protein preparation[16,24], as functional materials[76–78], and are medically relevant to toxic protein aggregation in disease[78]. While IBs have been characterized using various low-resolution methods, their insolubility and heterogeneity[14,78] 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*.

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.[47,48,51] 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.

Cu, Zn-Superoxide dismutase (SOD1, **Figure 3.1A**) serves as an ideal protein model to study structural impacts of mutations on globular proteins in IBs. 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[79,80]. 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.

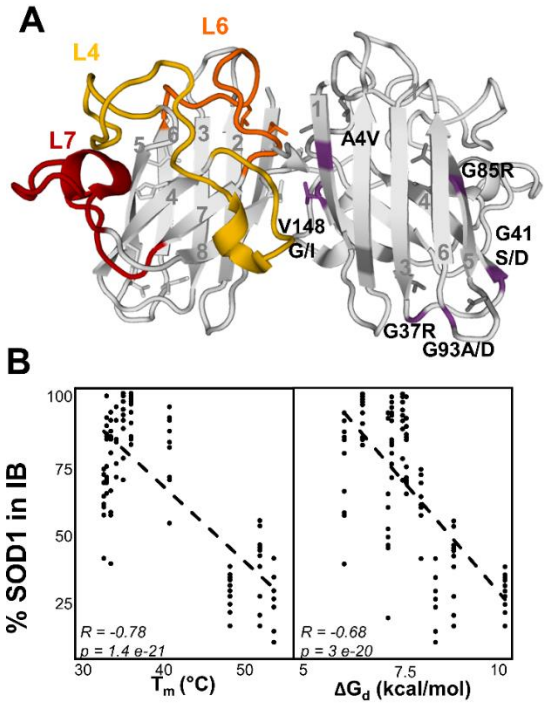


Figure 3.1: SOD1 Structure, Stability, and IB Formation.
(A) Native homodimer, with left subunit highlighting major structural features: β-strands 1-8 are numbered (β1- β8) and loops that make intermolecular interactions are coloured and numbered; PDB entry 1HL5. Right subunit shows mutations examined herein: grey/purple analyzed for % SOD1 in IB, purple also analyzed by qHDX, FTIR, and CR. (B) % cellular SOD1 in IB correlates with reduced apo T_m (*left*) and ΔG_d (*right*). Points correspond to separate samples for SOD1 mutants at 4 and 6 hours after induction of expression at 25 °C and 30 °C (**Table 3.1**).

First, IB formation was measured for 18 chemically diverse fALS point mutations (**Table 3.1**) that are distributed throughout the protein structure (**Figure 3.1A**, *right*) in *E. coli* cells grown in metal-free media at 25 °C, 30 °C and 37 °C (**Figure 3.1A**, **Appendix Figure A3.1**, **Figure A3.2**). Under these conditions, the IBs are formed by immature SOD1, i.e., with reduced cysteines and unmetallated (rApo) (**Figure A3.1**). Native rApo SOD1 is a marginally stable predominantly monomeric 8-stranded β-barrel with dynamic loops[25,40,81]. Similar to some proteins[82,83] but not

others[24,72] , 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) (**Figure 3.1B** *left*, **Table 3.1**)[40]. 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[84]. Notably, increased formation of mutant monomers with disrupted interfaces is also correlated with IB formation (**Figure 3.1B** *right*).

Mutant	Apo 25°C %SOD1 in IB	Apo 30°C %SOD1 in IB	ΔG (dimer)	ΔG (monomer)	Ox. Apo T_m	red. Apo T_m	Holo T_m
WT	26	20	10.3	2.8	59.1	47.6	92.7
A4V	90	91	6.4	2.3	50.7	36.3	86.7
G16S	64	83					
G37R	53	61	7.6	3.2	50.1	33.5	
G41D	67	57			45.2		86
G41S	65	75					84.4
H43R	77	68	7.5	1.7	48.1	35.4	86.3
H46R	31	26	8.4	5.3			
L84V	57	31			55.6	52.6	
G85R	77	79			54.7	40.7	77.5
D90A	35	24					77.5
G93A	71	45	7.2	1.7	47.9		87.7
G93D	77	39			45.6		85.1
E100G	63	37	8.0	1.6	51.2	33.2	86.2
L106V	80	81					
I113T	60	45	7.1	1.4	47.1		88.2
L144F	31	45					
V148G	67	44	5.9	2.2	49.3	34	86.9
V148I	29	40	8.9	5.7	60.5	51	92.7

Table 3.1: Summary of Aggregation Propensity and Biophysical Parameters of SOD1
%IB values are an average of 3 biological replicates measured previously by members of our group using SDS-PAGE and densitometry. ΔG_{dimer} and $\Delta G_{monomer}$ are from Broom *et al.*, 2015[84]. Fitted apparent midpoint of thermal unfolding transition (T_m) for reduced apo T_m's are fromVassal *et al.*, 2011[40], oxidized apo T_m's

(ox apo, metal free with disulphides oxidized) are from Vassal *et al.,* 2006[85], and holo T_m's are from Stathopulos *et al.,* 2006[39].

3.1.2 SOD1 Aggregation Prediction

As their name suggests, aggregation predictors can be utilized to predict the aggregation prone regions of a protein sequence. In addition to aggregation, there are prediction algorithms that can be utilized to predict other aspects of a protein sequence as well, including both hydrophobicity and secondary structure, among others. Aggregation was predicted for wild-type and the 9 mutant SOD1s characterized by qHDX using TANGO (sequence-based)[86], ZipperDB (sequence-based)[87], and CamSol (sequence- and structure-based)[88,89] predictors. The Kyte-Doolittle hydropathy index was utilized to determine the hydrophobicity of each mutant across their linear sequence[90]. Secondary structure and hydrogen bond predictions were obtained using RaptorX[91].

In general, the aggregation predictions differed in various details; however, all predicted a notable decrease in aggregation propensity within the β8 strand of V148G relative to WT (**Figure 3.2B**). Normalized prediction results for every other mutant relative to WT are shown in **Figure A3.3**. TANGO predicted a pronounced decrease in aggregation propensity for V148G, a moderate increase for A4V, a slight increase for V148I and no change for the other mutants (**Table 3.2**). ZipperDB predicted substantial decreases in aggregation propensity in the vicinity of G37R and G85R, moderate decreases for G93D and V148G, moderate increase for G41S, and slight increase for A4V, G93A, and V148I (**Table 3.2**). CamSol predicted a pronounced decrease in aggregation propensity for V148G, moderate decrease for G37R, G41S, G41D, G85R and G93D, slight decrease for V148I, pronounced increase for A4V, and moderate increase for G93A (**Table 3.2**). RaptorX uses an 8-point prediction to obtain the probability that any given residue is a part the

following structures: 4 turn helix, 3 turn helix, 5 turn helix, β-strand, β-bridge, hydrogen bonded turn, bend, or a loop. Only very small changes near sites of mutation were observed. Various residues that are protected in loops in IBs are predicted by RaptorX to form H bonded turns (**Figure 3.2C**). Residues that are predicted and observed to form β-strands in native SOD1 tend to show higher average protection than loop residues (**Figure 3.2**), consistent with packing of native-like monomers giving rise to protection in loops.

Mutant	CamSol	TANGO	ZipperDB
WT	0.45	0.63	0.61
A4V	0.00	1.00	0.60
G37R	0.81	0.64	0.81
G41D	0.80	0.63	0.41
G41S	0.42	0.63	0.00
G85R	0.66	0.64	1.00
G93A	0.30	0.63	0.80
G93D	0.82	0.63	0.80
V148G	1.00	0.00	0.85
V148I	0.53	0.58	0.40

Table 3.2: Normalized Aggregation Propensity Predictions.
The average total predictor score for each mutant was normalized using the following equation: Score = (average – minimum) / (maximum – minimum), for each individual predictor for them all to be on a scale of 0-1. For the predictors CamSol and ZipperDB, a higher number means the mutant is less aggregation prone. For the predictor TANGO, a higher number means the mutant is more aggregation prone.

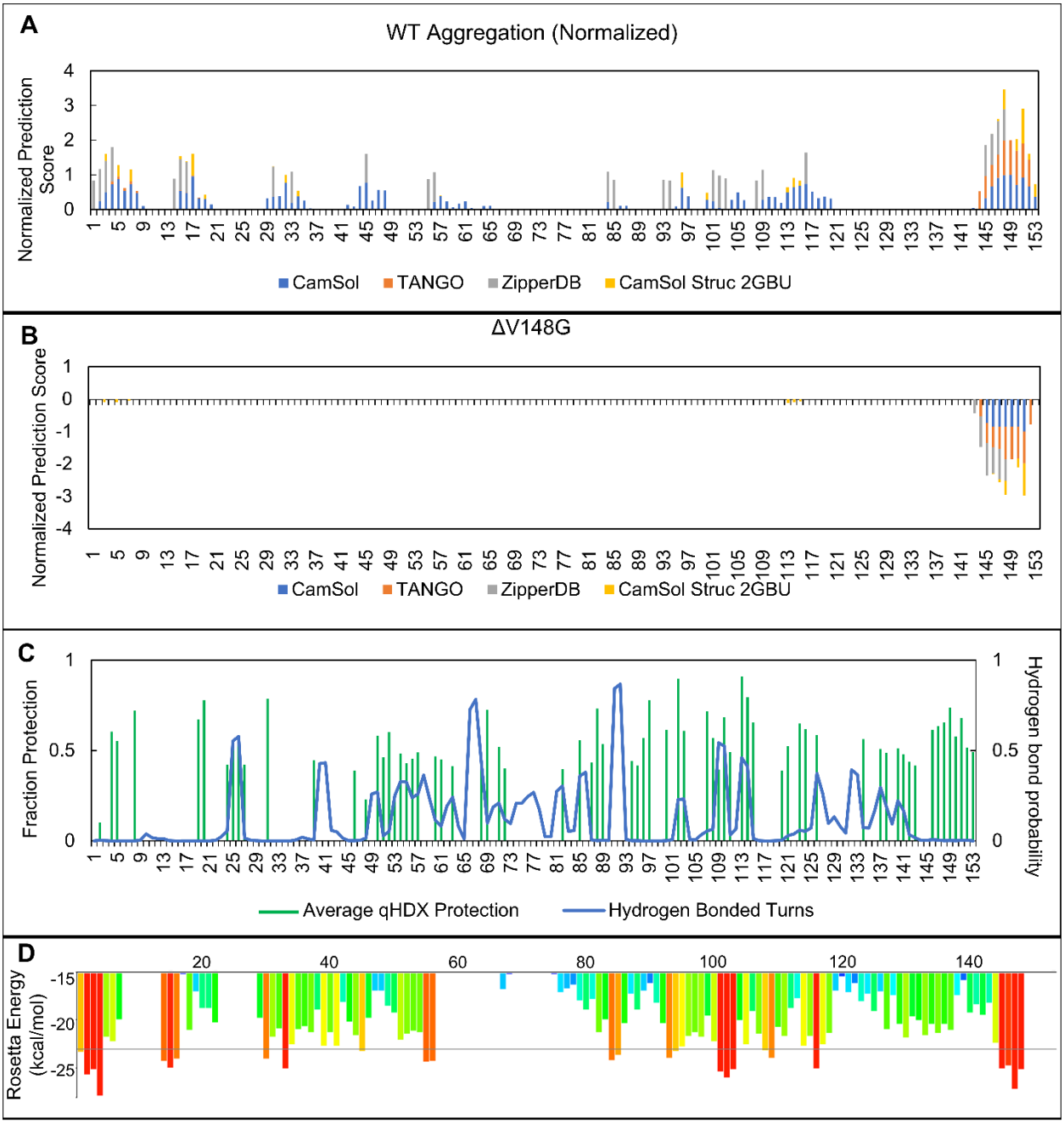


Figure 3.2: Aggregation and Secondary Structure Prediction for SOD1.
(A) Stacked, normalized aggregation predictor scores for WT SOD1 using CamSol, TANGO, ZipperDB, and CamSol Structured[86–88]. Scores for each predictor are normalized to a range of 0-1, corresponding to lowest and highest predicted aggregation propensities, respectively (see 3.2.2). (B) Differences in normalized predicted aggregation score for V148G with values for wild-type subtracted. See 3.2.2 Eq. 5 on calculating ΔResidue Aggregation Score. Negative values indicate regions predicted to aggregate less in V148G than in WT and positive values are predicted to aggregate more. Thus, V148G is predicted to be markedly less prone to aggregate at the C-terminus compared to WT, consistent with the observed qHDX data. (C) Hydrogen bonded residues

predicted by RaptorX (blue line) overlaid with average observed qHDX protection (green bars). Some residues in regions predicted to form turns that are not protected against exchange in native SOD1 and not predicted to aggregate are protected in IBs (E24-G27 in L2, N53-C57, H63, S68 in L4, G108-C111 in L6, L126, S134, G141 in L7). (D) ZipperDB aggregation propensity for wild-type SOD1. Regions with a Rosetta energy lower than -23 kcal mol^{-1} are highlighted in orange/red and are predicted to assemble into steric zippers.

3.2 Materials and Methods

All experiments, data acquisition, and data analysis are as described in Chapter 2. Covered in this section are the different growth protocol used for expressing SOD1 in the presence of Zn, and the normalization of aggregation predictor data.

3.2.1 Protein Expression, IB Preparation, and NMR Data Acquisition

Using 10X M9 minimal media stocks 1 L of 1X media was prepared and incubated overnight at 37 °C with shaking at 200 RPM. At the time of inoculation, 0.4% Glucose, 100 μM $CaCl_2$, 2 mM $MgSO_4$, 500 mg L^{-1} ^{15}N-NH_4Cl (or ^{14}N-NH_4Cl for non-isotopically labelled samples), and 0.1% Thiamine-HCl, were added to the medium of reduced apo (rApo) samples, together with 100 μg mL^{-1} ampicillin and 30 μg mL^{-1} chloramphenicol. For Zn-bound samples, the same 1X M9 medium was prepared as specified in Chapter 2. However, in addition to these reagents, 1 mM $ZnSO_4$ was added at this time as well. All Zn-containing inoculated M9 flasks were grown at 37 °C with shaking at 200 rpm for 4 hours, and expression was induced using 0.1 mM IPTG once an A_{600} of 0.6-0.8 was reached. All remaining steps, including expression, harvesting, sample preparation, NMR data acquisition, and NMR data analysis remain the same as stated in Chapter 2.

3.2.2 Aggregation Prediction

The servers TANGO (sequence-based)[86], ZipperDB (sequence-based)[87], and CamSol (sequence- and structure-based)[92] 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. In order to directly compare aggregation propensities predicted with different methods (which have different ranges of values), the prediction scores for each predictor were normalized on a scale of 0 to 1 as follows:

Normalized Aggregation Score for Residue i

$$= (\text{PredictionScore}_{\text{Residue i}} - \text{PredictionScore}_{\text{Lowest}})_i / (\text{PredictionScore}_{\text{Highest}} - \text{PredictionScore}_{\text{Lowest}}) \quad \textbf{Eq. 4}$$

where $\text{PredictionScore}_{\text{Residue i}}$ is the score for residue i, $\text{PredictionScore}_{\text{Lowest}}$ is the lowest predicted score for the protein sequence, and $\text{PredictionScore}_{\text{Highest}}$ is the highest predicted score for the protein sequence.

For the change (Δ) in aggregation propensity for individual residues in mutant relative to wild type, the difference in the residue scores obtained from each predictor were normalized using the absolute value of the difference for the residue with the maximum change:

ΔResidue Aggregation Score

$$= (\text{PredictionScore}_{\text{Mutant}} - \text{PredictionScore}_{\text{WildType}})_{\text{Residue i}} / |(\text{PredictionScore}_{\text{Mutant}} - \text{PredictionScore}_{\text{WildType}})_{\text{Max}}| \quad \textbf{Eq. 5}$$

Thus, the largest Δ values will be +1 or -1, with positive values indicating an increase in aggregation propensity and negative values indicating a decrease.

Secondary structure predictions including hydrogen bonded turns were obtained using RaptorX Property with SOD1 primary sequence and default prediction parameters[91,93].

3.3 SOD1 qHDX Results

3.3.1 Reduced apo SOD1 IBs expressed at 37 ℃ for 4 hours

Here we illustrate qHDX results for IB samples of the enzyme human Cu, Zn superoxide dismutase 1 (SOD1). Like many proteins, SOD1, a protein associated with the devastating neurodegenerative disease ALS, forms IBs when overexpressed in *E. coli*[94]. To define the IB structures, we performed qHDX experiments for nine fALS SOD1s (**Figure 3.1A**, *right*) containing point mutations that alter residue charge, hydrophobicity, secondary structure propensity, and increase or decrease aggregation propensity; the mutant T_ms range from ~51 °C (for V148I) to ~33 °C (for G37R) (**Table 3.1**). The qHDX experiments require a high level of IB formation, thus we measured qHDX for high aggregating reduced apo (rApo), the most immature form of SOD1 with reduced cysteines and no bound metals, mutants at 37 °C. Using resonance assignments for resolved amides (**Figure A2.6**, **Table A3.1**), residue assignments were transferred to the acquired ^{1}H-^{15}N HSQC of the qHDX sample for intensity analysis.

The observed exchange is biphasic, as also reported for other IBs, and indicating conformational heterogeneity (**Figure A3.7**, **Figure A3.8**). The measured qHDX data show extensive protection throughout the protein, with the general protection pattern very largely maintained across mutants (**Figure 3.3A**). Notably, extensive protection is observed throughout the β-strands of the protein, implying the β-barrel structure within SOD1 may remain in-tact within the aggregate. This extensive protection (**Figure 3.3A**) is in sharp contrast to the much more limited localized protection reported for IBs of other proteins.[51,95] Remarkably, the mutants display similar protection including and extending beyond the extensive protection observed for soluble native SOD1 (**Figure 3.3G**), despite large differences in protein stabilities and mutation type. One

notable exception is V148G SOD1, which, while still displaying extensive protection throughout the primary protein sequence, has decreased protection in β8 and loop L4, and increased protection in L6 (**Figure 3.3A,H**). The ratios of protection between all mutants and the average protection for this condition was determined in order to highlight any potential mutant-dependent differences in protection (**Figure A3.4**). Only V148G was found to differ significantly from the average protection of the other mutants.

