2019-09-11

Exploring the Complex Folding Free Energy Landscapes of a Series of β-rich Proteins

Noah R. Cohen
University of Massachusetts Medical School

EXPLORING THE COMPLEX FOLDING FREE ENERGY LANDSCAPES OF A SERIES OF β-RICH PROTEINS

A Dissertation Presented by

NOAH ROBERT COHEN

Submitted to the Faculty of the
University of Massachusetts Graduate School of Biomedical Sciences, Worcester in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

September 11, 2019

EXPLORING THE COMPLEX FOLDING FREE ENERGY LANDSCAPES OF A SERIES OF β-RICH PROTEINS

A Dissertation Presented

By

NOAH ROBERT COHEN

This work was undertaken in the Graduate School of Biomedical Sciences Biochemistry and Molecular Pharmacology

Under the mentorship of

C. Robert Matthews, Ph.D., Thesis Advisor

Lawrence Hayward, MD, Ph.D., Member of Committee

David G. Lambright, Ph.D., Member of Committee

Celia Schiffer, Ph.D., Member of Committee

Susan Marqusee, MD, Ph.D., External Member of Committee

Francesca Massi, Ph.D., Chair of Committee

September 11, 2019

Dedication

To my parents for teaching me to always learn more.

To the rest of my family for reminding me how far I've come.

To my friends for keeping me laughing.

Thank you all, for the help along the way.

Acknowledgements

This thesis is the culmination of the efforts of myself and numerous others, whom I'd like to acknowledge here.

The first is my advisor and mentor, Bob Matthews. Bob, it has truly been an honor to work in your lab and to learn from and with you. Your passion for science is, and always has been, a great source of inspiration to me.

Osman Bilsel has been a constant source of knowledge and was always willing to have a long ~~debate~~ *discussion* about just about anything. He also taught me, through numerous lessons, that the best experiments don't start until after 4 PM.

Jill Zitzewitz was always ready to listen and help in any way she could. Jill advised me when I was a new research technician in lab and she was vital in my becoming a graduate student at UMassMed. I owe her a great deal as without that initial support and encouragement I would not have written this thesis. Beyond that support, Jill was one of the few people with whom I could discuss my health issues and feel that she truly understand and empathized.

I'd also like to thank all the Matthews lab members past and present, who helped me in this endeavor. Can, Paul, Kevin, Yvonne, Sagar, Brian, Dan, Meme, Sujit, Rohit, Divya, Ganga, and the many others (too many to list!): it was a privilege to know and work with all of you. My lab mom, Karen, was always ready to provide me with the candy necessary to keep my science moving forward.

My committee members, Celia Schiffer, Daryl Bosco, Francesca Massi and David Lambright, were always helpful and ready to provide guidance and support. Larry Hayward was very accommodating in serving as a stand-in for Daryl during my defense. Susan Marqusee graciously agreed to be my outside member The rest of the BMP department has provided a wonderful environment for the past 10 years. I owe the entire graduate school of UMassMed a good deal of gratitude for allowing me to represent both the Graduate School Committee and Three School Committee in my time as a student representative.

My family was always there to provide the moral and culinary support necessary as a starving graduate student.

And lastly, Janelle Hayes, I'm very lucky that I got to meet you during all of this. I hope I helped you as much as you did me.

Abstract

Protein aggregation is deleterious to human health and detrimental to therapeutic shelf-life. The physical processes that induce aggregation are the same processes that drive productive folding reactions. As such, protein aggregation is a non-productive form of protein folding. To gain insight into the steps that serve as a partition between the folding and aggregation reactions, the folding mechanisms of several β-rich proteins with links to human disease or medicine were examined.

In the ALS-linked protein, SOD1, a subpopulation of the unfolded ensemble is found to be a common source of both nonnative structure and frustrated folding. These behaviors are only observed upon the reduction of the intrinsic disulfide bond, indicating that this covalent interaction wards against aggregation. The nonnative structure presents an attractive target for the development of new therapeutic agents.

In V_H domains from therapeutic mAbs, the intramolecular disulfide bond protects against aggregation. However, it can also introduce complexity to the folding mechanism. This complexity is linked to the formation of a strained orientation of the disulfide bond. This strained orientation of the disulfide in certain V_H domains is energetically unfavorable enough to disrupt the formation of the disulfide in the full length mAbs. The novel relationship observed between disulfide orientation, folding complexity, and incomplete oxidation warrants further examination in other Ig domains.

Overall, these results demonstrate that mapping the folding free energy landscape for proteins with roles in human disease or therapeutics can provide valuable insights for developing and improving treatment options.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

SOD1 – Superoxide Dismutase 1
ALS – Amyotrophic Lateral Sclerosis
kDa – kilodaltons
PDB – Protein Data Bank
Apo – Metal free
2SH –Reduced Disulfide
SS –Oxidized Disulfide
LC – Light Chain
HC – Heavy Chain
V_H – Variable Heavy Domain
C_H – Constant Heavy Domain
VL – Variable Light Domain
CL – Constant Light Domain
CDR – Complementarity Determining Region
FR – Framework Region
FDA – Food and Drug Administration
mAb – Monoclonal Antibody
FRET – Forster
trFRET
DO – Donor Only
Trp – Tryptophan
EDANS - 5-((2-Aminoethyl)amino)naphthalene-1-sulfonic acid
DA – Donor Acceptor
MEM – Maximum Entropy Modeling
k_d – Donor decay rate constant
UV – Ultraviolet
CD – Circular Dichroism
2DMEM – Two Dimensional MEM
κ_{ET} – Energy transfer rate constant
REED – End-to-End Distance
NATA – N-acetyl-tryptophanamide
mSOD1 – Monomeric SOD1
Gdn-HCl - Guanidine Hydrochloride
C_m – Midpoint
ΔG – Free energy of folding
m_{eq} – Denaturant dependence of ΔG
η – solvent viscosity
T_β – Tanford β
TSE – Transition State Ensemble
σ – Internal Friction
MD – Molecular Dynamics

Preface

The work presented in Chapter II is in press as *Cohen NR, Zitzewitz JA, Bilsel O, and Matthews CR. Nonnative structure in a peptide model of the unfolded state of superoxide dismutase 1 (SOD1): Implications for ALS-linked aggregation Journal Biological Chemistry, 2019.* Dr. Matthews, Dr. Bilsel, Dr. Zitzewitz and I were responsible for the experimental design. I performed the experimental work. I performed the data analysis with input and guidance from Dr. Bilsel. All authors interpreted the results collaboratively. I wrote the manuscript with input from all parties.

The work presented in Chapter III is currently under review at Biophysical Journal as *Cohen NR, Kayatekin C, Zitzewitz JA, Bilsel O, and Matthews CR*. *Friction-limited folding of disulfide-reduced SOD1: Implications for ALS-linked aggregation*. Dr. Matthews, Dr. Kayatekin, Dr. Ztizewitz and I were responsible for the experimental design. Dr. Kayatekin and I performed the experimental work. Dr. Bilsel and I designed the novel analysis method. I performed the data analysis. All authors interpreted the results collaboratively. I wrote the manuscript with input from all parties.

The work presented in Chapter IV is a preliminary report of the results of a study that explores the folding mechanism of several immunoglobulin variable heavy domains from therapeutic antibodies. The chapter is adapted from a manuscript draft in prep targeted to Protein Science. The original project was started as a collaboration between our group and Dr. Yung-Hsiang Kao's group at Genentech. Dr Xiren Yu provided me with small amounts of the purified domains for initial experiments. After the initial period she provided cell pellets from which I could purify more as well as the methodology to do so, which I scaled up as necessary. Dr. Xiren Yu crystallized the

domains under the direction of Dr. Kao. The experimental design was performed by Dr. Matthews, Dr. Kao and myself. I performed all the experimental work and the data analysis. Dr. Matthews and I interpreted the results and formed conclusions with input from Dr. Kao. I wrote the manuscript draft with input from Dr. Matthews.

Chapter I – Introduction

BACKGROUND

Proteins are structurally and functionally diverse macromolecules that mediate many of the biological processes responsible for life. Ranging from enzymes that catalyze metabolic reactions to cytoskeletal components that give cells structure and support while also providing intracellular transport mechanisms to signal receptors that offer a means of intercellular communication and a system for adaptive response to external stimuli, proteins play crucial roles throughout biology . In order to specifically and efficiently perform such wide-ranging functions, proteins must be able to adopt unique structures, which they do through a process termed “folding”.

Despite the wide array of folded structures all proteins begin in a similar form, as unfolded and expanded polypeptide chains of varying lengths. From this initial state, each protein is able to fold into a unique, stable and functional native state. How can various sequences of the same 20 amino acids encode such a multitude of three-dimensional protein structures? How does the primary sequence influence the mechanism by which a protein reaches its native state? These are the central questions that the protein folding field seeks to answer.

The landscape view of the protein folding

In seminal work that was later awarded the Nobel prize, Christian Anfinsen discovered that Ribonuclease was able to spontaneously regain its enzymatically active form when removed from a denaturing environment[1]. This work demonstrated that a polypeptide chain contains the information required to reach its native form, suggesting

that a protein's native state is the conformation with the lowest free energy and thus the most thermodynamically favored state. Concurrently, Cyrus Levinthal noted that due to the length of polypeptide chains and the number of potential phi and psi bond angle arrangements, a protein would have an inordinately large number of possible conformations[2]. Consequently, sampling each one until finding the one native state, even if done rapidly, would take timescales far beyond the limits imposed by biology. This thought experiment implies that protein folding is a sequential process through a defined pathway, rather than a random search process.

By combining these two basic principles the energy landscape model of protein folding was developed[3–5]. In this model the energy landscape is represented with a funnel-shaped diagram in which the width corresponds to the conformational entropy and the height denotes the free energy of the molecule. At the top of the funnel is the ensemble of unfolded states of the polypeptide; these unfolded states lack order or structure and thus have large conformational entropies and high free energies. The bottom of the funnel denotes the folded, native state of the protein; this well-defined structure represents the free energy minima of the system and is conformationally constrained. The sides of the landscape represent the pathway a polypeptide chain may take on its route to the native state and they can be either smooth, i.e., two-state folding in which unfolded chains fold directly to the native state, or rugged, when unfolded chains form partially-folded intermediates prior to the native state[5]. Mapping the energy landscape of proteins can provide insights into how different regions of sequence drive the distinct steps of folding. An example of a folding landscape is shown in Figure 1.1.

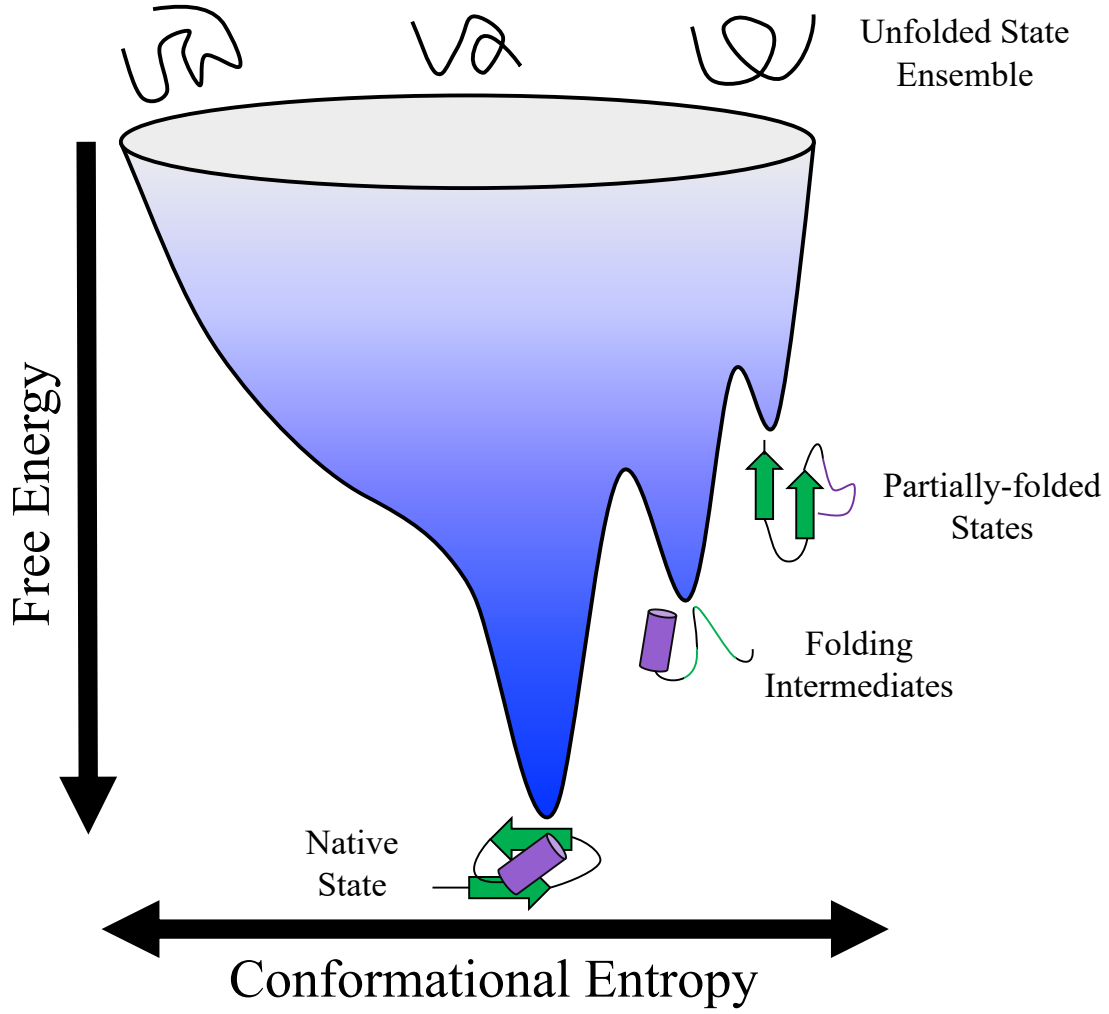

Figure 1.1: Folding free energy landscape
A representative folding free energy landscape is shown. During folding polypeptide chains move from the highest energy expanded unfolded state (top) towards the lowest energy native state (bottom). During this process partially-folded states (upper right) or stable intermediates (lower right) can form prior to reaching the native state.

PROTEIN AGGREGATION

While the properly folded form of a protein is the biologically active and relevant structure, it is not the only potential outcome of folding reactions. Proteins can also aggregate. Aggregation is the term used to describe the mechanism through which multiple polypeptide chains form intermolecular associations resulting in higher order nonnative structures. There is an array of nonnative structures that may be formed during the process of aggregation that range from small soluble oligomers[6] to macroscope aggregates that can form either highly ordered structures, such as amyloid fibrils[7], or remain amorphous and lack well-defined structural components[8].

Aggregation is detrimental to human health and to therapeutics

Aggregation has significant impacts on human disease and therapy. Large insoluble protein aggregates are a hallmark of multiple forms of devastating human diseases such as neurodegeneration and amyloidosis[9]. The neurodegenerative diseases Alzheimer's, amyotrophic lateral sclerosis, Parkinson's, as well as prion diseases have all been linked to misfolding and subsequent aggregation of proteins, and recent evidence suggests similar processes may play a role in other maladies. For example, amyloidosis is a group of diseases that affect a range of tissues in the body all of which have the hallmark of forming large amyloid fibril aggregates that interfere with normal biological functions[10]. Protein aggregation disorders can even be instigated by other ailments, such as the light chain amyloidosis that often accompanies multiple myeloma[11]. Table 1.1 lists some aggregation-linked diseases and the related proteins.

Table 1.1: Diseases linked to protein aggregation and the relevant proteins

Disease	Protein
Alzheimer's Disease	Amyloid β[12]
Parkinson's Disease	α-synuclein[13]
Tauopathies	τ-proteins[14]
Frontotemporal Lobar Degeneration	TDP-43 and FUS-TLS[15]
Amyotrophic Lateral Sclerosis	SOD1, C9ORF72[16]
Huntington's	Huntingtin[17]
Prion Diseases	Prion Protein[18]
Type II Diabetes	Islet Amyloid Polypeptide[19]
Transthyretin amyloidosis	Transthyretin[20]
Light Chain amyloidosis	Immunoglobulin Light Chains[21]
Heavy Chain Disease	Immunoglobulin Heavy Chains[22]
AA amyloidosis	Amyloid A protein[23]
Dialysis-related amyloidosis	β_2 microglobulin[24]
Cataracts	Crystallin[25]
Retinitis Pigmentosa	Rhodopsin[26]
Medullary thyroid carcinoma	Calcitonin[27]
Multiple Myeloma	Immunoglobulin Light Chains[11]

Aggregation also interferes with human health by negatively impacting therapeutic molecules used in treating human disease. Aggregated therapeutics can lead to an adaptive immune response in patients resulting in the production of anti-drug antibodies that lead to therapeutic immunity or, rarely, development of an auto-immune disease[28–30]. Beyond that, aggregation can occur prior to delivery of the therapeutic, such as during the production process or storage, leading to a loss of activity over time indicating that the therapeutic has a poor shelf life[31–33]. Short shelf lives result in increased production costs and these, in turn, increase medical costs for patients.

Why do proteins aggregate?

The formation of protein aggregates seems to contradict the landscape model, as two inherent assumptions of the model are that folding to native state is the most kinetically accessible and most thermodynamically favorable state available to a polypeptide chain. If these assumptions were true, then how could a protein aggregate? Examination of the steps in aggregation, the structures formed, and how these are related to productive folding, begins to reveal an answer.

Some aggregates, especially those that form fibrillar structures, have been shown to be remarkably stable and difficult to resolubilize[34,35]. High resolution structural data has demonstrated that amyloid fibrils are composed of cross-β sheets stabilized by strong hydrogen bonding networks inside of hydrophobic cores[36,37] and that introduction of electrostatic residues[38] or proline residues[39] disrupt these cores and destabilizes the fibrils. These behaviors are remarkably similar to that of native states, which are themselves stabilized by dense packing of hydrophobic cores[40] and favorable surface

electrostatic interactions[41]. The shared behavior of native states and end stage aggregates indicates that the same physiochemical forces that drive productive folding also drive aggregation. This observation along with the unusual stability of aggregates suggests that not only are they thermodynamically stable, but that they are even more thermodynamically favorable than the native state.

The earliest step in some aggregation reactions has been shown to be the formation of partially-folded or misfolded states[42–47]. These aggregation-inducing states are formed during search for the native conformation and may even contain residual aspects of the native structure, suggesting that these steps are also a part of the productive folding mechanism. After these early steps, aggregation is generally irreversible, meaning end stage aggregates cannot be returned to the native conformation[48].

The observations that aggregation can (1) be thermodynamically favorable, (2) is driven by the same forces as productive folding and (3) may involve some of the steps early in the productive folding reaction leads to the concept that the two processes are inexorably linked. What this means is that protein aggregation is another portion of the free energy landscape a diagram of which is shown in Figure 1.2. Consequently, mapping the nexus between the productive folding landscape and the aggregation landscape can provide information on the structures or species that partition the two competing reactions.

Understanding how the primary sequence guides a protein molecule towards aggregation instead of productive folding and how to inhibit the former and promote the latter can be thought of as another goal of the protein folding field.

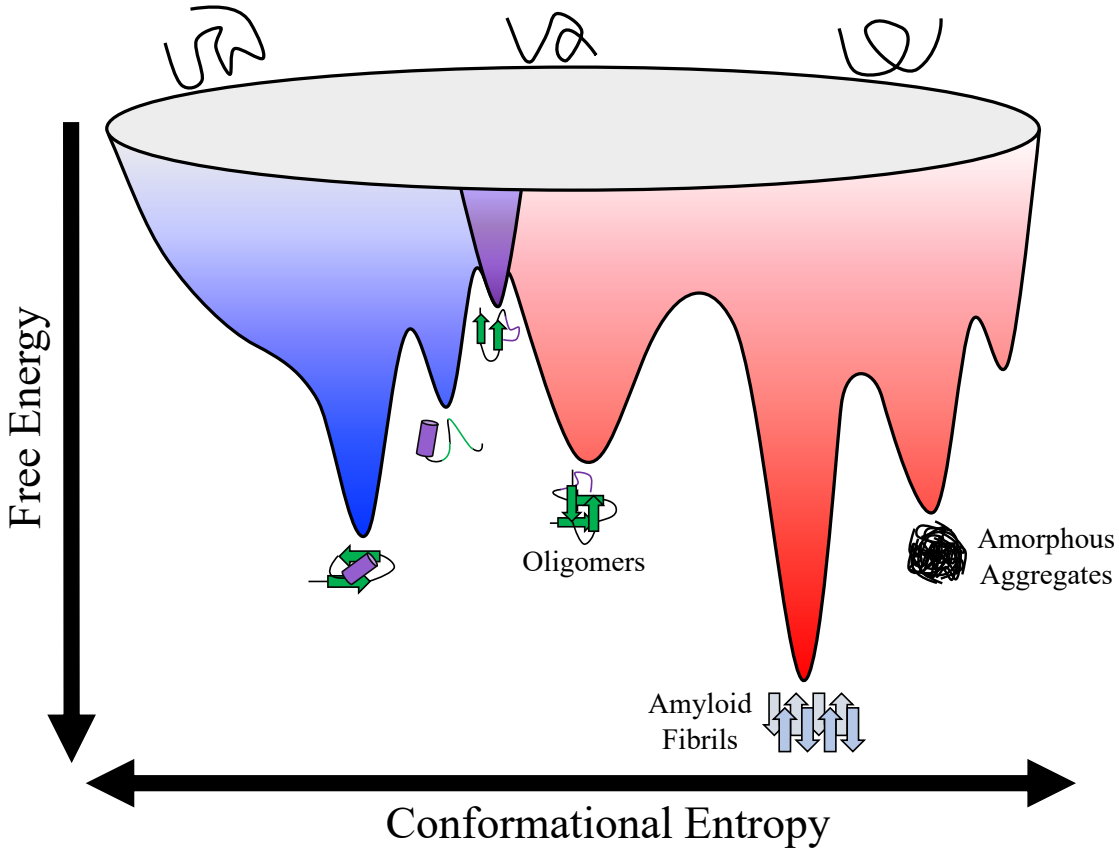

Figure 1.2: Folding and aggregation free energy landscape
A representative combined folding (blue) and aggregation (red) free energy landscape is shown. Not only can polypeptide chain fold productively (blue landscape, as in Fig. 1.1) but it may form intermolecular interactions resulting in the formation of oligomers (left), amyloid fibrils (middle) or amorphous aggregates (right). This view also highlights how partially-folded states may serve as a link between the two landscapes (purple section).

Can aggregation be prevented?

Ideally, to disrupt an aggregation reaction, we would first have a complete understanding of the steps in the process as well as high resolution structural information of the species formed at each step. Unfortunately, this information has proven to be challenging to obtain and such a detailed understanding of the aggregation process remains elusive.

A commonly accepted, while fairly nonspecific, mechanism describing the process of aggregation is shown in Figure 1.3[49]. This mechanism contains three major regions: a lag phase, a growth phase, and a plateau phase. The lag phase, which is predominantly defined by a search for a nucleating conformation, occurs after initiation of the reaction but prior to the formation of macromolecular aggregates[50]. The growth phase occurs when a critical concentration of the nucleating species is formed and large aggregates begin to rapidly form. The plateau phase is the period in which all of the polypeptide chains have been incorporated into aggregates and they can no longer grow. To further the goal of inhibiting aggregation, the lag phase is the period of most interest; the growth phase is past the critical point in which the reaction becomes irreversible and the plateau phase is defined by the behavior of the end stage aggregates. The lag phase, however, contains the earliest steps of the aggregation process such as misfolding and/or oligomerization that could be targeted to prevent the aggregation from proceeding further. While such a general mechanism of aggregation may be broadly applicable, the structural conformations formed during the process may be unique to individual proteins and are difficult to characterize.

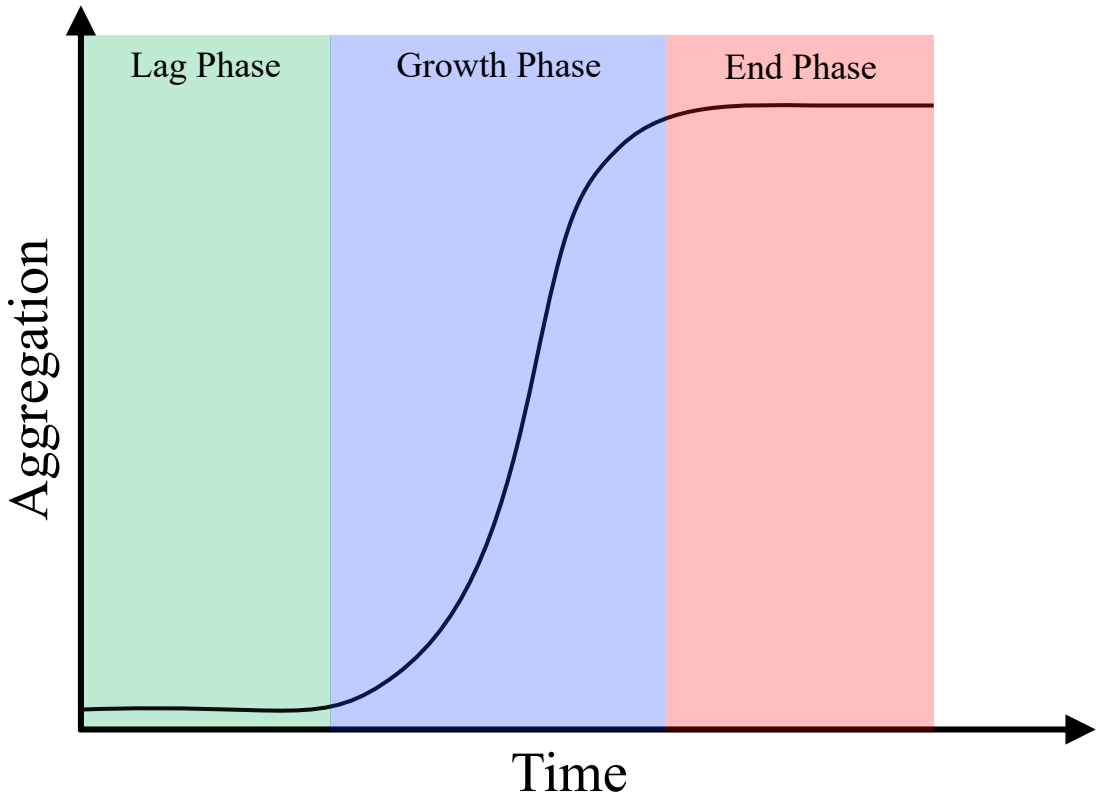

Figure 1.3: Simple aggregation mechanism
A simple mechanism showing aggregation as a function of time is shown. This mechanism is separated into three major regions: a Lag Phase (green), a Growth Phase (blue) and an End Phase (red).

A primary cause of this difficulty is the complexity involved in monitoring aggregation experimentally. The species populated during aggregation reactions are difficult to track as it involves monitoring the behavior of multiple species that make up a minute fraction of a highly heterogeneous sample. Furthermore, rapidly initiating aggregation reactions generally requires the use of powerful external stimuli such as the addition of large amounts of organic solvents[51] or the introduction of mechanical shearing forces by agitating the sample[52]. Such harsh *in vitro* treatments may preferentially induce aggregation pathways unrelated to the aggregation pathways *in vivo*, making direct comparisons between the two processes difficult. Once aggregation has been initiated, the state of the reaction must be probed without perturbing the process, which is also problematic given the transient nature and marginal stability of many of the species formed during the process. This problem is exacerbated by common probes of aggregation, such as thioflavin T[53], that are more sensitive to end stage aggregates than the early formed structures . The multitude of experimental challenges continues to make characterization of aggregation landscapes problematic, making the development of direct disruptors of aggregation difficult. Fortunately, there is another method by which we can inhibit aggregation.

Preventing aggregation by enabling productive folding

Because aggregation is intimately linked to folding, altering one process will inevitably affect the other. Therefore, rather than directly impede aggregation, we can instead facilitate productive folding. The caveat being that doing so requires a detailed understanding of the productive folding pathway. Fortunately, obtaining detailed folding

mechanisms is more thoroughly understood than doing the same for aggregation. Standard equilibrium and kinetic analyses can identify the major species populated during folding and map the pathways between these species and the native state as well provide information regarding the stability of each state[54]. The structural components of these species can then be determined using a wide-array of higher resolution techniques[55–59].

Identification of a marginally stable native state, persistent intermediate or unfolded states, or severely frustrated folding events all present potential sources of aggregation. Each potential source of aggregation also suggests possible methods for disrupting such aggregation. Fragile native states could be stabilized, thereby preventing unfolding events and reducing the opportunities for misfolding. Enhancing the rate of folding would decrease the time a polypeptide chain spends in the aggregation-prone portion of the free energy landscape. And disruption of the structures that frustrate folding would result in a smoother path to the native conformation.

The work described below will map the folding free energy landscapes of several proteins with links to aggregation in order to determine methods to promote productive folding or disrupt early steps in aggregation.

MODEL SYSTEMS

Superoxide Dismutase 1 as a model system

The first protein examined in the work presented below is the human protein copper, zinc Superoxide Dismutase 1 (SOD1). SOD1 is a ubiquitously expressed enzyme that catalyzes the dismutase reaction of superoxide radicals into hydrogen peroxide[60].

SOD1 is an attractive candidate for studies mapping the bridge between the folding and misfolding landscapes as mutant variants of SOD1 are known to aggregate and this aggregation is linked to the neurodegenerative disease amyotrophic lateral sclerosis (ALS)[61].

ALS is an invariably fatal neurodegenerative disease for which there is no cure. Worldwide, ALS affects ~5 in 100,000 people and the average time between initial onset of symptoms and death is only 2-3 years[62]. The symptoms of ALS are a progressive loss of muscle control that is caused by the death of motor neuron cells with eventual death caused by respiratory failure. While the exact mechanism of toxicity to motor neuron cells in ALS remains unclear, a hallmark of ALS and many other neurodegenerative diseases, is the formation of large protein-rich aggregates in affected tissues isolated from patients[63]. The observation that multiple neurodegenerative diseases share a congruent endpoint in which large insoluble inclusions are formed suggests protein misfolding and the subsequent aggregation may play a role in neurodegenerative disease pathology.

Mutant forms of SOD1 were first linked to ALS by the Brown lab in 1993[61] and since then >180 mutations in SOD1 have been linked to ALS[64]. Additionally, mutations to other proteins such as; FUS-TLS, TDP-43, c9ORF72, PFN1, VAPB, and numerous others that cause ALS have been identified[65]. A map of the known missense mutations to SOD1 linked to ALS is shown along the sequence of the protein in Figure 1.4.

SOD1 is a 32 kDa homodimeric protein in which each monomer is comprised of a 153-residue chain that forms eight anti-parallel β-strands and two long loops. Each of the

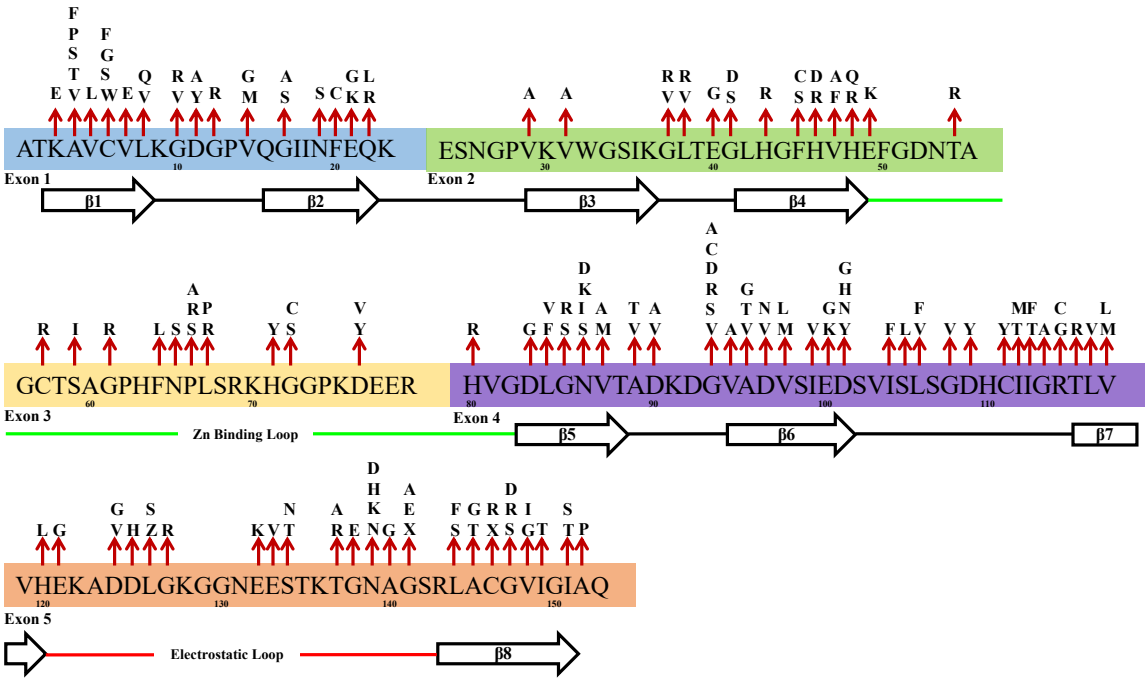

Figure 1.4: Map of SOD1 mutations linked to hereditary ALS

A map of the ALS-linked mutations is shown along the sequence of SOD1. ALS-linked mutations are indicated by red arrows pointing from the sequence to the known mutant residues. The sequence is separated into the five exons and the secondary structure elements are shown along the bottom of the sequence. This figure is adapted from Figure 1.1 from the thesis, " The coupling between folding, zinc binding, and disulfide bond status of human Cu, Zn superoxide dismutase" by Can Kayatekin.

loops in a monomer subunit binds one of two metal ion cofactors, zinc or copper. Binding the zinc ion stabilizes the protein[66], while the copper ion is used in the catalysis reaction[67]. Each monomer subunit is further stabilized by an intramolecular disulfide bond. This disulfide bond is a critical component used in copper ion loading[68] and is exceptional as disulfide bonds rarely exist in the reducing environment of the cytoplasm[69].

SOD1 adopts a fold similar to those seen in immunoglobulin domains that consists of a β-sandwich formed by two large antiparallel β-sheets, each themselves formed by 4 antiparallel β-strands. One sheet contains a sequence local hairpin-hairpin between β1-β2-β3 and a connection to the sequence distal β6, while the other sheet contains β4-β5 and β7-β8. β4-β5 are linked by the first of the long loops, which is responsible for coordinating the zinc cation and is accordingly termed the zinc-binding loop. β7-β8 are linked by the second long loop that is responsible for targeting the anionic superoxide substrate to the active site designated the electrostatic loop. β8 also links to Loop IV through the disulfide bridge between C146 and C57. A ribbon diagram and a topology map of SOD1 are shown in Figure 1.5.

The source of mutants SOD1-linked ALS toxicity remains unclear. While a loss of the enzymatic function might be assumed to play a role in ALS toxicity, studies involving transgenic mouse models instead revealed that the toxicity arises from in a gain of function manner[70]. This result is consistent with the ALS-linked mutations being distributed throughout the entire sequence as shown in Figure 1.4, rather than clustered around the enzymatic residues. Furthermore, a loss of SOD1 enzymatic activity would not explain the hallmark of the disease, the formation of large insoluble aggregates.

Figure 1.5: Ribbon diagram and topology map of SOD1
A ribbon diagram (top) and topology map (bottom) of SOD1 are shown. The ribbon diagram is adapted from PDB: 2C9V[71] and shows the zinc-binding loop (green), electrostatic loop (purple), zinc ion (blue), copper ion (orange) and the disulfide bond (yellow). The same colors are used in the topology map and the eight β-strands are numbered and their beginning and ending residues are shown.

The most apparent function gained by SOD1 upon mutation is aggregation, but are the aggregates themselves toxic? End stage SOD1 aggregates are often described as amorphous, as they lack the well-defined order of fibrillar aggregates and the morphology of the aggregates is variable[72]. While the toxicity of aggregates is not fully understood, the amorphous aggregates are not thought to drive ALS pathology as they appear too late relative to the onset of symptoms to be completely responsible[73,74]. The toxicity is instead attributed to the smaller, soluble oligomers that precede the formation of the larger insoluble inclusions[74–78]. Similar mechanisms of toxicity have been proposed for other protein misfolding and aggregation disorders, such as Alzheimer's[79,80] and Parkinson's[81,82].

The large number of mutations and their distribution throughout the sequence of SOD1 (Fig. 1.4) suggests that they act in a common manner to give rise to ALS. If so, then they likely share a common precursor that gives rise to the aggregation and toxicity. While some studies have identified aberrant nonnative and/or misfolded states that exist as branches off of the native folding landscape[44,83–86], others have proposed that the unfolded ensemble of SOD1, a feature common to all variants, is the source of ALS-linked aggregation[47,87,88].

While fully metalated and disulfide oxidized (SS) SOD1 dimers are extraordinarily stable, with a melting temp ~90° C[89] and enzymatic activity maintained in 8 M urea or 4 M guanidine hydrochloride[90], this remarkable stability is only gained during the maturation process as metal-free (apo) disulfide reduced (2SH) SOD1 monomers are much less stable[91]. The presence of ALS-linked mutations further

destabilizes the native state leading to significant increases in the population of the unfolded ensemble[88,92].

While a nascent SOD1 chain is able to fold in the absence of post-translational disulfide oxidation or metal ion binding, it does so exceptionally slowly. Apo-2SH-SOD1 takes >10 seconds to fold in native conditions, making it one of the slowest folding two-state proteins known[93]. The persistent lifetime of the SOD1 unfolded state coupled with the increase in the unfolded population of ALS-linked variants suggests that the unfolded state of SOD1 is the common source of aggregation in the >180 ALS-linked SOD1 mutations.

The work detailed below supports this suggestion by characterizing the behavior of the unfolded state of SOD1 and further proposes mechanisms through which these behaviors play a role in ALS-linked aggregation. The studies below demonstrate a unique role for the C-terminal region of the protein, especially the residues in β8 surrounding C146 from the disulfide bond, in bridging the folding and aggregation landscapes (Fig. 1.2). Mapping the connection between the folding and aggregation of SOD1 is done with the hope to gain insight into the mechanism of ALS-linked aggregation and identify potential novel therapeutic targets.

Variable Heavy Domains as a model system

The human adaptive immune response is a complex biological system through which organisms challenged with foreign pathogens use those same pathogens to rapidly develop unique, novel protein molecules that are astonishingly specific. These protein molecules are antibodies. Not only do they have specificity for distinct targets, but

antibodies also have distinct downstream immunologic effects based on their subclass. The substantial malleability in both target recognition and post-binding effects is how human immune systems are able to correctly recognize and counter the many pathogens that give rise to human diseases. The assembly and secretion of mature antibody molecules is a tightly regulated quality control process used to ensure antibodies are not only selective, but also non-harmful[94,95].

