BIOPHYSICAL MECHANISMS OF PROTEIN AGGREGATION

Sagar Khare

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Biochemistry and Biophysics (School of Medicine).

Chapel Hill
2005

Approved By

Advisor: Nikolay V. Dokholyan, PhD

Committee Chair: Sharon L. Campbell, PhD

Reader: Michael Caplow, PhD

Reader: Brian A. Kuhlman, PhD

Reader: Max L. Berkowitz, PhD

ABSTRACT

SAGAR KHARE: Biophysical mechanisms of protein aggregation

(Under the direction of Nikolay V. Dokholyan, PhD)

Protein aggregation related toxicity is implicated in a variety of neurodegenerative diseases including Alzheimer's, Huntington's, prion and Amyotrophic Lateral Sclerosis (ALS). The proteins or peptides known to aggregate in disease are unrelated in their amino acid sequence and native structure but form structurally similar aggregates – amyloids. Studies outlined in this dissertation were aimed at uncovering the underlying biophysical mechanisms of amyloid formation, and lay the groundwork to develop rational strategies to combat neurodegenerative diseases.

More than 100 point mutations in the homodimeric metalloenzyme Cu, Zn superoxide (SOD1) dismutase are involved in a genetically inherited familial form of ALS (FALS). We have discovered a mechanism of *in vitro* SOD1 aggregation in which the native SOD1 dimer dissociates, metals are lost from the monomers and the resulting apo-monomers oligomerize in a rate-limiting step. Further, we have computationally estimated that a majority of FALS-associated point mutants in SOD1 (70 out of the 75 studied) decrease dimer stability and/or increase dimer dissociation propensity. Thus, we have proposed that the underlying biophysical basis of FALS-linked SOD1 aggregation is the mutation-induced increase in the propensity to form apo-monomers.

To uncover the molecular determinants of SOD1 apo-monomer oligomerization, the rate-limiting step in aggregation, we have developed two complementary *in silico* approaches: (a) we have identified sequence fragments of SOD1 that have a high self-association propensity, and (b) we have performed molecular dynamics simulations of model SOD1 monomer and dimer folding and misfolding. In both cases, we have identified key residue-residue interactions in SOD1 responsible for maintaining fidelity to its native state. We have proposed that the disruption of one or more of these key interactions ("hot spots") is implicated in non-native oligomerization.

To understand the effect of FALS mutations on the key interactions involved in maintaining native-state fidelity, we have studied the nanosecond dynamics of wild type SOD1 and 3 FALS-associated mutant apo-dimers and apo-monomers. We found that in wild type SOD1 the motions of the dimer interface are mechanically coupled to the motions of the structurally distal metal-coordinating loops of both monomeric subunits. We further found that the strain induced in the protein by dimer dissociation, point mutations, or by exposure to high temperature is transmitted to a specific hairpin in the protein, previously found to be implicated in maintaining fold fidelity. The altered dynamics of mutant SOD1 dimers and monomers provides structural insights into the flexibility required for oligomerization.

Collectively, findings in this dissertation have enhanced our understanding of the complex mechanisms of protein aggregation. Mechanisms established and structural insights obtained herein may facilitate rational design of small molecules to prevent protein aggregation, hence provide a therapeutic intervention strategy in neurodegenerative diseases.

DEDICATION

To my parents,

Aai and Baba.

ACKNOWLEDGEMENTS

I would like to first thank my advisor, Dr. Nikolay Dokholyan, whose enthusiasm, guidance, support, patience, insightful ideas, and friendship have allowed me to carry out this work. To Professor Michael Caplow, I owe a great deal of thanks for his sharp insights, designing some elegant experiments, and for patiently teaching me the rigors of lab work. I would also like to thank the other members of my committee, Drs. Sharon Campbell, Brian Kuhlman, and Max Berkowitz for their helpful suggestions and feedback. To Drs. Jan Hermans, Barry Lentz, Andrew Lee, Ashutosh Tripathy, Alexander Makhov, John Sheehan, Glenn Butterfoss, Hao Hu, Jainab Khatun, Lanette Fee and Brian Dominy, I would like to express my appreciation for their help and time.

I have been fortunate to work with many outstanding people during my graduate years. I would like to thank Dr. Feng Ding for the many wonderful discussions and collaborations. To Kyle Wilcox, my thanks for being a great neighbor both in the lab and out of it. Thanks are also due to Yiwen Chen for the many late-night discussions on science, and for the jovial back-n-forth that kept the lab atmosphere lively. My friends Murali, Arka, and Pooja, were responsible for me not missing India. My appreciation is also due to other friends in Chapel Hill – too many to list here – for making my life out of lab exciting. Last, but not the least, a heartfelt thanks to my family, immediate and extended, in India and the US, for all that they have done over the years to help me pursue my dreams.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ALS	Amyotrophic lateral sclerosis
AUC	Analytical ultracentrifugation
CCS	Copper chaperone for SOD
DMD	Discrete molecular dynamics
EDTA	Ethylene diamine tetraacetic acid
FALS	Familial amyotrophic lateral sclerosis
MD	Molecular dynamics
PDB	Protein databank
RMSD	Root mean square deviation
SALS	Sporadic amyotrophic lateral sclerosis
SEC	Size exclusion chromatography
SOD1	Cu, Zn superoxide dismutase
SPC	Simple point charge
SPR	Surface plasmon resonance
T_F	Folding transition temperature
T_m	Melting temperature

CHAPTER 1

INTRODUCTION

Protein aggregation-related toxicity is implicated in a plethora of neurodegenerative diseases including Alzheimer's, Parkinson's, prion, and motor neuron diseases[1-5]. The proteins found to aggregate in diseases are unrelated in their native structures and primary sequences, but form similar fibrils with characteristic cross-β sheet morphologies (amyloid[6]) when aggregated. The aggregates can be intra- or extra-cellular; however aggregation and the induced cytotoxicity are tissue-specific. A small number of the cases of the diseases, called familial cases, are linked to genetically inherited mutations in the respective aggregating proteins, in contrast with most sporadic cases.

1.1 Aggregation and cytotoxicity in disease.

The existence of a causative link between protein aggregation and cytotoxicity in neurodegeneration remains controversial. While aggregation is usually co-incident with cell death, it is not clear if aggregation is the cause of cytotoxicity or an effect – it may represent a protective response of the cell which depletes the pool of the cytotoxic soluble protein. Evidence has accumulated that soluble oligomeric species, rather than the insoluble fibrils, may cause cytotoxicity. For example, in various animal models[7-11], it was observed that the triggering of cell-death mechanisms precedes the appearance of insoluble plaque[12,13]. The injection of soluble assemblies into cultured neuronal cells causes substantial neuronal dysfunction before aggregation[14]. In models of Huntington's disease, it was found that soluble oligomers accumulate in the nucleus and inhibit the function of key transcription factors[15]. Inhibition of oligomerization by the azo-dye Congo red, or by overexpression of the Hsp70/Hsp40 chaperone system, exerts marked protective effects *in vivo* and *in vitro*[16,17,14]. Thus, toxicity may arise by aberrant interactions with and/or sequestration of key

cellular components, such as transcription factors and the proteasomal machinery[18]. Another mechanism of oligomers-induced cytotoxicity is the permeabilization of membranes, which disrupts ion-homeostasis and results in cell death[19]. While membrane permeabilization is known to occur by the spontaneous formation of oligomeric ion-channels in mixtures of disease-associated peptides and artificially reconstituted membrane bilayers *in vitro*[20], evidence for ion-channels *in vivo*, in studies where ion homeostasis was found to be disrupted upon aggregation, remains controversial[21,22]. Finally, in some instances, cell death can be ascribed to mechanical disruption of the cell by the large quantities of insoluble material[23]. Thus, several scenarios linking aggregation and cytotoxicity have been proposed, and cytotoxicity may well be multi-factorial involving simultaneously more than one mechanism of cell death[24]. It is clear, however, that aggregation is an important event during neurodegeneration, and given the proposed direct involvement of oligomers in cytotoxicity, the mechanism of aggregation provides insights into the molecular basis of disease.

It has been shown that many proteins unrelated to any known disease can form disease-like ß-sheet amyloid aggregates *in vitro*, indicating that amyloid-formation is a common property of polypeptides chains[25]. Amyloid fibrils of different peptides share common structural features: they bind to dyes such as Congo Red, and Thioflavin T, and display a characteristic X-ray diffraction pattern called the cross-β pattern[25]. An antibody raised against oligomers of one peptide cross-reacts with oligomers of several other proteins, suggesting that the oligomers, like the amyloid fibrils, may also share a common structural epitope[26]. A better understanding of the underlying biophysical principles of aggregation may lead to therapeutic intervention aimed at inhibiting aggregation and cytotoxicity. For example, small-molecule inhibitors of transthyretin aggregation that were designed based on

the knowledge of its in vitro aggregation mechanism, are under clinical trials for treating transthyretin aggregation-associated systemic amyloidosis[27].

1.2 Amyotrophic Lateral Sclerosis and mutant Cu, Zn superoxide dismutase.

Amyotrophic lateral sclerosis (ALS) is a fatal neurodegenerative disease that targets the motor neurons in the spinal cord, brain stem, and cortex[24,28]. It is typically a disease of mid-life, with an average age-of-onset of 55±15, and is the most common motor neuron disease[24]. The symptoms start first with weakness in peripheral limbs and death generally occurs because of respiratory failure within 2-5 years of the appearance of the first symptoms. Most (90-95%) of the disease cases have no known cause and are called sporadic ALS (SALS), in the remaining (5-10%) cases there is a family history associated with the disease and these cases are called familial ALS (FALS). Of the familial cases, 20-25% of the cases are associated with inheritable mutations in a gene on chromosome 21 that codes for the ubiquitous cytoplasmic enzyme Cu, Zn superoxide dismutase (SOD1)[29]. More than 100 point mutations in SOD1 have been identified as linked to FALS[30]. The FALS mutations are scattered throughout the three-dimensional structure and primary sequence of the protein and do not show any obvious chemical patterns in their distribution.

Proteinaceous aggregates rich in mutant SOD1 have been found in tissues from ALS patients, mutant SOD1 transgenic mice, and in cell culture model systems, suggesting that that SOD1-associated FALS is a protein conformational disorder[31]. Mutant SOD1 also forms amyloid-like aggregates *in vitro* more readily then the wild type[32,33]. The inclusions in SOD1-linked FALS contain neurofilament proteins, ubiquitin, and a variety of other components in addition to SOD1[24], but it is not known if copper, zinc, or any other metal

ions are present in the inclusions or are involved in their formation *in vivo* (it is known, though, that SOD1 aggregates formed *in vitro* are not active[32]). It is also unknown if the SOD1 polypeptide is fragmented or covalently modified in the aggregates *in vivo*. As in the case of other aggregation-associated diseases, it is not clear of the final insoluble aggregates themselves are toxic or represent a dysfunction of the protective response of the cell where the toxic misfolded SOD1 protein cannot be degraded sufficiently by the proteasomal machinery. Soluble oligomerized species of SOD1, which may be toxicity-related, are found in the spinal cords of mice expressing mutant SOD1 well before disease onset or the appearance of the much larger microscopically visible fibrils or inclusions[34]. Nevertheless, it is well-accepted that mutant SOD1 proteins represent a toxic gain-of-function that facilitates misfolding and/or oligomerization in the disease[24,29].

Mitochondrial dysfunction coupled with excitotoxicity has been postulated as a mechanism of tissue-specific toxicity in ALS[35]. A fraction of the SOD1 is transported to the mitochondria[36], and mitochondrial abnormalities have been identified in transgenic mouse models of FALS expressing mutant SOD1[37]. It was found that these abnormalities begin prior to the clinical and pathological onset of the disease, suggesting that mitochondrial dysfunction may be causally involved in pathogenesis[38]. Although the mechanisms by which mutant SOD1 damages mitochondria remain to be fully understood, mutant SOD1 is known to form aberrant aggregates and protein interactions in mitochondria[39]. It was found that misfolded mutant SOD1, but not wild type, was selectively associated with mitochondrial membranes in motor neurons[40]. Motor neurons are known to be particularly sensitive to the disruption of mitochondria and Ca^{2+}-dependent signaling cascades[41]. Thus, in one scenario of cytotoxicity, misfolded/oligomerized mutant SOD1-induces disruption of mitochondrial and

Ca^{2+}-related cascades (resulting in excitotoxicity) and thereby, tissue-specific neurodegeneration.

There is evidence that the metal-content of SOD1, in which each subunit contains one copper and one zinc ion, may be implicated in the mechanism of toxicity[33]. The copper ion is the catalytically active constituent, and zinc is believed to stabilize the protein. Since metallation has a large stabilizing effect on the protein *in vitro*[42], it has been argued that the metallation-state of the protein is important for its misfolding and oligomerization *in vivo* – it has been postulated that the metal-free forms of the protein may be the oligomerization precursors[33]. Furthermore, transport of SOD1 into mitochondria requires the loss of metals from the protein, suggesting that the metallation-state of the protein may vary between compartments in the cell. It is also established that some portion of the SOD1 protein is normally present in cells in a copper-deficient state. In human lymphoblasts, the pool of copper-free protein was estimated to be about 35% of the total SOD1 protein[33], but there are no estimates of the population-size of apo-SOD1 available for motor neurons. The case for the involvement of the apo-state in FALS is strengthened by two observations: (i) FALS associated mutations were able to provoke the disease even in the absence of the metal loading chaperone CCS indicating that a copper-free precursor state of SOD1 is responsible for neurotoxicity[43]; (ii) several FALS mutants, in the holo-enzyme state, have wild-type-like stability and activity both *in vitro* and *in vivo* but are markedly destabilized in apo-forms[33]. The copper chaperone for SOD1 (CCS) is known to be responsible for the delivery of copper to the protein, but the details of zinc delivery and homeostasis are unknown. However, in rats fed with a copper-deficient diet, there was a significant amount of protein that was both copper- and zinc-free[44], suggesting that there may be coupling between the delivery and

homeostasis of both metals. A possible dysfunction of metal homeostasis may lead to a higher proportion of misfolding- and aggregation-prone apo-state of the protein.

Thus, the picture that emerges from studies of FALS suggests that a toxic gain-of-function that is related to the misfolding and/or oligomerization of mutant SOD1 is causally implicated in the disease. The molecular mechanism of misfolding and aggregation is not well-understood.

1.3 Statement of the problem.

In view of the important biological role of protein misfolding and/or aggregation in neurodegenerative diseases in general, and mutant SOD1 misfolding and/or aggregation in FALS in particular, this dissertation attempts to answer the following questions: What is(are) the molecular mechanism(s) of SOD1 aggregation? What, if any, key sequence and structural elements of the SOD1 dimer determine the aggregation pathway of mutant SOD1? Why do more than 100 structurally and chemically diverse genotype changes (point mutations) in SOD1 all result in an identical aggregate phenotype?

1.4 Structural frameworks for the mechanism of polypeptide aggregation.

It is well known that large conformational transitions occur during the aggregation of proteins from their native states into amyloid fibrils, suggesting that partial unfolding and/or misfolding [45-51] are necessary for aggregation. The thermodynamic stability of a protein determines the propensity of the protein to adopt native-like structures. Thus, the formation of stable protein aggregates is linked to changes in the thermodynamic stability of

aggregating proteins, and conditions which promote aggregation are expected to destabilize the native-state. Such conditions include mutations in the protein [52-55], changes in the local cell environment that affect processing of the polypeptide chain during localization/transport into organelles, post-translational modifications, interactions with reactive metabolite species, and aberrant interactions with the proteolytic or chaperone machinery [56].

While it is long established that destabilization or (at least partial) unfolding is necessary for aggregation, the driving forces for formation of a specific aggregate topology, the cross-β amyloid fibril, are less well understood. There are three main proposed driving forces for amyloid formation from partially unfolded or misfolded states of polypeptides:

Edge-strand hydrogen bonding. In the non-diseased cell, a large fraction of proteins folds to its native state, avoiding off-pathway processes that might lead to aggregation. It has been postulated that in the course of evolution, nature has selected sequences which not only fold into stable native states but also avoid non-native aggregation into amyloid fibrils. This is the so-called *principle of negative design*[57], which offers structural clues into protein aggregation. Upon a systematic analysis of the database of protein structures, it was postulated that all β-sheet proteins can form edge-to-edge aggregates unless they employ some 'blocking' strategies to prevent aggregation. There are two such 'blocking' strategies: 1. Minimization of unsatisfied hydrogen bonds (H-bonds) 2. Inward-pointing charged residues that block aggregation[58].

In the folded states of naturally occurring proteins, the number of unsatisfied H-bonds is minimized, and the presence of a large number of unsatisfied H-bonds promotes aggregation. It was shown that for pathological monomeric conformations of the prion proteins from mouse and Syrian hamster the proportion of unsatisfied H-bonds (14%) is

significantly larger than the average number of unsatisfied H-bonds in naturally occurring proteins (6%)[59]. Furthermore, the β-helix, which is a putative topology of amyloid fibrils, does not contain unsatisfied H-bond donors/acceptors. Thus, the formation of H-bonds may be the driving force for protein aggregation. In this scenario, formation of protein aggregates is governed by a competition between specific amino acid interactions within proteins, and non-specific hydrogen bond formation between the unprotected edge-strands of different proteins[54,60].

The template hypothesis of aggregation[61,62] states that preformed oligomers or aggregation nuclei provide edge-strands, and these interaction centers cause the conversion of a normally soluble protein into an aggregation-prone conformation. The docking of the aggregation-prone conformation to the pre-formed nucleus results in growth of the oligomer. Consistent with the template hypothesis, computational studies have shown that pre-formed or spontaneously-formed templates with hydrogen-bond donors/acceptors can cause the inter-conversion of α-helical peptides to β-strands that propagate the aggregate[63].

Domain-swapping. In this scenario, aggregates are formed from partially unfolded protein chains by a mechanism called "runaway domain swapping". Three-dimensional domain swapping is an event by which a monomer exchanges part of its chain with other identical monomers to form an oligomer in which each subunit has a similar structure to the monomer, albeit the subunit is now made of more than one chains. Domain swapping, initially proposed as a mechanism of functional regulation, has also been proposed to lead to misfolding and aggregation[64-66]. Although there is little direct evidence for domain swapping as a mechanism for aggregation and amyloid formation, several experimental (Ref.[65] and references therein) and computational[67-69] studies support the role of domain swapping in aggregation. For

example, a correlation between domain swapping propensity of the protein p13suc1 was found to be correlated with its rate of aggregation[70]. Eisenberg and coworkers have designed both domain-swapped dimers and high-order oligomers from the same three-helix bundle structural motif but with different topologies[64,71,72]. Furthermore, domain-swapped forms of both the human prion protein and the amyloidogenic human cystatin C[67,73] have been crystallized.

Domain-swapping has been most clearly implicated in the aggregation of polyglutamine (polyQ)-containing model systems. The aggregation of polyQ-containing proteins is implicated in at least nine neurodegenerative diseases, and model systems in which polyQ stretches are inserted into small, well-characterized, non-disease-associated proteins have been used to understand the structural basis of aggregation. It has been found that at small repeat lengths of polyQ, the destabilization induced by their insertion can be overcome by domain-swapped dimer formation by the chimeric protein[74]. For longer repeat lengths, the presence of polyQ stretches can destabilize the protein sufficiently to form higher-order oligomers and fibrils. A striking example of this phenomenon was recently demonstrated in the model system ribonuclease A by Eisenberg and co-workers[75]. They had previously found that ribonuclease A chains form domain-swapped dimers and trimers by exchanging identical structural elements on either side of a hinge region[64,72]. Upon a Q_{10} insertion in this hinge loop of ribonuclease A, polyQ self-association induces a "runaway" domain swap i.e. higher-order oligomerization and amyloid fibril formation.

Computational studies[67-69] using simplified native-structure based Gō-models[76,77] have shown that the monomeric protein topology alone is sufficient for predicting how a protein will form domain-swapped complexes, including higher-order oligomers. In

computational domain-swapping studies of the SH3 domain[69,78], two types of topologies have been detected: "closed" domain-swapped dimers which are observed in X-ray crystal structures[79], and more "open" oligomers which can be propagated to form fibrils by a "runaway domain swap". Structural features of the computationally obtained fibrils agree with the X-ray diffraction pattern of amyloid fibrils obtained in experiments[80].

Thus, in the domain-swapping framework, the strong bias towards the native-state of the protein leads to "closed" domain swapped dimers or trimers upon small destabilization. As the degree of destabilization increases, lower-order oligomers can no longer stabilize the protein chain in a native-like conformation. Instead, a "runaway" domain-swap occurs, resulting in the formation of amyloid fibrils, in which elements of native structure may still be retained.

Self-association of amyloidogenic fragments. In this scenario, unfolding may result in the exposure of amyloidogenic sequence fragments which, in turn, self-associate to induce oligomerization. This scenario is supported by the findings that several small (>5 residues long) sequence fragments of many aggregating proteins themselves form amyloid fibrils *in vitro*[81]. In some cases, these amyloidogenic sequence fragments have been found in the aggregated state to be arranged as parallel β-strands in a sheet in which the amino acid sequence is in exact register[82]. Thus, it has been argued that unfolding makes the self-association of amyloidogenic "hotspots" possible, and the self-association of amyloidogenic fragments nucleates the aggregation of the entire chain.

Mutational analyses of fragment aggregation *in vitro* and *in silico* suggest that hydrophobicity, net charge and β-sheet propensity modulate fragment self-association propensity[83]. Algorithms have been developed to predict the location of amyloidogenic

fragments based on polypeptide sequence alone, and it has been found that in well folded globular protein sequences amyloidogenic fragments are surrounded by residues that have a very low aggregation propensity ("amyloid-breakers")[84]. Analysis of protein structures also suggests that natural selection has led to amyloidogenic sequence fragments being protected in the native states of protein structures found in nature[57,85]. Therefore, the ability of amyloidogenic sequences to induce aggregation is modulated by the global stability and the structure of proteins. Protein aggregation propensity is then, the interplay between the stability of the native structure, which prevents protein aggregation, and the self-association of amyloidogenic sequence fragments from different polypeptide chains into in-register structures, which promotes protein aggregation. Consequently, mutations associated with familial forms of neurodegenerative diseases may promote aggregation by either destabilizing the native state and/or increasing the self association propensities of exposed sequence fragments under destabilizing conditions.

It should be noted that the three frameworks for aggregation outlined above are not mutually exclusive – for example, as we will see in Chapter 3, it is possible self-association-prone regions of the protein are in regions that also have a high propensity to domain-swap, i.e. in this case, domain swapping and self-association of amyloidogenic fragments occurs simultaneously during the oligomerization process, and this may lead to the formation of stabilizing hydrogen bonding networks in the oligomers. Further, it is also possible that the underlying driving forces are protein-dependent – i.e. a protein with a high content of hinge-regions aggregates *via* a runaway domain-swap, whereas a protein with a high content of amyloidogenic fragments aggregates primarily by the self-association of these fragments.

These different scenarios provide statistical mechanical frameworks to understand the molecular mechanism of polypeptide aggregation.

1.5 Biophysical properties of mutant and wild type SOD1

Stability. SOD1 is an unusually stable protein, both *in vitro* and *in vivo*. The FALS mutations that occur throughout the SOD1 polypeptide have been shown to decrease the half-life of the protein *in vivo*. Mouse wild-type SOD1, expressed in human kidney cells, had a half-life of 100 hours whereas mouse A4V and I113T had half-lives of 14 and 45 hours, respectively[86]. Human SOD1 overexpressed in COS1 cells was extremely stable with a half-life of 30 hours, whereas FALS mutants (I113T, G93C, G37R, G41D, G85R, and A4V) exhibited decreased half-lives from 20 hours for I113T to 7.5 hours for G85R and A4V[87]. A4T also exhibited a significantly reduced half-life of 18 hours when expressed in COS7 cells compared to human wild-type SOD1 which had a half-life of 25 hours[88].

The SOD1 dimer is also exceptionally stable *in vitro*. It retains its native 32 KDa molecular weight in 8 M urea for 72 hours at 25^{o}C and is active in 4% SDS[89]. The conformational stability of mutant forms of SOD1 has been analyzed using differential scanning calorimetry by measuring the irreversible unfolding of the distinctly metal-bound species of SOD1: a study by Rodriguez et al. revealed that some FALS mutants, isolated as is from an insect-baculovirus expression system, are destabilized relative to the wild type[42]. The mutations they studied are scattered over the three-dimensional structure of SOD1: in the active site region, the dimer interface, one pole of the β-barrel, and associated loops. Thus, the mutant-induced destabilization is not confined to particular structural regions of the proteins. Chemical denaturation studies of SOD1 carried out by Cardoso et al.[114] show that

the most common mutant SOD1 found in FALS patients (A4V) unfolds at a guanidine-hydrochloride concentration 1M lower than the wild type enzyme.

The metal-content of SOD1 is a crucial determinant of its stability. It was shown that for five FALS mutants – A4V and C6F (dimer disrupting), G93A and G93C (decreasing backbone flexibility) and D90A (surface mutant) – the apo-state of the mutants is markedly destabilized compared to the wild-type while the holo-state is not significantly affected (except for the dimer disrupting mutations)[33] suggesting that for some classes of mutants, the apo-state of the enzyme may be the species contributing to FALS linked pathology. Further support for the involvement of the apo-state in FALS was found in a weak correlation between the *in vitro* stabilities of 15 mutant apo-dimers and average survival times of patients with these mutations[90]. However, the stability of the apo-state may not be the "common denominator" responsible for FALS, as postulated by Lindeberg et al[90] based on their folding studies on fifteen mutants. It was shown by Valentine and co-workers that other mutations in the metal-binding regions of SOD1, distinct from the ones considered by Lindeberg et al., have similar or, in some cases, higher stabilities compared to the wild type apo-dimer, but cause FALS nevertheless[91]. Therefore, although decreased stability of FALS may be an important factor in disease, it is not the single underlying factor. These finding suggest that events downstream of unfolding affect the aggregation-propensity of mutants or the toxicity of oligomers, at least in some cases. A more complete understanding of the aggregation pathway, on the lines of the one developed in Chapter 2, is required to elucidate the common biophysical denominator of mutant SOD1 in causing FALS.

Monomer-dimer equilibria. SOD1 is a homodimer, with a dissociation constant, $K_d \sim 10^{-11}$ M, making it a very stable dimer[92]. FALS-associated mutations are known to affect the

monomer-dimer equilibrium (see Chapter 3), and as shown in Chapter 2, dimer dissociation is required for the aggregation of protein. A range of parameters affects the dimer dissociation properties of SOD1 *in vitro*[93], most importantly, the metal content of the protein and the presence of an intact disulfide between Cys57 and Cys146. It is therefore possible that mutant SOD1 may affect the monomer-dimer equilibrium either by structurally destabilizing the disulfide bond or by diminishing the ability for the protein to bind metals.

Crystal structures. Several crystal structures of mutant and wild type SOD1 have been solved, and in all known crystal structures mutant SOD1 adopts the same fold as wild type, the differences between them are limited to subtle changes in backbone and side-chain packing[29]. This observation is not surprising given that most mutants are known to be active, and are therefore expected to maintain the active-site geometry required for activity. However, subtle packing differences between the wild type and mutant, and, as we will see in Chapter 6, the dynamic propagation of these packing defects may be responsible for the enhanced aggregation propensities of the mutants. A comparison of the crystal structures shows that the A4V mutation causes significant shifts in side-chain positions in the vicinity of the mutation site causing defects in local packing[94]. In addition, a crystal structure of the G37R mutant shows higher atomic displacement parameters (*B*-factors) for the side-chains compared to the wild type, but the backbone conformation is not significantly different, thus indicating greater molecular flexibility in some portions of the structure possibly explaining the observations that mutant SOD1s exhibit accelerated turn-over *in vivo* or increased susceptibility to proteolytic digestion[95]. Thus, one scenario for SOD1 aggregation evident from an examination of the crystal structures of mutant SOD1 is the decreased conformational stability of mutants by a so-called "framework destabilization".

Protein dynamics. Mutations in the SOD1 polypeptide may destabilize the protein by altering the protein dynamics. It was found that the apo-state of the A4V FALS mutant had much higher rates of H/D exchange compared to the wild-type SOD1[91]. The higher exchange rate of the mutant persisted even at low temperatures (4^{o}C), indicating that the mutant has significantly higher flexibility. Using Protein NMR to probe motions on the *ps-ns* timescale, it was also found that holo-G93A shows a general increase in backbone mobility which is localized to two specific β-strands of the SOD1 fold important for maintaining fidelity to its native fold[96] (Chapter 5). Moreover, as we will see in Chapter 6, the motion of the two β-strands may correspond to a transient opening of the β-barrel fold of SOD1, and is observed in molecular dynamics simulations of apo-SOD1 wild type mutants, suggesting a specific dynamic signature of the mutations on the *ps-ns* timescale. Mutant SOD1 was also found to be more prone to disulfide-reduction[97]. Since the native disulfide bond is well-buried in the structure of SOD1, the observation of increased reduction susceptibility implies a greater dynamic tendency of mutant SOD1 to open. The increased flexibility of mutant SOD1 is also reflected in the observation that mutant SOD1 shows significantly higher affinity to bind hydrophobic beads *in vitro* compared to wild type[98]. However, the details of the motions implicated in misfolding and aggregation are not fully understood, especially in the *μs-ms* or higher timescales.

1.6 Outline of the Dissertation

The studies outlined in this dissertation were aimed at uncovering the molecular mechanism of SOD1 aggregation implicated in FALS. In Chapter 2, *in vitro* studies leading to the discovery of a mechanism for the aggregation of SOD1 are outlined. We

found that aggregation involves dimer dissociation and metal-loss from the resulting monomers, suggesting that FALS mutations enhance the dimer dissociation and/or metal-loss propensities of SOD1. In Chapter 3, computational estimates of the effect of FALS mutations on SOD1 thermodynamic stability and dimer dissociation propensity are presented, where we show that most FALS mutations destabilize the dimer and increase the dimer dissociation propensity. To identify events downstream of apo-monomer formation in the aggregation mechanism, in Chapter 4, studies leading to the identification of putative sequence and structural determinants of SOD1 aggregation are presented. In Chapter 5, a molecular mechanism for SOD1 monomer folding is outlined, in which regions of the SOD1 molecule responsible for maintaining fold fidelity are identified. In Chapter 6, the effect of mutations on native-state SOD1 dynamics are studied, and we found that the aggregation-prone "hot-spot" regions identified in Chapters 4 and 5 undergo significantly altered dynamics, leading us to a proposal for an atomic-resolution molecular scenario for the conformational conversion of SOD1 into fibrils. Finally, in Chapter 7, conclusions of the dissertation and recommendations for further investigation are discussed.

1.6 References

1. Dobson, C. M. Protein misfolding, evolution and disease. *Trends in Biochemical Sciences* **24**, 329-332 (1999).

2. Dobson, C. M. Protein folding and its links with human disease. *From Protein Folding to New Enzymes* 1-26 (2001).

3. Dobson, C. M. Experimental landscapes for protein folding and misfolding. *Abstracts of Papers of the American Chemical Society* **221**, U392 (2001).

4. Dobson, C. M., Ellis, R. J. & Fersht, A. R. Protein misfolding and disease - Preface. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences* **356**, 129-131 (2001).

5. Dobson, C. M. The structural basis of protein folding and its links with human disease. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences* **356**, 133-145 (2001).

6. Sipe, J. D. & Cohen, A. S. Review: History of the amyloid fibril. *Journal of Structural Biology* **130**, 88-98 (2000).

7. Hsia, A. Y., Masliah, E., Mcconlogue, L., Yu, G. Q., Tatsuno, G., Hu, K., Kholodenko, D., Malenka, R. C., Nicoll, R. A. & Mucke, L. Plaque-independent disruption of neural circuits in Alzheimer's disease mouse models. *Proc. Natl. Acad. Sci. U. S. A* **96**, 3228-3233 (1999).

8. Mucke, L., Masliah, E., Yu, G. Q., Mallory, M., Rockenstein, E. M., Tatsuno, G., Hu, K., Kholodenko, D., Johnson-Wood, K. & Mcconlogue, L. High-level neuronal expression of abeta 1-42 in wild-type human amyloid protein precursor transgenic mice: synaptotoxicity without plaque formation. *J. Neurosci.* **20**, 4050-4058 (2000).

9. Dodart, J. C., Bales, K. R., Gannon, K. S., Greene, S. J., DeMattos, R. B., Mathis, C., DeLong, C. A., Wu, S., Wu, X., Holtzman, D. M. & Paul, S. M. Immunization reverses memory deficits without reducing brain Abeta burden in Alzheimer's disease model. *Nat. Neurosci.* **5**, 452-457 (2002).

10. Westerman, M. A., Cooper-Blacketer, D., Mariash, A., Kotilinek, L., Kawarabayashi, T., Younkin, L. H., Carlson, G. A., Younkin, S. G. & Ashe, K. H. The relationship between Abeta and memory in the Tg2576 mouse model of Alzheimer's disease. *J. Neurosci.* **22**, 1858-1867 (2002).

11. Walsh, D. M., Klyubin, I., Fadeeva, J. V., Cullen, W. K., Anwyl, R., Wolfe, M. S., Rowan, M. J. & Selkoe, D. J. Naturally secreted oligomers of amyloid beta protein

potently inhibit hippocampal long-term potentiation in vivo. *Nature* **416**, 535-539 (2002).

12. Yankner, B. A. Mechanisms of neuronal degeneration in Alzheimer's disease. *Neuron* **16**, 921-932 (1996).

13. Selkoe, D. J. Neuroscience - Alzheimer's disease: Genotypes, phenotype, and treatments. *Science* **275**, 630-631 (1997).

14. Yang, W., Dunlap, J. R., Andrews, R. B. & Wetzel, R. Aggregated polyglutamine peptides delivered to nuclei are toxic to mammalian cells. *Human Molecular Genetics* **11**, 2905-2917 (2002).

15. Schaffar, G., Breuer, P., Boteva, R., Behrends, C., Tzvetkov, N., Strippel, N., Sakahira, H., Siegers, K., Hayer-Hartl, M. & Hartl, F. U. Cellular toxicity of polyglutamine expansion proteins: Mechanism of transcription factor deactivation. *Molecular Cell* **15**, 95-105 (2004).

16. Sanchez, I., Mahlke, C. & Yuan, J. Y. Pivotal role of oligomerization in expanded polyglutamine neurodegenerative disorders. *Nature* **421**, 373-379 (2003).

17. Muchowski, P. J., Schaffar, G., Sittler, A., Wanker, E. E., Hayer-Hartl, M. K. & Hartl, F. U. Hsp70 and Hsp40 chaperones can inhibit self-assembly of polyglutamine proteins into amyloid-like fibrils. *Proceedings of the National Academy of Sciences of the United States of America* **97**, 7841-7846 (2000).

18. Chen, S., Berthelier, V., Yang, W. & Wetzel, R. Polyglutamine aggregation behavior in vitro supports a recruitment mechanism of cytotoxicity. *Journal of Molecular Biology* **311**, 173-182 (2001).

19. Kroemer, G. Mitochondrial control of apoptosis: an introduction. *Biochemical and Biophysical Research Communications* **304**, 433-435 (2003).

20. Kourie, J. I. & Henry, C. L. Ion channel formation and membrane-linked pathologies of misfolded hydrophobic proteins: The role of dangerous unchaperoned molecules. *Clinical and Experimental Pharmacology and Physiology* **29**, 741-753 (2002).

21. Kourie, J. I., Culverson, A. L., Farrelly, P. V., Henry, C. L. & Laohachai, K. N. Heterogeneous amyloid-formed ion channels as a common cytotoxic mechanism - Implications for therapeutic strategies against amyloidosis. *Cell Biochemistry and Biophysics* **36**, 191-207 (2002).

22. Kayed, R., Sokolov, Y., Edmonds, B., McIntire, T. M., Milton, S. C., Hall, J. E. & Glabe, C. G. Permeabilization of lipid bilayers is a common conformation-dependent activity of soluble amyloid oligomers in protein misfolding diseases. *J. Biol. Chem.* **279**, 46363-46366 (2004).

23. Kyriakides, T., Marquez, B., Panousopoulos, A., Kyriacou, E. & Kyriacou, K. Amyloid myopathy: evidence for mechanical injury to the sarcolemma. *Clinical Neuropathology* **21**, 145-148 (2002).

24. Cleveland, D. W. & Rothstein, J. D. From Charcot to Lou Gehrig: Deciphering selective motor neuron death in ALS. *Nature Reviews Neuroscience* **2**, 806-819 (2001).

25. Dobson, C. M. Protein-misfolding diseases: Getting out of shape. *Nature* **418**, 729-730 (2002).

26. Kayed, R., Head, E., Thompson, J. L., McIntire, T. M., Milton, S. C., Cotman, C. W. & Glabe, C. G. Common structure of soluble amyloid oligomers implies common mechanism of pathogenesis. *Science* **300**, 486-489 (2003).

27. Adamski-Werner, S. L., Palaninathan, S. K., Sacchettini, J. C. & Kelly, J. W. Diflunisal analogues stabilize the native state of transthyretin. Potent inhibition of amyloidogenesis. *Journal of Medicinal Chemistry* **47**, 355-374 (2004).

28. Cleveland, D. W. From Charcot to SOD1: Mechanisms of selective motor neuron death in ALS. *Neuron* **24**, 515-520 (1999).

29. Valentine, J. S., Doucette, P. A. & Potter, S. Z. Copper-zinc superoxide dismutase and amyotrophic lateral sclerosis. *Annual Review of Biochemistry* **74**, 563-593 (2005).

30. Rosen, D. R., Siddique, T., Patterson, D., Figlewicz, D. A., Sapp, P., Hentati, A., Donaldson, D., Goto, J., Oregan, J. P., Deng, H. X., Rahmani, Z., Krizus, A., Mckennayasek, D., Cayabyab, A., Gaston, S. M., Berger, R., Tanzi, R. E., Halperin, J. J., Herzfeldt, B., Vandenbergh, R., Hung, W. Y., Bird, T., Deng, G., Mulder, D. W., Smyth, C., Laing, N. G., Soriano, E., PericakVance, M. A., Haines, J., Rouleau, G. A., Gusella, J. S., Horvitz, H. R. & Brown, R. H. Mutations in Cu/Zn Superoxide-Dismutase Gene Are Associated with Familial Amyotrophic-Lateral-Sclerosis. *Nature* **362**, 59-62 (1993).

31. Bruijn, L. I., Miller, T. M. & Cleveland, D. W. Unraveling the mechanisms involved in motor neuron degeneration in ALS. *Annual Review of Neuroscience* **27**, 723-749 (2004).

32. DiDonato, M., Craig, L., Huff, M. E., Thayer, M. M., Cardoso, R. M. F., Kassmann, C. J., Lo, T. P., Bruns, C. K., Powers, E. T., Kelly, J. W., Getzoff, E. D. & Tainer, J. A. ALS mutants of human superoxide dismutase form fibrous aggregates via framework destabilization. *Journal of Molecular Biology* **332**, 601-615 (2003).

33. Lindberg, M. J., Tibell, L. & Oliveberg, M. Common denominator of Cu/Zn superoxide dismutase mutants associated with amyotrophic lateral sclerosis: Decreased stability of the apo state. *PNAS* 262527099 (2002).

34. Johnston, J. A., Dalton, M. J., Gurney, M. E. & Kopito, R. R. Formation of high molecular weight complexes of mutant Cu,Zn-superoxide dismutase in a mouse model for familial amyotrophic lateral sclerosis. *Proceedings of the National Academy of Sciences of the United States of America* **97**, 12571-12576 (2000).

35. Manfredi, G. & Xu, Z. S. Mitochondrial dysfunction and its role in motor neuron degeneration in ALS. *Mitochondrion* **5**, 77-87 (2005).

36. Higgins, C. M. J., Jung, C. W., Ding, H. L. & Xu, Z. S. Mutant Cu, Zn superoxide dismutase that causes motoneuron degeneration is present in mitochondria in the CNS. *Journal of Neuroscience* **22**, (2002).

37. Higgins, C. M. J., Jung, C. W. & Xu, Z. S. ALS-associated mutant SOD1G93A causes mitochondrial vacuolation by expansion of the intermembrane space and by involvement of SOD1 aggregation and peroxisomes. *Bmc Neuroscience* **4**, (2003).

38. Higgins, C. M. J., Jung, C. W., Ding, H. L. & Xu, Z. S. Mutant Cu, Zn superoxide dismutase that causes motoneuron degeneration is present in mitochondria in the CNS. *Journal of Neuroscience* **22**, (2002).

39. Higgins, C. M. J., Jung, C. W. & Xu, Z. S. ALS-associated mutant SOD1G93A causes mitochondrial vacuolation by expansion of the intermembrane space and by involvement of SOD1 aggregation and peroxisomes. *Bmc Neuroscience* **4**, (2003).

40. Liu, J., Lillo, C., Jonsson, P. A., Velde, C. V., Ward, C. M., Miller, T. M., Subramaniam, J. R., Rothstein, J. D., Marklund, S., Andersen, P. M., Brannstrom, T., Gredal, O., Wong, P. C., Williams, D. S. & Cleveland, D. W. Toxicity of familial ALS-linked SOD1 mutants from selective recruitment to spinal mitochondria. *Neuron* **43**, 5-17 (2004).

41. Xu, Z. S., Jung, C. W., Higgins, C., Levine, J. & Kong, J. M. Mitochondrial degeneration in amyotrophic lateral sclerosis. *Journal of Bioenergetics and Biomembranes* **36**, 395-399 (2004).

42. Rodriguez, J. A., Valentine, O. S., Eggers, D. K., Roe, J. A., Tiwari, A., Brown, R. H. & Hayward, L. J. 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).

43. Wang, J., Xu, G. L., Gonzales, V., Coonfield, M., Fromholt, D., Copeland, N. G., Jenkins, N. A. & Borchelt, D. R. Fibrillar inclusions and motor neuron degeneration in transgenic mice expressing superoxide dismutase 1 with a disrupted copper-binding site. *Neurobiology of Disease* **10**, 128-138 (2002).

44. Rossi, L., Marchese, E., Demartino, A., Rotilio, G. & Ciriolo, M. R. Purification of a fully metal-depleted Cu, Zn superoxide dismutase from copper-deficient rat liver. *Biometals* **10**, 257-262 (1997).

45. Fandrich, M. & Dobson, C. M. Protein folding aspects of amyloid fibril formation. *Amyloid-Journal of Protein Folding Disorders* **8**, 26 (2001).

46. Guijarro, J. I., Sunde, M., Jones, J. A., Campbell, I. D. & Dobson, C. M. Amyloid fibril formation by an SH3 domain. *Proceedings of the National Academy of Sciences of the United States of America* **95**, 4224-4228 (1998).

47. Baldwin, M. A., James, T. L., Cohen, F. E. & Prusiner, S. B. The three-dimensional structure of prion protein: implications for prion disease. *Biochemical Society Transactions* **26**, 481-486 (1998).

48. Baskakov, I. V., Legname, G., Baldwin, M. A., Prusiner, S. B. & Cohen, F. E. Pathway complexity of prion protein assembly into amyloid. *J. Biol. Chem.* **277**, 21140-21148 (2002).