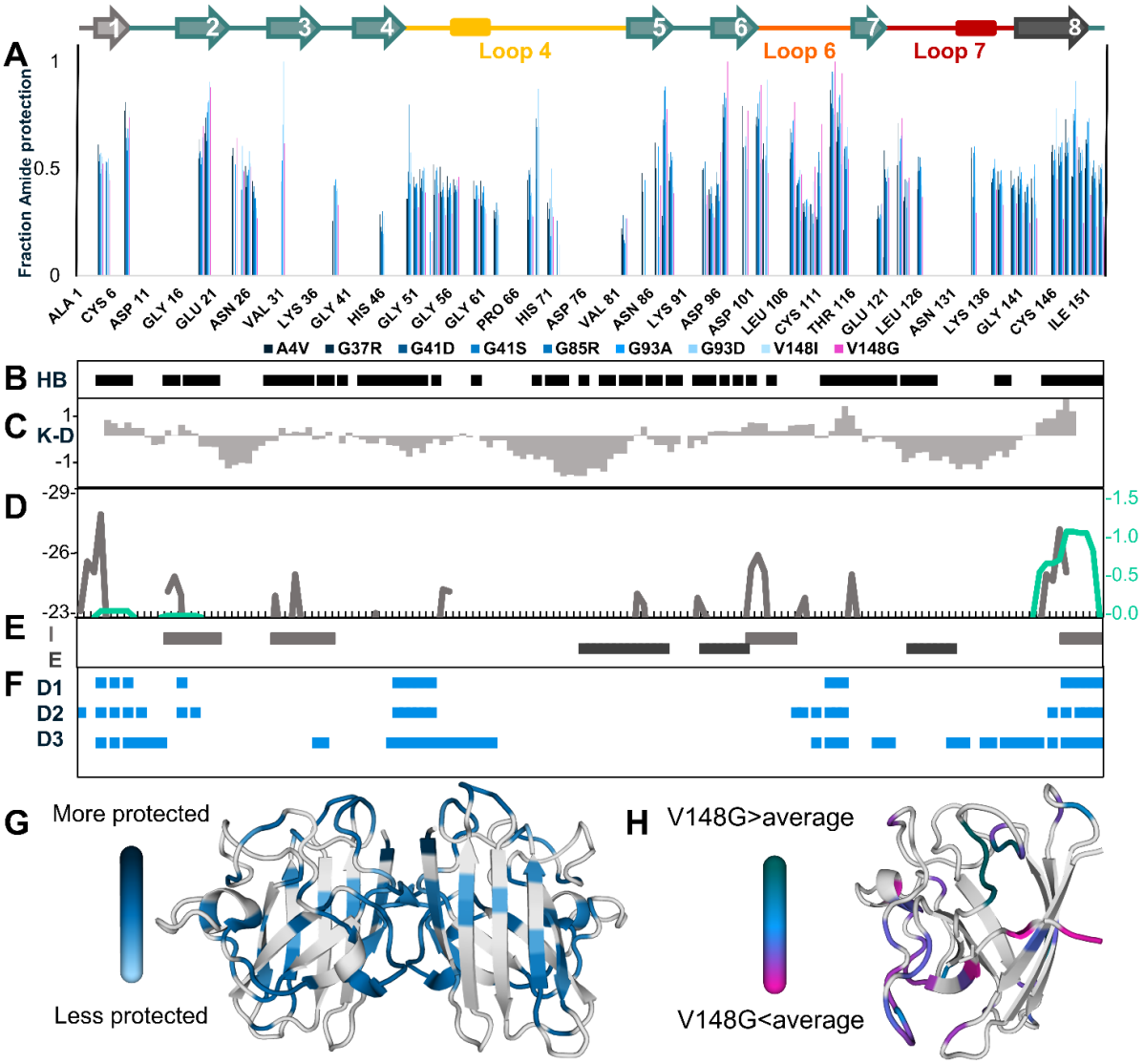


Figure 3.3: Structure of Mutant SOD1 IBs from qHDX Analysis.
(A) Fraction amide protected for 9 SOD1 mutants. Each bar is an average of two biological replicates. Gaps indicate areas of no data. (B) WT holo SOD1 traditional amide HDX protection.[96] (C) Kyte-Doolittle hydropathy index of WT SOD1.[90] (D) TANGO[86] (green) and ZipperDB[87] (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 (**Figure 3.2B**). (E) Gain of interaction regions observed for steric zipper peptides by Ivanova et al.[97] (light grey, top) and filaments of native dimers by Elam et al.[98] (dark grey, bottom). (F) Residues that engage in native dimerization (D1) and non-native dimerization (D2, D3)[25]; see also **Figure 3.10**. (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).

3.3.2 Reduced apo SOD1 IBs expressed at 25 ℃ overnight

We further investigated the effect of expression time and expression temperature on the structure of IBs formed by SOD1. As shown in **Figure A3.2**, the proportion of SOD1 that forms IBs is decreased in cells grown at 25 °C relative to cells grown at 37 °C. With these differences in expression, we wanted to determine if there are structural differences in the IB under the two different conditions. To define these IB structures, we performed qHDX experiments for four fALS associated SOD1 mutants: A4V, G85R, G93A, and V148G (**Figure 3.1A**, *right*). These mutants were selected due to their lower thermal stability and ability to form sufficient IBs at a lower expression temperature to allow for quantitative analysis (**Table 3.1**). This is of interest, as extent of IB formation is correlated to decreasing reduced apo T_m (**Figure 3.1B,** *left*). As in 3.3.1, crosspeak intensities using resolved amides were measured for SOD1 expressed at 25 ℃ overnight, roughly 18 hours post-induction. Attempts were made to express SOD1 for 6 hours at 25 ℃, however this resulted in both inconsistent and insufficient levels of aggregation for qHDX.

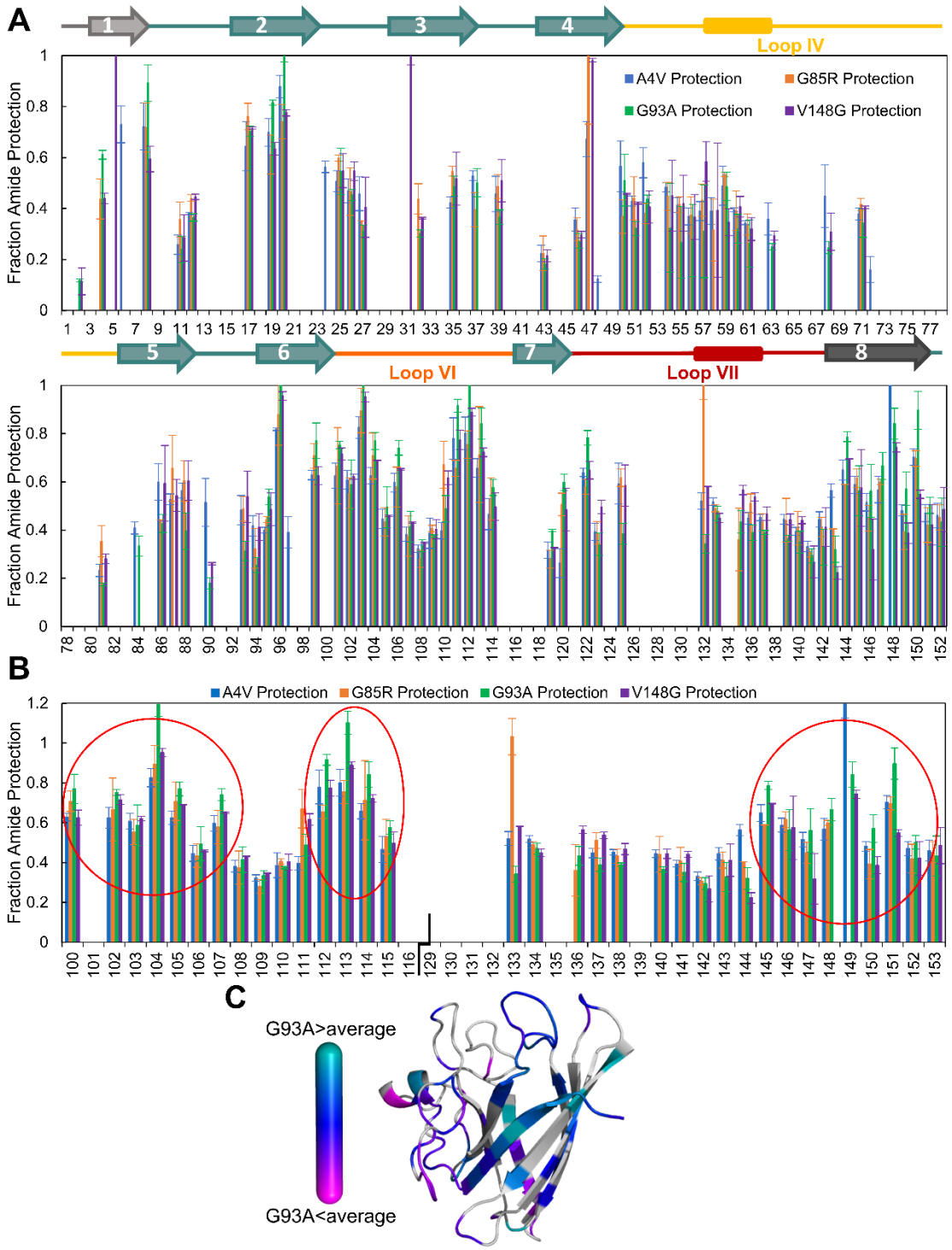


Figure 3.4: qHDX amide protection of mutant SOD1s expressed at 25 °C overnight.
(A) Fraction amide protected for four SOD1 mutants, A4V, G85R, G93A, and V148G, expressed at 25 °C overnight. Each bar is an average of two biological replicates. Gaps indicate areas of no data. The sequence is split into two rows for readability. The secondary structure of the protein is displayed above each row. (B) Zoom of regions of interest from the previous graph, highlighting

with red circles higher L6 and β8 protection in G93A relative to the other mutants expressed at the same conditions. Residues 117-128 were removed for legibility (split marked with a black squiggle). (C) Ratio of G93A protection relative to the average qHDX protection. Ratio is colored from 0.5 to 1.5 in magenta to teal, respectively (PDB 1HL4).

Similar to IBs of SOD1 expressed at 37 °C for 4 hours, the measured qHDX data show extensive protection throughout the protein sequence, with the general pattern maintained throughout the different mutants measured (**Figure 3.4**). The observed pattern of amide protection is also, surprisingly, comparable to what was measured for IBs expressed at 37 °C for 4 hours, and the average protection of all mutants generally does not differ significantly between the two conditions (**Figure 3.5**). The fraction amide protection of each mutant was directly compared for each mutant across each condition, and the general pattern was largely maintained within experimental uncertainty between each expression condition (**Figure 3.8**, **Figure A3.9**). The exceptions to this trend are the mutants G93A and V148G. G93A shows increased protection in L6 and β8, and less protection in L4 relative to the average of the mutants expressed at 25 °C for 4 hours (**Figure 3.4B**,**C**). Interestingly, V148G, which had a noticeably different amide protection pattern compared to the average protection of all other mutants at 37°C, did not display differences larger than experimental uncertainties relative to the average of the other mutants when expressed at 25 °C. The amide protection ratios between each mutant and the average protection for this condition was determined to highlight any potential mutant-dependent differences in protection (**Figure A3.5**). Only G93A was found to differ from the average protection of this expression condition. As highlighted above, the average fraction amide protected between the two conditions are comparable within experimental uncertainty (**Figure 3.5**), therefore the mutants G93A and V148G both show differences in their extent of amide protection depending on the expression time and temperature. These temperature dependent differences are further explored in section 3.4 and are highlighted in **Figure 3.8**.

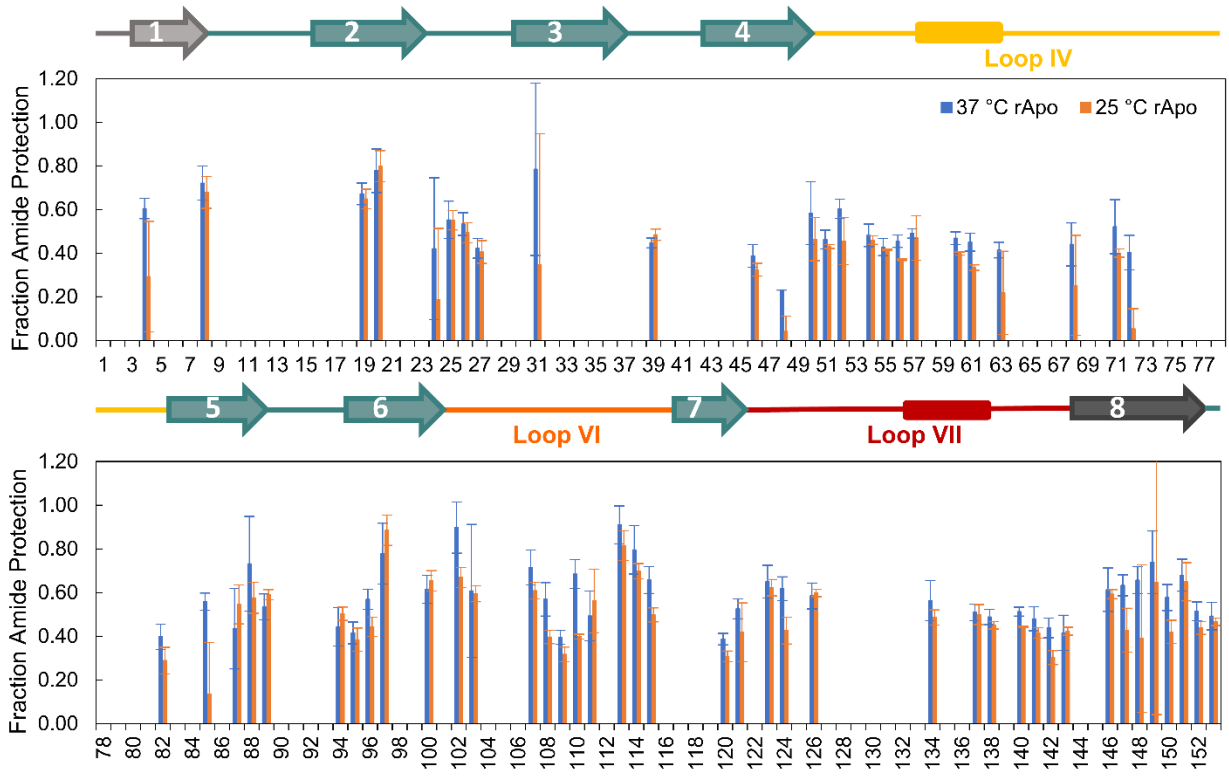


Figure 3.5: Comparison of average fraction amide protection between expression temperatures.
Comparison of normalized protection between two different expression conditions. The 37 °C dataset is the average of 2 biological replicates across eight different mutant SOD1s: A4V, G37R, G41D, G41S, G85R, G93A, G93D, and V148I (**Figure 3.3**). The 25 °C dataset is the average value of two biological replicates for each of A4V, G85R, and V148G SOD1. G93A is not included in this average due to differences in amide protection outside of experimental uncertainties. Fraction amide protection is calculated as per Chapter 2 (see 2.4.2 for quantitation of amide protection). Error bars represent one standard deviation of the averaged data. The overall protection profiles between the 37 °C data and the 25 °C data are similar, indicating similar ensembles of SOD1 conformations in the IBs.

3.3.3 Zn-bound SOD1 IBs expressed at 37 °C for 4 hours

In addition to measuring the aggregate structure of SOD1 expressed at a lower temperature, we investigated the effect of the presence of Zn during the expression of SOD1 IBs. Zn addition during the expression of SOD1 has been shown to stabilize the soluble native state; it was necessary to express SOD1 IBs at a higher temperature (37 °C) to ensure sufficient aggregation for qHDX analysis. Despite the increased expression temperature, not all mutants could be

analyzed by qHDX. Previous aggregation propensity studies performed in our lab have shown some mutant SOD1s remain largely soluble after 4 hours of expression, despite the increased expression temperature.

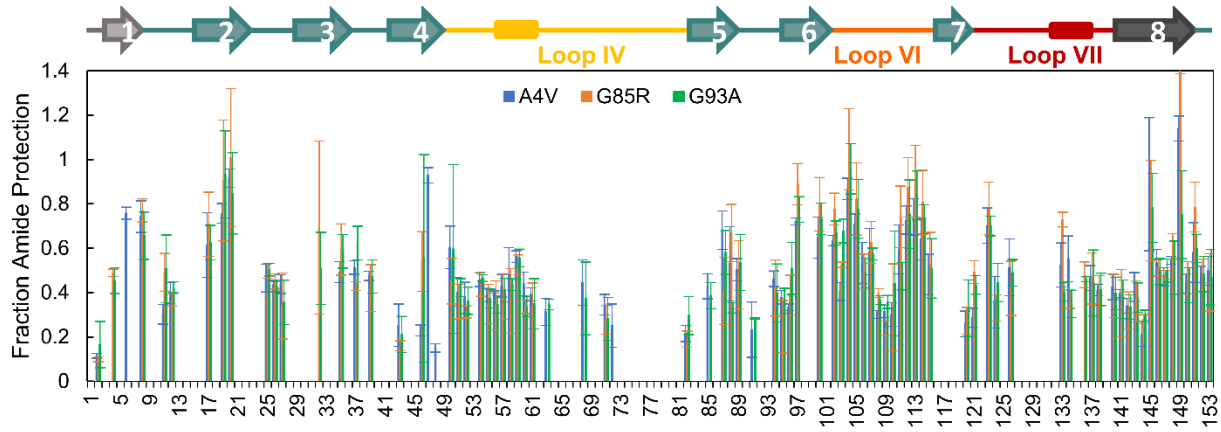


Figure 3.6: qHDX amide protection of mutant SOD1s expressed at 37 °C for 4 hours with Zn.
Fraction amide protected for 3 SOD1 mutants, A4V, G85R, and G93A, each expressed for 4 hours at 37 °C with Zn. Each bar is an average of two biological replicates. Error bars represent one standard deviation. Gaps indicate areas of no data. Normalized data determined from scaled signal of the first ^{1}H-^{15}N HSQC of the acquired series (see Chapter 2, 2.2.6).

As seen with the rApo SOD1 IBs expressed at 37 °C for 4 hours and the rApo SOD1 IBs expressed at 25 °C overnight, SOD1 IBs grown in the presence of Zn show extensive protection throughout the protein sequence, with the general pattern in the protection profile maintained throughout all the mutants measured (**Figure 3.6**). Unlike the other two conditions, none of the mutants studied stand out as an outlier from the average fraction amide protection pattern of SOD1 expressed at 37 °C for 4 hours with Zn outside of experimental uncertainty (**Figure 3.6**). Following the same trend as the 25 °C dataset, each mutant was compared to the average amide protection measured for rApo SOD1 expressed at the same conditions except without addition of Zn, and no considerable difference in protection was observed outside of experimental uncertainty (**Figure 3.7A**). This trend also holds true when the same mutant is compared with itself at different

expression conditions (**Figure A3.9**), except for the previously mentioned exceptions – rApo V148G expressed at 37 C for 4 hours and rApo G93A expressed at 25 C overnight (**Figure 3.8**). This information is also plotted onto a ribbon structure to further highlight the aggregation prone regions of the protein, L4, L6, and β8 being the most protected (**Figure 3.7B**). As noted above, no differences in amide protection outside of experimental uncertainty estimates are observed between the metalated and rApo conditions. The amide protection ratios between each mutant and the average protection for this condition was determined to highlight any potential mutant-dependent differences in protection, however no significant differences in amide protection were observed using these ratios (**Figure A3.6**).

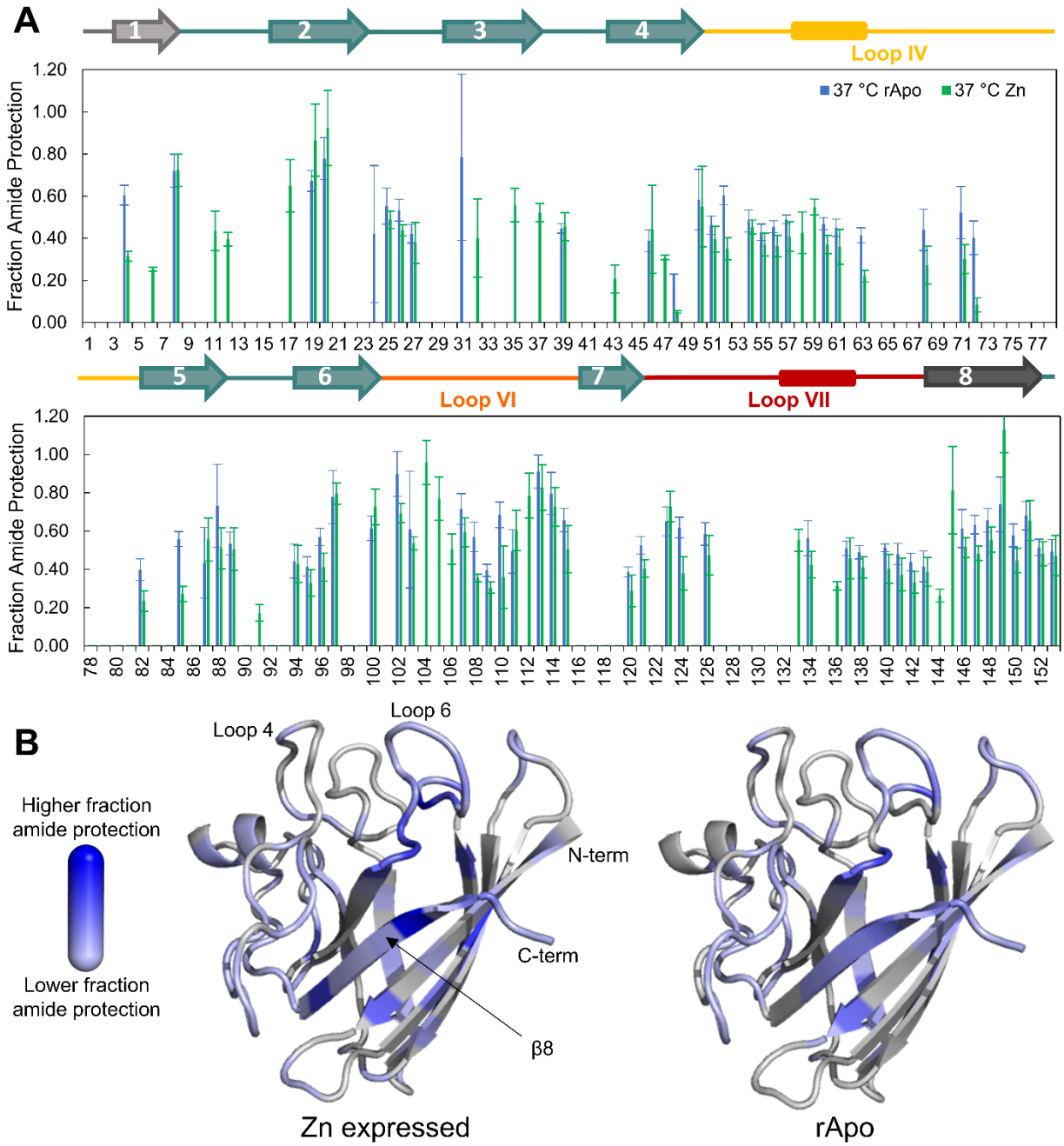


Figure 3.7: Average qHDX protection of mutant SOD1 expressed with and without Zn. (A) Average qHDX protection for two biological replicates of the mutant SOD1s expressed at 37 °C for 4 hours with Zn: A4V, G85R, and G93A, compared to the average qHDX protection for two biological replicates of mutant SOD1s expressed at 37 °C for 4 hours without Zn: A4V, G37R, G41D, G41S, G85R, G93A, G93D, and V148I (**Figure 3.3**). The sequence has been divided across two rows for readability. Error bars represent one standard deviation. Normalized data determined from scaled signal of the first ^{1}H-^{15}N HSQC of the acquired series (see Chapter 2, 2.2.6). (B) *Left* Average qHDX protection of all mutants (37 °C, 4 hours, with Zn) shown on SOD1 monomer

(PDB 2GBU). *Right* Average qHDX protection of all mutants (37 °C, 4 hours, rApo), expect V148G, shown on SOD1 monomer (PDB 2GBU). Protection from 0.5 to 1.0 is colored as light to dark blue, respectively. Grey indicates no data.