Human antibodies consist of four polypeptide chains; two copies of a Heavy Chain (HC) and two copies of a Light Chain (LC)[96]. These chains connect by forming intermolecular disulfide bond crosslinks resulting in a macromolecular complex roughly shaped like the letter "Y" as shown in Figure 1.6. Each of these chains is comprised of several small tandem immunoglobulin domains. The nomenclature of the HC and LC are a result of these domains as the HC contains either 4 or 5 domains giving it a higher molecular weight than the LC, which contains only two domains. The HC includes a variable heavy domain (V_H) linked to several constant heavy domains ($C_{H1-3/4}$). The LC contains a variable light domain (V_L) linked to a single constant light domain (C_L).

Each of the individual immunoglobulin domains adopts a structure termed the immunoglobulin-fold that is comprised of a β-sandwich formed by 2 β-sheets formed by 7-9 β-strands with a highly conserved intramolecular disulfide bond bridging the two β-sheets[97,98]. A ribbon diagram of such a domain is shown in the inset of Figure 1.6. The constant domains, as the name suggests, are identical across antibodies of the same subclass and these domains are responsible for the immune responses that occur downstream of antigen binding [96]. The variable domains, again, as their name suggests, are unique to each antibody as they are the domains responsible for antigen

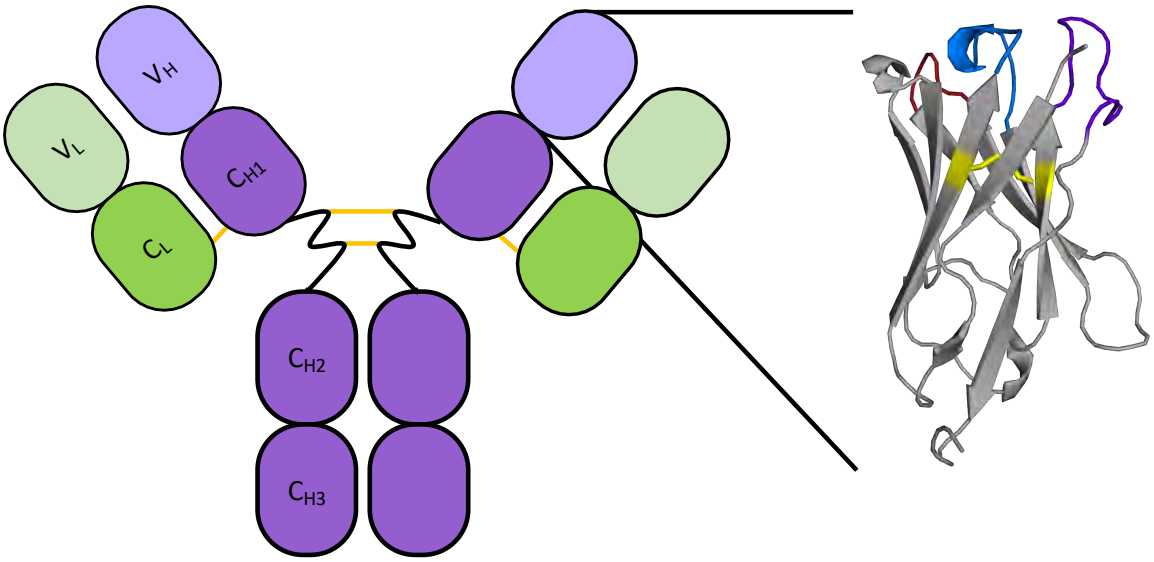

Figure 1.6: Antibody molecule illustration and ribbon diagram of a variable domain
A cartoon diagram of an antibody is shown (left). Each chain is shown as a series of domains (ovals). The domains of the HC (purple) and LC (green) are labelled on the left side. The variable domains are shown in light colors and the constant domains are shown in dark colors. The intermolecular disulfide bonds are shown as orange lines. A ribbon diagram of a variable domain, adapted from PDB: 1N8Z[99], is shown to the right. This diagram highlights the intramolecular disulfide bond (yellow) and CDR1 (blue), CDR2 (red), and CDR3 (purple).

recognition and binding[96]. The specificity of the variable domains is determined by the sequence of three complementarity determining regions (CDR1-3). The remainder of the variable domain sequence is referred to as framework regions (FR). The CDRs are altered through the processes of V(D)J recombination[100] and somatic hypermutation[101].

Recognizing the amazing therapeutic potential of antibodies, scientists have been developing them as treatments for human diseases for over 30 years[102]. Currently, therapeutic antibodies are the dominant portion of the biopharmaceutical marketplace[103–105] and ~60 antibodies are FDA approved as therapeutics (https://www.antibodysociety.org/resources/approved-antibodies/, Drugs@FDA) and hundreds more are under clinical development[106]. Due to the complicated nature of development, testing, and regulating human therapeutics, the time, cost and effort spent to develop antibody therapeutics is immense. On average, it takes 12 years and ~$2,600,000,000 to successfully develop a new therapeutic[107,108]

Until recently, using the fully evolved adaptive immune system to develop natural antibodies was difficult or unmanageable; as such, many current therapeutic antibodies are engineered or designed molecules rather than naturally evolved antibodies. Natural development of an antibody requires a host be challenged with the antigen target of interest and obvious ethical concerns prohibit the direct challenging of humans. Instead, animals, usually mice, are challenged and the resulting naturally develop murine antibodies are collected. While mouse antibodies are usually immunogenic and cannot be used directly as therapeutics[109], the sequences of the variable domains can be used in two forms of therapeutics, chimeric and humanized antibodies. Chimeric antibodies are created by replacing the human variable domains of an antibody with the developed

murine variable domains[110]. Humanized antibody domains are created by grafting the CDRs from the murine domains into a FR consensus sequence generated from human variable domains[111]. Alternatively, antibody domains can be engineered completely *in vitro* by using display technology platforms to rapidly screen millions of different sequences for binding potential[112].

Whichever method is used during development, the resulting antibodies are not natural molecules. Humanized and *in vitro* engineered antibodies are distinctly non-natural as they contain variable domains whose sequences were developed completely outside of the selection and maturation filter of natural human antibody development. In either case, these domains may contain "defects" that would have been screened out during natural maturation. Such defects would likely remain undetected, as current evaluation of potential antibody therapeutics is often limited to analysis of stability via thermal melt and measurement of binding affinity[113,114]. If present, such defects in the variable domains would be particularly problematic. Variable domain defects could potentially disrupt the antibody's specificity or binding efficacy and this would result in a large detrimental effect on the therapeutic value and potentially introduce harmful side effects.

In the work described below, three V_H domains from three therapeutic antibodies mAb A, mAb B, and mAB C will be examined. mAb A and mAb B, when produced in Chinese Hamster Ovary cells, were found to have a significant population that lacked the intramolecular disulfide bond in the V_H domain, while mAb C had no such issue [115,116]. mAb B lacking this disulfide bond was also found to have a significant loss of binding affinity, demonstrating a clear role for this issue in therapeutic efficacy[116].

mAb A is a humanized anti-CD20 antibody[115]. A topology map of the V_H domain of mAb A (V_{HA}) is shown in Figure 1.7. mAb B is a humanized anti-IgE antibody[116]. A topology map of of the V_H domain of mAb B (V_{HB}) is shown in Figure 1.8. mAb C is a humanized antibody that targeting HER2 and is used in the treatment of HER2 receptor positive breast cancer[117]. A topology map of of the V_H domain of mAb C (V_{HC}) is shown in Figure 1.9.

These three V_H domains from humanized antibodies share a high sequence identity (Fig. 1.10). The differences mainly arise in the residues of, or directly adjacent to, CDR1-3, which is to be expected as they all have distinct targets and they share a common framework consensus sequence[118–120]. Interestingly, their sequences do diverge in a few framework residues that are not directly adjacent to CDR1-3, but the overall hydrophobicity at these positions is largely conserved (Fig. 1.10).

Despite the level of similarity between the domains, the unique behavior of V_{HC} with regards to disulfide bond formation indicates that their folding pathways may not be conserved, unlike behavior seen in other protein families [121–123]. Mapping the folding mechanisms of these domains will not only provide insight into what gives rise to the varying levels of disulfide bond formation in these domains, but will also provide a unique opportunity to explore how human-directed development efforts may impact folding mechanisms when the natural filters of evolution and antibody maturation are removed. It may also provide insights into potential methods for re-engineering future

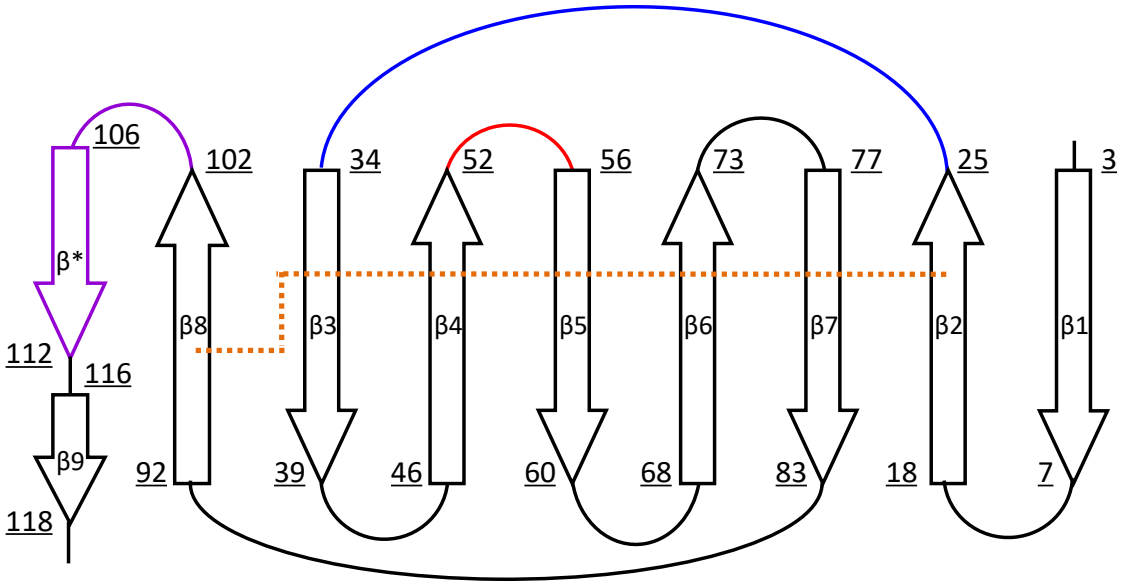

Figure 1.7: Topology map of V_{HA} domain
A topology map of the V_H domain of mAB A is shown. The β-strands are numbered and their beginning and ending residues are shown. The intramolecular disulfide bond is shown by the orange dotted line and CDR1-3 (blue, red, and purple, respectively) are shown.

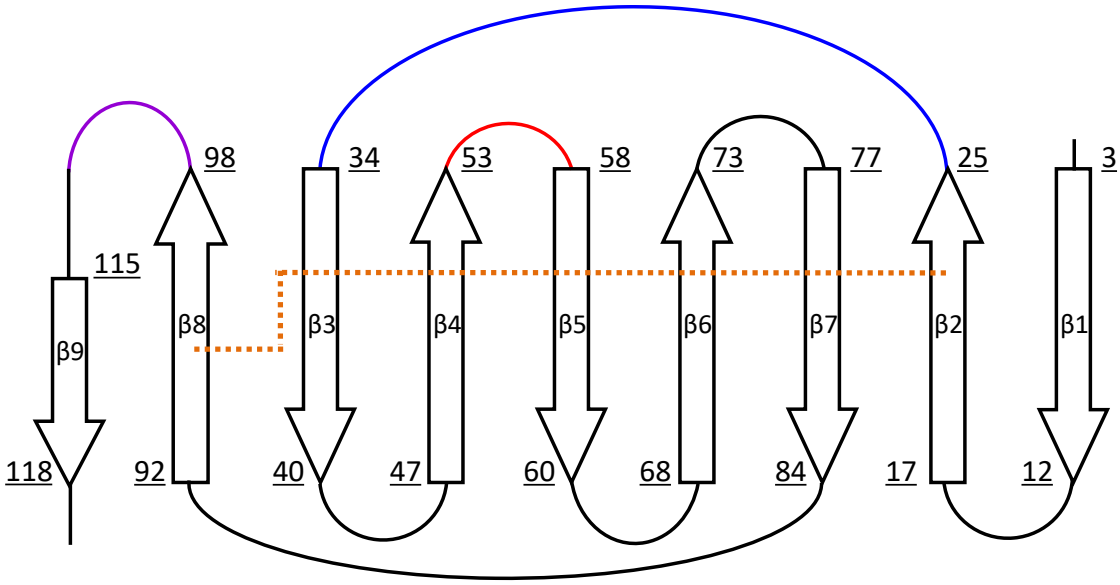

Figure 1.8: Topology map of V_{HB} domain
A topology map of the V_H domain of mAB B is shown. The β-strands are numbered and their beginning and ending residues are shown. The intramolecular disulfide bond is shown by the orange dotted line and CDR1-3 (blue, red, and purple, respectively) are shown.

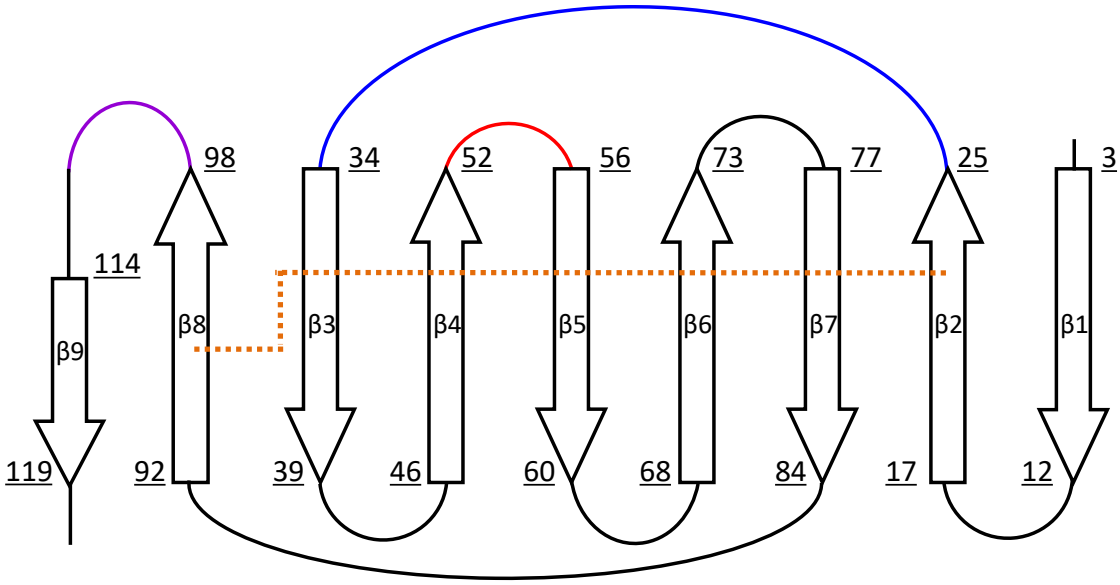

Figure 1.9: Topology map of V_{HC} domain
A topology map of the V_H domain of mAB C is shown. The β-strands are numbered and their beginning and ending residues are shown. The intramolecular disulfide bond is shown by the orange dotted line and CDR1-3 (blue, red, and purple, respectively) are shown.

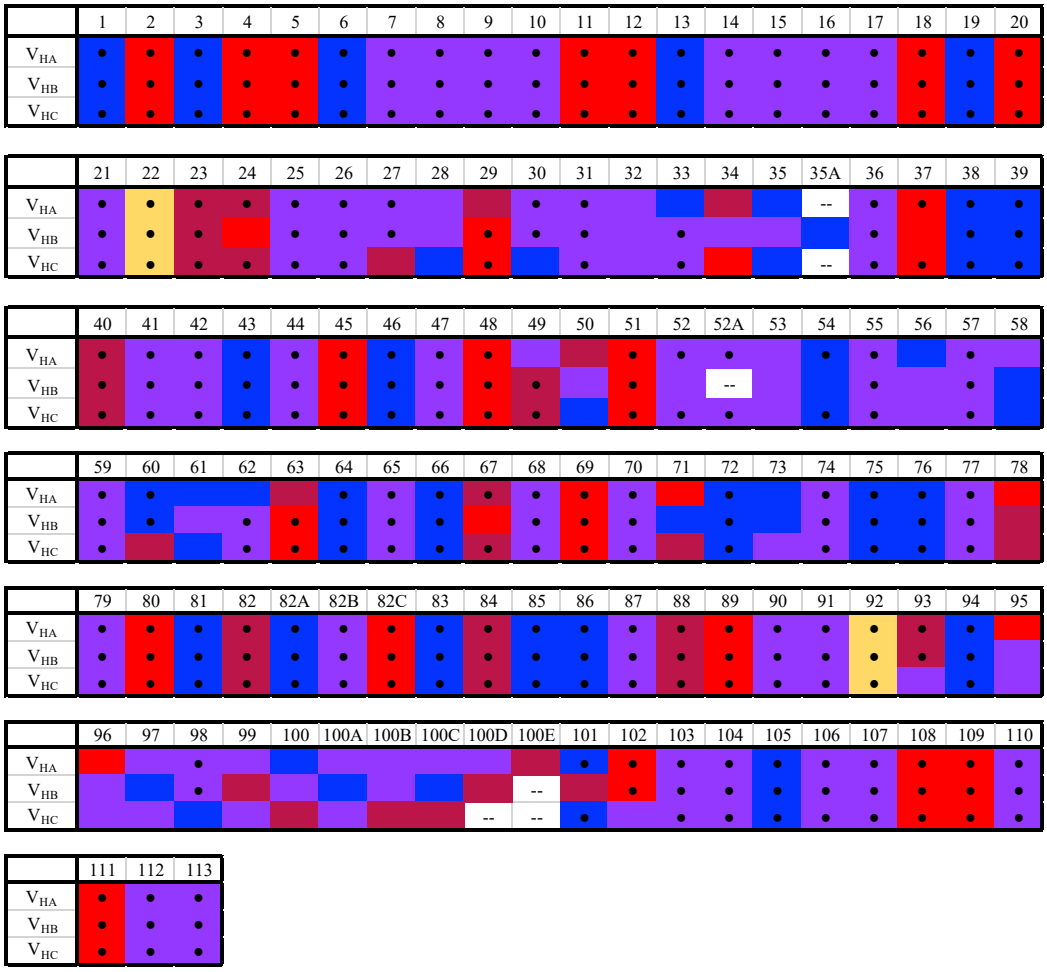

Figure 1.10: Sequence alignment of mAb A, B, and C V_H domains

A sequence alignment map of the V_H domains from mAB A, B, and C is shown. The alignment was performed using the abYsis software[124]. The exact residues are not shown to preserve intellectual property of Genentech Inc. Identical residues are indicated with a dot. Relative hydrophobicity of residues is indicated by the color scale with red-most hydrophobic to blue-most polar. The conserved cysteines that for the disulfide bond are shown in yellow. Gaps in the sequences are indicated by --. The residues are numbered according to the Kabat numbering system[125].

therapeutic sequences to favor fully oxidized disulfide bonds and simple folding behaviors.

SCOPE OF THESIS

The scope of this thesis is to describe the folding free energy landscapes of several β-sandwich proteins that have direct links to human disease and medicine. While each protein is a unique background in which to study the competition between aggregation and folding, the goal is to identify general mechanisms or early events in protein misfolding and/or aggregation that could present novel targets for therapeutics to disrupt disease-linked aggregation processes. Doing so might also allow early and rapid identification of therapeutic molecules that may be susceptible to misfolding and aggregation.

Chapter 2 will present work examining the behavior of short peptides of the SOD1 sequence in solution with potential implications in the neurodegenerative disease ALS. Using a version of maximum entropy modeling to analyze the time-resolved lifetime decays of a series of FRET donor-only and donor-acceptor labeled peptides revealed that only the peptide corresponding to the C-terminal region of SOD1 forms local nonnative structure. Surprisingly, truncation of the polar residues at the N-terminus and mutation of a hydrophobic residue at the C-terminus of this peptide both disrupted the nonnative structure, indicating that both regions are involved its formation. We propose that this nonnative state could be a precursor to the pathological aggregation of SOD1 and may present a target for therapeutic interventions.

Chapter 3 is an overview of how, through modulation of solvent viscosity, the folding and unfolding of disulfide-reduced monomeric SOD1 was limited by internal chain friction, but the same phenomenon was not observed in the disulfide-intact monomer of SOD1. While the use of an established isostability method suggested the presence of such internal friction, the errors in the linear regression typically used to determine the existence and magnitude of internal friction were large. To address this, we developed an alternative method to globally analyze the kinetic data to obtain the magnitude of the chain friction. The presence of internal chain friction only in the reduced form of SOD1 suggests that the regions near the disulfide bond, Loop IV and β8, drive the chain friction. It is possible that the friction arises from these regions sample nonnative structures, which must be disrupted before proper folding can occur and that the formation of the disulfide bond disrupts these structures.

In Chapter 4, preliminary work detailing the folding mechanisms of V_H domains of three therapeutic antibodies and the important role played by the disulfide bond in driving these processes is examined. Despite the three domains being developed using similar consensus framework sequence, they fold through mechanisms of varying complexities. Interestingly, the observed complexities in these mechanisms were disrupted by reduction of the disulfide bond but this also produced significant issues with aggregation. The observation of complexities in the folding of these therapeutic molecules demonstrates that thorough biophysical analyses performed early in the development process may provide more criteria to use in the selection of potential therapeutics. Furthermore, these results may help guide future antibody design and engineering endeavors.

Chapter 5 summarizes the results of these studies and discusses them in the broader context of understanding the competition between folding, misfolding and aggregation as well as describing the future directions of these projects. The insights gained from these works provide a basis for future work focused on the development of therapeutic approaches acting through the disruption of early states populated in disease-linked aggregation. The same framework of thought applied to the selection and design of therapeutics may result in the selection and development of treatments with simple, well-characterized biophysical behaviors. Thorough examination of the biophysics of molecules related to disease will help create new therapeutics and improve those that currently exist.

Chapter II – Nonnative structure in the unfolded ensemble

This chapter is adapted from an article currently *in press*:

Nonnative structure in a peptide model of the unfolded state of superoxide dismutase 1 (SOD1): Implications for ALS-linked aggregation

Noah R. Cohen, Jill A. Zitzewitz, Osman Bilsel, and C. Robert Matthews
Journal of Biological Chemistry, 2019
doi: 10.1074/jbc.RA119.008765

INTRODUCTION

Amyotrophic Lateral Sclerosis (ALS) is a neurodegenerative disorder affecting motor neuron cells, where it causes progressive loss of muscle control and eventual death[62,126]. Patients afflicted with ALS often only survive 2-3 years after the initial onset of symptoms, demonstrating the rapid and invariably fatal effects of the disease. The majority of all ALS cases occur in a sporadic nature, and only a small fraction (~10%) are linked to heritable mutations[62,126]. The first of the familial ALS proteins to be discovered was Cu, Zn superoxide dismutase 1(SOD1)[61]. SOD1 is a homodimeric protein in which each monomer is comprised of a 153-residue chain that adopts a β-sandwich fold. Each monomer is stabilized by an intramolecular disulfide bridge and contains two metal ion cofactors, zinc and copper (Fig. 2.1A). The zinc ion has been shown to stabilize the protein, while the copper ion is crucial to the enzymatic activity of SOD1[127]. Intriguingly, >170 mutations in the *sod1* gene linked to ALS are located throughout the sequence of the protein (http://alsod.iop.kcl.ac.uk/)[64].

Although the mechanism of pathogenesis by mutations in SOD1 is poorly understood, a hallmark of ALS and many other neurodegenerative diseases is the observation of large protein-rich aggregates in affected tissues[128]. The observation of a common endpoint of large cellular inclusions has given rise to the hypothesis that ALS may result from toxicity related to protein misfolding and subsequent aggregation[129,130]. While the disulfide-intact, metal-bound SOD1 dimer is remarkably stable and resistant to aggregation[131], mutations are known to destabilize the native dimer and increase the populations of folded and unfolded forms of the monomeric protein that may nucleate aggregation[92,132–135]. Several studies have suggested specific structures for such misfolded

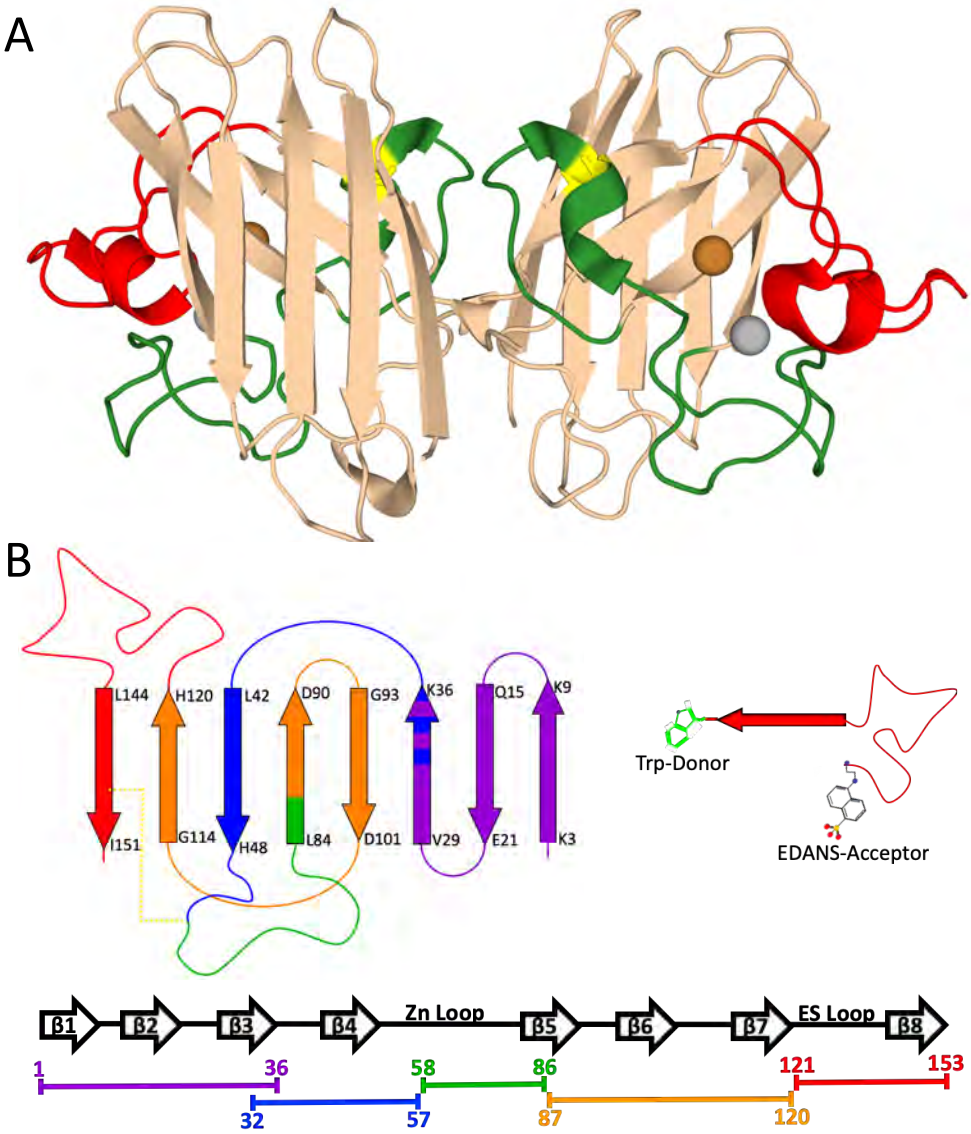

Figure 2.1: SOD1 ribbon diagram and peptide design

A, Ribbon diagram of Cu,Zn SOD1, adapted from PDB: 2C9V, is shown[71]. The bound copper and zinc metal cofactors are shown as orange and grey spheres, and the internal disulfide bond is shown in yellow. The zinc binding and electrostatic loops are shown in green and red, respectively. *B*, Each of the examined SOD1 peptides is mapped onto the topology diagram as a unique color in the top left. A sketch demonstrating how the FRET donors-acceptor pairs were placed at the termini of the peptides is shown on the top right. The exact residue length of each peptide is shown along a linear map of the SOD1 structure using the same colors as the topology map above. The colors associated with each peptide are carried forward.

monomers and nonnative dimers[44,83,84,136]. However, the large number of ALS-linked SOD1 mutations and, in particular, their distribution throughout the sequence, motivated us to propose a general explanation for the aggregation phenomenon.

Work from our lab on ALS variants of SOD1 have revealed that: (A) the destabilization of the metal-free native dimeric state increases the populations of both the folded and unfolded forms of the monomer[132], (B) the unfolded form of the disulfide-reduced monomer is favored at physiological temperatures[88], and (C) the lifetime of the unfolded monomer, prior to folding, is remarkably long, >10 s[137]. These observations support the hypothesis that the unfolded form of SOD1, after synthesis on the ribosome and prior to folding and maturation, is the source of aggregation in motor neurons. Also supporting this hypothesis are the results of Lang *et al.*[87] which showed that the lag time prior to fibrillar aggregation of SOD1 *in vitro* was linked to the magnitude of population of SOD1 in the unfolded state. Increasing the unfolded state population with denaturant led to a decrease in the lag time and increase in the fibrillation rate, implicating the unfolded state as a common precursor for the aggregation of ALS variants. This work was expanded upon by comparison of the calculated *in vitro* aggregation rates to the *in vivo* aggregation rates determined for several mouse models[47]. The good agreement argues for a common mechanism of aggregation *in vitro* and *in vivo*.

Eisenberg and colleagues have found that two heptapeptides derived from SOD1, one from the loop between β6-β7 and the other from β8, are capable of forming cross-beta structures that could serve as precursors to large scale aggregates[138]. The concept that discrete segments of SOD1 could serve as critical nucleation sites is further supported by the results of proteolytic digestion experiments on *in vitro* and *in vivo* derived aggregates

of SOD1[139]. Three different continuous segments of SOD1, spanning β1-β3 (an interlocking pair of β-hairpins), β5-β6 (a β-hairpin) and Loop VII-β8 (a region that lacks native self-contacts) were shown to be strongly protected against proteolysis *in vitro*; *in vivo*, only the Loop VII-β8 segment was protected.

Building upon these results, we investigated the propensity for spontaneous collapse in five peptides that span the entire sequence of SOD1. We found that only the peptide corresponding to the Loop VII-β8 segment of SOD1 was capable of adopting a compact structure. Surprisingly, the compact form is separated from the expanded form by an energy barrier that gives rise to an apparent two-state behavior when exposed to increasing concentrations of denaturant, rather than a gradual swelling of the segment as would be expected for a randomly collapsed state. We propose that the Loop VII-β8 segment as a promising candidate to be the primary nucleation site for the aggregation of ALS variants of SOD1 in ALS.

RESULTS

Strategy

We adopted the time-resolved FRET (trFRET) technique to measure the compaction of a set of peptides spanning the SOD1 sequence: the β1-β3 peptide (residues 1-36), the β4 peptide (residues 32-57), the Loop IV peptide (residues 58-86), the β5-β7 peptide (residues 87-120) and the Loop VII-β8 peptide (residues 121-153) (Fig. 2.1B). Three of these peptides were chosen on the basis of their resistance to proteolysis in *in vitro* and *in vivo* aggregates[139], supposing that these segments might be capable of nucleating aggregation reactions. The remaining two peptides, spanning residues 32-57 and 58-86,

completed the set spanning the entire sequence. The peptides were purchased in pairs such that one of the pair contained only a single tryptophan (Trp) donor residue at or near one terminus, serving as the donor-only peptide (DO). The second member of the pair contained a tryptophan residue as well as a 5-((2-aminoethyl)amino)naphthalene-1-sulfonic acid (EDANS) moiety conjugated to a glutamic acid residue at the other terminus, serving as the donor-acceptor peptide (DA). The Förster distance for Trp-EDANS is ~22 Å and is suitable for measuring the dimensions of peptides of these lengths.

FRET Efficiency from time-resolved fluorescence data

To test the state of compaction for each peptide we observed the excited state decay for each DO and DA pair as a function of denaturant using time-correlated single photon counting as previously described[140]. We used Maximum Entropy Modeling (MEM)[141,142] to fit these decays to rate distributions using the equation:

$$I_{DO/DA}(t) = \int p(k_d) e^{-k_d \cdot t} dk_d \qquad \text{(Eq. 2.1)}$$

where $I_{DO/DA}$ is the observed excited state decay for the DO or DA peptide, $p(k_d)$ is the distribution of excited-state decay rates, k_d is the donor decay rate constant and is also the inverse of the donor lifetime, 1/τ, and t is the time. From these distributions, we determined the quantum yield weighted average lifetime of the Trp, τ^{QY}, using the equation:

$$\tau^{QY} = \frac{\Sigma_i \alpha_i \tau_i^2}{\Sigma_i \alpha_i \tau_i} \qquad \text{(Eq. 2.2)}$$

where α is the amplitude associated with a distinct lifetime, τ, from the rate distribution. We then determined the quantum yield weighted averaged trFRET efficiency for each peptide and using the equation:

$$E_{trFRET} = 1 - \frac{\tau_{DA}^{QY}}{\tau_{DO}^{QY}} \qquad \text{(Eq. 2.3)}$$

Denaturant dependence of average trFRET efficiency

The quantum yield weighted average trFRET efficiency for the set of five peptides spanning the SOD1 sequence is plotted as a function of denaturant concentration in the top panel of Figure 2.2. Strikingly, the trFRET efficiencies the Loop IV and Loop VII-β8 peptides were dependent on the urea concentration; the trFRET efficiencies of the remaining peptides were independent of the urea concentration. The trFRET efficiencies of both the Loop IV and Loop VII-β8 peptides decreased with increasing urea concentration. The decrease for the Loop VII-β8 peptide was distinctly nonlinear and larger than that for the linear decrease for the Loop IV peptide. The expansion of the Loop VII-β8 peptide with denaturant is independent of the peptide concentration over the range from 2 to 20 μM, eliminating peptide-peptide interactions as an explanation for its behavior.

Noting that the β5-β6 segment, residues 82-102, forms a short β-hairpin in the native structure and was also found to be protected from proteolysis in *in vitro* aggregates[139], we examined the trFRET response of a peptide corresponding to this hairpin to increasingly denaturing environments (Fig. 2.3). This peptide was also found to be insensitive to denaturant.

As an independent test of the formation of structure in these peptides, we measured the far-UV CD spectra of each in the absence and presence of denaturant (Fig. 2.4). With

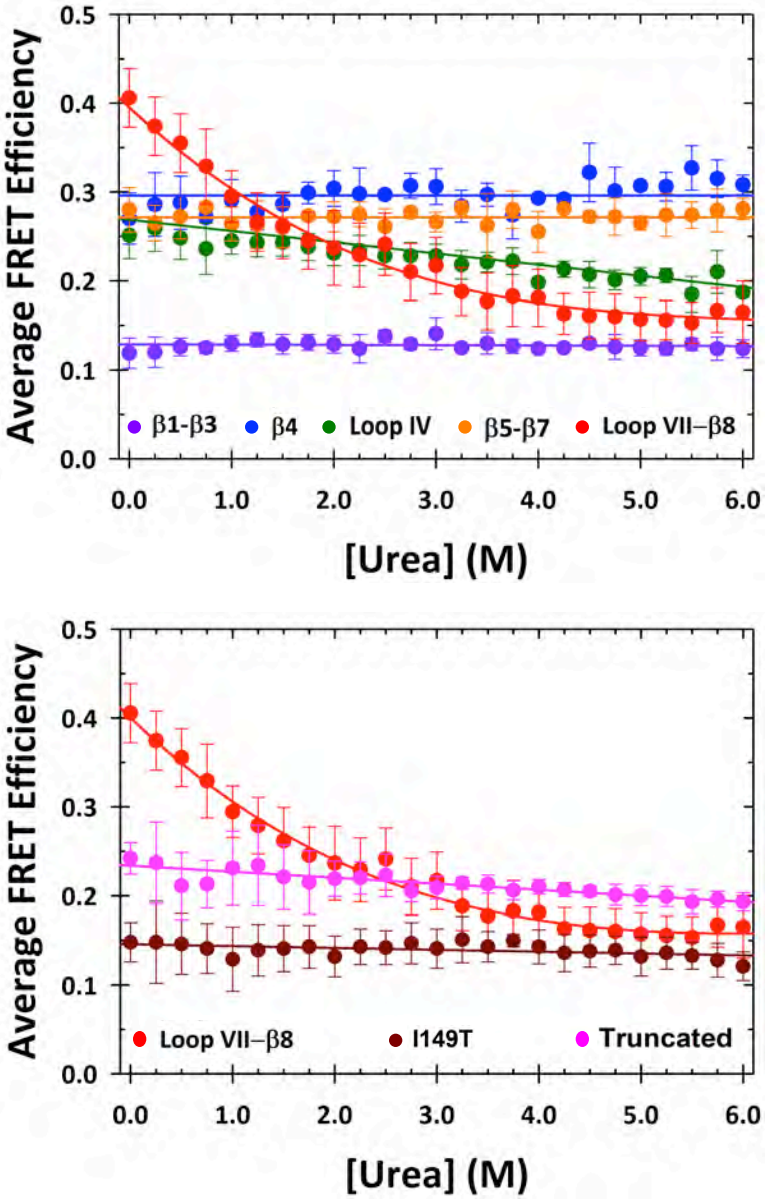

Figure 2.2: trFRET efficiency of SOD1 peptides as a function of denaturant
The quantum yield weighted average trFRET efficiencies for each of the five SOD1 peptides are shown as a function of denaturant (top, colors as in Figure 2.1B), and trend lines are shown to demonstrate the linearity or nonlinearity of the FRET change for each peptide. The FRET efficiencies for the Loop VII-β8 peptide (red) and the I149T (maroon) and truncated (pink) variants are shown as a function of denaturant (bottom) and trend lines are shown to demonstrate the linearity or nonlinearity of the FRET change for each peptide. The data points represent the average of ≥3 separate experimental runs, and the error bars are the standard deviation thereof.