49. Cohen, F. E. & Prusiner, S. B. Pathologic conformations of prion proteins. *Annual Review of Biochemistry* **67**, 793-+ (1998).

50. Petkova, A. T., Ishii, Y., Balbach, J. J., Antzutkin, O. N., Leapman, R. D., Delaglio, F. & Tycko, R. A structural model for Alzheimer's beta -amyloid fibrils based on experimental constraints from solid state NMR. *PNAS* **99**, 16742-16747 (2002).

51. Ding, F., Borreguero, J. M., Buldyrev, S. V., Stanley, H. E. & Dokholyan, N. V. A Mechanism for the alpha-helix to beta-hairpin transition. *Proceedings of the National Academy of Sciences of the United States of America* submitted (2002).

52. Bullock, A. N., Henckel, J., DeDecker, B. S., Johnson, C. M., Nikolova, P. V., Proctor, M. R., Lane, D. P. & Fersht, A. R. Thermodynamic stability of wild-type and mutant p53 core domain. *Proceedings of the National Academy of Sciences of the United States of America* **94**, 14338-14342 (1997).

53. Canet, D., Last, A. M., Tito, P., Sunde, M., Spencer, A., Archer, D. B., Redfield, C., Robinson, C. V. & Dobson, C. M. Local cooperativity in the unfolding of an amyloidogenic variant of human lysozyme. *Nature Structural Biology* **9**, 308-315 (2002).

54. Chiti, F., Taddei, N., Bucciantini, M., White, P., Ramponi, G. & Dobson, C. M. Mutational analysis of the propensity for amyloid formation by a globular protein. *Embo Journal* **19**, 1441-1449 (2000).

55. Villegas, V., Zurdo, J., Filimonov, V. V., Aviles, F. X., Dobson, C. M. & Serrano, L. Protein engineering as a strategy to avoid formation of amyloid fibrils. *Protein Sci* **9**, 1700-1708 (2000).

56. Betancourt, M. R. & Thirumalai, D. Exploring the kinetic requirements for enhancement of protein folding rates in the GroEL cavity. *Journal of Molecular Biology* **287**, 627-644 (1999).

57. Richardson, J. S. & Richardson, D. C. Natural beta-sheet proteins use negative design to avoid edge-to-edge aggregation. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 2754-2759 (2002).

58. Otzen, D. E., Kristensen, O. & Oliveberg, M. Designed protein tetramer zipped together with a hydrophobic Alzheimer homology: A structural clue to amyloid assembly. *Proceedings of the National Academy of Sciences of the United States of America* **97**, 9907-9912 (2000).

59. Dima, R. I. & Thirumalai, D. Exploring the propensities of helices in PrPc to form beta sheet using NMR structures and sequence alignments. *Biophys. J.* **83**, 1268-1280 (2002).

60. Chiti, F. Amyloid formation by a protein not associated with amyloid diseases. *Italian Journal of Biochemistry* **49**, 4 (2000).

61. Harrison, P. M., Chan, H. S., Prusiner, S. B. & Cohen, F. E. Conformational propagation with prion-like characteristics in a simple model of protein folding. *Protein Sci* **10**, 819-835 (2001).

62. Kelly, J. W. Towards an understanding of amyloidogenesis. *Nature Structural Biology* **9**, 323-325 (2002).

63. Malolepsza, E., Boniecki, M., Kolinski, A. & Piela, L. Theoretical model of prion propagation: A misfolded protein induces misfolding. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 7835-7840 (2005).

64. Liu, Y. S., Gotte, G., Libonati, M. & Eisenberg, D. A domain-swapped RNase A dimer with implications for amyloid formation. *Nature Structural Biology* **8**, 211-214 (2001).

65. Rousseau, F., Schymkowitz, J. W. H. & Itzhaki, L. S. The unfolding story of three-dimensional domain swapping. *Structure* **11**, 243-251 (2003).

66. Liu, Y. & Eisenberg, D. 3D domain swapping: As domains continue to swap. *Protein Sci* **11**, 1285-1299 (2002).

67. Ding, F., Dokholyan, N. V., Buldyrev, S. V., Stanley, H. E. & Shakhnovich, E. I. Molecular dynamic simulation of the SH3 domain aggregation suggests a generic amyloidogenesis mechanism. *Journal of Molecular Biology* **324**, 851-857 (2002).

68. Yang, S. C., Cheung, M. S., Onuchic, J. N. & Levine, H. Molecular dynamics simulations on domain swapping. *Biophys. J.* **86**, 267A-268A (2004).

69. Yang, S. C., Cho, S. S., Levy, Y., Cheung, M. S., Levine, H., Wolynes, P. G. & Onuchic, J. N. Domain swapping is a consequence of minimal frustration. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 13786-13791 (2004).

70. Rousseau, F., Schymkowitz, J. W. H., Wilkinson, H. R. & Itzhaki, L. S. Three-dimensional domain swapping in p13suc1 occurs in the unfolded state and is controlled by conserved proline residues. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 5596-5601 (2001).

71. Liu, Y. S., Hart, P. J., Schlunegger, M. P. & Eisenberg, D. The crystal structure of a 3D domain-swapped dimer of RNase A at a 2.1-angstrom resolution. *Proceedings of the National Academy of Sciences of the United States of America* **95**, 3437-3442 (1998).

72. Liu, Y. S., Gotte, G., Libonati, M. & Eisenberg, D. Structures of the two 3D domain-swapped RNase A trimers. *Protein Sci* **11**, 371-380 (2002).

73. Janowski, R., Kozak, M., Jankowska, E., Grzonka, Z., Grubb, A., Abrahamson, M. & Jaskolski, M. Human cystatin C, an amyloidogenic protein, dimerizes through three-dimensional domain swapping. *Nature Structural Biology* **8**, 316-320 (2001).

74. Chen, Y. W., Stott, K. & Perutz, M. F. Crystal structure of a dimeric chymotrypsin inhibitor 2 mutant containing an inserted glutamine repeat. *Proceedings of the National Academy of Sciences of the United States of America* **96**, 1257-1261 (1999).

75. Sambashivan, S., Liu, Y. S., Sawaya, M. R., Gingery, M. & Eisenberg, D. Amyloid-like fibrils of ribonuclease A with three-dimensional domain-swapped and native-like structure. *Nature* **437**, 266-269 (2005).

76. Go, N. & Abe, H. Noninteracting local-structure model of folding and unfolding transition in globular proteins. I. Formulation. *Biopolymers* **20**, 991-1011 (1981).

77. Abe, H. & Go, N. Noninteracting local-structure model of folding and unfolding transition in globular proteins. II. Application to two-dimensional lattice proteins. *Biopolymers* **20**, 1013-1031 (1981).

78. Ding, F., Dokholyan, N. V., Buldyrev, S. V., Stanley, H. E. & Shakhnovich, E. I. Molecular dynamics simulation of the SH3 domain aggregation suggests a generic amyloidogenesis mechanism. *Journal of Molecular Biology* **324**, 851-857 (2002).

79. Kishan, K. V. R., Scita, G., Wong, W. T., DiFiore, P. P. & Newcomer, M. E. The SH3 domain of Eps8 exists as a novel intertwined dimer. *Nature Structural Biology* **4**, 739-743 (1997).

80. Sunde, M., Serpell, L. C., Bartlam, M., Fraser, P. E., Pepys, M. B. & Blake, C. C. F. Common core structure of amyloid fibrils by synchrotron X-ray diffraction. *Journal of Molecular Biology* **273**, 729-739 (1997).

81. Thirumalai, D., Klimov, D. K. & Dima, R. I. Emerging ideas on the molecular basis of protein and peptide aggregation. *Current Opinion in Structural Biology* **13**, 146-159 (2003).

82. Balbach, J. J., Petkova, A. T., Oyler, N. A., Antzutkin, O. N., Gordon, D. J., Meredith, S. C. & Tycko, R. Supramolecular structure in full-length Alzheimer's beta-amyloid fibrils: Evidence for a parallel beta-sheet organization from solid-state nuclear magnetic resonance. *Biophys. J.* **83**, 1205-1216 (2002).

83. Chiti, F., Stefani, M., Taddei, N., Ramponi, G. & Dobson, C. M. Rationalization of the effects of mutations on peptide and protein aggregation rates. *Nature* **424**, 805-808 (2003).

84. Fernandez-Escamilla, A. M., Rousseau, F., Schymkowitz, J. & Serrano, L. Prediction of sequence-dependent and mutational effects on the aggregation of peptides and proteins. *Nature Biotechnology* **22**, 1302-1306 (2004).

85. Dobson, C. M. Protein folding and misfolding. *Nature* **426**, 884-890 (2003).

86. Hoffman, E. K., Wilcox, H. M., Scott, R. W. & Siman, R. Proteasome inhibition enhances the stability of mouse Cu/Zn superoxide dismutase with mutations linked to familial amyotrophic lateral sclerosis. *Journal of the Neurological Sciences* **139**, 15-20 (1996).

87. Borchelt, D. R., Guarnieri, M., Wong, P. C., Lee, M. K., Slunt, H. S., Xu, Z. S., Sisodia, S. S., Price, D. L. & Cleveland, D. W. Superoxide-Dismutase-1 Subunits with Mutations Linked to Familial Amyotrophic-Lateral-Sclerosis do Not Affect Wild-Type Subunit Function. *J. Biol. Chem.* **270**, 3234-3238 (1995).

88. Nakano, R., Inuzuka, T., Kikugawa, K., Takahashi, H., Sakimura, K., Fujii, J., Taniguchi, N. & Tsuji, S. Instability of mutant Cu/Zn superoxide dismutase (Ala4Thr) associated with familial amyotrophic lateral sclerosis. *Neuroscience Letters* **211**, 129-131 (1996).

89. Forman, H. J. & Fridovich, I. On the stability of bovine superoxide dismutase. The effects of metals. *J. Biol. Chem.* **248**, 2645-26499 (1973).

90. Lindberg, M. J., Bystrom, R., Boknas, N., Andersen, P. M. & Oliveberg, M. Systematically perturbed folding patterns of amyotrophic lateral sclerosis (ALS)-associated SOD1 mutants. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 9754-9759 (2005).

91. Rodriguez, J. A., Shaw, B. F., Durazo, A., Sohn, S. H., Doucette, P. A., Nersissian, A. M., Faull, K. F., Eggers, D. K., Tiwari, A., Hayward, L. J. & Valentine, J. S. Destabilization of apoprotein is insufficient to explain Cu,Zn-superoxide dismutase-linked ALS pathogenesis. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 10516-10521 (2005).

92. Stroppolo, M. E., Malvezzi-Campeggi, F., Mei, G., Rosato, N. & Desideri, A. Role of the tertiary and quaternary structures in the stability of dimeric copper,zinc superoxide dismutases. *Archives of Biochemistry and Biophysics* **377**, 215-218 (2000).

93. Doucette, P. A., Whitson, L. J., Cao, X. H., Schirf, V., Demeler, B., Valentine, J. S., Hansen, J. C. & Hart, P. J. Dissociation of human copper-zinc superoxide dismutase dimers using chaotrope and reductant - Insights into the molecular basis for dimer stability. *J. Biol. Chem.* **279**, 54558-54566 (2004).

94. Cardoso, R. M. F., Thayer, M. M., DiDonato, M., Lo, T. P., Bruns, C. K., Getzoff, E. D. & Tainer, J. A. Insights into Lou Gehrig's disease from the structure and instability of the A4V mutant of human Cu,Zn superoxide dismutase. *Journal of Molecular Biology* **324**, 247-256 (2002).

95. Hart, P. J., Liu, H. B., Pellegrini, M., Nersissian, A. M., Gralla, E. B., Valentine, J. S. & Eisenberg, D. Subunit asymmetry in the three-dimensional structure of a human CuZnSOD mutant found in familial amyotrophic lateral sclerosis. *Protein Sci* **7**, 545-555 (1998).

96. Shipp, E. L., Cantini, F., Bertini, I., Valentine, J. S. & Banci, L. Dynamic properties of the G93A mutant of copper-zinc superoxide dismutase as detected by NMR spectroscopy: Implications for the pathology of familial amyotrophic lateral sclerosis. *Biochemistry* **42**, 1890-1899 (2003).

97. Tiwari, A. & Hayward, L. J. Familial amyotrophic lateral sclerosis mutants of copper/zinc superoxide dismutase are susceptible to disulfide reduction. *J. Biol. Chem.* **278**, 5984-5992 (2003).

98. Tiwari, A., Xu, Z. S. & Hayward, L. J. Aberrantly increased hydrophobicity shared by mutants of Cu,Zn-superoxide dismutase in familial amyotrophic lateral sclerosis. *J. Biol. Chem.* **280**, 29771-29779 (2005).

CHAPTER 2

AN *IN VITRO* MECHANISM FOR SOD1 AGGREGATION

This chapter has been published as a manuscript in the *Proceedings of the National Academy of Sciences USA* and is formatted in the journal style.

2.1 Abstract

Mutation-induced aggregation of the dimeric enzyme Cu, Zn superoxide dismutase (SOD1) has been implicated in the familial form of the disease ALS (FALS), but the mechanism of aggregation is not known. Here, we show that *in vitro* SOD1 aggregation is a multi-step reaction that minimally consists of dimer dissociation, metal loss from the monomers, and oligomerization of the apo-monomers:

$$D \underset{k_{\text{on}}}{\overset{k_{\text{off}}}{\rightleftarrows}} 2M_{\text{holo}} \underset{k_{\text{m}}^{+}}{\overset{k_{\text{m}}^{-}}{\rightleftarrows}} 2M_{\text{apo}} \xrightarrow{k_{\text{agg}}} A$$

where D_{holo}, M_{holo}, M_{apo} and A are the holo-dimer, holo-monomer, apo-monomer and aggregate respectively. Under aggregation-promoting conditions (pH 3.5), the rate and equilibrium constants corresponding to each step are: (1) dimer dissociation: $K_d \approx 1\mu M$; $k_{off} \approx 1\times10^{-3}$ s^{-1}, $k_{on} \approx 1\times10^{3}$ $M^{-1}s^{-1}$; (2) metal loss: $K_m \approx 0.1\mu M$, $k_m^- \approx 1\times10^{-3}$ s^{-1}, $k_m^+ \approx 1\times10^{4}$ $M^{-2}s^{-1}$; (3) assembly (rate-limiting step): $k_{agg} \approx 1\times10^{3}$ $M^{-1}s^{-1}$. In contrast, under near-physiological conditions (pH 7.8), where aggregation is drastically reduced, dimer dissociation is less thermodynamically favorable: $K_d \approx 0.1$ nM, and extremely slow: $k_{off} \approx 3\times10^{-5} s^{-1}$, $k_{on} \approx 3\times10^{5}$ $M^{-1}s^{-1}$. Our results suggest that FALS-linked SOD1 aggregation occurs by a mutation-induced increase in dimer dissociation and/or increase in apo-monomer formation.

2.2 Introduction

Amyotrophic Lateral Sclerosis (ALS) involves selective motor neuron death in the brain and the spinal cord (1-6) initiating a progressive paralysis in mid-life. Point mutations in the cytoplasmic homodimeric enzyme Cu, Zn superoxide dismutase (SOD1, M_w≈32kDa) were identified as the primary cause of approximately 20% cases of the familial form of ALS (FALS) (7, 8) in contrast with the sporadic form (SALS). Approximately 90 distinct FALS mutations are known (9). The toxic gain-of-function of the mutants is believed to be associated with either intracellular misfolding and aggregation or oxidative damage caused by mutant SOD1-catalyzed aberrant reactions, although the two scenarios may not be mutually exclusive (10, 11). The aggregation hypothesis is supported by the observations that in both mice and cell culture models, death of motor neurons is preceded by formation of cytoplasmic aggregates containing mutant SOD1 (12-15), that SOD1 knockout mice do not develop motor neuron disease (16), and that even Cu-depleted SOD1 mutants cause the disease in mice (17). Although aggregation of SOD1 has also been found in some cases of SALS (18-20), the sporadic disease is believed to have a different molecular basis (1). FALS and SOD1-linked SALS may belong to the general class of protein conformational disorders in which perturbation of protein folding leads to a relatively higher population of misfolded or partially folded (21) protein molecules which then aggregate into regular (22) structured fibrils. Toxicity may arise due to the saturation of the cellular chaperone machinery by the misfolded or aggregated SOD1 molecules (23).

ALS is an age-related disease occurring in mid-life or later, and it is not surprising that the *in vitro* aggregation of SOD1 solutions under physiological conditions is slow. Therefore, it

is common to study SOD1 aggregation by perturbing the environmental conditions – by lowering the pH to 3.5 (24), by adding mild denaturants such as trifluoroethanol, or by heat-treatment (25) – such that disease-like aggregates can be detected on experimental time-scales. Aggregation-prone FALS mutations and/or loss of metals also decrease SOD1 stability (26, 27), indicating that exposure to non-physiological conditions, such as low pH, mimics the effect of mutation by similarly lowering the barrier for SOD1 aggregation. Although the rate of aggregation can be enhanced *in vitro*, the molecular mechanism of SOD1 aggregation is not well-understood.

We show here that the relatively rapid aggregation of SOD1 that results from perturbations in the environmental conditions occurs because these promote dimer dissociation and/or the loss of metals, and that the thus formed apo-monomers undergo multimeric assembly. We propose that this same reaction path holds for some or all FALS mutant forms of SOD1 and that increased aggregation results from perturbation of one or more of the steps in the pathway demonstrated here: dimer dissociation, metal loss from monomers, aggregation of apo-monomers.

2.3 Methods

Protein and buffers.

Bovine and human erythrocyte Cu, Zn superoxide dismutase (E.C. 1.15.1.1) were obtained from Sigma chemicals. Protein concentration was measured spectrophotometrically at 258 nm for the bovine enzyme (28). Buffers used in SPR at various pH values were: pH 7.8, pH 6.6, pH 5.8: 50mM phosphate, 150 mM NaCl; pH 4.8, pH 3.5: 100mM acetate, 150mM

NaCl, pH2.5, pH2.7, pH3.0. pH 10.8, pH 11.85, pH 12.5: 50mM phosphate buffer,150mM NaCl. All incubations were at 37 ^{0}C.

Biotinylation reaction.

Biotinylated SOD1 was prepared at low stoichiometry (1 mol equivalents of biotinylating agent/SOD1 dimer) for Surface Plasmon Resonance (SPR) studies and high (10 fold excess of biotinylating agent/SOD1 dimer) stoichiometry for Size Exclusion Chromatography (SEC) studies as previously described (29). Biotinylated SOD1 was found to be fully active after biotinylation ((30); data not shown) and had identical chromatographic properties compared to the unmodified protein (Supplementary Material S1).

Analytical ultracentrifugation (AUC).

The K_d for SOD1 dimer dissociation was determined using a Beckmann-XLA ultracentrifuge at a speed of 16,000 rpm at 4^0 C or 25^0C using a Ti-50 rotor. Radial UV absorbance (258 nm) was monitored every 2h in cells with a path length of 12 mm. One blank buffer sample was used for each protein sample to obtain baselines. A total of 7 scans were obtained for each sample and the corresponding blank cell to reach equilibrium, followed by over-spinning at 45,000 rpm to obtain another estimate of baseline absorbance (typically 0.01 units). The UV-absorbance profiles were superimposed to ensure equilibrium, analyzed for residuals and the equilibrium profile(s) were fit to the Lambert equation to obtain estimates of K_d and apparent molecular weight M_w.

Size exclusion chromatography.

SEC was performed with a Pharmacia Akta chromatography system using an Amersham-Pharmacia Superdex HR 10/30 column (Piscataway, NJ), with a 200-µl injection loop, total volume of 23.56 ml and a void volume estimate of 7.8 ml, working at 5°C. The column flow rate was 0.45 ml/min. Fractions (150-250 µl) were collected in glass tubes and these were analyzed immediately after completing the chromatogram. In reactions with low (nanomolar) SOD1 concentrations, the reaction mixture and the column buffer contained BSA at 10 mg/l to prevent nonspecific binding of SOD1 to test tubes and to the column matrix. UV absorbance was monitored spectrophotometrically at 254nm.

Dot blotting.

Column fractions were analyzed by a dot-blot assay, with Avidin-AP, as previously described (29) and/or by immunostaining with anti-SOD1 antibody (Supplementary Material S2).

Surface Plasmon Resonance measurements.

Plasmon resonance sensor chips, pre-coated with strepavidin, were obtained from Biacore Corp. (Piscataway, NJ) and used with a Biacore Model 2000 instrument. The chips were pretreated with NaOH/50mM NaCl (pH 12.7) before biotin-SOD1 was immobilized. The volume of each flow cell (4 per chip) was 7nl and a buffer flow of 2µl/min was used at 25^0C. Low stoichiometry (≤1 biotin/dimer) biotin-SOD1 was immobilized on one flow cell surface such that an increase of approximately 3000 Resonance Units (RU) resulted. After binding was complete, SOD1 dimer dissociation was induced with protein-free buffer. A small correction for the signal drift was applied using the signal from the flow-cells that had not received biotin-

SOD1 but received protein-free buffer. To obtain the pH dependence of the rate constant, the zero-order dissociation reaction was followed sequentially with increasing or decreasing pH, starting at pH 7.8.

2.4 Results

We show that the minimal sequence of steps constituting the aggregation pathway is:

$$D \underset{k_{on}}{\overset{k_{off}}{\rightleftarrows}} 2M_{holo} \underset{k_m^+}{\overset{k_m^-}{\rightleftarrows}} 2M_{apo} \xrightarrow{k_{agg}} A \qquad (1)$$

where D_{holo}, M_{holo}, M_{apo} and A are the holo-dimer, holo-monomer, Zn-free monomer (31) and aggregate respectively. Next, we show that: (i) the K_d and k_{off} for SOD1 dimer dissociation are enhanced under aggregation-promoting conditions; (ii) aggregation requires dissociation of the SOD1 dimer to monomers and a subsequent loss of metals. Based on the observed aggregation kinetics, we calculated the rate and equilibrium constants for each step in the reaction sequence.

The K_d for SOD1 dimer dissociation is 6×10^3-fold higher under aggregation-enhancing conditions than under near-physiological conditions.

The K_d and molecular weight M_w of SOD1 under aggregation-enhancing conditions were determined using analytical ultracentrifugation (AUC, Table I), fitting the absorbance profiles to a dimer-monomer equilibrium. 30μM SOD1 was entirely dimeric (apparent M_w≈36kDa) at pH 7.8 and was entirely monomeric (apparent M_w≈16kDa) when metals were chelated by incubation with 10mM EDTA at pH 3.5. A K_d ≈1μM was calculated for the dimer-monomer equilibrium by a combined fit to the AUC profiles of 30, 20 and 10μM SOD1. Thus, at pH 3.5, where aggregation was previously observed (24), 30μM SOD1 was ≈84% dimer and

≈16% monomer. In contrast, 30μM SOD1 was ≈100% dimer under near-physiological conditions (pH 7.8), where aggregation does not occur.

The K_d for SOD1 dimer dissociation at pH 7.8 could not be measured by AUC because the optical system used to detect the protein is limited to micromolar concentrations. However, nM SOD1 concentrations could be analyzed by using size exclusion chromatography (SEC) and a dot-blot assay. Based on the calibration of the Superdex HR 10/30 column (Fig. 1A), the SOD1 dimer and monomer are expected to elute at 15.5 and 17.0 ml respectively; SOD1 at 30μM in pH 7.8 buffer eluted as ≈100% dimer at 15.6 ml (Fig. 1B), in accord with AUC measurements (Table I). The K_d at pH 7.8 was estimated with 5nM SOD1 which gave column fractions at the limit of detection. The bulk of the protein eluted as a dimer at 15.3 ml with a small reproducible monomer peak at 16.8 ml, comprising ≈20% of the total protein (Fig. 1C). The dimer and monomer were expected to elute as distinct peaks because they were in slow equilibrium: the half-time for dimer dissociation was ≈6 h at pH 7.8 and 25^0C (see below), whereas the time required for chromatographic separation was ≈10 minutes at 5^0C. A K_d≈0.16 nM was calculated from a double-Gaussian least-square fit to the data, and is in reasonable agreement with the K_d≈1nM estimated from the unfolding studies of human SOD1 (32). Thus, the K_d≈1×10^{-6} M at pH 3.5 is larger than the K_d≈1.6×10^{-10} M at pH 7.8 by a factor of 6×10^3. This corresponds to a large (≥5.2 kcal/mol) reduction in dimer stability, indicating that a higher fraction of SOD1 is monomeric under aggregation-enhancing conditions than under near-physiological conditions.

Since identical M_w estimates were obtained for 10μM SOD1 at pH 3.5 with and without a 24 h incubation (Table I) incubation, dimer dissociation is presumably reversible and the irreversible (33),(34) step in the aggregation lies downstream of dimer dissociation. The

reversibility of dimer dissociation under aggregation-promoting conditions was directly demonstrated by SEC (Supplementary Material S4).

The k_{off} for SOD1 dimer dissociation is 30-fold higher under aggregation-enhancing conditions than under near-physiological conditions.

Having determined that the K_d for dimer dissociation was enhanced at least 20,000-fold at pH 3.5, we used Surface Plasmon Resonance (SPR) to determine the kinetics of dimer dissociation. SOD1, biotinylated at low stoichiometry (≤1 biotin/SOD1 dimer), was immobilized on a strepavidin-coated gold surface. As shown in Fig. 2A, the biotinylated dimer has one subunit bound to strepavidin and the non-biotinylated subunit is progressively lost by flowing SOD1-free buffer through the flow-cell, leading to a SPR signal decrease of approximately 50%. The dissociation is made irreversible by a 2000 nl/minute flow of protein-free buffer through the 7 nl flow-cell. The rate constant for dissociation k_{off} was determined by an exponential fit to the entire signal decay corresponding to a first-order reaction (Figs. 2B and 2C), or by a linear fit to the initial signal decay corresponding to a zero-order reaction (Fig. 2D). The dissociation at pH 3.5 had a half-time of ≈600 s (k_{off} ≈1.0×10^{-3} s^{-1}; Fig. 2B) whereas the dissociation at pH 7.8 had a half time of ≈6 h (k_{off}≈ 3.1×10^{-5} s^{-1}; Fig. 2C). Therefore, under conditions where aggregation was observed (pH 3.5), the rate of dimer dissociation is a 30-fold faster than under near-physiological conditions. This corresponds to an approximately 2.0 kcal/mol reduction in the energy barrier for dissociation.

SOD1 dimer dissociation may be induced by the disruption of salt-linkages.

The dramatic increase in k_{off} at pH 3.5 suggests that insights into the forces stabilizing the SOD1 dimer could be obtained from the pH-dependence of the reaction. The k_{off} increased sharply on the acidic and basic ends of the pH range 2.5-12.7 (Fig. 2D and 2E). In the simplest unified model to account for this effect, dimer dissociation is linked with the titration of carboxyl (Asp, Glu) and amine (Lys, Arg) groups forming a salt-linkage. At any given pH, the observed k_{off} reflects the dissociation of three sub-populations of the SOD1 dimer, in which the salt-linkage is: (i) intact; (ii) disrupted because of the protonation of the carboxylic acid; and (iii) disrupted because of the de-protonation of the amine. The species from sub-populations (i), (ii), and (iii) have intrinsic dissociation rate constants k_{ind}, k_{acid}, and k_{basic} respectively, which are weighted by their pH-dependent population size, so that the observed k_{off} is:

$$k_{off}^{obs} = k_{acid}\left\{\frac{[H^+]}{[H^+]+K_a}\right\} + k_{basic}\left\{\frac{K_b}{K_b+[H^+]}\right\} + k_{ind} \qquad (2)$$

where K_a and K_b are the equilibrium constants for the de-protonation of the carboxyl and amine groups involved. The first and the second term dominate at low and high pH respectively, and the rate is pH-independent at intermediate pH values. According to Eq.(2), in the low pH regime the titration-induced increase in $\log k_{off}^{obs}$ is predicted to have a slope of -1 when $pH >> pK_a$. Similarly, a slope of 1 is predicted for the corresponding increase in the high pH regime when $pH << pK_b$, and such increases were observed at both low and high pH (Fig. 2E). The apparent pK_a and pK_b values of <2.5 and >12.5, presumably correspond to carboxyl (Asp; free pK=3.86) and amine (Arg; free pK=12.48) groups forming a highly stabilizing salt-linkage. The linkage of the protonation of the carboxylic acid and the de-protonation of the amine to the dimer dissociation leads to the abnormal apparent pK values. The 100-fold change in the observed k_{off} in the intermediate pH range 4.8-10.8 is indicative of

other titrations, possibly of charged amino acid side-chains which may not form salt-linkages, so that they have a smaller effect on the rate k_{ind}.

Human SOD1 (hSOD1), which is highly homologous (82% sequence identity) to the bovine enzyme, exhibits a similar pH-dependence of dimer dissociation. The k_{off} values were similar at pH 6.6 (Fig. 2D, inset) and at pH 7.8: 2.74×10^{-5} s^{-1} for hSOD1 (data not shown) and 3.13×10^{-5} s^{-1} for bovine SOD1 (Fig. 2B).

It is interesting that there are no salt-linkages on the dimer interface of either bovine or human SOD1 so that the enhanced dissociation rate is caused indirectly by the titration of a salt-linkage that is not in the proximity of the dimer interface. The Arg79-Asp101 interaction was identified as a candidate for such a salt-linkage (Supplementary Material S5). This salt-linkage mechanism provides a unified framework that accounts for the steep pH-dependence of the reaction at low and high pH, and is consistent with the previous finding that charge-charge interactions in SOD1 contribute substantially to its stability and folding kinetics (21). An alternative mechanism in which ionization of surface exposed residue(s) dramatically influences dimer stability is less likely since it does not simultaneously predict the behavior at extremes of pH, and the modest pH-dependence of the rate at intermediate pH values.

Rapid oligomerization of SOD1 occurs under conditions where it is monomeric.

To test directly whether SOD1 monomer or dimer is the aggregation-competent species, we characterized the initial phase of aggregation as a function of the degree of dimer dissociation at pH 3.5. Based on a $K_d \approx 1\mu M$, as determined by AUC (Table I), a 30μM SOD1 solution was 84% (25μM) dimeric, while a 1μM solution was 39% (0.39μM) dimeric, corresponding to a ≈60-fold excess of dimer in the 30μM sample compared to the 1μM sample.

30μM SOD1 eluted predominantly as a dimer (elution volume 16.5 ml; Supplementary material Fig. S2A) on incubation for 40 minutes (Fig. 3A, top panel). 40 minutes was sufficient for the establishment of equilibrium since the dimer dissociation has a half-time of ≈600s at this pH (Fig. 2B). The elution profiles remained largely unaltered for 48 h (Fig. 3A, middle and bottom panels) and there was little aggregation since nearly identical peak heights and areas in the UV-absorbance profiles were obtained after incubation. This agreed with AUC results in which 30μM SOD1 gave nearly identical UV-absorbance signals and apparent M_w estimates (Table I). After 48 h, most of the protein remained dimeric or monomeric (Fig. 3A), and less than 10% was aggregated.

Relatively rapid aggregation occurred with 1μM SOD1 (Fig. 3B). After 40 minutes of incubation, the protein eluted in a broad peak centered at 17.2 ml corresponding to a 40:60 mixture of dimer and monomer, as expected from K_d≈1μM. The rate of aggregation was initially rapid and after 24 and 48 h approximately 60% of the protein was aggregated. The aggregation rate is expected to decrease markedly with time, and the failure to observe a significant increase in the fraction of aggregated protein at 48 h can be attributed to experimental error. These results show that at a given concentration, SOD1 aggregation is proportional to the fraction of monomer, rather than dimer, in solution.

Rapid oligomerization of SOD1 occurs under conditions where metals are lost irreversibly.

Since the chelation of metals is known to destabilize SOD1 (26) and promote aggregation (24), we postulated that the reactive species for aggregation is the apo-monomer. To test this, we determined the time-course of aggregation of 30μM SOD1 under conditions

where metal loss was made irreversible by dialysis against metal free buffer. Dialysis at pH 3.6 results in a rapid loss of 95% of the Zn (31) and the unbound metals were expected to be removed rapidly since we had found that the loss of cupric acetate from the dialysis bag has a concentration-independent (range 100μM-10mM) half-time of 10 minutes under the conditions of the experiment (data not shown). A substantial amount of aggregation was expected if the loss of metals from SOD1 induces formation of a species that is aggregation-competent.

During incubation under dialysis conditions the amount of dimer rapidly decreased, the degree of dimer dissociation increased, and the protein was almost completely aggregated in 21 h (Fig. 3D). Electron micrographs of the aggregates obtained by negative staining, show that they have fibrillar morphology (Fig. 3D, inset). Despite the fact that the dimer concentration decreased monotonically, the monomer concentration initially increased and subsequently gradually decreased (Fig. 3E), indicating that under dialysis conditions, dimer dissociation is coupled to metal loss and aggregation of the apo-monomers in the reaction sequence:

$$D \underset{k_{on}}{\overset{k_{off}}{\rightleftarrows}} 2M_{holo} \underset{k_m^+}{\overset{k_m^-}{\rightleftarrows}} (Zn)_{dialysed} + 2M_{apo} \xrightarrow{k_{agg}} A \qquad (3)$$

where D_{holo}, M_{holo}, M_{apo} and A are the holo-dimer, holo-monomer, Zn-free monomer and aggregate respectively. In contrast, in AUC studies described above, the dominant species after incubation remained D_{holo} and M_{holo} and there was no detectable aggregation. The coincidence between the increase in M_{apo} and the increased rate of aggregation shows that the apo-monomer is the reactive species for this process. Moreover, since the apo-monomer accumulated before aggregation, it is concluded that the oligomerization of apo-monomers is the rate-limiting step in Eq.(3).

For a more detailed characterization of SOD1 aggregation at pH 3.5, the rate-constants in the reaction sequence Eq. (3) were obtained. The value $k_{off} \approx 1\times10^{-3}$ s^{-1} (Fig. 2B) was

determined from SPR and $k_{on} \approx 10^3$ $M^{-1}s^{-1}$ was calculated from $K_d \approx 1\mu M$. The constants $k_m^- \approx 1.2\times10^{-4}$ s^{-1}, $k_m^+ \approx 5.2\times10^4$ $M^{-1}s^{-1}$, and $k_{agg} \approx 0.1\times10^3$ $M^{-1}s^{-1}$ were estimated from the observed kinetics of dialysis-induced aggregation (Supplementary material S3). The estimated value of $k_m^- \approx 1.2\times10^{-4}$ s^{-1} agrees well with the previously measured $k_m^- = 4.2\times10^{-4}$ s^{-1} (31) for the loss of Zn from SOD1 at pH 3.6 (half-time of ≈27 minutes at 20°C). The differential master equations corresponding to Eq.(3) were solved numerically (Supplementary Material S3) and the above rate constant estimates were refined, fitting to the observed time-profile of dialysis-induced aggregation (Fig. 3E). The values that best account for the experimental results are: $k_{off} = 3\times10^{-3}$ s^{-1}, $k_{on} = 1.6\times10^3$ M^{-1} s^{-1}, $k_m^- = 1\times10^{-3}$ s^{-1}, $k_m^+ = 9.8\times10^3$ M^{-1} s^{-1} and $k_{agg} = 1\times10^3$ $M^{-1}s^{-1}$. The rate constant values were found to be robust within an order of magnitude: the fit worsened dramatically on variations in rate constants greater than an order of magnitude. In accord with the above rate constants, which predict 12% aggregation in 30μM SOD1 after 24 h under non-dialysis conditions, we had found little aggregation (Fig. 3A). Similarly, 14% and 63% aggregation is predicted at 30 and 1μM in 48 h (Fig.3F), which is in good agreement with the observed values of 10% and 60% respectively.

2.5 Discussion

To create amyloidogenic conditions under which the aggregation of proteins can be readily induced, acidic conditions are routinely employed. For example, amyloidogenic intermediates of transthyretin (TTR), β_2-microglobulin, prion and lysozyme have been characterized at pH 4.5 (35), pH 3.6 (36), pH 4.4-6 (37), and pH 2 (38) respectively. Under conditions where SOD1 aggregation is experimentally accessible (pH 3.5), we have demonstrated an aggregation pathway in which the dimer dissociates, metal is lost from the

resultant monomer, and the apo-monomers aggregate in a rate-limiting step. The absence of any discontinuity in the rate of dimer dissociation when the pH is reduced from pH 7.8 to 3.5 (Fig. 2E) provides evidence that this same pathway holds at physiological pH, but at a drastically reduced rate (Fig. 4). Similarly, the pH effects on the metal content of human, bovine and yeast wild type and yeast mutant SOD1 (in the range pH 3-9) have previously been shown to be continuous, reversible, and related to conformational changes in the protein (39), further indicating that the holo- and apo-monomer equilibrium is also a representative property of the protein.

The sequence of events and the reactive species in the aggregation of SOD1.

To identify the sequence of events in the aggregation reaction and the aggregation-prone reactive species, we characterized the concentration and metal-content dependence of SOD1 aggregation. For the former, conditions such that there was a preponderance of either dimer or monomer were chosen, and the profiles of the initial phase of SOD1 aggregation under these conditions were compared. Total SOD1 concentrations of 30μM and 1μM were chosen so that at equilibrium there is an approximately 60-fold excess of dimer at the higher concentration. If aggregation is dependent on a bimolecular dimer-dimer reaction, a 3600-fold greater aggregation rate is expected at the higher concentration. Similarly, a 60-fold greater aggregation rate is expected if a unimolecular isomerization reaction i.e. a conformation change in dimer, leading to an aggregation-competent dimeric species, is rate-limiting. In stark contrast, we found rapid aggregation at the lower of the two concentrations, demonstrating that the dimer is not the reactive species for aggregation. The amount of aggregate at time t is $k_{agg}[M]^2t$ where $[M]$ is the concentration of the aggregating species.

Although there is an 8-fold excess of the total monomer at 30μM compared to 1μM, the concentration of the aggregating apo-monomer is 3-fold higher at 30μM. The total protein concentration is 30-fold higher in the 30μM sample and therefore, the fraction of the protein in aggregates, $f_{agg} = k_{agg}[M]^2 t/c_T$, where c_T is the total concentration of the protein, will be higher at 1μM than at 30μM, in agreement with the data in Fig. 3C.

To ascertain whether the reactive species for aggregation is a metal-bound or metal-free monomer, aggregation kinetics were determined with 30μM SOD1 under conditions where the loss of metals from the protein was irreversible. We found that the dimer concentration rapidly decreased whereas the monomer concentration initially increased and then gradually decreased (Fig. 3E), and that aggregation into a fibrillar product was nearly complete in 21 h. A comparison with the aggregation kinetics of 30μM SOD1 in the absence of dialysis (Fig. 3A) shows that the monomeric species appearing under dialysis conditions is the apo-monomer. These findings constitute direct evidence that aggregation occurs *via* an apo-monomer intermediate. The finding that the apo-monomer accumulated before aggregation shows that the final assembly step is rate-limiting in the reaction Eq. (1). The mechanism in Eq. (1) is minimal since the final step leading to fibril formation is likely to be a complex multi-step process, possibly involving higher-order nuclei and oligomers (40, 41). k_{agg} is an apparent-second order rate constant that is a convolution of these underlying higher-order reactions. Also, $k_{agg} \approx 10^3$ $M^{-1}s^{-1}$ is significantly smaller than that expected for a diffusion-controlled reaction, $k_{diff} = 10^6$-10^9 $M^{-1}s^{-1}$, further suggesting that additional steps that reduce the effective concentration of the aggregating species, such as a rate-limiting conformational change in the monomer or formation of oligomeric nuclei, occur downstream of the apo-monomer formation.

The monomer intermediate in the proposed mechanism was recently observed for oxidized-SOD1 aggregation (42) and for the A4V FALS mutant (43). Also, mutants that cause monomerization of the homotetrameric protein TTR (44) were found to aggregate rapidly and were cytotoxic. The dissociation of multimeric proteins into their monomeric subunits, followed by misfolding and oligomerization, may be a general theme in the amyloidosis of proteins.

The thermodynamics and kinetics of dimer dissociation and SOD1 stability.

The low value of $K_d \approx 1.6\times10^{-10}$ M at pH 7.8 reported here is consistent with previous studies showing the remarkable stability of the SOD1 dimer. The dimer retains its native 32 kDa molecular weight in 8 M urea for 72 hours at 25^{o}C (45) and is active in 4% SDS (46). This stability was also seen in pulse-chase studies showing that the protein has a relatively long (20-78 h) half-life in cells ((47), (48), (12)). The $K_d \approx 1.6\times10^{-10}$M (pH 7.8, 37^{0}C) corresponds to a free energy, $\Delta G_d \approx$ -13.5 kcal/mol, for the dimer dissociation reaction: $D \rightleftharpoons 2M$. Consistent with our estimate, a three-state fit to SOD1 unfolding (32) of the homologous human SOD1 (pH 7.4, 20^{0}C) gave: $D \rightleftharpoons 2M \rightleftharpoons U$; $\Delta G_d \approx$ -12.5 kcal/mol and $\Delta G_m \approx$ -8 kcal/mol, where U is the unfolded state and ΔG_m is the free energy of unfolding of both monomers. The total stability of SOD1 ($\Delta G_d + \Delta G_m$) is approximately 20.5 kcal/mol (32) and the dimer dissociation free energy represents the largest contribution to the stability of the protein. Therefore, conditions that lead to aggregation must destabilize the dimer and/or reduce the dissociation barrier. In accord with this, we found that the rate and equilibrium constants for dimer dissociation are enhanced 30- and 6×10^3-fold at pH 3.5 compared to pH 7.8. These values imply that low pH reduces the dimer dissociation barrier by 2.0 kcal/mol, while the dimer

stability is reduced by 5.2 (=13.5-8.3) kcal/mol (Fig. 4). Thus, the rarity and the late age-of-onset of ALS may occur because under physiological conditions, the K_d and k_{off} are 6,000- and 30-fold smaller, and the H^+-concentration expected to induce metal loss is 10,000-fold smaller, compared to pH 3.5. The effect of factors related to aging, such as the decreased efficacy of chaperone-mediated proteosomal degradation (23), is likely to increase the demetallation and/or the dimer dissociation *in vivo*.

A model for the effect of FALS mutations on SOD1 aggregation

There is no discernable pattern in the primary, secondary or tertiary structure of FALS-associated SOD1 mutants, suggesting that they affect a global property of the protein. We postulate that mutations affect one or more steps in the multi-step pathway for SOD1 aggregation, and this suggests a unified framework to account for the effect of FALS mutants on SOD1 aggregation (Fig. 4). Dimer dissociation generates a monomeric form of SOD1, which loses its metal(s) to generate the aggregation-competent species M_{apo}. We have computationally estimated (Khare et al., unpublished results) that 70 of 75 known FALS-associated point mutants in SOD1 decrease the dimer stability and/or increase dissociation propensity, and propose that mutant forms of SOD1 result in FALS because of one or more of the following: (1) mutations destabilize the SOD1 dimer resulting in higher than normal concentrations of M_{holo} and consequently M_{apo}; (2) mutations enhance the loss of metals from M_{holo}, possibly by affecting the interaction with the copper chaperone CCS, resulting in higher than normal concentrations of M_{apo}; (3) SOD1 mutations reduce the barrier for fibril formation. In all cases, the probability for forming protein aggregates is significantly increased. Important questions concerning the role, if any, of aggregates in the cell-specific biological origin of

FALS i.e. why the aggregates are found only in motor neurons and the origin of cytotoxicity in the disease, remain a formidable challenge.