3.4 SOD1 qHDX Discussion

To elucidate the molecular basis for the observed extensive protection across each sample, we examined experimental and predicted biophysical properties of SOD1 (**Figure 3.3B**,**C**). The primary sequence distribution of protected amides in IBs generally resembles that in native rApo and holo SOD1, noting that generally protection is conferred primarily by intramolecular hydrogen bonding for globular proteins[96]. For each expression condition analyzed, all mutant SOD1s exhibited extensive protection throughout the 8 strands of the β-barrel (β1-β8) (**Figure 3.3**, **Figure 3.4**, **Figure 3.6**). Protection in the loops connecting the strands is much more pronounced for rApo in IBs than expected for the soluble monomer[25,99]. The loop protection in the IBs resembles that in dimeric holo SOD1, which includes some intermolecular hydrogen bonding. Notably, the long loops, 4 and 7, are more extensively protected in the IBs than in the native dimer despite their strongly hydrophilic character[90] (**Figure 3.3C**).

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[86] and ZipperDB[87] (**Figure 3.2**, **Figure 3.3D**). The C-terminal strand, β8, is generally predicted as the most highly prone to aggregate, followed by β1. These strands as well as β2, β3 and L6 contain short stretches of residues with high predicted propensity to form steric zippers, assemblies consisting of hydrogen bonded antiparallel or parallel β-strands[97]. Previous experimental studies observed zipper formation by short peptides corresponding to β2, β3, β8, and L6, whereas a β2-β3 peptide formed a non-amyloid corkscrew oligomer[97,100]. It is important to note that native structure can modulate or prevent these

short sequences from being available to assemble into zippers[101]; also, the steric zipper regions in SOD1 comprise only a fraction of the many protected residues in the IBs.

We therefore must consider additional contributions of native-like structure in IBs. A model based on the β2-β3 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 inter-subunit packing and hydrogen bonding between L7 and β5-β6 [98,102,103], regions that also show HDX protection in IBs (**Figure 3.3E**). Furthermore, NMR experiments suggest rApo monomers of wild-type and fALS mutant SOD1s can form native dimer (D1) and 2 non-native dimers (a symmetric one, D2, and an asymmetric one, D3) in solution[25,104] (**Figure 3.3F**). Importantly, the interfaces for these dimers include intermolecular interactions of loops L4, L6 and L7, which are also protected against exchange in IBs (**Figure 3.3**, **Figure 3.4**, **Figure 3.6**). Although the extent of IB formation is decreased at lower expression temperatures and with the addition of Zn, in all cases similar protection is observed throughout the mutant proteins (**Figure 3.4**, **Figure 3.6**). These findings are consistent with native-like conformations in SOD1 contributing to the IB structure. Due to the extensive similarities in amide protection observed across the 3 different conditions, the data suggest contributions of native-like structures in the IBs, and varying expression conditions affects the extent of aggregation rather than the structure itself.

The altered HDX protection in V148G SOD1 (**Figure 3.3H**) further supports contributions of D1, D2 and D3 to IB structure. V148G is situated in β8, packing in the center of the native D1 dimer interface (**Figure 3.1A**), which is much weakened by the V148G mutation[84]. Increased dissociation of native D1 would decrease protection in β8 and L4, and the non-native asymmetric

D3 interface is also likely to be disrupted by V148G[25] thereby decreasing protection of L7. Thus, the relative protection of L6 due to a higher relative population of symmetric D2 may be increased, as observed in V148G. Interestingly, this altered protection is only observed at 37 °C, and not at 25 °C (**Figure 3.8**). V148G has one of the lowest thermal stabilities of all mutant SOD1 analyzed in this study and at 37 °C, it is more unfolded than at 25 °C. As the protein becomes more dynamic, it is believed more energetic pathways for assembly may become available[105]. This could result in changes for key structural regions involved in processes such as in the native D1 dimer interface. Multiple pathways are likely accessible to SOD1 mutants but to varying extents. For instance, if the packing of residue 148 is affected by this mutation, it could make other pathways more accessible, like the formation of non-native dimer interfaces D2 and D3. At 25 °C, where the protein itself is more folded and less structurally dynamic, there may be less pathways accessible. For instance, the loops involved in the native interface may be less flexible and therefore more likely to encourage the formation of the native D1 dimer interface relative to the non-native D2 dimer interface. However, at 37 °C with more dynamics we may allow for more non-native pathways to be accessible, the protection pattern as observed by qHDX may change.

The other mutant with observed altered protection compared to other mutants expressed at the same conditions, G93A at 25 °C (**Figure 3.4C**), is a highly conserved residue situated within L5 and shows increased protection in L6 and β8, and less protection in L4. As with V148G expressed at 37 °C, this could imply that increased populations of non-native dimer D2 may be observed for rApo G93A expressed at 25 °C. As well, G93A has been shown to be able to weaken the native D1 dimer interface, despite being far from the dimer interface[84]. Additionally, like V148G, this perturbed structure is only present at one temperature, however in this case the altered protection is observed at the lower temperature, 25 °C, as opposed to the higher temperature, 37 °C (**Figure**

3.8). The mutant G93A, being situated in L5 close to β6, may be disrupting the packing of L6 in the native-like structure, resulting in the aggregate ensemble shifting towards a structure with a more highly protected L6. As for why this change in protection only occurs at the lower of the two expression temperatures, some population of G93A SOD1 could be trapped in a non-native local minimum in the energy landscape funnel model of folding and cannot escape at the lower of the two expression temperatures. At 37 °C, where more thermal energy is available to the system, the protein may be able to surpass this minimum and form, predominantly, its native structure. This could be a potential explanation as to why the mutant at 37 °C appears to have a more native-like D1 dimer interface than at 25 °C. However, specific explanations as to why differences in protection are observed for the G93A mutation warrants further investigation.

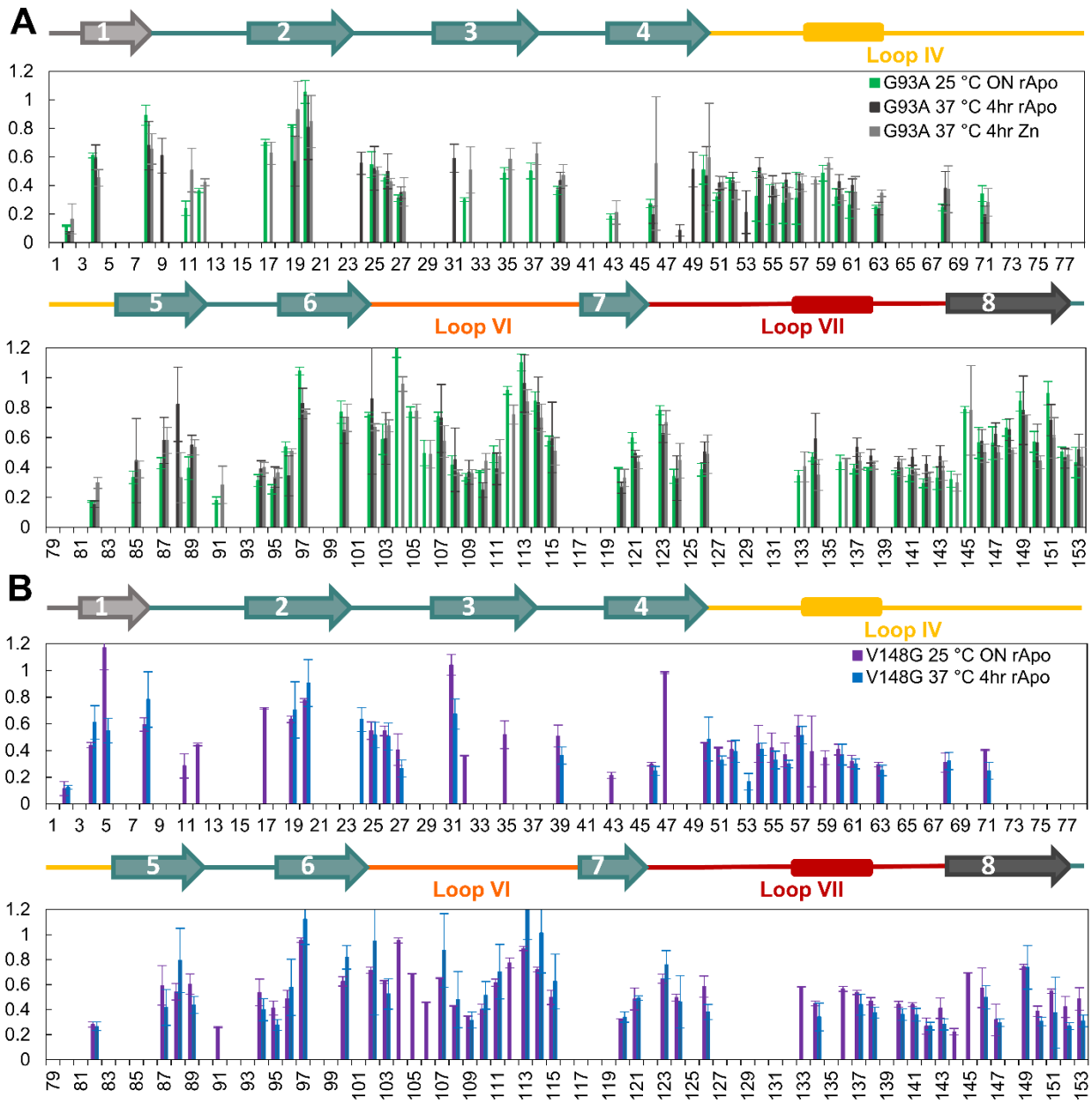


Figure 3.8: Comparison of fraction amide protection between expression conditions.
(A) Bar plot comparison of the fraction amide protection of G93A across all three expression conditions. Each bar is the averaged protection of two biological replicates. Error bars represent propagated error and were calculated using Eq. 6 (see below). (B) Bar plot comparison of the fraction amide protection of V148G across both rApo expression conditions. Each bar is the average of two biological replicates.

$$\frac{\Delta z}{z} = \sqrt{\left(\frac{\Delta x}{x}\right)^2 + \left(\frac{\Delta y}{y}\right)^2} \qquad \textbf{Eq. 6}$$

Where:

x = average scaled qHDX signal for a condition

y = averaged scaled protonated signal for a condition

z = ratio of the x/y

The IBs were further analysed using FTIR and CR spectroscopies. These data were acquired and analysed by Dalia Naser, Anna Schaefer, and Bruna Siebeneichler, all members of our group with whom I worked in collaboration. These data are included within this document with permission to add complementary structural information on SOD1 IBs.

FTIR and CR spectroscopies were employed to measure the IBs secondary structure and amyloid content, respectively. Consistent with the qHDX results, the spectra of the SOD1 mutant IBs coincide closely, demonstrating extensive structural similarities among the mutant IBs. The spectra and secondary structure analyses of IBs are akin to those of native rApo and holo SOD1 in solution (**Figure 3.9A,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[106]. Amyloid typically exhibits absorbance at low wavenumbers arising from short intermolecular hydrogen bonds (<1630 cm^{-1}), which is prominent only for sonicated SOD1. Significant absorbance at high wavenumbers (>1680 cm^{-1}) for IBs, native SOD1, and aggregates of rApo SOD1 in rats overexpressing human G93A SOD1[107] is characteristic of antiparallel β-sheet and turns. Conversely, parallel β-amyloid lacks absorbance in this region[108–110], as is also apparent for sonicated SOD1. Curve fitting of secondary structure components agrees with the preceding observations (**Figure 3.9B**). Overall, the FTIR

data indicate a lack of extensive amyloid structure, some disorder, and substantial presence of native-like conformations in the IBs.

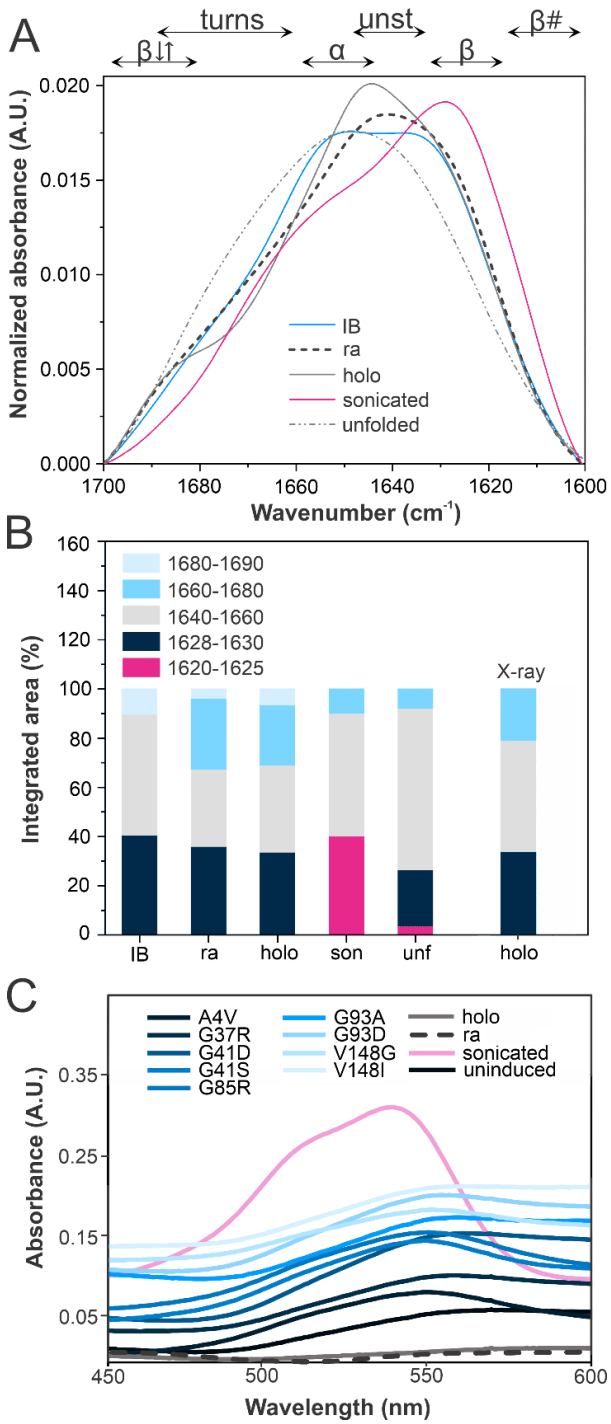


Figure 3.9: FTIR and CR Spectra of IBs exhibit Native and Amyloid Features.
(A) ATR-FTIR spectra in the amide I region for average IB, pure native reduced apo (ra) and holo pWT, ra A4V sonicated, and heat unfolded at 55 ºC. Spectral regions for antiparallel β-sheet (β⇓↑),

β-turn, α-helix (α), unstructured (unst), β-sheet (β), and β-amyloid with strong intermolecular hydrogen bonds (β#) are indicated. (B) Integrated area of secondary structure bands from curve-fitting analysis and the holo SOD1 x-ray structure[111]. (C) CR difference absorbance spectra for mutant IBs, uninduced cells, and pure holo, ra and sonicated amyloid SOD1**.**

The binding of CR to amyloid causes a red-shift in the CR absorbance spectrum, with maximal difference at ~541 nm[112]. Sonicated SOD1 exhibits typical CR spectral features corresponding to amyloid[106]. The spectra of mutant IBs are muted in comparison (**Figure 3.9C**), 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[113]. CR spectroscopy is therefore also consistent with limited amyloid structure in the SOD1 IBs.

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[84] 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[25,98,102–104] will confer protection of residues in loops, which are unstructured in native reduced apo SOD1[25,99]. 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[104,114]. 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 3.10**).

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, while increasing SOD1 stability through the addition of zinc decreases the proportion of SOD1 in IBs (**Figure 3.10, Figure A3.1**); why and how this occurs is unclear and requires further investigation.

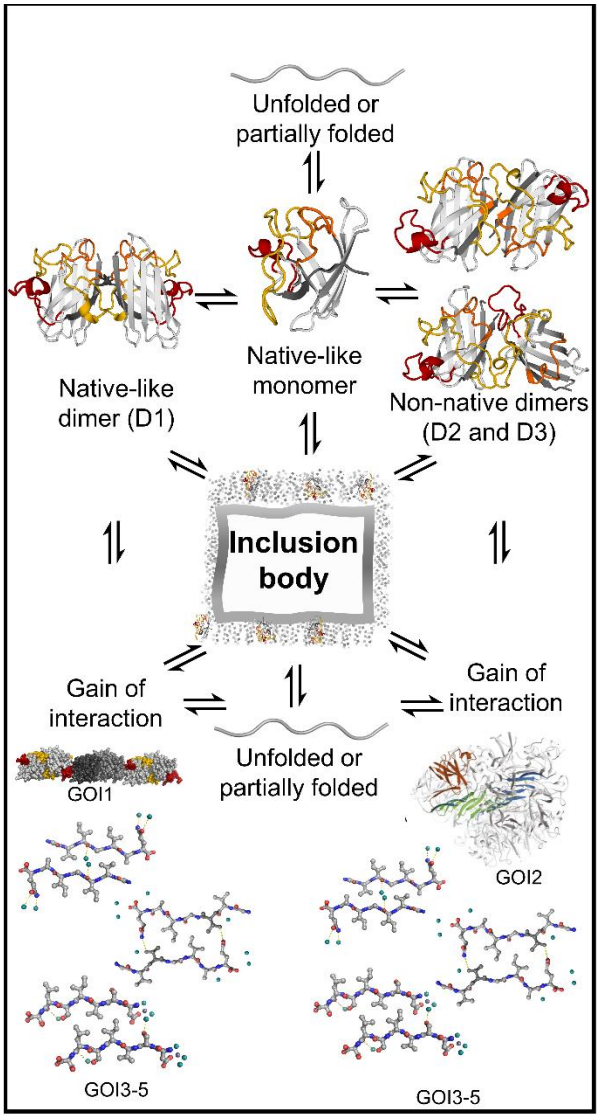


Figure 3.10: SOD1 IB Formation Involves Multiple Pathways that Include Native and Non-Native Interactions.
Loops are coloured as in **Figure 3.3**. D1, D2, D3 form in solution[25,104]. Gain of interactions in crystal structures include: GOI1, an "amyloid-like" filament of near-native dimers where L7 (red) packs against the β5-β6 edge strands of the next dimer including intermolecular hydrogen bonding (PDB entry 1OZU) [98,102]; GOI2 is a corkscrew oligomer model (5IIW)[100]; 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 [97].