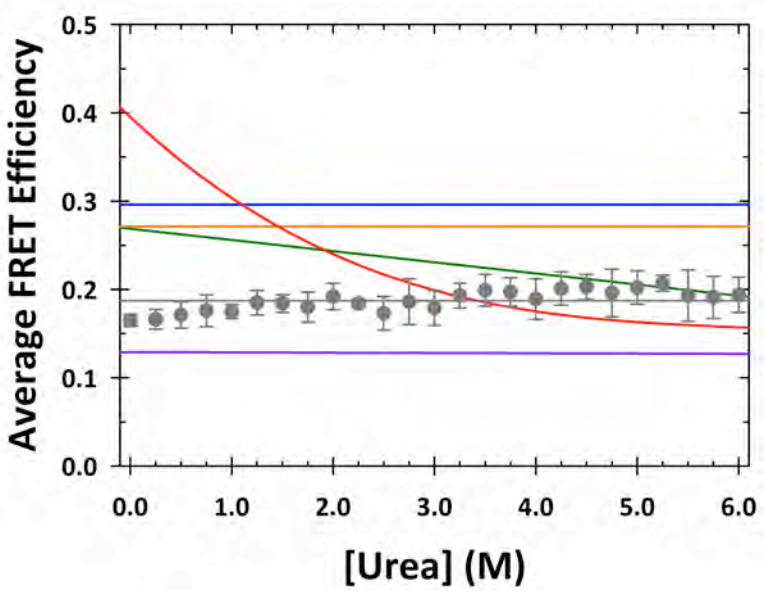

Figure 2.3: trFRET efficiency of the β5-β6 peptide as a function of denaturant
The quantum yield weighted average trFRET efficiency for the β5-β6 peptide is shown as a function of denaturant (grey circles) and trend lines for this and the other examined peptides are shown for reference (colors as in Figure 2.2). The data points represent the average of 3 separate experimental runs and the error bars are the standard deviation thereof.

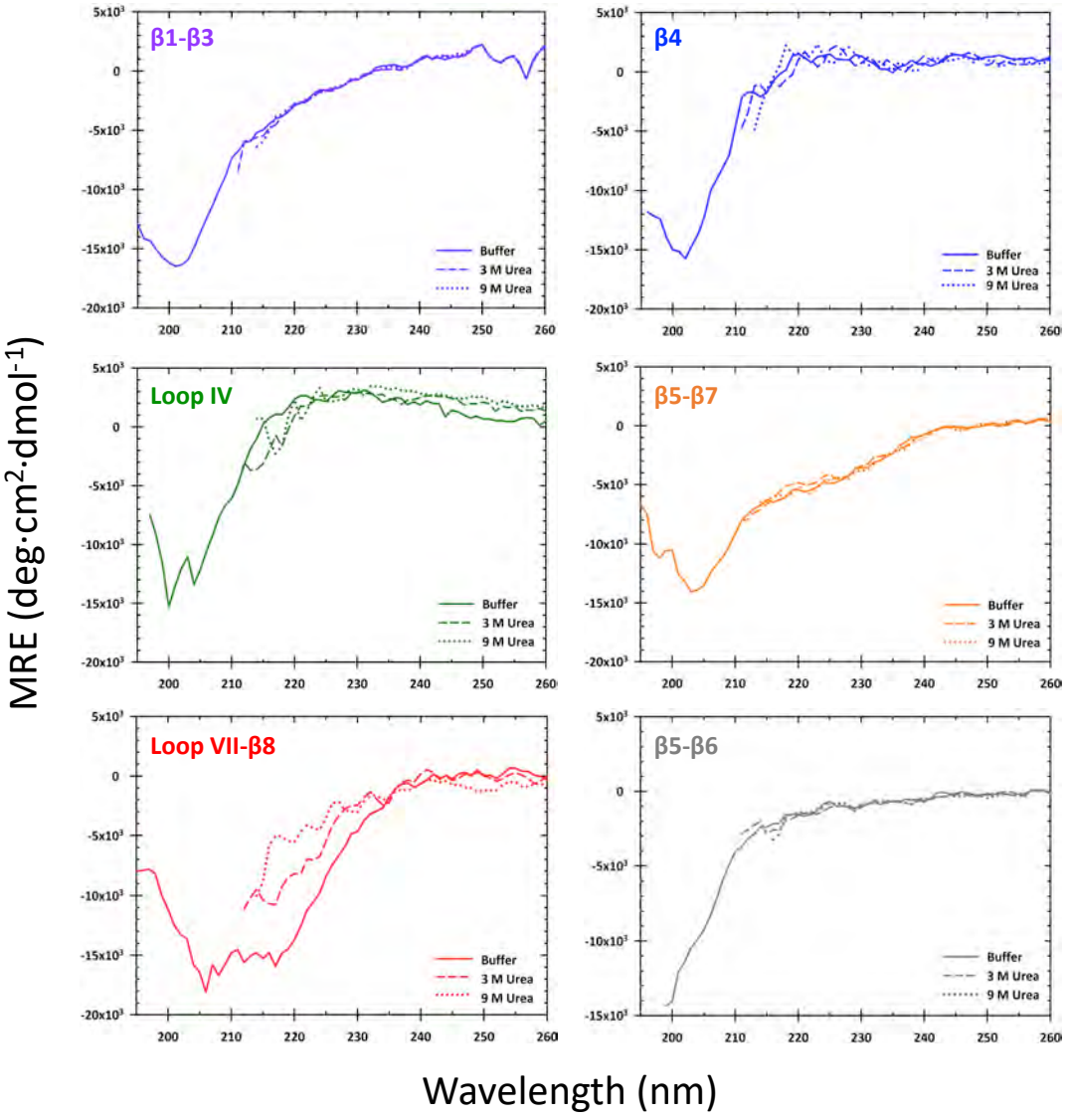

Figure 2.4: CD spectra of the examined peptides
The CD spectra of each peptide (colors as in Figures 2.2, 2.3), in aqueous (solid lines), mildly denaturing (dashed lines), and strongly denaturing (dotted lines) conditions are shown. Each panel is labeled with the peptide examined in a corner.

the exception of the Loop VII-β8, the spectra in aqueous and denaturing conditions are coincident. We conclude that none of the other peptides formed detectable secondary structure.

The Loop VII-β8 peptide is unique in that the loop segment is of low complexity and polar, typical of an intrinsically disordered protein[143], while the β8 sequence is almost entirely nonpolar. To test the involvement of these distinctly different sequences at the termini of this peptide in its trFRET efficiency, we examined the behavior of two variants. One replaces isoleucine 149 in the β8 region with a threonine, an ALS-linked mutation that eliminates protection against proteolysis for *in vitro* aggregates of SOD1[139]. The other variant truncates five residues, EKADD, from the N-terminus (Fig 2.5). We found that the trFRET efficiency for both variants in water is similar to that of the WT peptide in highly denaturing conditions (Fig. 2.2, bottom). The trFRET efficiency for the I149T variant is independent of denaturant concentration while that of the truncated variant experiences a small linear decrease, similar to the behavior of the Loop IV peptide. Surprisingly, these results show that both the polar N- and nonpolar C-termini of the Loop VII-β8 peptide are involved in stabilizing the compact form of the peptide.

Gradual swelling or discrete unfolding?

The sensitivity of the trFRET efficiency to denaturant for the Loop IV and Loop VII-β8 peptides could reflect a gradual expansion from a nonspecific collapsed state[144–146] or a two-state-like expansion from a discrete compact state to an expanded unfolded state[147]. To resolve this issue, we utilized a two-dimensional form of Maximum Entropy Modeling (2DMEM)[140]. 2DMEM is unbiased in that it makes no assumptions regarding

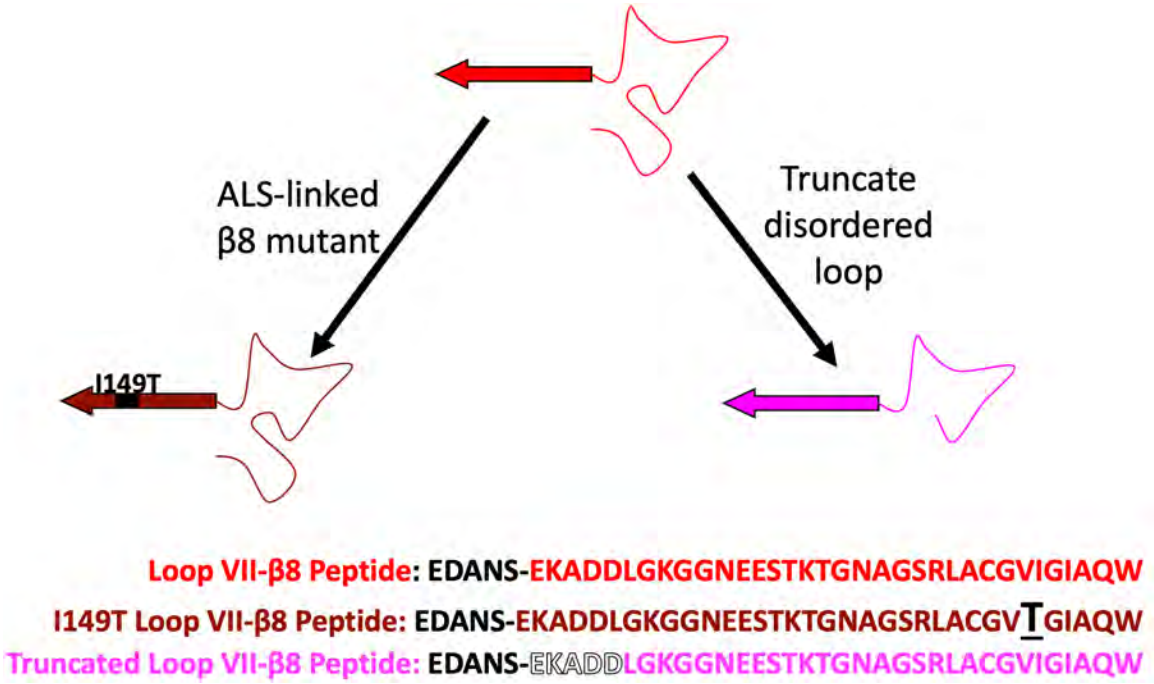

Figure 2.5: Loop VII-β8 peptide variant diagram
Diagram demonstrating the changes made to the Loop VII-β8 peptide to form the examined variants is shown at the top. How the primary sequence was changed in these variants is shown at the bottom. Colors used as in Figure 2.2.

the number of subpopulations or their lifetimes and allows for the possibility that each could have a distinct distance distribution[140]. By analyzing our time-resolved fluorescence data with MEM, we can distinguish states that are resolved on the nanosecond time-scale. In 2DMEM, the raw excited state decays of the DO and DA peptides are globally fit using a 2D rate distribution matrix of donor decay, k_d and energy transfer, k_{ET}, rate constants using Eq. 2.1 and the following:

$$I_{DA}(t) = \iint p(k_d, k_{ET}) e^{-(k_d + k_{ET})\cdot t} dk_d dk_{ET} \quad \text{(Eq. 2.4)}$$

There are two striking features in the results of the 2DMEM analysis of the Loop VII-β8 peptide (Fig. 2.6A). First there are multiple values of k_d and k_{ET} operative in the excited state decay. The three donor decay rate constants at 0.3, 1, and 8 ns^{-1} are characteristic of tryptophan[148], but only the slower two are FRET competent. The presence of both low and high FRET states with k_{ET} values of <1 and >1 ns^{-1}, respectively, indicate the presence of both expanded and compact states of the peptide. As the urea concentration increases, the population of the compact, high FRET state decreases while the population of the expanded, low FRET state increases in a concerted manner (Fig. 2.6B-3E). The results show that the change in trFRET efficiency for the Loop VII-β8 peptide reflects a two-state-like process between compact and expanded states separated by a kinetic barrier that is long relative to the nanosecond timescale of Trp excited-state decay. A denaturant-induced expansion of a randomly collapsed peptide would have produced a continuous shift of the high FRET state to the low FRET state. As a control, 2DMEM analysis of the β1-β3 peptide data in the absence of denaturant reveals only an expanded, low FRET state (Fig. 2.6F). The observation of a single donor decay rate constant for this

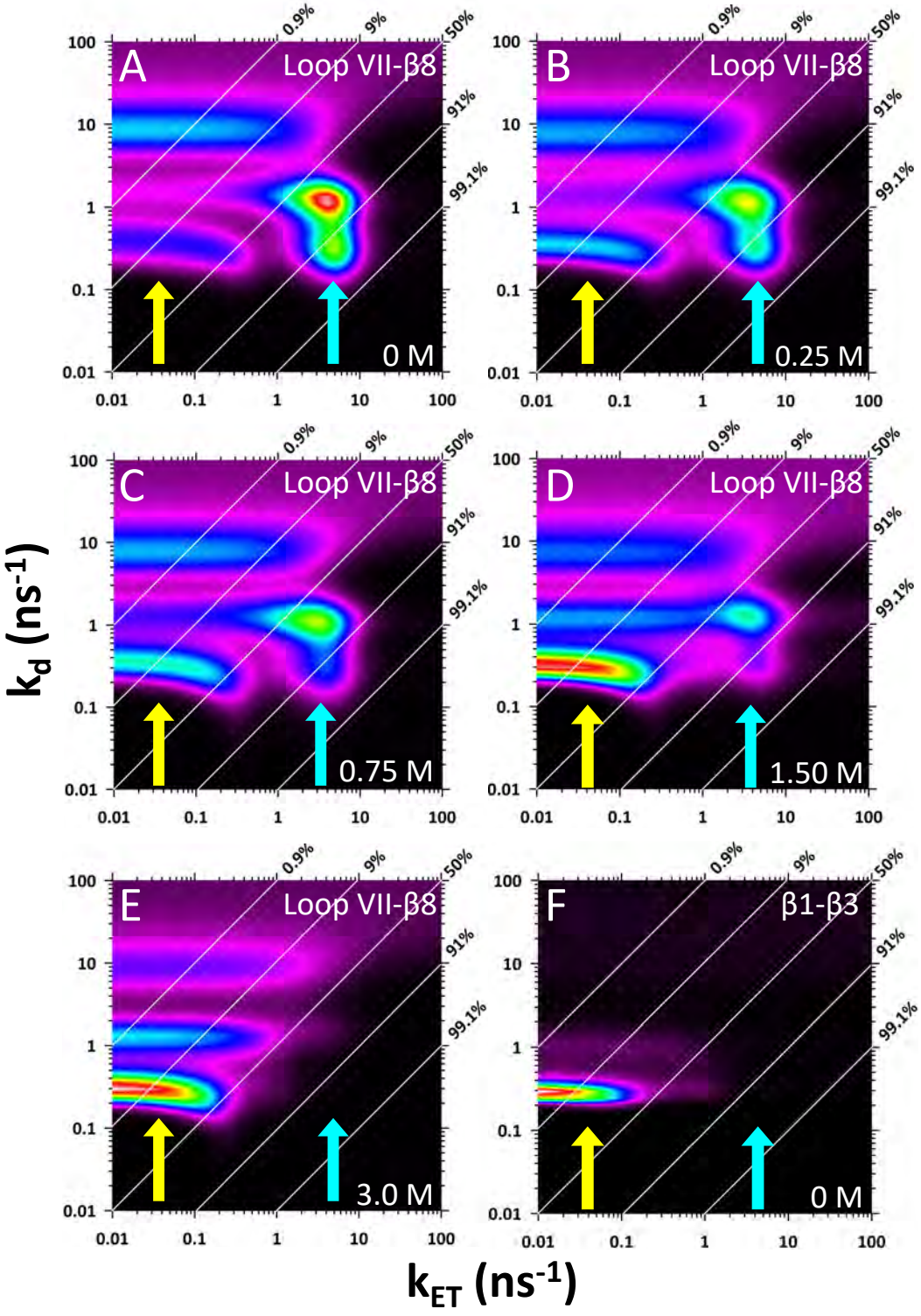
A
Loop VII-β8
0 M
B
Loop VII-β8
0.25 M
C
Loop VII-β8
0.75 M
D
Loop VII-β8
1.50 M
E
Loop VII-β8
3.0 M
F
β1-β3
0 M
0.9%
9%
50%
91%
99.1%
100
10
1
0.1
0.01
0.01
0.1
1
10
100
k_d (ns^{-1})
k_{ET} (ns^{-1})

Figure 2.6: 2DMEM distributions for the Loop VII-β8 peptide as a function of denaturant

A-E, The concentrations of urea examined are shown in the lower right corners of each panel, and the upper right corner of each panel is labeled with the examined peptide. The amplitude of the distributions in *B-E* was normalized to the amplitude in *A* to emphasize the change in response to denaturant. *F*, The 2DMEM distribution for the β1-β3 peptide in the absence of denaturant is shown for reference. The diagonal white reference lines are marked with the FRET efficiency along each line. The yellow arrows indicate the amplitude associated with extended conformations, and the cyan arrows indicate the amplitude associated with compact conformations.

peptide suggests that sequence differences in proximity to the indole side chain influence the rotamer distribution, which is often offered as an explanation for the multiple lifetimes for Trp[149,150]. The β1-β3 peptide is unique among the examined peptides as the Trp donor is not at a terminus of the peptide, and the presence of residues before and after the Trp may give rise to the single lifetime distribution.

We also used 2DMEM analysis to examine the behavior of the Loop VII-β8 peptide variants in water (Fig. 2.7A-C). Neither the truncated nor the I149T variants were found to populate a compact, high FRET state, and both had trFRET efficiencies comparable to that for the WT Loop VII-β8 peptide under strongly denaturing conditions. We conclude that both the polar residues in the electrostatic loop and the nonpolar residues of β8 are involved in the formation of the compact state. The absence of a compact state for the truncated variant eliminates the possibility that other, intervening segments of the Loop VII-β8 peptide are involved in the compact state and indicates that the donor-acceptor probes do not drive the formation of the compact state.

We also used 2DMEM analysis to examine the Loop IV peptide in water and found that it does not populate a compact, high FRET state distinct from the expanded low FRET state (Fig. 2.7D). The continuous intensity from the low to high FRET region reflects the breadth of the conformational ensemble sampled by this peptide, and the small shift in response to increasing amounts denaturant (Fig. 2.8) is evidence that the change in trFRET observed in Fig. 2.2 reflects a gradual expansion of the molecule.

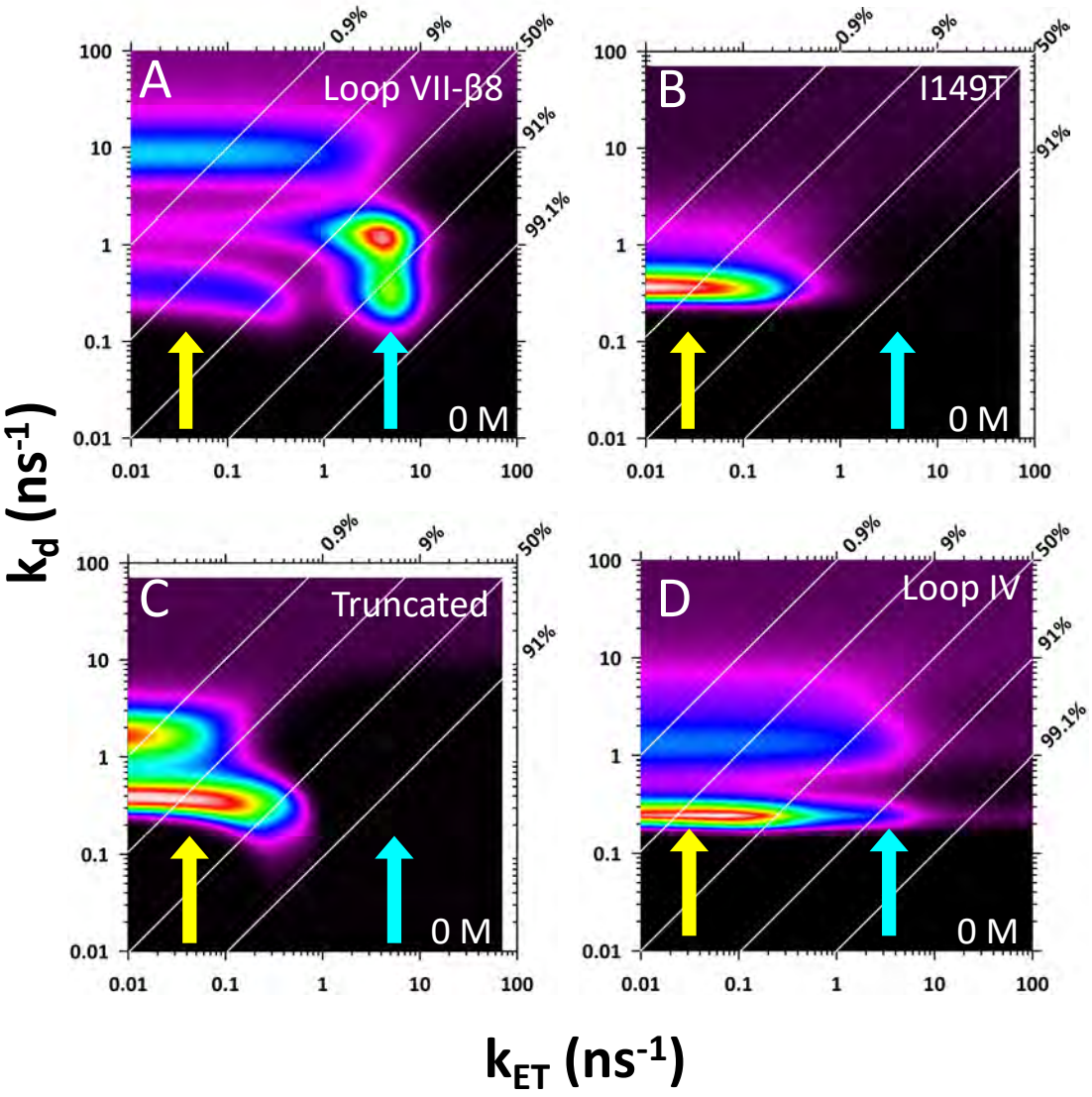

Figure 2.7: The 2DMEM distributions for the Loop VII-β8 peptide variants
A-D, The concentrations of urea examined are shown in the lower right corners of each panel, and the upper right corner of each panel is labeled with the examined peptide. The diagonal white reference lines are marked with the FRET efficiency along each line. The yellow arrows indicate the amplitude associated with extended conformations, and the cyan arrows indicate the amplitude associated with compact conformations.

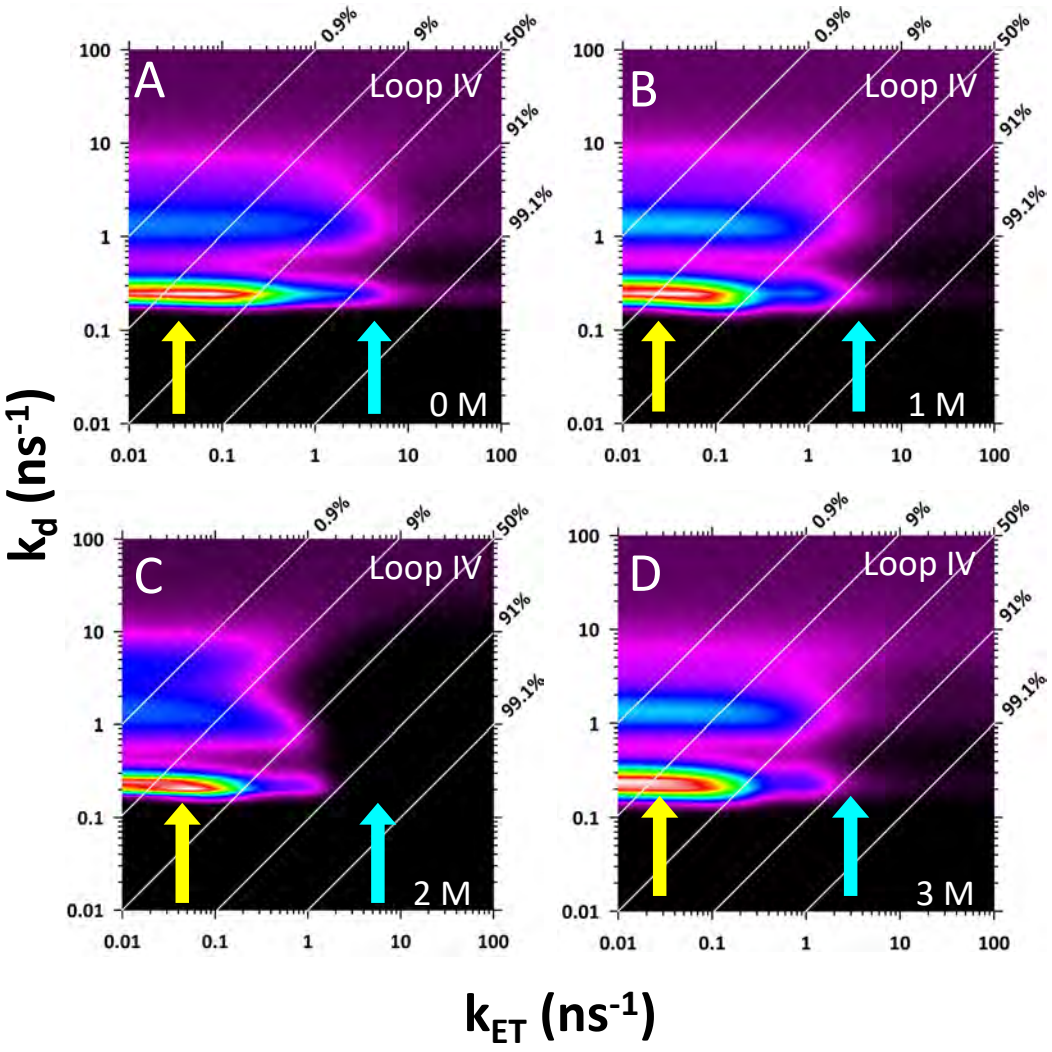

Figure 2.8: 2DMEM distributions for the Loop IV peptide as a function of denaturant *A-D*, The concentrations of urea examined are shown in the lower right corners of each panel and the upper right corner of each panel is labeled with the examined peptide. The diagonal white reference lines are marked with the FRET efficiency along each line. The yellow arrows indicate the amplitude associated with extended conformations and the cyan arrows indicate the amplitude associated with compact conformations.

Dimensional analysis of the compact state for the Loop VII-β8 peptide in water

The energy transfer rate constant, k_{ET}, axis of the 2DMEM distributions (Figs. 2.7A and 2.8A) can be transformed into end-to-end distances through the Förster equation (see Materials and Methods). The transformation yields a matrix that provides the lifetime distribution of the donor residue, the end-to-end distance distribution for the peptide, as well as information about the donor lifetimes present in each ensemble of states. The result for the WT Loop VII-β8 peptide in the absence of denaturant is shown as a three-dimensional surface in Fig. 2.9, and the lifetime distribution of the Trp donor and the distance distribution are projected along the XZ and YZ axes, respectively.

The lifetime distribution contains three peaks at ~0.3, ~1, and ~7 nanoseconds, in good agreement with typical complex lifetime distributions of Trp[148]. The distance distribution shows a significant fraction of the population is compact with an end-to-end distance (R_{EED}) of ~14 Å that is only apparent from the two slower lifetimes. The conformational ensemble of the expanded state is very broad, with an estimated R_{EED} >30 Å. For reference, the expected R_{EED} for a random coil of 35 residues is ~45 Å[151].

DISCUSSION

The C-terminus of SOD1 forms a compact state

Our trFRET data show that only the Loop VII-β8 peptide from SOD1 is capable of forming a compact thermodynamic state in water. The concerted transition from a compact to an expanded state with urea implies an intervening energy barrier; the sensitivity to urea reflects the burial of surface area from solvent, and the presence of comparable populations of the compact and expanded states are consistent with marginal stability, that is, $\Delta G^0 \approx 0$

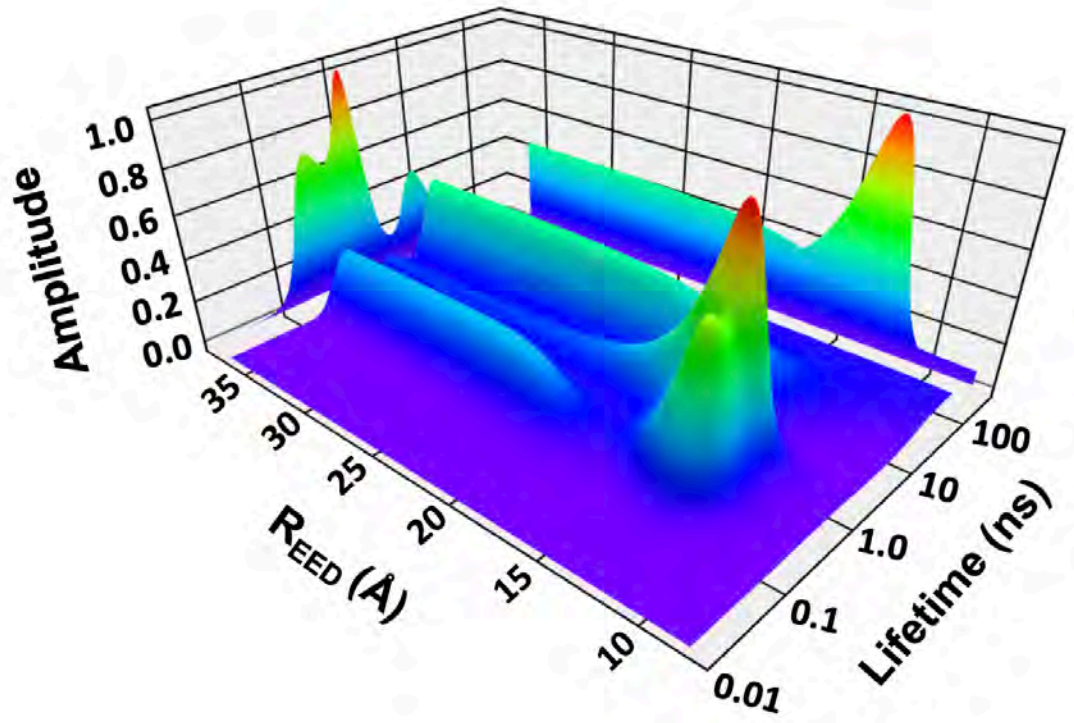

Figure 2.9: Three-dimensional surface of the transformed 2DMEM distribution for the WT Loop VII-β8 peptide in the absence of denaturant

The two-dimensional Trp lifetime distribution and the end-to-end distance distribution are shown as projections along the XZ and YZ axes, respectively.

kcal/mol. The sensitivity of the compact state to the truncation of five residues (EKAAD) from the N-terminus and the I149T mutation near the C-terminus demonstrates their mutual involvement in its structure. These same segments are not in contact in the native conformation (Fig. 2.1), suggesting that nonnative interactions are responsible for its formation. Interestingly, a peptide spanning residues 131-153 of SOD1 was used to develop an antibody recognizing misfolded, but not native, SOD1, consistent with our interpretation[152,153].

One possibility for the nonnative structure is the formation of a β-hairpin by the nonpolar residues that form β8 and a pair of glycines at positions 27 and 30 that could enable a turn (see Materials and Methods for the sequence). A small β-hairpin would explain the far-UV CD spectra observed for the Loop VII-β8 peptide in water (Fig. 2.4). This hairpin could serve to sequester the hydrophobic side chains from solvent and serve as a platform to interact with the polar residues at the N-terminus of the peptide.

Prior studies on the folding free energy surface of SOD1

Classic thermodynamic and kinetic studies of the folding of the monomeric chain of SOD1 have concluded that the reaction is well described by a two-state process, with a high barrier between the native and unfolded states[93]. However, ensemble measurements may miss complexities that could appear in the folding of an individual chain. Woodside and colleagues have examined the folding free energy surface of a single molecule of SOD1 when subjected to a constant force exerted on the N- and C-termini[44]. They found a complex set of partially-unfolded states that reflect the progressive unraveling of the β-sandwich architecture by the loss of individual β-strands. The initial step was the rupture

of the β8 edge strand, and the final steps involved the disruption of structure in the β1-β2-β3 set of interlocking β-hairpins in one face of the sandwich and β4 in the opposing face. A contrasting view of the folding free energy surface of SOD1 in water at the individual amino acid level was obtained by hydrogen exchange (HDX) NMR spectroscopy[154]. HDX-NMR is exquisitely sensitive to the presence of high-energy, partially-unfolded states of the sort predicted by the pulling experiments. The HDX-NMR results were consistent with a native basin defined by an intact β-sandwich, several of whose β-strands frayed at their termini. The majority of the main chain amide hydrogens in the β-strands exchanged in the unfolded state, i.e., not on the native side of the barrier to unfolding. There was no evidence for the partially-folded structures detected by the pulling experiment. Consistent with these results are those of kinetic folding studies of SOD1 by CD spectroscopy that did not detect the formation of secondary structure prior to the rate-limiting step in folding[137], i.e., on the unfolded side of the barrier to the native state. With the exception of the C-terminal peptide, the failure to detect compaction in peptides modeled after obvious elements of substructure in SOD1 in our trFRET experiments are in agreement with the HDX-NMR and CD studies. In other work, the partially-unfolded states observed by pulling experiments have been found to depend on the location of the tethers to exert the pulling force[155,156]. It remains to be seen how to relate the energy surfaces generated by pulling experiments on a single molecule with ensemble measurements using hydrogen exchange or chemical denaturation.

A mechanism for the aggregation of SOD1 by the Loop VII-β8 segment

The >10 s lifetime of the unfolded state of SOD1[93], prior to folding and maturation, would promote the formation of the compact structure at the C-terminus. A conceptual model for the mechanism by which this nonnative structure could lead to aggregation is shown in Figure 2.10. The intrinsically long lifetime of the unfolded state combined with its being favored under physiological conditions by ALS mutations[88], provides an environment in which the intramolecular interactions in a single chain are replaced by intermolecular interaction between pairs of chains. These dimers could then progressively add chains to form a collection of soluble oligomers. The high local concentration of SOD1 chains in these oligomers would enhance the subsequent maturation to fibrils or amorphous aggregates. The observation that only the C-terminus is protected against proteolysis in aggregates extracted from mouse motor neurons suggests that this region is the primary driver of aggregation.

We used the aggregation prediction algorithms Aggrescan[157], WALTZ[158], and TANGO[159] to examine which of the peptides would be predicted to be prone to aggregation (Fig. 2.11). While two peptides, β4 and Loop IV, were predicted to have little or no aggregation prone stretches, the other peptides, β1-β3 and β5-β7 and Loop VII-β8, were predicted to have several regions that could be prone to aggregation. Interestingly, the only peptide where the three different algorithms show consensus is the Loop VII-β8 peptide and it is also the most hydrophobic region of the SOD1 sequence[160].

This mechanism does not account for the properties of the I149T mutation. I149T disrupts the compact structure in the C-terminus and eliminates protection against proteolysis for *in vitro* aggregates[139]; however, it causes ALS. The observation that the

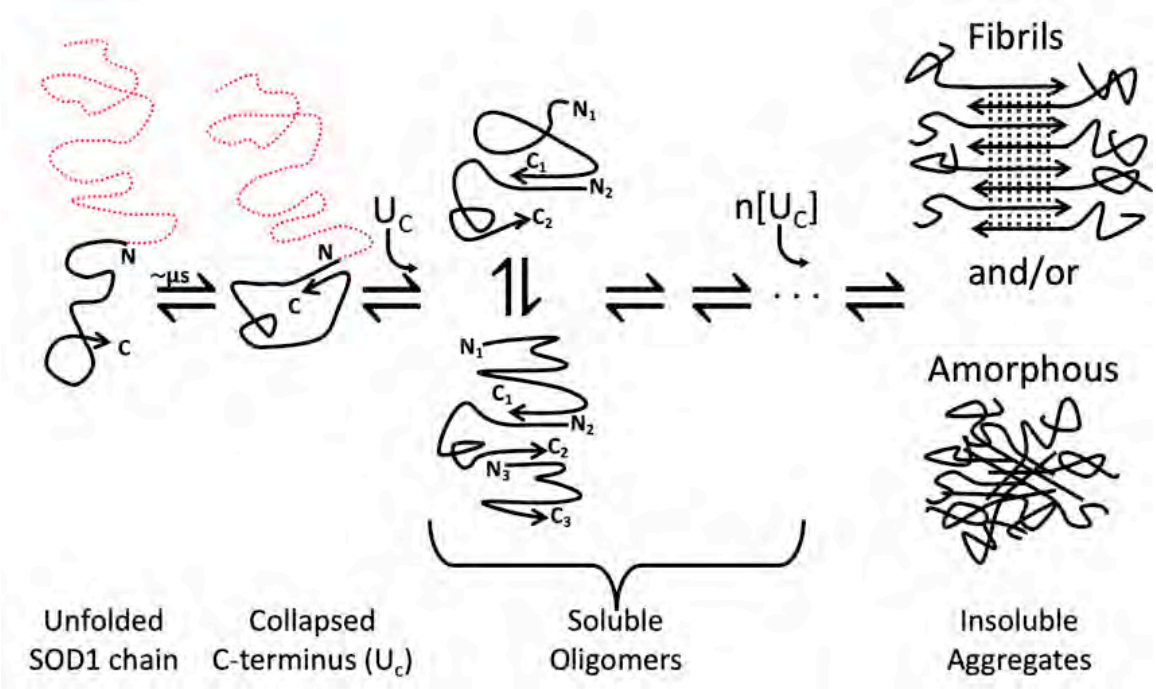

Figure 2.10: A model for the role of the compact Loop VII-β8 region in ALS-linked SOD1 aggregation

In this model, the C-terminal region (black line) is able to form the compact state, U_c, while the rest of the polypeptide chain remains unfolded (red dotted line). Once the compact state has formed, it can meet other compact states and form lower order, soluble oligomers. These oligomers could then further recruit polypeptide chains until insoluble aggregates, either fibril-like or amorphous, are formed.