2.6 References

1. Cleveland, D. W. & Rothstein, J. D. (2001) *Nature Reviews Neuroscience* **2**, 806-819.

2. Ratovitski, T., Corson, L. B., Strain, J., Wong, P., Cleveland, D. W., Culotta, V. C. & Borchelt, D. (2000) *Brain Pathology* **10**, 725.

3. Beckman, J. S., Estevez, A. G. & Crow, J. R. (2001) *Trends in Neurosciences* **24**, S15-S20.

4. Rowland, L. P. & Shneider, N. A. (2001) *New England Journal of Medicine* **344**, 1688-1700.

5. Siddique, T., Nijhawan, D. & Hentati, A. (1996) *Neurology* **47**, S27-S34.

6. Siddique, T., Deng, H. X., Hentati, A., Tandon, R., PericakVance, M. A., Laing, N. G., Roos, R. P., Cayabyab, A., Gerzfeldt, B., Hung, W. Y. *et al.* (1993) *Annals of Neurology* **34**, 303-304.

7. Rosen, D. R., Siddique, T., Patterson, D., Figlewicz, D. A., Sapp, P., Hentati, A., Donaldson, D., Goto, J., Oregan, J. P., Deng, H. X. *et al.* (1993) *Nature* **362**, 59-62.

8. Gurney, M. E., Pu, H. F., Chiu, A. Y., Dalcanto, M. C., Polchow, C. Y., Alexander, D. D., Caliendo, J., Hentati, A., Kwon, Y. W., Deng, H. X. *et al.* (1994) *Science* **264**, 1772-1775.

9. Gaudette, M., Hirano, M. & Siddique, T. (2000) *Amyotrophic Lateral Sclerosis and Other Motor Neuron Disorders* **1**, 83-89.

10. Rakhit, R., Cunningham, P., Furtos-Matei, A., Dahan, S., Qi, X. F., Crow, J. P., Cashman, N. R., Kondejewski, L. H. & Chakrabartty, A. (2002) *J. Biol. Chem.* **277**, 47551-47556.

11. Valentine, J. S. & Hart, P. J. (2003) *Proceedings of the National Academy of Sciences of the United States of America* **100**, 3617-3622.

12. Johnston, J. A., Dalton, M. J., Gurney, M. E. & Kopito, R. R. (2000) *Proceedings of the National Academy of Sciences of the United States of America* **97**, 12571-12576.

13. Bruijn, L. I., Houseweart, M. K., Kato, S., Anderson, K. L., Anderson, S. D., Ohama, E., Reaume, A. G., Scott, R. W. & Cleveland, D. W. (1998) *Science* **281**, 1851-1854.

14. Durham, H. D., Roy, J., Dong, L. & Figlewicz, D. A. (1997) *Journal of Neuropathology and Experimental Neurology* **56**, 523-530.

15. Wang, J., Xu, G. & Borchelt, D. R. (2002) *Neurobiology of Disease* **9**, 139-148.

16. Reaume, A. G., Elliott, J. L., Hoffman, E. K., Kowall, N. W., Ferrante, R. J., Siwek, D. F., Wilcox, H. M., Flood, D. G., Beal, M. F., Brown, R. H. *et al.* (1996) *Nature Genetics* **13**, 43-47.

17. Wang, J., Xu, G. L., Gonzales, V., Coonfield, M., Fromholt, D., Copeland, N. G., Jenkins, N. A. & Borchelt, D. R. (2002) *Neurobiology of Disease* **10**, 128-138.

18. Shibata, N., Hirano, A., Kobayashi, M., Sasaki, S., Kato, T., Matsumoto, S., Shiozawa, Z., Komiri, T., Ikemoto, A., Umahara, T. *et al.* (1994) *Neuroscience Letters* **179**, 149-152.

19. Matsumoto, S., Kusaka, H., Ito, H., Shibata, N., Asayama, T. & Imai, T. (1996) *Clinical Neuropathology* **15**, 41-46.

20. Shibata, N., Asayama, K., Hirano, A. & Kobayashi, M. (1996) *Developmental Neuroscience* **18**, 492-498.

21. Khare, S. D., Ding, F. & Dokholyan, N. V. (2003) *Journal of Molecular Biology* **334**, 515-525.

22. Elam, J. S., Taylor, A. B., Strange, R., Antonyuk, S., Doucette, P. A., Rodriguez, J. A., Hasnain, S. S., Hayward, L. J., Valentine, J. S., Yeates, T. O. *et al.* (2003) *Nature Structural Biology* **10**, 461-467.

23. Okado-Matsumoto, A. & Fridovich, I. (2002) *Proceedings of the National Academy of Sciences of the United States of America* **99**, 9010-9014.

24. DiDonato, M., Craig, L., Huff, M. E., Thayer, M. M., Cardoso, R. M. F., Kassmann, C. J., Lo, T. P., Bruns, C. K., Powers, E. T., Kelly, J. W. *et al.* (2003) *Journal of Molecular Biology* **332**, 601-615.

25. Stathopulos, P. B., Rumfeldt, J. A. O., Scholz, G. A., Irani, R. A., Frey, H. E., Hallewell, R. A., Lepock, J. R. & Meiering, E. M. (2003) *PNAS* 1237797100.

26. Lindberg, M. J., Tibell, L. & Oliveberg, M. (2002) *PNAS* 262527099.

27. Rodriguez, J. A., Valentine, O. S., Eggers, D. K., Roe, J. A., Tiwari, A., Brown, R. H. & Hayward, L. J. (2002) *J. Biol. Chem.* **277**, 15932-15937.

28. Inouye, K., Osaki, A. & Tonomura, B. (1994) *Journal of Biochemistry* **115**, 507-515.

29. Caplow, M. & Fee, L. (2001) *Molecular Biology of the Cell* **12**, 430A.

30. Mccord, J. M. & Fridovich, I. (1969) *J. Biol. Chem.* **244**, 6049-6055.

31. Pantoliano, M. W., Valentine, J. S., Mammone, R. J. & Scholler, D. M. (1982) *Journal of the American Chemical Society* **104**, 1717-1723.

32. Stroppolo, M. E., Malvezzi-Campeggi, F., Mei, G., Rosato, N. & Desideri, A. (2000) *Archives of Biochemistry and Biophysics* **377**, 215-218.

33. Lepock, J. R., Frey, H. E. & Hallewell, R. A. (1990) *J. Biol. Chem.* **265**, 21612-21618.

34. Rodriguez, J. A., Valentine, J. S., Eggers, D. K., Roe, J. A., Brown, R. H. & Hayward, L. J. (2001) *Journal of Inorganic Biochemistry* **86**, 403.

35. Liu, K., Cho, H. S., Lashuel, H. A., Kelly, J. W. & Wemmer, D. E. (2000) *Nature Structural Biology* **7**, 754-757.

36. McParland, V. J., Kalverda, A. P., Homans, S. W. & Radford, S. E. (2002) *Nature Structural Biology* **9**, 326-331.

37. Swietnicki, W., Petersen, R., Gambetti, P. & Surewicz, W. K. (1997) *J. Biol. Chem.* **272**, 27517-27520.

38. Blanch, E. W., Morozova-Roche, L. A., Cochran, D. A. E., Doig, A. J., Hecht, L. & Barron, L. D. (2000) *Journal of Molecular Biology* **301**, 553-563.

39. Goto, J. J., Zhu, H. N., Sanchez, R. J., Nersissian, A., Gralla, E. B., Valentine, J. S. & Cabelli, D. E. (2000) *J. Biol. Chem.* **275**, 1007-1014.

40. Lomakin, A., Teplow, D. B., Kirschner, D. A. & Benedek, G. B. (1997) *PNAS* **94**, 7942-7947.

41. Walsh, D. M., Hartley, D. M., Kusumoto, Y., Fezoui, Y., Condron, M. M., Lomakin, A., Benedek, G. B., Selkoe, D. J. & Teplow, D. B. (1999) *J. Biol. Chem.* **274**, 25945-25952.

42. Rakhit, R., Crow, J. P., Lepock, J. R., Kondejewski, L. H., Cashman, N. R. & Chakrabartty, A. (2004) *J. Biol. Chem.* **279**, 15499-15504.

43. Ray, S. S., Nowak, R. J., Strokovich, K., Brown, R. H., Walz, T. & Lansbury, P. T. (2004) *Biochemistry* **43**, 4899-4905.

44. Reixach, N., Deechongkit, S., Jiang, X., Kelly, J. W. & Buxbaum, J. N. (2004) *Proceedings of the National Academy of Sciences of the United States of America* **101**, 2817-2822.

45. Malinowski, D. P. & Fridovich, I. (1979) *Biochemistry* **18**, 5055-5060.

46. Forman, H. J. & Fridovich, I. (1973) *J. Biol. Chem.* **248**, 2645-26499.

47. Borchelt, D. R., Guarnieri, M., Wong, P. C., Lee, M. K., Slunt, H. S., Xu, Z. S., Sisodia, S. S., Price, D. L. & Cleveland, D. W. (1995) *J. Biol. Chem.* **270**, 3234-3238.

48. Nakano, R., Inuzuka, T., Kikugawa, K., Takahashi, H., Sakimura, K., Fujii, J., Taniguchi, N. & Tsuji, S. (1996) *Neuroscience Letters* **211**, 129-131.

Table 2.1 Measurement of the K_d for SOD1 dimer dissociation using AUC.

Conditions		M_w, Da	Comment
Concentration	pH/incubation		
30μM	pH 7.8	36180	Dimeric
30μM	pH 3.5+ EDTA (4^0C)[a]	16417	Monomeric
30μM	pH 3.5	29402	K_d =0.90μM
20μM	pH 3.5	27481	K_d =1.3μM
10μM	pH 3.5	27425	K_d =0.82μM
30μM	pH 3.5+ 24 h	30269	K_d =1.29μM
10μM	pH 3.5+ 24 h	29969	K_d =1.29μM

[a] Low temperature was used with EDTA in order to avoid aggregation of the apo-monomer.

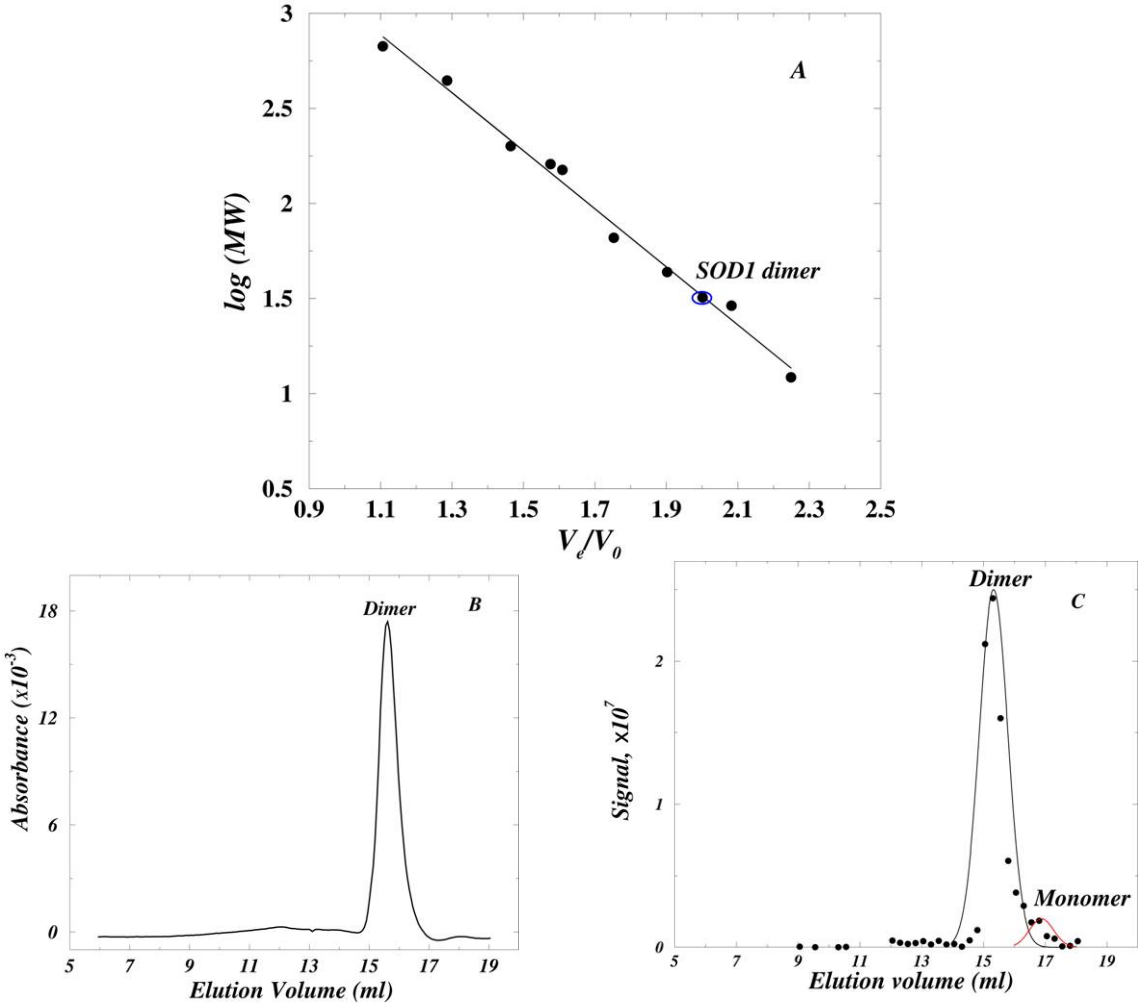


Figure 2.1: Size exclusion profiles of SOD1 at pH 7.8. A: The Calibration curve for the column obtained by using with globular protein standards at pH 7.8 was: log (M_w/1000) = 4.56-1.526·(V_e/V_0), where M_w is the molecular weight, V_e is the elution volume and V_0 is the void volume (7.8 ml) of the column. SOD dimer (32kDa) and monomer (16kDa) are expected to elute at 15.5 and 17.0 ml respectively. **B**: Elution profile at 30μM: SOD1 elutes at 15.6 ml and is therefore dimeric. **C**: Immunostaining dot-blot profile of a 5nM SOD1 sample after 24h of incubation. The peaks are Gaussian fits and correspond to the dimer and monomer respectively.

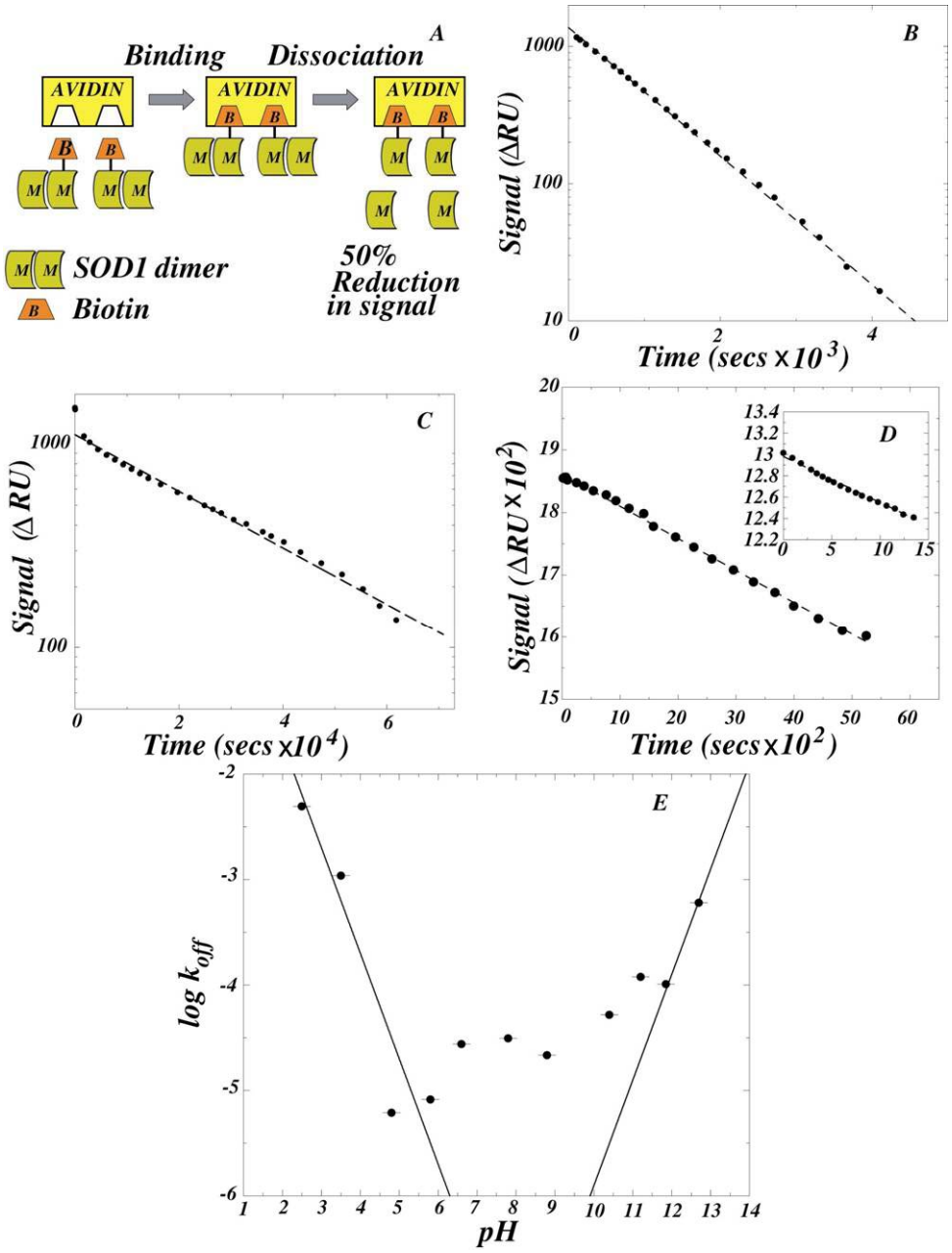


Figure 2.2: The dimer dissociation rates of SOD1 measured by SPR.
A: Scheme for SPR measurement of SOD1 dimer dissociation rate. The rate constant at pH 3.5 was calculated as 1.07×10^{-3} s^{-1} (**B**) and 3.13×10^{-5} s^{-1} at pH 7.8 (**C**), by fitting to a single exponential (dotted lines). The initial rate was followed in the pH range 2.5 to 12.7. **D**: For dissociation at pH 6.6, the rate constant was found to be 2.75×10^{-5} for bovine SOD1 and 1.53×10^{-5} for human SOD1 (inset). **E**: The k_{off} dependence on pH follows a titration curve with apparent pK_a values of <2.5 and >12.5 at the acidic and basic ends of the pH range respectively. The lines have slopes of 1 and -1 (Eq. (2)).

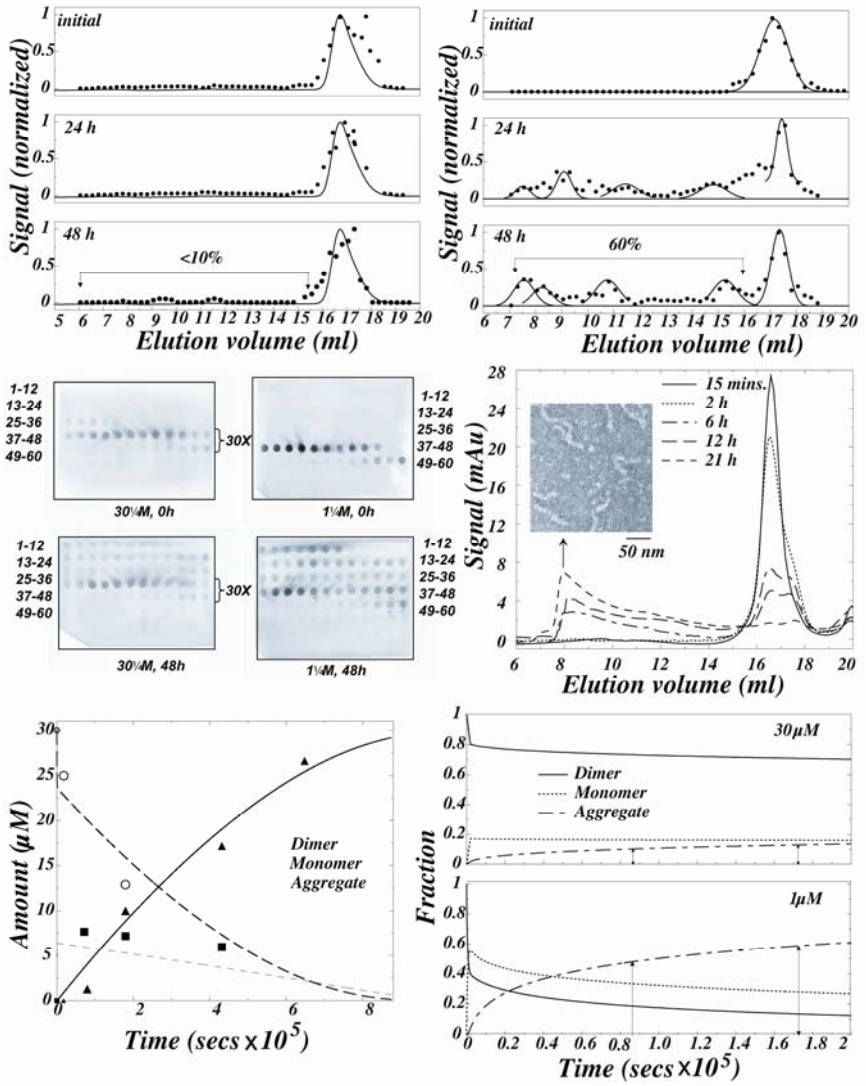


Figure 2.3: Kinetics of aggregation of SOD1.

A: Aggregation at 30µM. SOD1 elution profiles were obtained at initial time and at 24 and 48 hours by immunostaining dot-blots and UV-absorbance (solid lines). The profiles were normalized according to the fraction with the highest signal intensity and the monomer is more pronounced in the dot-blot elution profile. Aggregate is defined as the protein not eluting as dimer or monomer. **B**: Aggregation at 1µM. SOD1 was detected by immunobloting column fractions with Gaussian fits to the spot-intensity profile. **C**: Dot blots showing the elution profile at 30µM and 1µM before and after incubation. The fractions corresponding to the dimer and monomer at 30µM were diluted 30-fold to ensure linearity of signal. **D**: Time course of 30µM SOD1 under dialysis conditions. The inset shows an image of the fibrilar product after 21 h, obtained by negative staining electron microscopy. **E**: Fit of the model in Eq. (3), to the results obtained in D. The continuous curves are obtained from numerical simulations with the optimized rate constants. **F**: Predicted aggregation profiles with 30 and 1 µM using the parameters derived in E. The predicted fraction of aggregates in 30 and 1µM SOD1 after 48 h was 14% and 63% respectively, in close agreement with the experimentally measured values in A and B.

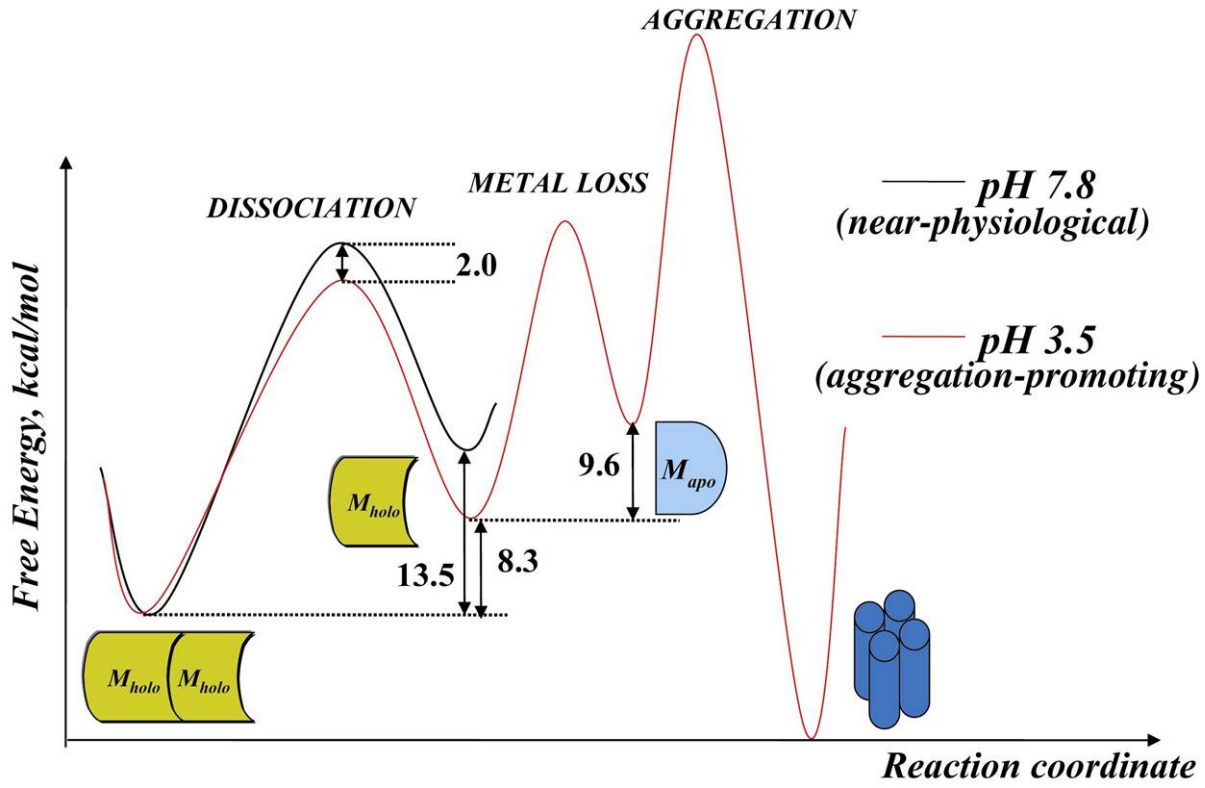


Figure 2.4: A scheme for the aggregation of SOD1. The more rapid aggregation at pH 3.5 compared to pH 7.8 results from the increased K_d and k_{off} values at pH 3.5. Exposure to low pH decreases the stability of SOD1 by 5.2 (=13.5-8.3) kcal/mol, while the dissociation barrier is reduced by approximately 2 kcal/mol. At pH 3.5, the dissociation free energy of Zn from SOD1 monomer, M_{holo}, is approximately 9.6 kcal/mol. In the minimal reaction sequence (Eq. (1)), the apo-monomers, M_{apo}, aggregate to form insoluble fibrils; this is the rate-limiting step for aggregation.

2.7 Supplementary Material

S1: SOD1 dimer dissociation equilibrium detected by Size Exclusion chromatography.

We confirmed that chemical modification by biotinylation of SOD1 does not affect its chromatographic or aggregation behavior (Fig. S1).

S2: Dot blotting.

Column fractions were analyzed by a dot-blot assay, with Avidin-AP, as previously described (1) and/or by immunostaining with anti-SOD1 antibody. At high (>1μM) protein concentrations, column fractions were diluted before blotting. For example, for a 30μM protein sample, 5μl of the column fraction (volume 250μl) was diluted into a volume of 300μl with the flow buffer, and 100μl of the diluted sample was blotted. For immunostaining, the membrane was blocked for 1h with 5% BSA, then incubated for 1h with rabbit anti-SOD1 (Abcam, diluted 4000-fold in PBS-Tween, 3%BSA), followed by another hour of anti-rabbit-AP secondary antibody (Sigma, diluted 40,000-fold in PBS-Tween, 1%BSA). After two 10 min. washes with PBS-Tween (0.1%) and one 10 min. wash with Tris-buffered saline, the membrane was treated with Amersham Pharmacia ECF reagent for 5-20 mins. The resulting signal was detected by a phosphorimager and quantitated with the software ImageQuant®. Background correction was applied by local averaging of the spot intensities and was used as an estimate of the baseline. It was found that while the relative intensity of spots on the dot-blot for a given set of column fractions was highly reproducible, their absolute values varied from blot to blot. Therefore, for comparison of dot-blot elution profiles obtained at different time-points, the spot intensities were normalized by the maximum value detected on a given blot.

S3: **Estimation of rate constants and numerical integration.**

The constants k_m^-, k_m^+ and k_{agg} in Eq. (1) and Eq. (3) were estimated from the data in Figs. 3A-3E. (i) k_m^-: The changes in SOD1 dimer and monomer concentrations under dialysis conditions were used to estimate $k_m^- = 1.2\times10^{-4}$. As shown in Fig. 3E, in the initial 6300 s of the reaction (900 s to 7200 s), D_{holo} decreased from 25 to 21μM. Based on K_d=1μM, M_{holo} is expected to concomitantly decrease from 5 to 4.5μM, whereas the observed total monomer (M_{holo}+M_{apo}) increased from 5 to 8.1μM. Therefore, M_{apo}=8.1-4.5=3.6μM was formed during this period from M_{holo}=5μM in a first-order process: $3.6\times10^{-6} = k_m^- \cdot 5\times10^{-6} \cdot 6300$. (ii) k_{agg}: The apparent second-order rate constant for aggregation, k_{agg} =0.12×10^3 M^{-1} s^{-1}, was estimated from the observed 9μM aggregation in 10800 s (7200 s to 18000 s; Fig. 3E), from a nearly constant M_{apo}=3.6μM during this period: $9\times10^{-6} = 0.5 \cdot k_{agg} \cdot (3.6\times10^{-6})^2 \cdot 10800$. (iii) k_m^+: The value k_m^+=5.2×10^4 $M^{-1}s^{-1}$ was obtained by taking into account the observation (Figs. 3A-B) that aggregation was more rapid with 1μM SOD1 than with 30μM SOD1. With 1μM SOD1, there was 0.6μM aggregation in 24 h, and based on k_{agg}=0.12×10^3 M^{-1} s^{-1}, it was calculated that this occurred from an apo-monomer concentration M_{apo}=0.28μM. This means that the equilibrium composition of 1μM SOD1 was D_{holo}=0.39μM, M_{holo}=0.33μM, and M_{apo}=0.28μM, yielding an equilibrium constant for Zn-loss $K_m = \frac{M_{apo} \cdot Zn}{M_{holo}}$ =2.3×10^{-7} M; this value and k_m^-= 1.2×10^{-4} s^{-1} were used to estimate k_m^+. These values were used for numerical simulation (Supplementary Material S3).

The differential master equations corresponding to the aggregation reaction under dialysis conditions are:

$$\begin{cases} \dot{D}_{holo} = -k_{off} D_{holo} + \frac{1}{2} k_{on} M_{holo}^2 \\ \dot{M}_{holo} = 2k_{off} D_{holo} - k_{on} M_{holo}^2 - k_m^- M_{holo} + k_m^+ M_{apo} Zn \\ \dot{M}_{apo} = k_m^- M_{holo} - k_m^+ M_{apo} Zn - k_{agg} M_{apo}^2 \\ \dot{A} = \frac{1}{2} k_{agg} M_{apo}^2 \\ \dot{Zn} = -k_{dialysis} Zn - k_m^+ M_{apo} Zn + k_m^- M_{holo} \\ \left[A + \frac{M_{holo}}{2} + \frac{M_{apo}}{2} + \frac{Zn}{2} + D_{holo} \right](t) = D_{holo}(0) \end{cases}$$

where the symbols are as defined in Eq. (3). In the absence of dialysis, the rate of loss of Zn, $\dot{Zn} = 0$. These equations were solved numerically and a time step of 10^{-2} was sufficient for the convergence of the solution.

S4: **SOD1 dimer dissociation under aggregation-promoting conditions is reversible.**

The reversibility of dimer dissociation at pH 3.5 was evaluated by SEC. A SOD1 sample at 6.5μM eluted as a mixture of approximately 73% dimer and 27% monomer (Fig. S2B, black curve). When this sample was concentrated approximately 4-fold (final concentration 24.1μM), it eluted as a mixture of approximately 87% dimer and 13% monomer (Fig. S2B, blue curve). The fraction of protein eluting as monomer reduced 14% upon concentration, showing that the dimer was re-formed from the monomer. Thus, SOD1 dimer dissociation at pH 3.5 is reversible.

S5: Arg79-Asp101 salt-linkage.

We propose that the salt-linkage governing dimer dissociation rate is between Arg79 and Asp101 (numbering corresponding to hSOD1) based on the observations that: (i) a structure-based alignment of 76 SOD1 sequences obtained from the FSSP database (2) showed

that this salt-linkage Arg79-Asp101 is 100% evolutionarily conserved in the eukaryotic SOD1 fold (Fig. S3), while being absent in prokaryotic SOD1, which has a structurally distinct dimer interface, (ii) Arg79 and Asp101 are located in the Zn-loop and an edge-strand of the β-barrel of SOD1 respectively and their interaction is crucial for the stability and the folding kinetics of SOD1 (3). It is likely that the titration of the Arg79-Asp101 salt-linkage promotes greater disorder in the Zn-binding region, thereby disrupting its interactions with the β-barrel and the disorder is propagated in the molecule leading to dimer dissociation. The mutation of Asp101 to Gly or Asn is implicated in FALS (http://www.alsod.org) and the disordered Zn-loop of metal-deficient SOD1 has previously been found to be a site for SOD1 oligomerization (4). Therefore, we propose that salt-linkages in SOD1 are crucial for maintaining SOD1 dimer structural integrity, and that the disruption of these charge-charge interactions promotes misfolding into an aggregation-competent state.

References

1. Caplow, M. & Fee, L. (2001) *Molecular Biology of the Cell* **12**, 430A.

2. Holm, L. & Sander, C. (1996) *Nucl. Acids. Res.* **24**, 206-209.

3. Khare, S. D., Ding, F. & Dokholyan, N. V. (2003) *Journal of Molecular Biology* **334**, 515-525.

4. Elam, J. S., Taylor, A. B., Strange, R., Antonyuk, S., Doucette, P. A., Rodriguez, J. A., Hasnain, S. S., Hayward, L. J., Valentine, J. S., Yeates, T. O. *et al.* (2003) *Nature Structural Biology* **10**, 461-467.

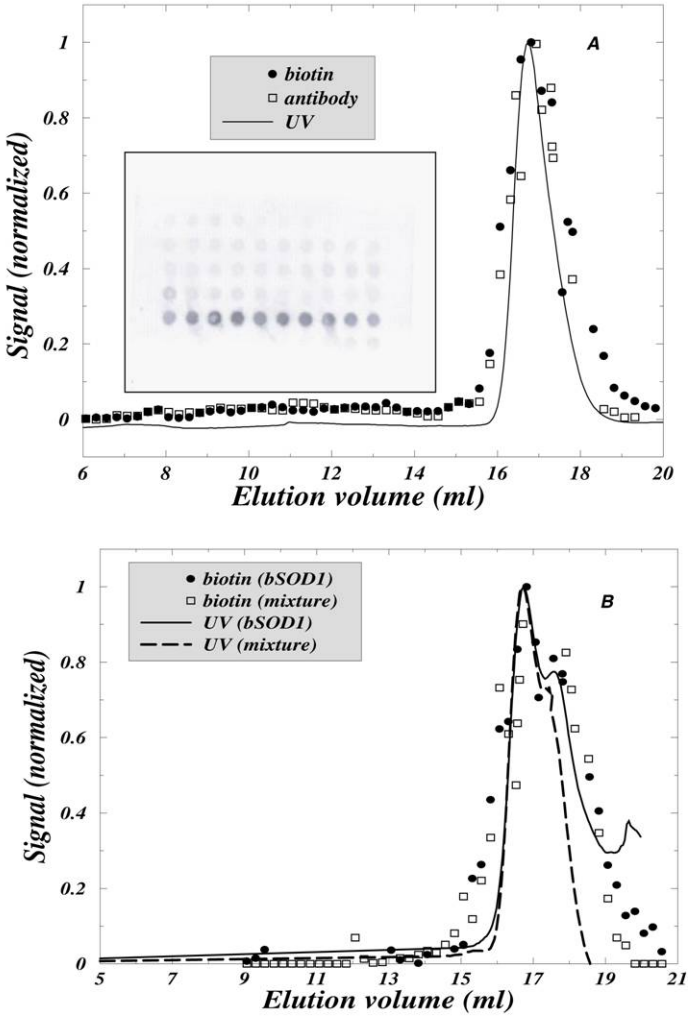


Figure S1: Size exclusion profiles of SOD1 and biotinylated SOD1 A: Size exclusion profiles of SOD1 sample (30μM, pH 3.5) a 1:4 mixture of the biontinylated and unmodified protein (bSOD1-SOD1), obtained by UV-absorbance and two dot-blotting methods: antibody to SOD1, and biotin detection; a typical dot-blot for this sample, stained by anti-SOD1 antibody, is shown in the inset. **B**: Size exclusion profiles of pure bSOD1 and a bSOD1-SOD1 under identical conditions (5μM, pH 3.5, 20 minute incubation) showing that biotinylated SOD1 has identical chromatographic properties as the unmodified SOD1.

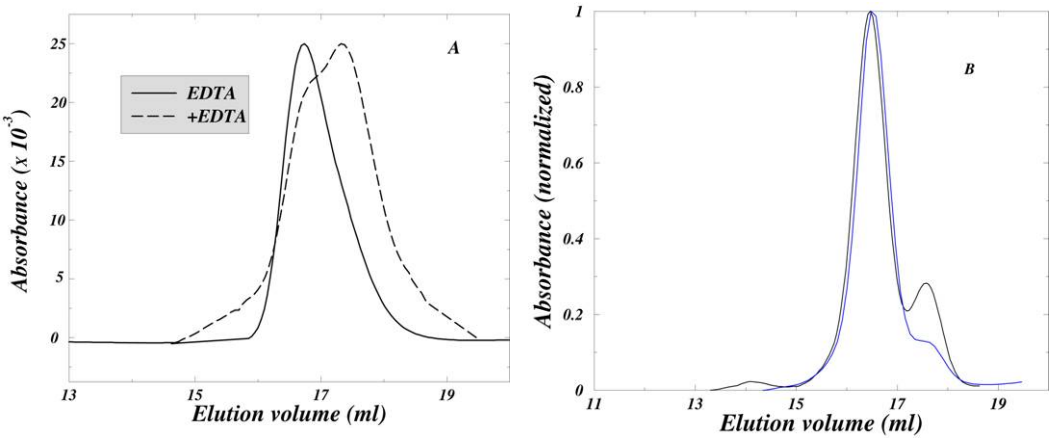


Figure S2: A. Column calibration at pH 3.5. 30μM SOD1 at pH 3.5 eluted at 16.5 ml and 17.3 ml, before and after incubation with 10mM EDTA. These samples were known to be primarily dimeric and primarily monomeric respectively, from AUC (Table I). Calibration at pH 3.5 was done as above because a set of globular protein standards at pH 3.5 was not available. **B.** The reversibility of SOD1 dimer dissociation. The elution profiles correspond to 6.5μM SOD1 (solid black curve), and 24.1μM SOD1 (solid blue curve) which was obtained by concentrating the 6.5μM sample. The fraction of monomer reduced upon concentration, indicating that dimer was formed from the monomer.

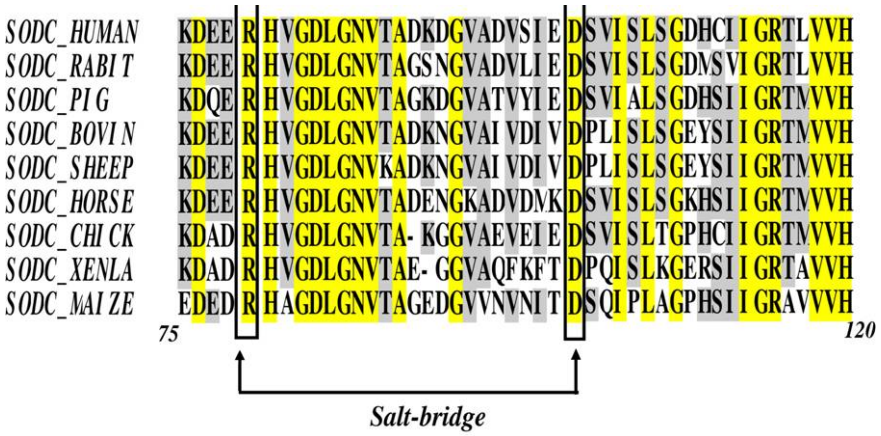


Figure S3: Sequence alignment of SOD1 sequences. A portion of the sequence alignment of 76 eukaryotic SOD1 sequences obtained from the FSSP database showing the conserved salt-linkage between Arg79 and Asp101. Positions with 100% and >50% conservation have yellow and grey backgrounds, respectively.

CHAPTER 3

ESTIMATION OF THE EFFECT OF MUTATION ON SOD1 STABILITY

This chapter was submitted as a part of a manuscript to *The Journal of Biological Chemistry*, and is formatted in the journal style.

3.1 Abstract

Mutations in the dimeric enzyme Cu, Zn superoxide dismutase (SOD1) leading to its aggregation are implicated in the toxicity in Familial Amyotrophic Lateral Sclerosis (FALS). We and others have previously shown that aggregation occurs by a pathway involving dimer dissociation, metal-loss from monomers and multimeric assembly of apo-SOD1 monomers. We postulate that FALS mutations cause enhanced aggregation by affecting one or more steps in the pathway, and computationally test this postulate for 75 known mis-sense FALS mutants of SOD1. Based on an extensive thermodynamic analysis of the stability of apo-dimer and apo-monomer forms of these mutants, we classify the mutations into the following groups: 70 out of 75 mutations in SOD1 lead to (i) decreased dimer stability, and/or (ii) increased dimer dissociation, compared to wild type, and 4 mutations lead to (iii) decreased monomer stability compared to wild type. Our results suggest that enhanced aggregation of mutant SOD1 in FALS occurs due to an increased population of mutant SOD1 apo-monomers compared to wild type.

3.2 Introduction

Point mutations in the cytoplasmic homodimeric enzyme Cu, Zn superoxide dismutase have been identified as the primary cause of approximately 20% cases of the disease Familial ALS (1, 2). More than 90 distinct and structurally diverse FALS mutations are known (3). The sporadic form of ALS (SALS), which is not associated with any heritable defect, constitutes ≈90% of all ALS cases and is symptomatically identical to SOD1-linked FALS. It is likely that understanding the molecular basis of SOD1-linked FALS will also provide insights for understanding SALS.

There is increasing evidence that the FALS does not arise due to a reduction in mutant SOD1 activity, and the motor neuron-specific toxic gain-of-function of the mutants is associated with intracellular aggregation, trafficking and/or degradation of misfolded SOD1. The mutation-induced destabilization of the native state of SOD1 may lead to a relatively higher population of misfolded or partially folded protein molecules that may (i) assemble into structured fibrils, (ii) interfere with mitochondrial transport, and/or (iii) saturate the protein degradation machinery, thereby leading to toxicity. Thus, the misfolding-induced aggregation of mutant SOD1 leading to the appearance of inclusion bodies is correlated with toxicity, and the inhibition of aggregation by over-expression of chaperones leads to increased cell viability.