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[115], which are desirable for the production of native proteins and as functional protein particles[76,77,116].

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[81,107] and patients[113] 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 Aβ[110,117], although some early stage aggregates of these proteins also contain significant antiparallel structure[118,119]. 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[75,78,117,120,121]. 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[122,123]. It is therefore crucial to be mindful of such factors when, for example, seeking to develop therapeutics to target protein aggregation. The qHDX method highlighted here has the advantage of minimal sample processing and associated changes to the protein structures formed in cells.

3.5 qHDX applied to other proteins

qHDX has been applied to Adnectin IBs by Anna Schaefer in our group. All data acquisition and analysis has been completed by her. Her data are included in this thesis with permission as an additional comparison to SOD1 IBs, as well as proof of this method successfully being applicable to another unrelated protein.

Adnectins, a family of engineered proteins derived from the human 10th fibronectin type 3 domain, have β-sandwich immunoglobulin-like structure with three loops analogous to the complementarity determining regions 1-3 of antibody variable domains. 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[53,124,125]. Adnectins 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[24]. Preliminary studies conducted for two Adnectin mutants reveal protection patterns resembling those of SOD1 (**Figure 3.11**). Furthermore, the general pattern is maintained between mutants, despite their widely different solubilities. While a comprehensive study is ongoing, widespread protection throughout the sequence agrees with the previous proposal that Adnectin aggregates retain native-like structure[24].

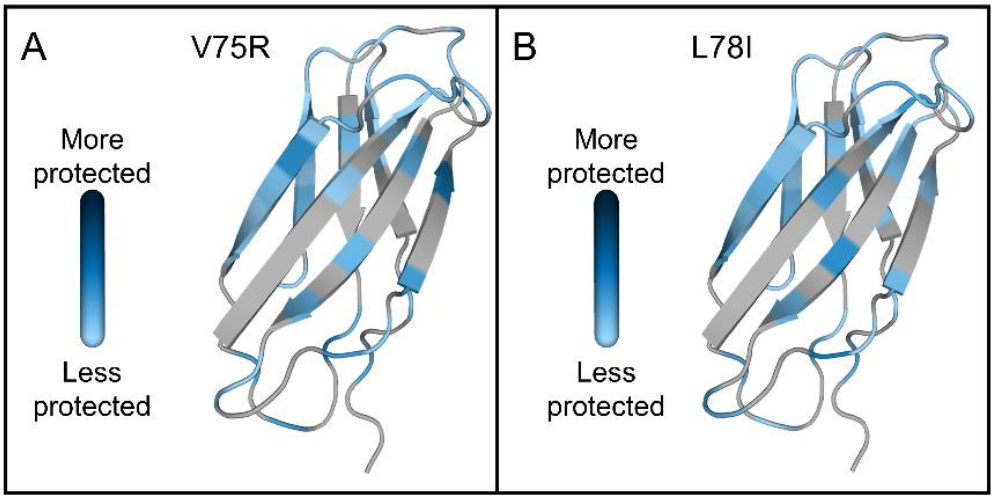


Figure 3.11: Structure of Adnectin IBs from qHDX analysis.
A) Average qHDX protection of two biological replicates of Adnectin V75R (model based on PDB 1FNF made with software Swiss-Model[126]). Protection fractions from 0.5 to 1.0 are coloured as light to dark blue, respectively. Grey indicates no data. B) Average qHDX protection of two biological replicates of Adnectin L78I (model based on PDB 1FNF made with software Swiss-Model[126]). The two mutants have very similar protection, implying similar IB structure, as observed also observed for SOD1 mutants.

3.6 Conclusions

In conclusion, this 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. qHDX observed by NMR 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.

Chapter 4

Conclusions and Future Work

4.1 Conclusions

IBs are a type of aggregate that can be found in recombinant bacteria with many practical medical, biotechnological, and industrial applications. Presented is a method, qHDX, that has been optimized to measure the high-resolution structure of IBs. High-resolution structural data on these aggregates provides the opportunity to improve upon these practical uses. Through measuring and controlling the structure of IBs, one could potentially engineer IBs with specific biophysical properties that can aid in these applications by developing more advanced recombinant expression systems.

qHDX NMR has been optimized for IB study on the model protein, SOD1, and has been shown to provide consistent, high-resolution information on individual amino acid amide protection. Interestingly, qHDX NMR with supporting FTIR and CR data display extensive protection throughout the aggregate, indicative of native-like structures within the IB. This conclusion holds true across the three measured expression conditions of SOD1: 4 hours at 37 ℃ (rApo), 4 hours at 37 ℃ (with Zn), and overnight at 25 ℃. Surprisingly, all conditions do not deviate in amide protection from each other, supporting the argument that SOD1 IB aggregation is through an ensemble of pathways independent of expression conditions or mutation.

As demonstrated, qHDX can be readily applied to explore many different recombinant expression systems and proteins. 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. This powerful method has the potential to lead to important advancements in the field of protein engineering, as it is an effective and efficient way to determine the high-resolution structure of protein aggregates formed *in vivo*.

4.2 Future Work

There are plenty of interesting avenues to follow in the future using qHDX on both SOD1 as well as other systems. One potential path to pursue would be the qHDX analysis of loop 6 mutants of SOD1. Potential mutants can include E100G and L106V, both being mutants shown to aggregate into IBs through previous analyses by members of the Meiering group, as well as being associated with the development of ALS. Loop 6 mutants would be of interest due to it being a hotspot of differences in our data sets. As mentioned in Chapter 3, loop 6 differs significantly from the average protection of V148G expressed for 4 hours at 37 °C, as well as G93A expressed overnight at 25 °C, having increased fraction amide protection in both cases.

Performing this analysis on a complete set of mutants grown and expressed in the presence of Zn would be of interest as well. While the three datasets that have been analysed do not show significant differences between the non-metalated aggregate species, it would be beneficial to include more mutants for analysis to state with confidence that the same species is aggregating in both the reduced apo and Zn-bound species. Ideally, the entire 9-set of mutants analysed for SOD1 IBs expressed for 4 hours at 37 °C would be sufficient to confirm if this is indeed the case.

Finally, it would be of interest to see the expansion of the qHDX method, and its application on entirely new proteins and their aggregates using this method. As highlighted in Chapter 3, our group has begun analysing the structure of Adnectins that form IBs, but it would be of interest to look at entirely new systems. Examples of such systems could include the disordered C-terminal domain of TDB43, as well as myoglobin. qHDX observed through NMR is a powerful tool that can serve as a key piece in understanding the puzzle that is protein aggregation. This method could help elucidate protein local stability, global stability, and secondary structures within *in-vivo* aggregates. I hope to see the application of qHDX widespread in the field of protein aggregation.

References

1. Voet D, Voet JG. *Biochemistry*. 4th ed. John Wiley and Sons; 2010.
2. Englander SW, Mayne L. The nature of protein folding pathways. *Proc Natl Acad Sci U S A*. 2014;111(45):15873-15880. doi:10.1073/pnas.1411798111
3. Hartl FU, Hayer-Hartl M. Converging concepts of protein folding in vitro and in vivo. *Nat Struct Mol Biol*. 2009;16(6):574-581. doi:10.1038/nsmb.1591
4. Levinthal C. Are there pathways for protein folding? *J Chim Phys*. 1968;65:44-45. doi:10.1051/jcp/1968650044
5. Anfinsen CB. Principles that Govern the Folding of Protein Chains. *Science*. 1973;181(4096):223-230. www.jstor.org/stable/1736447
6. Anfinsen CB, Haber E, Sela M, White FH. The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain. *Proc Natl Acad Sci U S A*. 1961;47(9):1309-1314. doi:10.1073/pnas.47.9.1309
7. Daggett V, Fersht A. The present view of the mechanism of protein folding. *Nat Rev Mol Cell Biol*. 2003;4(6):497-502. doi:10.1038/nrm1126
8. Dobson CM. Protein folding and misfolding. *Nature*. 2003;426:884-890. doi:https://doi.org/10.1038/nature02261
9. Ptitsyn OB. Protein folding: Hypotheses and experiments. *J Protein Chem*. 1987;6(4):273-293. doi:10.1007/BF00248050
10. Broom HR, Rumfeldt JAO, Meiering EM. Many roads lead to Rome? Multiple modes of Cu,Zn superoxide dismutase destabilization, misfolding and aggregation in amyotrophic lateral sclerosis. *Essays Biochem*. 2014;56(1):149-165. doi:10.1042/BSE0560149
11. Valentine JS, Doucette PA, Zittin Potter S. Copper-Zinc Superoxide Dismutase and Amyotrophic Lateral Sclerosis. *Annu Rev Biochem*. 2005;74(1):563-593. doi:10.1146/annurev.biochem.72.121801.161647
12. Eisenberg DS, Sawaya MR. Structural Studies of Amyloid Proteins at the Molecular Level. *Annu Rev Biochem*. 2017;86(1):69-95. doi:10.1146/annurev-biochem-061516-045104
13. Fink AL. Protein aggregation: Folding aggregates, inclusion bodies and amyloid. *Fold Des*. 1998;3(1):9-23. doi:10.1016/S1359-0278(98)00002-9
14. de Groot NS, Sabate R, Ventura S. Amyloids in bacterial inclusion bodies. *Trends Biochem Sci*. 2009;34(8):408-416. doi:10.1016/j.tibs.2009.03.009
15. García-Fruitós E, Vázquez E, Díez-Gil C, et al. Bacterial inclusion bodies: Making gold from waste. *Trends Biotechnol*. 2012;30(2):65-70. doi:10.1016/j.tibtech.2011.09.003
16. Singhvi P, Saneja A, Srichandan S, Panda AK. Bacterial Inclusion Bodies: A Treasure Trove of Bioactive Proteins. *Trends Biotechnol*. 2020;38(5):474-486. doi:10.1016/j.tibtech.2019.12.011

17. Rinas U, Garcia-Fruitós E, Corchero JL, Vázquez E, Seras-Franzoso J, Villaverde A. Bacterial Inclusion Bodies: Discovering Their Better Half. *Trends Biochem Sci*. 2017;42(9):726-737. doi:10.1016/j.tibs.2017.01.005

18. Cano-Garrido O, Sánchez-Chardi A, Parés S, et al. Functional protein-based nanomaterial produced in microorganisms recognized as safe: A new platform for biotechnology. *Acta Biomater*. 2016;43:230-239. doi:10.1016/j.actbio.2016.07.038

19. Ramón A, Señorale-Pose M, Marín M. Inclusion bodies: Not that bad... *Front Microbiol*. 2014;5(FEB):2010-2015. doi:10.3389/fmicb.2014.00056

20. Jäger VD, Kloss R, Grünberger A, et al. Tailoring the properties of (catalytically)-active inclusion bodies. *Microb Cell Fact*. 2019;18(1):1-20. doi:10.1186/s12934-019-1081-5

21. Schetters STT, Jong WSP, Kruijssen LJW, et al. Bacterial inclusion bodies function as vehicles for dendritic cell-mediated T cell responses. *Cell Mol Immunol*. 2020;17(4):415-417. doi:10.1038/s41423-019-0298-x

22. Invernizzi G, Aprile FA, Natalello A, et al. The Relationship between Aggregation and Toxicity of Polyglutamine-Containing Ataxin-3 in the Intracellular Environment of Escherichia coli. *PLoS One*. 2012;7(12). doi:10.1371/journal.pone.0051890

23. Sabaté R, Espargaró A, Saupe SJ, Ventura S. Characterization of the amyloid bacterial inclusion bodies of the HET-s fungal prion. *Microb Cell Fact*. 2009;8:1-10. doi:10.1186/1475-2859-8-56

24. Trainor K, Gingras Z, Shillingford C, et al. Ensemble Modeling and Intracellular Aggregation of an Engineered Immunoglobulin-Like Domain. *J Mol Biol*. 2016;428(6):1365-1374. doi:10.1016/j.jmb.2016.02.016

25. Sekhar A, Rumfeldt JAO, Broom HR, et al. Thermal fluctuations of immature SOD1 lead to separate folding and misfolding pathways. *Elife*. 2015;4(JUNE):1-33. doi:10.7554/eLife.07296

26. Wang Q, Johnson JL, Agar NYR, Agar JN. Protein aggregation and protein instability govern familial amyotrophic lateral sclerosis patient survival. *PLoS Biol*. 2008;6(7):1508-1526. doi:10.1371/journal.pbio.0060170

27. Vaquer-Alicea J, Diamond MI. Propagation of Protein Aggregation in Neurodegenerative Diseases. *Annu Rev Biochem*. 2019;88:785-810. doi:10.5772/intechopen.81224

28. Nonaka T, Masuda-Suzukake M, Arai T, et al. Prion-like Properties of Pathological TDP-43 Aggregates from Diseased Brains. *Cell Rep*. 2013;4(1):124-134. doi:10.1016/j.celrep.2013.06.007

29. Macedo B, Sant'Anna R, Navarro S, Cordeiro Y, Ventura S. Mammalian prion protein (PrP) forms conformationally different amyloid intracellular aggregates in bacteria. *Microb Cell Fact*. 2015;14(1):1-16. doi:10.1186/s12934-015-0361-y

30. Ayers JI, Diamond J, Sari A, et al. Distinct conformers of transmissible misfolded SOD1 distinguish human SOD1-FALS from other forms of familial and sporadic ALS. *Acta Neuropathol*. 2016;132(6):827-840. doi:doi:10.1007/s00401-016-1623-4.

31. Ayers JI, Fromholt S, Koch M, et al. Experimental transmissibility of mutant SOD1 motor neuron disease. *Acta Neuropathol*. 2014;128(6):791-803. doi:10.1007/s00401-014-1342-7

32. Ayers JI, Fromholt SE, O'Neal VM, Diamond JH, Borchelt DR. Prion-like propagation of mutant SOD1 misfolding and motor neuron disease spread along neuroanatomical pathways. *Acta Neuropathol*. 2016;131(1):103-114. doi:10.1007/s00401-015-1514-0

33. Lang L, Zetterström P, Brännström T, et al. SOD1 aggregation in ALS mice shows simplistic test tube behavior. *Proc Natl Acad Sci U S A*. 2015;112(32):9878-9883. doi:10.1073/pnas.1503328112

34. Rakhit R, Chakrabartty A. Structure, folding, and misfolding of Cu,Zn superoxide dismutase in amyotrophic lateral sclerosis. *Biochim Biophys Acta - Mol Basis Dis*. 2006;1762(11-12):1025-1037. doi:10.1016/j.bbadis.2006.05.004

35. McCord J, Fridovich I. Superoxide Dismutase an enzymatic function for Erythrocuprein (hemocuprien). *J Biol Chem*. 1969;244(22).

36. Deol H. *Characterizing the Dimer Interface and Thermal Unfolding of Reduced, Zinc Bound Superoxide Dismutase-1 (Master's Thesis)*.; 2018. https://uwspace.uwaterloo.ca/handle/10012/13203

37. Banci L, Bertini I, Cantini F, et al. Human superoxide dismutase 1 (hSOD1) maturation through interaction with human copper chaperone for SOD1 (hCCS). *Proc Natl Acad Sci U S A*. 2012;109(34):13555-13560. doi:10.1073/pnas.1207493109

38. Svensson AKE, Bilsel O, Kayatekin C, Adefusika JA, Zitzewitz JA, Robert Matthews C. Metal-free ALS variants of dimeric human Cu,Zn-superoxide dismutase have enhanced populations of monomeric species. *PLoS One*. 2010;5(4). doi:10.1371/JOURNAL.PONE.0010064

39. Stathopulos PB, Rumfeldt JAO, Karbassi F, Siddall CA, Lepock JR, Meiering EM. Calorimetric analysis of thermodynamic stability and aggregation for Apo and Holo amyotrophic lateral sclerosis-associated Gly-93 mutants of superoxide dismutase. *J Biol Chem*. 2006;281(10):6184-6193. doi:10.1074/jbc.M509496200

40. Vassall KA, Stubbs HR, Primmer HA, et al. Decreased stability and increased formation of soluble aggregates by immature superoxide dismutase do not account for disease severity in ALS. *Proc Natl Acad Sci U S A*. 2011;108(6):2210-2215. doi:10.1073/pnas.0913021108

41. Chan PK, Chattopadhyay M, Sharma S, et al. Structural similarity of wild-type and ALS-mutant superoxide dismutase-1 fibrils using limited proteolysis and atomic force microscopy. *Proc Natl Acad Sci U S A*. 2013;110(27):10934-10939. doi:10.1073/pnas.1309613110

42. Furukawa Y, Kaneko K, Yamanaka K, Nukina N. Mutation-dependent polymorphism of Cu,Zn-superoxide dismutase aggregates in the familial form of amyotrophic lateral sclerosis. *J Biol Chem*. 2010;285(29):22221-22231. doi:10.1074/jbc.M110.113597

43. Hwang YM, Stathopulos PB, Dimmick K, et al. Nonamyloid aggregates arising from mature copper/zinc superoxide dismutases resemble those observed in amyotrophic lateral

sclerosis. *J Biol Chem*. 2010;285(53):41701-41711. doi:10.1074/jbc.M110.113696

44. Rule G, Hitchens TK. *Fundamentals of Protein NMR Spectroscopy*. 5th ed. Springer Netherlands; 2006.

45. Doucleff M, Hatcher-Skeers M, Crane NJ. *Pocket Guide to Biomolecular NMR*.; 2011. doi:10.1017/CBO9781107415324.004

46. Zhang Y, Roder H, Paterson Y. Rapid amide proton exchange rates in peptides and proteins measured by solvent quenching and two-dimensional NMR. *Protein Sci*. 1995;4(4):804-814. doi:10.1002/pro.5560040420

47. Hoshino M, Katou H, Hagihara Y, Hasegawa K, Naiki H, Goto Y. Mapping the core of the β2-microglobulin amyloid fibril by H/D exchange. *Nat Struct Biol*. 2002;9(5):332-336. doi:10.1038/nsb792

48. Alexandrescu AT. An NMR-based quenched hydrogen exchange investigation of model amyloid fibrils formed by cold shock protein A. *Pac Symp Biocomput*. 2001;78:67-78. doi:10.1142/9789814447362_0008

49. Alexandrescu AT. Quenched Hydrogen Exchange NMR of Amyloid Fibrils. *Methods Mol Biol*. 2016;1345:211-222. doi:10.1007/978-1-4939-2978-8_14

50. Hoshino M, Katou H, Yamaguchi K, Goto Y. Dimethylsulfoxide-quenched hydrogen/deuterium exchange method to study amyloid fibril structure. *Biochim Biophys Acta - Biomembr*. 2007;1768(8):1886-1899. doi:10.1016/j.bbamem.2007.03.001

51. Wang L, Maji SK, Sawaya MR, Eisenberg D, Riek R. Bacterial inclusion bodies contain amyloid-like structure. *PLoS Biol*. 2008;6(8):1791-1801. doi:10.1371/journal.pbio.0060195

52. Wasmer C, Benkemoun L, Sabaté R, et al. Solid-state NMR spectroscopy reveals that E. coli inclusion bodies of HET-s(218-289) are amyloids. *Angew Chemie - Int Ed*. 2009;48(26):4858-4860. doi:10.1002/anie.200806100

53. Lipovšek D. Adnectins: engineered target-binding protein therapeutics. *Protein Eng Des Sel*. 2011;24(1-2):3-9. doi:10.1093/protein/gzq097

54. Kay LE, Ikura M, Tschudin R, Bax A. Three-dimensional triple-resonance NMR spectroscopy of isotopically enriched proteins. *J Magn Reson*. 1990;89(3):496-514. doi:10.1016/0022-2364(90)90333-5

55. Grzesiek S, Bax A. Correlating backbone amide and side chain resonances in larger proteins by multiple relayed triple resonance NMR. *J Am Chem Soc*. 2002;114(16):6291-6293. doi:10.1021/JA00042A003

56. Muhandiram DR, Kay LE. Gradient-Enhanced Triple-Resonance Three-Dimensional NMR Experiments with Improved Sensitivity. *J Magn Reson Ser B*. 1994;103(3):203-216. doi:10.1006/JMRB.1994.1032

57. Clubb RT, Thanabal V, Wagner G. A Constant-Time Three-Dimensional Triple-Resonance Pulse Scheme to Correlate Intraresidue 1H N, 15N, and 13C' Chemical Shifts

in 15N-13C-Labeled Proteins.