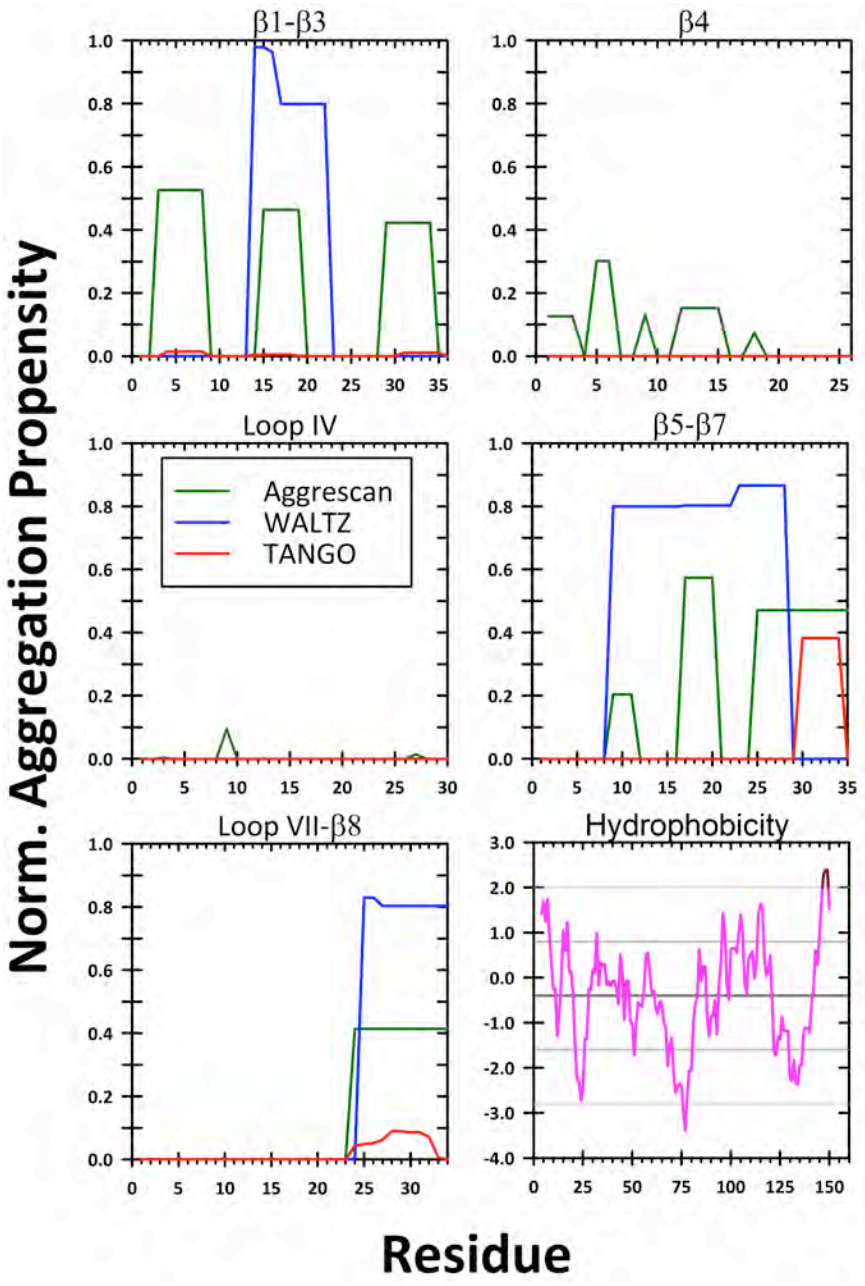

Figure 2.11: Aggregation propensity of the SOD1 peptides

The aggregation propensity score for each of the examined peptides determined by the three algorithms, Aggrescan (Green), WALTZ (blue) and TANGO (red) is shown in the first five panels. The final panel shows the Kyte-Doolittle hydrophobicity along the SOD1 sequence (pink). The mean (dark grey), one standard deviation (grey), and two standard deviation (light grey) lines are shown for reference. The most hydrophobic region, corresponding to part of the β8 sequence, is colored in dark red.

β1-β3 and the β5-β6 segments are protected against proteolysis for *in vitro* aggregates of I149T shows that secondary nucleation sites come into play when the primary site is destabilized. Further support for alternative nucleation sites is the existence of the ALS-linked variants L126X, G127X, G141X, and C146X which truncate varying amounts of the C-terminus of SOD1[64]. The L126X variant was also shown to form *in vitro* aggregates with protection patterns similar to I149T[139]. The existence of multiple pathways through which SOD1 can aggregate may also explain how different ALS-inked mutants give rise to morphologically distinct aggregates[72].

Another noteworthy element of the mechanism we propose is that it is not dependent upon the presence of an ALS-linked SOD1 mutation. The long lifetime of the unfolded chain, >10 sec, would provide an opportunity for the C-terminal region to sample the nonnative structure even in the WT-SOD1 background. The chain containing this nonnative structure could then be recognized by other misfolded protein partners, thereby providing an explanation for the recent observation of aggregated WT-SOD1 in tissue from patients with pathological ALS mutations in other genes[72].

Implications for therapeutic targets

The swap of intramolecular interactions for intermolecular interactions between C-terminal segments of unfolded SOD1 is the key step in the progression of events leading to soluble oligomers and macroscopic aggregation (Fig. 2.10). One therapeutic approach would be to screen for small molecules that preferentially stabilize the intramolecular, compact state and favor productive folding to the native dimer. Most ALS variants of SOD1 are extremely stable in their fully mature dimeric states[161,162]. An alternative

approach, shown to be effective in non-human primates[163], would be to knockdown the expression of SOD1 with siRNAs and, thereby, slow the bimolecular steps crucial to the formation of soluble oligomers and aggregates. While not offering a cure for this tragic disease, these approaches or others might delay the onset and/or slow the progression of ALS.

MATERIALS AND METHODS

Peptide Design and Preparation

The peptides spanning the SOD1 sequence used in the FRET experiments were ordered form New England Peptide Inc., where the proper molecular weight was checked by mass spectrometry and the purity was ensured to be >95% by HPLC. The sequences corresponding to the donor-only peptides follow:

β1-β3 peptide: TKAVCVLKGDGPVQGIINFEQKESNGPVKVWGSIK

β4 peptide: WGSIKGLTEGLHGFHVHEFGDNTAGC

Loop IV peptide: WTSAGPHFNPLSRKHGGPKDEERHVGDLGN

β5-β7 peptide: VTADKDGVADVSIEDSVISLSGDHCIIGRTLVVHW

Loop VII-β8 peptide: EKADDLGKGGNEESTKTGNAGSRLACGVIGIAQW

I149T peptide: EKADDLGKGGNEESTKTGNAGSRLACGVTGIAQW

Truncated peptide: LGKGGNEESTKTGNAGSRLACGVTGIAQW

β5-β6 peptide: GDLGNVTADKDGVADVSIEDSW

The donor-acceptor peptides are comprised of the same sequences with a Glu-EDANS moiety at the terminus opposite the Trp residue.

Small amounts (~1-2 mg) of the peptides were weighed out and resuspended in our experimental buffer: 20 mM HEPES, 1mM EDTAN, and 1 mM TCEP at pH=7.2. Each peptide stock was filtered through a polyethersulfone 0.22 μm filter and the concentrations were then determined using the 280 absorbance peak from the Trp donor (ε_{280} = 5690 $M^{-1}cm^{-1}$) using a Cary 100-UV/Vis spectrophotometer. For the CD analysis, the same method was used, but the buffer used was 10 mM KPi, 1mM EDTA, and 1 mM TCEP at pH=7.2.

CD Spectroscopy

All CD measurements were performed with a Jasco-810 spectropolarimeter (Jasco Inc., Easton, MD) using a water-cooled Peltier unit. The spectra of the SOD1 peptides from 190 to 260 nm were monitored using a synthetic fused silica 0.1-cm path-length cuvette at a scan rate of 20 nm per min and an 8 second response time. The same methods were used to collect denatured spectra of the peptides, yet from only 215-260 nm to avoid scattering.

Time-Resolved FRET Measurements

The excitation was provided by the tripled output of a Ti:sapphire laser at 3.8 MHz repetition rate as previously described[140,147]. Excitation power was kept at several hundred μWatts at 292 nm. Counting rates in the detector channel were limited to 1×10^5 counts per second. A 357 nm 40 nm wide bandpass filter (Semrock, Rochester, NY) was used to select the Trp emission.

The samples were prepared from stocks of peptide in buffer and peptide in ≥ 6 M urea and mixed precisely using a Hamilton Microlab 500 titrator and in-house software. Equilibrium measurements were performed by interfacing a home-built autosampler running custom LabVIEW software (available at www.osmanbilsel.net) to the lifetime instrument. Each sample in a 96-well microplate was brought into a flow cell (1mm×1mm) using a Hamilton Microlab 560 dual syringe pump and flowed in an oscillating pattern at a flow rate of 10 μL/s during data collection. Doing so reduced continuous exposure of the sample to the beam to <1 s and minimized photobleaching. An N-acetyl tryptophanamide (NATA) control was obtained prior to each titration in order to obtain relative quantum yields for all of the samples.

Maximum Entropy Modeling

Kumar *et al.*[164] have shown that MEM can recover the distribution of decay rates from time-resolved data. To determine the trFRET values as a function of denaturant (Fig. 2.2 and 2.3), the excited-state decay of the Trp donor in the DO and DA peptides at each urea concentration were fit using Eq. 2.1. These distributions were used to calculate the quantum-yield weighted averaged lifetimes using Eq. 2.2 from which the trFRET efficiencies were calculated using Eq. 2.3.

The 2DMEM used to further analyze the time-resolved data used Eq. 2.1 and Eq. 2.4 to simultaneously fit the excited-state decays of the DO and DA samples, respectively. The fitting of all the MEM data was done with in-house software coded in LabVIEW (National Instruments). The $p(k_d,k_{ET})$ distributions were represented by 50×50 grids of rates in logarithmic rate space. No scaling of the DO and DA traces was used in the global

fitting except in cases where a concentration difference was known to be present. Additional terms accounting for scattered light and infinite time offsets were included in these analyses, but these were negligible and exclusion of these terms did not affect the resulting distributions.

These distributions were converted into efficiency histograms and distance distributions using another in-house software package that applies a quantum-yield correction to the R_0 of the donor-acceptor pair. The output of this software was then plotted in Figure 2.9 as a three-dimensional surface of R_{EED}, k_d, and normalized Amplitude using the online open source software plot.ly. Two-dimensional projections of the R_{EED} *vs* normalized amplitude and k_d *vs* normalized amplitude distributions were also plotted along their respective axes.

Chapter III – Folding limited by internal chain friction

This chapter is currently under review at Biophysical Journal as:

Friction-limited folding of disulfide-reduced SOD1: Implications for ALS-linked aggregation

Noah R. Cohen, Can Kayatekin, Jill A. Zitzewitz, Osman Bilsel, and C. Robert Matthews

INTRODUCTION

The proper folding of polypeptides has becomes an area of intense interest related to human health, as protein misfolding and subsequent aggregation are associated with numerous neurological diseases[9]. Aggregates containing an antioxidant protein, Cu, Zn superoxide dismutase (SOD1), were identified in the brain and spinal cord tissues of patients suffering from amyotrophic lateral sclerosis (ALS), an invariably fatal disease caused by the atrophy of motor neurons[61]. Subsequently, 160 mutations at 88 positions have been discovered in the *sod1* gene in patients suffering from ALS (http://alsod.iop.kcl.ac.uk/), making SOD1 variants responsible for a large fraction of familial ALS[64].

SOD1 is a 153-residue homodimeric β-sandwich protein containing eight anti-parallel β-strands in each monomer (Fig. 3.1). Each subunit binds a structural zinc and a catalytic copper ion and contains an intramolecular disulfide bond between C57 in Loop IV following β4 and C146 in β8. Loop IV binds the zinc ion, while Loop VII, the electrostatic loop, serves to support Loop IV and guides the anionic substrate to the redox active copper ion. The mature protein is remarkably stable and soluble in its native dimeric, metal bound, disulfide-intact form. As such, loss of metals, reduction of the disulfide-bond, and dissociation of the dimer, have been considered necessary to induce aggregation[88,132,165–168]. Consequently, attention has turned to the disulfide-reduced monomer as a potential driver of the aggregation of SOD1[87,88,134,135,162,169].

Although the folded and unfolded forms of reduced and oxidized versions of a stable monomeric form, mSOD1[170], dominate the denaturant-induced unfolding profiles, minor partially-folded states could represent a source of aggregation-prone

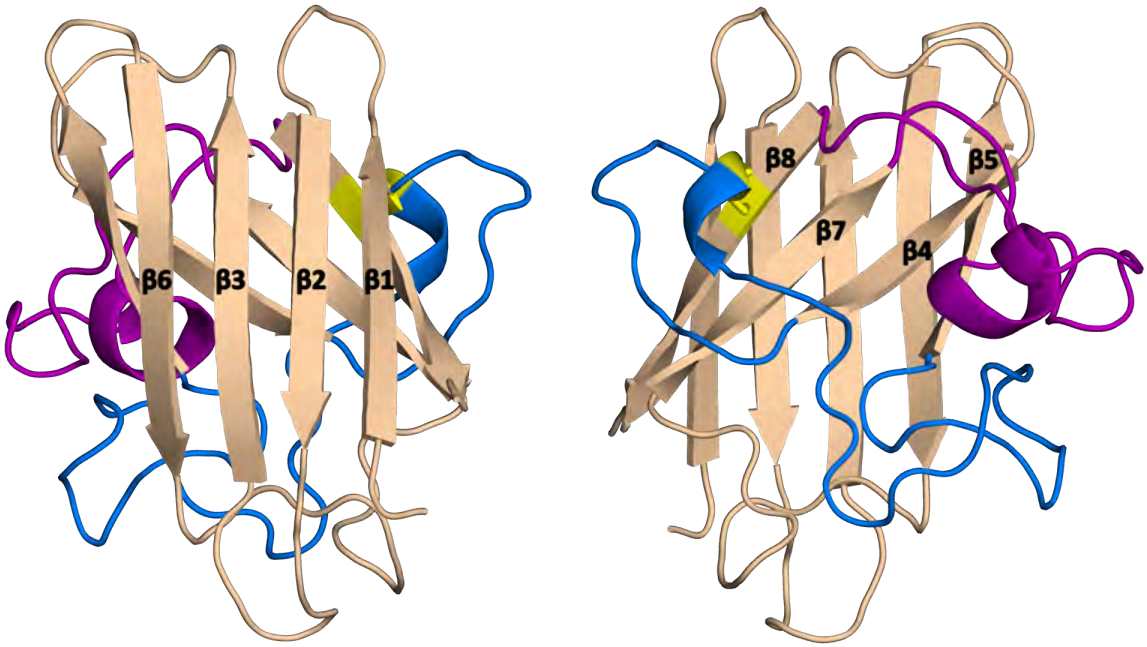

Figure 3.1: Ribbon diagram of mSOD1
The view of mSOD1 along each of the β-sheets is shown, adapted from PDB: 2C9V[71]. The eight β-strands are labeled and the zinc-binding and electrostatic loops are shown in blue and purple, respectively. The disulfide bond between C57-C146 is shown in yellow.

material[44,84,85,88,132,136]. Another source of aggregation-prone material might be found on the unfolded side of the rate-limiting barrier in the form of local condensations of the chain. By sampling the variety of conformational species present in the unfolded state ensemble, the formation of non-native contacts might be able to interfere with the productive folding of SOD1. These putative nonnative intramolecular interactions, which could be replaced by intermolecular interactions that nucleate aggregation, might be detected by their propensity to slow folding below the diffusion limit[171,172].

A general approach towards seeking these nonnative states of SOD1 is to introduce small molecule polyols into the solvent. Glycerol, glucose and other kosmotropes increase the free energy differences between folded and unfolded states by mechanisms that have been ascribed to preferential hydration of nonpolar surfaces or a decrease in the solubility of peptide bonds[173–175]. This strategy has been utilized to reveal an intermediate in the folding reaction of barstar and offers an opportunity to highlight partially-folded states in SOD1[176]. Through their capacity to increase the solvent viscosity, glycerol and glucose also provide a kinetic test for nonnative structures that can limit the rates of the folding reaction. Diffusional processes involved in folding are expected to slow with increasing viscosity and the existence of intramolecular chain friction can be determined by a less than unitary dependence of the relative folding rate constant on the relative solvent viscosity[175–177].

We found that reduced and oxidized monomeric SOD1 closely adhere to a two-state mechanism in glycerol and glucose when examined by both equilibrium and kinetic methods. Using a novel method to globally fit the kinetic data, we found that the folding and unfolding reactions of only the disulfide-reduced protein were partially limited by

internal chain friction. The simple diffusional processes observed for its oxidized counterpart suggest that the internal friction experienced by reduced SOD1 involves the formation of structure involving one or both of the segments containing the C57 and C146 partners in the disulfide bond.

MATERIALS AND METHODS

SOD1 monomer model and expression/purification

The SOD1 monomer variant, mSOD1, has a pair of nonpolar to polar mutations at the dimer interface, F50E and G51E, that prevents dimerization[170]. The two cysteines at positions 6 and 111 were replaced with alanine and serine, respectively, to eliminate disulfide scrambling with the C57-C146 disulfide bond in the unfolded state. Recombinant metal-free SS-mSOD1 was expressed and purified according to the procedure previously described[132], and 2SH-mSOD1 was created by incubation with 1 mM of TCEP for several hours at room temperature[93].

Thermodynamic and Kinetic CD experiments

CD spectroscopy was performed on a Jasco-810 spectropolarimeter (Jasco Inc., Easton, MD) equipped with a water-cooled Peltier temperature control system. The equilibrium Gdn-HCl induced unfolding transitions were monitored from 215-240 nm in a 0.2 cm path length synthetic fused silica cuvette using a scan rate of 20 nm per min and a response time of 8 s. Manual-mixing kinetic folding and unfolding reactions, with a deadtime of 3-5 seconds, were monitored at 230 nm with a response time of 1-2 s in a 1 cm cuvette. Gdn-HCl concentrations were determined by refractive index on a Leica

Mark II refractometer. The protein concentrations were 15 µM for equilibrium experiments and 5-10 µM for kinetic experiments. The buffer conditions were 20 mM HEPES, 1mM EDTA for SS-mAS-SOD1 experiments and 20 mM HEPES, 1mM EDTA, and 1 mM TCEP for 2SH-mAS-SOD1 experiments. The temperature was 20° C for all experiments. The data analyses were performed with an in-house software Savuka available online at: https://osmanbilsel.com/.

Solvent viscosity measurements

The solvent viscosity at each polyol concentration was determined using a Gilmont falling ball viscometer (Thermo Scientific). The time of travel was measured 10 times for each sample and was repeated a minimum of 2 times with fresh sample preparations. All buffers were filtered and degassed prior to the experiment.

RESULTS

Equilibrium studies

To test for partially-folded states in mSOD1 we employed circular dichroism (CD) spectroscopy to monitor the influence of glycerol and glucose on the folding free energy surface revealed by chemical denaturation. Both metal-free disulfide oxidized-mSOD1 (SS-mSOD1) and metal-free disulfide reduced-mSOD1 (2SH-mSOD1) exhibited increased stability with increasing concentrations of glycerol, as evidenced by the shift of the midpoint of the equilibrium titration, C_m, to higher concentrations of guanidine hydrochloride (Gdn-HCl)(Fig. 3.2, left panels). A similar trend was observed

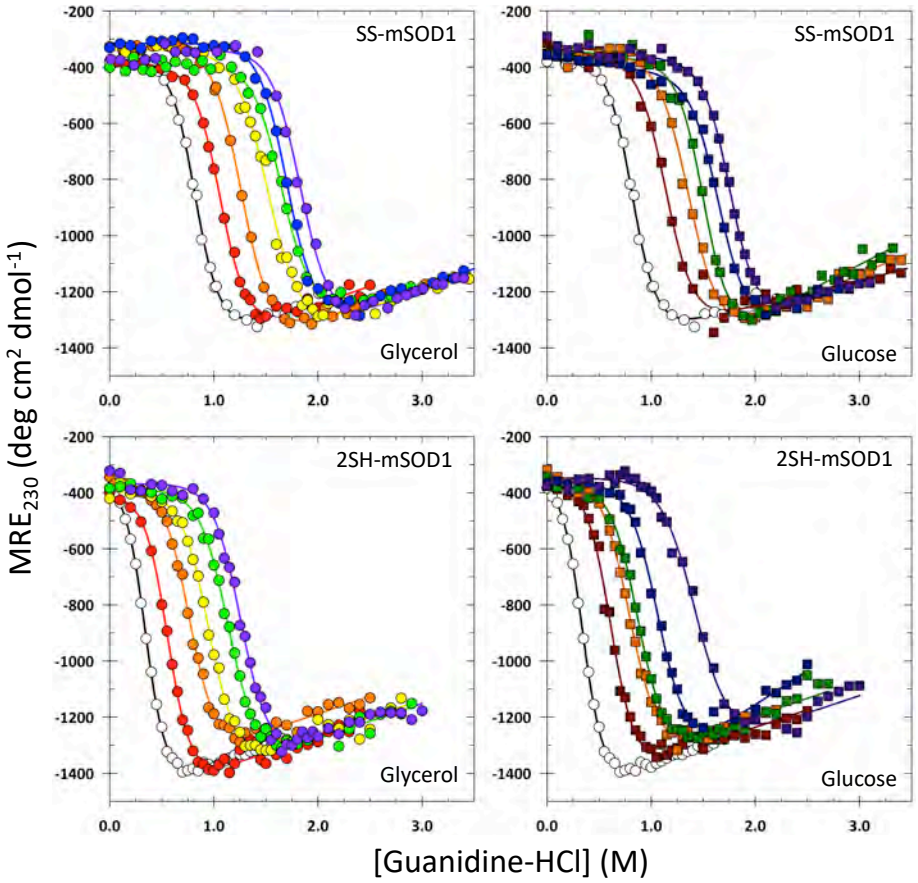

Figure 3.2: Polyol dependence of equilibrium titrations
The equilibrium titration of SS-mSOD1 (top) and 2SH-mSOD1 (bottom) at increasing concentrations of glycerol (circles) and glucose (squares), monitored by CD spectroscopy at 230 nm, are shown with a two-state fit to the data (solid lines). SS-mSOD1: glycerol percent by volume (top left): 0% (white), 10% (red), 16% (orange), 24% (yellow), 28% (green), 30% (blue) and 32% (purple). SS-mSOD1: glucose percent by volume (top right): 0% (white), 11% (dark red), 17% (dark orange), 20% (dark green), 24% (dark blue) and 26% (dark purple). 2SH-mSOD1: glycerol percent by volume (bottom-left) and glucose percent by volume (bottom-right) follow the same color scheme.

for glucose (Fig. 3.2, right panels), demonstrating that this phenomenon is not specific to the choice of kosmotrope.

The data were fit to a two-state model assuming a linear dependence of the free energy of folding on the denaturant concentration[137], and quantitative estimates of the free energy of folding are provided in Table 3.1. Over the range of cosolvents explored, 0% to 32% by volume, the unfolding transitions for both forms of mSOD1 were well-described by two-state models, and the stabilities were found to increase by ~2-fold (Tables 3.1, 3.2). Interestingly, the trend is reversed for the denaturant dependence of the free energy of folding, the m_{eq}-values, that decrease by ~25% over this range in glucose or glycerol concentrations for both SS-mSOD1 and 2SH-mSOD1 (Tables 3.1, 3.2). Although the decrease in the m_{eq}-value could reflect the presence of partially-folded states whose population is increased by the kosmotropes, companion kinetic studies of the folding reactions offer an alternative explanation (see below).

Kinetic studies

As an orthogonal test for the presence of partially-folded states, we performed kinetic folding experiments on SS-mSOD1 and 2SH-mSOD1. The unfolding and refolding reactions for metal-free SS-mSOD1 and 2SH-mSOD1 were monitored by CD spectroscopy at 230 nm. The resulting kinetic traces were well-fit by single exponential functions that provided the apparent rate constant, k, and its reciprocal, the relaxation time, τ. The excellent agreement between the ellipticities at the beginning of both unfolding and refolding reactions and the estimated ellipticities of the native and unfolded states under the same conditions, eliminates the possibility of partially-folded

Table 3.1: Thermodynamic Parameters for the folding reaction of SS-mSOD1			
Solvent Condition	**ΔG°_{eq}**	**m_{eq}**	**C_m**
0% Cosolvent	-4.10 ± 0.30	4.98 ± 0.36	0.82 ± 0.08
10% Glycerol	-4.86 ± 0.22	4.56 ± 0.23	1.07 ± 0.07
16% Glycerol	-5.40 ± 0.10	4.25 ± 0.08	1.27 ± 0.03
24% Glycerol	-5.73 ± 0.18	3.76 ± 0.12	1.53 ± 0.07
28% Glycerol	-7.04 ± 0.27	4.26 ± 0.17	1.65 ± 0.09
30% Glycerol	-7.42 ± 0.59	4.35 ± 0.36	1.71 ± 0.19
32% Glycerol	-7.64 ± 0.18	4.15 ± 0.10	1.84 ± 0.06
11% Glucose	-5.08 ± 0.33	4.41 ± 0.29	1.15 ± 0.11
17% Glucose	-5.58 ± 0.38	4.04 ± 0.28	1.38 ± 0.13
20% Glucose	-6.54 ± 0.88	4.36 ± 0.56	1.51 ± 0.28
24% Glucose	-7.09 ± 0.64	4.28 ± 0.41	1.66 ± 0.22
26% Glucose	-8.26 ± 0.22	4.68 ± 0.13	1.77 ± 0.07

The thermodynamic parameters obtained from fitting the Gdn-HCl titrations of SS-mSOD1 are shown for each cosolvent concentration examined. The equilibrium denaturant dependence, m_{eq}, is shown in kcal mol^{-1} $[D]^{-1}$ where [D] is the molar concentration of denaturant. The units of the folding free energy, ΔG°_{Eq}, are in kcal mol^{-1}. The midpoint, C_m, of each titration is shown as the [Gdn-HCl] (M).

Table 3.2: Thermodynamic Parameters for the folding reaction of 2SH-mSOD1			
Solvent Condition	**ΔG°_{eq}**	**m_{eq}**	**C_m**
0% Cosolvent	-2.04 ± 0.05	5.92 ± 0.12	0.34 ± 0.01
10% Glycerol	-2.96 ± 0.20	5.28 ± 0.31	0.56 ± 0.05
16% Glycerol	-3.57 ± 0.11	4.65 ± 0.14	0.77 ± 0.03
24% Glycerol	-4.23 ± 0.18	4.42 ± 0.18	0.96 ± 0.06
28% Glycerol	-4.85 ± 0.25	4.24 ± 0.21	1.14 ± 0.08
32% Glycerol	-5.56 ± 0.11	4.40 ± 0.08	1.26 ± 0.03
11% Glucose	-3.11 ± 0.10	5.02 ± 0.15	0.62 ± 0.03
17% Glucose	-3.42 ± 0.10	4.30 ± 0.12	0.80 ± 0.03
20% Glucose	-4.04 ± 0.28	4.61 ± 0.29	0.87 ± 0.08
24% Glucose	-4.56 ± 0.21	4.21 ± 0.21	1.08 ± 0.07
26% Glucose	-5.15 ± 0.16	3.59 ± 0.11	1.43 ± 0.06

The thermodynamic parameters obtained from fitting the Gdn-HCl titrations of 2SH-mSOD1 are shown for each cosolvent concentration examined. The equilibrium denaturant dependence, m_{eq}, is shown in kcal mol^{-1} $[D]^{-1}$ where [D] is the molar concentration of denaturant. The units of the folding free energy, ΔG°_{Eq}, are in kcal mol^{-1}. The midpoint, C_m, of each titration is shown as the [Gdn-HCl] (M).

species with significant amounts of secondary structure forming within the deadtime of the experiment (Fig. 3.3). The chevron plots, the observed relaxation times as a function of final denaturant concentration, are shown in Figure 3.4. These data demonstrate that increasing concentrations of glycerol and glucose stabilize the native state of the protein by decreasing the refolding times and increasing the unfolding relaxation times. The net effect of these changes moves the maximum relaxation time, where $k_u = k_f$, to progressively higher concentrations of denaturant, concordant with the C_m values from the equilibrium studies.

The chevron data were well described by a simple two-state kinetic model linking the native and unfolded states. The predicted free energy differences in the absence of denaturant between the folded and unfolded states were calculated from $\Delta G^0 = -RT \ln (k_u/k_f)$, where the rate constants in the absence of denaturant were obtained by linear extrapolations of the unfolding and refolding legs of the chevron plot to determine τ_f/τ_u, the reciprocal of k_f/k_u. The m_{eq}-value can also be determined by summing the absolute values of the $m_u^{\ddagger}$- and $m_f^{\ddagger}$-values, the slopes of the two legs of the chevron plot[137]. The resulting values are listed in Tables 3.3 and 3.4. The agreement between the kinetic and equilibrium estimates of stability, the m_{eq}-values, and the C_m values for both SS-mSOD1 and 2SH-mSOD1 over the range of kosmotrope concentrations examined (Fig. 3.5), and the absence of burst-phases (Fig. 3.3) rule out the presence of partially-folded states as an explanation for the reduction of m_{eq}-values at increasing kosmotrope concentrations.

It is noteworthy that the denaturant dependence of the unfolding and refolding relaxation times (Tables 3.3, 3.4), the $m_u^{\ddagger}$- and $m_f^{\ddagger}$- values, for SS-mSOD1 and 2SH-mSOD1, vary with glycerol and glucose concentration in a fashion that leaves their

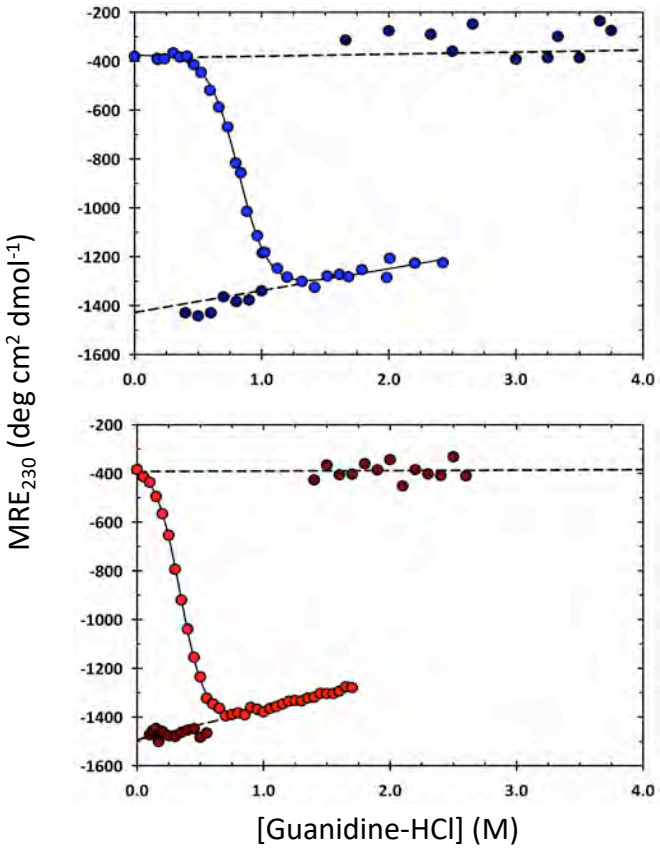

Figure 3.3: Burst phase analysis for the unfolding and refolding of SS- and 2SH-mSOD1

A comparison of the initial ellipticities in kinetic experiments (dark circles) are shown with the respective equilibrium titrations (light circles) for SS-mSOD1 (top, blue circles) and 2SH-mSOD1 (bottom, red circles). The coincidences of the initial ellipticities for unfolding jumps to high denaturant concentration with the extrapolated native baselines and the refolding jumps with the extrapolated unfolded baselines shows no loss in secondary structure or collapse in the dead-time of the mixing experiments.

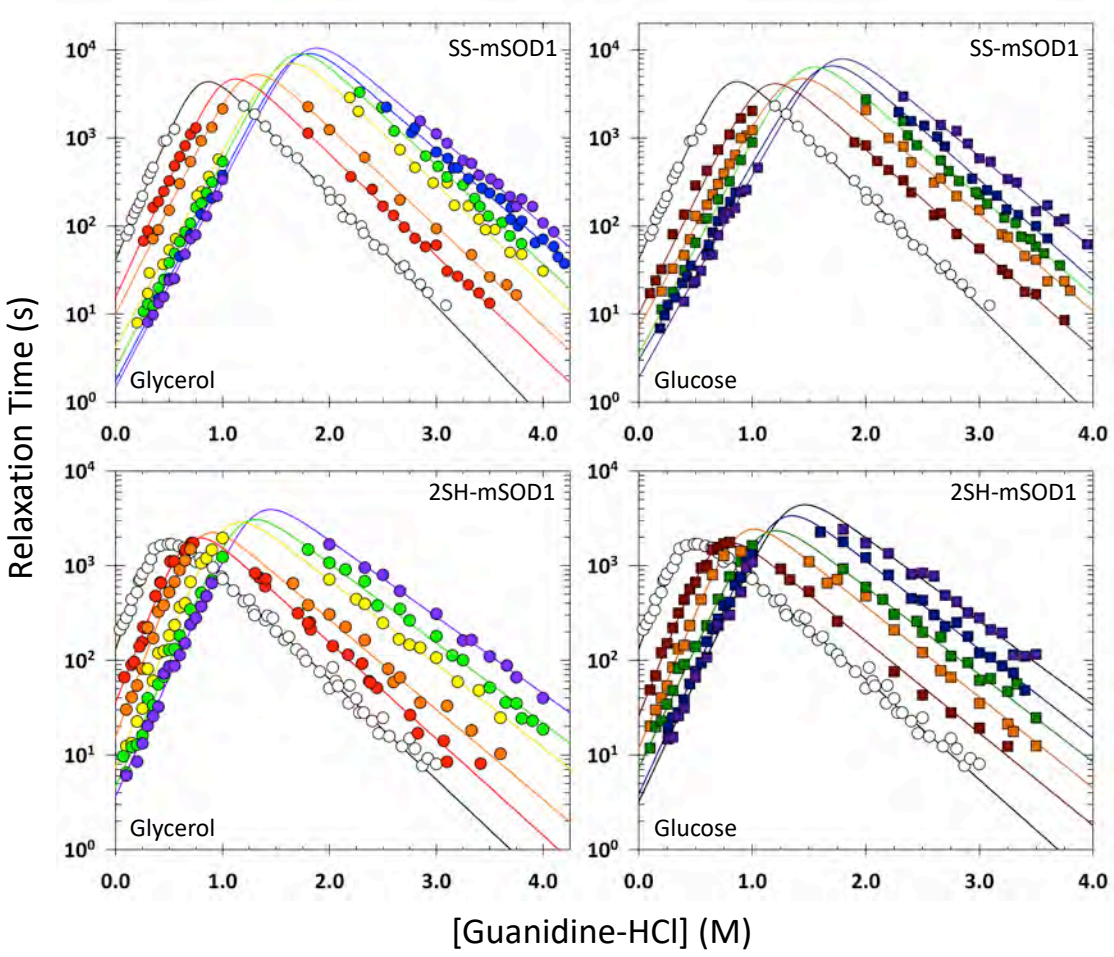

Figure 3.4: Polyol dependence of the folding kinetics
Chevron plots at increasing concentrations of glycerol (circles) and glucose (squares) of SS-mSOD1 and 2SH-mSOD1, obtained by CD spectroscopy at 230 nm, are shown with a two-state fit to the data (solid lines). Color schemes as in Figure 3.2.

Table 3.3: Kinetic Parameters for the folding reaction of SS-mSOD1

Solvent Condition	τ_f	$m_f^{\ddagger}$	τ_u	$m_u^{\ddagger}$	$\Delta G_{eq}^{\ddagger}$	$m_{eq}^{\ddagger}$	C_m	η	T_β
0% Cosolvent	41.4 ± 2.62	3.93 ± 0.13	(78.4 ± 5.48) × 10^3	-1.70 ± 0.02	-4.39 ± 0.06	5.63 ± 0.13	0.78 ± 0.02	1.01 ± 0.05	0.70 ± 0.03
10% Glycerol	15.6 ± 0.42	3.57 ± 0.11	(13.3 ± 1.60) × 10^4	-1.55 ± 0.03	-5.27 ± 0.09	5.12 ± 0.12	1.03 ± 0.03	1.30 ± 0.11	0.70 ± 0.03
16% Glycerol	9.99 ± 0.03	3.27 ± 0.11	(2.33 ± 0.42) × 10^5	-1.51 ± 0.04	-5.85 ± 0.12	4.78 ± 0.12	1.22 ± 0.04	1.66 ± 0.06	0.69 ± 0.03
24% Glycerol	4.20 ± 0.57	2.99 ± 0.10	(8.55 ± 1.55) × 10^5	-1.54 ± 0.04	-7.11 ± 0.13	4.54 ± 0.11	1.57 ± 0.05	2.34 ± 0.09	0.66 ± 0.03
28% Glycerol	2.44 ± 0.18	3.15 ± 0.05	(11.7 ± 2.75) × 10^5	-1.51 ± 0.04	-7.61 ± 0.14	4.66 ± 0.07	1.63 ± 0.04	2.65 ± 0.13	0.68 ± 0.01
30% Glycerol	1.75 ± 0.29	3.13 ± 0.11	(10.7 ± 1.65) × 10^5	-1.41 ± 0.03	-7.76 ± 0.13	4.54 ± 0.11	1.71 ± 0.05	2.87 ± 0.06	0.69 ± 0.03
32% Glycerol	1.48 ± 0.17	3.11 ± 0.08	(12.5 ± 2.50) × 10^5	-1.37 ± 0.04	-7.94 ± 0.13	4.48 ± 0.09	1.77 ± 0.05	3.19 ± 0.07	0.70 ± 0.02
11% Glucose	9.99 ± 0.96	3.50 ± 0.10	(13.5 ± 1.74) × 10^4	-1.51 ± 0.03	-5.53 ± 0.10	5.02 ± 0.11	1.10 ± 0.03	1.34 ± 0.06	0.70 ± 0.03
17% Glucose	7.08 ± 0.69	3.09 ± 0.07	(26.2 ± 5.10) ×10^4	-1.47 ± 0.04	-6.12 ± 0.13	4.56 ± 0.08	1.34 ± 0.04	1.74 ± 0.08	0.68 ± 0.02
20% Glucose	3.49 ± 0.42	3.26 ± 0.10	(51.8 ± 6.70) × 10^4	-1.51 ± 0.03	-6.93 ± 0.10	4.77 ± 0.10	1.45 ± 0.04	1.90 ± 0.05	0.68 ± 0.03
24% Glucose	2.99 ± 0.53	2.99 ± 0.15	(8.79 ± 1.71) × 10^5	-1.53 ± 0.04	-7.33 ± 0.15	4.52 ± 0.15	1.62 ± 0.06	2.26 ± 0.08	0.66 ± 0.04
26% Glucose	1.88 ± 0.34	3.08 ± 0.13	(8.92 ± 1.84) ×10^5	-1.41 ± 0.04	-7.61 ± 0.16	4.49 ± 0.14	1.69 ± 0.06	2.50 ± 0.19	0.69 ± 0.04

The parameters obtained from fitting the chevron plots of SS-mSOD1 are shown for each cosolvent concentration examined. The extrapolated refolding, τ_f, and unfolding, τ_u, relaxation times and the errors of the fit to the chevron plots (Fig. 3.3) are shown in seconds. The fits and errors for the refolding, $m_f^{\ddagger}$, and unfolding, $m_u^{\ddagger}$, portions of the chevron plots are shown in kcal mol^{-1} [D]$^{-1}$. The free energies of folding, $\Delta G^{\ddagger}_{eq}$, were calculated from the kinetic data and are shown in kcal mol^{-1}. The denaturant dependence of the free energy of folding, $m_{eq}^{\ddagger}$, calculated as, $m_{eq}^{\ddagger} = m_f^{\ddagger} - m_u^{\ddagger}$ is also shown in kcal mol^{-1} [D]$^{-1}$. The midpoint, C_m, of each titration is shown in [Gdn-HCl] (M). The viscosity of each buffer condition is shown in centipoises and was calculated using a Gilmont falling ball viscometer. The average and standard deviation of 10 trials is shown. The Tanford beta, T_β, for each viscogen concentration was calculated to be $T_\beta = m_f^{\ddagger}/(m_f^{\ddagger} - m_u^{\ddagger})$.