The molecular mechanism of misfolding and aggregation of SOD1 is not well-understood. We and others have previously provided evidence that the aggregation of SOD1 occurs via a pathway involving dissociation of the dimer and metal-loss, followed by multimeric assembly of the apo-monomeric SOD1. Mutations are likely to lead to enhanced aggregation by affecting one or more steps in the pathway. *In vivo*, FALS mutant SOD1s

have decreased solubility and half-life compared to wild type, indicating that they form less stable dimers. Also, SOD1 solvent denaturation occurs via a monomeric intermediate (4, 5), this monomeric intermediate was also detected in the oxidation-induced aggregation of wild type SOD1, and preventing dimer dissociation by engineering an inter-subunit disulfide bond abolished the aggregation of the A4V FALS SOD1 mutant. Evidence that the apo-form of mutant SOD1 is involved in FALS-linked pathology is (i) the *in vitro* stability of the mutant apo-dimers was correlated with the mean survival time of FALS patients carrying the mutation (6), (ii) treatment of ALS patients with chelators such as penicillamine, EDTA, and thioctic acid was found to sometimes accelerate the disease, (iii) the *in vitro* aggregation rate of A4V and H43R SOD1 was enhanced in the presence of EDTA, and (iii) some recombinant mutant SOD1s were found to have smaller metal content and stability than wild type. Taken together, these studies suggest a general framework for the effect of mutation on FALS-linked aggregation of SOD1, in which mutations cause enhanced aggregation by decreasing apo-SOD1 dimer and monomer stability and/or increasing dimer dissociation, compared to the wild type. An exhaustive analysis of FALS mutants is required to test this general framework. In spite of the recent significant advances(6-8), the challenging task of uncovering the effect of mutation on dimer-monomer equilibrium of apo-SOD1 is likely to remain experimentally unfeasible on the large scale required to ascertain the generality of the suggested framework.

Here, we perform exhaustive *in silico* mutagenesis of SOD1 and evaluate the dimer stabilities and dissociation propensities for 75 FALS-associated point mutants in SOD1 using molecular dynamic (MD) simulations and an implicit-solvation based scoring function. The effect of mutation on SOD1 dimer stability is evaluated as the difference in the calculated

conformational free energy between mutant and wild type SOD1, and the dissociation propensity is evaluated as the difference between the conformational free energies of the SOD1 dimer and two monomers. Based on these calculations, we classify mutants into four groups depending upon how mutation affects dimer stability and dimer dissociation. We find that 70 out of 75 FALS mutants decrease dimer stability and/or increase dimer dissociation, while four of the remaining five lead to decreased monomer stability, suggesting that FALS-linked aggregation may be mediated by the apo-state of mutant SOD1.

3.3 Methods

Generating mutant SOD1 structures

The available structures of mutant SOD1 molecules show remarkable similarity and are almost identical to the wild-type (RMSD≤1.25Å); the difference between wild-type and mutant SOD1 molecules is largely due to side-chain packing(9-12). Thus, we generate mutant structures by *in silico* mutation using the package Sybyl (Tripos Inc.) and the wild-type SOD1 structure (PDB accession code: 1SPD) keeping the backbone co-ordinates fixed and replacing the appropriate side-chains. This is followed by 200 steps of steepest descent energy minimization in the vicinity of mutation to remove steric clashes, while the rest of the protein is rigid.

Estimating Conformational Free Energy

We determine the conformational free energy of a structure by the Explicit Solvent Implicit Solvent (ES/IS) method as described by Vorobjev et al.(13-16). Applied to SOD1, the method consists of a short molecular dynamics simulation (196 ps) of SOD1 (wild-type

or mutant) with explicit solvent (ES) to obtain microstates of the system. The conformational free energy of the SOD1 structure is written as a sum of six contributions:

$$G = \langle U_{m,sh} \rangle + \langle U_{m,coul} \rangle - TS_{conf} + \langle G_{cav} \rangle + \langle G_{s,vdw} \rangle + \langle G_{pol} \rangle, \quad (1)$$

where $\langle \ldots \rangle$ denotes average over the MD trajectory, U_m is the intra-protein conformational energy which can be written as the sum of short-range, $U_{m,sh}$, and coulombic, $U_{m,coul}$, terms, S_{conf} is the determined conformational entropy, and T is the absolute temperature. The solvation free energy, G_{sol}, is a sum of the last three terms : the free energy of creating the empty protein-sized cavity in the solvent G_{cav}, the free energy of inserting the protein molecule into the protein-sized cavity $G_{s,vdw}$, and the free energy of solvent polarization G_{pol}. $\langle U_m \rangle$ and $\langle G_{s,vdw} \rangle$ is accumulated as averages during the molecular dynamics simulation. S_{conf} is calculated from the position covariance matrix of the protein during the dynamics trajectory in the quasi-harmonic approximation. G_{cav} and G_{pol} is calculated with models in which the solvent is treated implicitly, as described by Vorobjev and Hermans(16). Thus, this gives a physically realistic energy function that effectively discriminates between wild-type and mutant SOD1 structures.

Estimation of the effect of a mutation on SOD1 stability requires knowledge of the free energy of the reference (unfolded) states of the wild type G_U^{WT} and mutant G_U^X molecules. The stability of the wild type or a mutant SOD1 is then determined by the following:

$$\Delta G^{WT} = G_N^{WT} - G_U^{WT}, \quad (2)$$

and

$$\Delta G^X = G_N^X - G_U^X \,. \tag{3}$$

We thus determine the change in SOD1 stability due to a mutation:

$$\Delta\Delta G^X = \Delta G^X - \Delta G^{WT} = \left(-G_N^{WT} + G_N^X\right) + \left(G_U^{WT} - G_U^X\right) \tag{4}$$

We assume that the reference (unfolded) state can be modeled as a fully solvated chain. We determine these by computing the conformational free energy G_U of the small peptides Ace-$(Ala)_2$-Gly-Ala-x-Ala-Gly-$(Ala)_2$-NMet using the ES/IS method, where x denotes any of 20 amino acids (Supplementary material Table SI). Thus, we calculate the conformational free energy difference between wild-type and mutant SOD1 for monomeric $\Delta\Delta G_m^X$, dimeric $\Delta\Delta G_d^X$.

The dimer dissociation is characterized by comparing the free energies of monomers and the dimer:

$$\Delta G_{dissoc} \approx -(G_{dim} - 2G_{mono}) \tag{5}$$

where ΔG_{dissoc} is the free energy of dimer dissociation, G_{dim} and G_{mono} are the ES/IS free energies of the dimer and monomer respectively.

All molecular dynamic (MD) calculations are performed using the Sigma molecular dynamics program(17) and the CEDAR force field using a protocol involving relaxation and production, as described elsewhere(18)

Scaling of free energy values and error-bars

We scale the free-energy values obtained according to experimental values(5) of the measured free energy of unfolding of SOD1. The free energy of dimer dissociation has been

measured to be 12.5 kcal/mol (5) for the wild-type. We estimate it to be 93.8 kcal/mol from our calculations. This difference is likely to arise due to (i) systematic errors in the force-field, (ii) limited sampling (iii) the difficulties in modeling the unfolded state. The differences in the stabilities are relative to the wild type and we expect that these systematic errors will be compensated. Therefore, we scale the calculated free energies by a factor of 12.5/93.8=0.13. The uncertainty (error-bar) in the estimated free energy is ideally obtained by performing multiple simulations and calculating the dispersion of the average values thus obtained. For the D90A mutant, the average free energies obtained in two independent simulations are -7353.5 and -7332.0 kcal/mol (difference = 21.5 kcal/mol). The (unscaled) difference between the D90A mutant and wild type is 175.8 kcal/mol, and therefore we estimate the values reported have an intrinsic uncertainty of approximately 15%.

3.4 Results & Discussion

MD trajectories

Molecular dynamic simulations of mutant SOD1 structures generated from the wild type are performed for 196 ps and snap shots are recorded every 2 ps yielding a total of 98 snapshots, averaged over which the conformational free energy of the mutant is calculated. All the SOD1 MD trajectories are stable in our simulations and sample conformations that are near-native (RMSD < 2.5 Å). For four representative mutant dimers and monomers, RMSD *versus* time is shown in Fig 1(a). We find that the monomers have a larger RMSD, which is expected since monomerization exposes a large hydrophobic patch on the surface of SOD1 to the solvent. However, both the dimers and monomers are stable on the time-scale of the simulation (200 ps). This is in accord with the observation that the SOD1 unfolding

occurs *via* a monomeric intermediate, and the stability of the monomer is ~4 kcal/mol, thus the monomers are expected to be stable on a picosecond timescale.

Conformational free energy calculations and classification of mutants

Based on a free-energy cycle thermodynamic analysis (Methods) we estimate the change in free energy change ($\Delta\Delta G_{dim} = \Delta G_{\mathrm{dim}}^{mut} - \Delta G_{\mathrm{dim}}^{WT}$) for unfolding between the mutant and Wild-type dimer structures. $\Delta\Delta G_{dim}$ is a measure of the tendency of the mutant dimer to unfold relative to the wild type. The effect of mutation on the stability can be construed as arising from both effects on intra-protein interactions, and protein-water interactions (solvation). We find that in mutations such as D90A and G93A (Fig. 1(b)), the destabilizing effect of mutation is due to the decreased solvation energy relative to wild type(shown as red curves in Fig 1(b)); the intra-protein interactions (in vacuum) are not significantly affected. For other mutants such as H43R and I112T, which stabilize the dimer, the intra-protein interaction energy is higher than the wild type, whereas the solvation free energy is lower or of similar magnitude (Fig. 1(b)). This is in accord with the experimental finding that SOD1 FALS mutants have lower solubility *in vivo* than the wild type protein. We find that the differences between mutant and wild type dimers are statistically significant and we obtain distinct distributions in the Kolomogorov-Smirnov test (Fig. 1(c)).

We evaluate the tendency of the mutant SOD1 to dissociate into its monomeric subunits, which we estimate by change in dissociation free energy change ($\Delta\Delta G_{dissoc}$) between mutant and wild type. The more negative this value, the higher the dissociation tendency and the higher is the chance that the mutant forms the monomeric state. This apo-monomer, in turn, is likely to form misfolded/unfolded structures that aggregate. Thus, we

characterize each mutant by its $\Delta\Delta G_{dim}$ and $\Delta\Delta G_{dissoc}$ values (Fig. 1(d)) and classify the mutants into four different classes based on their effect on dimer stability and dissociation relative to wild type (Fig. 1(d) and Supplementary material Table SII): Class I (lower dimer stability, increased dimer dissociation, e.g. D90A), Class II (higher dimer stability, increased dimer dissociation e.g. H43R), Class III (lower dimer stability, decreased dimer dissociation e.g. G93A), and Class IV (higher dimer stability, decreased dimer dissociation e.g. I112T), mutants respectively. In total, all but 15 out of the 75 mutants destabilize the dimer, while all but 10 are more prone to dissociate than the wild type. 5 of the 75 mutants marginally stabilize the dimer as well as increase dimer dissociation propensity, but the monomers of these mutants have lower stability than the wild type monomer. Therefore, we conclude that FALS mutations in SOD1 may induce aggregation by one or more of the following: (i) decreasing dimer stability, (ii) decreasing monomer stability, and (iii) increasing dimer dissociation propensity. A mutation belonging to each class, affects the dimer-monomer equilibrium such that the overall equilibrium is shifted towards the apo-monomer, and thus increases the population of the aggregation-prone species.

Comparison with experiments

We compare our calculations with the experiments of Rodriguez et al. (Fig. 2) and Lindeberg et al. Rodriguez et al. report 4 peaks in the DSC scans of the SOD1 mutants, and the peak with the lowest melting temperature (T_m) corresponds to the metal-free SOD1 species. We compare trends in the calculated $\Delta\Delta G_{dim}$ values with the ΔT_m values of the corresponding 15 FALS mutants and find excellent agreement. We correctly identify the mutations that do not have any effect on stability (H48Q), as well as those which have a large

effect on stability, thus covering a range of stability values. Similarly, for the data of Lindeberg et al., we find good qualitative agreement between the estimated free energy values, and the change in mid-point concentration of urea, for the melting curve obtained in their experiments. This indicates that the simulation is successful in estimating the thermodynamic effect of mutation. Our simulations rely on the assumption that there is no major structure change upon mutation, and in accord with this assumption, we find that the simulated structures are within 1.8 Å RMSD, of the experimentally determined FALS mutants (Fig. 3). Thus, our simulations agree with experimental data in terms of both the thermodynamic and the structural effects of mutation.

3.5 Conclusions

We demonstrate that the effect of mutations on SOD1 can be characterized in terms of the relative stabilities and dimer dissociation constants of mutant proteins compared to wild type. We characterize all the known SOD1 point mutants into classes based on this scheme. We find that 70 out of 75 mutants modulate aggregation propensity by affecting the stability and/or dimer dissociation in a way that makes either the mutant dimer less stable and/or more prone to dissociate. A graph theoretic analysis of SOD1 interactions reveals that perturbations in the protein caused by mutations are likely to be effectively transmitted to both the dimer interface and metal-coordinating centers. Thus, dimer dissociation and/or metal-loss may provide the intervention points to prevent the aggregation and consequent toxicity of SOD1.

3.6 References

1. Rosen, D. R., Siddique, T., Patterson, D., Figlewicz, D. A., Sapp, P., Hentati, A., Donaldson, D., Goto, J., Oregan, J. P., Deng, H. X. *et al.* (1993) *Nature* **362**, 59-62.

2. Gurney, M. E., Pu, H. F., Chiu, A. Y., Dalcanto, M. C., Polchow, C. Y., Alexander, D. D., Caliendo, J., Hentati, A., Kwon, Y. W., Deng, H. X. *et al.* (1994) *Science* **264**, 1772-1775.

3. Gaudette, M., Hirano, M. & Siddique, T. (2000) *Amyotrophic Lateral Sclerosis and Other Motor Neuron Disorders* **1**, 83-89.

4. Mei, G., Rosato, N., Silva, N., Rusch, R., Gratton, E., Savini, I. & FinazziAgro, A. (1992) *Biochemistry* **31**, 7224-7230.

5. Stroppolo, M. E., Malvezzi-Campeggi, F., Mei, G., Rosato, N. & Desideri, A. (2000) *Archives of Biochemistry and Biophysics* **377**, 215-218.

6. Lindberg, M. J., Tibell, L. & Oliveberg, M. (2002) *PNAS* 262527099.

7. Rakhit, R., Crow, J. P., Lepock, J. R., Kondejewski, L. H., Cashman, N. R. & Chakrabartty, A. (2004) *J. Biol. Chem.* **279**, 15499-15504.

8. DiDonato, M., Craig, L., Huff, M. E., Thayer, M. M., Cardoso, R. M. F., Kassmann, C. J., Lo, T. P., Bruns, C. K., Powers, E. T., Kelly, J. W. *et al.* (2003) *Journal of Molecular Biology* **332**, 601-615.

9. Ferraroni, M., Rypniewski, W., Wilson, K. S., Viezzoli, M. S., Banci, L., Bertini, I. & Mangani, S. (1999) *Journal of Molecular Biology* **288**, 413-426.

10. Parge, H. E., Hallewell, R. A. & Tainer, J. A. (1992) *Proceedings of the National Academy of Sciences of the United States of America* **89**, 6109-6113.

11. Mcree, D. E., Redford, S. M., Getzoff, E. D., Lepock, J. R., Hallewell, R. A. & Tainer, J. A. (1990) *J. Biol. Chem.* **265**, 14234-14241.

12. Hart, P. J., Liu, H. B., Pellegrini, M., Nersissian, A. M., Gralla, E. B., Valentine, J. S. & Eisenberg, D. (1998) *Protein Sci* **7**, 545-555.

13. Vorobjev, Y. N. & Hermans, J. (2001) *Protein Sci* **10**, 2498-2506.

14. Vorobjev, Y. N. & Hermans, J. (1999) *Biophysical Chemistry* **78**, 195-205.

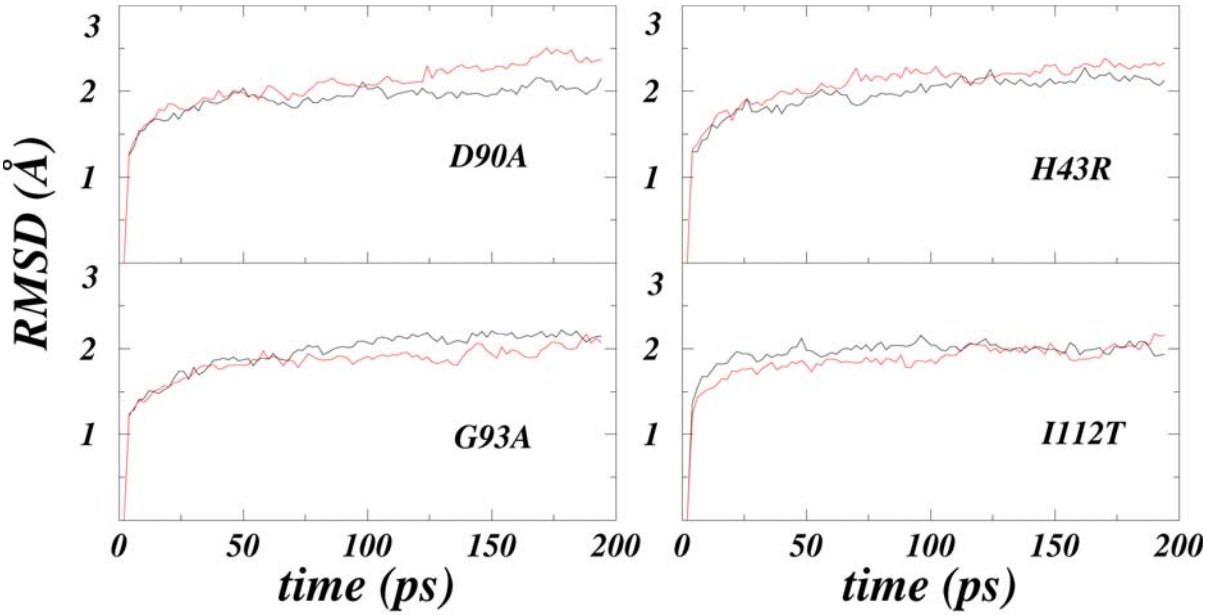
15. Vorobjev, Y. N., Almagro, J. C. & Hermans, J. (1998) *Proteins-Structure Function and Genetics* **32**, 399-413.

16. Vorobjev, Y. N. & Hermans, J. (2002) *Protein Sci* **11**, 994.

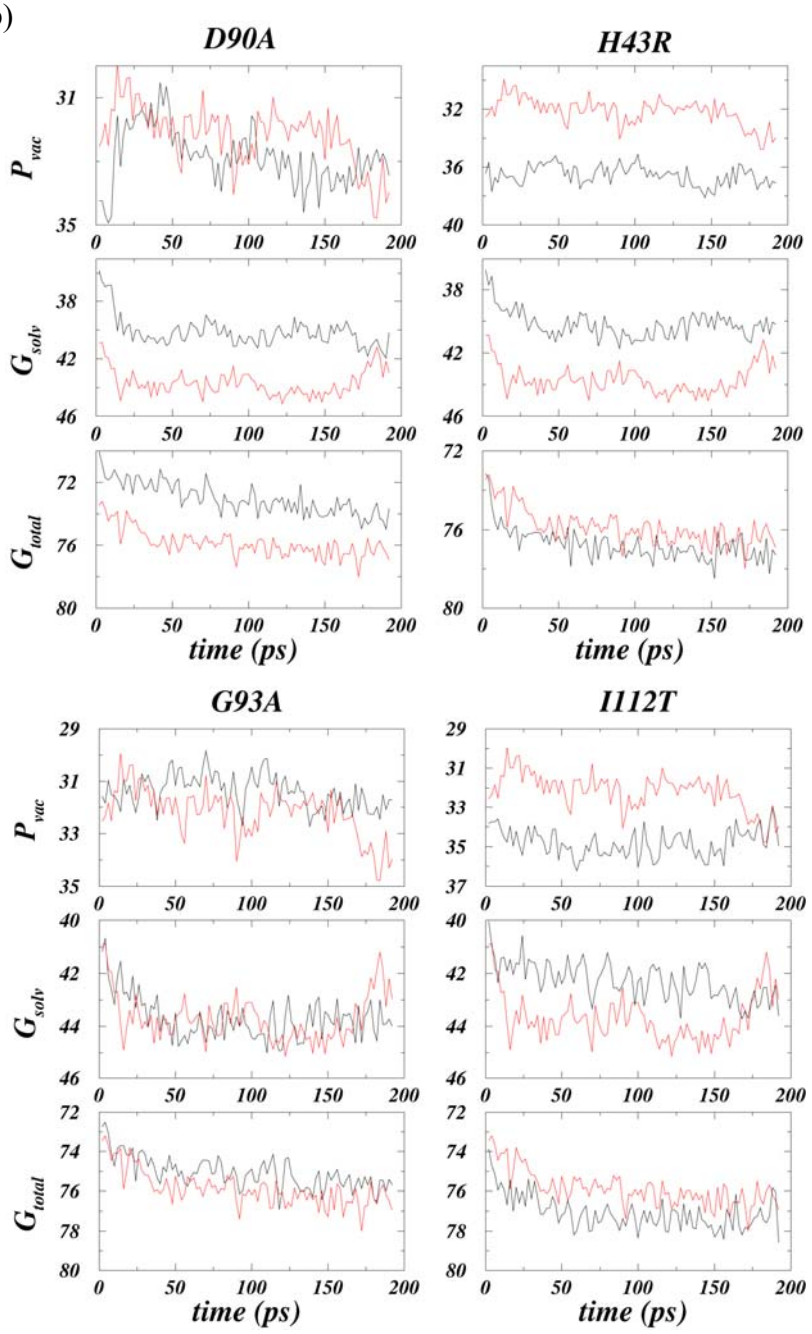
17. Mann, G., Yun, R. H., Nyland, L., Prins, J., Board, J. & Hermans, J. The Sigma MD program and a generic interface applicable to multi-functional programs with complex, hierarchical command structure. eds. Schlick, T. & Gan, H. H. 129-145. 2002. New York, Springer-Verlag. Computational Methods for Macromolecules: Challenges and Applications - Proceedings of the 3rd International Workshop on Algorithms for Macromolecular Modelling. 10-12-2000.

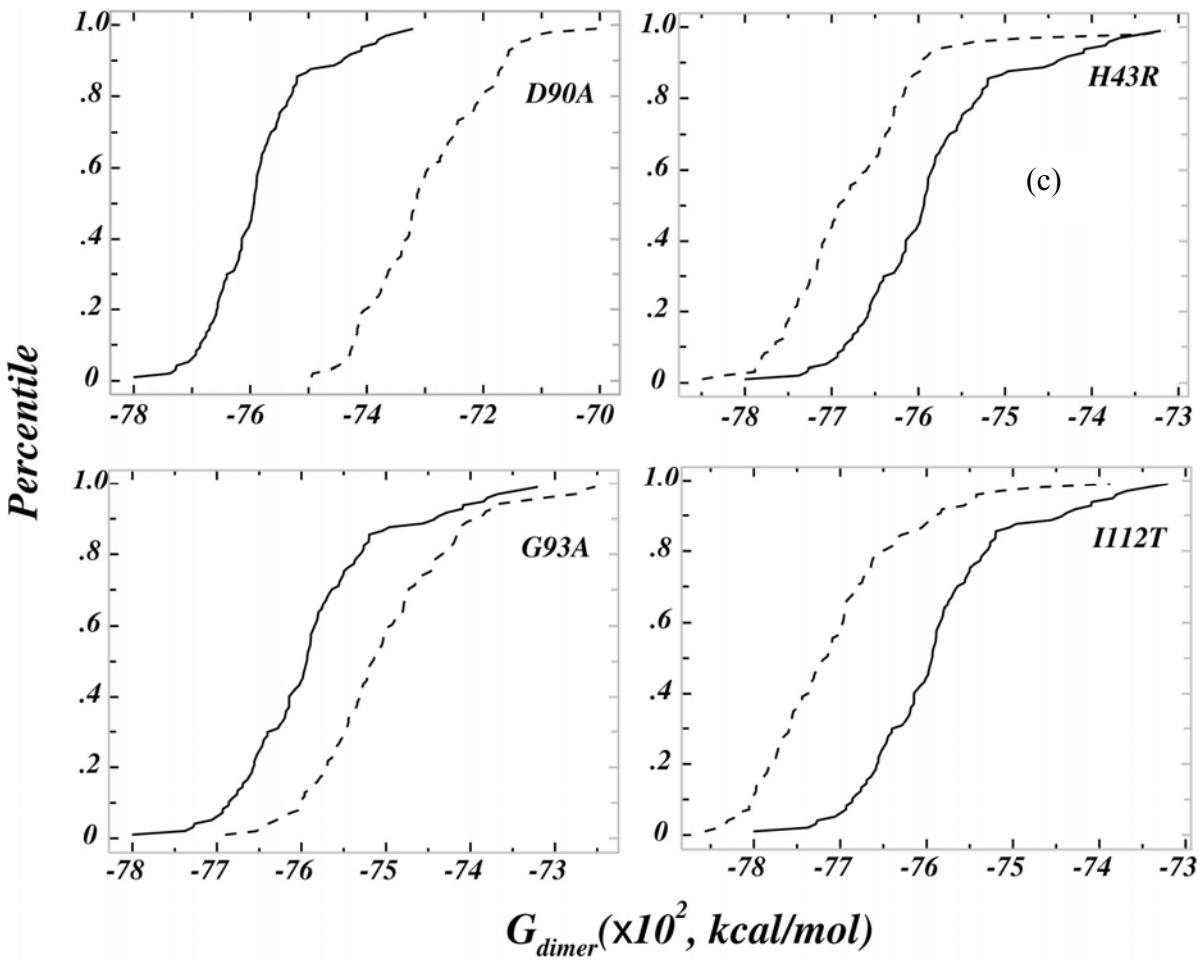
18. Urbanc, B., Cruz, L., Ding, F., Sammond, D., Khare, S., Buldyrev, S., Stanley, H. E. & Dokholyan, N. V. (2004) *Biophys. J.*

(a)



(b)


D90A
H43R
G93A
I112T
P vac
G solv
G total
time (ps)
0
50
100
150
200


Percentile
D90A
H43R
(c)
G93A
I112T
1.0
.8
.6
.4
.2
0
-78
-76
-74
-72
-70
-77
-75
-73
G_{dimer}(x10^2, kcal/mol)

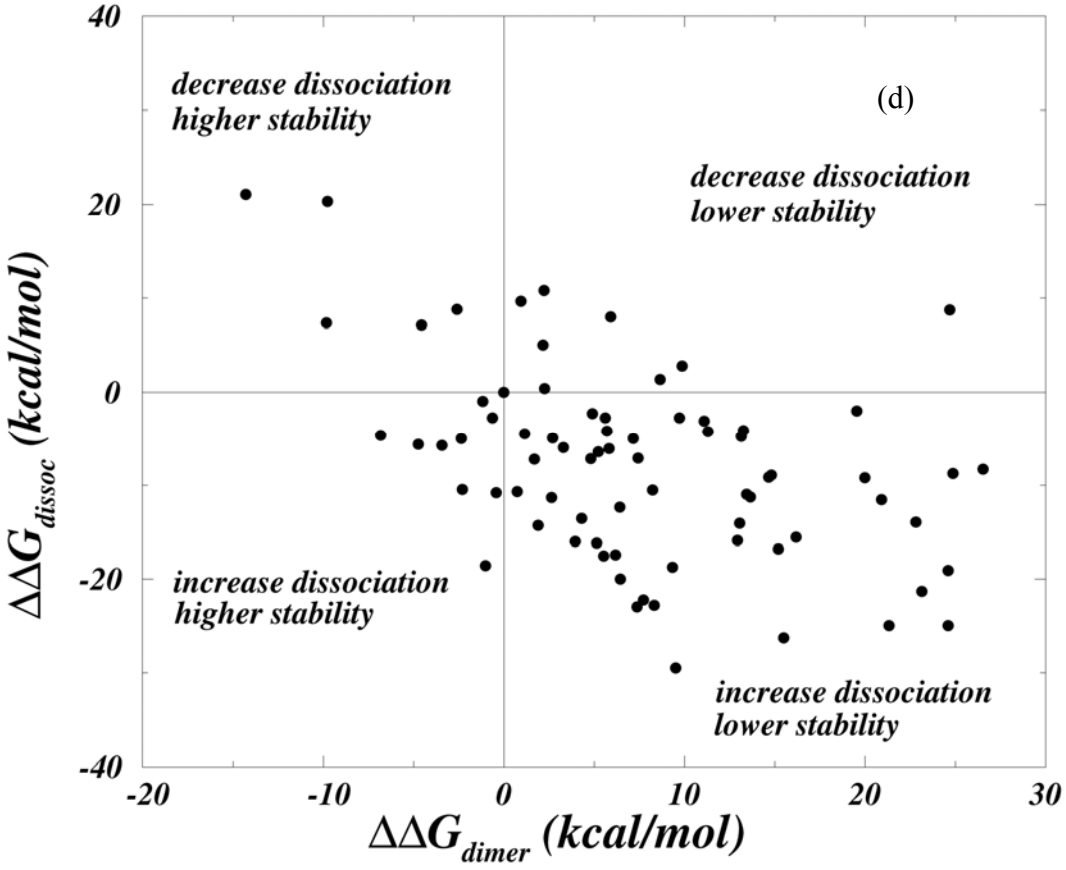


Figure 3.1 Conformational free energies of 75 FALS point mutants. (a) The C_α-RMSD vs. time for four representative mutant trajectories. The dimer is in black and monomer in red. (d) The internal energy, solvation free energy and total conformational free energy for the four mutants in (a). For comparison, the wild type energies are shown in red. (c) Distributions of free energy values obtained in the Kolmogorov-Smirnov test, showing that the differences observed between the wild type (solid) and mutant (dashed) are significant. The differences between mutant and wild type dimers are statistically significant [42]. (d) The effect of mutation on SOD1 dimer stability and dissociation for all 75 FALS mutants considered in this study. 70 out of 75 mutations lower dimer stability and/or increase dimer dissociation, whereas 5 out of 75 lead to marginally higher stability and decreased dimer dissociation.

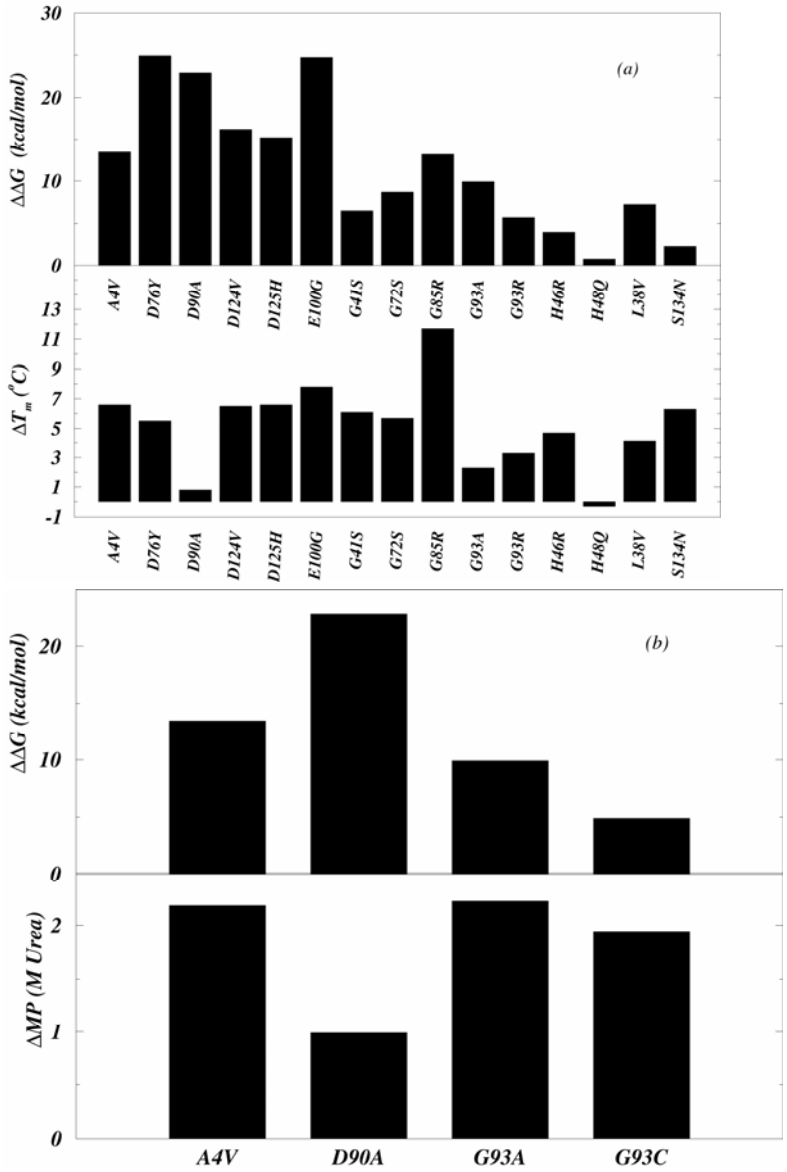


Figure 3.2 Comparison of calculated free energies with experiments (a) Comparison of calculated $\Delta\Delta G$ with change in melting temperature (ΔT_m) for 15 FALS mutants. ΔT_m is obtained from Rodriguez et al [43]. (b) Comparison of calculated ΔΔG with change in mid-point Urea concentration during solvent denaturation (ΔMP^{apo}) for 4 FALS mutants. ΔMP^{apo} is obtained from Lindberg et al [26].

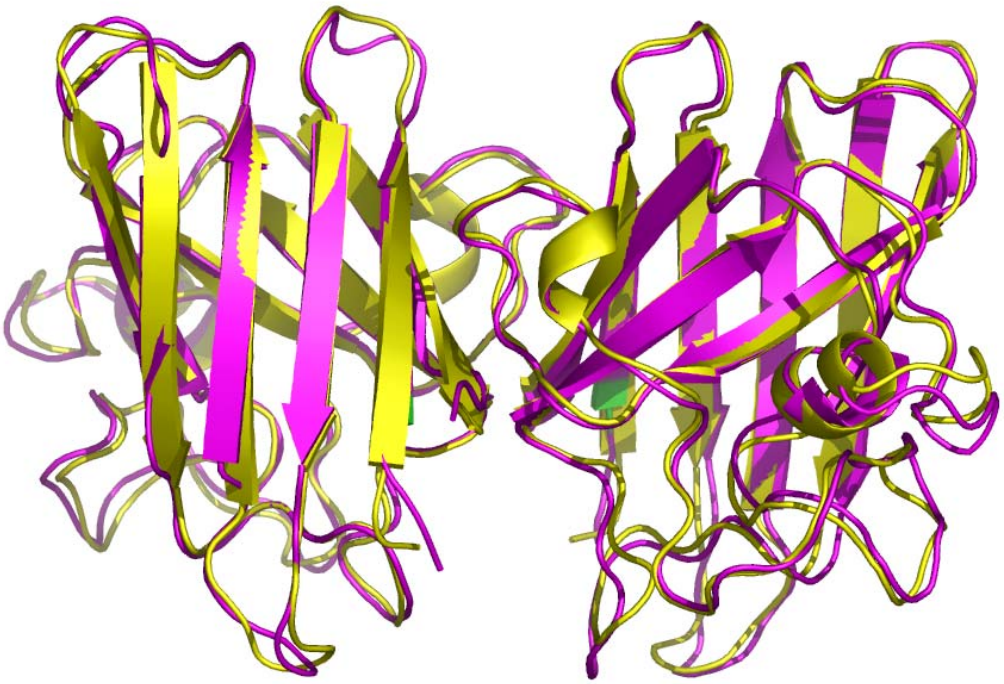


Figure 3.3 Superimposition of SOD1 structures. Structural identity is seen between solved crystal structures of A4V (1UXM, pink) and the modeled structure (yellow). The site of mutation, Ala4, is shown in green.

3.7 Supplementary Material

Table 3.S.1 Free energies of the unfolded state Free energies of peptides (Ace-$(Ala)_2$-Gly-Ala-x-Ala-Gly-$(Ala)_2$-NMet) relative to x=Gly

x	$U_{m,\,sh}$	$U_{m,\,coul}$	G_{cav}	$G_{s,vdw}$	G_{pol}	G_{solv}	TS_{conf}	G
Ala	31	0.8	43.2	-35.2	-39.3	-31.3	-18.70	3.76
Arg	39	-14	49.4	-37.7	-96.3	-84.7	-19.77	-57.50
Asn	33.1	-40.9	45.5	-35.2	-46.2	-35.9	-18.10	-39.92
Asp	58	-9.7	46.1	-29.1	-122.8	-105.8	-21.42	-56.94
Cys	55.5	-2.5	45.5	-37	-39	-30.5	-18.63	25.836
Gln	35.1	-36.4	47.4	-35.9	-50.9	-39.4	-18.92	-37.65
Glu	34.6	-5.4	45.8	-29.7	-127.5	-111.4	-23.77	-84.00
Gly	31.6	-2.8	43.3	-36	-38.2	-30.9	-19.77	0
His	32.9	-17.3	46.8	-38.2	-46.2	-37.5	-19.03	-18.95
Ile	38.5	-5.2	45.8	-38.3	-34.9	-27.4	-18.17	9.69
Leu	36.6	-5.6	46.6	-38.9	-34.3	-26.6	-20.39	5.97
Lys	67	1	48.9	-31.1	-133.2	-115.3	-20.50	-45.83
Met	34	-6.1	47.8	-40.2	-34.5	-26.9	-19.32	3.64
Phe	35	-0.3	47.1	-41.5	-38.9	-33.2	-19.61	3.86
Pro	44.4	-0.6	45.7	-38.6	-38.5	-31.4	**-22.36**	12.0
Ser	32.7	-3.7	44.1	-32.8	-46.5	-35.2	-21.37	-5.60
Thr	35	-7.5	45.9	-36	-43.3	-33.4	-22.60	-6.53
Trp	36.2	-2.9	50.1	-44.9	-41.4	-36.2	-19.38	-0.31
Tyr	35.2	-9.9	48.7	-40.1	-41	-32.3	-19.31	-4.34
Val	62.1	-4	46.3	-36.9	-37.4	-28	-20.09	31.97

Table 3.S.2: Conformational Free Energies of SOD1 mutants

Res(wild type)	Pos	Res(Mut)	ΔΔG(dim-s)	ΔΔG(dissoc-s)
Class I				
Wild type			0	0
A	4	S	9.737	-2.73
A	4	T	11.128	-3.068
A	4	V	13.468	-10.985
V	7	E	9.529	-29.471
L	8	V	2.73	-4.94
G	12	R	6.201	-17.446
V	14	G	14.69	-9.126
V	14	M	2.197	5.005
G	16	A	9.373	-18.72
G	16	S	2.236	10.816
E	21	G	11.323	-4.199
G	37	R	6.461	-19.994
L	38	R	1.924	-14.287
L	38	V	7.189	-4.953
G	41	S	6.435	-12.311
H	46	R	3.965	-16.003
H	48	Q	0.754	-10.66
E	49	K	7.397	-22.932
N	65	S	13.286	-4.134
D	76	V	20.007	-9.204
D	76	Y	24.908	-8.723
L	84	F	4.329	-13.546
L	84	V	15.535	-26.299
G	85	R	13.182	-4.719
N	86	S	14.833	-8.931
A	89	V	13.078	-14.027
D	90	A	22.854	-13.936
D	90	V	23.179	-21.294
G	93	C	4.914	-2.288
G	93	R	5.629	-2.756
G	93	V	13.676	-11.258
A	95	T	5.837	-6.071
D	96	N	19.565	-1.976
D	101	G	26.559	-8.268
D	101	N	24.635	-24.986
S	105	L	4.823	-7.137
L	106	V	7.449	-7.085
G	108	V	5.252	-6.422
D	109	N	21.346	-24.973
I	112	M	8.359	-22.802

I	113	F	3.315	-5.915
I	113	T	2.652	-11.336
G	114	A	12.948	-15.873
R	115	G	20.943	-11.544
D	124	V	16.198	-15.509
D	125	H	15.21	-16.77
L	126	S	5.564	-17.524
S	134	N	2.262	0.39
L	144	F	1.183	-4.472
A	145	G	8.268	-10.517
A	145	T	24.622	-19.084
C	146	R	7.748	-22.191
V	148	G	5.733	-4.147
V	148	I	5.161	-16.146
I	149	T	1.703	-7.189
Res(wild type)	Pos	Res(Mut)	ΔΔG(dim-s)	ΔΔG(dissoc-s)
Class II				
L	8	Q	-0.611	-2.73
G	41	D	-1.001	-18.564
H	43	R	-6.812	-4.602
L	67	R	-3.419	-5.694
H	80	R	-4.732	-5.59
V	97	M	-0.416	-10.79
I	104	F	-1.131	-0.949
N	139	K	-2.327	-4.966
L	144	S	-2.288	-10.426
Res(wild type)	Pos	Res(Mut)	ΔΔG(dim-s)	ΔΔG(dissoc-s)
Class III				
G	10	V	5.941	7.995
G	72	S	8.684	1.378
G	93	A	9.906	2.795
E	100	G	24.726	8.736
I	151	T	0.975	9.633
Res(wild type)	Pos	Res(Mut)	ΔΔG(dim-s)	ΔΔG(dissoc-s)
Class IV				
E	21	K	-4.55	7.163
G	93	D	-9.815	7.358
G	93	S	-2.561	8.775
E	100	K	-9.737	20.358
I	112	T	-14.274	21.086

CHAPTER 4

SEQUENCE AND STRUCTURAL DETERMINANTS OF SOD1 AGGREGATION

This chapter has been published as a manuscript in the *Proteins: Structure, Function, Bioinformatics* and is formatted in the journal style.

4.1 Abstract

Diverse point mutations in the enzyme Cu, Zn superoxide dismutase (SOD1) are linked to its aggregation in the familial form of the disease Amyotrophic Lateral Sclerosis. The disease-associated mutations are known to destabilize the protein, but the structural basis of the aggregation of the destabilized protein and the structure of aggregates are not well understood. Here, we investigate *in silico* the sequence and structural determinants of SOD1 aggregation: (a) we identify sequence fragments in SOD1 that have a high aggregation propensity, using only the sequence of SOD1, and (b) we perform molecular dynamics simulations of the SOD1 dimer folding and misfolding. In both cases, we identify identical regions of the protein as having high propensity to form intermolecular interactions. These regions correspond to the two cross-over loops and two β-strands in the Greek-key native fold of SOD1. Our results suggest that the high aggregation propensity of mutant SOD1 may result from a synergy of two factors: the presence of highly amyloidogenic sequence fragments ("hot-spots"), and the presence of these fragments in regions of the protein that are structurally most likely to form inter-molecular contacts under destabilizing conditions. Therefore, we postulate that the balance between the self-association of aggregation-prone sequences and the specific structural context of these sequences in the native state determines the aggregation propensity of proteins.