58. Wittekind M, Mueller L. HNCACB, a High-Sensitivity 3D NMR Experiment to Correlate Amide-Proton and Nitrogen Resonances with the Alpha- and Beta-Carbon Resonances in Proteins. *JMRB*. 1993;101(2):201-205. doi:10.1006/JMRB.1993.1033

59. Weisemann R, Rüterjans H, Bermel W. 3D Triple-resonance NMR techniques for the sequential assignment of NH and 15N resonances in 15N- and 13C-labelled proteins. *J Biomol NMR*. 1993;3(1):113-120. doi:10.1007/BF00242479

60. Skinner SP, Fogh RH, Boucher W, Ragan TJ, Mureddu LG, Vuister GW. CcpNmr AnalysisAssign: a flexible platform for integrated NMR analysis. *J Biomol NMR*. 2016;66(2):111-124. doi:10.1007/S10858-016-0060-Y

61. Damo SM, Phillips AH, Young AL, Li S, Woods VL, Wemmer DE. Probing the conformation of a prion protein fibril with hydrogen exchange. *J Biol Chem*. 2010;285(42):32303-32311. doi:10.1074/jbc.M110.114504

62. Vilar M, Wang L, Riek R. *Structural Studies of Amyloids by Quenched Hydrogen-Deuterium Exchange by NMR*. Vol 849.; 2012. doi:10.1007/978-1-61779-551-0_13

63. Bordwell FG. Equilibrium Acidities in Dimethyl Sulfoxide Solution. *Acc Chem Res*. 1988;21(12):456-463. doi:10.1021/ar00156a004

64. Beckner C, Kantor S, Schmitt R. Studies on Repository Compound Stability. *J Biomol Screen*. 2009;8(3):292-304. doi:10.1177/1087057103254366

65. Moorthy B, Iyer L, Topp E. Characterizing Protein Structure, Dynamics and Conformation in Lyophilized Solids. *Curr Pharm Des*. 2015;21(40):5845-5853. doi:10.2174/1381612821666151008150735

66. Moorthy BS, Schultz SG, Kim SG, Topp EM. Predicting protein aggregation during storage in lyophilized solids using solid state amide hydrogen/deuterium exchange with mass spectrometric analysis (ssHDX-MS). *Mol Pharm*. 2014;11(6):1869-1879. doi:10.1021/mp500005v

67. Wishart DS, Bigam CG, Holm A, Hodges RS, Sykes BD. 1H, 13C and 15N random coil NMR chemical shifts of the common amino acids. I. Investigations of nearest-neighbor effects. *J Biomol NMR 1995 51*. 1995;5(1):67-81. doi:10.1007/BF00227471

68. Karsisiotis AI, Deacon OM, Macdonald C, Blumenschein TMA, Moore GR, Worrall JAR. Near-complete backbone resonance assignments of acid-denatured human cytochrome c in dimethylsulfoxide: a prelude to studying interactions with phospholipids. *Biomol NMR Assign*. 2017;11(2):165-168. doi:10.1007/s12104-017-9740-0

69. Kulminskaya N V., Yoshimura Y, Runager K, et al. Near-complete 1H, 13C, 15N resonance assignments of dimethylsulfoxide-denatured TGFBIp FAS1-4 A546T. *Biomol NMR Assign*. 2016;10(1):25-29. doi:10.1007/s12104-015-9630-2

70. Chandak MS, Nakamura T, Makabe K, et al. The H/D-exchange kinetics of the Escherichia coli co-chaperonin GroES studied by 2D NMR and DMSO-quenched exchange methods. *J Mol Biol*. 2013;425(14):2541-2560. doi:10.1016/j.jmb.2013.04.008

71. Williams D, Van Frank R, Muth W, Burnett J. Cytoplasmic inclusion bodies in Escherichia coli producing biosynthetic human insulin proteins. *Science (80-).* 1982;215(4533):687-689. doi:10.1126/SCIENCE.7036343

72. Chrunyk BA, Evans J, Lillquist J, Young P, Wetzel R. Inclusion body formation and protein stability in sequence variants of interleukin-1 beta. *J Biol Chem*. 1993;268(24):18053-18061. doi:10.1016/S0021-9258(17)46810-4

73. Mitraki A, King J. Protein Folding Intermediates and Inclusion Body Formation. *Nat Biotechnol*. 1989;7(7):690-697. doi:10.1038/nbt0789-690

74. Rinas U, Garcia-Fruitós E, Corchero JL, Vázquez E, Seras-Franzoso J, Villaverde A. Bacterial Inclusion Bodies: Discovering Their Better Half. *Trends Biochem Sci*. 2017;42(9):726-737. doi:10.1016/j.tibs.2017.01.005

75. Ke PC, Zhou R, Serpell LC, et al. Half a century of amyloids: Past, present and future. *Chem Soc Rev*. 2020;49(15):5473-5509. doi:10.1039/c9cs00199a

76. Krauss U, Jäger VD, Diener M, Pohl M, Jaeger KE. Catalytically-active inclusion bodies—Carrier-free protein immobilizates for application in biotechnology and biomedicine. *J Biotechnol*. 2017;258(February 2017):136-147. doi:10.1016/j.jbiotec.2017.04.033

77. Céspedes MV, Cano-Garrido O, Álamo P, et al. Engineering Secretory Amyloids for Remote and Highly Selective Destruction of Metastatic Foci. *Adv Mater*. 2020;32(7):1-9. doi:10.1002/adma.201907348

78. Chiti F, Dobson CM. Protein Misfolding, Amyloid Formation, and Human Disease: A Summary of Progress Over the Last Decade. *Annu Rev Biochem*. 2017;86:27-68. doi:10.1146/ANNUREV-BIOCHEM-061516-045115

79. Broom HR, Rumfeldt JAO, Meiering EM. Many roads lead to Rome? Multiple modes of Cu,Zn superoxide dismutase destabilization, misfolding and aggregation in amyotrophic lateral sclerosis. *Essays Biochem*. 2014;56:149-165. doi:10.1042/bse0560149

80. Wright GSA, Antonyuk S V., Hasnain SS. The biophysics of superoxide dismutase-1 and amyotrophic lateral sclerosis. *Q Rev Biophys*. 2019;52:e12. doi:10.1017/S003358351900012X

81. Furukawa Y, O'Halloran T V. Amyotrophic lateral sclerosis mutations have the greatest destabilizing effect on the apo- and reduced form of SOD1, leading to unfolding and oxidative aggregation. *J Biol Chem*. 2005;280(17):17266-17274. doi:10.1074/jbc.M500482200

82. Calloni G, Zoffoli S, Stefani M, Dobson CM, Chiti F. Investigating the effects of mutations on protein aggregation in the cell. *J Biol Chem*. 2005;280(11):10607-10613. doi:10.1074/jbc.M412951200

83. Castillo V, Espargaró A, Gordo V, Vendrell J, Ventura S. Deciphering the role of the thermodynamic and kinetic stabilities of SH3 domains on their aggregation inside bacteria. *Proteomics*. 2010;10(23):4172-4185. doi:10.1002/pmic.201000260

84. Broom HR, Rumfeldt JAO, Vassall KA, Meiering EM. Destabilization of the dimer interface is a common consequence of diverse ALS-associated mutations in metal free SOD1. *Protein Sci*. 2015;24(12). doi:10.1002/pro.2803

85. Vassall KA, Stathopulos PB, Rumfeldt JAO, Lepock JR, Meiering EM. Equilibrium Thermodynamic Analysis of Amyotrophic Lateral Sclerosis-Associated Mutant Apo Cu,Zn Superoxide Dismutases. *Biochemistry*. 2006;45(23):7366-7379.

86. Fernandez-Escamilla AM, Rousseau F, Schymkowitz J, Serrano L. Prediction of sequence-dependent and mutational effects on the aggregation of peptides and proteins. *Nat Biotechnol*. 2004;22(10):1302-1306. doi:10.1038/nbt1012

87. Thompson MJ, Sievers SA, Karanicolas J, Ivanova MI, Baker D, Eisenberg D. The 3D profile method for identifying fibril-forming segments of proteins. *Proc Natl Acad Sci U S A*. 2006;103(11):4074-4078. doi:10.1073/pnas.0511295103

88. Sormanni P, Aprile FA, Vendruscolo M. The CamSol Method of Rational Design of Protein Mutants with Enhanced Solubility. *J Mol Biol*. 2015;427(2):478-490. doi:10.1016/J.JMB.2014.09.026

89. Trainor K, Broom A, Meiering EM. Exploring the relationships between protein sequence, structure and solubility. *Curr Opin Struct Biol*. 2017;42:136-146. doi:10.1016/j.sbi.2017.01.004

90. Kyte J, Doolittle RF. A simple method for displaying the hydropathic character of a protein. *J Mol Biol*. 1982;157(1):105-132. doi:10.1016/0022-2836(82)90515-0

91. Källberg M, Wang H, Wang S, et al. Template-based protein structure modeling using the RaptorX web server. *Nat Protoc*. 2012;7(8):1511-1522. doi:10.1038/nprot.2012.085

92. Sormanni P, Aprile FA, Vendruscolo M. The CamSol method of rational design of protein mutants with enhanced solubility. *J Mol Biol*. 2015;427(2):478-490. doi:10.1016/j.jmb.2014.09.026

93. Wang Z, Zhao F, Peng J, Xu J. Protein 8-class secondary structure prediction using conditional neural fields. *Proteomics*. 2011;11(19):3786. doi:10.1002/PMIC.201100196

94. Naser D, Tarasca M V, Siebeneichler B, Schaefer A, Deol HK. High Resolution NMR H / D Exchange of Human Cu , Zn Superoxide Dismutase Inclusion Bodies Reveals Significant Native Features Despite Structural Heterogeneity. *Under Rev Angew Chemie - Int Ed*. Published online 2021.

95. Wang L, Schubert D, Sawaya MR, Eisenberg D, Riek R. Multidimensional structure-activity relationship of a protein in its aggregated states. *Angew Chemie - Int Ed*. 2010;49(23):3904-3908. doi:10.1002/anie.201000068

96. Museth AK, Brorsson A-C, Lundqvist M, Tibell LAE, Jonsson B-HB. The ALS-associated mutation G93A in human copper-zinc superoxide dismutase selectively destabilizes the remote metal binding region. *Biochemistry*. 2009;48(37):8817-8829. doi:10.1021/BI900703V

97. Ivanova MI, Sievers SA, Guenther EL, et al. Aggregation-triggering segments of SOD1

fibril formation support a common pathway for familial and sporadic ALS. *Proc Natl Acad Sci U S A*. 2014;111(1):197-201. doi:10.1073/pnas.1320786110

98. Elam JS, Taylor AB, Strange R, et al. Amyloid-like filaments and water-filled nanotubes formed by SOD1 mutant proteins linked to familial ALS. *Nat Struct Biol*. 2003;10(6):461-467. doi:10.1038/nsb935

99. Banci L, Bertini I, Cramaro F, Del Conte R, Viezzoli MS. Solution Structure of Apo Cu,Zn Superoxide Dismutase: Role of Metal Ions in Protein Folding †. *Biochemistry*. 2003;42(32):9543-9553. doi:10.1021/bi034324m

100. Sangwan S, Zhao A, Adams KL, et al. Atomic structure of a toxic, oligomeric segment of SOD1 linked to amyotrophic lateral sclerosis (ALS). *Proc Natl Acad Sci U S A*. 2017;114(33):8770-8775. doi:10.1073/pnas.1705091114

101. Goldschmidt L, Teng PK, Riek R, Eisenberg D. Identifying the amylome, proteins capableof forming amyloid-like fibrils. Published online 2010. doi:doi/10.1073/pnas.0915166107

102. Strange RW, Antonyuk S, Hough MA, et al. The structure of holo and metal-deficient wild-type human Cu, Zn superoxide dismutase and its relevance to familial amyotrophic lateral sclerosis. *J Mol Biol*. 2003;328(4):877-891. doi:10.1016/S0022-2836(03)00355-3

103. Nordlund A, Leinartaitė L, Saraboji K, et al. Functional features cause misfolding of the ALS-provoking enzyme SOD1. *Proc Natl Acad Sci*. 2009;106(24):9667-9672. doi:10.1073/PNAS.0812046106

104. Sekhar A, Rumfeldt JAO, Broom HR, et al. Probing the free energy landscapes of ALS disease mutants of SOD1 by NMR spectroscopy. *Proc Natl Acad Sci U S A*. 2016;113(45):E6939-E6945. doi:10.1073/pnas.1611418113

105. Serio T, Cashikar A, Kowal A, Sawicki G, Moslehi J, Lindquist S. Nucleated Conformational Conversion and the Replication of Conformational Information by a Prion Determinant. *Am Assoc Adv Sci*. 200AD;289. http://www.jstor.org/stable/3077621%0D

106. Stathopulos PB, Scholz GA, Hwang Y-M, Rumfeldt JAO, Lepock JR, Meiering EM. Sonication of proteins causes formation of aggregates that resemble amyloid. *Protein Sci*. 2008;13(11):3017-3027. doi:10.1110/ps.04831804

107. Andjus P, Stamenković S, Dučić T. Synchrotron radiation-based FTIR spectro-microscopy of the brainstem of the hSOD1 G93A rat model of amyotrophic lateral sclerosis. *Eur Biophys J*. 2019;48(5):475-484. doi:10.1007/s00249-019-01380-5

108. Barth A. Infrared spectroscopy of proteins. *Biochim Biophys Acta - Bioenerg*. 2007;1767(9):1073-1101. doi:10.1016/j.bbabio.2007.06.004

109. Sarroukh R, Goormaghtigh E, Ruysschaert JM, Raussens V. ATR-FTIR: A "rejuvenated" tool to investigate amyloid proteins. *Biochim Biophys Acta - Biomembr*. 2013;1828(10):2328-2338. doi:10.1016/j.bbamem.2013.04.012

110. Annamalai K, Liberta F, Vielberg MT, et al. Common Fibril Structures Imply Systemically Conserved Protein Misfolding Pathways In Vivo. *Angew Chemie - Int Ed*.

2017;56(26):7510-7514. doi:10.1002/anie.201701761

111. Levitt M, Greer J. Automatic identification of secondary structure in globular proteins. *J Mol Biol*. 1977;114(2):181-239. doi:10.1016/0022-2836(77)90207-8

112. Klunk WE, Jacob RF, Mason RP. Quantifying amyloid by congo red spectral shift assay. *Methods Enzymol*. 1999;309(1974):285-305. doi:10.1016/S0076-6879(99)09021-7

113. Kerman A, Liu HN, Croul S, et al. Amyotrophic lateral sclerosis is a non-amyloid disease in which extensive misfolding of SOD1 is unique to the familial form. *Acta Neuropathol*. 2010;119(3):335-344. doi:10.1007/s00401-010-0646-5

114. Banci L, Barbieri L, Bertini I, Cantini F, Luchinat E. In-cell NMR in E. coli to monitor maturation steps of hSOD1. *PLoS One*. 2011;6(8):1-8. doi:10.1371/journal.pone.0023561

115. Wu W, Xing L, Zhou B, Lin Z. Active protein aggregates induced by terminally attached self-assembling peptide ELK16 in Escherichia coli. *Microb Cell Fact*. 2011;10:1-8. doi:10.1186/1475-2859-10-9

116. De Marco A, Ferrer-Miralles N, Garcia-Fruitós E, et al. Bacterial inclusion bodies are industrially exploitable amyloids. *FEMS Microbiol Rev*. 2019;43(1):53-72. doi:10.1093/femsre/fuy038

117. Fändrich M, Schmidt M. Methods to study the structure of misfolded protein states in systemic amyloidosis. *Biochem Soc Trans*. 2021;49(2):977-985. doi:10.1042/BST20201022

118. Ruggeri FS, Longo G, Faggiano S, Lipiec E, Pastore A, Dietler G. Infrared nanospectroscopy characterization of oligomeric and fibrillar aggregates during amyloid formation. *Nat Commun*. 2015;6(7831). doi:10.1038/ncomms8831

119. Röhr D, Boon BDC, Schuler M, et al. Label-free vibrational imaging of different Aβ plaque types in Alzheimer's disease reveals sequential events in plaque development. *Acta Neuropathol Commun*. 2020;8(1). doi:10.1186/s40478-020-01091-5

120. Ghosh U, Thurber KR, Yau WM, Tycko R. Molecular structure of a prevalent amyloid-β fibril polymorph from Alzheimer's disease brain tissue. *Proc Natl Acad Sci U S A*. 2021;118(4). doi:10.1073/pnas.2023089118

121. Bergh J, Zetterström P, Andersen PM, et al. Structural and kinetic analysis of protein-aggregate strains in vivo using binary epitope mapping. *Proc Natl Acad Sci U S A*. 2015;112(14):4489-4494. doi:10.1073/pnas.1419228112

122. Smirnovas V, Baron GS, Offerdahl DK, Raymond GJ, Caughey B, Surewicz WK. Structural organization of brain-derived mammalian prions examined by hydrogen-deuterium exchange. *Nat Struct Mol Biol*. 2011;18(4):504-506. doi:10.1038/nsmb.2035

123. Li Q, Wang F, Xiao X, et al. Structural attributes of mammalian prion infectivity: Insights from studies with synthetic prions. *J Biol Chem*. 2018;293(48):18494-18503. doi:10.1074/jbc.RA118.005622

124. Donnelly DJ, Adam Smith R, Morin P, et al. *Synthesis and Biologic Evaluation of a Novel*

18 F-Labeled Adnectin as a PET Radioligand for Imaging PD-L1 Expression. Vol 59.; 2018. doi:10.2967/jnumed.117.199596

125. Chandler PG, Buckle AM. Development and Differentiation in Monobodies Based on the Fibronectin Type 3 Domain. *Cells*. 2020;9(3). doi:10.3390/cells9030610

126. Waterhouse A, Bertoni M, Bienert S, et al. SWISS-MODEL: Homology modelling of protein structures and complexes. *Nucleic Acids Res*. 2018;46(W1):W296-W303. doi:10.1093/nar/gky427

127. Bai Y, Milne JS, Mayne L, Englander SW. Primary Structure Effects on Peptide Group Hydrogen Exchange. *Proteins*. 1993;17(1). doi:10.1002/prot.340170110

Appendix

Appendix-Chapter 2

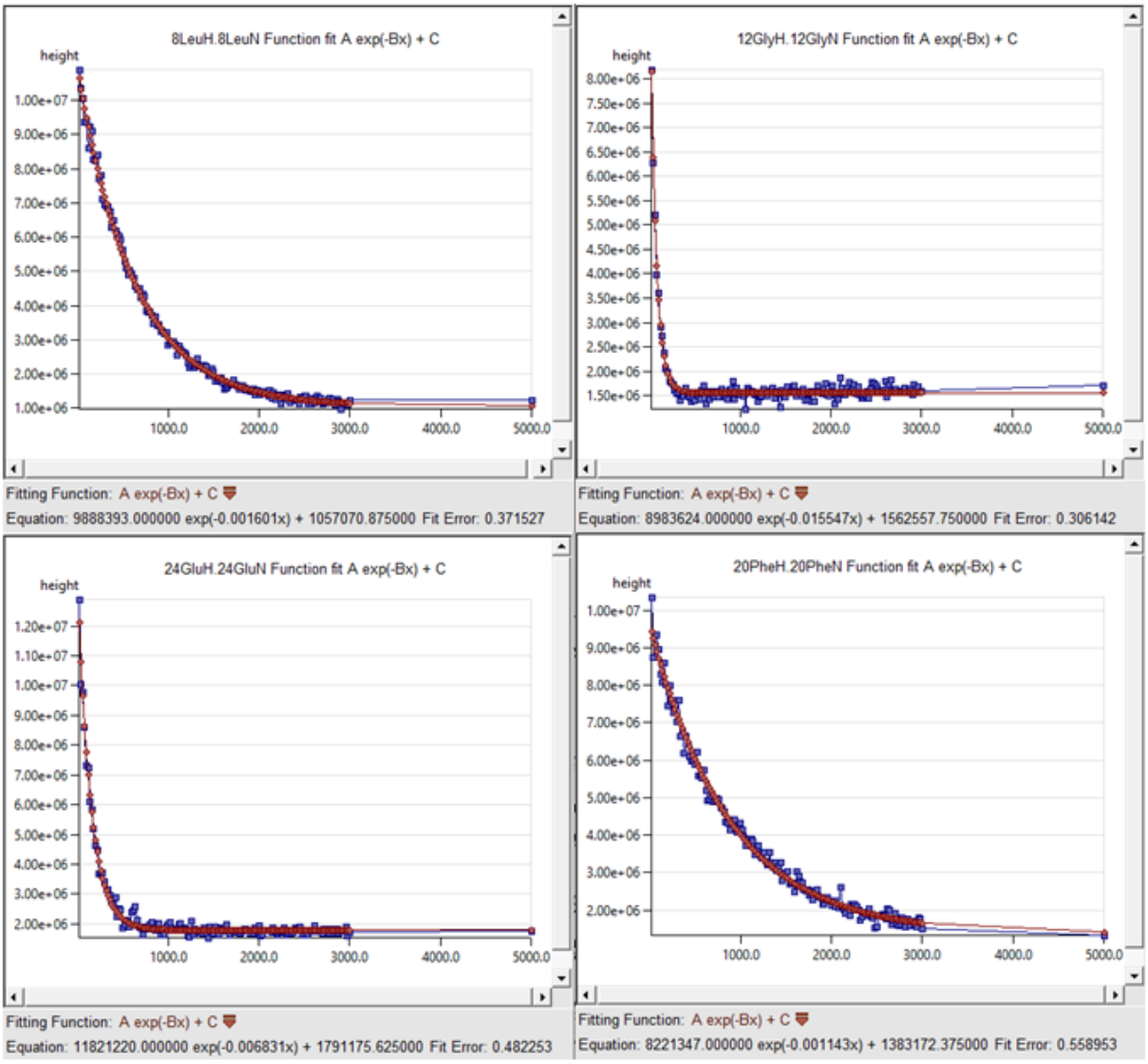


Figure A2.1: Amide exchange decay fits in CCPTNMR software.
CCPNMR produces an output of the intensities of each assigned cross peak in the series of ^{1}H-^{15}N HSQC spectra fit as a function of time. Here, a series of 150 ^{1}H-^{15}N HSQC spectra were obtained, assignments were transferred automatically throughout the series after manual transfer to one spectrum. The output produces a fit in the form of ($I = I_o * e^{(-k*x)} + C$), with I being the intensity at time x, I_o being the initial intensity, -k being the k_{int}, and C being the signal of the cross peak at HD exchange equilibrium.