Table 3.4: Kinetic Parameters for the folding reaction of 2SH-mSOD1

Solvent Condition	τ_f	$m_f^{\ddagger}$	τ_u	$m_u^{\ddagger}$	$\Delta G_{eq}^{\ddagger}$	$m_{eq}^{\ddagger}$	C_m	η	T_β
0% Cosolvent	133 ± 7.49	4.59 ± 0.17	(8.48 ± 0.33) × 10^3	-1.43 ± 0.02	-2.42 ± 0.04	6.02 ± 0.17	0.40 ± 0.01	1.01 ± 0.05	0.70 ± 0.03
10% Glycerol	35.9 ± 3.18	3.94 ± 0.13	(1.69 ± 1.63) × 10^4	-1.37 ± 0.03	-3.58 ± 0.08	5.31 ± 0.13	0.67 ± 0.02	1.30 ± 0.11	0.70 ± 0.03
16% Glycerol	15.6 ± 1.79	4.13 ± 0.16	(2.12 ± 0.92) × 10^4	-1.27 ± 0.11	-4.20 ± 0.27	5.40 ± 0.19	0.78 ± 0.06	1.66 ± 0.06	0.69 ± 0.03
24% Glycerol	7.42 ± 0.81	3.66 ± 0.11	(4.13 ± 0.79) × 10^4	-1.19 ± 0.04	-5.02 ± 0.13	4.85 ± 0.12	1.04 ± 0.04	2.34 ± 0.09	0.66 ± 0.03
28% Glycerol	4.57 ± 0.75	3.52 ± 0.20	(5.21 ± 0.54) × 10^4	-1.13 ± 0.03	-5.44 ± 0.11	4.65 ± 0.20	1.17 ± 0.06	2.65 ± 0.13	0.68 ± 0.01
32% Glycerol	3.66 ± 0.34	3.37 ± 0.07	(7.75 ± 1.32) × 10^4	-1.09 ± 0.04	-5.80 ± 0.11	4.45 ± 0.08	1.30 ± 0.03	3.19 ± 0.07	0.70 ± 0.02
11% Glucose	25.6 ± 2.86	4.05 ± 0.15	(1.42 ± 0.24) × 10^4	-1.31 ± 0.05	-3.68 ± 0.12	5.36 ± 0.16	0.69 ± 0.03	1.34 ± 0.06	0.70 ± 0.03
17% Glucose	11.6 ± 1.37	3.86 ± 0.13	(3.05 ± 0.47) × 10^4	-1.28 ± 0.04	-4.58 ± 0.11	5.14 ± 0.14	0.89 ± 0.03	1.74 ± 0.08	0.68 ± 0.02
20% Glucose	7.24 ± 0.53	3.42 ± 0.06	(4.17 ± 0.38) × 10^4	-1.24 ± 0.02	-5.04 ± 0.07	4.66 ± 0.07	1.08 ± 0.02	1.90 ± 0.05	0.68 ± 0.03
24% Glucose	3.84 ± 0.28	3.50 ± 0.06	(8.33 ± 1.11) × 10^4	-1.25 ± 0.03	-5.81 ± 0.09	4.75 ± 0.06	1.22 ± 0.02	2.26 ± 0.08	0.66 ± 0.04
26% Glucose	3.19 ± 0.46	3.42 ± 0.10	(1.16 ± 0.31) × 10^5	-1.19 ± 0.06	-6.11 ± 0.18	4.60 ± 0.12	1.33 ± 0.05	2.50 ± 0.19	0.69 ± 0.04

The parameters obtained from fitting the chevron plots of 2SH-mSOD1 are shown for each cosolvent concentration examined. The extrapolated refolding, τ_f, and unfolding, τ_u, relaxation times and the errors of the fit to the chevron plots (Fig. 3.3) are shown in seconds. The fits and errors for the refolding, $m_f^{\ddagger}$, and unfolding, $m_u^{\ddagger}$, portions of the chevron plots are shown in kcal mol^{-1} [D]$^{-1}$. The free energies of folding, $\Delta G^{\ddagger}_{eq}$, were calculated from the kinetic data and are shown in kcal mol^{-1}. The denaturant dependence of the free energy of folding, $m_{eq}^{\ddagger}$, calculated as, $m_{eq}^{\ddagger} = m_f^{\ddagger} - m_u^{\ddagger}$ is also shown in kcal mol^{-1} [D]$^{-1}$. The midpoint, C_m, of each titration is shown in [Gdn-HCl] (M). The viscosity of each buffer condition is shown in centipoises and was calculated using a Gilmont falling ball viscometer. The average and standard deviation of 10 trials is shown. The Tanford beta, T_β, for each viscogen concentration was calculated to be $T_\beta = m_f^{\ddagger}/(m_f^{\ddagger} - m_u^{\ddagger})$.

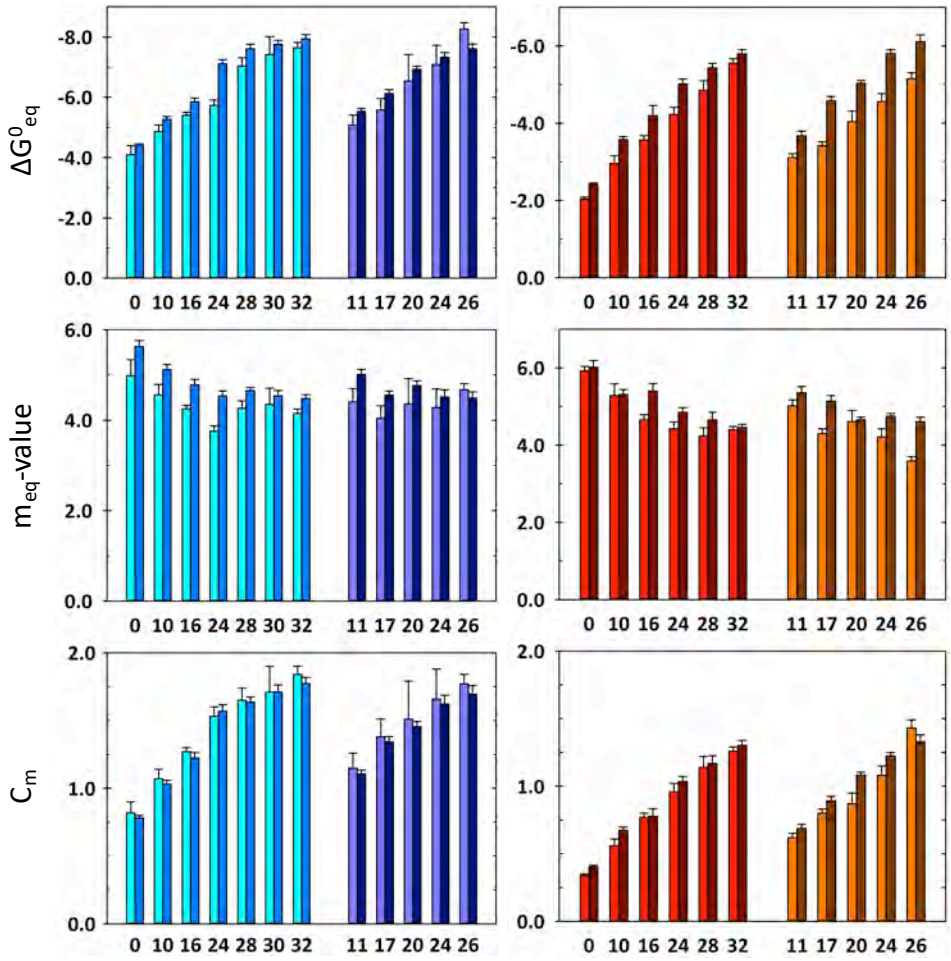

Figure 3.5: The thermodynamic parameters of SS-mSOD1 and 2SH-mSOD1 are comparable when measured by kinetics or equilibrium

The folding free energies, ($\Delta G°_{eq}$, top), the denaturant dependence of the free energy of folding, (m_{eq}-value, middle), and the midpoints (C_m, bottom) from the equilibrium (light colors) or kinetic (dark colors) experiments are in good agreement for both SS-mSOD1 (blue and purple, left) and 2SH-mSOD1 (red and orange, right). This agreement is observed regardless of whether the cosolvent used was glycerol (blue and red colors) or glucose (purple and orange colors). The $\Delta G°_{eq}$ are shown in kcal mol^{-1}. The m_{eq}-values are shown in kcal mol^{-1} $[D]^{-1}$. The C_m are shown in [Gdn-HCl] (M).

Tanford beta, T_β, values unchanged across the range of glycerol and glucose concentrations explored (Fig. 3.6). T_β reflects the fraction of the buried surface area in the transition state ensemble (TSE) relative to the buried surface area in the native state and is calculated as $T_\beta = m_f^\ddagger/(m_f^\ddagger - m_u^\ddagger)$[178]. The T_β values for SS-mSOD1 and 2SH-mSOD1 are ~0.68 ± 0.02 and ~0.75 ± 0.01 kcal mol^{-1} M^{-1}, respectively, and are independent of kosmotrope concentration. The robust positions of the TSEs in terms of buried surface area for both oxidized and reduced mSOD1 demonstrate that these additives have a common but proportional effect on the structures and the free energies of the TSE and the unfolded state. The reduction in the m_{eq}-value in the presence of kosmotropes (Tables 3.1-3.4) is, therefore, a consequence of the solvent composition and does not reflect the presence of an additional thermodynamic state. This conclusion is consistent with a previous study that observed the compaction of the unfolded state of SNase in response to viscogen[179]. The consistently higher m_{eq}-value (Tables 3.1-3.4) and the greater T_β value for 2SH-mSOD1 (Fig. 3.6) likely reflect the increased solvent exposure of side chains and backbone in an unfolded form that is not constrained by the disulfide crosslink.

Viscosity Dependence of Folding

Although the presence of glycerol or glucose did not reveal partially-folded states for SS-mSOD1 or 2SH-mSOD1, these additives offer another probe of the folding reactions. Glycerol and glucose also act to substantially increase the solvent viscosity, providing a test for intramolecular friction in limiting diffusive folding reactions. Intramolecular chain friction could be exchanged for intermolecular interactions that could nucleate aggregation in SOD1.

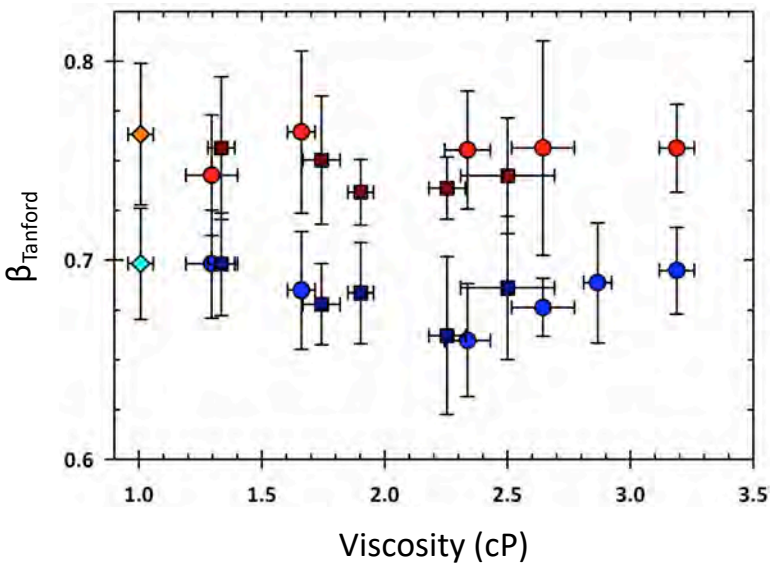

Figure 3.6: Tanford β is unaffected by a change in viscosity
The Tanford β value of each cosolvent is shown as a function of the resultant viscosity. The Tanford β values of SS- and 2SH-mSOD1 in buffer (blue and orange diamond, respectively) are shown. Neither SS-mSOD1 (blue) nor 2SH-mSOD1 (red) experiences a change in the Tanford β value when exposed to increasing concentrations of glycerol (circles) or glucose (squares).

The rate constants of diffusion-limited reactions are expected to vary linearly with solvent viscosity, accordingly to the Kramer's formalism[180–186]:

$$k = \frac{1}{\tau} = \frac{C}{\eta + \sigma} \exp\left(\frac{-\Delta G^{0\ddagger}}{RT}\right) \qquad Eq. 3.1$$

where k is the rate constant, τ is the corresponding relaxation time, C is an adjustable parameter, η is the viscosity of the solvent, σ is the internal friction, $\Delta G^{0\ddagger}$ is the intrinsic barrier height under standard conditions, R is the gas constant and T is the absolute temperature. Thus, in the absence of significant internal friction, the folding rate constant should vary inversely with the solvent viscosity, $k \propto 1/\eta$[187,188]. A less than unitary dependence of the observed rate constant on the solvent viscosity, i.e., a nonzero value for σ, has been taken as evidence for chain friction in limiting the folding of three helix bundle spectrin domains[189–192], a photolyzed cytochrome c[193], an ultrafast folding protein[194], and unfolded and intrinsically disordered proteins[195–197]. Because viscogens significantly alter the thermodynamic properties, an isostability approach, in which the refolding and unfolding rate constants were interpolated under strongly refolding and unfolding conditions, is typically employed[176,180,198]. This approach, however, does not account for the observed dependence of the $m_f^{\ddagger}$ and $m_u^{\ddagger}$-values for SOD1 on the viscogen concentration (Tables 3.1-3.4).

To address this issue, we developed a method of globally fitting the chevron data by introducing parameters that describe the viscogen dependence of the thermodynamic and kinetic properties of the folding reaction. As noted above, viscogens alter both the stability and the m_{eq}-value of SS- and 2SH-mSOD1. When plotted as a function of viscogen concentration, we observed that ΔG^0, $m_{eq}^{\ddagger}$, $m_f^{\ddagger}$ and $m_u^{\ddagger}$ all vary linearly as a

function of viscogen concentration (Figs. 3.7 and 3.8). Interestingly, the viscogen dependence of each parameter was unique to each of the examined viscogens and the linear fits of the data show that glucose has a 2-3 fold stronger effect on the stability and *m*-values than glycerol. Adding terms to Eq. 3.1 to describe the effect of viscogen on the *m*-values and ΔG^0 results in the following equation:

$$k_{obs} = \frac{C_f}{\eta + \sigma_f} \exp\left(\frac{\left(m_f^{\ddagger} + \alpha_f^{\ddagger}[Visc]\right)[Den] - m_{vf}^{\ddagger}[Visc]}{RT}\right) + \frac{C_u}{\eta + \sigma_u} \exp\left(\frac{\left(m_u^{\ddagger} + \alpha_u^{\ddagger}[Visc]\right)[Den] + m_{vu}^{\ddagger}[Visc]}{RT}\right) \qquad Eq.\,3.2$$

where $[Visc]$ and $[Den]$ are molar concentrations of viscogen and denaturant, k_{obs} is the observed rate constant at that viscogen and denaturant concentration, $C_{f/u}$ are adjustable parameters, $m_f^{\ddagger}$ and $m_u^{\ddagger}$ represent the denaturant dependence of the rate constants, $\alpha_f^{\ddagger}$ and $\alpha_u^{\ddagger}$ are the viscogen dependence of $m_f^{\ddagger}$ and $m_u^{\ddagger}$, $m_{vf}^{\ddagger}$ and $m_{vu}^{\ddagger}$ represent the viscogen dependence of the transition state free energy, η is the solvent viscosity, and σ_f and σ_u are the chain friction terms for folding and unfolding. Note that the viscogen dependence terms $\alpha_f^{\ddagger}$, $\alpha_u^{\ddagger}$, $m_{vf}^{\ddagger}$ and $m_{vu}^{\ddagger}$ are unique for the specific viscogens, while the other terms are universal parameters. A key assumption in this approach is the independent effects of denaturant and viscogen on the experimentally measured viscosity. A previous study supports this assumption in the range of Gdn-HCl and glucose concentrations used in our study[199].

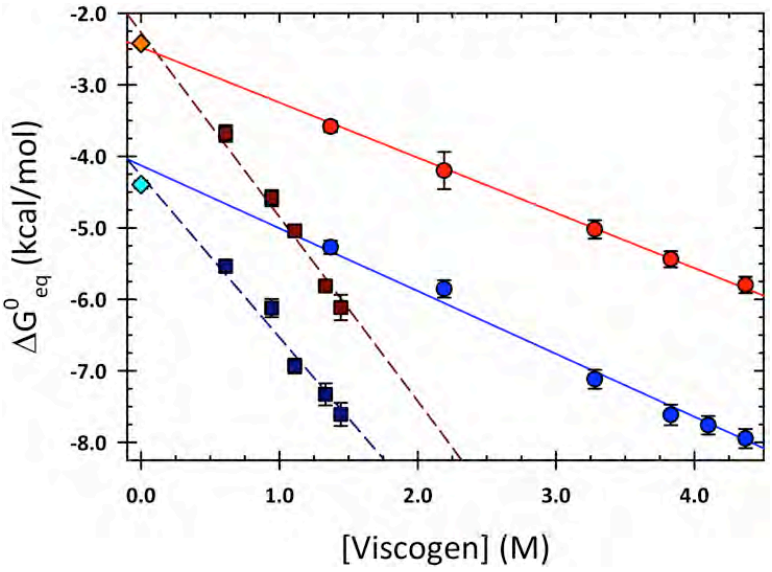

Figure 3.7: Viscogen dependence of ΔG^0_{eq} for SS-mSOD1 and 2SH-mSOD1
The ΔG^0_{eq} of SS-mSOD1 (blue) and 2SH-mSOD1 (red) as a function of either glycerol (M) (light circles) or glucose (M) (dark squares) is shown along with linear fits of the data (solid and dashed for glycerol and glucose, respectively). The ΔG^0_{eq} in the absence of viscogen is shown as a diamond for both SS-mSOD1 and 2SH-mSOD1.

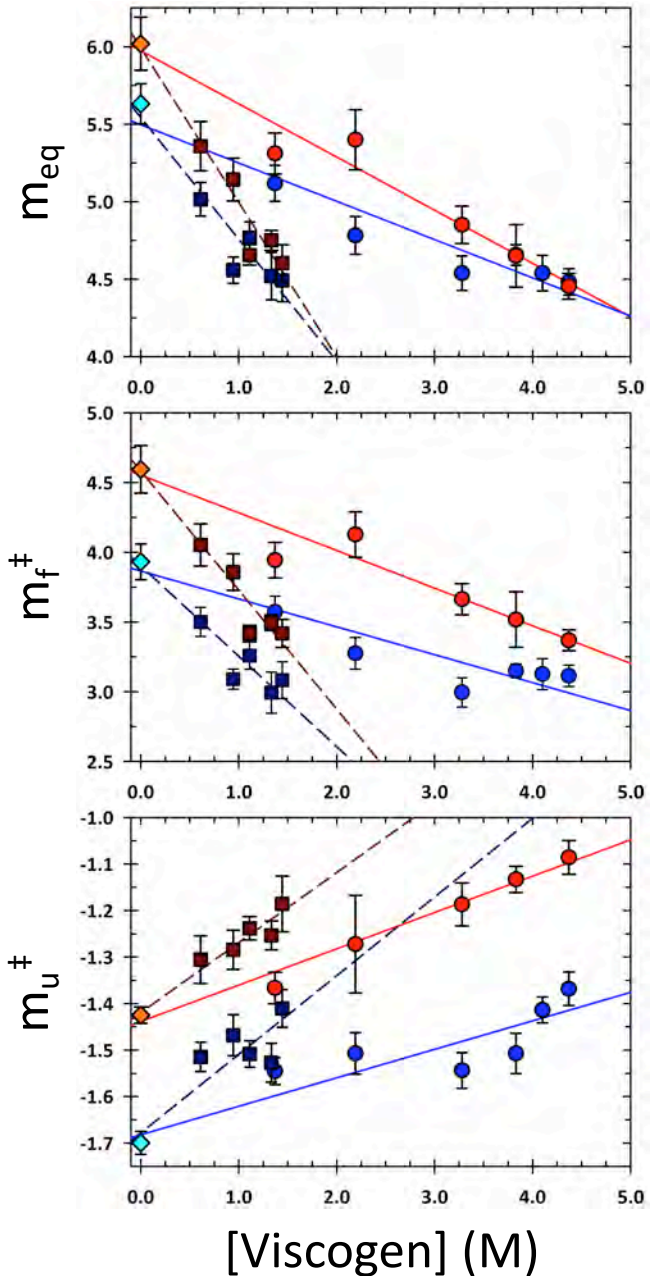

Figure 3.8: Viscogen dependence of m-values for SS-mSOD1 and 2SH-mSOD1
The m_{eq} (top), $m_f^{\ddagger}$ (middle), and $m_u^{\ddagger}$ (bottom) are shown for SS-mSOD1 (blue) and 2SH-mSOD1 (red) as a function of either glycerol (M) (light circles) or glucose (M) (dark squares) along with linear fits to the data (solid and dashed for glycerol and glucose, respectively). The units of m_{eq}, $m_f^{\ddagger}$, and $m_u^{\ddagger}$ are in kcal mol^{-1} [D]$^{-1}$. The m-values in the absence of viscogen are shown for SS-mSOD1 and 2SH-mSOD1 as blue and orange diamonds, respectively. m_{eq} was calculated as $m_f^{\ddagger}$-$m_u^{\ddagger}$.

Global fits of the chevron data for both SS-mSOD1 and 2SH-mSOD1 to Eq. 3.2 provided an excellent description of the results (Fig. 3.9). Supporting the validity of the analysis is the excellent agreement between (1) the results of the individual linear regression values for terms $\alpha_f^{\ddagger}$, $\alpha_u^{\ddagger}$, $m_{vf}^{\ddagger}$, and $m_{vu}^{\ddagger}$ and those obtained from the global fit (Table 3.5) and (2) the estimates of the friction coefficients for the unfolding and refolding reactions in the presence of glucose and glycerol (Table 3.6). A global fit of the combined glucose and glycerol data sets, subjected to a rigorous error analysis (Fig. 3.10) finds the folding and unfolding friction coefficients for SS-mSOD1 are $\sigma_f = 0.07 \pm 0.09$ and $\sigma_u = 0.00 \pm 0.01$. SS-mSOD1 experiences neither friction-limited folding nor unfolding. By distinct contrast, the friction coefficients for 2SH-mSOD1 are $\sigma_f = 1.02 \pm 0.24$ and $\sigma_u = 0.87 \pm 0.13$. We conclude that 2SH-mSOD1 experiences internal chain friction during both folding and unfolding reactions, slowing access to the productive TSE from both the native and the unfolded states of 2SH-mSOD1.

DISCUSSION

The introduction of glucose and glycerol to the solvent for mSOD1 enabled two distinct tests for partially-folded states that might serve to nucleate the aggregation reaction known to be a hallmark of SOD1 variants that cause ALS. Although a substantial perturbation of the folding free energy surface by a pair of kosmotropes did not reveal the presence of a distinct, high-energy thermodynamic state for either SS-mSOD1 or 2SH-mSOD1, the accompanying increase in solvent viscosity revealed friction-limited folding and unfolding for the disulfide-reduced protein. The contrasting behavior of SS-mSOD1 pinpoints unique roles for one or both of the segments adjacent

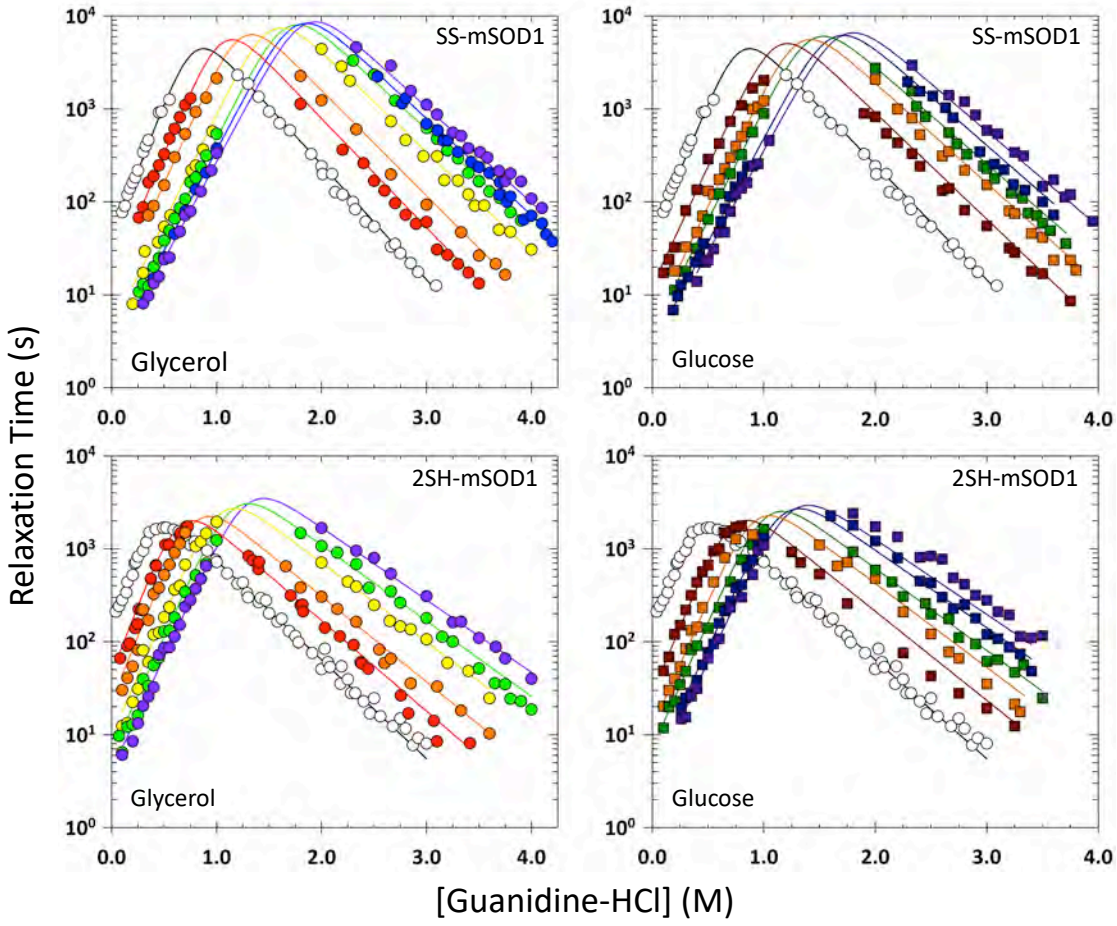

Figure 3.9: Global fit of folding kinetics
Global fits using Eq 3.2 are shown for SS-mSOD1 (top) and 2SH-mSOD1 (bottom) in Glycerol (left) and glucose (right). Colors as in Figure 3.2.

Table 3.5: Parameters extracted from the linear regression of the viscogen dependence are in excellent agreement with those obtained from the global fit

	Glycerol					Glucose				
SS-mSOD1	$\alpha_f^{\ddagger}$	$\alpha_u^{\ddagger}$	$m_{vf}^{\ddagger}$	$m_{vu}^{\ddagger}$	$m_{veq}^{\ddagger}$	$\alpha_f^{\ddagger}$	$\alpha_u^{\ddagger}$	$m_{vf}^{\ddagger}$	$m_{vu}^{\ddagger}$	$m_{veq}^{\ddagger}$
Linear Fit	-0.20 ± 0.02	0.06 ± 0.01	N/A	N/A	-0.88 ± 0.05	-0.65 ± 0.08	0.17 ± 0.04	N/A	N/A	-2.26 ± 0.16
Global Fit	-0.25 ± 0.02	0.07 ± 0.01	-0.54 ± 0.01	0.23 ± 0.02	-0.77 ± 0.02	-0.72 ± 0.05	0.17 ± 0.02	-1.41 ± 0.04	0.65 ± 0.05	-2.06 ± 0.06
2SH-mSOD1	$\alpha_f^{\ddagger}$	$\alpha_u^{\ddagger}$	$m_{vf}^{\ddagger}$	$m_{vu}^{\ddagger}$	$m_{veq}^{\ddagger}$	$\alpha_f^{\ddagger}$	$\alpha_u^{\ddagger}$	$m_{vf}^{\ddagger}$	$m_{vu}^{\ddagger}$	$m_{veq}^{\ddagger}$
Linear Fit	-0.27 ± 0.03	0.08 ± 0.00	N/A	N/A	-0.77 ± 0.01	-0.85 ± 0.08	0.15 ± 0.01	N/A	N/A	-2.58 ± 0.13
Global Fit	-0.20 ± 0.01	0.09 ± 0.00	-0.58 ± 0.01	0.16 ± 0.01	-0.74 ± 0.01	-0.54 ± 0.04	0.21 ± 0.02	-1.70 ± 0.04	0.58 ± 0.05	-2.28 ± 0.06

The parameters from fitting the viscogen dependence of the kinetic parameters to linear regressions are compared to the parameters obtained with the global fitting method. The units of $\alpha_{f/u}^{\ddagger}$ are shown in kcal mol^{-1} [D]$^{-1}$ [Visc]$^{-1}$ where [D] and [Visc] are the molar concentrations of denaturant and viscogen. The units of the $m_{vf}^{\ddagger}$-, $m_{vu}^{\ddagger}$-, and $m_{veq}^{\ddagger}$-values are kcal mol^{-1} [Visc]$^{-1}$. The $m_{veq}^{\ddagger}$-value for the global fit parameters was calculated as, $m_{veq}^{\ddagger} = m_{vf}^{\ddagger} - m_{vu}^{\ddagger}$.

Table 3.6: Comparison of $\sigma_{f/u}$ from glycerol-only, glucose-only, and global fits						
	glycerol		**glucose**		**global**	
	σ_f	**σ_u**	**σ_f**	**σ_u**	**σ_f**	**σ_u**
SS-mSOD1	0.06 ± 0.04	0.00 ± 0.03	0.06 ± 0.04	0.00 ± 0.03	0.07 ± 0.09	0.00 ± 0.03
2SH-mSOD1	0.97 ± 0.07	0.88 ± 0.04	0.86 ± 0.05	1.09 ± 0.09	1.02 ± 0.24	0.87 ± 0.13

The value of σ obtained using Eq 3.2 to fit the glycerol-only, glucose-only, and global data sets. σ is in units of cP.

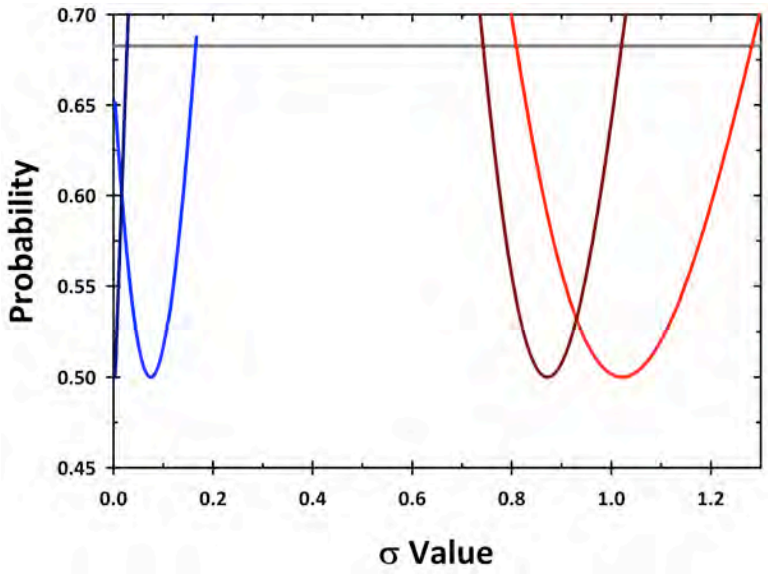

Figure 3.10: Rigorous error analysis of σ_f and σ_u for SS-mSOD1 and 2SH-mSOD1
The rigorous error analysis of σ_f (light lines) and σ_u (dark lines) is shown for SS-mSOD1 (blue) and 2SH-mSOD1 (red). The y-axis shows the probability of obtaining a better fit to the data as a function of the σ-values along the x-axis. The grey line denotes a p=0.68, which corresponds to a single standard deviation in the σ-value.

to C57 and C146, i.e., loop IV and β8, in introducing the chain friction that slows the refolding and unfolding reactions.

Molecular mechanisms of internal chain friction in proteins

Despite the observation of friction-limited folding in several proteins, the molecular mechanisms responsible for friction remain a subject of active discussion. Experimental observation of friction in spectrin domains has been attributed to nonnative docking of helices[189]. MD simulations on short peptides suggested that internal friction was due to isomerization of local torsional angels[200] and was closely linked to the α-helical propensity of the protein[201,202]. This interpretation is consistent with the observation of internal friction in several fast folding proteins[193,194,203,204]. However, this explanation may not account for the observation of internal friction in unfolded proteins[195,197] or the experimental observation of helical proteins lacking internal friction[177,186,198]. This hypothesis would not obviously apply to SOD1, a β-sandwich protein.

A clue to the source of friction in 2SH-mSOD1 is its absence in its oxidized counterpart, SS-mSOD1. The implication is that the disulfide bond effectively blocks the intramolecular interactions that cause friction. Examination of the sequences adjacent to the C57 and C146 partners in the disulfide bond reveals that C57 is embedded in the polar-rich sequence of Loop IV that is responsible for binding the Zn^{+2} ion. In the absence of a Zn^{+2} ion, Loop IV adopts a disordered conformation[205], which may serve as a source of friction as others have observed[195,197]. However, a more plausible source of friction could be the nonpolar sequence of β8 adjacent to C146. Perhaps this sequence,

when not constrained by the disulfide bond, forms fleeting interactions with one or more of the 7 remaining β-strands, giving rise to chain friction.

A recent peptide dissection analysis of SOD1 from our lab provides a surprising alternative[206]. The 35 residue C-terminal peptide, containing the polar sequence of Loop VII and the nonpolar sequence of β8, folds on itself to bring the polar and charged N-terminus and the nonpolar C-terminus into a nonnative juxtaposition. Deletion of the N-terminal 5 amino acids or the replacement of isoleucine 149 with threonine disrupts the interaction, and the peptide expands. The conjecture that β8 is primarily responsible for chain friction is consistent with the observation that both folding and unfolded reactions of 2SH-mSOD1 are slowed by friction. Previous work from the Oliveberg laboratory mapped the structure of the TSE onto β2, β3, β4, and β7, the internal strands on both faces of the β-sandwich (Fig. 3.1)[207]. Thus, Loop VII and β8 would be released prior to the TSE in unfolding and not engage after the formation of the TSE in folding, enabling the C-terminus of SOD1 to collapse on itself and introduce chain friction for both reactions.

The link between internal friction, aggregation and ALS in SOD1

Intramolecular interactions that give rise to friction have the potential to become intermolecular interactions that could drive aggregation under appropriate conditions. There are two lines of evidence supporting a primary role for the C-terminus of SOD1 in aggregation and toxicity in ALS. Work by Ivanova *et al.* showed that only a hexapeptide containing residues 147-153 from β8 accelerated fibril formation of both wild-type metal-free SOD1 and ALS-linked mutants of SOD1[138]. The relevance of a unique role for β8 in

the aggregation of SOD1 is also evident in the observations of Furukawa *et al.*[139], who found that the segment containing β8 and portions of the preceding Loop VII was the only region resistant to proteolysis in aggregates of ALS variants of SOD1 isolated from mouse motor neurons.

It should be noted that other segments of SOD1, e.g., β1-β3 and β5-β6, are also protected against proteolysis when the aggregates are artificially created *in vitro*[139]. These regions are also the only regions protected in the I149T variant and in mutants that truncate the C-terminus, such as L126Z. The observation that both of these variants cause ALS suggests that, in the absence of Loop VII-β8, other β-strands may serve as secondary sites for nucleating aggregation.

SOD1 and ALS

The nascent SOD1 chain prior to disulfide bond formation, dimerization, and full metalation has several properties that distinguish it from many proteins and would increase its propensity to aggregate. First, the reduced wild-type monomer is only marginally stable, resulting in ~1% of the population occupying the unfolded state at equilibrium[91]. ALS mutations dramatically destabilize the native state, resulting in a preponderance of the unfolded state at neutral pH and physiological temperatures[88]. The dramatic shift to the unfolded state of SOD1 is compounded by its exceedingly slow folding reaction, $\tau > 10$ s in the absence of denaturant[93]. The long lifetime and preponderance of the unfolded nascent chain in SOD1 ALS-variants would enhance the probability for the same sequences that give rise to intramolecular friction to form intermolecular interactions and nucleate aggregation[138,139]. The nascent chain, with its

nonnative collapsed structure at the C-terminus, would be a nearly universal target for therapeutics designed to ameliorate the deterioration of motor neurons in ALS caused by mutant forms of SOD1.

Chapter IV –Disulfide bonds protect against aggregation, but induce complex folding mechanisms

This chapter is adapted from a manuscript currently *in prep*:

Strained disulfide bonds in V_H domains create complex folding free energy surfaces
Cohen NR, Kao YH, Matthews CR.

INTRODUCTION

The broad family of antibody molecules are the crucial component of the adaptive immune response, and engineered variants represent a growing portion of the therapeutic market. Currently there are 80 FDA-approved antibody biologics for a host of human diseases (drugs@FDA). Of these 35 are natural murine or human antibodies, 10 are chimeric antibodies containing murine variable domains attached to human constant domains, and 35 are "humanized", meaning they contain recognition sequences from murine antibodies appended to human scaffolds[111]. The latter group of hybrid molecules presents an interesting class of proteins whose sequences were not evolved through the natural processes of antibody maturation. Removing these variable domains from the optimization of adaptive immunity may result in proteins with unanticipated and possibly deleterious biophysical properties. Our lab recently found a complex folding free energy surface in a *de novo* designed βα protein with a natural topology, highlighting the challenges that can arise in protein design and engineering[208].

Human antibodies are formed through the covalent association of four polypeptide chains, two Heavy Chains (HC) an two Light Chains (LC). Each LC pairs with a HC partner, forming a disulfide-crosslinked heterodimer, that then forms disulfide crosslinks with another heterodimer to form the full antibody. Each chain is comprised of a series of tandem domains, 4-5 in the HC and 2 in the LC. The N-terminal variable domain, V_L in the LC and V_H in the HC, is responsible for antigen recognition and binding. Following in sequence are the constant domains, C_L in the LC and $C_{H1\text{-}3/4}$ in the HC, that are responsible for the formation of the full antibody and also stimulate the immune response upon antigen binding. The constant domains are highly conserved

between different antibodies on the same class, such as the IgMs or IgGs, while the variable domains are unique to each antibody.

Each of the variable and constant domains contain the canonical immunoglobulin fold, Ig-fold, defined by 7-9 antiparallel β-strands that form a pair of opposing β-sheets linked by a highly conserved intramolecular disulfide bond (Figure 1)[97,98]. The V_H domains fold autonomously, while the V_L constant domains often require chaperones or partners to fold[209–213].