4.2 Introduction

The formation of protein aggregates is associated with cytotoxicity in more than 20 diverse human pathologies including amyotrophic lateral sclerosis, Alzheimer's, Parkinson's and prion diseases[1-3]. Compelling experimental evidence suggests that specific, contiguous sequence fragments in proteins are responsible for nucleating the conversion of native proteins to amyloids[4,5]. Peptides corresponding to sequence fragments of amyloid-forming polypeptides such as the amyloid β peptide, yeast and human prion proteins, calcitonin, insulin, transthyretin, β2 microglobulin, and tau protein, have been shown to form amyloid fibrils *in vitro* (Ref.[6] and references therein). Mutations found to diminish (or enhance) the aggregation of isolated peptide fragments also diminish (or enhance) the aggregation of the entire polypeptide chains[7]. Furthermore, the insertion of amyloidogenic fragments of human β2-microglobulin and the PI-SH3 domain into their respective non-aggregating homologs, mouse β2-microglobulin and the Src-SH3 domain respectively, causes the engineered homologs to readily aggregate, suggesting that the amyloidogenic sequence fragments may be necessary and sufficient for aggregation [8,9].

It is widely believed that protein aggregation requires partial or complete unfolding of the native state[6,10]. Unfolding may result in the exposure of amyloidogenic sequence fragments which, in turn, leads to oligomerization. It has been argued that natural selection has led to amyloidogenic sequence fragments being protected in the native states of protein structures found in nature [1,11]. Therefore, the ability of amyloidogenic sequences to induce aggregation is modulated by the global stability and the structure of proteins. In the aggregated state, these amyloidogenic sequence fragments have been found to be arranged as parallel β-strands in a sheet in which the amino acid sequence is in exact register. Protein

aggregation propensity is then, the interplay between the stability of the native structure, which prevents protein aggregation, and the self-association of amyloidogenic sequence fragments from different polypeptide chains into in-register structures, which promotes protein aggregation. Consequently, mutations associated with familial forms of neurodegenerative diseases may promote aggregation by either destabilizing the native state and/or increasing the self association propensities of exposed sequence fragments under destabilizing conditions. However, the molecular basis underlying protein aggregation and the effect of mutation on aggregation is not well-understood. To understand the physical basis of protein aggregation, we address the following questions: (i) which sequence fragments in a protein have high amyloidogenicity, and (ii) how is the association of these fragments modulated by the native structure and stability of the protein during its misfolding?

To address these questions, we develop an *in silico* method to identify sequence fragments and structural regions that have high amyloidogenic propensity. We apply our method for determination of the sequence and structural aggregation propensities to the homodimeric enzyme Cu, Zn superoxide dismutase (SOD1). The mutation-induced aggregation of SOD1 has been implicated in the familial form of the disease amyotrophic lateral sclerosis (FALS)[3]. In its native state, each SOD1 monomer adopts the classic Greek-key fold [12] – two β-sheets, composed of four β-stands each, and connected by two cross-over loops (Fig. 1(a)). More than 90 FALS-associated point-mutations are scattered throughout the structure of SOD1. A subset of these mutations is known to destabilize SOD1, both *in vitro* and *in vivo*[13-15]. It was previously demonstrated that *in vitro* SOD1 aggregation involves dissociation of the dimer, the loss of metal ions, and assembly of misfolded apo-monomers

into aggregates [16-18]. The first step in the SOD1 aggregation possibly occurs via the non-native dimerization of apo-monomers. The domain swapping mechanism [19-21] was suggested as a plausible mechanism for the formation of aberrant dimers[22]. Domain swapping (Fig. 1(b)) is known to occur in amyloid fibril-forming proteins[22,23].

To determine sequence determinants of SOD1 aggregation, we identify sequence fragments that have a high tendency to self-associate into in-register β-strands, based purely on the sequence of SOD1. To determine the structural determinants of SOD1 aggregation, we study the folding and misfolding of the SOD1 dimer, based purely on the geometry of the dimer. Both these complementary approaches identify the same amyloidogenic regions in the protein – corresponding to two β-strands and to the two cross-over loops in SOD1 (Fig. 1(a)) – indicating that the high aggregation propensity of SOD1 is a due to both amyloidogenic sequence "hot-spots" and their structural context. Our results suggest that aggregation of proteins requires both highly amyloidogenic sequence fragments and the presence of these fragments in specific aggregation-prone structural elements of the protein.

4.3 Methods

Preparation of peptide fragment structures

We obtain a total of 147 fragments of seven consecutive residues for the 153 residues in a SOD1 monomer. For each fragment, we construct the following five conformations: monomer, β-strand dimer (parallel and anti-parallel), and β-strand tetramer (parallel and anti-parallel). For constructing the dimer and tetramer, we mount the sequence of each fragment on idealized parallel and anti-parallel β-sheet structures using the package SCWRL. These template structures are constructed using the packages MOE (Schröedinger Inc.) and have

ideal parallel or anti-parallel β-sheet geometry. We cap the strands at the N- and C-termini by acetyl and N-methyl groups respectively, using the package SYBYL (Tripos, St. Louis, MO). For constructing the monomer, the sequence of each fragment is mounted on a random coil conformation generated using SYBYL.

MD simulations and free energy calculations

For each fragment, we perform all-atom MD simulations of each conformation and calculate the conformational free energy. For each MD trajectory, the simulation time is 496 picoseconds (ps) in explicit solvent (SPC water model [54]) using the SigmaX2.2 MD program. Conformational free energy is calculated using the ES/IS method [55] for every snapshot. Following the procedure for free energy calculations described in Ref. [56], each MD simulation consists of relaxation of the peptide(s) and water, followed by a production run of 496 picoseconds, where snapshots are collected at intervals of 1 ps. The conformational free energy is

$$G = < E > - TS_{conf} + < \Delta G_{solv} > , \qquad (1)$$

where <…> represents the average over the MD trajectory, E is the internal energy of the peptides in vacuum, T is the absolute temperature (set to 300 K), S_{conf} is the configurational entropy, and ΔG_{solv} is the solvation free energy, calculated using an implicit solvation model described by Vorobjev and Hermans [57].

For each fragment i, the difference between the average conformational free energies of the dimer and two times that of the monomer, ΔG_D^i, represents the free energy of dimerization:

$$\Delta G_D^i = \left(\overline{G_D^i} - 2 \times \overline{G_M^i} \right) \pm \sqrt{\sigma^2 \left(G_D^i \right) + 4 \times \sigma^2 \left(G_M^i \right)}, \qquad (2)$$

where $\overline{G_D^i}$, $\overline{G_M^i}$, $\sigma^2\left(G_D^i\right)$ and $\sigma^2\left(G_M^i\right)$ are the averages and standard deviations of the conformational free energies of the dimer and the corresponding monomer, respectively. The free energy of tetramerization ΔG_T^i is similarly the difference between the average conformational free energies of the tetramer and four times that of the monomer. The values ΔG_D^i and ΔG_T^i are measures of the amyloidogenicity of a given sequence fragment.

To calculate the per residue amyloidogenicity for the protein sequence, we assume that each residue in a given fragment contributes equally to its calculated amyloidogenicity. The amyloidogenicity of any residue is its average contribution to the amyloidogenicity of all fragments that include this residue. This averaging ensures that the amyloidogenicity of a given residue is modulated by its sequence neighbors. The amyloidogenicity, ΔG^j, of the residue *j* is

$$\Delta G^j = \left(\frac{\sum_{i=1}^{n} \Delta G_D^i}{7n} \right) \pm \frac{\sqrt{\sum_{i=1}^{n} \left(\sigma^2 \left(\Delta G_D^i \right) \right)}}{7n}, \qquad (3)$$

where the sum of ΔG_D^i is over all dimer fragments, *n* ($1 \leq n \leq 7$), that include a given residue *j*, and the second term in Eq. (3) is the standard deviation. We calculate a similar per-residue amyloidogenicity profile for the tetramer structures.

Discrete molecular dynamics simulations

We use a scaled Gō-model, based on the interaction scheme developed by Khare et al. [36], to model the folding and misfolding of the SOD1 dimer. We assign pair-wise, square-well interaction potentials between beads in a simplified polypeptide model according to contacts formed in the native state, i.e. two residues are said to be in contact if any of their atoms

(excluding hydrogen) are within 4.5 Å in the native state crystal structure (Protein DataBank accession code 1SPD). This procedure results in a matrix of contacts Δ_{ij}, in which each element is equal to 1 if residues *i* and *j* are in contact, and 0 otherwise (Fig. 1(c)). The contact map includes interactions between monomers at the dimer interface. In addition, we allow the formation of both intra- and inter-monomer contacts, for example, Ile 18 from chain A, which forms a native contact with Ala 4 from chain A, can also interact analogously with Ala 4 from chain B. Thus, the effective Gō-like energy of the model protein is

$$E = \sum_{i,j} \varepsilon \cdot \Delta_{ij} + \sum_{i,k} \varepsilon \cdot \Delta_{ik}^{\text{dim}} + \sum_{i,j^*} \alpha_{\text{domain-swap}} \cdot \varepsilon \cdot \Delta_{ij^*} , \qquad (4)$$

where ε is the strength of a contact, Δ_{ij}, Δ_{ik}^{dim}, and Δ_{ij^*} are the contact maps corresponding to a monomer, the dimer interface, and the domain swapping interactions, respectively. The value $\alpha_{\text{domain-swap}}$ in Eq. (4) is used to regulate the degree of inter-monomer overlap, or the effective concentration of the protein. When $\alpha_{\text{domain-swap}}>1$, interactions between residues from different monomers are stronger than intra-monomer interactions, resulting in a tendency of each chain to penetrate the pervaded volume of the other chain, rather than forming intra-chain contacts. This scenario corresponds to a polypeptide concentration in the semi-dilute regime[58]. There is an additional translational entropic contribution associated with the interactions between amino acids from different protein chains compared to the analogous interactions within the protein chain. Therefore, the magnitude of the inter-protein interactions may differ from the intra-protein interactions between analogous amino acids depending on environmental conditions such as the protein concentration. To capture this effect, we study the dynamics of misfolding at different values of the scaling factor $\alpha_{\text{domain-swap}}$.

4.4 Results

Sequence determinants of SOD1 amyloidogenesis

Since the formation of β-sheets is a necessary condition for aggregation[1,24], to identify amyloidogenic sequence fragments in SOD1, we use idealized dimer and tetramer β-strand templates, in both parallel and anti-parallel conformations. We mount overlapping heptapeptide sequences from the SOD1 sequences on each template, and calculate the conformational free energies of the 147 heptapeptide sequence oligomers using explicit solvent 496-ps MD simulations and an implicit solvation energy function (Eq. (1), Fig 2 (a)). By subtracting the free energy of two and four monomers from parallel and anti-parallel dimers and tetramers, respectively, we obtain the free energies of dimerization and tetramerization into these parallel and anti-parallel β-structures (Eq. (2)). We use these values to obtain a sequence-profile for amyloidogenicity (Eq. (3), Fig. 2(b) and (c)). The free energy contribution per residue ranges from -5.9 to +2.5 kcal/mol for dimerization (Fig. 2(b)), and from -18.3 to +6.1 kcal/mol (Fig. 2(c)). Although for a given residue, the magnitude of the amyloidogenicity is larger in the tetramer compared to the dimer, we obtain similar free energy profiles for parallel, anti-parallel, dimer and tetramer (Fig. 2(b) and (c)) structures. Thus, we conclude that the free-energy profiles adequately represent the oligomerization propensity of the SOD1 sequence fragments.

Based on the sequence amyloidogenicity profile, we identify sequence regions with high and low amyloidogenicity. Four regions of the protein have high amyloidogenicity: the N- and the C-termini, and the residues 35-45 and 110-120. The residue sequence 35-45 (IKGLTEGLHGF) is a cross-over loop, and connects the two β-strands, β3 and β4, of the sheets in the barrel. The residue sequence 110-120 (HCIGRTLVVH) corresponds to a loop

between the strands β6 and β7 (residues 110-114), and the strand β7 (residues 115-120). Residues 113-115 are also part of the dimer interface, and in the native dimer structure are symmetrically placed such that residues 113 and 114 from both subunits form contacts with each other. Thus, we find that specific regions of the SOD1 sequence have a high propensity to oligomerize into in-register, parallel and anti-parallel β-strands.

Folding thermodynamics of SOD1 monomer and dimer

In the context of the entire polypeptide chain, the oligomerization of the identified amyloidogenic "hot-spot" sequences is dependent on the degree to which these sequences are exposed in unfolded or partially folded SOD1. To identify the regulation of amyloidogenic sequences during the misfolding of SOD1, we first model the folding of SOD1 monomer and dimer to reproduce the experimentally observed thermodynamics of folding. Protein folding pathways are largely determined by the topology of the native state[25-30], and, therefore, the native topology of SOD1 determines the specific sub-structures of the protein that are exposed in the aggregation-prone partially folded states (i.e. the structural determinants of SOD1 aggregation). The principal difficulty in studying the folding of proteins *in silico* is the lack of accurate information about the energetics of interaction between amino-acids. Since protein structures have *a posteriori* information about the amino-acid interactions[31], simplified native-structure based models such as the Gō model are used to study folding. In the Gō model[32,33], the energy of the protein is expressed as a sum of pairwise native contact energies. A native contact exists between amino acid residues if they are closer to each other than a cut-off distance in the native state, and folding process is regarded as the acquisition of all native contacts. Typically, in these simplified models of folding, a coarse-grained

representation of the protein is used in which each amino acid is represented by one or more beads and the protein chain is represented as beads-on-a-string[34]. For modeling SOD1 folding, we use a coarse-grained representation of the SOD1 polypeptide chain developed by Ding et al. [35] (6-bead model, 4 backbone and 2 side chain beads). We have previously shown that a scaled Gō-model of native-state based inter-residue interactions can reproduce the two-state folding of the SOD1 monomer[36] (see Methods). Starting from a stretched conformation, we perform equilibrium simulations at constant temperatures, T, ranging from T=0.1 to 1.0 (Fig. 3(a)). We find that the 6-bead polypeptide model with scaled-Gō interactions reproduces the experimentally observed two-state folding thermodynamics of SOD1. For a two-state protein, the heat capacity is expected to have a single peak at the folding transition temperature, T_F. A MD trajectory at $T=T_F$ consists of two distinct populations – folded and unfolded – at equilibrium with other, characteristic of a two-state protein (Fig 3(a)). At $T<T_F$ and $T>T_F$, on the other hand, only the folded and the unfolded states, respectively, are populated.

Next, we model the folding of the SOD1 dimer. In addition to the intra-monomeric contacts of the two monomers, the contact map of the dimer contains contacts corresponding to the native dimer interface. Starting from two stretched chains, we perform equilibrium simulations at constant temperature ranging from T=0.1 to T=1.0. The heat capacity *versus* temperature curve shows the existence of two transitions, corresponding to the temperatures T_1=0.66 and T_2=0.72 (Fig. 3(b)). The MD trajectories at T_1 consist of two distinct populations, corresponding to the dimer and two folded monomers respectively, whereas at T_2, the two populations correspond to the folded monomers and unfolded monomers respectively (Fig. 3(b)). At low temperatures ($T<T_1$), the native dimer is formed, where as at

high temperatures ($T>T_2$), both chains are unfolded. This is in accord with experimental studies, in which the folding of the dimer has been shown to be a three-step process[16,17,37]:

$$D \rightleftarrows 2M \rightleftarrows 2U \,, \tag{5}$$

where D is the dimer, M is the monomer and U is the unfolded state Thus, our simulations reproduce the folding thermodynamics of the SOD1 dimer in agreement with the experimentally observed thermodynamics.

Structural determinants of SOD1 misfolding and aggregation

After verifying that the SOD1 monomer and dimer models fold to the correct native state with experimentally-observed two-state and three-state thermodynamics respectively, we study the misfolding of the SOD1 dimer. In Eq. (4), when $\alpha_{\text{domain-swap}}>1$, the relatively higher strength of inter-monomer domain-swapped interactions (see Methods) compared to the intra-monomer interactions mimics an increase in the protein concentration. At high concentrations, native structure formation within a monomer competes with the formation of domain-swapped[19,20] topologies (Fig. 1(b) and (c)). Further, non-specific hydrogen bonds can form between the backbones of the two monomers. Under these conditions, we expect to observe non-native topologies. We perform dimer simulations over a range of α-values, α = 0.5, 1, and 1.5, to screen for non-native dimeric forms of SOD. To obtain the aggregation propensity of each residue, we calculate the cumulative frequency of inter-monomer contacts formed by each residue at the temperatures T=0.65 and T=0.75. For the native dimer, the residues on the dimer interface are 5, 7, 9, 50-53, 113-115, and 150-153, in each monomer. Since aggregation has been shown to occur in a narrow temperature range around the folding transition temperature, we perform simulations with domain-swapped interactions in the

range T=0.65 to T=0.75 (the transition temperatures for the native dimer are T_1=0.66 and T_2=0.72).

At α=0.5, the folding of the dimer is native-like, as evidenced by frequency profile (Fig. 4(a)) where residues that are involved in the native dimer interface form inter-monomer contacts with a high frequency at both T=0.65 (red bars) and T=0.75 (green bars). However, other residues in the vicinity of the dimer interface residues, such as 3, 55, 107-110 also form inter-monomer contacts with a high probability. Thus, although weak domain-swapping interactions do not significantly alter the folding of the dimer, several residues, especially near the native dimer interface and near residue 110, can induce alternate dimer conformations.

At α=1, we observe significant non-native contact formation, especially at T=0.65, which is near the transition temperature T_1=0.66 for dimer dissociation. The residues that form inter-monomer contacts include the native dimer residues and several other residues, especially at the N- and the C-termini, and the regions 25-40 and 100-120. The formation of inter-monomer contacts by both the native dimer interface and specific non-native residues suggests that the formation of an alternate dimer interface competes with the formation of the native dimer interface.

At α=1.5 and T=0.65, the native dimer interface is not seen as evidenced by a disappearance of inter-monomer contacts by native interface residues. Instead, a new set of residues, 55-65, 90-100, 110-120 is observed to form the most inter-monomer contacts. Thus, at increasing concentrations of the destabilized dimer, simulated by increasing α-values, we observe decreased formation of the native dimer interface and an increase in the formation of an alternative dimer interface.

An examination of the structures with the altered dimer interface shows a diversity of conformations. This is expected to be the case if dimer formation is under kinetic control and further oligomerization, i.e. the formation of a trimeric or a higher-order species, causes thermodynamic stabilization of the oligomers. These observations are in agreement with recent computational and experimental studies[38,39] suggesting that kinetic effects play an important role in the early stages of aggregation. Based on the structures of non-native dimers observed in simulation, we propose several plausible, kinetically competing aggregation pathways. For example, two representative structures are shown in Fig. 4 (b) and 4(c). Each of these two non-native dimeric states persists throughout the entire simulation time of 1.5×10^5 time units once it is formed. Fig. 4(b) shows a domain swapping interaction to form a native-like beta-barrel using half of the strands from each monomer, while the remainder of each monomer undergoes smaller-scale strand-by-strand pairing between the chains. Interestingly, residues 35-45 from each monomer associate with each other at the interface between the canonical domain-swapped region and the strand-swapped region. Domain swapping of this nature results in the apposition of identical, aggregation-prone sequences from each monomer. Fig. 4(c) shows a non-native SOD dimer in which the native strand pairings form but the monomers do not collapse into the barrel topologies, resulting in the formation of a continuous β-sheet. At the interface between the two flattened sheets, residues 110-120 from the opposite monomers directly interact in a parallel fashion. The residues 35-45 are located on the outside edges of this dimeric β-sheet where they are available for propagating the aggregate further. In this alternative model for multimeric β-sheet association, flattened β-sheets associate via edge-edge interactions mediated by fragment 110-120 and propagate unidirectionally by interactions between the fragments 35-

45. Thus, the dimers exhibit a rich variety of structures in which the residues 35-45 and 110-120 form key intermolecular contacts.

4.5 Discussion

Finding aggregation "hot-spots" in SOD1 using peptide fragments

To obtain the aggregation-propensity of SOD1 sequence fragments, we use the free energy of oligomerization of isolated sequence fragments mounted onto idealized β-sheet templates, composed of two and four β-strands. In a similar approach, de la Paz et al.[24] mounted hexapeptide sequences on each strand of a six-stranded antiparallel β-sheet template to design highly amyloidogenic peptide sequences. Using this simplified design procedure, they found that, for short peptide sequences, the formation of β-structures is necessary, but not sufficient, for the formation of amyloid fibrils.

The β-sheet templates we used included both parallel and anti-parallel conformations. Although protein and peptide aggregation involves conversion to β-sheets, whether the β-sheets are parallel or anti-parallel has not been understood[6]. It has been argued that the early oligomers formed during protein and peptide aggregation can be of either topology, although for small peptides, anti-parallel topologies are preferably formed[40]. In recent study of the dynamics of peptide dimerization, Hwang et al.[38] found that the β-strand peptide dimers exhibited both parallel and anti-parallel β-sheets, with an overall preference for anti-parallel arrangements. Since both parallel and anti-parallel topologies are likely in the early stages of aggregation, we used both topologies for constructing the templates, and also found that anti-parallel β-sheets have lower free energies for a given sequence (Fig. 2(b) and (c)). We

postulate that the lower free energy of the anti-parallel structures is due to less strained hydrogen bonds in the anti-parallel structures compared to the parallel structures.

We find that the thermodynamic contribution of each residue to the aggregation propensity (ΔG) is greater in the tetramer compared to the dimer (Fig. 2(b) and (c)). This finding indicates that the free energy gain upon aggregation increases non-linearly as the size of the aggregate increases, i.e. greater stabilization occurs as higher-order oligomers are formed. Thus, aggregation may be reversible for smaller oligomers, but as the size and the stability of the oligomer of the increases, the disaggregation is less likely. This postulate is in agreement with recent findings of the stability of amyloid β dimers, which suggest that dimers are only marginally stabilized compared to free monomers[41].

The use of overlapping peptide fragments to determine amyloidogenicity of residues ensures that sequence neighbors of a residue modulate its amyloidogenicity. Inclusion of local interactions by use of such peptide models is known to be a successful strategy for modeling protein structure[42]. In protein modeling algorithms, for example the ROSETTA program developed by Baker and co-workers[43,44], the propensity of peptide fragments of a protein to adopt specific secondary structures has been used to successfully predict protein structures. Our approach of predicting aggregation propensities also similarly relies on evaluating the propensities of short peptide fragments to adopt specific structures, and on evaluating the effect of neighboring residues on the amyloidogenicity of a given residue.

Dependence of amyloidogenicity on the hydrophobicity, β-sheet propensity, and net charge of the sequence

The residue fragments that we identify as amyloidogenic have a high content of hydrophobic residues, and strategically placed polar (T, H) and charged residues (K, E, R). The charged residues may provide further stabilizing interactions by electrostatic and/or hydrogen bonds. Both hydrophobic and charged interactions are known to be critical to stabilize aggregates[24,45,46]. The residue sequence 100-109 (EDSVISLSGD) is also amyloidogenic, albeit less than the two highly amyloidogenic fragments 35-45 and 110-120, and shows a similar distribution of charged, polar and hydrophobic residues. Residues fragments that have comparatively low amyloidogenicity are 18-22 (INFEQ), 52-55 (DNTA), 84-88 (LGNVT), and 135-140 (TKTNA). These fragments have a high content of polar, uncharged amino-acids (N, Q, T). The fragments containing these residues are stable in tetrameric and dimeric oligomeric states (Fig. 2 (a)), but the free energy penalty for oligomerization is high (Fig. 2(b) and (c)). We argue that this high penalty is a result of more favorable interactions of the monomers with water compared to the interactions with other peptides in the oligomeric state.

To understand if the per residue amyloidogenicity is simply a reflection of some intrinsic physico-chemical property of the sequence, such as β-strand propensity and hydrophobicity, we compare the amyloidogenicity profile with these physico-chemical properties. Recently, a number of phenomenological approaches have been developed in which these properties and experimental conditions are used as variables to predict aggregation rates of polypeptide chains. We use the model developed by Chiti et al.[47] in which average hydrophobicity, β-sheet propensity, and net charge is calculated for each sequence and a linear regression fit to the aggregation rate is obtained. Following Chiti et

al.,[47] we assume that the free energy of multimer formation calculated in our MD simulations is:

$$\Delta G = C_{\beta} p_{\beta} + C_{hydrophob} p_{hydrophob} + C_{\pm} p_{\pm}, \quad (6)$$

where p_{β}, $p_{hydrophob}$, and $p_{\pm}$ are the β-sheet propensity, hydrophobicity, and net charge of the sequence at pH 7, C_{β}, $C_{hydrophob}$ and $C_{\pm}$ are scaling coefficients, and ΔG is free energy of dimer or tetramer formation determined in MD simulations. First, we find that none of the individual properties correlate with the free energy of multimer formation (Fig. 5(a)-(c)). Second, for each template structure, namely the dimer and the tetramer with parallel and anti-parallel topology, we perform a linear regression analysis using Eq. (6) to determine the set of coefficients C_{β}, $C_{hydrophob}$ and $C_{\pm}$. We find that in each case the correlation between the ΔG-values and the sequence properties is poor (Fig. 5(d)-(g)), and conclude that the ΔG of multimerization is not sufficiently explained by a linear combination of the β-sheet propensity, hydrophobicity, or the net charge of the sequence. Our finding is in agreement with the recent work of Serrano and co-workers in which mean-field aggregation-profiles for protein sequences calculated using their TANGO algorithm did not correlate with the β-strand propensity of the sequence [48].

The determinants of SOD1 aggregation in FALS

To determine the sequence and structural determinants of SOD1 aggregation, we evaluate (1) the propensities of oligomerization of all sequence fragments of SOD1, and (2) the structural propensities of different parts of SOD1 to self-associate during misfolding. We use no sequence information during the identification of structural regions that are likely to self-associate, and vice versa. Since we find identical regions of the SOD1 molecule using

these disparate methods, we argue that aggregation of SOD1 is a consequence of both having aggregation-prone sequence fragments, and the topological context of these fragments during misfolding.

We have previously shown[16] that a minimal mechanism for the aggregation of SOD1 is

$$D \rightleftarrows 2M_{apo} + metals \xrightarrow{k_{agg}} A, \qquad (7)$$

where D is the native dimer, M_{apo} is the apo-monomer and A is the aggregate. The mechanism in Eq. (5) is minimal because the second reaction, $M_{apo} \rightarrow A$, is likely to be a multi-step process. The first step in this process is the non-native dimerization of apo-monomers, followed by the further addition of misfolded monomers. A common mechanism for non-native dimerization and aggregation is domain swapping [20,49]. The domain swapping of two monomers leads to the formation of structures which are either "open" and lead to further elongation of the aggregate, or are "closed" and serve as dead-ends that cannot propagate further. It has been shown that domain swapping has a purely topological origin, i.e. it is a consequence of the competition between the formation of native contacts, and the formation of symmetric inter-monomer contacts [49,50]. Therefore, we study the misfolding of SOD1 dimer using domain-swapping.

We propose that FALS mutations induce aggregation by affecting the rate and equilibrium constants of dimer dissociation and metal-loss in Eq. (5), and by affecting the degree of domain swapping. To qualitatively model the differential effects of mutations, we introduce the scaling parameter $\alpha_{\text{domain-swap}}$, with which we control the relative of strengths of intra- and inter-chain interactions. Under a range of these destabilizing conditions, we observe the preferential self-association of residues fragments 35-45 or 110-120, suggesting

that conversion to the aggregate is a specific process governed by the self-association of specific regions of SOD1. We find a number of misfolded dimeric species in which these key interactions formed, but there is considerable conformational diversity in the overall structure of the misfolded dimers. This conformational diversity was absent in the misfolded dimers of the SH3-domain generated using a similar approach, and only two dominant "open" and "closed"topologies were found[49]. Thus, the conformational diversity of the misfolded dimers is an intrinsic property of the SOD1 structure. We expect that once the key interactions for aggregation are formed, different dimer conformations are stabilized, and therefore, initiate aggregation under different experimental conditions. This phenomenon may be responsible for the experimentally observed variety of aggregate morphologies as a function of environmental conditions, such as denaturants, pH and temperature, that have been used to generate the aggregates[18,51-53].

The sequence profile of amyloidogenicity represents an upper bound for the aggregation propensity of SOD1 molecule because it is a measure of the intrinsic property of the SOD1 sequence to self-associate. However, the degree to which these different sequences can self-associate is determined by the structural dynamics during misfolding. The high aggregation-propensity of the destabilized mutant SOD1 may be a result of the synergy between structural dynamics and sequence propensity, such that highly aggregation-prone sequences are also topologically most likely to form inter-chain contacts.

4.6 References

1. Dobson,C.M. Protein folding and misfolding. *Nature* **426**, 884-890 (2003).

2. Dobson,C.M. In the footsteps of alchemists. *Science* **304**, 1259-+ (2004).

3. Cleveland,D.W. & Rothstein,J.D. From Charcot to Lou Gehrig: Deciphering selective motor neuron death in ALS. *Nature Reviews Neuroscience* **2**, 806-819 (2001).

4. Reches,M., Porat,Y. & Gazit,E. Amyloid fibril formation by pentapeptide and tetrapeptide fragments of human calcitonin. *J. Biol. Chem.* **277**, 35475-35480 (2002).

5. Tenidis,K. *et al.* Identification of a penta- and hexapeptide of islet amyloid polypeptide (IAPP) with amyloidogenic and cytotoxic properties. *Journal of Molecular Biology* **295**, 1055-1071 (2000).

6. Thirumalai,D., Klimov,D.K. & Dima,R.I. Emerging ideas on the molecular basis of protein and peptide aggregation. *Current Opinion in Structural Biology* **13**, 146-159 (2003).

7. Villegas,V. *et al.* Protein engineering as a strategy to avoid formation of amyloid fibrils. *Protein Sci* **9**, 1700-1708 (2000).

8. Ventura,S. *et al.* Short amino acid stretches can mediate amyloid formation in globular proteins: The Src homology 3 (SH3) case. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 7258-7263 (2004).

9. Ivanova,M.I., Sawaya,M.R., Gingery,M., Attinger,A. & Eisenberg,D. An amyloid-forming segment of beta 2-microglobulin suggests a molecular model for the fibril. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 10584-10589 (2004).

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

11. Richardson,J.S. & Richardson,D.C. Natural beta-sheet proteins use negative design to avoid edge-to-edge aggregation. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 2754-2759 (2002).

12. Tainer,J.A., Getzoff,E.D., Beem,K.M., Richardson,J.S. & Richardson,D.C. Determination and Analysis of the 2A Structure of Copper, Zinc Superoxide-Dismutase. *Journal of Molecular Biology* **160**, 181-217 (1982).

13. 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).

14. Lindberg,M.J., Tibell,L. & Oliveberg,M. Common denominator of Cu/Zn superoxide dismutase mutants associated with amyotrophic lateral sclerosis: Decreased stability of the apo state. *PNAS* 262527099 (2002).

15. Johnston,J.A., Dalton,M.J., Gurney,M.E. & Kopito,R.R. Formation of high molecular weight complexes of mutant Cu,Zn-superoxide dismutase in a mouse model for familial amyotrophic lateral sclerosis. *Proceedings of the National Academy of Sciences of the United States of America* **97**, 12571-12576 (2000).

16. Khare,S.D., Caplow,M. & Dokholyan,N.V. Determination of the rate and equilibrium constants for a multi-step reaction for aggregation of superoxide dismutase in ALS. *PNAS* **101**, 15094-15099 (2004).

17. Ray,S.S. *et al.* An intersubunit disulfide bond prevents in vitro aggregation of a superoxide dismutase-1 mutant linked to familial amytrophic lateral sclerosis. *Biochemistry* **43**, 4899-4905 (2004).

18. Rakhit,R. *et al.* Monomeric Cu,Zn-superoxide dismutase is a common misfolding intermediate in the oxidation models of sporadic and familial amyotrophic lateral sclerosis. *J. Biol. Chem.* **279**, 15499-15504 (2004).

19. Bennett,M.J., Choe,S. & Eisenberg,D. Domain Swapping - Entangling Alliances Between Proteins. *Proceedings of the National Academy of Sciences of the United States of America* **91**, 3127-3131 (1994).

20. Schlunegger,M.P., Bennett,M.J. & Eisenberg,D. Oligomer formation by 3D domain swapping: A model for protein assembly and misassembly. *Advances in Protein Chemistry, Vol 50* **50**, 61-122 (1997).

21. Liu,Y.S., Gotte,G., Libonati,M. & Eisenberg,D. A domain-swapped RNase A dimer with implications for amyloid formation. *Nature Structural Biology* **8**, 211-214 (2001).

22. Ding,F., Dokholyan,N.V., Buldyrev,S.V., Stanley,H.E. & Shakhnovich,E.I. Molecular dynamic simulation of the SH3 domain aggregation suggests a generic amyloidogenesis mechanism. *Journal of Molecular Biology* **324**, 851-857 (2002).

23. Janowski,R. *et al.* Human cystatin C, an amyloidogenic protein, dimerizes through three-dimensional domain swapping. *Nature Structural Biology* **8**, 316-320 (2001).

24. de la Paz,M.L. & Serrano,L. Sequence determinants of amyloid fibril formation. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 87-92 (2004).

25. Dokholyan,N.V., Li,L., Ding,F. & Shakhnovich,E.I. Topological determinants of protein folding. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 8637-8641 (2002).

26. Dokholyan,N.V. *et al.* Identifying importance of amino acids for protein folding from crystal structures. *Macromolecular Crystallography, Pt D* **374**, 616-+ (2003).

27. Pande,V.S., Grosberg,A.Y. & Tanaka,T. Heteropolymer freezing and design: Towards physical models of protein folding. *Reviews of Modern Physics* **72**, 259-314 (2000).

28. Bryngelson,J.D., Onuchic,J.N., Socci,N.D. & Wolynes,P.G. Funnels, Pathways, and the Energy Landscape of Protein-Folding - A Synthesis. *Proteins-Structure Function and Genetics* **21**, 167-195 (1995).

29. Shea,J.E., Onuchic,J.N. & Brooks,C.L. Molecular dynamics study of the mechanism and thermodynamics of folding of the src-SH3 protein. *Abstracts of Papers of the American Chemical Society* **221**, U406-U407 (2001).

30. Plotkin,S.S. & Onuchic,J.N. Understanding protein folding with energy landscape theory - Part I: Basic concepts. *Quarterly Reviews of Biophysics* **35**, 111-167 (2002).

31. Dokholyan,N.V. *et al.* Identifying importance of amino acids for protein folding from crystal structures. *Macromolecular Crystallography, Pt D* **374**, 616-+ (2003).

32. Go,N. & Abe,H. Noninteracting local-structure model of folding and unfolding transition in globular proteins. I. Formulation. *Biopolymers* **20**, 991-1011 (1981).

33. Abe,H. & Go,N. Noninteracting local-structure model of folding and unfolding transition in globular proteins. II. Application to two-dimensional lattice proteins. *Biopolymers* **20**, 1013-1031 (1981).

34. Dokholyan,N.V., Buldyrev,S.V., Stanley,H.E. & Shakhnovich,E.I. Discrete molecular dynamics studies of the folding of a protein-like model. *Folding & Design* **3**, 577-587 (1998).

35. Ding,F., Borreguero,J.M., Buldyrey,S.V., Stanley,H.E. & Dokholyan,N.V. Mechanism for the alpha-helix to beta-hairpin transition. *Proteins-Structure Function and Genetics* **53**, 220-228 (2003).

36. Khare,S.D., Ding,F. & Dokholyan,N.V. Folding of Cu, Zn superoxide dismutase and familial amyotrophic lateral sclerosis. *Journal of Molecular Biology* **334**, 515-525 (2003).

37. Mei,G. *et al.* Denaturation of Human Cu/Zn Superoxide-Dismutase by Guanidine-Hydrochloride - A Dynamic Fluorescence Study. *Biochemistry* **31**, 7224-7230 (1992).

38. Hwang,W., Zhang,S.G., Kamm,R.D. & Karplus,M. Kinetic control of dimer structure formation in amyloid fibrillogenesis. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 12916-12921 (2004).

39. Tcherkasskaya,O., Sanders,W., Chynwat,V., Davidson,E.A. & Orser,C.S. The role of hydrophobic interactions in amyloidogenesis: Example of prion-related polypeptides. *Journal of Biomolecular Structure & Dynamics* **21**, 353-365 (2003).

40. Tycko,R., Petkova,A., Oyler,N., Chan,C.C. & Balbach,J. Probing the molecular structure of amyloid fibrils with solid state NMR. *Biophys. J.* **82**, 187A (2002).

41. Urbanc,B. *et al.* Molecular dynamics simulation of amyloid beta dimer formation. *Biophys. J.* **87**, 2310-2321 (2004).

42. Doyle,R., Simons,K., Qian,H. & Baker,D. Local interactions and the optimization of protein folding. *Proteins-Structure Function and Genetics* **29**, 282-291 (1997).

43. Simons,K.T. *et al.* Improved recognition of native-like protein structures using a combination of sequence-dependent and sequence-independent features of proteins. *Proteins-Structure Function and Genetics* **34**, 82-95 (1999).

44. Bystroff,C. & Baker,D. Prediction of local structure in proteins using a library of sequence-structure motifs. *Journal of Molecular Biology* **281**, 565-577 (1998).

45. Gazit,E. A possible role for pi-stacking in the self-assembly of amyloid fibrils. *Faseb Journal* **16**, 77-83 (2002).

46. Chiti,F., Stefani,M., Taddei,N., Ramponi,G. & Dobson,C.M. Rationalization of the effects of mutations on peptide and protein aggregation rates. *Nature* **424**, 805-808 (2003).

47. Chiti,F. *et al.* Studies of the aggregation of mutant proteins in vitro provide insights into the genetics of amyloid diseases. *PNAS* **99**, 16419-16426 (2002).

48. Linding,R., Schymkowitz,J., Rousseau,F., Diella,F. & Serrano,L. A comparative study of the relationship between protein structure and beta-aggregation in globular and intrinsically disordered proteins. *Journal of Molecular Biology* **342**, 345-353 (2004).

49. Ding,F., Dokholyan,N.V., Buldyrev,S.V., Stanley,H.E. & Shakhnovich,E.I. Molecular dynamics simulation of the SH3 domain aggregation suggests a generic amyloidogenesis mechanism. *Journal of Molecular Biology* **324**, 851-857 (2002).

50. Yang,S.C., Cheung,M.S., Onuchic,J.N. & Levine,H. Molecular dynamics simulations on domain swapping. *Biophys. J.* **86**, 267A-268A (2004).

51. DiDonato,M. *et al.* ALS mutants of human superoxide dismutase form fibrous aggregates via framework destabilization. *Journal of Molecular Biology* **332**, 601-615 (2003).

52. Stathopulos,P.B. *et al.* Cu/Zn superoxide dismutase mutants associated with amyotrophic lateral sclerosis show enhanced formation of aggregates in vitro. *PNAS* **100**, 7021 (2003).

53. Rakhit,R. *et al.* Oxidation-induced misfolding and agagaregation of superoxide dismutase and its implications for amyotrophic lateral sclerosis. *J. Biol. Chem.* **277**, 47551-47556 (2002).

54. Hermans,J., Berendsen,H.J.C., Vangunsteren,W.F. & Postma,J.P.M. A Consistent Empirical Potential for Water-Protein Interactions. *Biopolymers* **23**, 1513-1518 (1984).

55. Vorobjev,Y.N., Almagro,J.C. & Hermans,J. Discrimination between native and intentionally misfolded conformations of proteins: ES/IS, a new method for calculating conformational free energy that uses both dynamics simulations with an explicit solvent and an implicit solvent continuum model. *Proteins-Structure Function and Genetics* **32**, 399-413 (1998).

56. Urbanc,B. *et al.* Molecular dynamics simulations of Amyloid beta dimer formation. *Biophys. J.* (2004).

57. Vorobjev,Y.N. & Hermans,J. Free energies of protein decoys provide insight into determinants of protein stability. *Protein Sci* **10**, 2498-2506 (2001).

58. Rubinstein,M. & Colby,R.H. Polymer Physics. Oxford University Press, New York (2003).

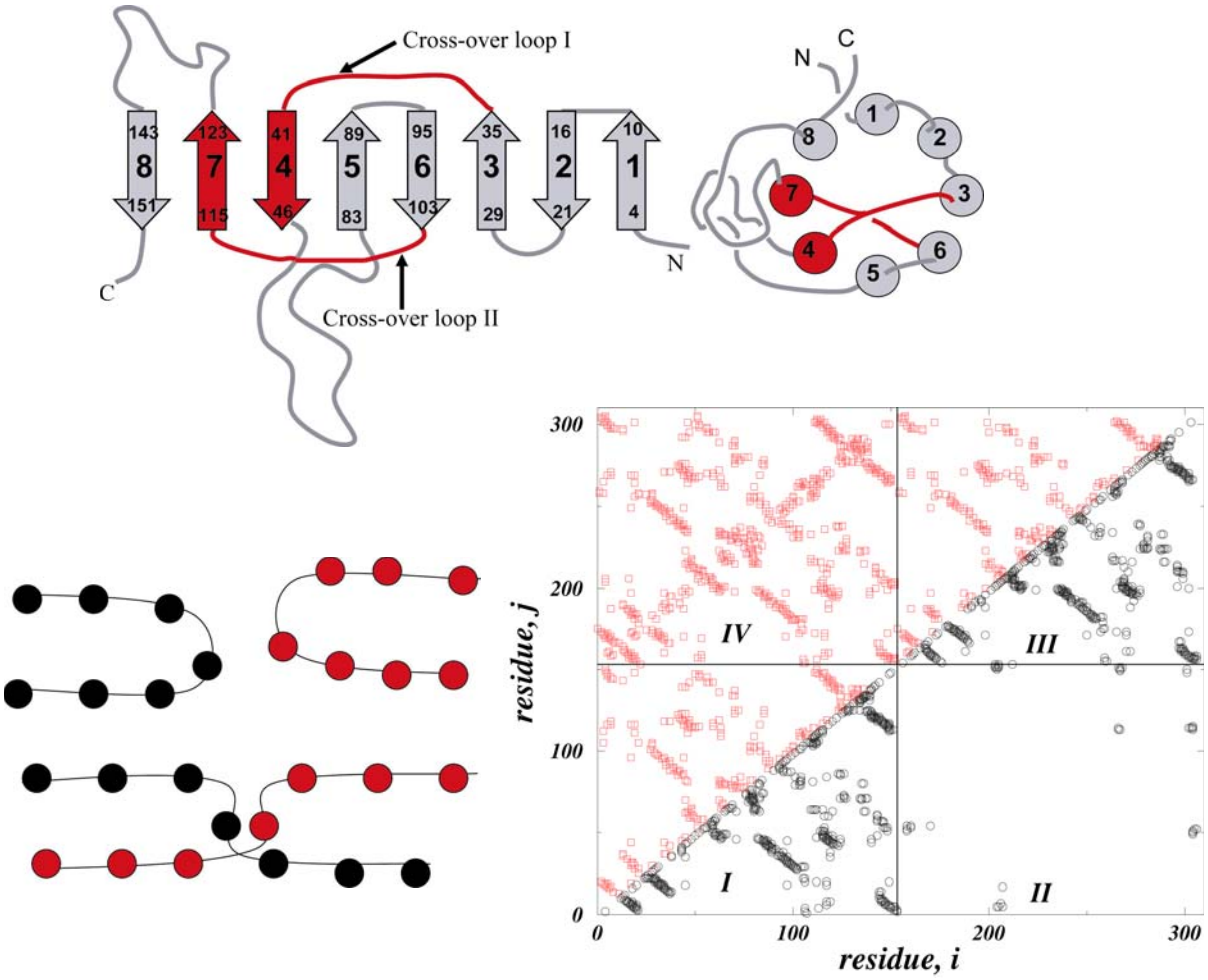


Figure 4.1 Schematic and contact map of SOD1. (a) A schematic of SOD1. The regions identified as amyloidogenic are shown in red. These constitute the β-strands 4 and 7 and the two cross-over loops which connect the two β-sheets in the SOD1 barrel. (b) A schematic of domain-swapped interactions and (c) the domain-swapped contact map used for simulations. In the contact map, regions I and III correspond to the intra-monomer contacts of the two monomers, region II corresponds to contacts on the dimer interface, and region IV corresponds to the domain-swapped interactions between the two monomers.