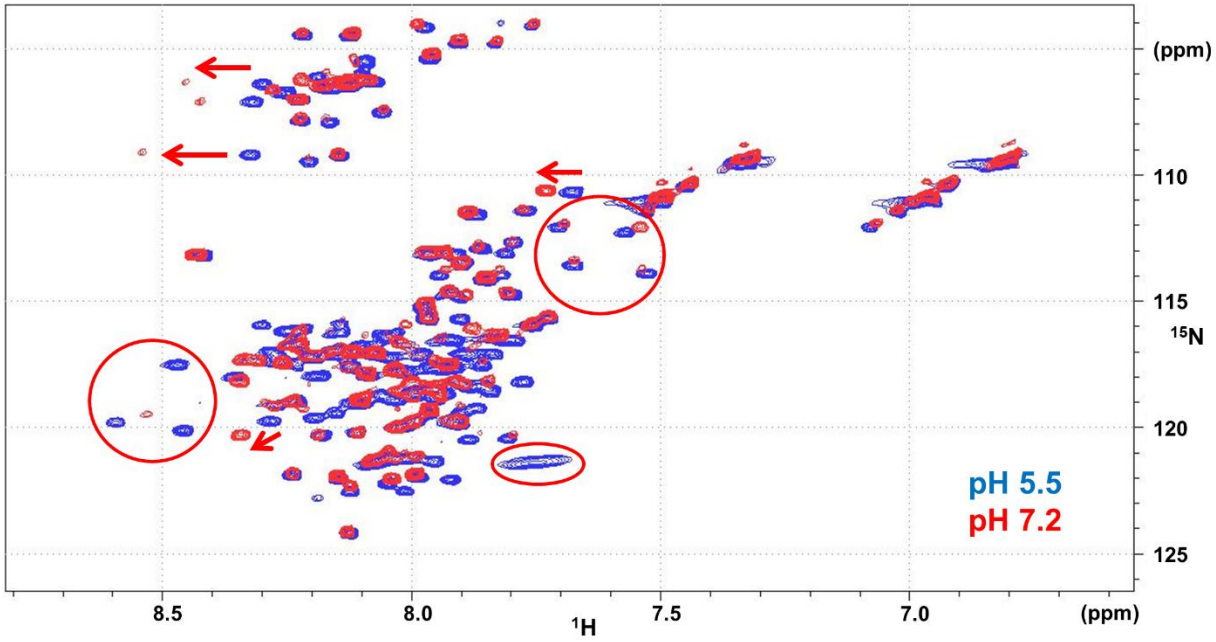


Figure A2.2: Effects of pH on ^{1}H-^{15}N HSQC cross peak intensity and chemical shift.
Replicate samples with different pH values. Areas where signal intensities differ greatly are circled, and areas where there are larger differences in chemical shifts are marked with arrows. The random coil region (~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.

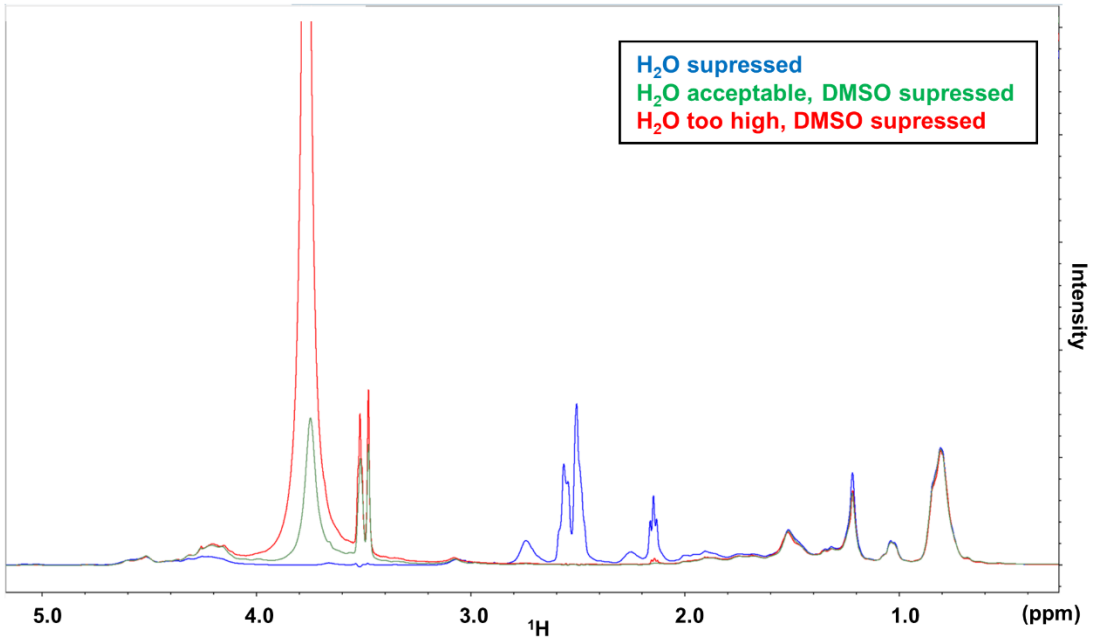


Figure A2.3: ^{1}H 1D NMR spectra overlaid to highlight differences in water content across samples.

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

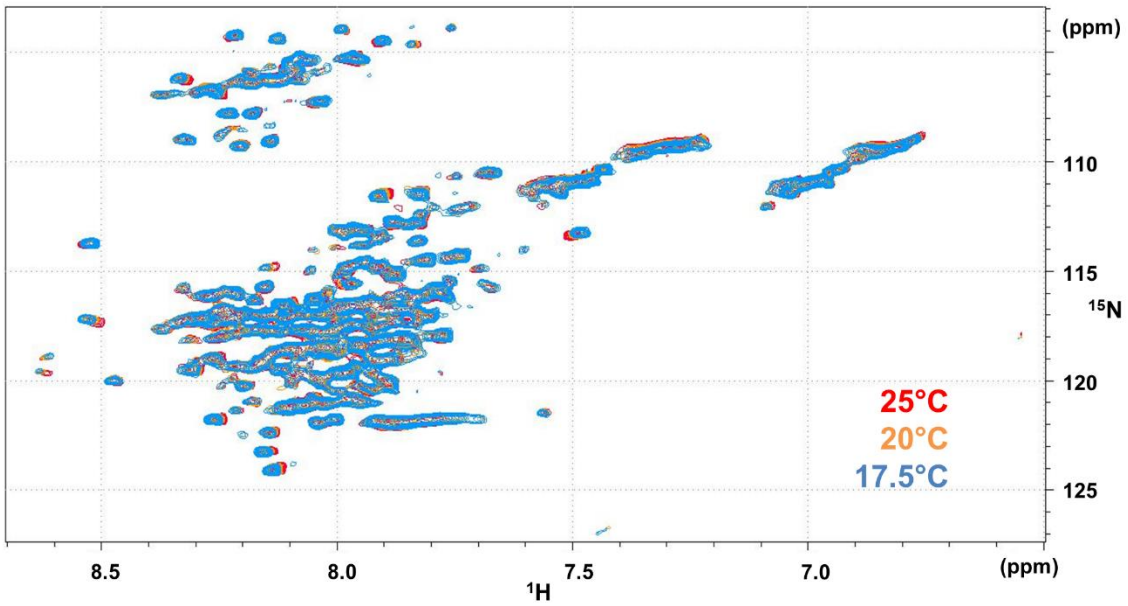


Figure A2.4: Overlay of ^{1}H-^{15}N spectra at different acquisition temperatures.
A series of temperature controlled ^{1}H-^{15}N HSQC spectra show that cross peaks move mostly uniformly in the unfolded protein, allowing for resonance assignments of a spectrum measured at 25 °C to be easily transferred to one measured at lower temperatures, being 19 °C for our experiments.

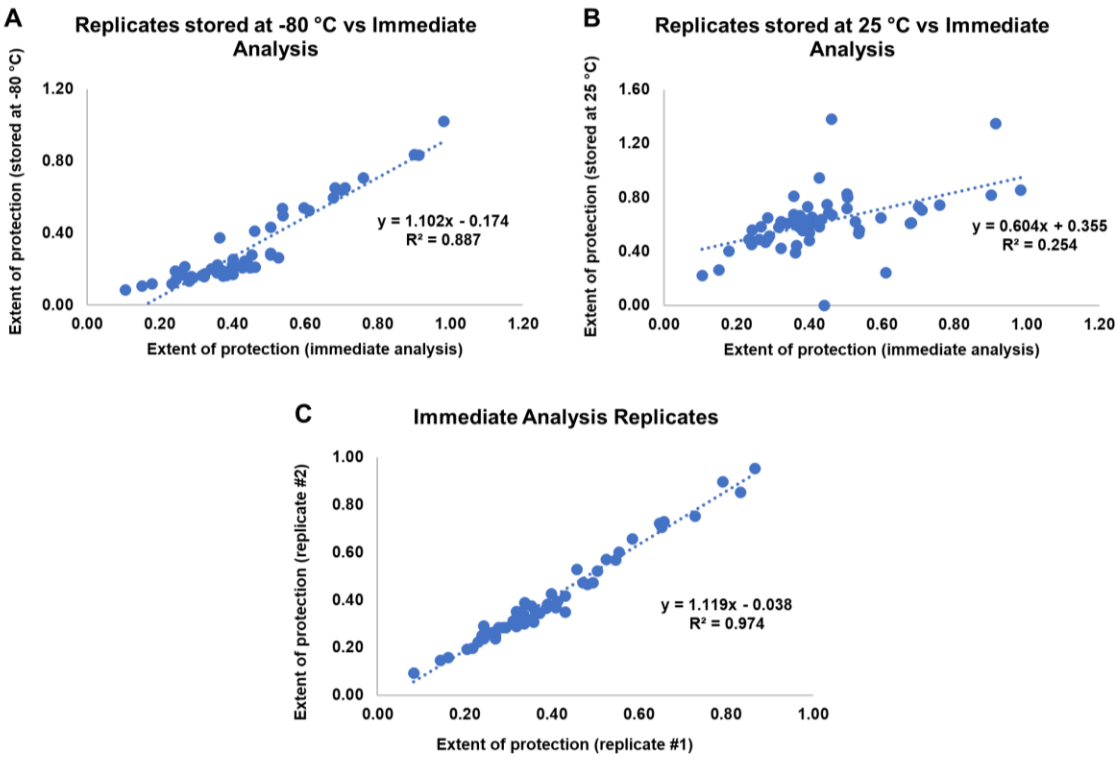


Figure A2.5: Effects of varying IB storage.
A) Replicate samples, one analyzed immediately after lyophilization compared with one analyzed after storing the sample at -80 °C for three months. Some changes are observed, notably for residues with lower protection. B) Replicate samples analyzed immediately 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.

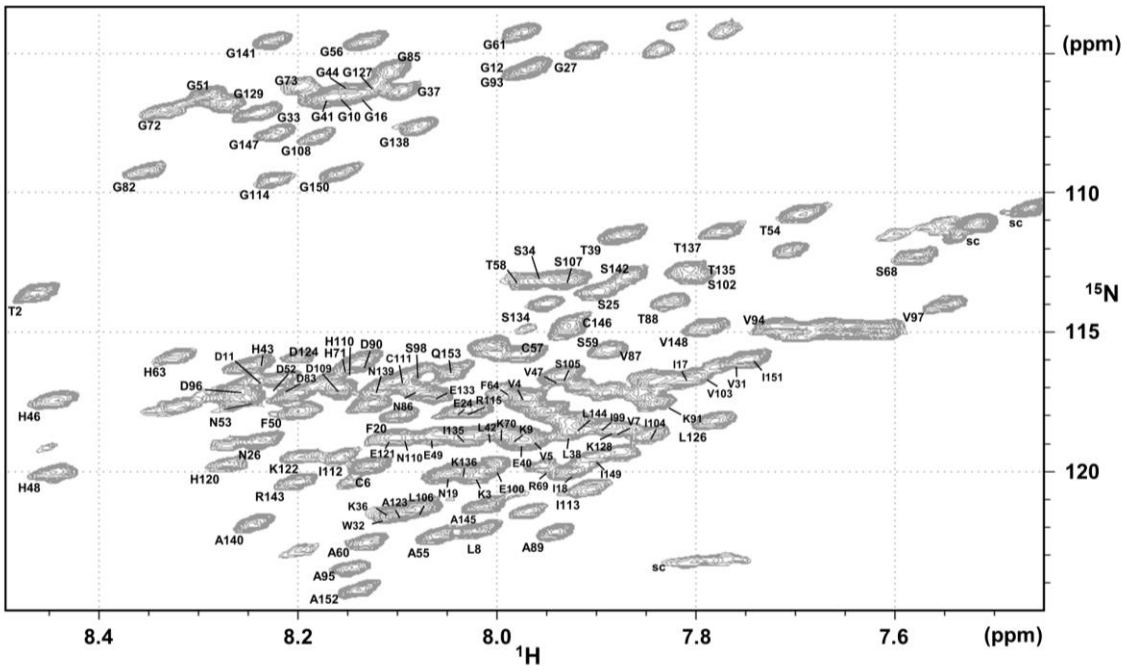


Figure A2.6: ^{1}H-^{15}N spectrum of A4V SOD1 unfolded in 95% DMSO / 5% H_2O at pH 5.5 with backbone residue assignments.

Sequence specific resonance assignments made at 25 °C as described in 2.5.3. Assignments were obtained for 126 of 148 non-Pro residues. Unassigned cross peaks for sidechains are labelled as 'sc'. Note the low chemical shift dispersion in the ^{1}H dimension relative to a folded protein.

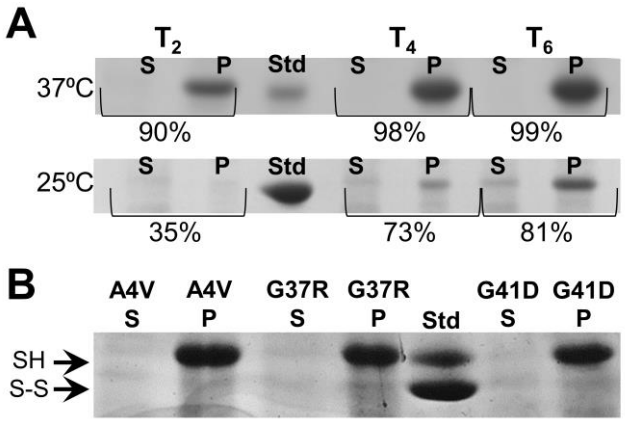


Figure A3.1: SDS-PAGE of mutant SOD1 IB formation and disulphide status.
(A) Samples of cell culture taken at 2, 4, and 6 hours post induction at 37 °C (top) and 25 °C (bottom) were lysed and centrifuged, then soluble (S) and insoluble (P) fractions were analyzed by 15% SDS-PAGE. Gel densitometry was 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 x 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

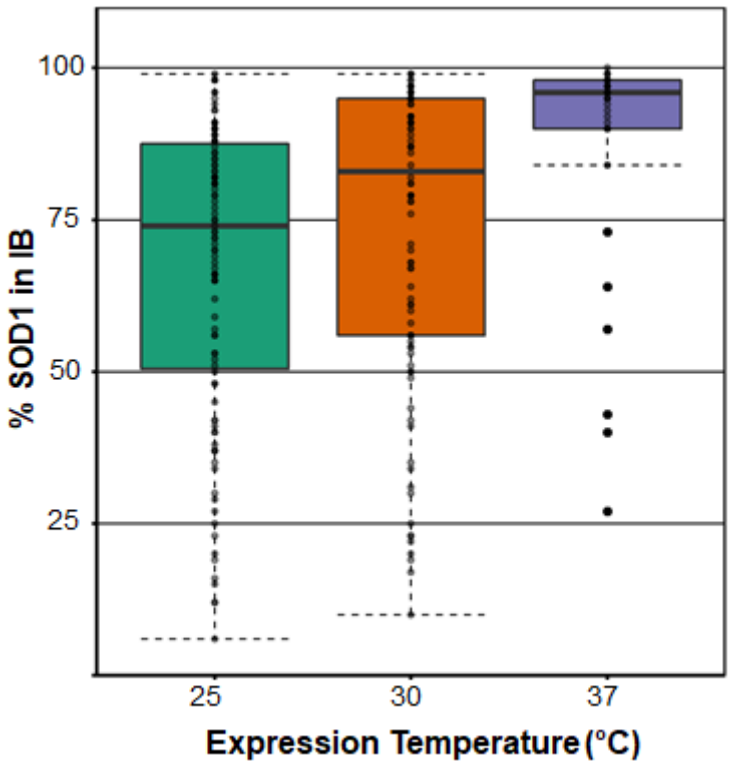
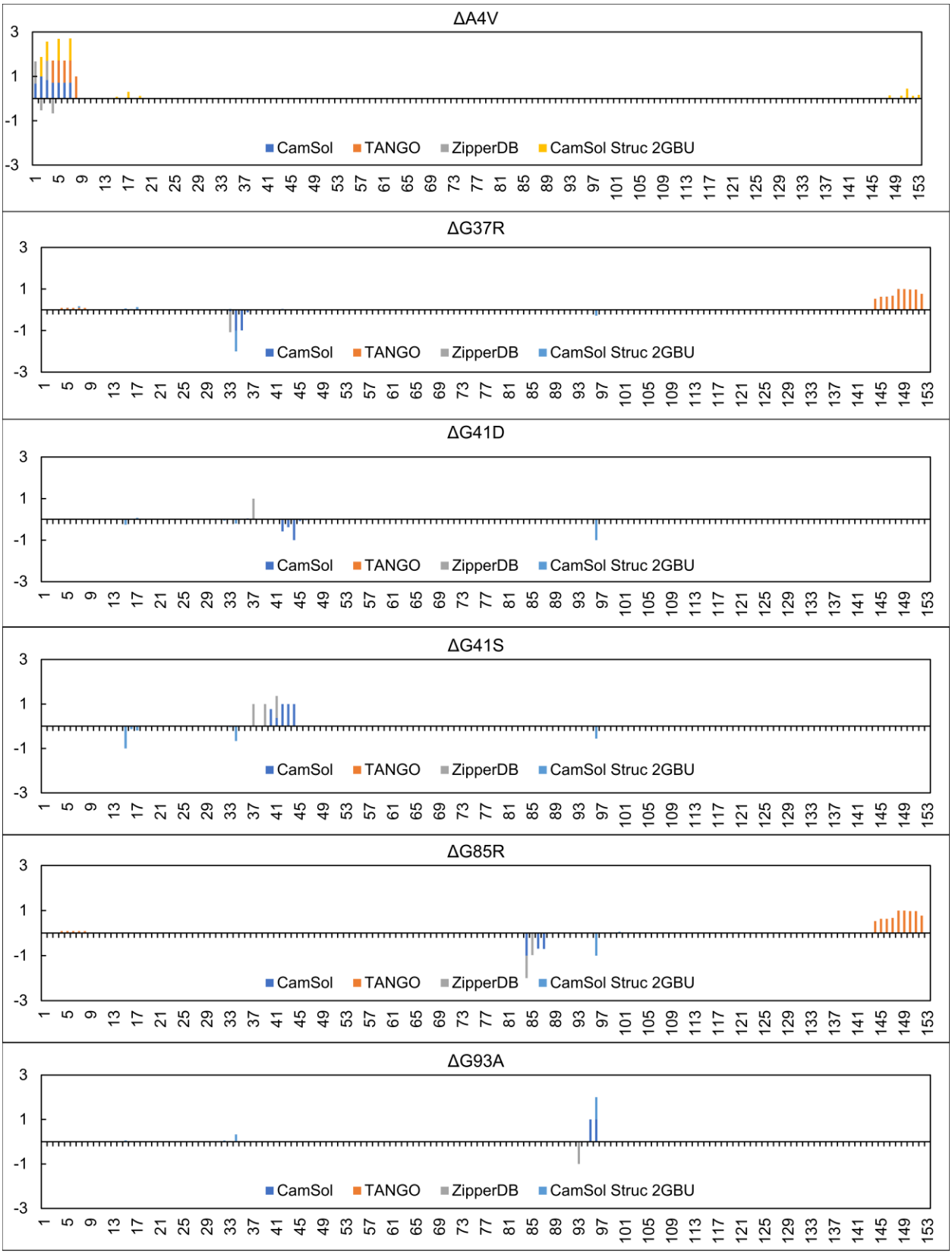


Figure A3.2: The % of SOD1 that forms insoluble IBs increases at higher growth temperature.