A recent study of a humanized IgG1 monoclonal antibody (mAb) found a significant population lacked an intramolecular disulfide (SS) bond in the V_H domain [115,214]. Similar behavior has been observed in another IgG1 antibody[215], an anti-IgE molecule[116] and several other antibodies[216,217] suggesting this may be a common phenomenon in engineered antibodies. In the anti-CD20 molecule, the absence of the SS bond did not have an impact on binding, however, the anti-IgE molecule experienced a significant reduction in binding efficacy. One could imagine that the absence of the SS bond would also alter the stability of the V_H domain and potentially lead to unfolding and aggregation.

To understand the heterogeneity in the SS bond content in these mAbs, we chose to examine the isolated V_H domains of the two antibodies that do not completely form the SS bond as well as one that does. Approximately 30% of the V_H domains from the IgG1 mAb A[115] (V_{HA}) and 50% of the V_H domains from the anti-IgE mAb B (V_{HB})[116] do not form intramolecular SS bonds while the V_H domains of mAb C (V_{HC}) contain a fully formed SS bond[116]. The synthesis of a V_H domain prior to the remaining HC sequence and its autonomous folding reaction rationalized this strategy. We propose that the failure

Figure 4.1: Structure of a V_H domain
A ribbon diagram of a V_H domain, adapted from PDB 1N8Z[99], is shown. The complementarity determining regions 1-3 are shown in red, blue, and purple. The disulfide bond is shown in yellow.

to completely form the SS bond in the V_{HA} and V_{HB} domains is best understood as a consequence of the unanticipated strain required to form the SS bond. The introduction of the SS bond in these domains also results in a complex folding reaction involving residual structure in the unfolded state, a folding intermediate and parallel folding channels.

RESULTS

The failure to completely form the intramolecular disulfide in the two V_H domains could reflect the thermodynamic and/or kinetic properties of the disulfide reduced (2SH-) or disulfide oxidized (SS-) forms of these domains. To gain a comprehensive understanding of the interplay between folding and SS bond formation we performed classical kinetic and thermodynamic analyses of both the 2SH- and SS-forms of these V_H domains.

Analysis of the folding and stability of the 2SH-V_H domains

To determine if the biophysical properties of the 2SH-V_H domains explain the heterogeneity of the SS bond formation observed in mAb A and mAb B, we examined the thermodynamic and kinetic folding properties of V_{HA} and V_{HB} and those of V_{HC} that completely forms the SS bond. The 2SH forms of the 3 domains were prone to aggregation at the micromolar concentrations required for circular dichroism (CD) analysis, requiring fluorescence spectroscopy (FL) to monitor the folding reactions. Although the reduced proteins were amenable to kinetic refolding studies at nanomolar concentrations, prolonged exposure to refolding conditions for equilibrium studies also

led to aggregation. Thus, we were only able to obtain a partial chevron plot, i.e. log relaxation time vs. final denaturant concentration, by refolding from highly denaturing urea solutions and monitoring the reaction before the onset of aggregation. The chevron plots for 2SH-V_{HA}, 2SH-V_{HB} and 2SH-V_{HC} are shown in Figure 4.2A-C.

The refolding reactions of 2SH-V_{HA}, 2SH-V_{HB} and 2SH-V_{HC} were well described by single exponential responses and the denaturant dependences yielded simple chevrons consistent with a two-state mechanism, N↔U. Fits of these chevrons to a two-state model provided the refolding and unfolding relaxation times in the absence of denaturant as well as the denaturant dependence of the relaxation times, the *m*-values (Table 4.1). We found 2SH-V_{HA} and 2SH-V_{HB} fold within 100's of milliseconds, while 2SH-V_{HC} folds far slower, on the order of 10's of seconds. The maxima in chevron plots corresponds closely with the midpoint of the equilibrium unfolding reactions, C_m, an indicator of the stability of the protein to chemical denaturation. Comparison of the maxima of the partial chevrons shows the rank order of stability to be 2SH-V_{HA} > 2SH-V_{HB} > 2SH-V_{HC}.

Analysis of the folding and stability of the SS-V_H domains

Thermodynamic Analysis

To gain a comprehensive understanding of the differential behavior of these engineered proteins, we also examined the thermodynamic and kinetic properties of the three SS-V_H domains. We collected equilibrium denaturation curves for SS-V_{HA}, SS-V_{HB}, and SS-V_{HC} using FL lifetime detection (Fig. 4.3A-C). The equilibrium curves of both

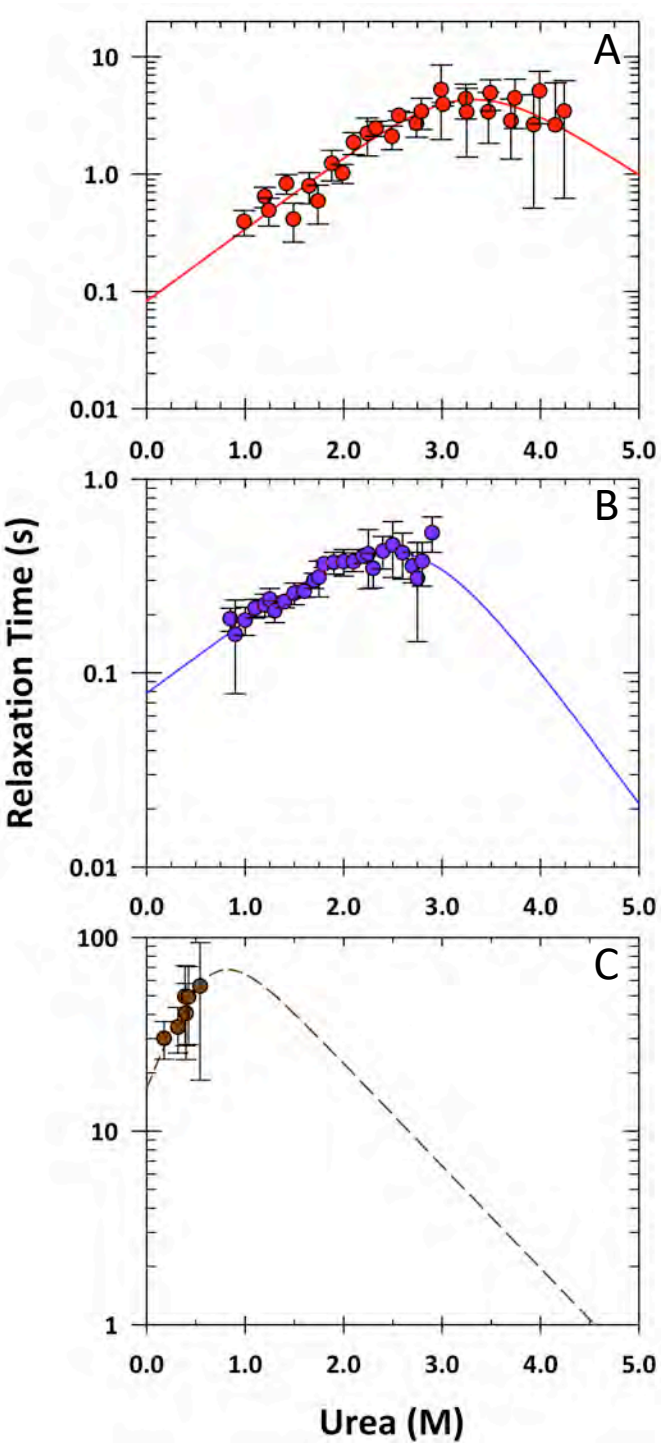

Figure 4.2: Chevron plots of the 2SH-V_H domains
Chevron plots of 2SH-V_{HA} (A, red) and 2SH-V_{HB} (B, purple), obtained with stopped-flow FL, are shown with two-state fit to the data (solid line). Preliminary chevron of 2SH-V_{HC} (C, brown), obtained with manual-mixing FL, is shown with an approximate fit to the data (dashed line).

Table 4.1: Kinetic Parameters for the folding reaction of the 2SH-V_H domains							
	τ_f	$m_f^{\ddagger}$	τ_u	$m_u^{\ddagger}$	$\Delta G_{eq}^{\ddagger}$	$m_{eq}^{\ddagger}$	C_m
V_{HA}	0.082 ± 0.024	0.84 ± 0.10	530 ± 1000	-0.73 ± 0.40	5.11 ± 1.49	1.57 ± 0.41	3.3 ± 1.3
V_{HB}	0.079 ± 0.019	0.50 ± 0.14	60 ± 170	-0.92 ± 0.81	3.85 ± 2.40	1.42 ± 0.82	2.7 ± 2.3
V_{HC}[1]	17.8 ± 2.4	1.87 ± 0.36	250 ± 110	-0.71 ± 0.22	1.55 ± 0.28	2.58 ± 0.43	0.6 ± 0.2

The parameters obtained from fitting the chevron plots of 2SH-V_H domains are shown. The extrapolated refolding, τ_f, and unfolding, τ_u, relaxation times and the errors of the fit to the chevron plots (Fig. 4.2) are shown in seconds. The fits and errors for the refolding, $m_f^{\ddagger}$, and unfolding, $m_u^{\ddagger}$, portions of the chevron plots are shown in kcal mol^{-1} [D]$^{-1}$. The free energies of folding, $\Delta G^{\ddagger}_{eq}$, were calculated from the kinetic data and are shown in kcal mol^{-1}. The denaturant dependence of the free energy of folding, $m_{eq}^{\ddagger}$, calculated as, $m_{eq}^{\ddagger} = m_f^{\ddagger} - m_u^{\ddagger}$ is also shown in kcal mol^{-1} [D]$^{-1}$. The midpoint, C_m, of each titration is shown in [Urea] (M).

[1]The V_{HC} parameters are from an approximate fit to preliminary data

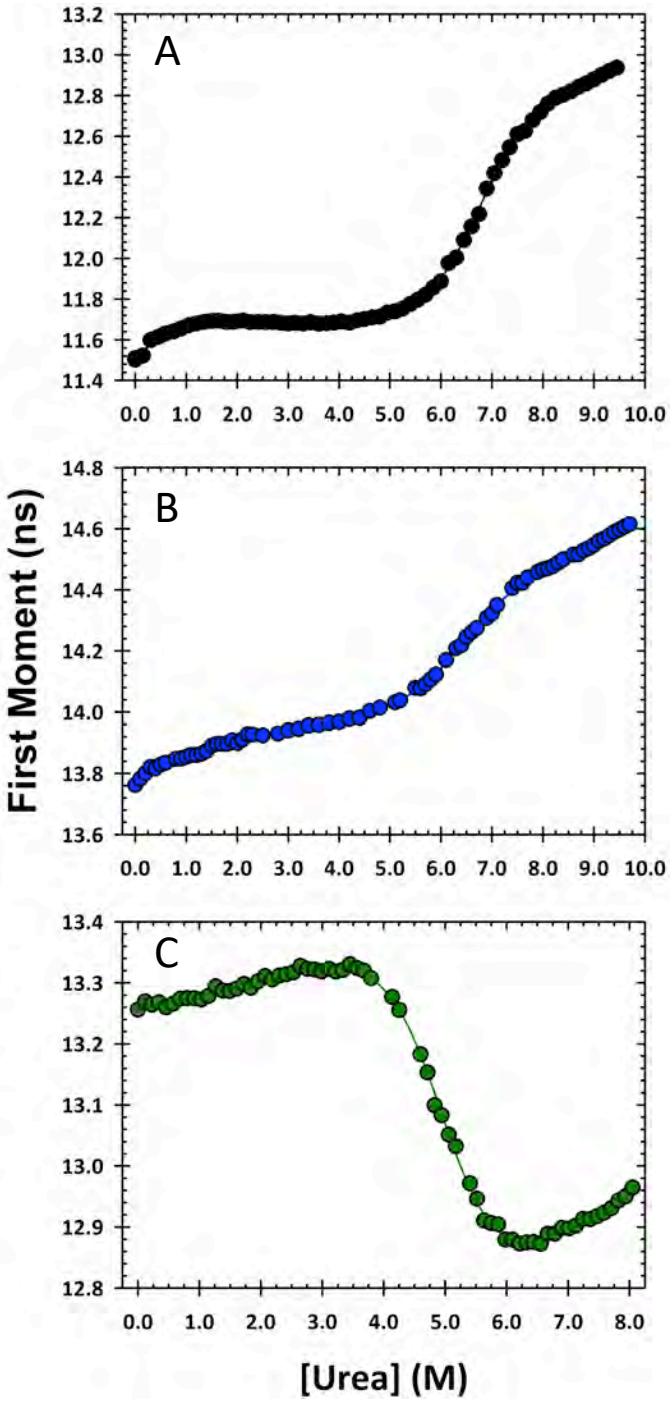

Figure 4.3: Equilibrium titrations of SS-V_H domains by FL Lifetime
The equilibrium titration of SS-V_{HA} (black, A), SS-V_{HB} (blue, B) and SS-V_{HC} (green, C) monitored by tryptophan lifetime detection are shown with fits to the data (solid lines). The parameters of the fits are shown in Table 4.2.

SS-V_{HA} and SS-V_{HB} show two transitions; one occurring at < 1 M urea and the other with a midpoint of ~6-7 M urea. These curves were fit to a 3-state model, N↔I↔U ,and the extracted thermodynamic parameters are reported in Table 4.2. Similar results were observed using CD spectroscopy ~235 nm but the midpoint of the early transition was different, indicating that this step does not reflect a global folding event (Fig. 4.4A-B, Table 4.2). By contrast, SS-V_{HC} FL (Fig. 4.3C) and CD (Fig. 4.4C) equilibrium denaturation curves show only a single transition with a midpoint of ~5 M urea. These curves were fit to a two-state model, N↔U, and the thermodynamic parameters are reported in Table 4.2.

The pair of transitions in the equilibrium denaturation of SS-V_{HA} and SS-V_{HB} demonstrates that the SS bond introduces a partially-folded state not observed for SS-V_{HC} or any of the 2SH-forms (Fig. 4.2). The lack of a fully folded baseline for the N↔I transition precluded a fully quantitative fit the of the data to a 3-state model. However the data above 2M urea could be fit to a 2-state model describing I↔U. The excellent agreement between the thermodynamic parameters for the FL and CD curves (Table 4.2) is consistent with a global folding reaction. Further support for this conclusion was provided by analyzing the CD data at 220 nm (Fig. 4.5), where the secondary structure is the principle contributor to the ellipticity. The ellipticity at 230 nm is thought to reflect aromatic packing[218] suggesting a more subtle rearrangement of secondary or tertiary structure. Conversely, the thermodynamic parameters determined from the CD and FL denaturation curves for SS-V_{HC} are in good agreement (Table 4.2), confirming the N↔U mechanism.

Table 4.2: Equilibrium Parameters for the folding reactions of the SS-V_H domains						
	$\Delta G^{\circ}_{NI\text{-}app}$	$m_{Ni\text{-}app}$	$C_{mNI\text{-}app}$	ΔG°_{IU}	m_{IU}	C_{mIU}
Tryptophan Lifetime						
V_{HA}	0.08 ± 0.53	1.92 ± 0.51	0.04 ± 0.28	7.44 ± 0.22	1.12 ± 0.04	6.65 ± 0.31
V_{HB}	-0.95 ± 2.33	1.24 ± 0.41	-0.80 ± -1.9	8.36 ± 1.23	1.29 ± 0.20	6.48 ± 1.40
V_{HC}	N/A	N/A	N/A	6.30 ± 0.12	1.26 ± 0.03	5.00 ± 0.15
Circular Dichroism						
V_{HA}	0.84 ± 0.09	0.39 ± 0.02	2.15 ± 0.26	7.16 ± 0.21	1.06 ± 0.03	6.75 ± 0.27
V_{HB}	-0.09 ± 0.17	0.42 ± 0.03	-0.21 ± 0.41	8.46 ± 0.09	1.24 ± 0.01	6.82 ± 0.09
V_{HC}	N/A	N/A	N/A	6.31 ± 0.30	1.28 ± 0.061	4.92 ± 0.34

The thermodynamic parameters obtained from fitting the urea titrations of the three SS-V_H domains are shown. The apparent equilibrium denaturant dependence of the N⟷I transition, $m_{NI\text{-}app}$, is shown in kcal mol^{-1} $[D]^{-1}$ where [D] is the molar concentration of denaturant. The units of the apparent folding free energy of the N⟷I transition, $\Delta G^{\circ}_{NI\text{-}app}$, are in kcal mol^{-1}. These are apparent values as the lack of a well-defined native baseline precluded their full quantification. The equilibrium denaturant dependence and folding free energy of the I⟷U transition, m_{IU} and ΔG°_{IU}, are shown in the same units. The midpoint, C_m, of each transition is shown as the [Urea] (M)

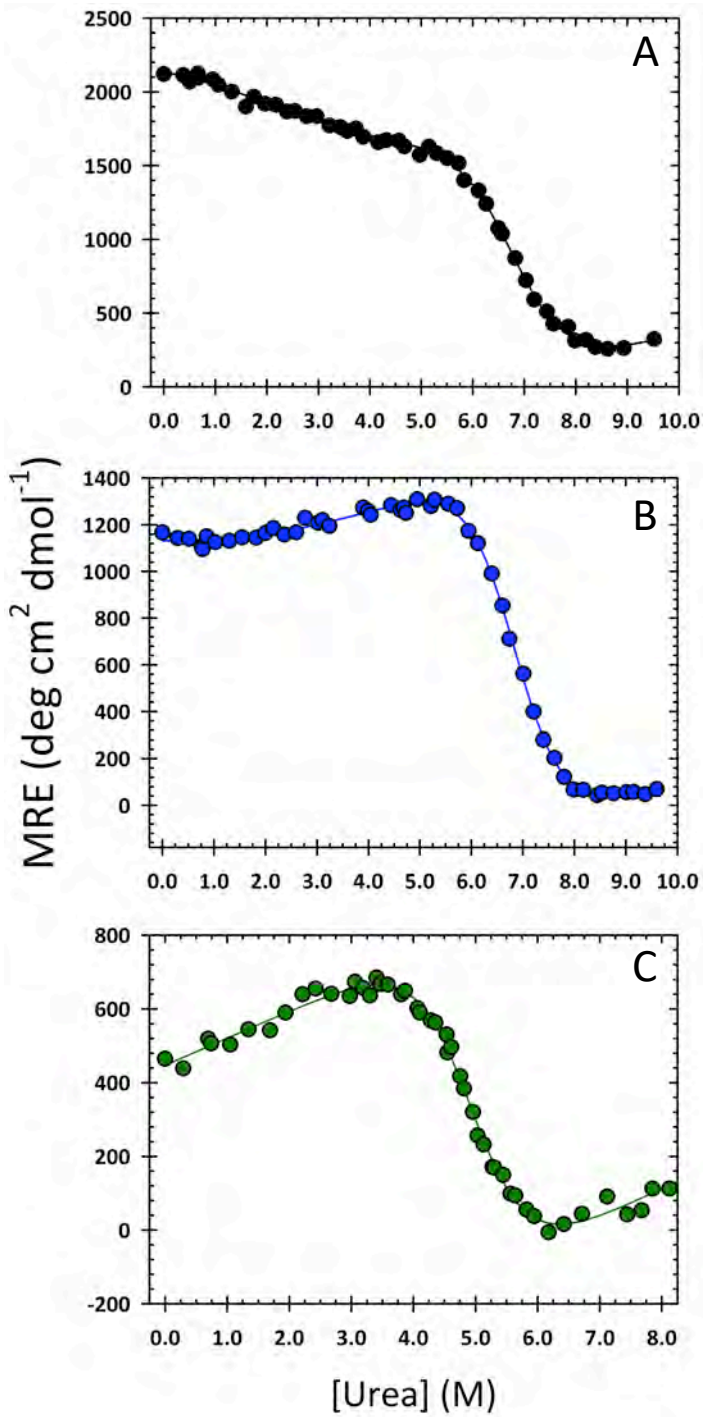

Figure 4.4: Equilibrium titrations of SS-V_H domains by CD
The equilibrium titration of SS-V_{HA} (black, A), SS-V_{HB} (blue, B) monitored by CD at 235 nm are shown and SS-V_{HC} (green, C) monitored CD are shown with fits to the data (solid lines). SS-V_{HA} and SS-V_{HB} were monitored at 235 nm and SS-V_{HC} was monitored at 236 nm to monitor the ~235 nm peak characteristic of antibody domains[219,220]. The parameters of the fits are shown in Table 4.2.

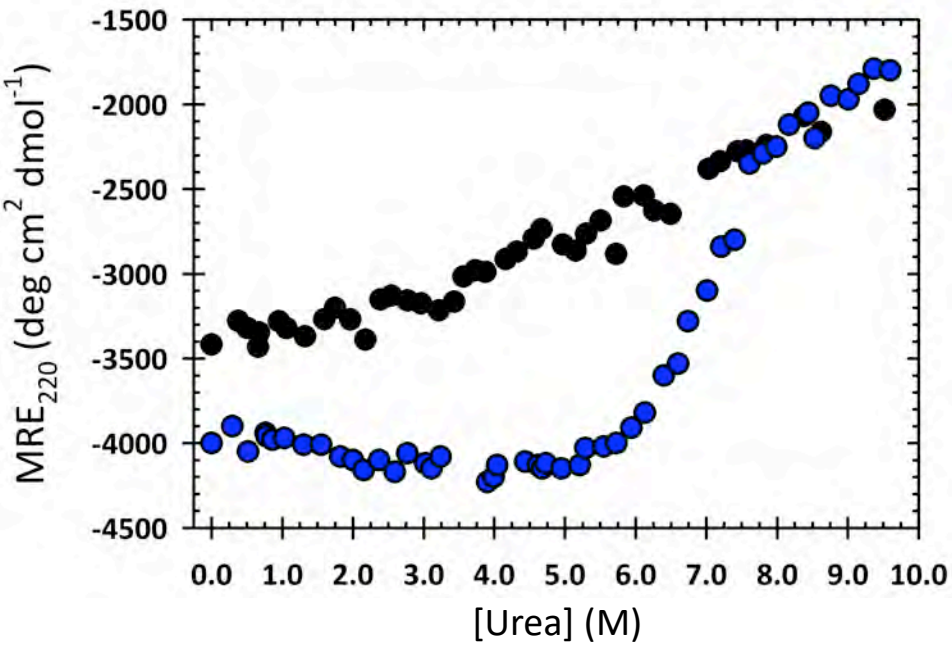

Figure 4.5: Equilibrium titrations of SS-V_{HA} and SS-V_{HB} by CD at 220mn
The equilibrium titration of SS-V_{HA} (black, A), SS-V_{HB} (blue, B) monitored by CD at 220 nm are shown. The disruption of the β-sheets is only evident at higher concentrations of urea.

Kinetic Analysis

Further insights into the folding free energy surfaces were obtained from a kinetic analysis of the folding reactions of the 3 SS-domains using FL intensity and fitting the traces to one or more exponentials. The resulting chevron plots are shown in Figure 4.6A-C and the parameters from the fits of these chevron plots are shown in Table 4.3. Unlike their 2SH-forms, SS-V_{HA} and SS-V_{HB} have complex chevron plots with multiple phases in both unfolding and refolding while SS-V_{HC} shows simple two-state behavior.

3 phases were observed when jumping SS-V_{HA} from native to denaturing conditions; a sub-second phase with a small denaturant dependence, and two slower phases that display typical chevron behavior. For refolding jumps between ~2-7 M urea, only two phases were observed and they correspond to the refolding legs of the 2 slower unfolding phases. Under strongly folding conditions, < 2 M urea, the faster refolding chevron rolls-over to be similar to the relaxation time for the denaturant independent faster phase in unfolding. The "nested" chevrons apparent between 5-9 M urea are evidence of parallel folding pathways[221].

We confirmed the parallel pathways by performing unfolding jumps to the midpoint of the nested chevrons, ~7 M urea, from different initial concentrations of denaturant(Fig 4.7, circles). The amplitude of the fast, urea-independent unfolding phase decreases progressively while the amplitudes of the faster unfolding chevron phase increases proportionally up to 5 M urea. Regardless of initial conditions, the faster and slower chevron phases maintain a constant ratio of ~80:20. This behavior is consistent with a pair of parallel folding reactions, an apparent 3-state reaction and a 2-state reaction:

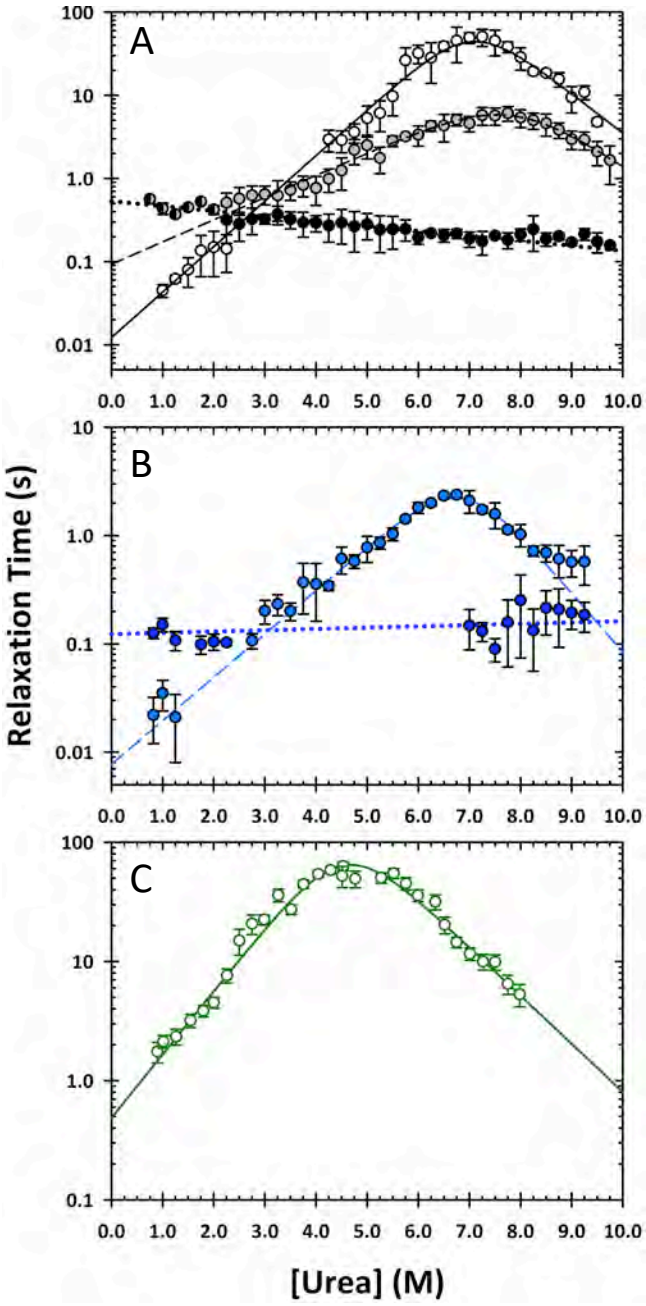

Figure 4.6: Chevron plots of the SS-V_H domains
Chevron plots of SS-V_{HA} (A, black) and SS-V_{HB} (B, blue), obtained with stopped-flow FL, are shown with fits to the data (solid lines). For SS-V_{HA}, the N⟷I transition is shown in black circles, the faster I⟷U transition is shown in grey circles and the slower I⟷U transition is shown in white circles. For SS-V_{HB}, the N⟷I transition is shown in dark blue circle and the I⟷U transition is shown in light blue circles. The chevron plot of SS-V_{HC} (C, green), obtained with manual-mixing FL and CD, is shown with a 2-state fit to the data (solid line).

Table 4.3: Kinetic Parameters for the folding reactions of the SS-V_H domains

	τ_f	$m_f^{\ddagger}$	τ_u	$m_u^{\ddagger}$	$\Delta G_{eq}^{\ddagger}$	$m_{eq}^{\ddagger}$	C_m
V_{HA} N↔I*	~0.53	~0.06	N/A	N/A	N/A	N/A	N/A
V_{HA} I↔U fast	0.09 ± 0.01	0.36 ± 0.02	$1.8x10^4$ ± $1.9x10^4$	-0.55 ± 0.07	7.10 ± 0.62	0.91 ± 0.07	7.8 ± 0.9
V_{HA} I↔U slow	0.01 ± 0.001	0.73 ± 0.02	$1.9x10^5$ ± $2.1x10^5$	-0.64 ± 0.07	9.66 ± 0.62	1.37 ± 0.08	7.1 ± 0.6
V_{HB} N↔I*	~0.123	~0.01	N/A	N/A	N/A	N/A	N/A
V_{HB} I↔U	$7.7x10^{-3}$ ± $1.5x10^{-3}$	0.54 ± 0.02	$2.5x10^4$ ± $2.5x10^4$	-0.73 ± 0.07	8.72 ± 0.59	1.27 ± 0.07	6.9 ± 0.6
V_{HC}	0.49 ± 0.06	0.71 ± 0.03	$8.9x10^3$ ± $3.8x10^3$	-0.54 ± 0.04	5.71 ± 0.26	1.25 ± 0.05	4.6 ± 0.3

The parameters obtained from fitting the chevron plots of SS-V_H domains are shown. The extrapolated refolding, τ_f, and unfolding, τ_u, relaxation times and the errors of the fit to the chevron plots (Fig. 4.6) are shown in seconds. The fits and errors for the refolding, $m_f^{\ddagger}$, and unfolding, $m_u^{\ddagger}$, portions of the chevron plots are shown in kcal mol^{-1} $[D]^{-1}$. The free energies of folding, $\Delta G^{\ddagger}_{eq}$, were calculated from the kinetic data and are shown in kcal mol^{-1}. The denaturant dependence of the free energy of folding, $m_{eq}^{\ddagger}$, calculated as, $m_{eq}^{\ddagger} = m_f^{\ddagger} - m_u^{\ddagger}$ is also shown in kcal mol^{-1} $[D]^{-1}$. The midpoint, C_m, of each titration is shown in [Urea] (M). *The N↔I transitions show little to no denaturant dependence so the parameters are from an estimated fit to data.

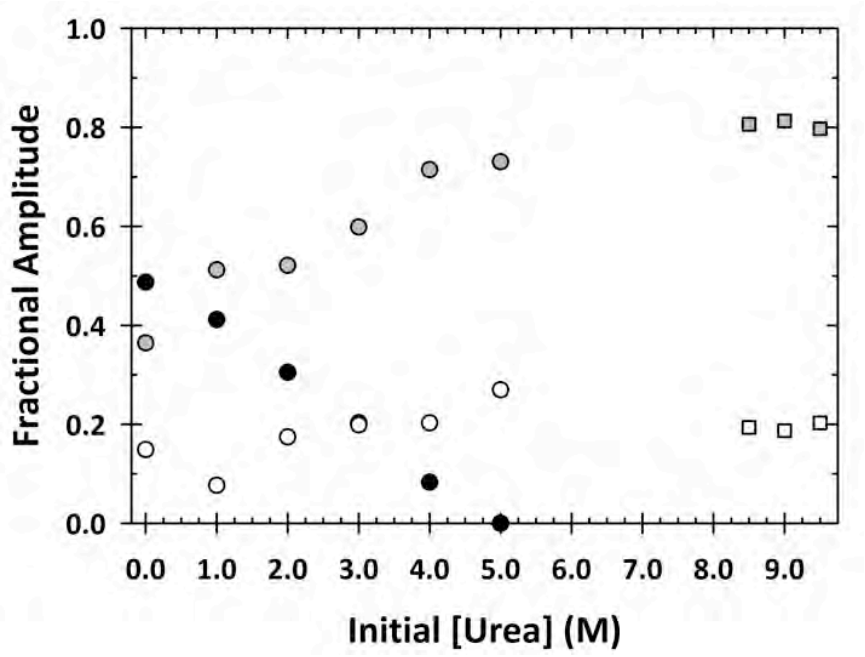

Figure 4.7: Fractional amplitudes of the three phases in SS-V_{HA} unfolding and refolding
The fractional amplitude of each phase of SS-V_{HA} unfolding (circles) and refolding (squares) is shown as a function of initial denaturant concentration. The colors correspond to those in Fig. 4.6A.

$$[N_1 \rightleftharpoons N^*] \rightleftharpoons U_1 \qquad \text{SS} - \text{V}_{HA} \text{ major pathway}$$

$$N_2 \rightleftharpoons U_2 \qquad \text{SS} - \text{V}_{HA} \text{ minor pathway}$$

The discrepancy between the midpoint of the early transition by FL lifetime (Fig. 4.3), CD (Figs. 4.4 and 4.5) and the relative amplitude data (Fig. 4.6 and 4.7) highlights the heterogeneity of this reaction and suggests that this early step reflects an ensemble of states within the native basin, [$N_1 \rightleftharpoons N^*$]. The independence of the relative amplitude of the slower chevron phase on the initial urea concentration (Fig 4.7) reflects the absence of a dynamic equilibrium between the N_1 and N_2 states prior to unfolding.

Evidence for parallel folding channels in refolding is provided by the observation of two phases under strongly refolding conditions, <2 M urea. The relaxation time of the faster refolding phase decreases with decreasing urea concentration in a manner that directly connects to the refolding leg of the slower chevron phase (Fig 4.6A). The slower refolding phase links directly to the urea-independent fast unfolding phase, identifying it as the $N^* \rightarrow N_1$ reaction, the reverse of the $N_1 \rightarrow N^*$ unfolding reactions. The persistence of the faster urea-dependent refolding phase at low denaturant concentrations rules out a single sequential reaction where the faster phase leads directly to the slower refolding phase. The absence of a dynamic equilibrium between the 2 unfolded states was demonstrated by varying the initial concentration of urea in the >7 M range and jumping to the same final condition at 7 M (Fig 4.7, squares).

The distinct kinetic *m*-values for the chevrons (Fig 4.6A and Table 4.3) must mean the SS-V_{HA} has two distinct unfolded conformations: (1) a compact unfolded state, U_1, with residual structure and (2) a fully expanded unfolded state, U_2. This conclusion is based on the observation that *m*-values reflect the change in buried surface area between 2 states in

a folding reaction[178]. In kinetic studies, the transition state ensemble, TSE, is the target state for unfolding and refolding reactions from folded and unfolded thermodynamic states, respectively. The sum of the absolute values of the kinetic *m*-values, m_f and m_u, are equivalent to the equilibrium *m*-value. The total *m*-value for the 3-state channel, 0.91±0.07 kcal mol^{-1} M^{-1}, is less than the ~1.45 kcal mol^{-1} M^{-1} expected for a fully unfolded chain of 125 residues with a disulfide bond[178]. By contrast, the total *m*-value for the 2-state channel is 1.37±0.08 kcal mol^{-1} M^{-1} in excellent agreement with the predicted value of 1.45 kcal mol^{-1} M^{-1}.

The folding kinetics of the SS- V_{HB} (Fig 4.6B) domain resemble, in several respects, those for the SS-V_{HA} domain (Fig 4.6A). For unfolding jumps a ~100 ms phase with almost no denaturant dependence precedes what appears to be a single, slower phase with a chevron dependence on the urea concentration. Refolding jumps reveal the ~100 ms phase seen in unfolding as well as faster phase whose relaxation time decreases at lower urea concentration. Although definitive evidence for parallel channels in unfolding via nested chevrons is lacking, the pair of phases under strongly refolding conditions make a compelling argument for parallel channels in refolding. It may be the case that the relaxation times of the putative pair of chevrons in SS-V_{HB} are too similar to resolve with exponential fitting of the FL intensity data. With the unfolding data consistent with an apparent 3-state reaction and the refolding data with parallel channels, we tentatively conclude that the same 2 channel folding mechanism holds for both SS-V_{HA} and SS-V_{HB}.

The kinetic folding analysis for the SS-V_{HC} domain revealed a single chevron response consistent with a simple 2-state mechanism (Fig. 4.6C).

$$N \rightleftharpoons U \qquad \text{SS} - V_{HC} \text{ mechanism}$$

The stability and *m*-value extracted from the kinetic data were in excellent agreement with those from the equilibrium data (Tables 4.2 and 4.3), further confirmation the 2-state nature of the reaction.

DISCUSSION

We study the folding mechanisms of the 2SH- and SS-forms of three V_H domains from mAbs with the goal of understanding why 2 of these domains can exist in both the 2SH and SS-form in the full length mAb while the third only exists in the SS-form[115,116].

Do the biophysical properties of 2SH-forms of the V_H domains determine the size of the incomplete disulfide population in the full length mAbs?

The 2SH-forms of the domains were prone to aggregation but refolding jumps at nanomolar concentrations revealed that these domains fold via a simple two-state mechanism (Figure 4.2A-C). The extrapolated refolding rates of the three domains (Table 4.1) indicated that the folding of 2SH-V_{HA} and 2SH-V_{HB} are ~100-fold faster than that of 2SH-V_{HC}. It is tempting to speculate that the rapid folding of 2SH-V_{HA} and 2SH-V_{HB} is able to compete with SS bond oxidation, while the much slower folding of 2SH-V_{HC} is unable to do so thereby providing an explanation for the presence of disulfide-reduced populations in mAb A and mAb B but not in mAb C. However, previous studies have indicated that folding drives rather than inhibits disulfide formation in Ig-fold proteins[222,223]. A possible explanation for this difference is that the folding times of the Ig-fold proteins in the previous studies, β_2-microglobulin[222] and titin[222,223], are on the order of seconds, similar to 2SH-V_{HC}, rather than the hundreds of milliseconds observed

for 2SH-V_{HA} and 2SH-V_{HB}. The large difference in folding times might allow kinetic competition in one system, while preventing it in another. Another potential explanation is that antibody molecules are produced through overexpression in CHO cells[224]. This overexpression combined with the rapid folding of $V_{HA/HB}$ may overwhelm the natural disulfide isomerases leading to the formation of the disulfide-reduced population in the full length mAbs. However, the complexity observed in folding reactions of the SS-V_{HA} and SS-V_{HB} domains indicates there may be a preferable alternative explanation.

Analysis of the biophysical properties of the SS-forms of the V_{HA} and V_{HB} domains revealed unexpected complexities

Both equilibrium and kinetic analyses revealed that SS-V_{HA} and SS-V_{HB} domains contained surprising amounts of complexity in their folding mechanisms, while the folding of SS-V_{HC} was more simple and similar to that of the 2SH-domains (Figs. 4.2A-C and 4.6A-C). Kinetic analysis of V_{HA} and V_{HB} revealed that both exhibit distinct parallel folding pathways that do not appear to interconvert (Figs. 4.6-4.7). The reactions of the SS-V_{HA} pathways were separate and had distinct rates likely due to the presence of residual structure in the unfolded state of one pathway that was not present in the other. The folding reactions of the SS-V_{HB} domain were only distinguishable under strongly refolding conditions, and therefore there was no indication of residual structure in the unfolded state. This is evidence that the residual structure in SS-V_{HA} was not the sole cause of the parallel channels. In stark contrast, the kinetic analysis of V_{HC} was in excellent agreement with the equilibrium data (Tables 4.2 and 4.3) confirming that it folds in a simple, two-state manner.