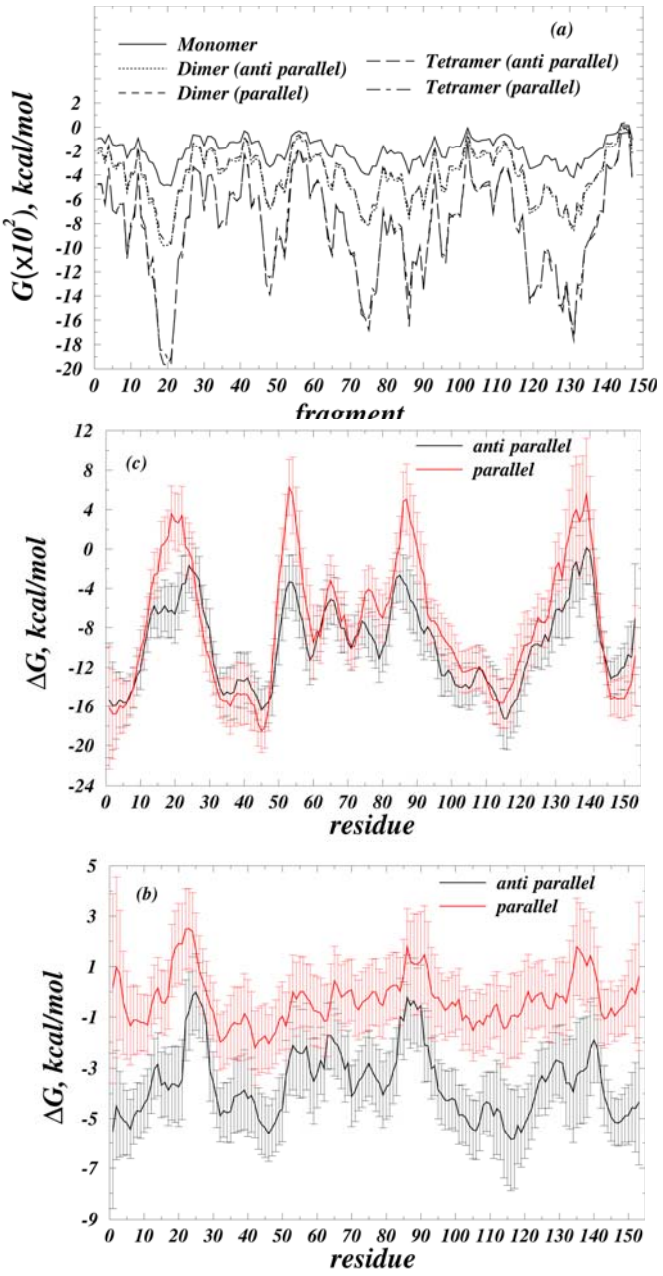
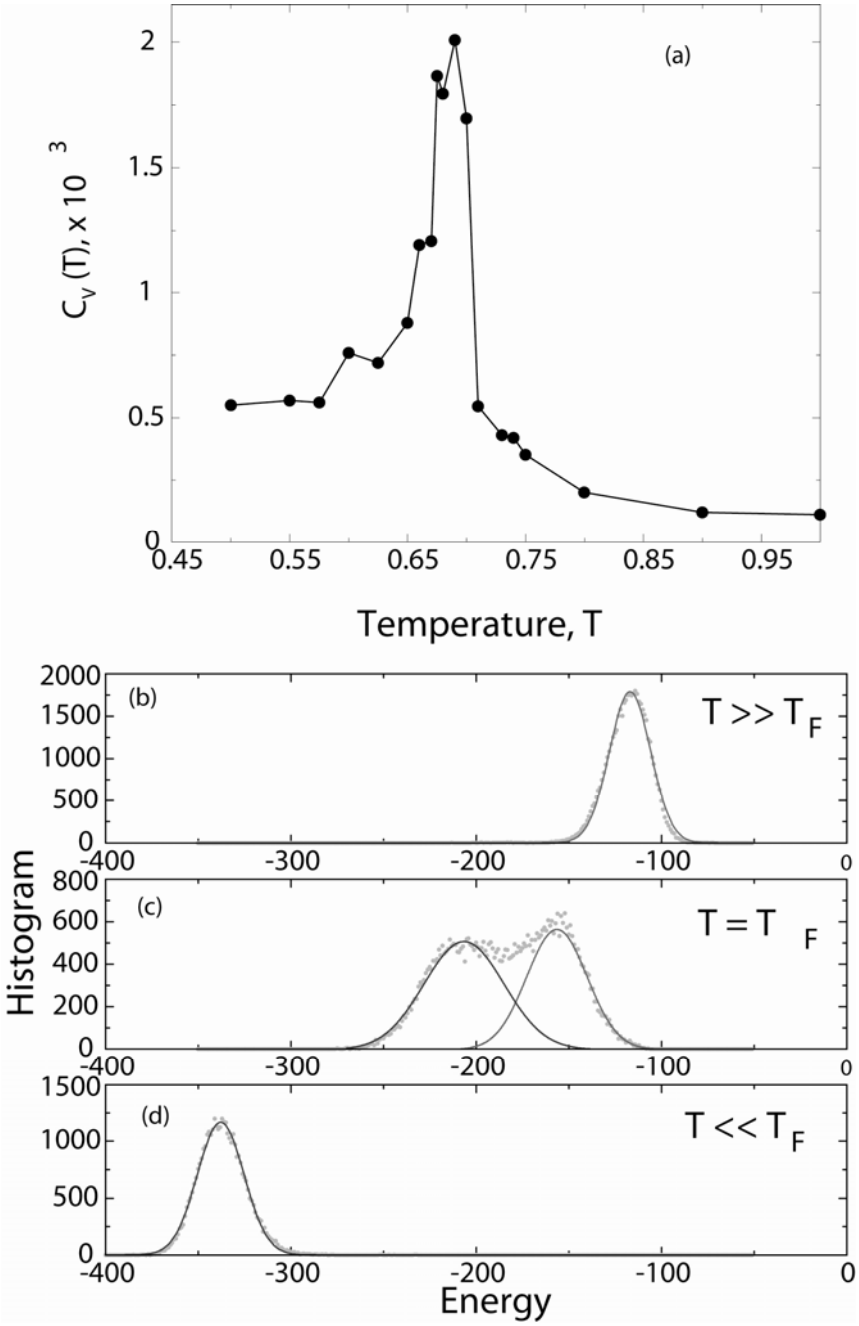


Figure 4.2. **Free energies of SOD1 regions obtained in MD simulations**. (a) The free energy of the 147 overlapping heptamer sequence fragments of SOD1, mounted on various template backbones. We use the average contribution of each residue to the free energy of all fragments containing the residue to evaluate the free energy of oligomerization of the residue. (b) The free energy of dimerization for each residue (c) The free energy of tetramerization for each residue.


(a)
C_v (T), x 10 ^{3}
Temperature, T
(b)
$T >> T_F$
(c)
$T = T_F$
(d)
$T << T_F$
Histogram
Energy

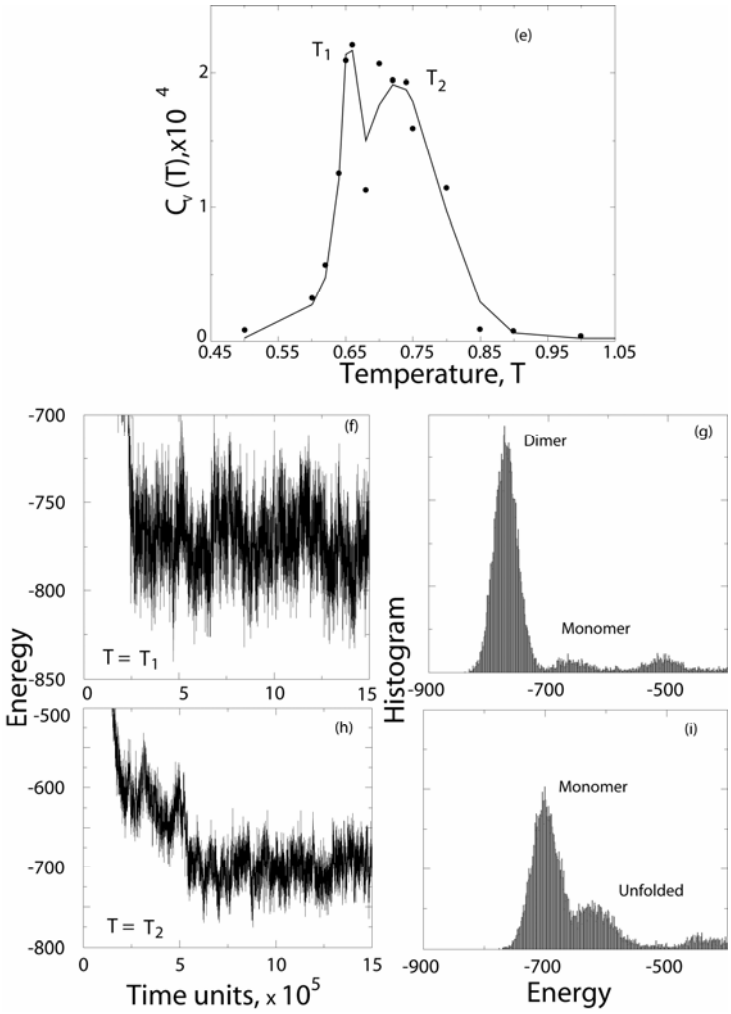


Figure 4.3. **Folding of SOD1 monomer and dimer** (a)-(d) Folding of the SOD1 monomer. (a) The heat capacity as a function of temperature shows a single peak at the folding transition temperature, T_F. The histograms of populations at (b) $T<<T_F$, (c) $T=T_F$ and (d) $T>T_F$ show that the folding transition is two state, as found in experiments. (e)-(i) The folding of the SOD1 dimer. (e) The heat capacity *versus* temperature shows two peaks (at T_1=0.66 and T_2=0.72) indicating the existence of multiple transitions. Trajectories and population histograms at the two transition temperatures, (f) and (g) at T_1, and (h) and (i) at T_2 show that the folding/unfolding of the dimer is a three-state process, involving folded dimer, folded monomers and unfolded monomers, as found in experiments.

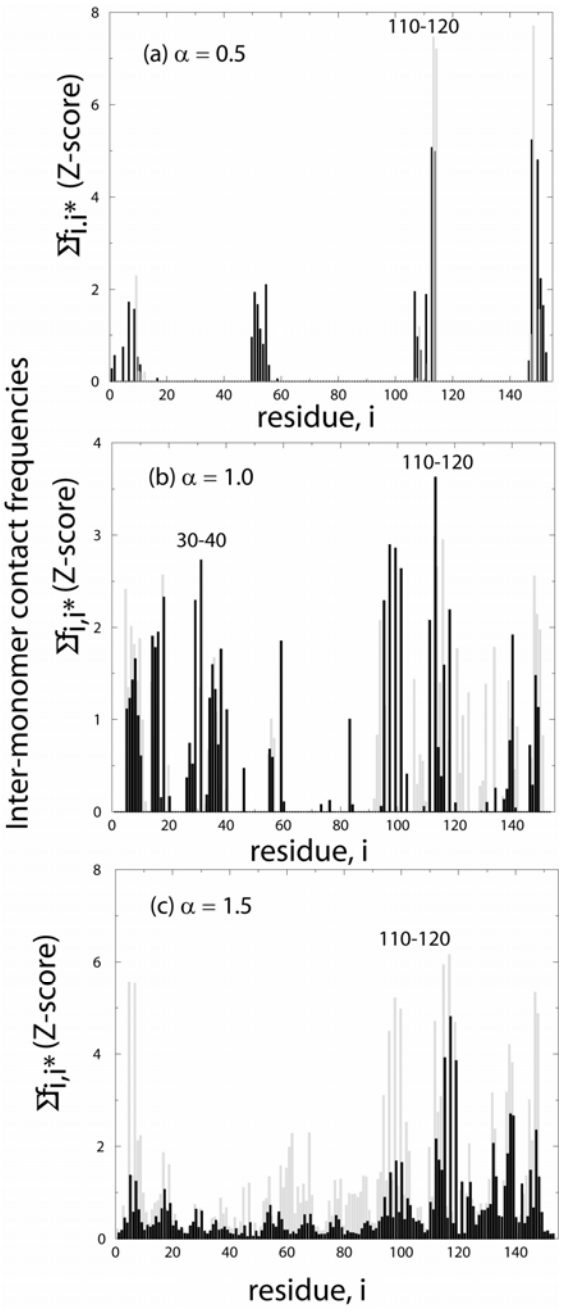


Figure 4.4. **Inter-monomer interactions during misfolding**. (a) The cumulative frequency of inter-monomer contacts at (a) α=0.5, (b) α=1, and (c) α=1.5 at *T*=0.65 (black) and *T*=0.75 (grey). As the strength of domain swapped interactions increases, the native dimer interface is lost and several competing dimeric structures are populated.

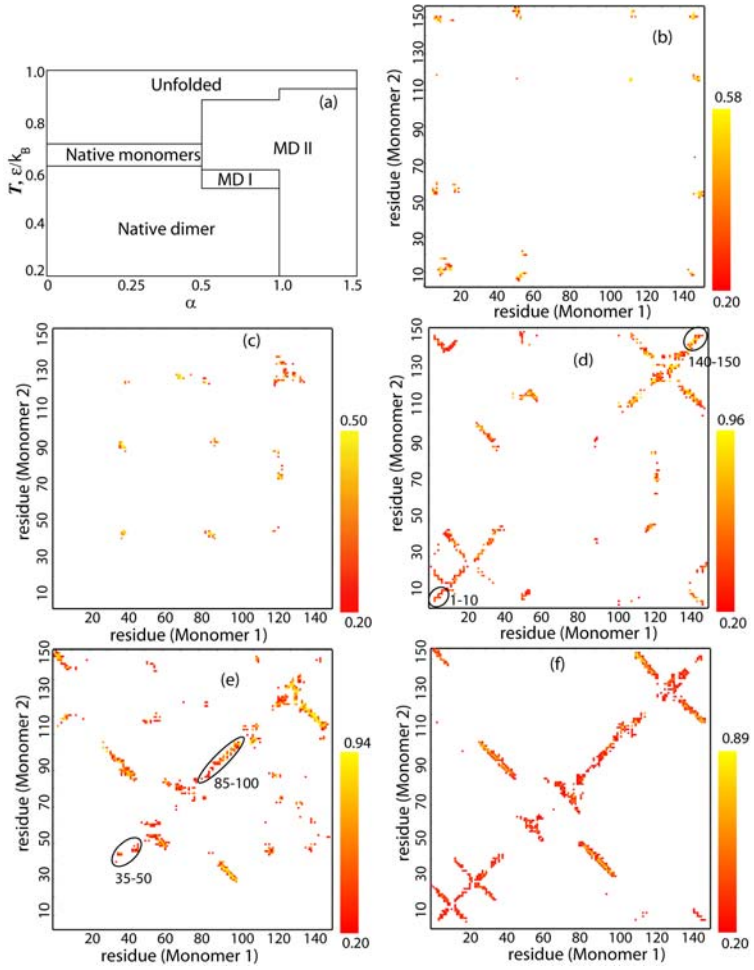


Figure 4.5. **Self-association and domain-swapping.** (a) A "phase diagram" summarizing the dimer topologies observed as a function of α and *T*. At low temperatures and α-values, native topologies (dimer, monomer) are observed, whereas at higher α-values (α=1.0 and 1.5), domain-swapping interactions lead to the appearance of modified dimers (MD I and MD II). Contact frequency maps corresponding to the inter-monomer contacts in (b) native and (c)-(f) modified dimers. Domain-swapping induced by high values of α and *T* is associated with the formation of self-association contacts (between corresponding identical elements of the two chains) which lie on the diagonal of the contact map. (c) MD I, for example at α=1.0 and T=0.5 is characterized by a small number of inter-monomer contacts in isolated elements of the structures, whereas MD II, for example at (d) α=1.5, T=0.4 (e) α=1.5, T=0.5 and (f) α=1.0, T=0.8 is characterized by extensive interactions between β-strands from the two monomers. Topologies observed for both MD I and MD II are diverse. Two representative structures of modified are shown in (g) and (h). These structures are formed by in-register interactions between residues 110-120 (red) or residues 35-45 (blue). We propose that there are multiple mechanisms by which these modified dimers can further propagate the aggregate structures and eventually form fibrils.

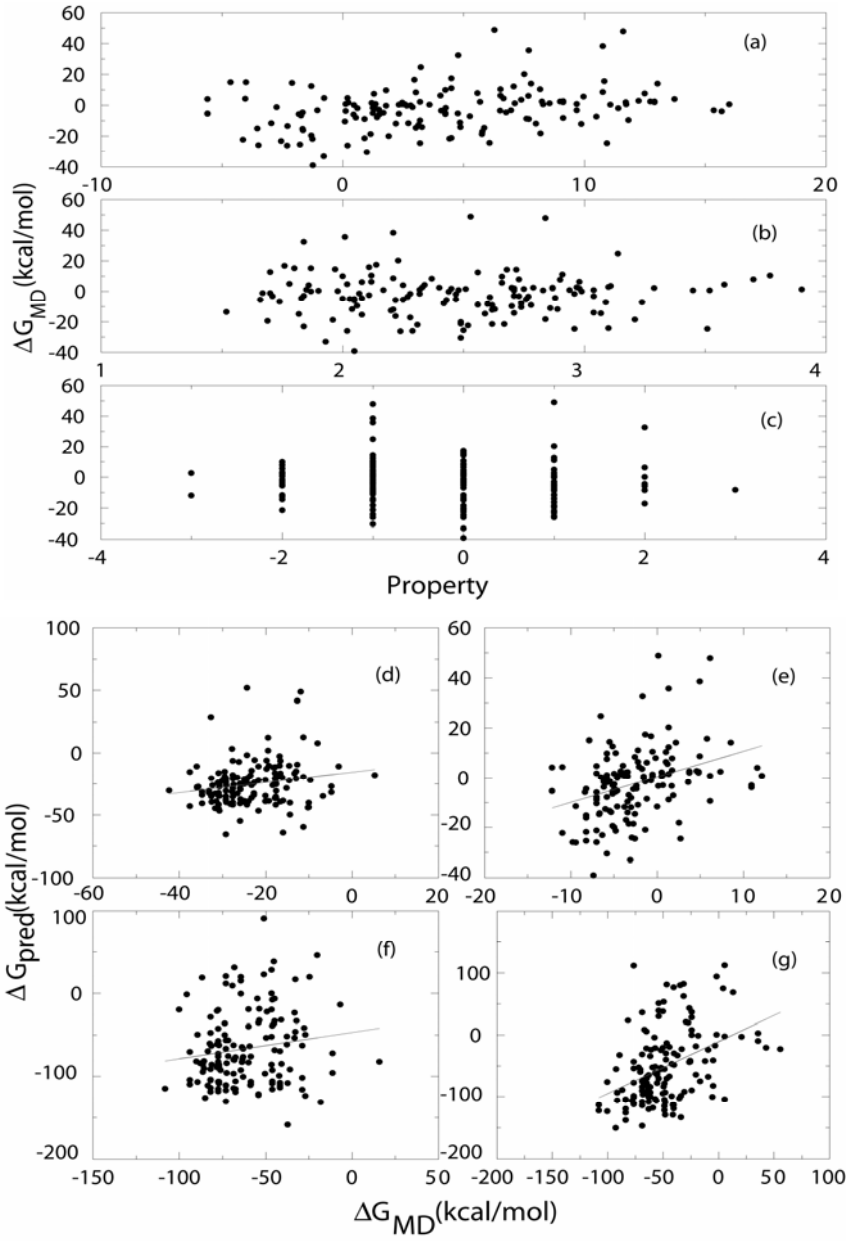


Figure 4.6. **The correlation between the free energies of dimerization into anti-parallel β-strands obtained in MD simulations with physical-chemical properties of the fragment.** Correlations of free-energy of association with (a) the average hydrophobicity (b) β-sheet propensity (c) net charge at pH 7, of the fragment. A linear combination of the three properties also does not correlate well with the free energies calculated from MD simulations of fragments mounted on (e) anti-parallel dimeric (f) parallel dimeric (g) anti-parallel tetrameric (h) parallel tetrameric template structures. We conclude that the free energy of oligomerization calculated in the MD simulation is not sufficiently explained by these physico-chemical properties of the sequence.

CHAPTER 5

FOLDING AND MISFOLDING OF SOD1 MONOMER

This chapter has been published as a manuscript in *The Journal of Molecular Biology*, and is formatted in the journal style.

5.1 Summary

Cu, Zn superoxide dismutase (SOD1) has been implicated in the familial form of the neurodegenerative disease Amyotrophic Lateral Sclerosis (ALS). It has been suggested that mutant mediated SOD1 misfolding/aggregation is an integral part of the pathology of ALS. We study the folding thermodynamics and kinetics of SOD1 using a hybrid molecular dynamics approach. We reproduce the experimentally observed SOD1 folding thermodynamics and find that the residues which contribute the most to SOD1 thermal stability are also crucial for apparent two-state folding kinetics. Surprisingly, we find that these residues are located on the surface of the protein and not in the hydrophobic core. Mutations in some of the identified residues are found in patients with the disease. We argue that the identified residues may play an important role in aggregation. To further characterize the folding of SOD1, we study the role of cysteine residues in folding and find that non-native disulfide bond formation may significantly alter SOD1 folding dynamics and aggregation propensity.

5.2 Introduction

ALS is the most common motor neuron disease in human adults that is characterized by selective motor neuron death [1-4]. In approximately 10% of the cases of ALS, the disease is inherited, thus called Familial ALS (FALS) [1]. Mutations in the cytoplasmic enzyme SOD1 were identified as the primary cause of approximately 20% of FALS cases [5, 6]. More than 90 mutations have been identified so far. The FALS mutations are scattered throughout the primary sequence and three-dimensional structure of SOD1, which exists as a homo-dimer in native state (each monomer is 153 amino acids in length) [7]. One hypothesis to explain the toxic gain-of-function of the mutants is that the toxicity is derived from intracellular aggregates and/or failure of SOD1 degradation, which is supported by the observation that in both mice and cell culture models, death of mutant neurons is preceded by formation of cytoplasmic aggregates containing mutant SOD1 [8-10]. In addition, SOD1 knockout mice do not develop motor neuron disease [11]. Aggregation or misfolding is therefore a characteristic of SOD1 mediated FALS. Toxicity may arise through (i) aberrant chemistry due to misfolded mutants [12-14], (ii) saturation of essential cellular machinery such as chaperones and proteasome components [15, 16], or (iii) oligomeric forms of the aggregate may themselves be toxic to the cells, as has been suggested for other neurodegenerative diseases [17].

Aggregation of proteins is mediated by a variety of factors including native state stability [18, 19], β-sheet propensity [20], net charge and overall hydrophobicity [21]. In addition, folding and aggregation are kinetically competing processes [20, 22, 23]. Thus, it is crucial to study the folding thermodynamics and kinetics of SOD1 in order to understand the causes of SOD1 misfolding, aggregation and toxicity. A microscopic picture of SOD1 folding

and/or aggregation dynamics of SOD1 is not available from either theory or experiment. We reconstruct the SOD1 folding mechanism from molecular dynamics simulations.

Neither the structure of SOD1 aggregates nor the mechanism of aggregation is fully understood. Several experimental studies have characterized the folding thermodynamics of SOD1 and the associated FALS mutants [24-30]. Lepock et al. [27] and Rodriguez et al. [30] measured the stability of mutants in differential scanning calorimetric studies of the irreversible unfolding of the distinctly metallated species of the enzyme, and showed that mutants are destabilized relative to the wild type. In addition, a crystal structure of the G37R mutant [31] shows higher atomic displacement parameters (B-factors) for the side-chains compared to the wild type, but the backbone conformation is not significantly different, thus indicating greater molecular flexibility in some portions of the structure. It was recently shown that for some mutants, the apo-state (without metals) of the mutants is markedly destabilized compared to that of the wild-type, while the holo-state (with metals) of both mutants and wild-type is not significantly affected [32]. It is, therefore, possible that the apo-state of SOD1 – the subject of our studies – is implicated in misfolding and aggregation.

Most Cu, Zn superoxide dismutases, including human SOD1, have been shown to undergo irreversible aggregation on exposure to temperatures higher than their respective melting temperatures. The presence of free cysteines, those that do not form disulfide bonds in the native state, is known to be involved in this phenomenon. It is well-established that improper disulfide bond formation, concomitant with cysteine oxidation, possibly enhanced by metals in SOD1, is a cause of heat-denatured aggregation [33]. In the case of bovine SOD1, substitution by site-directed mutagenesis of free cysteine residues was found to greatly increase the reversibility of denaturation without substantially affecting the conformational stability [25,

27]. Yeast SOD1 [34] and *E. coli* SOD1 [35], neither of which contain free cysteines, show partial and full reversibility respectively upon denaturation. However, SOD1 from *P. leiognathi* also denatures irreversibly despite having no free cysteines [36]. Clearly, the presence of free cysteines is one of the crucial factors (but not the only factor) responsible for irreversible denaturation. One plausible reason for the effect of free cysteines is that the formation of improper disulfide bonds makes SOD1 more prone to kinetic traps. Therefore, we also study the effect of disulfide bonds on SOD1 folding reversibility.

An atomic resolution simulation of the folding process using traditional molecular mechanics force-fields is difficult by direct computational approaches because of the vast dimensionality of the protein conformational space [37]. Simplified models such as the Gō-model [38, 39] provide a powerful alternative to study folding because of their ability to simulate folding on reasonable time scales and to reproduce the basic thermodynamic and kinetic properties of proteins both on- [40-42] and off-lattices [43, 44]. In the Gō model, the energy of the protein is expressed as a sum of pair-wise native contact energies. A native contact exists between amino acids if they are closer to each other than a cut-off distance in the native state. Thus, in the Gō approximation, folding can be construed as a transition from a state with no or a small number of native contacts to a state with all native contacts. Despite their apparent simplicity, native topology-based approaches have yielded results in agreement with experiments, in particular for small proteins (< 100 amino acids) [45, 46] reflecting the underlying simplicity in the folding of these proteins. However, a Gō model of longer proteins may lead to discrepancy in protein folding dynamics between simulation and experiment because sequence-specific interactions may significantly alter the protein folding dynamics.

Thus, we develop methodology to incorporate sequence specific information of the relative contribution of various amino-acids into the Gō model of SOD1.

5.3 Results

We perform DMD simulations using the unscaled Gō model, to determine the thermodynamic properties of a model of SOD1 (as described in Methods). The temperature dependence of average potential energy is shown in Fig. 1a. There is a sigmoidal increase in the potential energy with increasing temperature. At low temperatures, SOD1 is present mostly in its native state, as evidenced by the small (2.83Å) root-mean-square deviation, RMSD, from the native state (Fig. 1b), and the small fluctuations in the radius of gyration, R_g, around its native state value (14.2 Å) (data not shown). At high temperatures, SOD1 is unfolded with the RM SD from the native state being greater than 40 Å, demonstrating the loss of any structural similarity with the native state. Also, R_g is approximately three times its value in the native state (data not shown), indicating that the average distance between any two amino acids is approximately three times that in the native state.

In a typical trajectory (Fig. 2a) at the temperature corresponding to the mid-point of the abrupt change in potential energy (Fig. 1a) dependence on temperature (defined as the folding transition temperature T_F), we observe three distinct states: folded, unfolded and intermediate. A histogram of potential energies of protein conformations near T_F is, therefore, tri-modal (Fig. 2c), instead of the expected bi-modal for a two-state protein, indicating the presence of a folding intermediate. The detection of the intermediate state is in contrast with the experimental observation of the two-state folding dynamics of the SOD1 monomer [28]. Thus, the unscaled Gō model for SOD1 does not accurately reproduce the folding dynamics

of SOD1. Next, we calculate the pair-wise interaction energies in the native state of SOD1 monomer and then employ the scaled Gō model, to uncover SOD1 folding thermodynamics and kinetics.

A few contacts make a large thermodynamic contribution to SOD1 native state stability.

We decompose the stability of SOD1 into pair-wise contributions (as described in Methods) that are plotted in the lower triangle in Fig. 3 (the energy map). In the energy map, some contacts, corresponding to both short-range and long- range interactions, have significantly stronger CHARMM interaction energies (from three to five times) than average. We find that there are 16 long-range interactions that have large (about five times larger than the average) interaction energies: Lys3-Glu21, Lys3- Glu153, Lys30-Glu100, Ser34-Asp96, Lys36-Asp92, Glu40-Lys91, Glu49-Arg115, Arg69- Glu77, His71-Asp83, Arg79-Asp83, Arg79-Asp101, His80-Asp83, Asp101-Val103, Glu121- Ser142, Asp125-Lys128, Glu133-Lys136. The residues involved in the strong long-range contacts are highlighted in Fig. 4 and are involved in contacts between oppositely charged amino acid pairs on the surface of the protein. Out of the 16 identified interactions, 6 have been found to be directly disrupted in patients with FALS as a result of point mutations (listed in Table I) in the residues making these contacts 1 . In most cases, the mutations correspond to the change of a charged amino acid for an oppositely charged or a neutral one, suggesting an important role these key contacts play in SOD1 folding.

Scaling thermodynamically important contacts modulates folding kinetics in the Gō model.

We test the importance of the identified amino acids in DMD simulations by scaling their contact strength in the Gō potential. We increase the depth of the interaction well of the identified key contacts to five times their original value in the unscaled Gō model (b = 5). Using this scaled Gō model, we perform DMD simulations to calculate the thermodynamic properties of the SOD1. We find that the potential energy, RM SD and Rg (data not shown) exhibit the same sigmoidal trend as for the unscaled Gō model, but the folding transition temperature increases to approximately 1.1 from its earlier value of 1.0. Remarkably, a typical trajectory near the folding transition temperature (T_F = 1.15) shows (Fig. 2b) that the intermediate state is no longer populated and the protein only exists in either the folded or the unfolded state (Fig. 2b and d) near T_F . Thus, by strengthening these key contacts we reproduce the experimentally observed two-state folding of SOD1 monomer [28, 29]. As control, we randomly select several combinations of sixteen contacts and scale them by the same procedure. In such control simulations, we do not observe such effect on the folding kinetics (Fig. 2b), thus demonstrating that the effect of strengthening the identified contacts is specific. Importantly, this result uncovers a crucial connection between the folding thermodynamics and kinetics of SOD1: in our model, the residues that contribute the most to the SOD1 thermodynamic stability of the native state are also crucial for keeping the folding kinetics two-state. Improper disulfide bonds cause kinetic trapping of SOD1. Human SOD1 is known to aggregate irreversibly on heat denaturation [30] in vitro. One of the reasons for the irreversibility is improper disulfide bond formation between free cysteines. There are four cysteine residues in the SOD1 monomer, at positions 6, 57, 111 and 146 of which residues 57

and 146 form a disulfide linkage in the native state. We study the effect of non-native disulfide bond formation by starting from a fully unfolded conformation, generated at T=2.3 ($>>T_F$), and cooling the system to low temperature T=0.4 ($<T_F$) slowly, while allowing the possibility of interactions between any of the four cysteines to form improper disulfide bond(s). We model disulfide bonds by an attractive potential (b = 2, $\gamma = -1$ in Eq. (2)). As a control, we repeat the annealing procedure without these non-native disulfide bonds (b = 1, $\gamma = 1$ in Eq. (2)). We observe that out of 20 trajectories with improper disulfide interactions, 4 result in SOD1 structures that are trapped in non-native conformations while all of the 20 control trajectories fold to the correct native structure on the same time scale, implying that occurrence of improper disulfide linkage increases the number of kinetic traps for the SOD1 chain en route to its native state. The above observation is statistically significant (the probability of observing this result by chance: p-value < 0.01). The increased number of kinetic traps makes intermediate states more populated and interaction of these intermediate states may make SOD1 prone to aggregation. We conclude that free cysteine residues play an important part in the kinetics of folding, and possibly, aggregation.

5.4 Discussion

Experimental evidence suggests that the unfolding of dimeric superoxide dismutase occurs in a three state equilibrium [28, 29].

$$D \rightleftarrows 2M \longrightarrow U \qquad (1)$$

where D, M and U are SOD1 dimer, monomer and unfolded state respectively. In order to understand the effect of mutations on SOD1 folding, it is important to study the effect of mutation on both processes in Eq. (1). Here, we study the second process (M$\leftrightarrow$U) and

report thermodynamically and kinetically important residues for SOD1 monomer folding. Using our hybrid method, which incorporates molecular mechanics energetics into the Gō model, we identify the key residues involved in the folding of the protein. Surprisingly, the key residues are found on the surface of the protein and not in its hydrophobic core (Fig. 4). By comparing the identified residues with a comprehensive list of mutations associated with FALS (http://www.alsod.org), we find that, 6 out of the 16 contacts identified in our simulations are found to be implicated in the disease. In Table I, we list those FALS mutations that disrupt contacts identified from our simulations. We find that these mutations replace a charged residue in the wild-type protein to a neutral or oppositely charged residue in the mutant, thereby disrupting the electrostatic interaction. The presence of mutations disrupting these contacts further underscores the importance of these key contacts and indicates that the disruption of even one of these contacts may potentially result in protein misfolding. There have been no experimental studies characterizing SOD1 monomer folding kinetics. However, other members of the immunoglobulin-like (Ig-like) superfold family [47], to which SOD1 belongs, are experimentally well-characterized. Clarke and colleagues have established that the folding of pathways of Ig-like proteins share some common features [48]. They observe that for five structurally similar Ig-like proteins, the folding rates are correlated with thermal stabilities, suggesting a similar folding pathway for all members of the family. Thus, for Ig-like proteins, the kinetic importance of residues in folding (Φ-values [49]) is largely determined by the topology of the fold. We, therefore, expect that the residues important for SOD1 folding are in structurally equivalent positions to other members of the Ig-like super-fold family, whose folding kinetics is experimentally well-studied [48]. We choose Tnf3, the fibronectin type III domain of human tenascin, for which, based on Φ-

values, eight residues in the hydrophobic core were identified as part of the folding nucleus [50]. We identify residues in SOD1 that are structurally equivalent to the nucleus residues in Tnf3 by performing a structural alignment of the two proteins using DALI [51] and list these residues in Table II. We discover that two of the residues corresponding to this putative nucleus correspond to the strongly interacting key residues identified by our simulations, while three others are nearest neighbors of other residues we identify. This is significant because (i) mutations in the nucleus residues, being in proximity to the identified key residues, may disrupt the network of interactions of key residues and affect folding kinetics and (ii) in the Gō model entropic contributions are included as effective energetic interactions. Therefore, the identified key residues may be kinetically crucial because of the entropic contribution of the corresponding region of the structure (which is reflected by high Φ-values of the neighboring nucleus residues). Therefore, kinetic evidence further supports our hypothesis that residues important for SOD1 native thermodynamics are important for folding kinetics. Richardson and Richardson have pointed out [52] that the interactions between edges of β-strands are crucial to regulate protein aggregation propensity. Naturally occurring β-sheet proteins prevent aggregation by protecting the edges of β strands through a multitude of fold- specific mechanisms. In particular, for Ig-like β-sandwiches, strategically placed side-chain charges at the edges and loop-crossing are dominant mechanisms which shield the edges of the sheets. A recently solved crystal structure of mutant apo-SOD1 supramolecular assembly [53] identified three regions of the SOD1 structure where non-native gain of interaction was observed leading to amyloid like arrangement of SOD1 in crystals: the cleft between strands S5 and S6 (residues 83-100), the zinc loop (residues 65-78) and the electrostatic loop (residues 125-142). It was noted that disorder in the zinc and

electrostatic loops promoted non-native interactions with the S5-S6 cleft, which is the interface between the two β-sheets of SOD1 and hence a likely site for edge-to-edge aggregation [52]. We find that most of the key contacts for proper folding are made by residues in the three regions identified above. In particular, we see that contacts on both the edge-strands S5 and S6 play a crucial role in folding kinetics. The contacts made by residues Asp92, Asp96 and Lys91 which form the turn in the S5-S6 hairpin are crucial to seal the interface between the two β-sheets and likely prevent gain of non-native interactions, as shown in Fig. 5. Thus, our simulations constitute another piece in the growing body of evidence which implicates edge protection as the dominant mechanism by which natural proteins avoid aggregation. We postulate that the disruption of key interactions is a plausible scenario for mutant mediated aggregation of SOD1. The re-organization of the S5-S6 edge may be the dominant feature which distinguishes the aggregation-prone misfolded state of SOD1, from the native state. The disruption of charge-charge interactions of the side-chains of the identified key residues, which stabilize the β-barrel architecture of SOD1, might expose the edges of β-strand main- chain to interact with other structures, thus leading to aggregation. Wang and Hecht [54] have shown that strategically designed mutations in β-strand overhangs disfavor aggregation of normally aggregating proteins, further supporting our hypothesis.

We show that the presence of oxidizable cysteines enhances the kinetic trapping during folding. In the presence of metals, cysteine oxidation is enhanced suggesting a plausible cause for the misfolding of SOD1. However, it has recently been suggested [32] that misfolding of SOD1 likely takes place in the apo-enzyme state, when the metals have not yet been delivered to SOD1. In the reducing environment of the cell, cysteine-mediated

misfolding/aggregation of the apo-enzyme is, therefore, not a likely mechanism of SOD misfolding/aggregation. However, our observations do suggest a generic mechanism for SOD1 misfolding especially for mutations that do not affect the native state conformational stability [30]. In the scenario similar to the one we have presented for free cysteines, other mutations may affect the folding kinetics of SOD1 by introducing non-native interactions along the folding pathway, leading to kinetic trapping without substantially affecting the native state stability.

5.5 Methods

The unscaled Gō model. We use the Discrete Molecular Dynamics (DMD) algorithm [43, 55, 56] to study the folding thermodynamics of SOD1. DMD has recently been used to study the folding kinetics [45, 46, 57, 58] and aggregation of proteins [20, 59]. We model the SOD1 chain by the beads-on-a-string model developed by Ding et al. [46] with beads corresponding to all Cα and Cβ atoms and constraints between neighboring beads to mimic real protein flexibility. We study the folding of one monomer derived from the crystal structure of SOD1 (Protein data bank access code: 1SPD). We use the Gō potential [38] to model the interaction energy, V_{ij}, between Cβ atoms (Cα for Gly) *i* and *j* of SOD1:

$$V_{ij} = \begin{cases} +\infty, & | r_i - r_j | \le r_0 \\ \gamma u_{ij}, & r_0 < | r_i - r_j | \le r_1 \\ 0, & | r_i - r_j | > r_1 \end{cases} \qquad (2)$$

where $|r_i - r_j|$ is the distance between atoms *i* and *j*. The parameters r_0 and r_1 are the hard-core diameter (set to 3.25 Å) and the cut-off distance (7.5 Å) between Cβ atoms (Cα for glycine) in the native state respectively. We assign attractive interaction for native contacts ($\gamma = -1$), defined to exist between pairs of residues whose Cβ atoms (Cα for glycine) are closer than

7.5 Å in the native state, and repulsive interaction ($\gamma = 1$) for non-native contacts, defined to exist between Cβ atoms (Cα for glycine) which are farther than 7.5 Å in the native state.

The depth of the attractive square-well u_{ij} in Eq. (2) determines the strength of the attractive interaction between two amino acids. First, we consider the case where all contacts are equally strong, i.e. the depth of each attractive well is u. We call this case the unscaled Gō model and use it to perform DMD simulations of SOD1. We find that our model undergoes a collapse transition to its native state through a meta-stable intermediate (see Results) in contrast with the experimental observation of two-state folding dynamics of SOD1 monomer [28, 29]. A possible reason for the existence of meta-stable intermediates is that in the unscaled Gō model, all native contacts have equal strength. While such an approximation holds for smaller proteins (such as Src SH3 [45]) with relatively low number of contacts (162), SOD1 monomer has 495 native contacts and the assumption of equal contact strength overestimates the population of intermediate species. Thus, in order to include sequence specificity of native interactions, we increase the strength of few specific native contacts according to an all-atom model of SOD1 described below by making the depth of the attractive well of some chosen native contacts bu_{ij}, where b is a scale factor. We call this the scaled Gō model. By scaling Gō potential according to an all atom model of the protein, we effectively include the sequence information.

The scaled Gō model. To determine the relative strength of each individual native contact in SOD1, we use an effective energy function to decompose the total energy of the protein as a sum of pairwise terms [60]:

$$E = \sum_i \sum_{j>i} E_{ij} \qquad (3)$$

where the summation is taken is over all pairs of residues of SOD1 and E_{ij} is the energy of interaction between residues *i* and *j*. To determine the terms E_{ij}, we perform simulations with the molecular mechanics force field CHARMM [61] to construct an all atom model of the SOD1 monomer in the native state, starting from its crystal structure (PDB accession code:1SPD). Recently, higher resolution crystal structures of apo-SOD1 have appeared in the literature [62], but our results are not qualitatively affected by starting with the higher resolution structure. We add hydrogen atoms and remove any steric clashes occurring in the structure by energy minimization. The interactions which contribute the most to the relative contact interaction strength are the non-bonded interactions and the solvent-induced interaction screening [60]:

$$E_{ij} = E_{ij}^{non-bonded} + G_{ij} \qquad (4)$$

We compute $E^{non-bonded}$ as the sum of the Lennard-Jones and electrostatic interaction energies. The solvation free energy, G_{ij}, measures the screening of the pairwise interaction due to polarization of the solvent. We calculate G_{ij} using an implicit description of the solvent [63]. We approximate the solvent as a continuous dielectric medium with a dielectric constant ($\varepsilon = 80$) of water and the protein as a continuous medium with dielectric constant ($\varepsilon = 1$) of vacuum. We compute the solvation energy by solving the Poisson-Boltzmann equation using the finite difference method [64] available in CHARMM. Following Dominy et al. [65], in order to decompose the solvation energy into pairwise contributions, we eliminate all charges in the protein except those corresponding to residue *i* and calculate its "self" solvation energy U_i. Similarly, we compute U_j and U_{ij} when charges corresponding to both *i* and *j* are present. Then, the solvent mediated screening of the interaction is:

$$G_{ij} = U_{ij} - [U_i + U_j] \qquad (5)$$

So, we calculate the effective pairwise energy contribution of a contact in the folded state. The E_{ij} values that we thus obtain correspond to the folded state of the protein. In order to calculate the pairwise contribution of residues to the stability, we subtract the value of the contact energy in the unfolded state from the E_{ij} values obtained above. The unfolded state is an ensemble of conformations. It has been noted that there is considerable native like structure in the unfolded state [66-68], meaning that the average environment of a contact is likely the same as in the folded state, but this contact is formed less frequently. Therefore, we model the unfolded state by simulating the unscaled Gō model near T_F and estimating the frequency of native contact formation. The total contribution of a given contact to the stability can be written as:

$$\Delta E_{ij} = E_{ij}(1 - p_{ij}), \qquad (6)$$

where E_{ij} is the contact energy in the folded state and p_{ij} is the probability of forming the contact in the unfolded state. The energy and contact maps are plotted in Fig. 3.