% SOD1 in IB is calculated as in **Figure A3.1** for samples taken at 4 and 6 hours post induction at 25 °C, 30 °C (both typically in triplicate) and 37 °C (single measurements). Data are shown as box-and-whisker plots, with horizontal lines at the median, box enclosing 50% of the data. Whiskers represent the minimum of the lower quartile and the maximum of the upper quartile, respectively. Outliers are greater than 1.5 times the interquartile range. Data obtained by Dalia Naser and Susan Kelso in the Meiering group.



[Continued on next page]

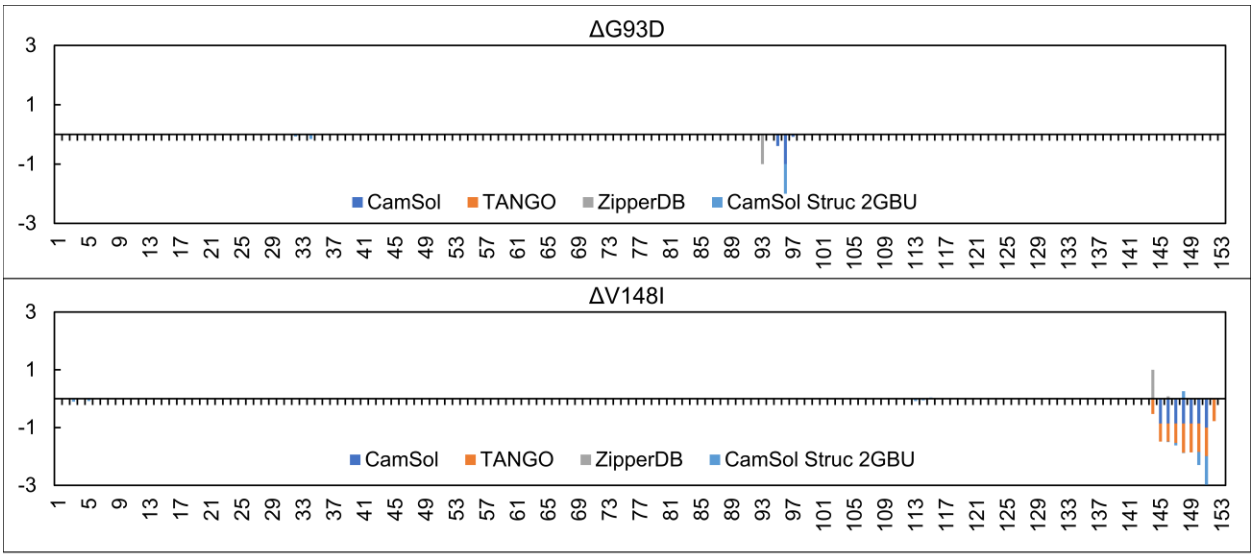
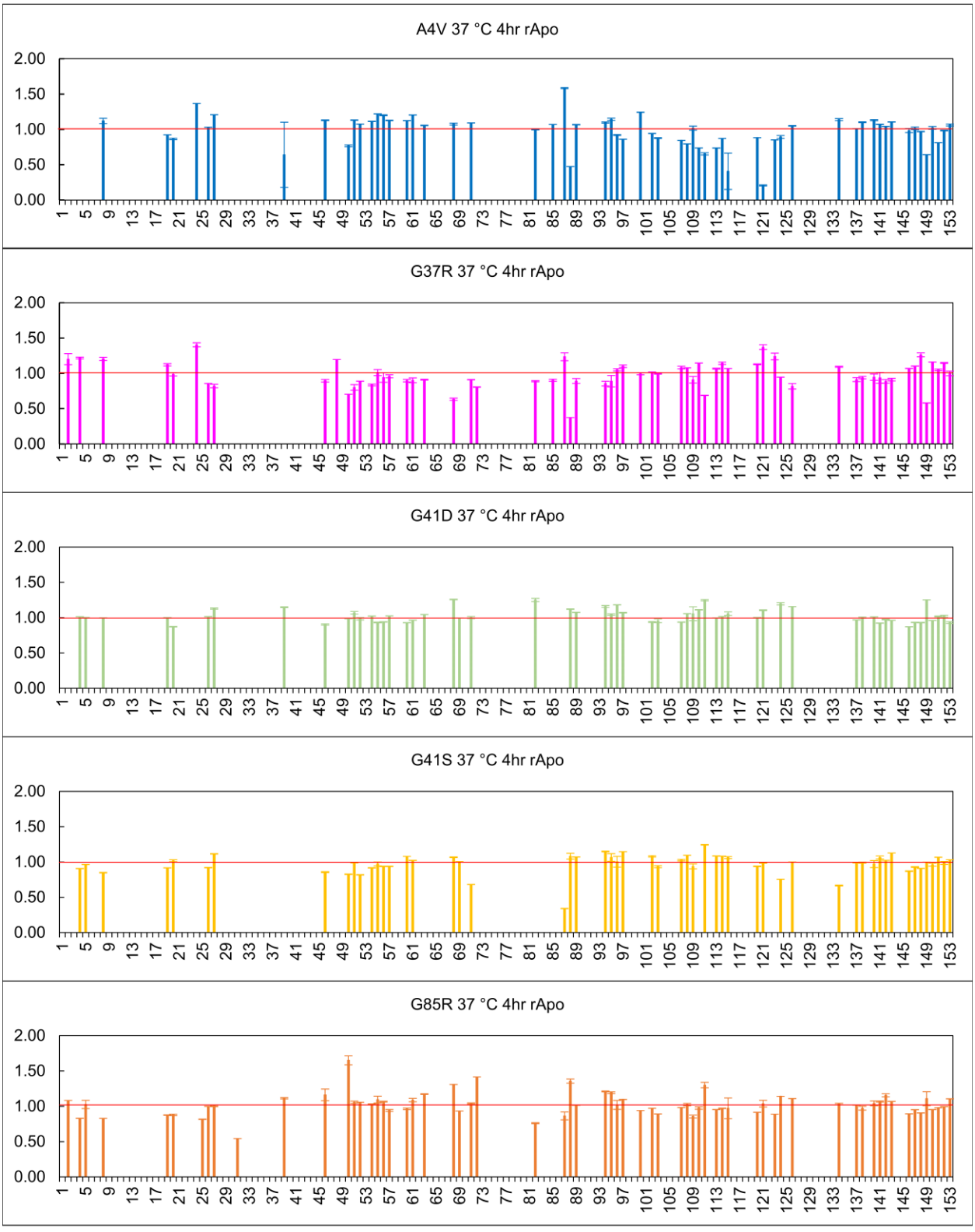


Figure A3.3: Normalized Delta Predictor Scores.
Delta prediction data are presented for all mutants, besides V148G which is included in the main text. See 3.3.2, for detailed explanation on how these values were obtained.



[Continued on next page]

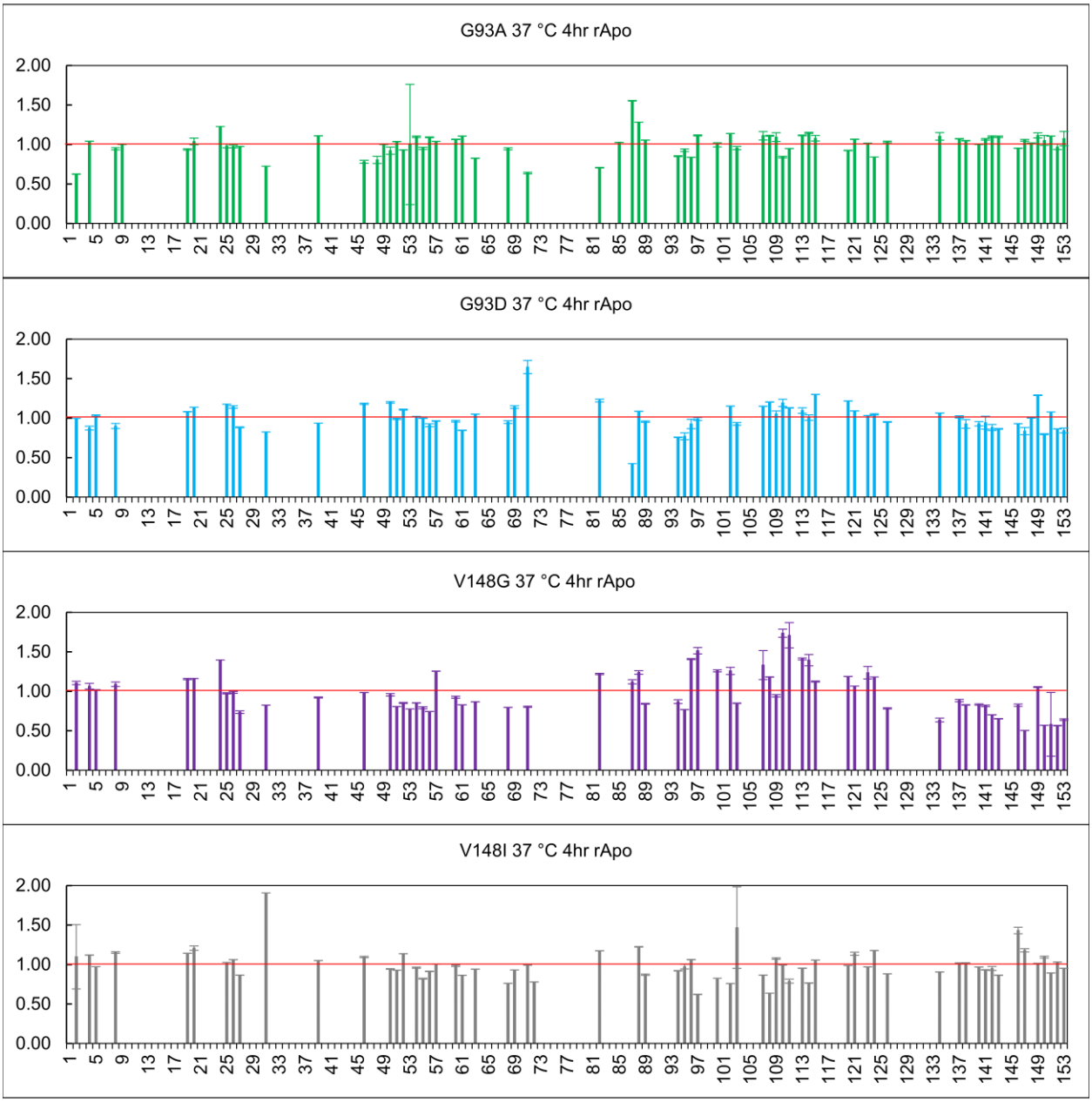


Figure A3.4: Protection ratios - rApo expressed at 37° C for 4 hours.
Ratio of the fraction amide protected for each mutant to the average fraction amide protected for all mutants expressed with these conditions (A4V, G37R, G41D, G41S, G85R, G93A, G93D, V148I). V148G is excluded from this average due to it having increased protection in loop L6, and decreased protection in loop L4 and β8. The red line represents a value of 1.0, any data points above this line have a higher observed protection relative to the average, and any points below this line have a lower observed protection.

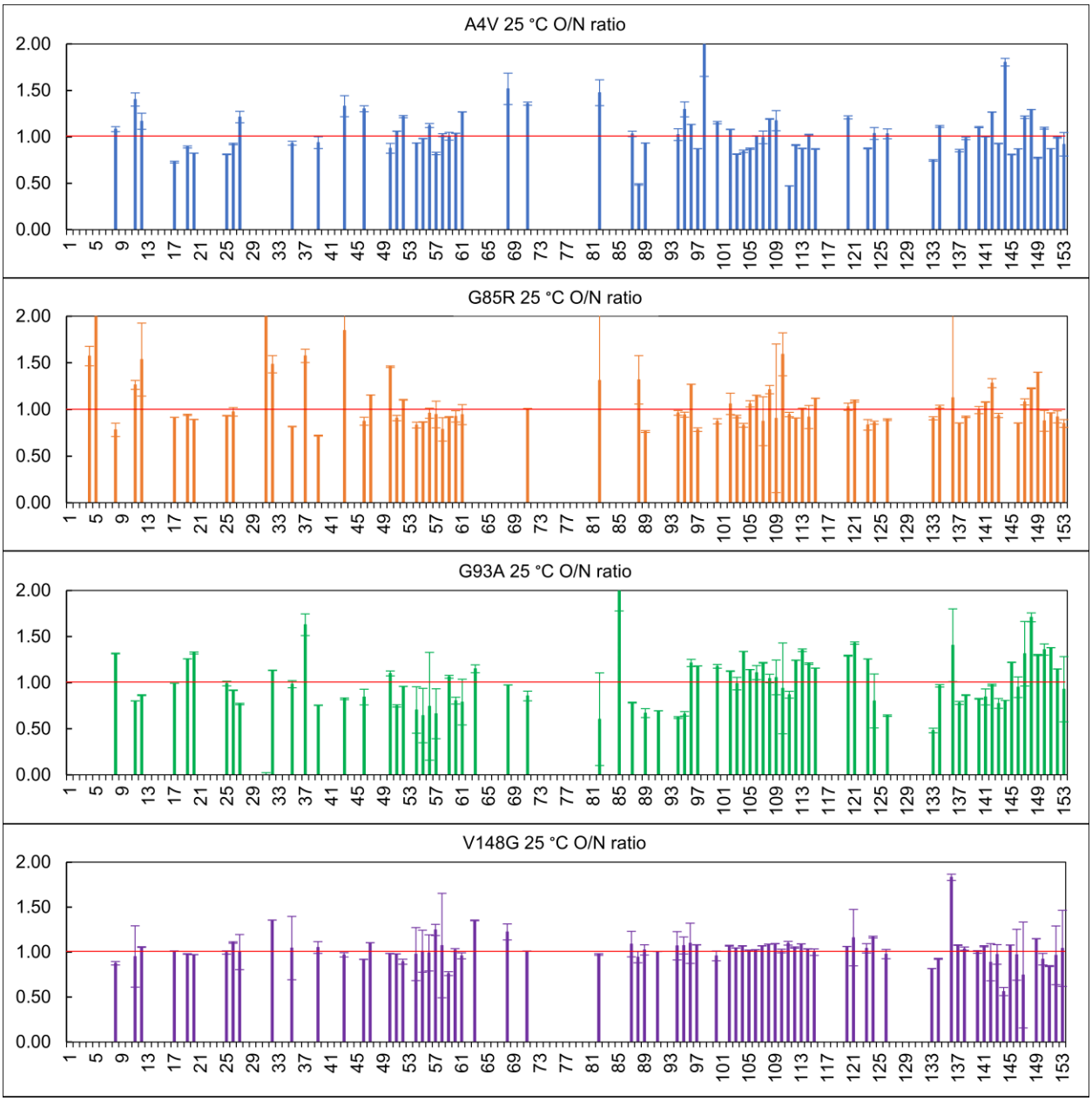


Figure A3.5: Protection ratios - rApo expressed at 25 °C overnight.
Ratio of the fraction amide protected for each mutant to the average fraction amide protected for all mutants expressed with these conditions (A4V, G85R, V148G). G93A is excluded from this average due to it having increased protection in loop L6 and β8, and decreased protection in loop L4. The red line represents a value of 1.0, any data points above this line have a higher observed protection relative to the average, and any points below this line have a lower observed protection.

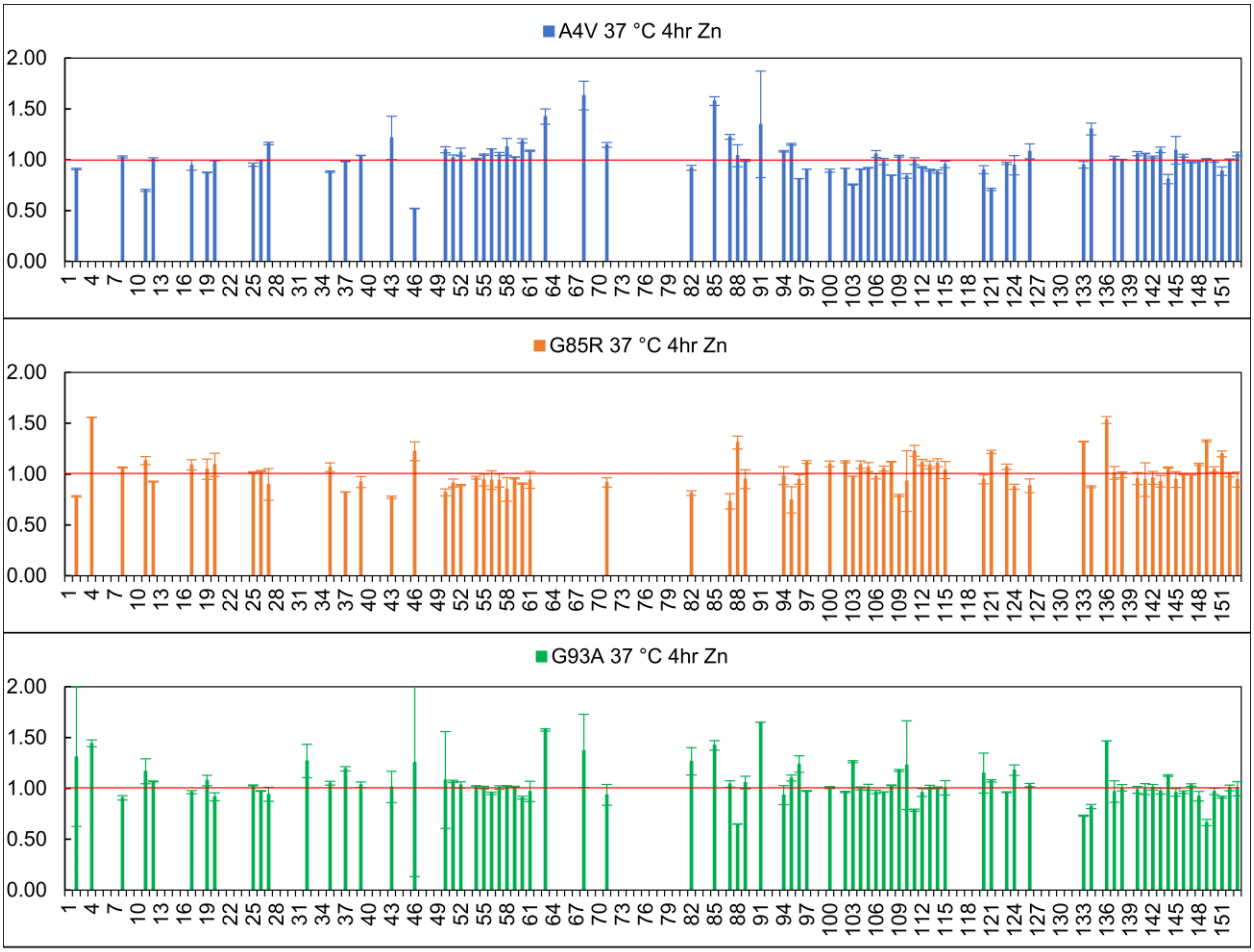


Figure A3.6: Protection ratios - Zn expressed at 37 °C for 4 hours.
Ratio of the fraction amide protected for each mutant to the average fraction amide protected for all mutants expressed with these conditions (A4V, G85R, G93A). The red line represents a value of 1.0, any data points above this line have a higher observed protection relative to the average, and any points below this line have a lower observed protection.