What gives rise to the complexities in the folding of the SS-$V_{HA/HB}$ domains?

The complexities observed in the folding of the SS-, but not 2SH-, domains were unexpected. This suggests that the presence of the disulfide bond is driving the parallel folding pathways in V_{HA} and V_{HB} but not V_{HC}. We aligned crystal structures of the three domains to determine if there was a structural basis for the complexity (Fig. 4.8). The RMSD of the alignment was excellent, 0.437 Å, due to the high sequence identity shared between the framework regions of three domains (Fig. 1.10). However, examination of the SS bonds revealed that the SS bonds of V_{HA} and V_{HB} are in a different orientation from that of the SS bond in V_{HC}. The SS bond of both V_{HA} and V_{HB} are right-handed disulfides, defined by a positive χ^3 angle, while the SS bond of V_{HC} is a left-handed disulfide, defined by a negative χ^3 angle[225].

Using an algorithm to calculate the disulfide strain energy (DSE) in these domains based on the χ-angles in the crystal structures[226,227], we found that the right-handed SS bonds of V_{HA} and V_{HB} have strains of 3.81 kcal mol^{-1} and 4.22 kcal mol^{-1}, respectively. In contrast, the left-handed SS bond in V_{HC} had a strain of only 1.49 kcal mol^{-1} (Table 4.4). Notably, the DSE of the right-handed SS bonds, ~4 kcal mol^{-1}, is nearly equivalent to the redox potential of Protein Disulfide Isomerase (PDI), -180 mV[228]. This would preclude a complete transfer of the disulfide from PDI to the V_H domain, suggesting that the disulfide-free population in the full length mAb is due to unfavorable thermodynamics of the redox reaction, rather than kinetic competition between folding and oxidation. It is also noteworthy that the cysteines forming the SS bond, C22 and C96, directly precede CDR1 and CDR3 in sequence. It seems plausible that the conformation of the CDRs could alter the disulfide orientation resulting in strain. This strain would then lead to both

Figure 4.8: Alignment of the three V_H domains
The three V_H domains were aligned in Pymol and the result is shown as ribbon diagrams. SS-V_{HA} is in black, SS-V_{HB} is in blue and SS-V_{HC} is in green. The disulfide bonds are shown for SS-V_{HA} and SS-V_{HB} in yellow and for SS-V_{HC} in orange. The crystal structures of the SS-V_{HA} and SS-V_{HB} domains were solved by Dr. Yiren Xu. The SS-V_{HC} domain is adapted from PDB: 1N8Z[99].

Table 4.4: Disulfide Bond Parameters in several V_H domains

	χ^1	χ^2	χ^3	Distance	χ^{-2}	χ^{-1}	DSE
V_{HA}	171.98	116.31	65.75	2.04	163.08	65.44	3.81
V_{HB}	177.94	118.07	76.83	2.05	156.55	40.36	4.22
V_{HC}	-177.69	166.16	-84.73	2.04	-165.16	166.99	1.49
3B9V[28]	178.26	165.1	-81.95	2.06	-179.76	-177.24	0.81
4U3X[29]	169.84	172.77	-76.01	2.08	179.8	-177.08	0.98
3P9W[30]	176.15	166.11	-74.77	2.04	173.16	179.85	1.01
5VH3[31]	167.36	-172.05	-72.37	2.07	-175.47	176.77	1.33
1MEL[32]	-170.46	172.9	-100.87	2.00	-172.12	-178.62	1.57
1OHQ[33]	161.91	165.96	-65.56	2.04	172.25	-172.81	2.53
5I0Z	176.09	111.07	74.66	2.04	169.24	58.81	2.75
1OL0[34]	178.49	123.96	62.12	2.10	167.19	60.54	3.73
1T2J	163.22	122.23	65.11	2.06	169.85	60.69	4.11
5FV2[35]	168.79	125.27	54.14	2.04	167.24	72.03	5.32

The parameters of the disulfide bonds in the crystal structures of the examined V_H domains are shown. The dihedral angles along the cysteine residues of the disulfide bond are shown in degrees. The distance between the sulfur atoms is shown in Å. The disulfide strain energy is shown in kcal mol^{-1}. The three V_H domains examined in this study are in the top rows while the other ten domains are listed as PDB numbers and sorted by DSE from lowest to highest. The dashed line separates the low and high DSE groups.

the disulfide-free population in full length mAbs and the complex folding free energy surfaces described above.

Are right-handed disulfide bonds common in VH domains?

We compared V_{HA}, V_{HB}, and V_{HC} to ten other V_H domains from crystal structures available in the PDB and found that 6 had left-handed SS bonds with DSE<2.5 kcal mol^{-1} and aligned well with the SS bond of V_{HC} (Table 4.4, Fig. 4.9A). The other four had DSE>2.5 kcal mol^{-1} and aligned well with the right-handed SS bonds of V_{HA} and V_{HB} (Table 4.4, Fig. 4.9B). In this cursory analysis, there appeared to be a slight preference for left-handed SS bonds in V_H domains, but a more thorough analysis would clarify any such trends.

Close examination of other mAbs containing V_H domains with right-handed SS bonds might reveal a direct connection between SS bond orientation and in the presence of the disulfide-free population. If so, then examination of the SS bonds of mAb crystal structures could provide a relatively rapid and accessible method to screen for the potential to form disulfide-free populations as well as complexities in the folding free energy surfaces.

MATERIALS AND METHODS

V_H domain purification

E. Coli cell pellets were provided by the Kao group at Genentech and Dr Xiren Yu developed the following purification method with minor adaptions by our group. Cell pellets were thawed at 4°C overnight. Cell pellets were then resuspended in 25 mM Tris

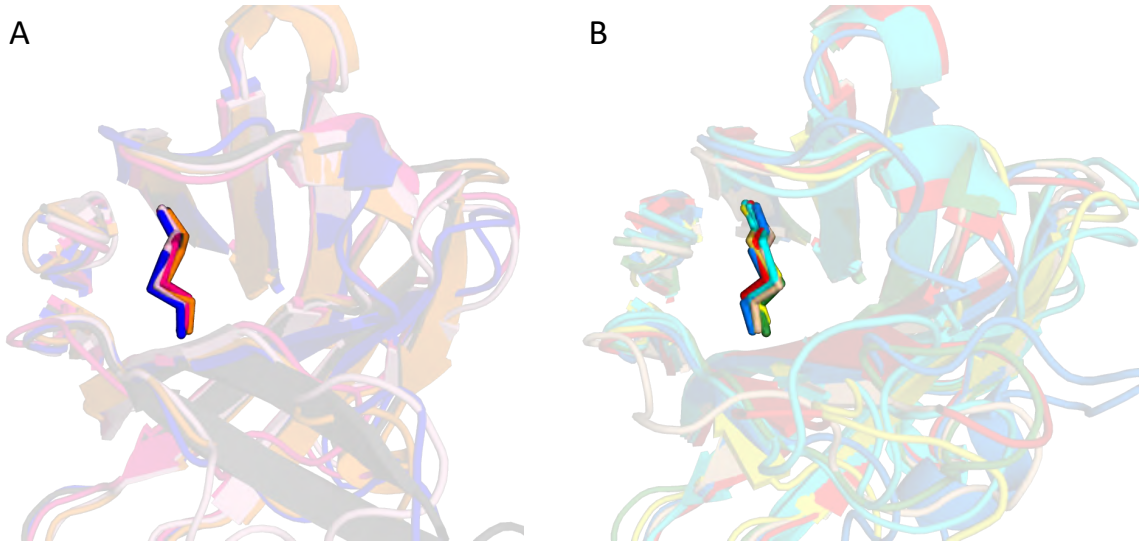

Figure 4.9: Disulfide Bond Orientation in other V_H domains compared to SS-$V_{HA/HB/HC}$
Ten V_H domains from the PDB (Table 4.4) were aligned with the three SS-V_H domains in Pymol and then sorted by disulfide bond orientation into right-handed (A) and left-handed (B) groups.

buffer pH=7.5 at a 1:10 w/v ratio. The resulting solution was passed through a microfluidizer twice to ensure complete lysis. The lysis solution was centrifuged at 12,000 g for 15 minutes and the supernatant was collected and centrifuged at 12,000 g for another 15 min and the supernatant was collected. This lysate solution was loaded onto a series of HiTrap SP FF 25 mL columns (GE Healthcare) at a 10:1 ratio of lysate volume-to-column volume using a BIO-RAD NGC system (BIO-RAD). In order to scale up the process multiple 25 mL columns were linked in series to allow greater load volumes. The proteins were eluted with a 0-25% gradient of 25 mM Tris, 2 M NaCl, pH=7.5 over ten column volumes and the peak fractions were collected. These fractions were combined and loaded onto a series pre-charged and pre-equilibrated HiTrap IMAC FF, 25 mL columns (GE Healthcare) using a BIO-RAD NGC system (BIO-RAD). The series of IMAC columns prevented overloading. The remaining impurities were washed off over 12 column volumes with a buffer of 85% 50 mM Tris, 50 mM NaCl, 5 mM Imidazole, pH=7.5 and 15% 50 mM Tris, 50 mM NaCl, 1 M Imidazole, pH=7.5. The V_H domains were then eluted with 15-40% gradient of the 50 mM Tris, 50 mM NaCl, 1 M Imidazole, pH=7.5 buffer over 15 column volumes. The resulting peak fractions were collected and dialyzed into the experimental buffer 20 mM MES, 50 mM NaCl pH=6.0.

At each step along this process small fractions were collected and run by both SDS-PAGE and western blots to confirm the presence of the V_H domains. Western blots were able to detect the domains using an antibody against the His-6-tag in the domains.

Disulfide Reduction

Cysteines of the disulfide in the V_H domains were inaccessible unless under strongly denaturing conditions[115], so reduction was carried out by shifting the protein into 9 M Urea, 20 mM MES, 50 mM NaCl, 1mM TCEP buffer pH=6.0 for an hour. For the reduced experiments all buffers used were the same with the addition of 1 mM TCEP to ensure reduction.

Stopped Flow Fluorescence Data Collection and Analysis

The folding kinetic experiments on V_H domains were performed using an Applied Photophysics system and the data was collected with the Applied Photophysics Pro-Data software. A 357 nm 40 nm wide bandpass filter (Semrock, Rochester, NY) was used to detect the emission peak. Jumps of 1:10 protein: buffer ratio at 20°C were performed to various final amounts of urea and final protein concentrations of 150-500 nM. The nM concentrations were used to decouple the refolding and aggregation processes of the 2SH-V_H domains. The observed reaction times did not change across the range of concentrations examined, indicating that they are unimolecular folding reactions. Similar experiments were performed for the refolding and unfolding of the SS-V_H domains. For accurate detection of the millisecond reactions the drive pressure was held constant throughout the experiment to prevent back mixing artifacts in the data. Each experimental run involved >10 jumps to each final urea concentration and each experiment was performed >3 times.

The initial folding and unfolding reactions were monitored on a log scale from milliseconds to 100's of seconds to ensure the entire reaction was captured. Once the

reaction time was assessed, smaller time windows were used to obtain denser data sets. The resulting traces were independently fit to the necessary exponentials until the data was well described using in-house software (available upon request). In many traces a linear term was required to account for photobleaching of the sample. The time constants of the from the >10 fits for each final concentration of denaturant in an was then averaged to obtain the relaxation time at that point. These were recorded and plotted as a function of final denaturant concentration to obtain the chevron plots in Figs. 4.2 and 4.5. The data point of the chevrons represent the average of >3 experiments to each final concentration of denaturant and the error bars denote the standard deviation.

Chevron Plot Analysis

The chevron plots were input and fit to the following equation using SigmaPlot software and the results of the fits were reported in Tables 4.1 and 4.3:

$$\frac{1}{\tau_{obs}} = k_{obs} = \frac{1}{k_f^0 e^{\left(\frac{-m_f[D]}{\mathrm{RT}}\right)} + k_u^0 e^{\left(\frac{-m_u[D]}{\mathrm{RT}}\right)}} \qquad Eq.\,4.1$$

Where τ_{obs} is the observed time constant from the fit and its inverse, k_{obs} is the observed rate constant, k_f^0 is the folding rate constant in the absence of denaturant, m_f is the denaturant dependence of the refolding rate constant, k_u^0 is the unfolding rate constant in the absence of denaturant, m_u is the denaturant dependence of the unfolding rate constant, [D] is concentration of urea, R is the ideal gas constant and T is temperature.

The fits of the chevrons were recorded and used to calculate the free energy of folding of each domain with the following equation

$$\Delta G^0 = -\mathrm{RTln}\left(\frac{k_u^0}{k_f^0}\right) \qquad Eq.4.2$$

The total denaturant dependence of the free energy of folding, m_{eq}, was calculated as m_f-m_u. The midpoint of each reaction, C_m, was calculated as $\Delta G^0/m_{eq}$.

To fit the 2SH-chevrons, a maximum value of m_{eq} was estimated using the methods described by Myers, Pace and Sholtz[178]. This was used as an upper limit to allow the fitting routine to function with an estimated value of m_u despute being based on solely refolding data.

Equilibrium FL Lifetime Data Collection and Analysis

The equilibrium FL lifetime data were collected using excitation by tripled output from a Ti:sapphire laser at 3.8 MHz repetition rate as previously described[140,147,206]. The excitation power was kept below several hundred μWatts at 292 nm. Counting rates in the detector channel were kept to less than 1×10^5 counts per second. A 357 nm 40 nm wide bandpass filter (Semrock, Rochester, NY) was used to select the emission peak.

The samples were prepared from protein stocks in buffer and in ≥9 M urea and mixed precisely into a 96-well microplate using a Hamilton Microlab 500 titrator and in-house software. The plates were allowed to equilibrate for several hours. Equilibrium measurements were performed with a home-built autosampler running custom LabVIEW software (available upon request) to the lifetime instrument. The samples were brought into a flow cell (1mm×1mm) using a Hamilton Microlab 560 dual syringe pump and oscillated at a flow rate of 10 μL/s during data collection to reduce exposure of the sample to the beam and minimize photobleaching. The FL lifetime at each point was

calculated by first moment analysis of the measured decay curves using in-house software. The resulting plots were fit to two-state or three-state mechanisms using in-house software to determine the ΔG_{eq}^{0} and m_{eq} for each transition. The midpoints of the transitions were calculated as described above.

Equilibrium CD Data Collection and Analysis

The equilibrium CD data were collected on a Jasco J-810 CD spectrophotometer (Jasco Inc.). Samples were prepared as described above and the far-UV CD spectra was monitored from 215-260 nm in 0.2 cm pathlength synthetic fused silica cuvette using a scan rate of 20 nm min^{-1} and a response of 8 seconds. The refractive index of each sample was measured and used to determine the final denaturant concentration of urea. The CD signal was converted to mean residue ellipticity and observed as a function of final urea concentration. The CD spectra of the V_H domains contained a positive peak ~235 nm which was used to monitor their unfolding as previously described[219,220]. The resulting plots were fit to two-state or three-state mechanisms using in-house software to determine the ΔG_{eq}^{0} and m_{eq} for each transition. The midpoints of the transitions were calculated as described above.

Fractional Amplitude Analysis

Jumps were performed as described above except the final denaturant concentration was held constant at 7 M urea and the initial concentration varied from (1) 0→5 M urea in unfolding and (2) 8.5→9 M urea in refolding. >10 jumps from each initial concentration were collected and fit as described above. The average amplitude

associated with each time constant was divided by the sum of all the amplitudes to obtain the fractional amplitude of each species. These fractional amplitudes was plotted as a function of initial concentration of urea.

V_H Domain Alignment

The crystal structures of the V_{HA} and V_{HB} domains were solved by Dr. Xiren Yu at Genentech which she graciously provided for our use. The domains were aligned in Pymol and the RMSD was recorded. The other V_H domains were obtained from the listed PDB entries and were aligned in the same way.

Disulfide Strain Energy Calculation

The disulfide strain energy of each domain was calculated using the following equation based on the AMBER force field[229,230]:

$$DSE = 2(1 + cos(3\chi^1)) + 2(1 + cos(3\chi^{-1}) + (1 + cos(3\chi^2)) + (1 + cos(3\chi^{-2})) + 3.5(1 + cos(2\chi^3)) + 0.6(1 + cos(3\chi^3) \quad Eq\ 4.3$$

Where DSE is the disulfide strain energy in kcal mol^{-1} and $\chi^{1,2,3,-2,-1}$ are the dihedral angles of the cysteine residues in the disulfide bond. The dihedral angles were determined from the PDBs using the "Disulfide Bond Dihedral Angle Energy Server" available at https://services.mbi.ucla.edu/disulfide/.

Chapter V – Discussion

REVIEW

The goal of the research presented above was to examine the folding of several proteins with similar topologies to gain insight into where in the folding process opportunities for misfolding and aggregation arise. We selected the series of proteins SOD1 and the V_H domains of mAb A,B, and C that all adopt an Ig-fold. The Ig-fold is a β-rich structure found throughout nature and is used in numerous proteins with diverse functions. Perhaps most notably, this fold is used by that antibody molecules that form the backbone of adaptive immunity. Ig-fold proteins are especially attractive candidates for this type of research as they are linked to human diseases typically characterized by aggregation. Furthermore, due to their natural role in the immune response, they are often used in protein engineering efforts focused on the creation of novel therapeutics. Any insights into the interplay between productive folding and deleterious aggregation of Ig-fold proteins could prove invaluable in disease treatment and therapeutic development.

In Chapter II, the link between the productive folding and harmful aggregation of the ALS-linked protein, SOD1, was explored through the characterization of a set of peptides serving as a model for the behavior of the nascent polypeptide chain. End labeling each peptide with a FRET donor-acceptor pair and monitoring the excited-state lifetime decay of the donor allowed the calculation of the trFRET efficiency. Examination of the response of the trFRET efficiency of each SOD1 peptide to denaturant found that the peptide containing the Loop VII-β8 region at the C-terminus of SOD1 demonstrated a unique sensitivity to denaturant. Utilizing a two-dimensional form of maximum entropy modeling to simultaneously fit the donor-only and donor-acceptor

lifetime decays allowed the amplitude to be assigned to species with discrete energy transfer and donor decay rate constants. This analysis revealed that the Loop VII-β8 peptide in aqueous conditions contains populations of both compact and expanded states. The compact state was only marginally stable and its population shifted into the expanded state in a two-state-like manner in response to even small amounts denaturant. Two variants of the Loop VII-β8 peptide, one containing the ALS-linked mutation I149T and another truncating the polar residues of Loop VII, were examined and the trFRET efficiency of both was found to be insensitive to denaturant. 2DMEM analysis further confirmed that neither variant formed the compact state seen in the WT peptide.

These results indicate that the very C-terminus of SOD1 was capable of forming nonnative structure in the absence of the rest of the sequence. Surprisingly, alteration of residues in either hydrophobic stretch of β8 or the polar region Loop VII was able to prevent the formation of this nonnative structure. The simplest explanation for this observation is that the nonnative state is defined by an atypical interaction between these two dissimilar regions. We proposed that the intramolecular interactions that form this species could swap with other polypeptide chains, thereby becoming intermolecular interactions that may be a precursor to toxic aggregation. This model could prove useful in searching for new therapies.

Chapter III discussed the work probing the behavior of the SOD1 polypeptide chain by monitoring the response of the SS- and 2SH-mSOD1 to denaturant in the presence of the kosmotropes, glycerol and glucose. The initial aim was to identify any other transient or marginally stable species populated during the unfolding and refolding of SOD1, yet no such species were observed. In all the glycerol and glucose

concentrations examined, the equilibrium titration and kinetic folding jumps were well-described by a two-state reaction. Interestingly, we observed that the free energy of folding and the denaturant dependence of the free energy of folding were found to vary linearly with increasing molar amounts of viscogen. By introducing parameters to describe the viscogen dependence of the free energy of folding and the viscogen dependence of the denaturant dependence, we developed a novel method of globally fitting chevron data in the presence of viscogens. Using this method to globally fit the SS-mSOD1 and 2SH-mSOD1 data, we found that the folding and unfolding of 2SH-mSOD1 was limited by internal chain friction but disulfide-oxidized SOD1 was not.

This finding implicates the regions constrained by the disulfide bond, Loop IV and β8, as drivers of the internal friction. When considered in context with the results of the peptide analysis from Chapter 2 discussed above, the sequence of β8 appears to be the most likely driver of internal friction, which may be a consequence of this region adopting a nonnative structure. In the broader context of SOD1-linked ALS, these results suggest that the C-terminal region of SOD1 is a potential origin of early misfolding and aggregation. Interestingly, this conclusion is supported by the observations that (1) a short peptide from β8 nucleates *in vitro* aggregation[138], (2) this region is the core region protected from proteolysis in *in vivo* aggregates[139] and (3) a peptide similar to the one used in our trFRET studies, spanning residues 131-153, was used to develop an antibody that specifically recognizes misfolded variants of SOD1[72,152].

Chapter IV describes the preliminary results of the examination of three V_H domains from therapeutic antibodies, two that contain 2SH-populations in the full length mAb and one that does not. Classical kinetic and thermodynamic analyses of these three

domains revealed surprising distinctions between their folding mechanisms. In each of the three domains the 2SH-forms were found to fold in simple, two-state manner. However the SS-forms of the domains that formed the 2SH- populations had complex folding free energy surfaces that were tied to the presence of the disulfide bond. Examination of the disulfide bond in the crystal structures of these domains suggested that the complexities arise from the formation of strained disulfide bonds.

Interestingly, these results indicate that the strained disulfide bond in the two V_H domains is both protective and disruptive. The disulfide bond protects the domains from aggregation at the cost of disrupting their otherwise simple folding reactions. We speculate that examination of the orientation of disulfide bonds in mAbs may provide a valuable tool for determining whether their individual domains have complex folding free energy surfaces.

IMPACT

The work described above could have substantial impact in the ALS and antibody development fields.

In the ALS field, the compaction of the C-terminus of SOD1 presents a new target for the development of therapeutics aimed at treating the source of disease. The currently available therapeutic options for ALS only modestly slow disease progression and their exact mechanism of action remains unclear. However, therapeutics specifically designed to disrupt the compact state might inhibit the early steps of aggregation, thereby significantly hindering the onset and progression of the disease. Such therapeutics might

even be useful as a form of early intervention in families with known ALS-linked SOD1 genetic variants.

The observation of a potential link between strained disulfides in VH domains and free thiols in mAb populations could have a remarkable impact on the future of antibody development. As screening methods and technology have advanced, there have been more frequent observations of heterogeneity in therapeutic antibody molecules. Such heterogeneity can have negative impacts on therapeutic efficacy and is likely involved in the huge financial and time costs of successfully developing new therapeutics. The population of mAbs containing unformed disulfide bonds discussed in Chapter 4 being an example of such heterogeneity that has been observed frequently in recent years[115,214,215,217,231,232]. By suggesting a potential source of the population lacking disulfide bonds, this work could allow future antibody engineering and development research to identify methods to avoid the formation of such populations entirely.

Beyond this potential impact, the work discussed in Chapter 4 above is, too this author's knowledge, the first noting the variability of the disulfide orientation in the domains of immunoglobulins and the first examining the role of strain in the disulfides. In current research, the concept of disulfide strain is primarily connected to redox or allosteric mechanisms[227]. Studies often classify the disulfide bonds of immunoglobulins as "structural disulfides" and do not examine the role of disulfide strain in this group further[233]. The above work suggests that even in disulfides viewed as "structural" high disulfide strain energies can have an impact with significant biological consequences. A more thorough examination of structural disulfides, especially those in immunoglobulin domains, may prove to be a novel research focus in the future.

FUTURE DIRECTIONS

Moving forward with these results presents several possibilities for future research studies.

One option is to explore the role of the nonnative structure at the C-terminus of SOD1 in driving aggregation. Mapping the point at which the intramolecular interactions are replaced with intermolecular interactions would provide further insights regarding how this region nucleates the aggregation process. Obtaining high resolution information about the structure of the nonnative species and early oligomers through experiment would likely be complicated and technically challenging. However, the short size of the peptides examined, ~30 residues, would be suitable for simulation based studies. Combinations of simulation and experimental approaches have been used to obtain structural details for other diseases linked peptides such as Aβ for Alzheimer's[234], α-synuclein for Parkinsons[235], and IAPP for Type II Diabetes[236]. The nonnative structure involving crosstalk between hydrophobic and polar regions could also present the opportunity to develop novel force fields, most of which are currently built to drive native folding.

The link between this nonnative structure and friction-limited folding in SOD1 could also be established further. Examining the folding of a variant known to disrupt the formation of the nonnative structure, such as I149T-mSOD1, for friction-limited folding could determine whether the two observations are directly linked. Furthermore, the internal friction and/or the formation of the nonnative state may present an opportunity for identification of therapeutic agents. The Loop VII-β8 peptide could be used in a screen of small molecule libraries to identify compounds that disrupt the nonnative

structure. A similar approach has been used in our lab to screen for small molecules to stabilize the native dimer of SOD1, and the results of such screens might provide precursors to new therapies.

Advancing the research into antibody domains could be done by performing comparably detailed analyses of the folding free energy landscapes of V_H domains from other therapeutic mAbs. Doing so could reveal further evidence for unexpected levels of complexity in these engineered proteins. Comparison of such folding data with the disulfide bond strains calculated for crystal structures could provide further evidence of a link between the observation of disulfide strain and complex folding mechanisms.

A similar approach could also be used to examine V_L domains, as it seems possible that they experience similar issues with disulfide oxidation and complexities during folding. A bioinformatics study of disulfide orientation in variable domains of antibody crystal structures in the PDB could allow a direct comparison of the distributions of disulfide orientations in engineered antibodies versus those in naturally evolved antibodies. Doing so might also establish the analysis of disulfide orientation as a novel and rapid method for the detection of antibody variable domains susceptible to loss of the intrachain disulfide.

A new form of V_H domains, single domain antibodies (nanonbodies, sdAbs), are a subject of much attention currently[237,238]. sdAbs are based on antibodies from camelids and cartilaginous fish (e.g. sharks) comprised of only heavy chains, which allows the variable domains to function as monomers. Utilizing sdAbs for therapeutics would remove many of the complicating factors of antibody production such as glycosylation and proper assembly. Currently, sdAbs are characterized in the same way as mAbs, i.e.

with thermal melts, which leaves their true biophysical properties poorly understood[239,240]. Performing analyses similar to those described in Chapter IV on sdAbs would not only provide more detailed information on their biophysical properties, but might also be used as a basis for future engineering studies.

PERSPECTIVE

The folding free energy landscape of a protein defines the stages through which a nascent unfolded polypeptide chain progresses towards its productive native state and this landscape is itself defined by the amino acid sequence of the protein. Analyzing the process of folding characterizes the sequence of events occurring between the unfolded and folded chain, thereby providing a description of the free energy landscape. This landscape can then provide insights into which steps might frustrate or disrupt productive folding, simultaneously enabling deleterious aggregation.

The link between folding/aggregation and human disease and medicine has long been established. Unfortunately, the exact mechanisms by which aggregation overwhelms productive folding and how this drives toxicity remains elusive. As the productive folding landscapes of more proteins are mapped and the mechanisms through which these proteins begin to explore deleterious aggregation landscapes are defined, insights into how to drive folding and/or disrupt aggregation will become more evident. Using these insights in the development and/or engineering of therapeutic agents will allow drastic improvements in the health and wellbeing of patients.

The work described above is an example of how a thorough biophysical analysis can reveal information that will guide the creation of new medicines and accelerate the advancement of those that currently exist.

References

1. Habert, E., Christiax, A. & Anfinsen, B. Side-chain interactions governing the pairing of half-cystine residues in ribonuclease. *J. Biol. Chem.* **237**, 1839–1844 (1962).
2. Levinthal, C. Are there pathways for protein folding? *J. Chim. Phys.* **65**, 44–45 (1968).
3. Wolynes, P. G., Luthey-Schulten, Z. & Onuchic, J. N. Fast-folding eriments and the topography of protein folding energy landscapes. *Chem. Biol.* **3**, 425–432 (1996).
4. Leopold, P. E., Montal, M. & Onuchic, J. N. Protein folding funnels: a kinetic approach to the sequence-structure relationship. *Proc. Natl. Acad. Sci.* **89**, 8721–8725 (1992).
5. A., D. K., S, C. H. & S., C. H. From Levinthal to Pathways to Funnels: The "New View" of Protein Folding Kinetics. *Nat. Struct. Biol.* **4**, 10 (1997).
6. Haass, C. & Selkoe, D. J. Soluble protein oligomers in neurodegeneration: Lessons from the Alzheimer's amyloid β-peptide. *Nat. Rev. Mol. Cell Biol.* **8**, 101–112 (2007).
7. Chuang, E., Hori, A. M., Hesketh, C. D. & Shorter, J. Amyloid assembly and disassembly. *J. Cell Sci.* **131**, jcs189928 (2018).
8. Borgia, M. B., Nickson, A. A., Clarke, J. & Hounslow, M. J. A mechanistic model for amorphous protein aggregation of immunoglobulin-like domains. *J. Am. Chem. Soc.* **135**, 6456–6464 (2013).
9. Chiti, F. & Dobson, C. M. Protein Misfolding, Amyloid Formation, and Human Disease: A Summary of Progress Over the Last Decade. *Ann. Rev. Biochem* **86**, 27–68 (2017).
10. D'Aguanno, V. *et al.* Systemic Amyloidosis: a Contemporary Overview. *Clin. Rev. Allergy Immunol.* 1–19 (2019). doi:10.1007/s12016-019-08759-4
11. Rajkumar, S. V., Gertz, M. A. & Kyle, R. A. Primary systemic amyloidosis with delayed progression to multiple myeloma. *Cancer* **82**, 1501–1505 (1998).
12. Agrawal, N. & Skelton, A. A. Structure and Function of Alzheimer's Amyloid βeta Proteins from Monomer to Fibrils: A Mini Review. *Protein J.* **1**, 3 (2019).
13. Kragh, C. L., Romero-Ramos, M., Halliday, G. & Jensen, P. H. Alpha synuclein in

parkinson's disease. in *Handbook of Neurotoxicity* **2**, 691–726 (Springer New York, 2014).

14. Orr, M. E., Sullivan, A. C. & Frost, B. A Brief Overview of Tauopathy: Causes, Consequences, and Therapeutic Strategies. *Trends in Pharmacological Sciences* **38**, 637–648 (2017).

15. Fiesel, F. C. & Kahle, P. J. TDP-43 and FUS/TLS: Cellular functions and implications for neurodegeneration. *FEBS Journal* **278**, 3550–3568 (2011).

16. Robert H. Brown & Ammar Al--Chalabi. Amyotrophic Lateral Sclerosis. *N. Engl. J. Med.* **377**, 162–72 (2017).

17. Arrasate, M. & Finkbeiner, S. Protein aggregates in Huntington's disease. *Experimental Neurology* **238**, 1–11 (2012).

18. Jucker, M. & Walker, L. C. Self-propagation of pathogenic protein aggregates in neurodegenerative diseases. *Nature* **501**, 45–51 (2013).

19. Westermark, P., Andersson, A. & Westermark, G. T. Islet Amyloid Polypeptide, Islet Amyloid, and Diabetes Mellitus. *Physiol. Rev.* **91**, 795–826 (2011).

20. Galant, N. J., Westermark, P., Higaki, J. N. & Chakrabartty, A. Transthyretin amyloidosis: an under-recognized neuropathy and cardiomyopathy. *Clin. Sci.* **131**, 395–409 (2017).

21. Bellotti, V., Mangione, P. & Merlini, G. Review: Immunoglobulin light chain amyloidosis - The archetype of structural and pathogenic variability. *J. Struct. Biol.* **130**, 280–289 (2000).

22. Bianchi, G., Anderson, K. C., Harris, N. L. & Sohani, A. R. The heavy chain diseases: clinical and pathologic features. *Oncology (Williston Park).* **28**, 45–53 (2014).

23. Westermark, G. T., Fändrich, M. & Westermark, P. AA Amyloidosis: Pathogenesis and Targeted Therapy. *Annu. Rev. Pathol. Mech. Dis.* **10**, 321–344 (2014).

24. Scarpioni, R. *et al.* Dialysis-related amyloidosis: Challenges and solutions. *International Journal of Nephrology and Renovascular Disease* **9**, 319–328 (2016).

25. Moreau, K. L. & King, J. A. Protein misfolding and aggregation in cataract disease and prospects for prevention. *Trends Mol. Med.* **18**, 273–282 (2012).

26. Park, P. S.-H. Rhodopsin Oligomerization and Aggregation. *J. Membr. Biol.* 1–11 (2019). doi:10.1007/s00232-019-00078-1

27. Sletten, K., Westermark, P. & Natvig, J. Characterization of amyloid fibril proteins from medullary carcinoma of the thyroid. *J. Exp. Med.* **143**, 993–998 (1976).

28. Ratanji, K. D., Derrick, J. P., Dearman, R. J. & Kimber, I. Immunogenicity of therapeutic proteins: Influence of aggregation. *J. Immunotoxicol.* **11**, 99–109 (2014).

29. Perez-Alvarez, R., Pérez-De-Lis, M. & Ramos-Casals, M. Biologics-induced autoimmune diseases. *Current Opinion in Rheumatology* **25**, 56–64 (2013).

30. Laptoš, T. & Omersel, J. The importance of handling high-value biologicals: Physico-chemical instability and immunogenicity of monoclonal antibodies. *Exp. Ther. Med.* **15**, 3161–3168 (2018).

31. Weiss, W. F., Young, T. M. & Roberts, C. J. Principles, approaches, and challenges for predicting protein aggregation rates and shelf life. *J. Pharm. Sci.* **98**, 1246–1277 (2009).

32. Roberts, C. J. Protein aggregation and its impact on product quality. *Curr. Opin. Biotechnol.* **30**, 211–217 (2014).

33. Roberts, C. J. Therapeutic protein aggregation: mechanisms, design, and control. *Trends Biotechnol.* **32**, 372–380 (2014).

34. Hirota-Nakaokal, N., Hasegawa, K., Naiki, H. & Goto, Y. Dissolution of β2-Microglobulin Amyloid Fibrils Hironobu by Dimethylsulfoxide. *J Biochem* **134**, 159–164 (2003).

35. Meersman, F. & Dobson, C. M. Probing the pressure-temperature stability of amyloid fibrils provides new insights into their molecular properties. *Biochim. Biophys. Acta - Proteins Proteomics* **1764**, 452–460 (2006).

36. Makin, O. S., Atkins, E., Sikorski, P., Johansson, J. & Serpell, L. C. Molecular basis for amyloid fibril formation and stability. *Proc. Natl. Acad. Sci.* **102**, 315–320 (2005).

37. Nelson, R. *et al.* Structure of the cross-β spine of amyloid-like fibrils. *Nature* **435**, 773–778 (2005).

38. Shammas, S. L. *et al.* Perturbation of the stability of amyloid fibrils through alteration of electrostatic interactions. *Biophys. J.* **100**, 2783–2791 (2011).

39. Williams, A. D. *et al.* Mapping Aβ amyloid fibril secondary structure using scanning proline mutagenesis. *J. Mol. Biol.* **335**, 833–842 (2004).

40. Kathuria, S. V., Chan, Y. H., Nobrega, R. P., Özen, A. & Matthews, C. R. Clusters of isoleucine, leucine, and valine side chains define cores of stability in high-energy states of globular proteins: Sequence determinants of structure and stability. *Protein Sci.* **25**, 662–675 (2016).

41. Tzul, F. O., Schweiker, K. L. & Makhatadze, G. I. Modulation of folding energy landscape by charge–charge interactions: Linking experiments with computational modeling. *Proc. Natl. Acad. Sci.* **112**, E259–E266 (2015).

42. Jahn, T. R. & Radford, S. E. Folding versus aggregation: Polypeptide conformations on competing pathways. *Arch. Biochem. Biophys.* **469**, 100–117 (2008).

43. Krebs, M. R. . *et al.* Formation and seeding of amyloid fibrils from wild-type hen lysozyme and a peptide fragment from the β-domain. *J. Mol. Biol.* **300**, 541–549 (2000).

44. Sen Mojumdar, S. *et al.* Partially native intermediates mediate misfolding of SOD1 in single-molecule folding trajectories. *Nat. Commun.* **8**, 1–11 (2017).

45. Colon, W. & Kelly, J. W. Partial Denaturation of Transthyretin Is Sufficient for Amyloid Fibril Formation in Vitro. *Biochemistry* **31**, 8654–8660 (1992).

46. Mackness, B. C., Tran, M. T., McClain, S. P., Matthews, C. R. & Zitzewitz, J. A. Folding of the RNA Recognition Motif (RRM) domains of the Amyotrophic Lateral Sclerosis (ALS)-linked protein TDP-43 reveals an intermediate state. *J. Biol. Chem.* **289**, 8264–8276 (2014).

47. Lang, L. *et al.* SOD1 aggregation in ALS mice shows simplistic test tube behavior. *Proc. Natl. Acad. Sci.* **112**, 9878–9883 (2015).

48. Fink, A. L. Protein aggregation: Folding aggregates, inclusion bodies and amyloid. *Fold. Des.* **3**, R9–R23 (1998).

49. Chaturvedi, S. K., Siddiqi, M. K., Alam, P. & Khan, R. H. Protein misfolding and aggregation: Mechanism, factors and detection. *Process Biochemistry* **51**, 1183–1192 (2016).

50. Rochet, J. C. & Lansbury, P. T. Amyloid fibrillogenesis: Themes and variations. *Current Opinion in Structural Biology* **10**, 60–68 (2000).

51. Anderson, V. L. & Webb, W. W. A desolvation model for trifluoroethanol-induced

aggregation of enhanced green fluorescent protein. *Biophys. J.* **102**, 897–906 (2012).

52. Torisu, T., Maruno, T., Hamaji, Y., Ohkubo, T. & Uchiyama, S. Synergistic Effect of Cavitation and Agitation on Protein Aggregation. *J. Pharm. Sci.* **106**, 521–529 (2016).

53. Nilsson, M. R. Techniques to study amyloid fibril formation in vitro. *Methods* **34**, 151–160 (2004).

54. Scholtz, J. M., Grimsley, G. R. & Pace, C. N. *Chapter 23 Solvent Denaturation of Proteins and Interpretations of the m Value*. *Methods in Enzymology* **466**, (Elsevier Inc., 2009).

55. Halloran, K. T. *et al.* Frustration and folding of a TIM barrel protein. *Proc. Natl. Acad. Sci.* 201900880 (2019). doi:10.1073/pnas.1900880116

56. Shi, J. *et al.* Atomistic structural ensemble refinement reveals non-native structure stabilizes a sub-millisecond folding intermediate of CheY. *Sci. Rep.* **7**, (2017).