5.6 Conclusions

We present the folding thermodynamic and kinetic analysis of SOD1 using a hybrid molecular dynamics method. We show that our simple model qualitatively reproduces the folding of SOD1 observed in experiments. We find that interactions at the edges of SOD1 β sheets and the zinc, electrostatic and cross-over loops are the key interactions that modulate the folding of SOD1 suggesting that mutants destabilizing these interactions may make the enzyme more prone to aggregation. In particular, we identify the S5-S6 cleft (residues 83-100) as a likely site for reorganization to yield an aggregation prone misfolded state, an observation consistent with recent experimental evidence [53] and the negative design

paradigm [52]. We also provide a plausible explanation for the phenomenon of increased reversibility of folding upon mutation of free cysteine residues. We find good agreement between the positions identified by us to be crucial for folding kinetics and those identified by experiments [50] on other members of the super-fold family (Ig-like proteins) to which SOD1 belongs. Further, we find that mutations at six of the residue contacts we identify are implicated in FALS.

5.7 References

1. Cleveland, D. W. & Rothstein, J. D. (2001). From Charcot to Lou Gehrig: deciphering selective motor neuron death in ALS. Nature Rev. Neurosci. 2, 806-819.

2. Beckman, J. S., Estevez, A. G., Crow, J. P. & Barbeito, L. (2001). Superoxide dismutase and the death of motorneurons in ALS. Trends Neurol. Sci. 24, S15-S20.

3. Rowland, L. P. & Shneider, N. A. (2001). Amyotrophic lateral sclerosis. N. Engl. J. Med. 344, 1688-1700.

4. Siddique, T. (1991). Molecular genetics of familial amyotrophic lateral sclerosis. Adv. Neurol. 56, 227-231.

5. Rost, D. R., Siddique, T., Patterson, D., Figlewicz, D. A., Sapp, P., Hentati, A., Donaldson, D., Goto, J., et al (1993). Mutations in Cu/Zn superoxide-dismutase gene are associated with familial amyotrophic-lateral-sclerosis. Nature, 362, 59-62.

6. Gurney, M. E., Pu, H., Chiu, A. Y., Canto, M. C. D., Polchow, C., Alexander, D. D., Caliendo, J., Hentati, A., Kwon, Y. W., Deng, H. X., Chen, W., Zhai, P., Sufit, R. L. & Siddique, T. (1994). Motor neuron degeneration in mice that express a human Cu,Zn superoxide dismutase mutation. Science, 264, 1772-1775.

7. Gaudette, M., R. Hirano, M. & Siddique, T. (2000). Current status of SOD1 mutations in familial amyotrophic lateral sclerosis. Amyotroph. Lateral Scler. Other Motor Neuron Disord. 1, 83-89.

8. Johnston, J. A., Dalton, M. J., Gurney, M. E. & Kopito, R. R. (2000). Formation of high molecular weight complexes of mutant Cu,Zn-superoxide dismutase in a mouse model for familial amyotrophic lateral sclerosis. Proc. Natl. Acad. Sci. U. S. A. 97, 12571-12576.

9. Bruijn, L. I., Houseweart, M. K., Kato, S., Anderson, K. L., Anderson, S. D., Ohama, E., Reaume, A. G., R. W. Scott, R. & Cleveland, D. W. (1998). Aggregation and motor neuron toxicity of an ALS-linked SOD1 mutant independent from wild-type SOD1. Science, 281, 1851-1854.

10. Durham, H. D., Roy, J., Dong, L. & Figlewicz, D. A. (1997). Aggregation of mutant Cu/Zn superoxide dismutase proteins in a culture model of ALS. J. Neuropathol. Exp. Neurol. 56, 523-530.

11. Reaume, A. G., Elliott, J. L., Hoffman, E. K., Kowall, N. W., Ferrante, R. J., Siwek, D. F., Wilcox, H. M., Flood, D. G., Beal, M. F., Jr, R. H. B., Scott, R. W. & Snider, W. D. (1996). Motor neurons in Cu/Zn superoxide dismutase-deficient mice develop normally but exhibit enhanced cell death after axonal injury. Nature Genet. 13, 43-47.

12. Bromberg, M. B. (1999). Pathogenesis of amyotrophic lateral sclerosis: a critical review. Curr. Opin. Neurol. 12, 581-588.

13. Cleveland, D. W. (1999). From Charcot to Lou Gehrig: mechanisms of selective motor neuron death in ALS. Neuron, 24, 515-520.

14. Elliot, J. L. (2001). Zinc and copper in the pathogenesis of amyotrophic lateral sclerosis. Prog. Neuro-Psychopharmacol. & Biol. Psychiat. 25, 1169-1185.

15. Tran, P. B. & Miller, R. J. (1999). Aggregates in neurodegenerative disease: crowds and power? Trends Neurol. Sci. 22, 194-197.

16. Bruening, W., Roy, J., Giasson, B., Figlewicz, D. A., Mushynski, W. E. & Durham, H. D. (2001). Up-regulation of protein chaperones preserves viability of cells expressing toxic Cu/Zn- superoxide dismutase mutants associated with amyotrophic lateral sclerosis. J. Neurochem. 72, 693-699.

17. Walsh, D. M., Klyubin, I., Fadeeva, J. V., Rowan, M. J. & Selkoe, D. J. (2002). Amyloid-beta oligomers: their production, toxicity and therapeutic inhibition. Biochem. Soc. Trans. 30, 552-557.

18. Chiti, F., Taddei, N., Bucciantini, M., White, P. M., Ramponi, G. & Dobson, C. M. (2000). Mutational analysis of the propensity for amyloid formation by a globular protein. EMBO J. 19, 1441-1449.

19. Canet, D., Last, A. M., Tito, P., Sunde, M., Spencer, A., Archer, D. B., Redfield, C., Robinson, C. V. & Dobson, C. M. (2002). Local cooperativity in the unfolding of an amyloidogenic variant of human lysozyme. Nature Struct. Biol. 9, 308-315.

20. Ding, F., Dokholyan, N. V., Buldyrev, S. V., Stanley, H. E. & Shakhnovich, E. I. (2002). Molecular dynamics simulation of C-Src SH3 aggregation suggests a generic amyloidogenesis mechanism. J. Mol. Biol. 324, 851-857.

21. Chiti, F., Calamai, M., Taddei, N., Stefani, M., Ramponi, G. & Dobson, C. M. (2002). Studies of the aggregation of mutant proteins in vitro provide insights into the genetics of amyloid disease. Proc. Natl. Acad. Sci. U. S. A. 6, 1005-1009.

22. Chiti, F., Taddei, N., Baroni, F., Capanni, C., Stefani, M., Ramponi, G. & Dobson, C. M. (2002). Kinetic partitioning of protein folding and aggregation. Nature Struct. Biol. 9, 137-143.

23. Dobson, C. M. (1999). Protein misfolding, evolution and disease. Trends Biochem. Sci. 24, 329-332.

24. Hallewell, R. A., Imlay, K. C., Lee, P., Fong, N. M., Gallegos, C., Getzoff, E. D., J. A. Tainer, J., Cabelli, D. E., Tekampolson, P., G. T. Mullenbach, G. & Cousens, L. (1991).

Thermosta- bilization of recombinant human and bovine Cu,Zn superoxide dismutases by replacement of free cysteines. Biochem. Biophys. Res. Commun. 181, 474-480.

25. McRee, D. E., Redford, S., Getzoff, E. D., Lepock, J. R., Hallewell, R. & Tainer, J. (1990). Changes in crystallographic structure and thermostability of a Cu,Zn superoxide-dismutase mutant resulting from the removal of a buried cysteine. J. Biol. Chem. 265, 14234-14241.

26. Parge, H. E., Hallewell, R. A. & Tainer, J. A. (1992). Atomic structures of wild-type and thermostable mutant recombinant human Cu,Zn superoxide-dismutase. Proc. Natl. Acad. Sci. U. S. A. 89, 6109-6113.

27. Lepock, J. R., Frey, H. E. & Hallewell, R. A. (1990). Contribution of conformational stability and reversibility of unfolding to the increased thermostability of human and bovine superoxide dismutase mutated at free cysteines. J. Biol. Chem. 265, 21612-21618.

28. Stroppolo, M. E., Malvezzi-Campeggi, F., Mei, G., Rosato, N. & Desideri, A. (2000). Role of the tertiary and quaternary structures in the stability of dimeric copper,zinc superoxide dismutases. Arch. Biochem. Biophys. 377, 215-218.

29. Mei, G., Rosato, N., Silva, N., Rusch, R., Gratton, E., Savini, I. & Finazzi-Agro, A. (1992). Denaturation of human copper-zinc superoxide dismutase by guanidine hydrochloride: a dy- namic fluorescence study. Biochemistry, 31, 7224-7230.

30. Rodriguez, J. A., Valentine, J. S., Eggers, D. K., Roe, J. A., Tiwari, A., Brown Jr., R. H. & Hayward, L. J. (2002). 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.

31. Hart, P. J., Liu, H., Pellegrini, M., Nersissian, A. M., Gralla, E. B., Valentine, J. S. & Eisenberg, D. (1998). Subunit asymmetry in the three-dimensional structure of a human CuZnSOD mutant found in familial amyotrophic lateral sclerosis. Protein Science, 7, 545-

32. Lindberg, M. J., Tibell, L. & Oliveberg, M. (2002). Common denominator of Cu/Zn superoxide dismutase mutants associated with amyotrophic lateral sclerosis: decreased stability of the apo state. Proc. Natl. Acad. Sci. U. S. A. 99, 16607-16612.

33. Perry, L. J. & Wetzel, R. (1987). The role of cysteine oxidation in the thermal inactivation of T4 lysozyme. Protein Eng. 1, 101-105.

34. Lepock, J. R., Arnold, L. D., Torrie, B. H., Andrews, B. & Kruuv, J. (1985). Structural-analyses of various Cu,Zn superoxide dismuatses by differential scanning clorimetry and Raman-spectroscopy. Arch. Biochem. Biophys. 241, 243-251.

35. Battistoni, A., Folcarelli, S., Cervoni, L., Polizio, F., Dessideri, A., Giartosio, A. & Rotilio, G. (1998). Role of the dimeric structure in Cu,Zn superoxide dismutase. J. Biol. Chem. 273, 5655-5661.

36. Bourne, Y., Redford, S. M., Steinman, H. M., Lepock, J. R., Tainer, J. A. & Getzoff, E. D. (1996). Novel dimeric interface and electrostatic recognition in bacterial Cu,Zn superoxide dismutase. Proc. Natl. Acad. Sci. U. S. A. 93, 12774-12779.

37. Karplus, M. & Shakhnovich, E. I. (1994). Protein folding: theoretical studies of thermodynamics and dynamics. In Protein Folding, (Creighton, T., ed.),. W. H. Freeman and Company New York.

38. Gō, N. & Abe, H. (1981). Noninteracting local-structure model of folding and unfolding transition in globular proteins. I. formulation. Biopolymers, 20, 991-1011.

39. Abe, H. & Gō, N. (1981). Noninteracting local-structure model of folding and unfolding transition in globular proteins. II. application to two-dimensional lattice proteins. Biopolymers, 20, 1013-1031.

40. Shimada, J., Kussell, E. L. & Shakhnovich, E. I. (2001). The folding kinetics and thermody- namics of crambin using an all-atom monte-carlo simulation. J. Mol. Biol. 308, 79-95.

41. Clementi, C., Nymeyer, H. & Onuchic, J. N. (2000). Topological and energetic factors: what determines the structural details of the transition state ensemble and \en-route" intermediates for protein folding? an investigation for small globular proteins. J. Mol. Biol. 278, 937-953.

42. Abkevich, V. I., Gutin, A. M. & Shakhnovich, E. I. (1994). Specific nucleus as the transition state for protein folding: evidence from the lattice model. Biochemistry, 33, 10026-10036.

43. Dokholyan, N. V., Buldyrev, S. V., Stanley, H. E. & Shakhnovich, E. I. (1998). Molecular dynamics studies of folding of a protein-like model. Folding & Design, 3, 577-587.

44. Dokholyan, N. V., Li, L., Ding, F. & Shakhnovich, E. I. (2002). Topological determinants of protein folding. Proc. Natl. Acad. Sci. U. S. A. 99, 8637-8641.

45. Borreguero, J. M., Dokholyan, N. V., Buldyrev, S., Stanley, H. E. & Shakhnovich, E. I. (2002). Thermodynamics and folding kinetics analysis of the SH3 domain from discrete molecular dynamics. J. Mol. Biol. 318, 863-876.

46. Ding, F., Dokholyan, N. V., Buldyrev, S. V., Stanley, H. E. & Shakhnovich, E. I. (2002). Direct molecular dynamics observation of protein folding transition state ensemble. Biophys. J. 83.

47. Murzin, A. G., Brenner, S. E., Hubbard, T. & Chothia, C. (1995). SCOP: A structural classification of proteins database for the investigation of sequences and structures. J. Mol. Biol. 247, 536-540.

48. Clarke, J., Cota, E., Fowler, S. B. & Hamill, S. J. (1999). Folding studies of immunoglobulin- like beta-sandwich proteins suggest that they share a common folding pathway. Structure, 7, 1145-1153.

49. Fersht, A. R. (1995). Characterizing transition states in protein folding: an essential step in the puzzle. Curr. Opinion Struc. Biol. 5, 79-84.

50. Hamill, S. J., Steward, A. & Clarke, J. (2000). The folding of an immunoglobulin-like greek key protein is defined by a common-core nucleus and regions constrained by topology. J. Mol. Biol. 297, 165-178.

51. Holm, L. & Sander, C. (1995). DALI: a network tool for protein structure comparison. Trends Biochem. Sci. 20, 478-480.

52. Richardson, J. S. & Richardson, D. C. (2002). Natural -sheet proteins use negative design to avoid edge-to-edge aggregation. Proc. Natl. Acad. Sci. U. S. A. 99, 2754-2759.

53. Elam, J. S., Taylor, A. B., Strange, R. W., Antonyuk, S., Doucette, P. A., Rodriguez, J. A., Hasnain, S. S., Hayward, L. J., Valentine, J. S., Yeates, T. O. & Hart, P. J. (2003). Amyloid- like filaments and water-filled nanotubes formed by SOD1 mutant proteins linked to familial ALS. Nature Struct. Biol. 10, 461-467.

54. Wang, W. X. & Hecht, M. H. (2002). Rationally designed mutations convert de novo amyloid- like fibrils into monomeric beta-sheet proteins. Proc. Natl. Acad. Sci. U. S. A. 99, 2760 -2765.

55. Zhou, Y., Karplus, M., Wichert, J. M. & Hall, C. K. (1997). Equilibrium thermodynamics of homopolymers and clusters: molecular dynamics and monte carlo simulations of system with square-well interactions. J. Chem. Phys. 107, 10691-10708.

56. Smith, S. W., Hall, C. K. & Freeman, B. D. (1997). Molecular dynamics for polymeric fluids using discontinuous potentials. J. Comput. Phys. 134, 16-30.

57. Zhou, Y. & Karplus, M. (1999). Interpreting the folding kinetics of helical proteins. Nature, 401, 400-403.

58. Zhou, Y. & Karplus, M. (1999). Folding of a model three-helix bundle protein: a thermody- namic and kinetic analysis. J. Mol. Biol. 293, 917-951.

59. Smith, A. V. & Hall, C. K. (2001). Protein refolding versus protein aggregation: computer simulations on an intermediate-resolution protein model. J. Mol. Biol. 312, 187-202.

60. Paci, E., Vendruscolo, M. & Karplus, M. (2002). Native and non-native interactions along protein folding and unfolding pathways. Proteins: Struc. Func. Genet. 47, 379-392.

61. Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S. & Karplus, M. (1983). CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. J. Comput. Chem. 4, 187-217.

62. Strange, R. W., Antonyuk, S., Hough, M. A., Doucette, P. A., Rodriguez, J. A., Hart, P. J., Hayward, L. J., Valentine, J. S. & Hasnain, S. S. (2003). The structure of holo and metal-deficient wild-type human Cu, Zn superoxide dismutase and its relevance to familial amy otrophic lateral sclerosis. J. Mol. Biol. 328, 877-891.

63. Sharp, K. A. (1998). Electrostatic interactions in macromolecules. Curr. Opinion Struc. Biol. 3, R108-R111.

64. Nicholls, A. & Honig, B. (1991). A rapid finite-difference algorithm, utilizing successive over-relaxation to solve the Poisson-Boltzmann equation. J. Comput. Chem. 12, 435-445.

65. Dominy, B. N., Perl, D., Schmid, F. X. & Brooks III, C. L. (2002). The effects of ionic strength on protein stability: the cold shock protein family. J. Mol. Biol. 319, 541-554.

66. Zagrovic, B., Snow, C. D., Khaliq, S., Shirts, M. R. & Pande, V. S. (2002). Native-like mean structure in the unfolded ensemble of small proteins. J. Mol. Biol. 323, 153-164.

67. Shortle, D. & Ackerman, M. S. (2001). Persistence of native-like topology in a denatured protein in 8 M urea. Science, 293, 487-489.

68. Elcock, A. H. (1999). Realistic modeling of the denatured states of proteins allows accurate calculations of the pH dependence of protein stability. J. Mol. Biol. 294, 1051-1062.

Table 5.1: **Mutations in the identified kinetically important residues also found in patients with FALS** In the first column, 6 of the 16 contacts that we identify are shown. The second column lists mutations that disrupt these contacts and are found in FALS patients.

Contact	Mutation(s)
Lys3-Glu21	Glu21Lys, Glu21Gly
Glu49-Arg115	Glu49Lys
Asp96-Ser34	Asp96Asn
Lys30-Glu100	Glu100Lys, Glu100Gly
Arg79-Asp101	Asp101Gly, Asp101Asn
Glu49-Arg115	Arg115Gly

Table 5.2 Comparison of kinetically important residues Kinetically important residues in SOD1 were determined by (i) structural comparison with the protein 1TEN [50] using DALI [51] and (ii) by our simulations. Nucleus represents high Φ-value residues in 1TEN, equivalent positions refer to structurally equivalent positions determined by DALI alignment and key residues in the vicinity of equivalent positions identified by our simulations are shown in column 3. N/A refers to no equivalent residue found.

Nucleus (1TEN)	Equivalent Position (1SPD)	Key Residue
Ile20	Phe31	Lys30
Glu49	Tyr36	His48
Ile48	Asp83	Asp83
Leu50	Asn86	N/A
Ile59	Asp101	Asp101
Thr66	His110	N/A
Tyr68	Ile112	N/A
Val70	Gly114	Arg115

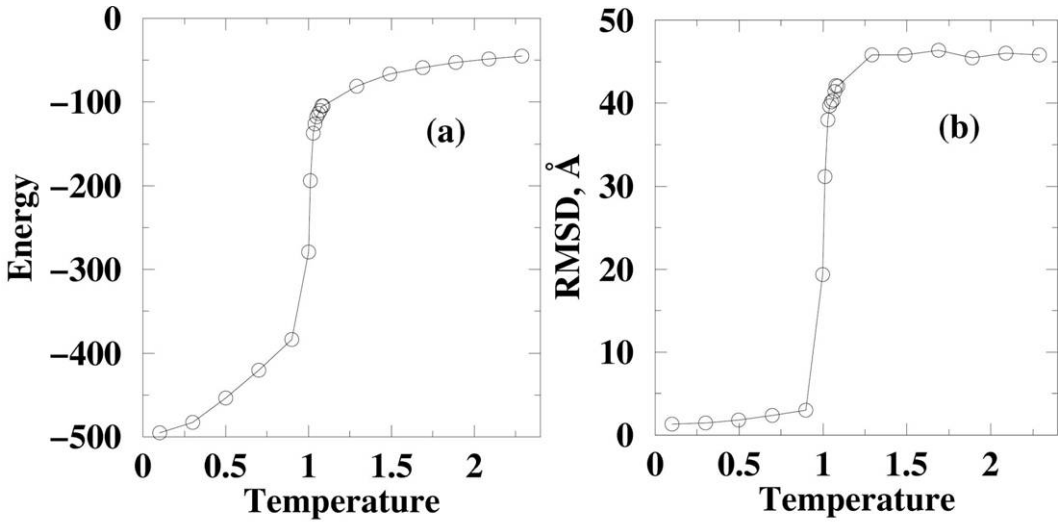


Fig. 5.1 SOD1 thermodynamic properties Dependence on temperature of (a) average energy (b) RMSD in DMD simulations with unscaled Gō potential. The sigmoidal curves in both (a) and (b) show that SOD1 undergoes a collapse transition. Similar curves are obtained for DMD simulations with the scaled Gō potential.

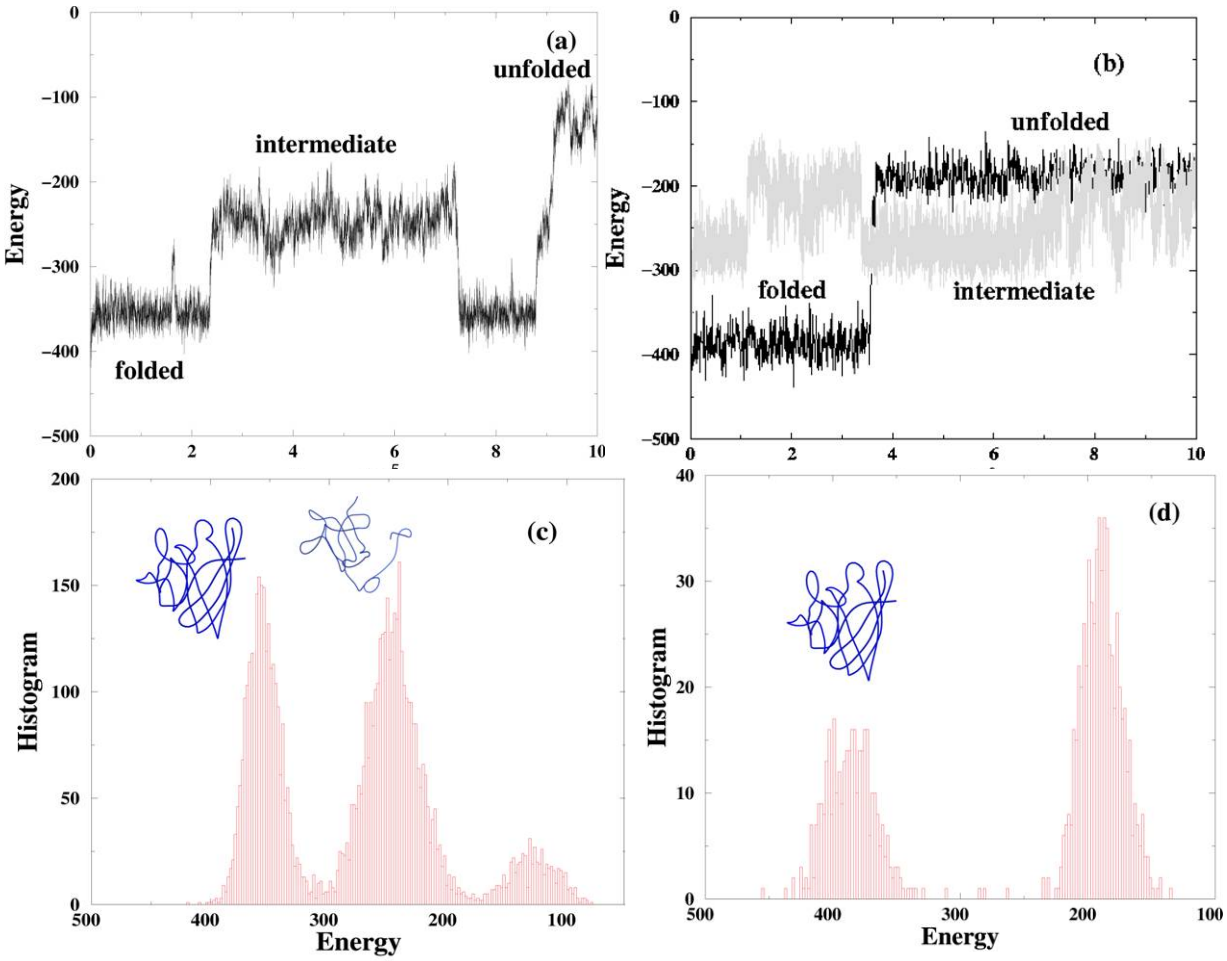


Figure 5.2 Typical DMD trajectories near T_F for (a) unscaled and (b) scaled Gō models. One trajectory is shown out of 4 for the unscaled Gō model and 5 for the scaled Gō model. In (b), a control trajectory with randomly chosen contacts (see text) strengthened is shown in grey. Histograms of energy values near T_F, for (c) unscaled and (d) scaled Gō models, averaged over 4 and 5 trajectories respectively, are also shown. From (b) and (d) we observe that the intermediate state is not populated in the scaled Gō model but is present in the unscaled Gō model, indicating that the kinetics of the scaled Gō model is indeed two-state as observed in experiments [28, 29].

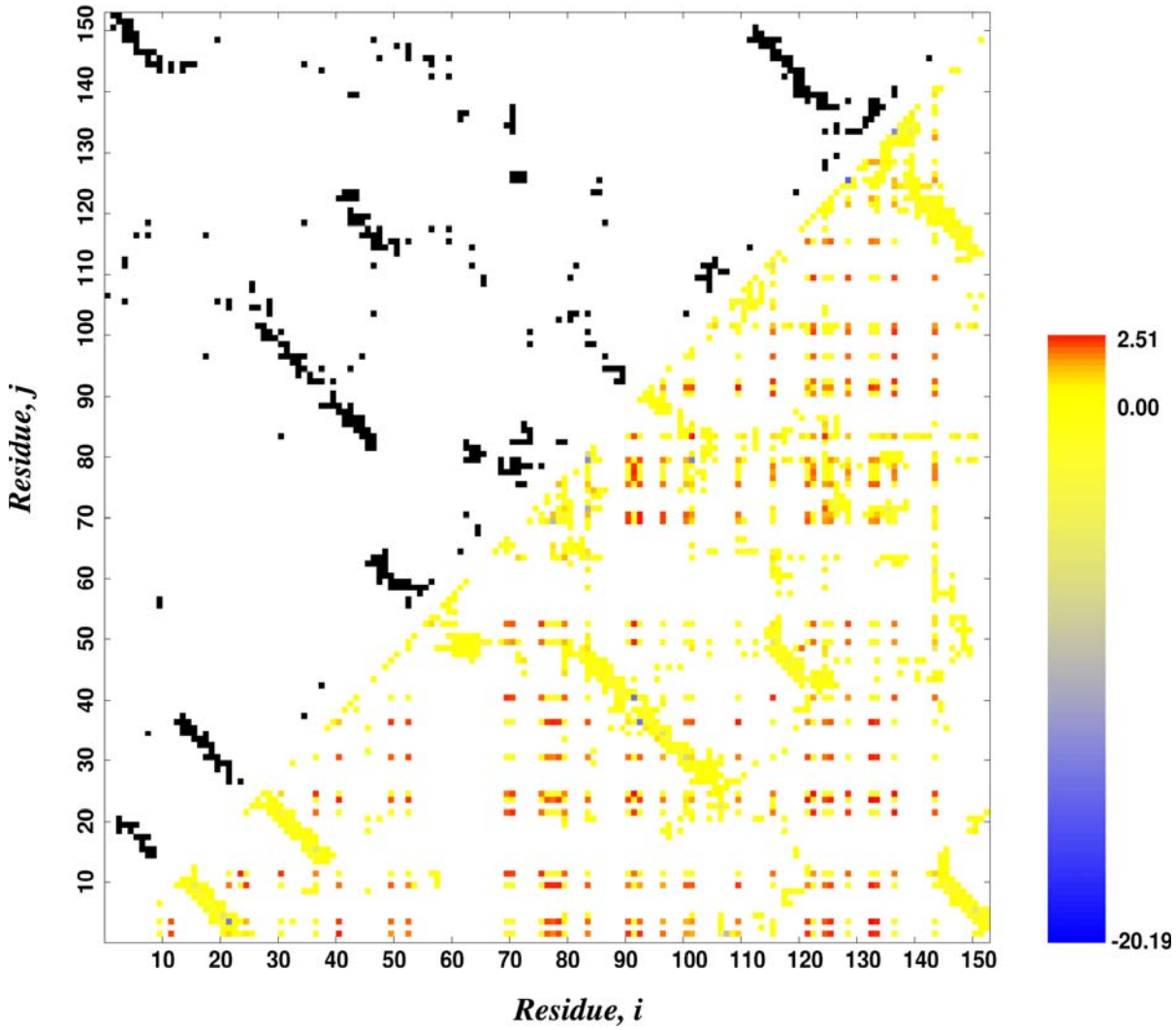


Figure 5.3 Contact map (upper triangle) and CHARMM generated energy map (lower triangle) of the SOD1 monomer. The color shading in the energy map is assigned according to the interaction energy values; red indicates repulsive interactions and blue indicates attractive interactions. Energy values are in kcal/mol.

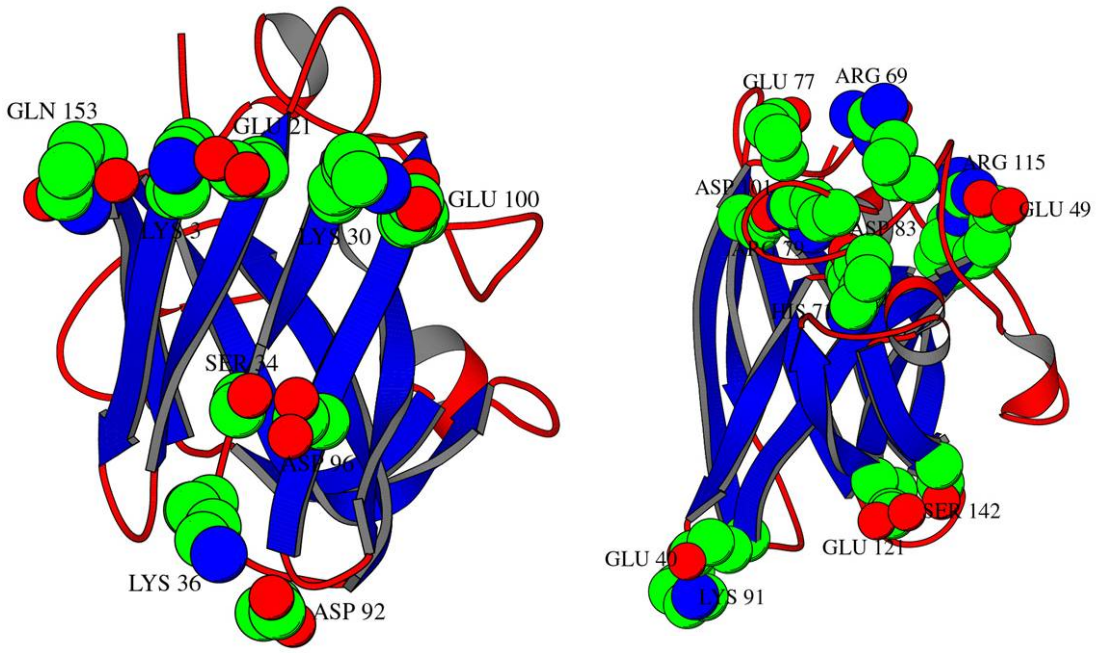


Fig. 5.4 Two projections of SOD1 monomer structure highlighting residues making the identified contacts important for the two-state folding of SOD1: Lys3-Glu21, Lys3-Glu153, Lys30-Glu100, Ser34-Asp96, Lys36-Asp92, Glu40-Lys91, Glu49-Arg115, Arg69-Glu77, His71-Asp83, Arg79-Asp83, Arg79-Asp101, His80-Asp83, Asp101-Val103, Glu121-Ser142, Asp125-Lys128, Glu133-Lys136. Figure generated using Kraulis' Molscript program and co-ordinates from PDB structure 1SPD.

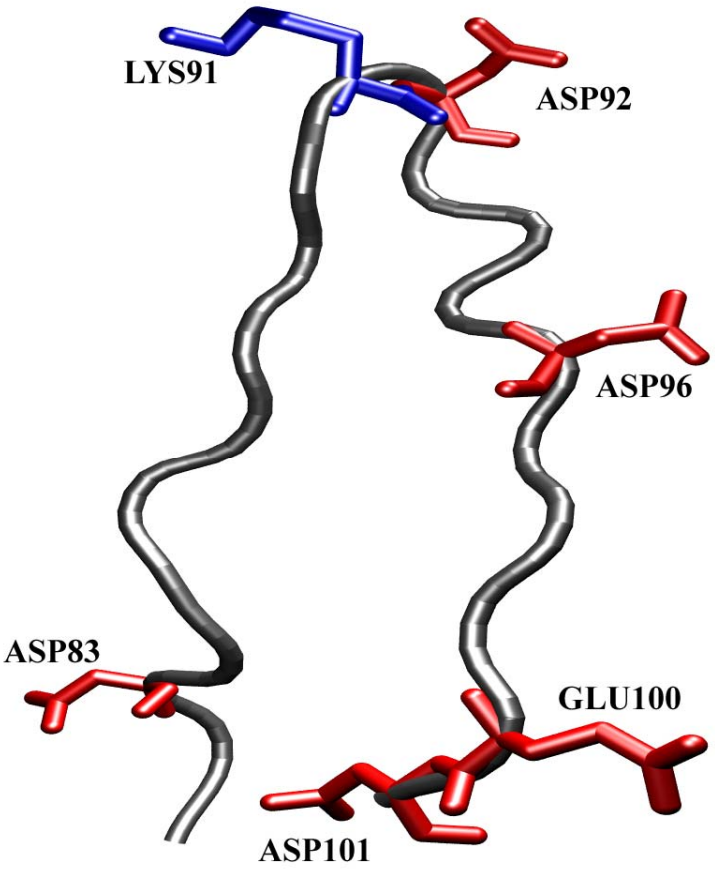


Figure 5.5 The S5-S6 cleft. The residues in the cleft which are identified as crucial for making the folding kinetics two-state in our model, are highlighted. This cleft represents one of the edges of the β-barrel and is a possible site for initiation of non-native contact formation as also found in Ref. [53], leading to aggregation.

CHAPTER 6

DYNAMICAL SIGNATURES OF FALS MUTATIONS

This chapter has been submitted as a manuscript to the journal *Proceedings of the National Academy of Sciences, USA*, and is formatted in the journal style.

6.1 Abstract

More than 100 structurally diverse point mutations in the dimeric enzyme Cu, Zn superoxide dismutase (SOD1) leading to its aggregation are implicated in Familial Amyotrophic Lateral Sclerosis (FALS). Although it is known that SOD1 dimer dissociation is required for its aggregation, the common structural basis for diverse FALS mutations resulting in aggregation is not fully understood. Here, in molecular dynamics simulations of wild type SOD1 and three structurally diverse FALS mutants (A4V, G37R and H46R), we find that a common effect of mutations on SOD1 dimer is the mutation-induced disruption of dynamic coupling between the monomers. In the wild type dimer, the principal coupled motion corresponds to a "breathing motion" of the monomers around an axis parallel to the dimer interface, and an opening-closing motion of the distal metal-binding loops. These global motions are disrupted in all three mutants independent of the structural location of the mutation. We also find that the loss of coupled motions in mutant dimers occurs with the increased disruption of a key stabilizing structural element – the β-barrel plug – leading to the de-protection of edge-strands, and may thus be implicated in early events in aggregation. To rationalize the disruption of coupling, which is independent of the effect of the mutation on SOD1 global stability, we analyze the network of residue-residue contacts formed in the SOD1 dimer fold. We find that the dimer interface and metal-binding loops, both involved in coupled motions, are regions of high connectivity in the network. Thus, we postulate that a perturbation such as mutation-induced strain in the protein is transmitted readily to these areas and thereby disrupts coupled motions. Our results suggest that altered protein dynamics due to long-range communication within its structure may underlie the aggregation of mutant SOD1 in FALS.

6.2 Introduction

Point mutations in the cytoplasmic homodimeric enzyme Cu, Zn superoxide dismutase (1) have been identified as the primary cause of approximately 20% cases of the disease Familial Amyotrophic Lateral Sclerosis (FALS) (2, 3). More than 100 distinct and structurally diverse FALS mutations, which result in identical symptoms are known (4), suggesting that all mutations affect an unknown common property of the protein, leading to toxicity. There is evidence that the motor neuron-specific toxic gain-of-function of the mutants is associated with intracellular aggregation, trafficking and/or degradation of misfolded SOD1 (5, 6). The inhibition of aggregation by over-expression of chaperones leads to increased cell viability (7).

The molecular mechanism of misfolding and aggregation of SOD1, and how structurally and chemically diverse mutations lead to aggregation, is not fully-understood. We and others have previously demonstrated that the *in vitro* aggregation of SOD1 occurs via a pathway involving dissociation of the dimer and the loss of metals, followed by multimeric assembly of the apo-monomeric SOD1 (8-10). FALS mutations are likely to lead to enhanced aggregation by affecting one or more steps in the aggregation pathway leading to an increase in the population of the aggregation-prone apo-monomer. Inhibition of mutant SOD1 dimer dissociation by engineering an inter-subunit disulfide bond on the interface, or by small molecule binding drastically reduces or abolishes aggregation (8). For a subset of mutants, the decrease in the stability of the mutant apo-dimer correlates with the average survival-time of patients carrying the mutation, suggesting that stability of the dimer may be a "common denominator" underlying FALS(11, 12). However, for other structurally-distinct mutants, mutation is found to stabilize the apo-state of the protein, while still causing disease, suggesting that the decreased stability is not sufficient for explaining the enhanced aggregation propensity of mutants(13).

The lack of specificity in mutations coincident with an identical aggregation phenotype indicates the high connectivity of the scattered mutation sites with sites that maintain the fold fidelity of SOD1. It is expected that mutants affect the global dynamics of SOD1 in the native state, and lead to greater sampling of aggregation-prone states. Indeed, several mutants are known to more dynamic – they are more susceptible to the reduction of the conserved disulfide bond in SOD1(14), display enhanced binding to hydrophobic beads(15), have greater H/D exchange rates(13), and have enhanced mobility in specific regions of the protein on the picosecond-nanosecond timescale, compared to wild type SOD1(16). Given the higher flexibility of the mutants, mutant SOD1 may induce aggregation by enhancing the propensity to locally (for both stabilizing and destabilizing mutants) and globally (for destabilizing mutants) unfold, leading to the exposure of aggregation-inducing structural elements. Therefore, understanding how structural perturbation associated with different FALS mutations influences the global and local dynamics of SOD1, particularly of the aggregation-prone regions, can provide clues for delineating the common underlying basis of mutation-associated aggregation.

In many cellular processes, energy change associated with perturbations such as mutations or ligand-binding is transduced through a protein structure, linking residues within or between subunits (17, 18). This long-range communication within protein structures has been uncovered using both experimental and theoretical approaches, and these approaches have shown that the free energy change at one site is not transduced uniformly through the protein structure; instead it is directed to specific regions of the protein (19) . These regions are determined by the geometry and the connectivity of the structural network of the protein (20).

Here we address how and why diverse FALS mutations affect the local, near-native dynamics of the SOD1 dimer and monomer, and how the altered dynamics leads to the

disruption of fold fidelity in SOD1 that may result in its aggregation. We choose three structurally and thermodynamically diverse mutations: (i) A4V, a highly destabilizing mutation on the dimer interface; (ii) G37R, a destabilizing mutation in the "β-plug" region close to the hydrophobic core; and (iii) H46R, a mildly stabilizing mutation in an active-site loop. We find that independent of the effect on global protein stability and the location of mutation, mutant SOD1 displays a loss of coupling between the motions of monomer subunits in the dimer. Furthermore, we find that the hydrogen bond network stabilizing the putatively aggregation-prone edge-strands and the hydrophobic β-plug region of the SOD1 is also weakened in the mutant dimers and monomers, and upon exposure of the wild type to aggregation-promoting conditions, compared to the native wild type. We rationalize the disruption of inter-monomer coupling in mutants by analyzing the connectivity properties of the SOD1 β-barrel fold, and find that the topology of the SOD1 fold may underlie the observed (loss of) coupling. Our results suggest that dynamics of SOD1 helps explain the observed diversity of mutations in SOD1 that lead to its aggregation in FALS.

6.3 Methods

Molecular dynamics simulations

We use the AMBER molecular dynamics (MD) simulation package with Parm99 set of parameters(21) to perform 5-ns long MD simulations of monomers and dimers of the wild type SOD (PDB accession code 1SPD) at 300 K and 400K (for the wild type dimer) and of three mutants A4V (PDB accession code 1N19), G37R (PDB accession code 1AZV) and H46R (PDB accession code 1OZT). Since apo-SOD1 has been postulated to be implicated in disease, we remove metals from structures in which metals are present, and use these as the starting

structures for simulations. In crystal structures of apo- SOD1 wild type native dimers, relatively minor structural rearrangements occur compared to the holo-protein, which mainly involve rearrangement of the metal-liganding His residues in the active site without changes in the overall fold of SOD1 (22). The prominent effect of metals is on the overall thermodynamic stability of the protein, which is considerably reduced by the loss of metals, and leads to an enhanced propensity to unfold (23). Since the unfolding of SOD1 dimer occurs on the millisecond-second timescale, therefore, on the picosecond-nanosecond timescale that are accessed by our simulations, the dynamical effects observed are likely valid for both the holo- and the apo-forms of SOD1.

The simulation protocol is as follows. Starting with the atomic co-ordinates from the crystal structures, hydrogen atoms are added and the protein is solvated in an octahedral box of TIP3P waters extending 10 Å from the protein. Counterions (Na^+) are added at appropriate places in the box to ensure charge-neutrality using the LEaP module in AMBER. Periodic boundary conditions are applied throughout the simulation, and a non-bonded cutoff of 10Å is used. Long-range electrostatic interactions are treated by particle mesh Ewald method as implemented in AMBER. The system is equilibrated using the following protocol. First, the protein is fixed with positional restraints (spring constant 500 kcal mol^{-1} $Å^{-2}$), and the solvent is energy minimized using 500 steps of steepest descent followed by 500 steps of conjugate gradient minimization with constant volume. After solvent equilibration, the system is energy-minimized using 2500 steps of steepest descent without any restraints. Next, the system is heated from 0K to 300K (or 400K) in 20 ps of constant volume MD using a time step of 1.5 fs, a Langevin thermostat (collision frequency $1ps^{-1}$), weak positional constraints for the protein atoms (spring constant 500 kcal mol^{-1} $Å^{-2}$), and SHAKE restrains for bonds with hydrogen .