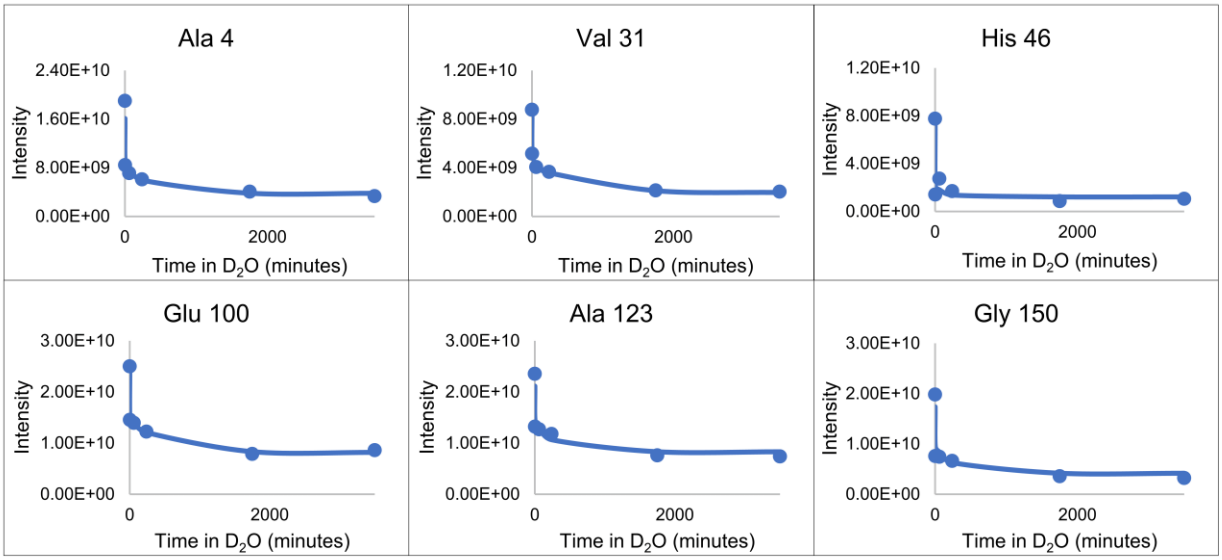


Figure A3.7: Biphasic exchange of IB amide protons in D_2O.
Aliquots of IBs from the same culture expressing G85R SOD1 grown at 37 °C and harvested at 4 hours post-induction were incubated in D_2O for 0 (protonated), 2, 60, and 240 at room temperature, and for 1 week and 2 weeks at 4 °C (converted to 1750 and 3480 minutes, respectively, at 20 °C using SPHERE[46,127]). To guide the eye, solid lines show fits of the biphasic time course of exchange in D_2O to a double exponential decay with an offset, Amide Peak Intensity = $A_1e^{(-x/t1)} + A_2e^{(-x/t2)}$ + offset, using OriginPro 7.

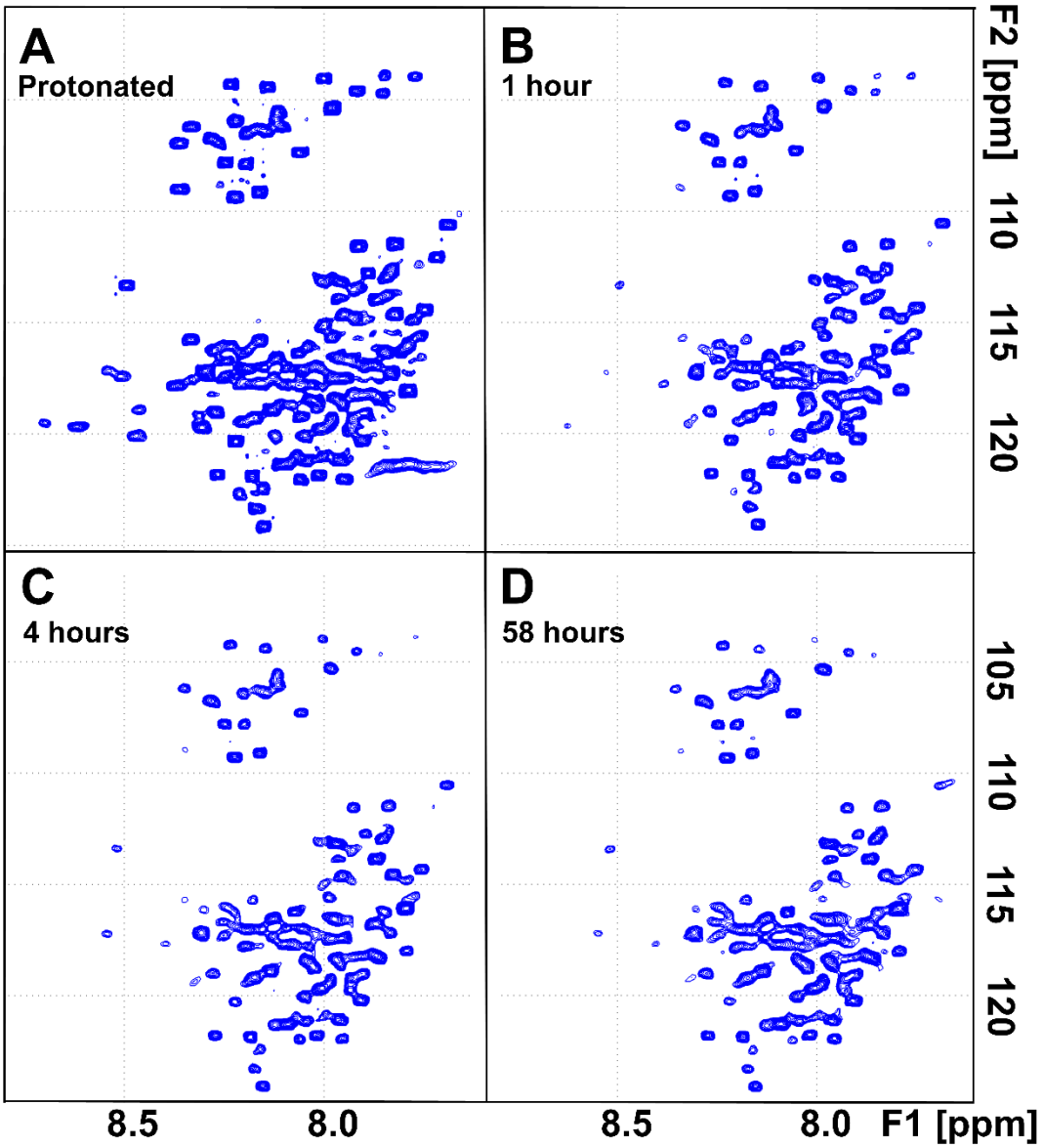


Figure A3.8: ^{1}H-^{15}N HSQC spectra for aliquots of G85R SOD1 IBs at different times of exchange in D_2O.

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 SPHERE[46,127])

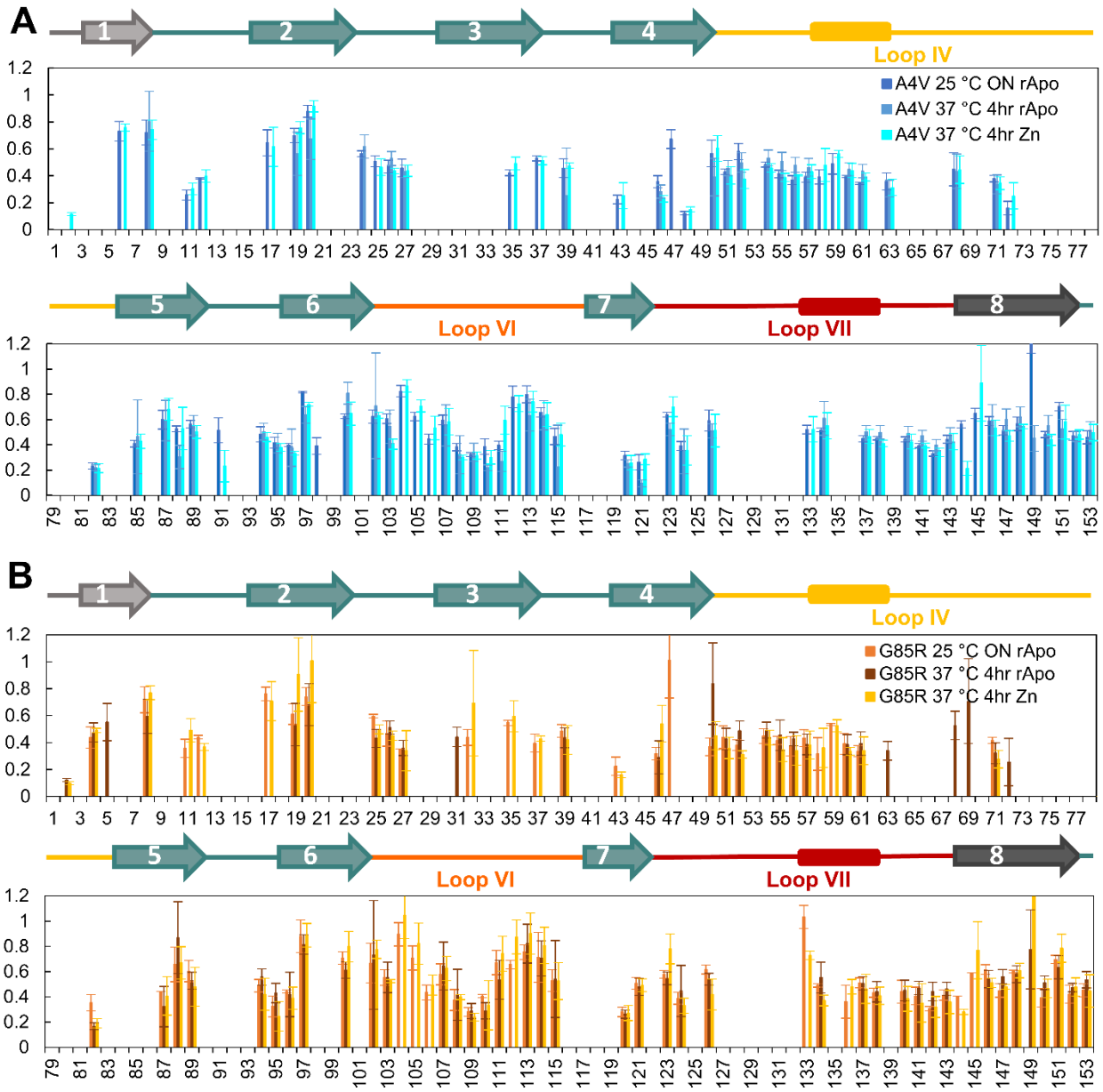


Figure A3.9: Comparison of amide protection between different expression conditions. (A) Fraction amide protection of A4V SOD1 at all three expression conditions. Each bar is the average of two biological replicates. No significant differences in protection were detected throughout each condition. Error bars represent propagated error and were calculated using Eq. 6. (B) Fraction amide protection of G85R SOD1 at all three expression conditions. Each bar is the average of two biological replicates. No significant differences in protection were detected throughout each condition. Error bars represent propagated error and were calculated using Eq. 6.

Residue		Nucleus	Chemical shift (ppm)
2	THR	H	8.456
2	THR	N	113.582
3	LYS	H	8.019
3	LYS	N	120.049
4	VAL	H	7.969
4	VAL	N	117.258
5	VAL	H	7.969
5	VAL	N	118.85
6	CYS	H	8.124
6	CYS	N	119.741
7	VAL	H	7.951
7	VAL	N	117.813
8	LEU	H	8.016
8	LEU	N	122.063
9	LYS	H	7.974
9	LYS	N	118.641
10	GLY	H	8.162
10	GLY	N	106.526
11	ASP	H	8.214
11	ASP	N	116.688
12	GLY	H	7.768
12	GLY	H	7.965
12	GLY	N	104.14
12	GLY	N	105.526
14	VAL	H	7.991
14	VAL	N	115.677
15	GLN	H	8.053
15	GLN	N	120.037
16	GLY	H	8.14
16	GLY	N	106.426
17	ILE	H	7.838
17	ILE	N	116.662
18	ILE	H	7.921
18	ILE	N	119.977
19	ASN	H	8.046
19	ASN	N	119.924
20	PHE	H	8.099
20	PHE	N	117.846
24	GLU	H	8.034
24	GLU	N	117.775
25	SER	H	7.899
25	SER	N	113.509
26	ASN	H	8.234
26	ASN	N	118.834
27	GLY	H	7.909
27	GLY	N	104.944
31	VAL	H	7.769
31	VAL	N	116.087
32	TRP	H	8.09
32	TRP	N	121.339
33	GLY	H	8.232
33	GLY	N	107.032
34	SER	H	7.942
34	SER	N	113.039
35	ILE	H	8.033
35	ILE	N	118.614

36	LYS	H	8.086
36	LYS	N	121.288
37	GLY	H	8.1
37	GLY	N	106.308
38	LEU	H	7.929
38	LEU	N	118.557
39	THR	H	7.868
39	THR	N	111.438
40	GLU	H	7.974
40	GLU	N	118.887
41	GLY	H	8.17
41	GLY	N	106.535
42	LEU	H	8
42	LEU	N	118.614
43	HIS	H	8.245
43	HIS	N	116.155
44	GLY	H	8.16
44	GLY	N	106.306
46	HIS	H	8.439
46	HIS	N	117.447
47	VAL	H	7.933
47	VAL	N	116.649
48	HIS	H	8.441
48	HIS	N	120.023
49	GLU	H	8.065
49	GLU	N	118.63
50	PHE	H	8.198
50	PHE	N	117.719
51	GLY	H	8.283
51	GLY	N	106.444
52	ASP	H	8.213
52	ASP	N	116.759
53	ASN	H	8.251
53	ASN	N	117.245
54	THR	H	7.687
54	THR	N	110.695
55	ALA	H	8.056
55	ALA	N	122.232
56	GLY	H	8.128
56	GLY	N	104.54
57	CYS	H	7.972
57	CYS	N	115.759
58	THR	H	7.968
58	THR	N	113.076
59	SER	H	7.93
59	SER	N	114.91
60	ALA	H	8.127
60	ALA	N	122.493
61	GLY	H	7.972
61	GLY	N	104.21
63	HIS	H	8.318
63	HIS	N	115.836
64	PHE	H	7.986
64	PHE	N	117.201
65	ASN	H	8.553
65	ASN	N	119.681
68	SER	H	7.575
68	SER	N	112.254

69	ARG	H	7.959
69	ARG	N	119.72
71	HIS	H	8.151
71	HIS	N	116.364
72	GLY	H	8.329
72	GLY	N	106.95
73	GLY	H	8.197
73	GLY	N	106.187
82	GLY	H	8.348
82	GLY	N	109.185
83	ASP	H	8.205
83	ASP	N	117.011
84	LEU	H	8.029
84	LEU	N	118.818
85	GLY	H	8.105
85	GLY	N	105.476
86	ASN	H	8.083
86	ASN	N	116.949
87	VAL	H	7.888
87	VAL	N	115.608
88	THR	H	7.825
88	THR	N	113.803
89	ALA	H	7.938
89	ALA	N	122.14
90	ASP	H	8.137
90	ASP	N	116.045
91	LYS	H	7.84
91	LYS	N	117.456
92	ASP	H	8.167
92	ASP	N	116.654
93	GLY	H	7.973
93	GLY	N	105.583
94	VAL	H	7.714
94	VAL	N	114.719
95	ALA	H	8.142
95	ALA	N	123.418
96	ASP	H	8.255
96	ASP	N	116.926
97	VAL	H	7.545
97	VAL	N	113.981
98	SER	H	8.08
98	SER	N	116.383
99	ILE	H	7.879
99	ILE	N	118.47
100	GLU	H	8.006
100	GLU	N	119.834
101	ASP	H	8.122
101	ASP	N	117.551
102	SER	H	7.802
102	SER	N	112.945
103	VAL	H	7.805
103	VAL	N	116.621
104	ILE	H	7.847
104	ILE	N	118.578
105	SER	H	7.932
105	SER	N	116.388
106	LEU	H	8.074
106	LEU	N	121.326

107	SER	H	7.924
107	SER	N	113.075
108	GLY	H	8.176
108	GLY	N	107.924
109	ASP	H	8.158
109	ASP	N	117.022
110	HIS	H	8.132
110	HIS	N	115.912
111	CYS	H	8.096
111	CYS	N	116.677
112	ILE	H	8.153
112	ILE	N	119.446
113	ILE	H	7.906
113	ILE	N	120.499
114	GLY	H	8.219
114	GLY	N	109.476
115	ARG	H	8.022
115	ARG	N	117.869
119	VAL	H	8.007
119	VAL	N	115.458
120	HIS	H	8.266
120	HIS	N	119.686
121	GLU	H	8.106
121	GLU	N	118.661
122	LYS	H	8.194
122	LYS	N	119.434
124	ASP	N	115.919
126	LEU	H	7.78
126	LEU	N	118.112
127	GLY	H	8.117
127	GLY	N	105.802
128	LYS	H	7.966
128	LYS	N	117.531
129	GLY	H	8.268
129	GLY	N	106.725
133	GLU	H	8.067
133	GLU	N	117.148
134	SER	H	7.949
134	SER	N	113.843
135	THR	H	7.807
135	THR	N	112.704
136	LYS	H	8.011
136	LYS	N	119.726
137	THR	H	7.771
137	THR	N	111.318
138	GLY	H	8.08
138	GLY	N	107.567
139	ASN	H	8.121
139	ASN	N	117.008
140	ALA	H	8.24
140	ALA	N	121.802
141	GLY	H	8.223
141	GLY	N	104.51
142	SER	H	7.868
142	SER	N	112.942
143	ARG	H	8.196
143	ARG	N	120.303
144	LEU	H	7.897

144	LEU	N	118.293	149	ILE	N	119.43
145	ALA	H	8.011	150	GLY	H	8.156
145	ALA	N	121.22	150	GLY	N	109.287
146	CYS	H	7.927	151	ILE	H	7.747
146	CYS	N	114.554	151	ILE	N	115.978
147	GLY	H	8.217	152	ALA	H	8.135
147	GLY	N	107.808	152	ALA	N	124.191
148	VAL	H	7.789	153	GLN	H	8.046
148	VAL	N	114.773	153	GLN	N	116.48
149	ILE	H	7.905				

Table A3.1: ^{1}H-^{15}N backbone assignments of A4V SOD1 in 95% DMSO / 5% H_2O at pH 5.5.
All A4V SOD1 resonance assignments were determined by Dalia Naser in the Meiering group.