57. Zhou, Z., Feng, H., Ghirlando, R. & Bai, Y. The High-Resolution NMR Structure of the Early Folding Intermediate of the Thermus thermophilus Ribonuclease H. *J. Mol. Biol.* **384**, 531–539 (2008).

58. Feng, H., Zhou, Z. & Bai, Y. A protein folding pathway with multiple folding intermediates at atomic resolution. *Proc. Natl. Acad. Sci.* **102**, 5026–5031 (2005).

59. Yu, H. *et al.* Direct observation of multiple misfolding pathways in a single prion protein molecule. *Proc. Natl. Acad. Sci.* **109**, 5283–5288 (2012).

60. Mccords, J. M. & Fridovich, I. Superoxide dismutase. An enzymic function for erythrocuprein (hemocuprein). *J. Biol. Chem.* **244**, 6049–6055 (1969).

61. Rosen, D. R. *et al.* Mutations in Cu/Zn superoxide dismutase gene are associated with familial amyotrophic lateral sclerosis. *Nature* **362**, 59–62 (1993).

62. Oskarsson, B., Gendron, T. F. & Staff, N. P. Amyotrophic Lateral Sclerosis: An Update for 2018. *Mayo Clin. Proc.* **93**, 1617–1628 (2018).

63. Forman, M. S., Trojanowski, J. Q. & Lee, V. M. Neurodegenerative diseases: a decade of discoveries paves the way for therapeutic breakthroughs. **10**, 1055–1063 (2004).

64. Abel, O., Powell, J. F., Andersen, P. M. & Al-Chalabi, A. ALSoD: A user-friendly online bioinformatics tool for amyotrophic lateral sclerosis genetics. *Hum. Mutat.*

33, 1345–1351 (2012).

65. Volk, A. E., Weishaupt, J. H., Andersen, P. M., Ludolph, A. C. & Kubisch, C. Current knowledge and recent insights into the genetic basis of amyotrophic lateral sclerosis. *Medizinische Genetik* **30**, 252–258 (2018).

66. Kayatekin, C., Zitzewitz, J. A. & Matthews, C. R. Zinc Binding Modulates the Entire Folding Free Energy Surface of Human Cu , Zn Superoxide Dismutase. *J. Mol. Biol.* **384**, 540–555 (2008).

67. Tainer, J. A., Getzoff, E. D., Beem, K. M., Richardson, J. S. & Richardson, D. C. Determination and analysis of the 2 Å structure of copper, zinc superoxide dismutase. *J. Mol. Biol.* **160**, 181–217 (1982).

68. Furukawa, Y., Torres, A. S. & O'Halloran, T. V. Oxygen-induced maturation of SOD1: A key role for disulfide formation by the copper chaperone CCS. *EMBO J.* **23**, 2872–2881 (2004).

69. Hwang, C., Sinskey, A. J., Lodish, H. F., Hwang, C. & Sinskey, A. J. State Redox Oxidized of Glutathione in the Endoplasmic Reticulum. *Science (80-.).* **257**, 1496–1502 (1992).

70. Bruijn, L. I. *et al.* Aggregation and Motor Neuron Toxicity of an ALS-Linked SOD1 Mutant Independent from Wild-Type SOD1. *Science (80-.).* **281**, 1851–1854 (1998).

71. Strange, R. W. *et al.* Variable metallation of human superoxide dismutase: Atomic resolution crystal structures of Cu-Zn, Zn-Zn and as-isolated wild-type enzymes. *J. Mol. Biol.* **356**, 1152–1162 (2006).

72. Forsberg, K. *et al.* Misfolded SOD1 inclusions in patients with mutations in C9orf72 and other ALS/FTD-associated genes. *J. Neurol. Neurosurg. Psychiatry* 1–9 (2019). doi:10.1136/jnnp-2018-319386

73. Wang, J., Xu, G. & Borchelt, D. R. High molecular weight complexes of mutant superoxide dismutase 1: Age-dependent and tissue-specific accumulation. *Neurobiol. Dis.* **9**, 139–148 (2002).

74. Gill, C. *et al.* SOD1-positive aggregate accumulation in the CNS predicts slower disease progression and increased longevity in a mutant SOD1 mouse model of ALS. *Sci. Rep.* **9**, (2019).

75. Zhu, C., Beck, M. V, Griffith, J. D., Deshmukh, M. & Dokholyan, N. V. Large SOD1 aggregates, unlike trimeric SOD1, do not impact cell viability in a model of

amyotrophic lateral sclerosis. *Proc. Natl. Acad. Sci.* **115**, 4661–4665 (2018).

76. Sangwan, S. *et al.* Atomic structure of a toxic, oligomeric segment of SOD1 linked to amyotrophic lateral sclerosis (ALS). *Proc. Natl. Acad. Sci.* **114**, 8770–8775 (2017).

77. Proctor, E. A. *et al.* Nonnative SOD1 trimer is toxic to motor neurons in a model of amyotrophic lateral sclerosis. *Proc. Natl. Acad. Sci.* **113**, 614–619 (2016).

78. Zetterstrom, P. *et al.* Soluble misfolded subfractions of mutant superoxide dismutase-1s are enriched in spinal cords throughout life in murine ALS models. *Proc. Natl. Acad. Sci.* **104**, 14157–14162 (2007).

79. Holmes, C. *et al.* Long-term effects of Aβ42 immunisation in Alzheimer's disease: follow-up of a randomised, placebo-controlled phase I trial. *Lancet* **372**, 216–223 (2008).

80. Sengupta, U., Nilson, A. N. & Kayed, R. The Role of Amyloid-β Oligomers in Toxicity, Propagation, and Immunotherapy. *EBioMedicine* **6**, 42–49 (2016).

81. Conway, K. A. *et al.* Acceleration of oligomerization, not fibrillization, is a shared property of both alpha-synuclein mutations linked to early-onset Parkinson's disease: implications for pathogenesis and therapy. *Proc. Natl. Acad. Sci.* **97**, 571–576 (2000).

82. Roberts, H. L. & Brown, D. R. Seeking a mechanism for the toxicity of oligomeric α-synuclein. *Biomolecules* **5**, 282–305 (2015).

83. Sekhar, A. *et al.* Probing the free energy landscapes of ALS disease mutants of SOD1 by NMR spectroscopy. *Proc. Natl. Acad. Sci. U. S. A.* **113**, E6939–E6945 (2016).

84. Sekhar, A. *et al.* Thermal fluctuations of immature SOD1 lead to separate folding and misfolding pathways. *Elife* **4**, e07296 (2015).

85. Teilum, K. *et al.* Transient structural distortion of metal-free Cu/Zn superoxide dismutase triggers aberrant oligomerization. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 18273–8 (2009).

86. Shaw, B. F. *et al.* Local unfolding in a destabilized, pathogenic variant of superoxide dismutase 1 observed with H/D exchange and mass spectrometry. *J. Biol. Chem.* **281**, 18167–76 (2006).

87. Lang, L., Kurnik, M., Danielsson, J. & Oliveberg, M. Fibrillation precursor of superoxide dismutase 1 revealed by gradual tuning of the protein-folding

equilibrium. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 17868–73 (2012).

88. Kayatekin, C., Zitzewitz, J. A. & Matthews, C. R. Disulfide-Reduced ALS Variants of Cu, Zn Superoxide Dismutase Exhibit Increased Populations of Unfolded Species. *J. Mol. Biol.* **398**, 320–331 (2010).

89. Lepock, J. R., Arnold, L. D., Torrie, B. H., Andrews, B. & Kruuv, J. Structural analyses of various Cu2+, Zn2+-superoxide dismutases by differential scanning calorimetry and Raman spectroscopy. *Arch. Biochem. Biophys.* **241**, 243–251 (1985).

90. Forman, H. J. & Fridovich, I. On the stability of bovine superoxide dismutase. The effects of metals. *J. Biol. Chem.* **248**, 2645–2649 (1973).

91. Lindberg, M. J., Normark, J., Holmgren, A. & Oliveberg, M. Folding of human superoxide dismutase: disulfide reduction prevents dimerization and produces marginally stable monomers. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 15893–8 (2004).

92. 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.* **280**, 17266–74 (2005).

93. Kayatekin, C., Cohen, N. R. & Matthews, C. R. Enthalpic barriers dominate the folding and unfolding of the human Cu, Zn superoxide dismutase monomer. *J. Mol. Biol.* **424**, 192–202 (2012).

94. Feige, M. J., Hendershot, L. M. & Buchner, J. How antibodies fold. *Trends Biochem. Sci.* **35**, 189–198 (2009).

95. Feige, M. J. & Buchner, J. Principles and engineering of antibody folding and assembly. *Biochim. Biophys. Acta - Proteins Proteomics* **1844**, 2024–2031 (2014).

96. Schroeder, H. W. & Cavacini, L. Structure and function of immunoglobulins. *J. Allergy Clin. Immunol.* **125**, S41–S52 (2010).

97. Bork, P., Holm, L. & Sander, C. The immunoglobulin fold: Structural classification, sequence patterns and common core. *Journal of Molecular Biology* **242**, 309–320 (1994).

98. Halaby, D. M. M., Poupon, A. & Mornon, J.-P. The immunoglobulin fold family: sequence analysis and 3D structure comparisons. *Protein Eng.* **12**, 563–571 (1999).

99. Cho, H.-S. *et al.* Structure of the extracellular region of HER2 alone and in complex with the Herceptin Fab. *Nature* **421**, 756–7560 (2003).

100. Schatz, D. G. & Ji, Y. Recombination centres and the orchestration of V(D)J recombination. *Nature Reviews Immunology* **11**, 251–263 (2011).

101. Di Noia, J. M. & Neuberger, M. S. Molecular Mechanisms of Antibody Somatic Hypermutation. *Annu. Rev. Biochem.* **76**, 1–22 (2007).

102. Singh, S. *et al.* Monoclonal Antibodies: A Review. *Curr. Clin. Pharmacol.* **13**, 85–99 (2017).

103. Walsh, G. Biopharmaceutical benchmarks 2018. *Nat. Biotechnol.* **36**, 1136–1145 (2018).

104. Elgundi, Z., Reslan, M., Cruz, E., Sifniotis, V. & Kayser, V. The state-of-play and future of antibody therapeutics. (2016). doi:10.1016/j.addr.2016.11.004

105. Ecker, D. M., Jones, S. D. & Levine, H. L. The therapeutic monoclonal antibody market. *MAbs* **7**, 9–14 (2015).

106. Kaplon, H. & Reichert, J. M. Antibodies to watch in 2019. *MAbs* **11**, 219–238 (2019).

107. Van Norman, G. A. Drugs, Devices, and the FDA: Part 1: An Overview of Approval Processes for Drugs. *JACC Basic to Transl. Sci.* **1**, 170–179 (2016).

108. DiMasi, J. A., Grabowski, H. G. & Hansen, R. W. Innovation in the pharmaceutical industry: New estimates of R&D costs. *J. Health Econ.* **47**, 20–33 (2016).

109. Klee, G. G. Human anti-mouse antibodies. *Arch. Pathol. Lab. Med.* **124**, 921–3 (2000).

110. Chames, P., Van Regenmortel, M., Weiss, E. & Baty, D. Therapeutic antibodies: Successes, limitations and hopes for the future. *British Journal of Pharmacology* **157**, 220–233 (2009).

111. Safdari, Y., Farajnia, S., Asgharzadeh, M. & Khalili, M. Antibody humanization methods - A review and update. *Biotechnology and Genetic Engineering Reviews* **29**, 175–186 (2013).

112. Winter, G., Griffiths, A. D., Hawkins, R. E. & Hoogenboom, H. R. Making Antibodies by Phage Display Technology. *Annu. Rev. Immunol.* **12**, 433–455 (1994).

113. Vermeer, A. W. P. & Norde, W. The Thermal Stability of Immunoglobulin : Unfolding and Aggregation of a Multi-Domain Protein. *Biophys. J.* **78**, 394–404

(2000).

114. Menzen, T. & Friess, W. High-throughput melting-temperature analysis of a monoclonal antibody by differential scanning fluorimetry in the presence of surfactants. *J. Pharm. Sci.* **102**, 415–428 (2013).

115. Zhang, T. *et al.* Identification and characterization of buried unpaired cysteines in a recombinant monoclonal IgG1 antibody. *Anal. Chem.* **84**, 7112–7123 (2012).

116. Harris, R. J. Heterogeneity of Recombinant Antibodies: Linking Structure to Function. *Dev Biol* **122**, 117–125 (2005).

117. Slamon, D. J. *et al.* Use of Chemotherapy plus a Monoclonal Antibody against HER2 for Metastatic Breast Cancer That Overexpresses HER2. *N. Engl. J. Med.* **344**, 783–792 (2002).

118. Knappik, A. *et al.* Fully synthetic human combinatorial antibody libraries (HuCAL) based on modular consensus frameworks and CDRs randomized with trinucleotides. *J. Mol. Biol.* **296**, 57–86 (2000).

119. Rothe, C. *et al.* The Human Combinatorial Antibody Library HuCAL GOLD Combines Diversification of All Six CDRs According to the Natural Immune System with a Novel Display Method for Efficient Selection of High-Affinity Antibodies. *J. Mol. Biol.* **376**, 1182–1200 (2008).

120. Prassler, J. *et al.* HuCAL PLATINUM, a synthetic fab library optimized for sequence diversity and superior performance in mammalian expression systems. *J. Mol. Biol.* **413**, 261–278 (2011).

121. Ferreiro, D. U., Cho, S. S., Komives, E. A. & Wolynes, P. G. The energy landscape of modular repeat proteins: Topology determines folding mechanism in the ankyrin family. *J. Mol. Biol.* **354**, 679–692 (2005).

122. Chi, C. N. *et al.* A conserved folding mechanism for PDZ domains. *FEBS Lett.* **581**, 1109–1113 (2007).

123. Carstensen, L. *et al.* Conservation of the folding mechanism between designed primordial (βα)8-barrel proteins and their modern descendant. *J. Am. Chem. Soc.* **134**, 12786–12791 (2012).

124. Swindells, M. B. *et al.* abYsis: Integrated Antibody Sequence and Structure—Management, Analysis, and Prediction. *J. Mol. Biol.* **429**, 356–364 (2017).

125. Wu, T. & Kabat, E. An Analysis of the Sequences of the Variable Regions of Bence Jones Proteins and Myeloma Light Chains and their Implications for

Antibody Complementarity. *J. Exp. Med.* **132**, 211–250 (1970).

126. Redler, R. L. & Dokholyan, N. V. The Complex Molecular Biology of Amyotrophic Lateral Sclerosis (ALS). *Prog. Mol. Biol. Transl. Sci.* **107**, 215–262 (2012).

127. Valentine, J. S., Doucette, P. A. & Potter, S. Z. Copper-Zinc Superoxide Dismutase and Amyotrophic Lateral Sclerosis. *Annu. Rev. Biochem* **74**, 563–93 (2005).

128. Ross, C. A. & Poirier, M. A. Protein aggregation and neurodegenerative disease. *Nat. Med.* **10**, S10–S17 (2004).

129. Chattopadhyay, M. & Valentine, J. S. Aggregation of Copper–Zinc Superoxide Dismutase in Familial and Sporadic ALS. *Antioxid. Redox Signal.* **11**, 1603–1614 (2009).

130. Ross, C. A. & Poirier, M. A. What is the role of protein aggregation in neurodegeneration? *Nature Reviews Molecular Cell Biology* **6**, 891–898 (2005).

131. Lelie, H. L. *et al.* Copper and zinc metallation status of copper-zinc superoxide dismutase from amyotrophic lateral sclerosis transgenic mice. *J. Biol. Chem.* **286**, 2795–806 (2011).

132. Svensson, A.-K. E. *et al.* Metal-Free ALS Variants of Dimeric Human Cu,Zn-Superoxide Dismutase Have Enhanced Populations of Monomeric Species. *PLoS One* **5**, 18–23 (2010).

133. Oztug Durer, Z. A. *et al.* Loss of Metal Ions, Disulfide Reduction and Mutations Related to Familial ALS Promote Formation of Amyloid-like Aggregates from Superoxide Dismutase. *PLoS One* **4**, 5004 (2009).

134. Furukawa, Y., Kaneko, K., Yamanaka, K., O'Halloran, T. V & Nukina, N. Complete loss of post-translational modifications triggers fibrillar aggregation of SOD1 in the familial form of amyotrophic lateral sclerosis. *J. Biol. Chem.* **283**, 24167–24176 (2008).

135. Chattopadhyay, M. *et al.* The Disulfide Bond, but Not Zinc or Dimerization, Controls Initiation and Seeded Growth in Amyotrophic Lateral Sclerosis-linked Cu,Zn Superoxide Dismutase (SOD1) Fibrillation. *J. Biol. Chem.* **290**, 30624–36 (2015).

136. Culik, R. M. *et al.* Effects of maturation on the conformational free-energy landscape of SOD1. *Proc. Natl. Acad. Sci.* **115**, E2546–E2555 (2018).

137. Svensson, A. K. E., Bilsel, O., Kondrashkina, E., Zitzewitz, J. A. & Matthews, C. R. Mapping the Folding Free Energy Surface for Metal-free Human Cu,Zn Superoxide Dismutase. *J. Mol. Biol.* **364**, 1084–1102 (2006).

138. Ivanova, M. I. *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.* **111**, 197–201 (2014).

139. 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.* **285**, 22221–22231 (2010).

140. Wu, Y., Kondrashkina, E., Kayatekin, C., Matthews, C. R. & Bilsel, O. Microsecond acquisition of heterogeneous structure in the folding of a TIM barrel protein. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 13367–72 (2008).

141. Skilling, J. & Bryan, R. K. Maximum entropy image reconstruction: general algorithm. *Mon. Not. R. astr. Soc* **211**, 111–124 (1984).

142. Livesey, A. K. & Brochon, J. C. Analyzing the Distribution of Decay Constants in Pulse-Fluorimetry Using the Maximum Entropy Method. *Biophys. J.* **52**, 693–706 (1987).

143. Das, R. K., Ruff, K. M. & Pappu, R. V. Relating sequence encoded information to form and function of intrinsically disordered proteins This review comes from a themed issue on Sequences and topology. *Curr. Opin. Struct. Biol.* **32**, 102–112 (2015).

144. Kohn, J. E., Gillespie, B. & Plaxco, K. W. Non-Sequence-Specific Interactions Can Account for the Compaction of Proteins Unfolded under “Native” Conditions. *J. Mol. Biol.* **394**, 343–350 (2009).

145. Chan, C. K. *et al.* Submillisecond protein folding kinetics studied by ultrarapid mixing. *Proc. Natl. Acad. Sci. U. S. A.* **94**, 1779–84 (1997).

146. Krantz, B. A., Mayne, L., Rumbley, J., Englander, S. W. & Sosnick, T. R. Fast and Slow Intermediate Accumulation and the Initial Barrier Mechanism in Protein Folding. *J. Mol. Biol.* **324**, 359–371 (2002).

147. Kathuria, S. V. *et al.* Microsecond Barrier-Limited Chain Collapse Observed by Time-Resolved FRET and SAXS. *J. Mol. Biol.* **426**, 1980–1994 (2014).

148. Engelborghs, Y. The analysis of time resolved protein fluorescence in multi-tryptophan proteins. *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* **57**, 2255–

2270 (2001).

149. Bajzer, Z. & Prendergast, F. G. A Model for Multiexponential Tryptophan Fluorescence Intensity Decay in Proteins. *Biophys. J. Vol.* **65**, 2313–2323 (1993).

150. Szabo, A. G. & Rayner, D. M. Fluorescence Decay of Tryptophan Conformers in Aqueous Solution. *J. Am. Chem. Soc.* **102**, (1980).

151. Goldenberg, D. P. Computational Simulation of the Statistical Properties of Unfolded Proteins. *J. Mol. Biol.* **326**, 1615–1633 (2003).

152. Forsberg, K. *et al.* Novel antibodies reveal inclusions containing non-native SOD1 in sporadic ALS patients. *PLoS One* **5**, 11552 (2010).

153. Forsberg, K., Andersen, P. M., Marklund, S. L. & Brännström, T. Glial nuclear aggregates of superoxide dismutase-1 are regularly present in patients with amyotrophic lateral sclerosis. *Acta Neuropathol.* **121**, 623–634 (2011).

154. Danielsson, J., Kurnik, M., Lang, L. & Oliveberg, M. Cutting off functional loops from homodimeric enzyme superoxide dismutase 1 (SOD1) leaves monomeric β-barrels. *J. Biol. Chem.* **286**, 33070–83 (2011).

155. Guinn, E. J., Jagannathan, B. & Marqusee, S. Single-molecule chemo-mechanical unfolding reveals multiple transition state barriers in a small single-domain protein. *Nat. Commun.* **6**, (2015).

156. Jagannathan, B., Elms, P. J., Bustamante, C. & Marqusee, S. Direct observation of a force-induced switch in the anisotropic mechanical unfolding pathway of a protein. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 17820–17825 (2012).

157. Conchillo-Solé, O. *et al.* AGGRESCAN: a server for the prediction and evaluation of 'hot spots' of aggregation in polypeptides. *BMC Bioinformatics* **8**, 65 (2007).

158. Maurer-Stroh, S. *et al.* Exploring the sequence determinants of amyloid structure using position-specific scoring matrices. *Nat. Methods* **7**, 237–42 (2010).

159. Fernandez-Escamilla, A.-M., Rousseau, F., Schymkowitz, J. & Serrano, L. Prediction of sequence-dependent and mutational effects on the aggregation of peptides and proteins. *Nat. Biotechnol.* **22**, 1302–6 (2004).

160. Kyte, J. & Doolittle, R. F. A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **157**, 105–132 (1982).

161. Rodriguez, J. A. *et al.* Familial amyotrophic lateral sclerosis-associated mutations decrease the thermal stability of distinctly metallated species of human copper/zinc

superoxide dismutase. *J. Biol. Chem.* **277**, 15932–15937 (2002).

162. Arnesano, F. *et al.* The unusually stable quaternary structure of human Cu,Zn-superoxide dismutase 1 is controlled by both metal occupancy and disulfide status. *J. Biol. Chem.* **279**, 47998–48003 (2004).

163. Borel, F. *et al.* Safe and effective superoxide dismutase 1 silencing using artificial microRNA in macaques. *Sci. Transl. Med.* **10**, eaau6414 (2018).

164. Kumar, A. T. N., Zhu, L., Christian, J. F., Demidov, A. A. & Champion, P. M. On the rate distribution analysis of kinetic data using the maximum entropy method: Applications to myoglobin relaxation on the nanosecond and femtosecond timescales. *J. Phys. Chem. B* **105**, 7847–7856 (2001).

165. Lindberg, M. J., Bystrom, R., Boknas, N., Andersen, P. M. & Oliveberg, M. Systematically perturbed folding patterns of amyotrophic lateral sclerosis (ALS)-associated SOD1 mutants. *Proc. Natl. Acad. Sci.* **102**, 9754–9759 (2005).

166. Vassall, K. A., Stathopulos, P. B., Rumfeldt, J. A. O., Lepock, J. R. & Meiering, E. M. Equilibrium thermodynamic analysis of amyotrophic lateral sclerosis-associated mutant Apo Cu,Zn superoxide dismutases. *Biochemistry* **45**, 7366–7379 (2006).

167. Rumfeldt, J. A. O., Stathopulos, P. B., Chakrabarrty, A., Lepock, J. R. & Meiering, E. M. Mechanism and thermodynamics of guanidinium chloride-induced denaturation of ALS-associated mutant Cu,Zn superoxide dismutases. *J. Mol. Biol.* **355**, 106–123 (2006).

168. Chattopadhyay, M. *et al.* Initiation and elongation in fibrillation of ALS-linked superoxide dismutase. *Proc. Natl. Acad. Sci.* **105**, 18663–18668 (2008).

169. Broom, H. R., Rumfeldt, J. A. O. & Meiering, E. M. Many roads lead to Rome? Multiple modes of Cu,Zn superoxide dismutase destabilization, misfolding and aggregation in amyotrophic lateral sclerosis. *Essays Biochem.* **56**, 2014 (2014).

170. Banci, L. *et al.* Solution structure of reduced monomeric Q133M2 copper, zinc superoxide dismutase (SOD). Why is SOD a dimeric enzyme? *Biochemistry* **37**, 11780–11791 (1998).

171. Srivastava, K. R., French, K. C., Tzul, F. O., Makhatadze, G. I. & Lapidus, L. J. Intramolecular diffusion controls aggregation of the PAPf39 peptide. *Biophys. Chem.* **216**, 37–43 (2016).

172. Lapidus, L. J. Understanding protein aggregation from the view of monomer

dynamics. *This J. is c R. Soc. Chem.* **9**, 29–35 (2013).

173. Bolen, D. W. & Rose, G. D. Structure and energetics of the hydrogen-bonded backbone in protein folding. *Annu. Rev. Biochem.* **77**, 339–362 (2008).

174. Timasheff, S. N. Protein hydration, thermodynamic binding, and preferential hydration. *Biochemistry* **41**, 13473–13482 (2002).

175. Pradeep, L. & Udgaonkar, J. B. Osmolytes induce structure in an early intermediate on the folding pathway of barstar. *J. Biol. Chem.* **279**, 40303–40313 (2004).

176. Pradeep, L. & Udgaonkar, J. B. Diffusional Barrier in the Unfolding of a Small Protein. *J. Mol. Biol.* **366**, 1016–1028 (2007).

177. Chrunyk, B. A. & Matthews, C. R. Role of diffusion in the folding of the alpha subunit of tryptophan synthase from Escherichia coli. *Biochemistry* **29**, 2149–2154 (1990).

178. Myers, J. K. K., Pace, C. N. N. & Scholtz, J. M. M. Denaturant m values and heat capacity changes: relation to changes in accessible surface areas of protein unfolding. *Protein Sci.* **4**, 2138–2148 (1995).

179. Schroer, M. A. *et al.* High-pressure SAXS study of folded and unfolded ensembles of proteins. *Biophys. J.* **99**, 3430–3437 (2010).

180. Jacob, M., Schindler, T., Balbach, J. & Schmid, F. X. Diffusion control in an elementary protein folding reaction. *Proc. Natl. Acad. Sci.* **94**, 5622–5627 (1997).

181. Ramos, C. H. I., Weisbuch, S. & Jamin, M. Diffusive motions control the folding and unfolding kinetics of the apomyoglobin pH 4 molten globule intermediate. *Biochemistry* **46**, 4379–4389 (2007).

182. Hagen, S. J. Solvent Viscosity and Friction in Protein Folding Dynamics. *Curr. Protein Pept. Sci.* **11**, 385–395 (2010).

183. Kramers, H. A. Brownian motion in a field of force and the diffusion model of chemical reactions. *Phys.* **7**, 284–304 (1940).

184. Ansari, A., Jones, C., Henry, E., Hofrichter, J. & Eaton, W. The role of solvent viscosity in the dynamics of protein conformational changes. *Science (80-.).* **256**, 1796–1798 (1992).

185. Kubelka, J., Hofrichter, J. & Eaton, W. A. The protein folding 'speed limit'. *Curr. Opin. Struct. Biol.* **14**, 76–88 (2004).

186. Jacob, M. & Schmid, F. X. Protein folding as a diffusional process. *Biochemistry* **38**, 13773–13779 (1999).

187. Klimov, D. K. & Thirumalai, D. Viscosity Dependence of the Folding Rates of Proteins. *Phys. Rev. Lett.* **79**, 317–320 (1997).

188. Zagrovic, B. & Pande, V. Solvent viscosity dependence of the folding rate of a small protein: Distributed computing study. *J. Comput. Chem.* **24**, 1432–1436 (2003).

189. Wensley, B. G. *et al.* Experimental evidence for a frustrated energy landscape in a three-helix-bundle protein family. *Nature* **463**, 685–688 (2010).

190. Borgia, A. *et al.* Localizing internal friction along the reaction coordinate of protein folding by combining ensemble and single-molecule fluorescence spectroscopy. *Nat. Commun.* **3**, 1195 (2012).

191. Wensley, B. G. *et al.* Separating the effects of internal friction and transition state energy to explain the slow, frustrated folding of spectrin domains. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 17795–9 (2012).

192. Wensley, B. G., Kwa, L. G., Shammas, S. L., Rogers, J. M. & Clarke, J. Protein folding: Adding a nucleus to guide helix docking reduces landscape roughness. *J. Mol. Biol.* **423**, 273–283 (2012).

193. Pabit, S. A., Roder, H. & Hagen, S. J. Internal friction controls the speed of protein folding from a compact configuration. *Biochemistry* **43**, 12532–12538 (2004).

194. Qiu, L. & Hagen, S. J. Internal friction in the ultrafast folding of the tryptophan cage. *Chem. Phys.* **307**, 243–249 (2004).

195. Soranno, A. *et al.* Quantifying internal friction in unfolded and intrinsically disordered proteins with single-molecule spectroscopy. *Proc. Natl. Acad. Sci.* **109**, 17800–17806 (2012).

196. Aznauryan, M. *et al.* Comprehensive structural and dynamical view of an unfolded protein from the combination of single-molecule FRET, NMR, and SAXS. *Proc. Natl. Acad. Sci. U. S. A.* **113**, E5389–E5398 (2016).

197. Soranno, A. *et al.* Integrated view of internal friction in unfolded proteins from single-molecule FRET, contact quenching, theory, and simulations. *Proc. Natl. Acad. Sci.* **114**, E1833–E1839 (2017).

198. Plaxco, K. W. & Baker, D. Limited internal friction in the rate-limiting step of a two-state protein folding reaction. *Proc. Natl. Acad. Sci.* **95**, 13591–13596 (1998).

199. Sato, S., Sayid, C. J., Raleigh, D. P. & Raleigh, D. P. The failure of simple empirical relationships to predict the viscosity of mixed aqueous solutions of guanidine hydrochloride and glucose has important implications for the study of protein folding. *Protein Sci.* **9**, 1601–1603 (2009).

200. Zheng, W., De Sancho, D. & Best, R. B. Modulation of Folding Internal Friction by Local and Global Barrier Heights. *J. Phys. Chem. Lett.* **7**, 1028–1034 (2016).

201. Zheng, W., De Sancho, D., Hoppe, T. & Best, R. B. Dependence of internal friction on folding mechanism. *J. Am. Chem. Soc.* **137**, 3283–3290 (2015).

202. Schulz, J. C. F., Miettinen, M. S. & Netz, R. R. Unfolding and Folding Internal Friction of β-Hairpins Is Smaller than That of α-Helices. *J. Phys. Chem. B* **119**, 4565–4574 (2015).

203. Qiu, L. & Hagen, S. J. A Limiting Speed for Protein Folding at Low Solvent Viscosity. *J. Am. Chem. Soc.* **126**, 3398–3399 (2004).

204. Cellmer, T., Henry, E. R., Hofrichter, J. & Eaton, W. A. Measuring internal friction of an ultrafast-folding protein. *Proc. Natl. Acad. Sci.* **105**, 18320–18325 (2008).

205. Banci, L., Bertini, I., Cramaro, F., Del Conte, R. & Viezzoli, M. S. Solution structure of apo Cu,Zn superoxide dismutase: Role of metal ions in protein folding. *Biochemistry* **42**, 9543–9553 (2003).

206. Cohen, N. R., Zitzewitz, J. A., Bilsel, O. & Matthews, C. R. Nonnative structure in a peptide model of the unfolded state of SOD1: Implications for ALS-linked aggregation. *J. Biol. Chem.* jbc.RA119.008765 (2019). doi:10.1074/jbc.RA119.008765

207. Nordlund, A. & Oliveberg, M. Folding of Cu/Zn superoxide dismutase suggests structural hotspots for gain of neurotoxic function in ALS: parallels to precursors in amyloid disease. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 10218–10223 (2006).

208. Basak, S. *et al.* Networks of electrostatic and hydrophobic interactions modulate the complex folding free energy surface of a designed βα protein. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 6806–6811 (2019).

209. Lee, Y.-K., Brewer, J. W., Hellman, R. & Hendershot, L. M. BiP and Immunoglobulin Light Chain Cooperate to Control the Folding of Heavy Chain and Ensure the Fidelity of Immunoglobulin Assembly. *Mol. Biol. Cell* **10**, 2209–2219 (1999).

210. Feige, M. J. *et al.* An Unfolded CH1 Domain Controls the Assembly and Secretion of IgG Antibodies. *Mol. Cell* **34**, 569–579 (2009).

211. Hellman, R., Vanhove, M., Lejeune, A., Stevens, F. J. & Hendershot, L. M. The In Vivo Association of BiP with Newly Synthesized Proteins Is Dependent on the Rate and Stability of Folding and Not Simply on the Presence of Sequences That Can Bind to BiP. *J. Cell Biol.* **144**, 21–30 (1999).

212. Kaloff, C. R. & Haas, I. G. Coordination of immunoglobulin chain folding and immunoglobulin chain assembly is essential for the formation of functional IgG. *Immunity* **2**, 629–637 (1995).

213. Davis, D. P., Khurana, R., Meredith, S., Stevens, F. J. & Argon, Y. Mapping the major interaction between binding protein and Ig light chains to sites within the variable domain. *J. Immunol.* **163**, 3842–50 (1999).

214. Liu, H., Jeong, J., Kao, Y.-H. & Zhang, Y. T. Characterization of free thiol variants of an IgG1 by reversed phase ultra high pressure liquid chromatography coupled with mass spectrometry. *J. Pharm. Biomed. Anal.* **109**, 142–149 (2015).

215. Ouellette, D. *et al.* Studies in serum support rapid formation of disulfide bond between unpaired cysteine residues in the VH domain of an immunoglobulin G1 molecule. *Anal. Biochem.* **397**, 37–47 (2010).

216. Chumsae, C., Gaza-Bulseco, G. & Liu, H. Identification and localization of unpaired cysteine residues in monoclonal antibodies by fluorescence labeling and mass spectrometry. *Anal. Chem.* (2009). doi:10.1021/ac900815z

217. Xiang, T., Chumsae, C. & Liu, H. Localization and quantitation of free sulfhydryl in recombinant monoclonal antibodies by differential labeling with 12C and 13C iodoacetic acid and LC-MS analysis. *Anal. Chem.* **81**, 8101–8108 (2009).

218. Woody, R. W. Aromatic side-chain contributions to the far ultraviolet circular dichroism of peptides and proteins. *Biopolymers* **17**, 1451–1467 (1978).

219. Ghose, A. C. & Jirgensons, B. Circular dichroism studies on the variable and constant halves of kappa-type Bence-Jones proteins. *Biochim. Biophys. Acta - Protein Struct.* **251**, 14–20 (1971).

220. Hagihara, Y., Mine, S. & Uegaki, K. Stabilization of an immunoglobulin fold domain by an engineered disulfide bond at the buried hydrophobic region. *J. Biol. Chem.* **282**, 36489–95 (2007).

221. Jennings, P. A., Finn, B. E., Jones, B. E. & Matthews, C. R. A Reexamination of

the Folding Mechanism of Dihydrofolate Reductase from Escherichia coli: Verification and Refinement of a Four-Channel Model. *Biochemistry* **32**, 3783–3789 (1993).

222. Robinson, P. J., Pringle, M. A., Woolhead, C. A. & Bulleid, N. J. Folding of a single domain protein entering the endoplasmic reticulum precedes disulfide formation. *J. Biol. Chem.* **292**, 6978–6986 (2017).

223. Kosuri, P. *et al.* Protein folding drives disulfide formation. *Cell* **151**, 794–806 (2012).

224. Koterba, K. L., Borgschulte, T. & Laird, M. W. Thioredoxin 1 is responsible for antibody disulfide reduction in CHO cell culture. *J. Biotechnol.* **157**, 261–267 (2012).

225. Srinivasan, N., Sowdhamini, R., Ramakrishnan, C. & Balaram, P. Conformations of disulfide bridges in proteins. *Int. J. Pept. Protein Res.* **36**, 147–155 (1990).

226. Katz, B. A. & Kossiakoff, A. The crystallographically determined structures of atypical strained disulfides engineered into subtilisin. *J. Biol. Chem.* **261**, 15480–15485 (1986).

227. Wong, J. W. H. & Hogg, P. J. Allosteric disulfide bonds. *Protein Rev.* **14**, 151–182 (2011).

228. Lundström, J. & Holmgren, A. Determination of the Reduction—Oxidation Potential of the Thioredoxin-like Domains of Protein Disulfide-Isomerase from the Equilibrium with Glutathione and Thioredoxin. *Biochemistry* **32**, 6649–6655 (1993).

229. Pearlman, D. A. *et al.* AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules. *Comput. Phys. Commun.* **91**, 1–41 (1995).

230. Haworth, N. L., Gready, J. E., George, R. A. & Wouters, M. A. Evaluating the stability of disulfide bridges in proteins: A torsional potential energy surface for diethyl disulfide. *Mol. Simul.* **33**, 475–485 (2007).

231. Chumsae, C., Gaza-Bulseco, G. & Liu, H. Identification and localization of unpaired cysteine residues in monoclonal antibodies by fluorescence labeling and mass spectrometry. *Anal. Chem.* **81**, 6449–6457 (2009).

232. Cheng, Y. *et al.* Domain-specific free thiol variant characterization of an IgG1 by

reversed-phase high-performance liquid chromatography mass spectrometry. *Anal. Biochem.* **519**, 8–14 (2017).

233. Pijning, A. E., Chiu, J., Yeo, R. X., Wong, J. W. H. & Hogg, P. J. Identification of allosteric disulfides from labile bonds in X-ray structures. *R. Soc. Open Sci.* **5**, (2018).

234. Nasica-Labouze, J. *et al.* Amyloid β Protein and Alzheimer's Disease: When Computer Simulations Complement Experimental Studies. *Chemical Reviews* **115**, 3518–3563 (2015).

235. Jamal, S., Kumari, A., Singh, A., Goyal, S. & Grover, A. Conformational ensembles of α-synuclein derived peptide with different osmolytes from temperature replica exchange sampling. *Front. Neurosci.* **11**, 684 (2017).

236. Moore, S. J., Sonar, K., Bharadwaj, P., Deplazes, E. & Mancera, R. L. Characterisation of the structure and oligomerisation of islet amyloid polypeptides (IAPP): A review of molecular dynamics simulation studies. *Molecules* **23**, 1–37 (2018).

237. Muyldermans, S. Nanobodies: Natural Single-Domain Antibodies. *Annu. Rev. Biochem.* **82**, 775–797 (2013).

238. Salvador, J. P., Vilaplana, L. & Marco, M. P. Nanobody: outstanding features for diagnostic and therapeutic applications. *Analytical and Bioanalytical Chemistry* **411**, 1703–1713 (2019).

239. Kunz, P. *et al.* Exploiting sequence and stability information for directing nanobody stability engineering. *Biochim. Biophys. Acta - Gen. Subj.* **1861**, 2196–2205 (2017).

240. Goldman, E. R., Liu, J. L., Zabetakis, D. & Anderson, G. P. Enhancing stability of camelid and shark single domain antibodies: An overview. *Frontiers in Immunology* **8**, 1–11 (2017).