Then, the whole system is equilibrated without constraints with 200 ps of NPT MD simulations using constant pressure periodic boundary with an average pressure of 1 atm and a relaxation time of 2 ps. SHAKE is used for bonds with hydrogen, the temperature is maintained at 300K (or 400K) through a Berendsen-thermostat, by coupling to a bath with temperature 300 K (or 400 K) and time constant 0.1 ps. Finally, we perform a 5-ns MD simulation of the whole system using identical parameters as the final equilibration. A snapshot of the trajectory is stored every 0.2 picosecond (100 timesteps). Each simulation is performed in parallel on 8-processors on a cluster of dual-processor nodes based on AthlonTM 2.8 GHz.

To compare the fluctuation of a given residue between various simulation trajectories, after superimposition of the trajectory on the starting structure and removal of rotational and translational degrees of freedom, we calculate a *Z*-score for the fluctuation value of the residue in the trajectory: $Z(i) = (f_i - \langle f_i \rangle)/\sigma$ where the fluctuation of a residue, i, is f_i, $\langle f_i \rangle$ the average fluctuation of all residues in the protein, and σ is the standard deviation of the distribution of fluctuations. Thus, residues with a positive or negative value of $Z(i)$ fluctuate more or less than the average fluctuation of the protein, respectively.

Covariance matrices and Essential dynamics analysis

We calculate a C_α-covariance matrix that describes the correlation of the positional shifts of C_α-atoms in proteins.

$$c_{ij} = \left\langle (r_i - \langle r_i \rangle)(r_j - \langle r_j \rangle) \right\rangle \qquad (1),$$

where r_i and r_j represent the coordinates of atoms i and j in a conformation, and <…> represents the average over the trajectory. The average is calculated over all structures after they are

superimposed on a reference (minimized crystal) structure to remove overall translational and rotational motion. To verify the robustness of the calculated correlation coefficients, we perform and the co-variance analysis over two non-overlapping 2-ns long sub-trajectories corresponding to the last 4 ns of simulation time and find that the two matrices thus obtained are highly similar to the matrix obtained for the complete trajectory (data not shown).

Diagonalizing the covariance matrix yields a set of eigenvectors and eigenvalues. The eigenvectors are directions in a $3N$-dimensional space (where N is the number of C_α-atoms), and motion along a single eigenvector corresponds to concerted displacements of groups of atoms in the Cartesian "essential" sub-space. The eigenvalues are a measure of the mean square fluctuation of the system along the corresponding eigenvectors. The eigenvectors are sorted according to their eigenvalue, the first eigenvector having the largest eigenvalue.

Graph theoretic analysis

We represent the SOD1 structures as an undirected, unweighted graph $G=\{N_i, E_{ij}\}$, with nodes N_i corresponding to each residue i, and edges E_{ij} between the nodes N_i and N_j, if these are in contact. A contact between two residues exists if the distance between any pair of atoms (excluding hydrogen atoms) belonging to these residues is less than 4.2 Å. Two connected nodes are called neighbors of each other. To estimate the effect of perturbation induced by mutation at each site, we calculate for each residue: (i) the average minimal path L_i to all other residues, a path between nodes N_i and N_j is defined as a set of edges that connect these nodes through neighbors. The length of this path is the number of edges on the path, and the minimal path is defined as the path of minimal length; (ii) the betweenness B_i, defined as the number of minimal

paths passing through the given residue. High betweenness of a node indicates that a large number of minimal paths pass through this node. The betweenness of a node N_i is therefore a measure of participation of this node in connectivity of the graph. For the residue-wise distributions of each of these properties P_i, we obtain a Z-score for each residue, which is a measure of the statistical significance: $Z(P_i) = (P_i - \langle P \rangle)/\sigma$, where P is any property i.e. L or B, $\langle P \rangle$ is the average over all nodes and σ is the standard deviation of the distribution $\{P_i\}$.

6.4 Results & Discussion

We perform all-atom molecular dynamics simulations of dimers and monomers of wild type SOD1 and three mutants for 5 nanoseconds each, in a box of water (see Methods). To understand the effect of mutation on the dynamics of the protein on the picosecond-nanosecond timescale, we compare dynamical properties, calculated from the simulation trajectories, of the mutants with the wild type, and the dimers with the monomers. The root mean square deviation (RMSD) of the backbone atoms from the respective minimized crystal structure in the protein is less than 2.5 Å for all dimers (Fig. 1(a)-(d)), and less than 2 Å for all monomers (Fig. 1(e)-(h)) for all monomers. The RMSD of the dimer interface mutant, A4V, monomer has an increasing trend with time, suggesting that a structural rearrangement occurs within this monomer occurs on the nanosecond timescale, and that a significant destabilization of the protein occurs as a result of the addition of hydrophobic groups (Ala to Val) on the water-exposed dimer interface of the monomer. However, with the exception of the A4V monomer, we find that all other trajectories do not show large drifts in RMSD, and sample near-native conformations (Fig.1(a)-(h)).

Altered fluctuation patterns of mutant SOD1 dimers

To understand whether and how mutation affects the dynamics of SOD1 dimer on a picosecond-nanosecond timescale, we calculate the fluctuations of each residue in the wild type and mutant dimers. We calculate the internal mobility of any given residue, *i*, i.e. fluctuations of the residue relative to the average fluctuation of all residues in the protein (see Methods). The resulting normalized fluctuation of a residue, $Z(i)$, can be compared between trajectories by calculating the change in the internal mobility of the residue, $\Delta Z = Z(i, \text{mutant}) - Z(i, \text{wild type})$. We find that mutants have a significantly altered pattern of mobility compared to wild type (Fig. 1(i)-(k)). The changes in the patterns of mobility are qualitatively similar in the two monomers of the dimer, but not identical (only monomer A is shown in Figs 1(i)-(k)). This asymmetry in the dimer has been reported in previous shorter MD simulations on the 300 ps timescale(24), and we find that the asymmetry is preserved on the nanosecond timescale and is present in both wildtype and mutant SOD1 dimers, suggesting that the asymmetry on the picosecond-nanosecond time scale is an intrinsic feature of the dimer. In the A4V, H46R and G37R mutants, 62%, 53%, 64% of the residues in the dimer are more flexible than the wild type, respectively, indicating that as a consequence of the altered packing, residues are globally more flexible in the mutants than in the wild type dimer (Figs. 1(i)-(k)). However, there are several residues that are less flexible in the mutants, for example, the residues 70-80 in the zinc loop are more flexible in the wild type dimer than in the A4V and H46R mutants. On the other hand, the N- and C-termini and the residues 133-138 in the electrostatic loop, which form a helix, are more flexible in all mutants, compared to the wild type. In good agreement with our calculations, these helix-forming residues 133-144 have also been found to be more flexible in the G37R mutant crystal structure. The altered flexibility of mutant dimers suggests that certain modes of motion occurring in the wild type dimer are absent in the mutant dimers, and vice-versa.

Distinct loss of inter-subunit coupling in mutant SOD1 dimers

To elucidate the types of motions that are altered by mutations in SOD1 dimer, for each mutant and wild type dimer, we calculate the covariance matrix of residues over the trajectories, which measures the coupling between the motion of any pair of residues on the picosecond-nanosecond timescale accessed in the simulation (see Methods). We find that there are distinctly pronounced patterns of correlation between residues in the wild type dimer (Fig. 2(a)) at 300 K, which are significantly diminished in the three mutant dimers (Figs 2(b)-(d)). These couplings in the wild type dimer have been observed in previous shorter timescale MD studies(25). In all trajectories, secondary elements of the native structure show a high correlation with each other, indicating that the all dimers sample near-native states. However, in the wild type dimer, there are a large number of long-range positively correlated (square III in Fig. 2(a)) and, more prominently, anti-correlated motions (square IV in Fig. 2(a)), including both intra- and inter-monomer motions exhibited by pairs of residues. While a majority of the positively correlated motions are between residues that are sequence neighbors or are in contact in the native state, a majority of the anti-correlated motions are between residues in the active site loops (60-80 and 120-140), and other regions of the protein, suggesting an opening-closing motion of the loops, which allows substrate diffusion to the active site of the enzyme. Consistent with our observations, previous studies of enzyme molecular dynamics by Bruice and co-workers (26) have also identified anti-correlated motions of the active sites to be signatures of enzymatic activity, since they represent an opening-closing motion of the active-site channel. To evaluate the timescale of the emergence of inter-monomer correlated and anti-correlated motions, we calculate the covariance matrices over ten non-overlapping sequence windows of 400 ps in each

trajectory, and find that on this timescale, only correlated motions between secondary structure elements within monomers persist in all dimers, and the differences between the mutants and the wild type dimer vanish. Thus, correlated motions between monomers arise on the nanosecond timescale, and are, remarkably, disrupted in the mutant dimers.

To quantify the degree of decoupling induced by the mutation, we calculate the contribution of inter-monomer correlated motions to the overall correlations in the dimer. This contribution is defined as

$$C_{\text{inter}} = \sum_{\text{i,j}} c_{ij}^{\text{inter}} \Big/ \left(\sum_{\text{i,j}} c_{ij}^{\text{intra1}} + \sum_{\text{i,j}} c_{ij}^{\text{intra2}} + \sum_{\text{i,j}} c_{ij}^{\text{inter}} \right), \qquad (1)$$

where c_{ij} is the absolute values of the correlation coefficient between residues *i* and *j*, the superscripts *intra1* and *intra2* represent correlations within the two monomers and *inter* represents correlations between the two monomers. The values of C_{inter}, the fraction of the correlation that is inter-monomer, are 0.26, 0.07, 0.12 and 0.14 for the wild type, A4V, H46R and G37R mutant dimers, respectively, showing that the coupling between the monomers is considerably reduced in all three mutants compared to the wild type. The values C_{intra1} and C_{intra2}, defined similarly as C_{inter}, also change upon mutation (Table I). We find that the wild type has greater intra-monomer correlations than the mutants, indicating that strain introduced by the mutation disrupts not only inter-monomer motions, but also intra-monomer concerted motions of secondary structural elements.

To further elucidate the dominant motions in the wild type that lead to the observed coupling and that are disrupted in the mutants, we use the essential dynamics method, which is based on projecting the correlated fluctuations in the protein to a principal-components

"essential" sub-space (see Methods). The starting point of the essential dynamics is the covariance matrix and results are eigenvectors and eigenvalues of the covariance matrix, which represent the directions and amplitudes of the principal motions in the principal components sub-space (Fig. 2(a)-(d)). We find that for the wild type, A4V, H46R and G37R mutants, the first 5 to 10 modes, arranged in decreasing order of their eigenvalues, capture a majority (≈75%) of the protein motions. The remaining motions of the protein are thermal noise. The trends in eigenvalues obtained for the wildtype are in good qualitative agreement with a previous, significantly shorter timescale MD study(25). The first eigenmode of the dimers is projected in their respective structures indicate there are pronounced differences in the first, most dominant, mode between the wild type and the mutant dimers (Fig.3). Motions observed in the lower modes also differ between the wild type and the mutant (see Supplementary material), but their contribution to the overall dynamics is smaller compared to the first eigenmode.

In the first eigenmode, the wild type dimer exhibits a "breathing" motion around an axis parallel to the dimer interface, along with an opening-closing motion of its loops. This motion, which is global and is a coupled motion between the two dimers, is absent in all the mutants. Instead, the dominant modes of motion of mutant dimers are confined to the active sites and in some cases other loops of the protein. We note that the loss of global coupling in the mutants is irrespective of the location of the mutation: the A4V mutation is on the dimer interface, the G37R is in the so-called "β-plug" region and the H46R is in the metal-binding site. Also, the lack of coupling is independent of the effect of mutation on stability, A4V and G37R are known to destabilize the protein whereas H46R is known to stabilize apo-SOD1.

We postulate that breathing motions provide a mechanism for the dissipation of energy perturbations induced by the environment. However in the mutants, due the evident absence of

this breathing global mode on the picosecond-nanosecond timescale, the energy perturbation may instead cause the local unfolding of key structural features of the protein, and may thus enhance aggregation propensity, independent of the effect of mutation on the global stability of the protein. The local melting of secondary structure elements in proteins is known to occur on the nanosecond timescale (as observed in H/D exchange NMR experiments), whereas global unfolding occurs on the millisecond-second or higher timescales.

Dynamical effects of dimer dissociation

Dimer dissociation is known to be required for the aggregation of SOD1, and SOD1 folding is known to be a three-state process involving the folding of individual dimers followed by their association. The folding timescale of SOD1 is in the millisecond-second range(12), and given their ~4 kcal/mol stability, monomers are expected to be stable on the picosecond-nanosecond timescale. To probe the dynamical effects of dimer dissociation on SOD1 we perform 5-ns long MD simulations of the wild type and 3 mutant SOD1 monomers, and compare monomer dynamics to that of the corresponding dimer on an identical timescale. We find that all monomer trajectories, except for the A4V monomer which has an increasing trend in RMSD, display lower or similar RMSD compared to the dimers, confirming that the monomers are stable on the picosecond-nanosecond timescale. Since monomers are on-pathway to aggregation, changes in the flexibility of monomer sub-structures upon dimer dissociation may provide structural clues to the propensities of specific areas of the monomer to induce aggregation. Therefore, we calculate the changes upon dimer dissociation in the flexibility of each residue relative to the rest of the protein (Fig. 1(l)-(o)). We find that, in the wild type protein, residues 9-13, 25-27, 38-62, 82-86 and 151-153 are more rigid in the dimer whereas residues 37-43, 70-80,

88-95 and 125-142 are more rigid in the monomer. These findings are in good agreement with NMR experiments of Banci et al.(27), in which they find that 47-59, 76-86, and 151-153 are more rigid in the dimer whereas residues 131-142 and 88-95 are more rigid in the monomer. The disagreement between our findings and the data of Banci et al. is in the change of flexibility of residues 76-80, which may be attributed to the cumulative error of MD simulations and NMR order-parameter estimation procedures.

The differences in the rigidities of dimer and monomers follow similar trends in the stabilizing H46R mutant as the wildtype, but follow a different pattern in the A4V and G37R mutants. In A4V, the residues 134-138, part of the active site Zn-loop, are found to be more flexible in the monomer than in the dimer, and the increasing RMSD of the A4V with time can be attributed to the greater disorder in the Zn-loop. The Zn-loop is known to be a site of non-native contacts in a crystal structure of multimeric SOD1 assembly(28). In G37R, the residues 48-62, which form a loop at the dimer interface (residues 50-53 make inter-monomer contacts) are found to be less flexible upon dimer dissociation compared to wild type and other mutants. However, in all mutants, the region 88-95, which forms a loop at the edge-strand of SOD1 and is part of the stabilizing "β-plug" group of residues is found to be nearly equally flexible in the mutant dimers compared to monomers, in contrast with the wild type protein where it is found to rigidify upon dimer dissociation.

Mutation- and dissociation-induced changes in the conformational stability of the β-plug

The group of residues 37-43, 90-92, 94 and 144 comprises the β-plug region of the SOD1 dimer(29). The residues 90-95 are part of the loop connecting strands β5 and β6 in SOD1, which is the site of a putative edge-strand of the β-barrel, where the cylindrical hydrogen bonding

patterns which maintain the barrel geometry are strained(30). Based on a survey of the protein database, Richardson and Richardson have postulated that edge-strands are likely sites for aggregation in proteins(30). Previous experimental and computational studies have also identified the residues 90-95 as possible sites for edge-strand mediated aggregation, and it was postulated that a transient opening of the β5-β6 strands may expose backbone hydrogen bond donors and acceptors to other chains leading to aggregation(27). Therefore, we evaluate the effect of mutation on the dynamics of the loop formed by residues 90-95. We find that the mutant SOD1 dimers show mildly enhanced rigidity of the residues in the 90-95 loop. Concomitant with this increase in rigidity is the loss of stabilizing hydrogen bonds within the loop and with other residues of the β-plug region 37-43 (Table II), indicating that although the mutants are more rigid, they adopt an altered, locally destabilized conformation. The rigidification effect is also observed in the wild type protein, in a more pronounced manner, upon dimer dissociation. An examination of the first dominant eigenmode of the wild type dimer and monomer indicates that in the dimer, these loop-forming residues undergo a twisting motion such that the strands they connect, β5 and β6, maintain their hydrogen bonds with each other. On the other hand, in the monomer, residues 90-95 are rigidified so that the β5 and β6 strands now move away from each other, resulting in a transient opening of the SOD1 β-barrel, as previously predicted by Banci et al.(16)

The loss of the structural fidelity of the β5-β6 edge results from both mutations and dimer dissociation. To further evaluate if a mildly destabilizing environment (which promotes aggregation *in vitro*(31)) also induces the destabilization of the β-plug residues and disrupts their characteristic twisting motion, we perform a 5-ns long MD simulation of the wild type dimer at 400K and perform essential dynamics analysis to obtain the principal components of motion. We

find that although there are large fluctuations in the protein as expected at high temperature, remarkably, the residues 90-95 are rigidified relative to the rest of the protein, and do not undergo a twisting motion observed at 300 K in the dominant eigenmode. Instead, the motion of the β5-β6 strands at 400 K is more monomer-like – involving the rigidification of the residues 90-95 and a transient opening of the β-barrel at the edge. The dihedral angles of the backbone populated by residues 91-94 during the trajectories further reinforce the rigidification of the loop under destabilizing conditions. Residues 91-93 in the dimer undergo transitions between dihedral angles only in the wild type, but not upon mutation, dimer dissociation or, remarkably, at 400K. These twisting motions allow favorable hydrogen bonding interactions within the loop and between the residues 90-95 and 37-43 in the wild type, and a rigidification of the loop disrupts the stabilizing hydrogen bond network (Table II). Instead the strain in the loop is transmitted to the β5 and β6 strands which show increased propensity to move in opposite directions. Therefore, we conclude that mutation, dimer dissociation, and exposure to mildly destabilizing conditions, conditions known to promote aggregation, promote the disruption of the structural fidelity of the β-plug, and result in an increased propensity of the SOD1 β-barrel to transiently open.

The destabilization of the β-plug was previously observed in the crystal structure of the G37R mutant(29). Similarly, in an NMR study of the G93A mutant SOD1 dimer, it was found that the edge-strands had higher mobility compared to the wild type(16). However, both the residue 37 and 93 are part of the plug region, so it was not clear if plug destabilization is a general mechanism, or is limited to these mutants. Our results indicate that mutations in residues distal from the plug (such as residues 4 and 46) may also dynamically destabilize it, even though they have differing effects on the global stability of SOD1, leading to a transient opening of the

β-barrel. We have previously identified the charged residues in the β5- β6 loop to be critical for the folding fidelity of the monomer(32), and also observed in simplified models of SOD1 aggregation that an opened SOD1 β-barrel has high propensity to swap strands with another similarly destabilized chain, particularly in the β5-βregion (Chapter 4). The enhanced destabilization of the β-plug region under three aggregation-promoting conditions reinforces a possible scenario of aggregation based on the opening of the SOD1 dimer. Locally destabilized SOD1 exposes the hydrogen bond donors and acceptors in the β5-β6 region, and destabilizes the β-plug of the protein by decreased interactions with the residues 38-43. Multiple open SOD1 monomers stack together at their opened exposed edges, and lead to the formation of flat sheet-like aggregates, which may retain the elements of the native structure (Figure 6). This mechanism is reminiscent of the "runaway domain-swapping" mechanism of aggregation proposed by Eisenberg and co-workers.

Rationalizing the loss of inter-subunit coupling in mutants

We observe that mutant SOD1 monomers do not undergo coupled motions in the dimer, in contrast with the wild type. Since the mutant and wild type structures are nearly identical (C_α-RMSD < 1.5 Å), the disruption of coupling suggests that the strain induced in the protein by mutation is transmitted specifically to regions that undergo coupled motions, without affecting the overall fold of the molecule.

In order to rationalize the observed loss of coupling between monomers in the mutants, we analyze the network of interactions formed in the structure of native SOD1 dimer and monomers (Fig. 5(a)), and characterize the distribution of minimal path lengths, L_{ij}, from residue *i* to the residue *j* (See Materials and Methods). For each residue *i*, we calculate the average

minimal path, a measure of the connectivity of the residue to the rest of the netwrok, to all other residues, $\langle L_i \rangle$, to the dimer interface $\langle L_{iI} \rangle$, and to the metal-binding site $\langle L_{iM} \rangle$, by averaging over the minimal paths to all other residues in the protein, to residues on the dimer interface and in the metal-binding site, respectively. For a monomer, the residues on the interface are 5-7, 17, 50-53, 114 and 148-153, and the residues in the metal-binding site are 46, 48, 63, 71, 80, 83, and 120 (Fig. 5(b)). For the dimer, we consider the corresponding residues from both monomers, as comprising the respective sites.

The value $\langle L_i \rangle$ is an indicator of the global connectivity of a given residue and a small value of $\langle L_i \rangle$ (corresponding to a negative value of $Z(\langle L_i \rangle)$) indicates that any perturbation, such as a mutation, at the residue *i* will be transmitted effectively to all other residues in the protein. For the dimer, we find that there are five regions of the sequence that have high connectivity to the rest of the protein (Fig. 5(c)): residues 2-9, 16-21, 43-65, 109-119 and 144-153. All the residues on the dimer interface lie in the identified regions of high connectivity. Surprisingly, of the residues chelating the metals, residues 46, 48 and 63 are highly connected, even though these are not in the protein core, and away from the dimer interface. Similar results are obtained for the betweenness B_i-profile (Fig. 5(d)). Residues with high betweenness are the hubs in the network through which a large number of paths linking various pairs of residues pass. These observations indicate that both the dimer interface and the metal-binding sites in the dimer are highly connected to the rest of the protein, and mutations in these residues would likely lead to a global destabilization of the protein. Conversely, a perturbation in the protein at any site due to mutation is likely to affect both the dimer interface and metal coordinating sites through the network of contact interactions.

To probe whether effect of a mutation on the dimer interface is correlated with its effect on the metal-coordinating sites, we calculate the average minimal path of all residues to the dimer interface and metal coordinating residues (Fig. 5(c)). We find that these minimal distances $<L_{iI}>$ and $<L_{iM}>$ are highly correlated to each other (Fig. 5(e)) in the dimer, but not in the monomer (Fig. 5(h)). As control, we choose the β-plug region of both monomers, composed of residues, 36-40, 88, 90-92, 94-96, and 144, since we find that in MD simulations the hydrogen bonding network of this region is destabilized in mutants and in monomers (Fig. 5(b)), and find that the distribution of average minimal paths to this surface does not correlate with either $<L_{iI}>$ and $<L_{iM}>$ (Figs. 5(f) and (g)). The correlations are also absent in the monomer (Figs. 5(i) and (j)). This shows that the correlation between the minimal paths to the dimer interface and the metal-binding site in the dimer is specific and significant. Thus, a perturbation such as mutation at any site in the protein is likely to disrupt the dimer interface, and the metal-sites. We postulate that this connectivity explains the disruption of the coupling between the two monomers, since the coupled elements, the interface and the metal-binding loops, are both likely perturbed by the mutation.

6.6 References

1. Mccord, J. M. & Fridovich, I. (1969) *J. Biol. Chem.* **244**, 6049-6055.

2. Rosen, D. R., Siddique, T., Patterson, D., Figlewicz, D. A., Sapp, P., Hentati, A., Donaldson, D., Goto, J., Oregan, J. P., Deng, H. X. *et al.* (1993) *Nature* **362**, 59-62.

3. Gurney, M. E., Pu, H. F., Chiu, A. Y., Dalcanto, M. C., Polchow, C. Y., Alexander, D. D., Caliendo, J., Hentati, A., Kwon, Y. W., Deng, H. X. *et al.* (1994) *Science* **264**, 1772-1775.

4. Gaudette, M., Hirano, M. & Siddique, T. (2000) *Amyotrophic Lateral Sclerosis and Other Motor Neuron Disorders* **1**, 83-89.

5. Cleveland, D. W. & Rothstein, J. D. (2001) *Nature Reviews Neuroscience* **2**, 806-819.

6. Okado-Matsumoto, A. & Fridovich, I. (2002) *Proceedings of the National Academy of Sciences of the United States of America* **99**, 9010-9014.

7. Bruening, W., Roy, J., Giasson, B., Figlewicz, D. A., Mushynski, W. E. & Durham, H. D. (1999) *Journal of Neurochemistry* **72**, 693-699.

8. Ray, S. S., Nowak, R. J., Strokovich, K., Brown, R. H., Walz, T. & Lansbury, P. T. (2004) *Biochemistry* **43**, 4899-4905.

9. Rakhit, R., Crow, J. P., Lepock, J. R., Kondejewski, L. H., Cashman, N. R. & Chakrabartty, A. (2004) *J. Biol. Chem.* **279**, 15499-15504.

10. Khare, S. D., Caplow, M. & Dokholyan, N. V. (2004) *PNAS* **101**, 15094-15099.

11. Lindberg, M. J., Tibell, L. & Oliveberg, M. (2002) *PNAS* 262527099.

12. Lindberg, M. J., Bystrom, R., Boknas, N., Andersen, P. M. & Oliveberg, M. (2005) *Proceedings of the National Academy of Sciences of the United States of America* **102**, 9754-9759.

13. Rodriguez, J. A., Shaw, B. F., Durazo, A., Sohn, S. H., Doucette, P. A., Nersissian, A. M., Faull, K. F., Eggers, D. K., Tiwari, A., Hayward, L. J. *et al.* (2005) *Proceedings of the National Academy of Sciences of the United States of America* **102**, 10516-10521.

14. Tiwari, A. & Hayward, L. J. (2003) *J. Biol. Chem.* **278**, 5984-5992.

15. Tiwari, A., Xu, Z. S. & Hayward, L. J. (2005) *J. Biol. Chem.* **280**, 29771-29779.

16. Shipp, E. L., Cantini, F., Bertini, I., Valentine, J. S. & Banci, L. (2003) *Biochemistry* **42**, 1890-1899.

17. Lockless, S. W. & Ranganathan, R. (1999) *Science* **286**, 295-299.

18. Suel, G. M., Lockless, S. W., Wall, M. A. & Ranganathan, R. (2003) *Nature Structural Biology* **10**, 59-69.

19. Fuentes, E. J., Der, C. J. & Lee, A. L. (2004) *Journal of Molecular Biology* **335**, 1105-1115.

20. Gunasekaran, K., Nussinov, R. & Ma, B. (2004) *Proteins-Structure Function and Bioinformatics* **57**, 433-443.

21. Cheatham, T. E., Cieplak, P. & Kollman, P. A. (1999) *Journal of Biomolecular Structure & Dynamics* **16**, 845-862.

22. Strange, R. W., Antonyuk, S., Hough, M. A., Doucette, P. A., Rodriguez, J. A., Hart, P. J., Hayward, L. J., Valentine, J. S. & Hasnain, S. S. (2003) *Journal of Molecular Biology* **328**, 877-891.

23. Rodriguez, J. A., Valentine, O. S., Eggers, D. K., Roe, J. A., Tiwari, A., Brown, R. H. & Hayward, L. J. (2002) *J. Biol. Chem.* **277**, 15932-15937.

24. Falconi, M., Melchionna, S. & Desideri, A. (1999) *Biophysical Chemistry* **81**, 197-205.

25. Chillemi, G., Falconi, M., Amadei, A., Zimatore, G., Desideri, A. & DiNola, A. (1997) *Biophys. J.* **73**, 1007-1018.

26. Luo, J. & Bruice, T. C. (2004) *Proceedings of the National Academy of Sciences of the United States of America* **101**, 13152-13156.

27. Banci, L., Bertini, I., Cramaro, F., Del Conte, R., Rosato, A. & Viezzoli, M. S. (2000) *Biochemistry* **39**, 9108-9118.

28. Elam, J. S., Taylor, A. B., Strange, R., Antonyuk, S., Doucette, P. A., Rodriguez, J. A., Hasnain, S. S., Hayward, L. J., Valentine, J. S., Yeates, T. O. *et al.* (2003) *Nature Structural Biology* **10**, 461-467.

29. Hart, P. J., Liu, H. B., Pellegrini, M., Nersissian, A. M., Gralla, E. B., Valentine, J. S. & Eisenberg, D. (1998) *Protein Sci* **7**, 545-555.

30. Richardson, J. S. & Richardson, D. C. (2002) *Proceedings of the National Academy of Sciences of the United States of America* **99**, 2754-2759.

31. Mcree, D. E., Redford, S. M., Getzoff, E. D., Lepock, J. R., Hallewell, R. A. & Tainer, J. A. (1990) *J. Biol. Chem.* **265**, 14234-14241.

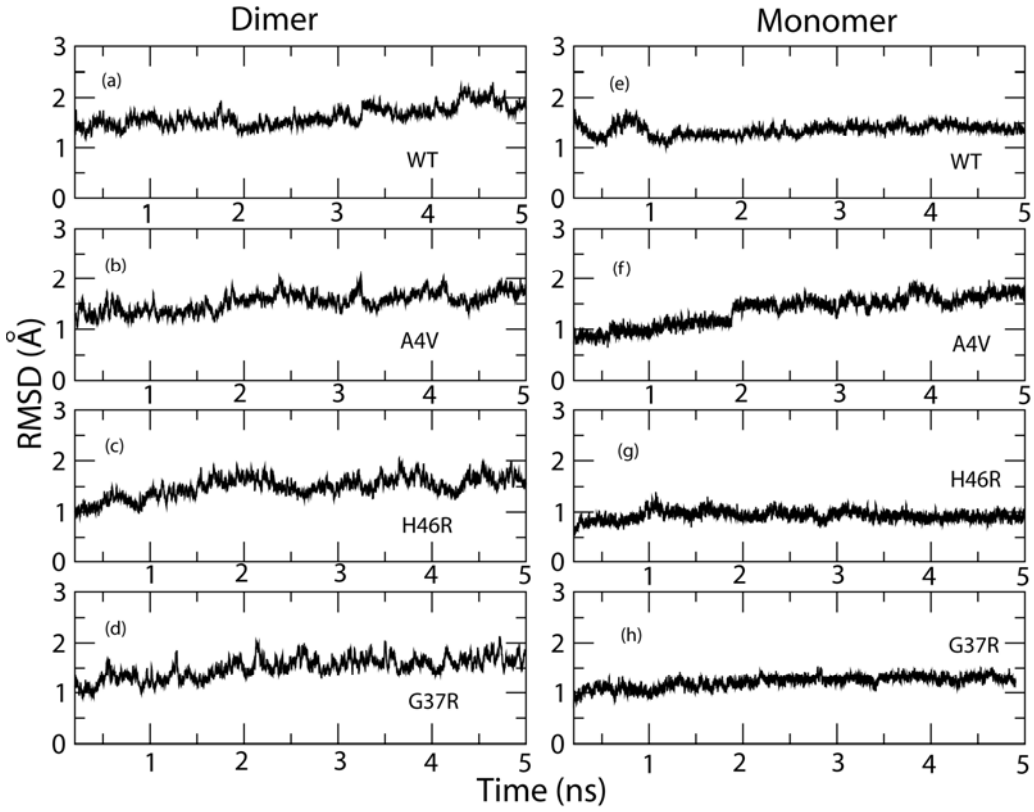
32. Khare, S. D., Ding, F. & Dokholyan, N. V. (2003) *Journal of Molecular Biology* **334**, 515-525.

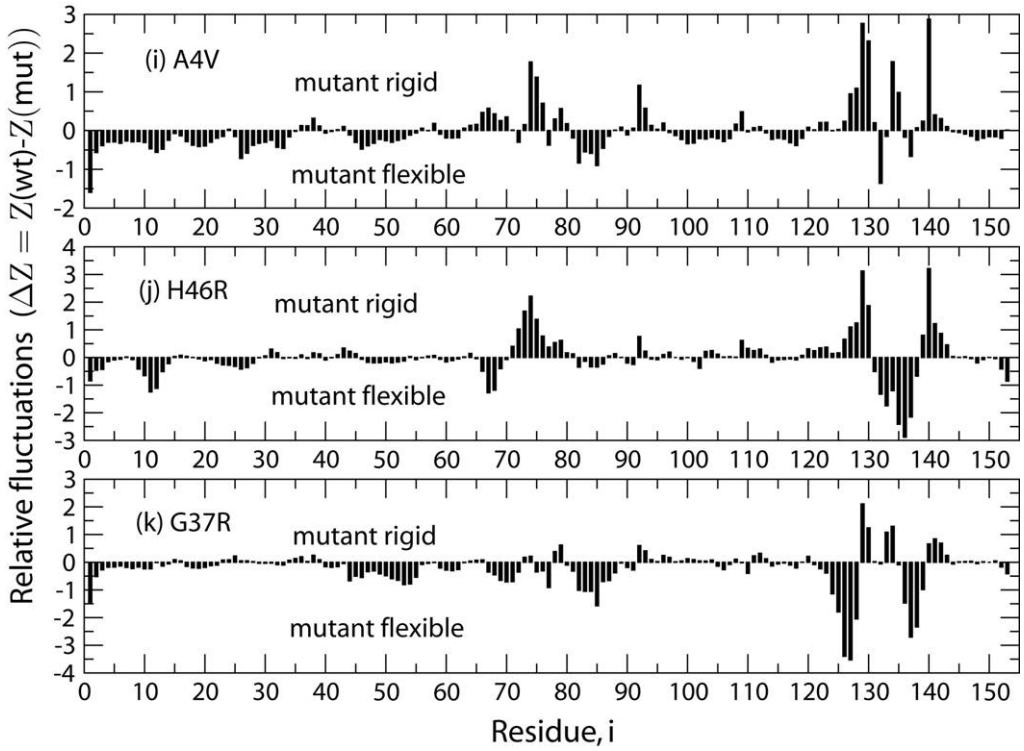
Table 6.1 Sum of correlations and anti-correlations within and between monomers

Dimer	Within Monomer1	Within Monomer2	Inter-monomer
wild type	3983.1 (34%)	4880.12 (40%)	3173.67 (26%)
A4V	1947.60 (49%)	1768.28 (44%)	290.62 (7%)
H46R	2075.61 (48%)	1734.55 (40%)	524.46 (12%)
G37R	2240.45 (46%)	1979.24 (40%)	724.84 (14%)

Table 6.2. Hydrogen bonds in the β-plug region in the MD trajectories

Donor residue	Donor atom	Receptor residue	Receptor atom	Frequency intact			
				wild type	A4V	H46R	G37R
39	O	43	HE2	0.97	0.96	0.93	0.96
90	OD1	91	H	0.60	-	-	-
90	OD2	92	H	0.31	0.79	0.76	0.81
90	O	93	H	0.89	0.93	0.95	0.85
90	O	94	H	0.90	-	-	-
93	O	37	H	0.81	0.36	0.26	-
93	O	38	H	-	0.93	0.89	0.97
94	O	90	H	0.99	-	-	-
39B	O	43B	HE2	0.97	0.97	0.97	0.96
90B	OD1	91B	H	0.28	-	0.37	-
90B	OD2	92B	H	0.79	0.51	-	0.69
90B	O	93B	H	0.99	0.97	0.96	0.94
90B	O	94B	H	0.41	-	-	-
93B	O	37B	H	0.86	0.42	-	-
93B	O	38B	H	0.44	0.90	0.96	0.95
94B	O	90B	H	0.96	0.55	-	0.31


Dimer
Monomer
(a)
WT
(b)
A4V
(c)
H46R
(d)
G37R
(e)
WT
(f)
A4V
(g)
H46R
(h)
G37R
RMSD (Å)
Time (ns)


(i) A4V
mutant rigid
mutant flexible
(j) H46R
mutant rigid
mutant flexible
(k) G37R
mutant rigid
mutant flexible
Relative fluctuations ($\Delta Z = Z(wt)-Z(mut)$)
Residue, i

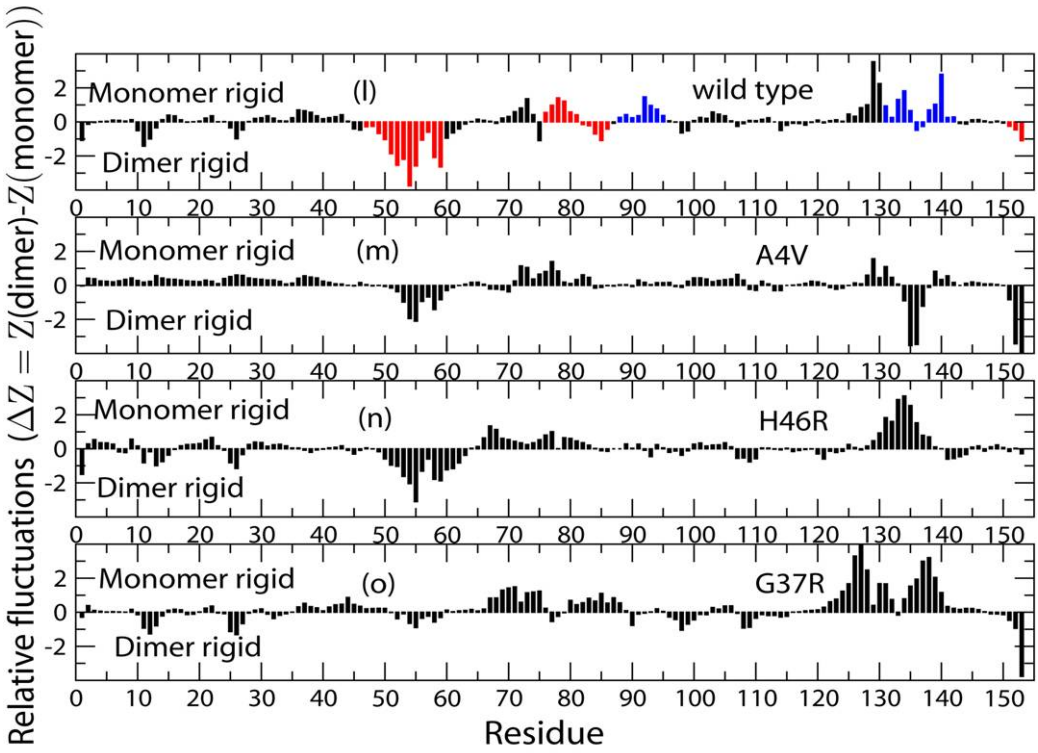


Figure 6.1 MD simulations of the SOD1 dimers and monomers (a)-(h) The root mean square deviation of the MD trajectories from the original minimized crystal structure. (i)-(k) Relative fluctuations of residues in the mutant dimers compared to fluctuation in the wild type. Only one monomer (A in the crystal structure of the wild type) is shown for comparison. (l)-(o) Relative fluctuations of residues in the wild type and mutant monomers compared to the fluctuation in the respective dimer. Residues found in NMR experiments to have enhanced flexibility in the monomer are shown in red, while residues identified as having enhanced flexibility in the monomer are shown in blue in (l)

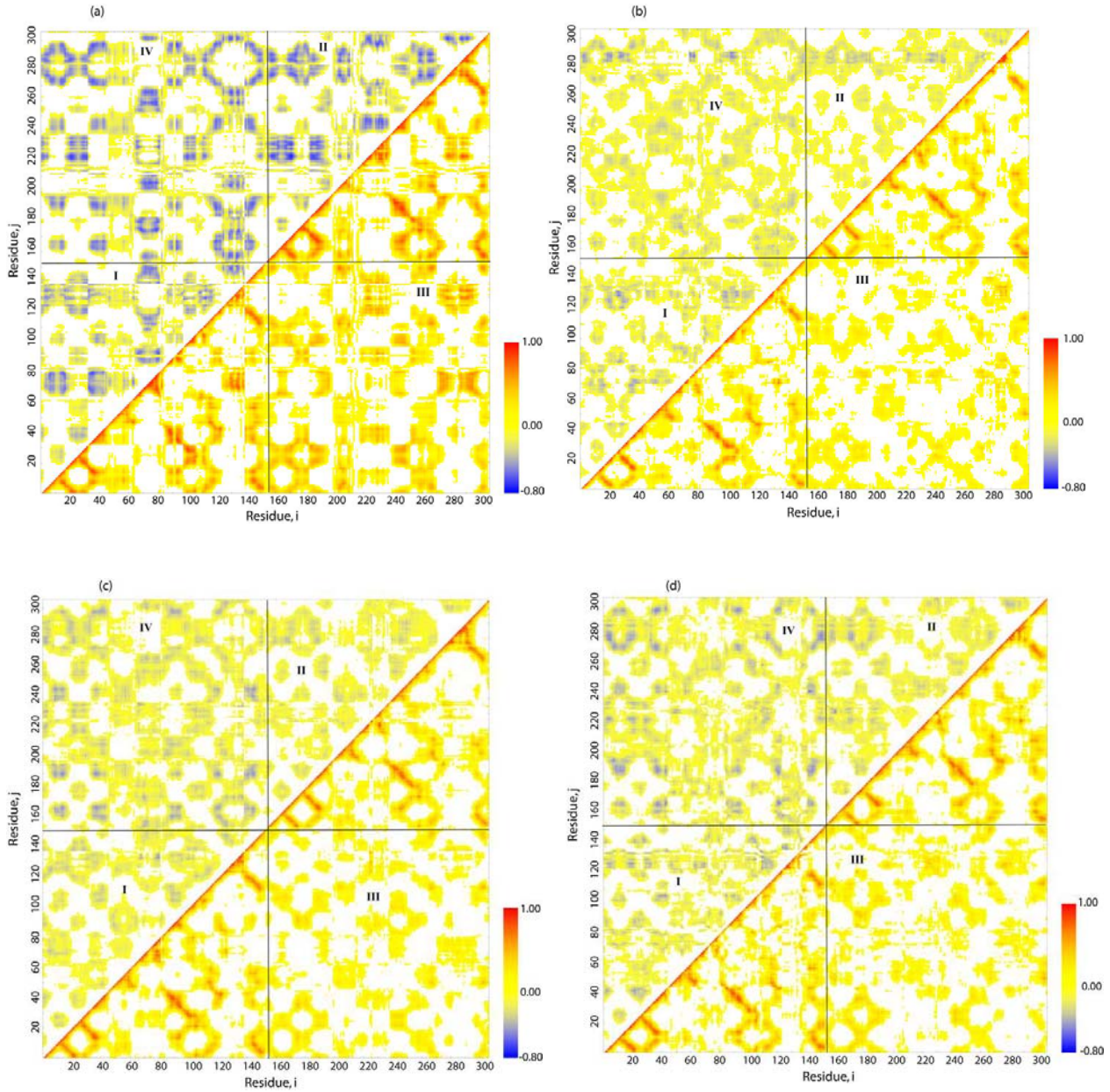


Figure 6.2 Covariance matrices of dimers (a) wild type (b) A4V (c) H46R and (d) G37R dimers. Positive correlations (red) are below the main diagonal whereas anti-correlations (blue) are above the main diagonal. Squares I and II represent the correlations within the two monomers, and III and IV represent the correlations and anti-correlations between the monomers respectively.

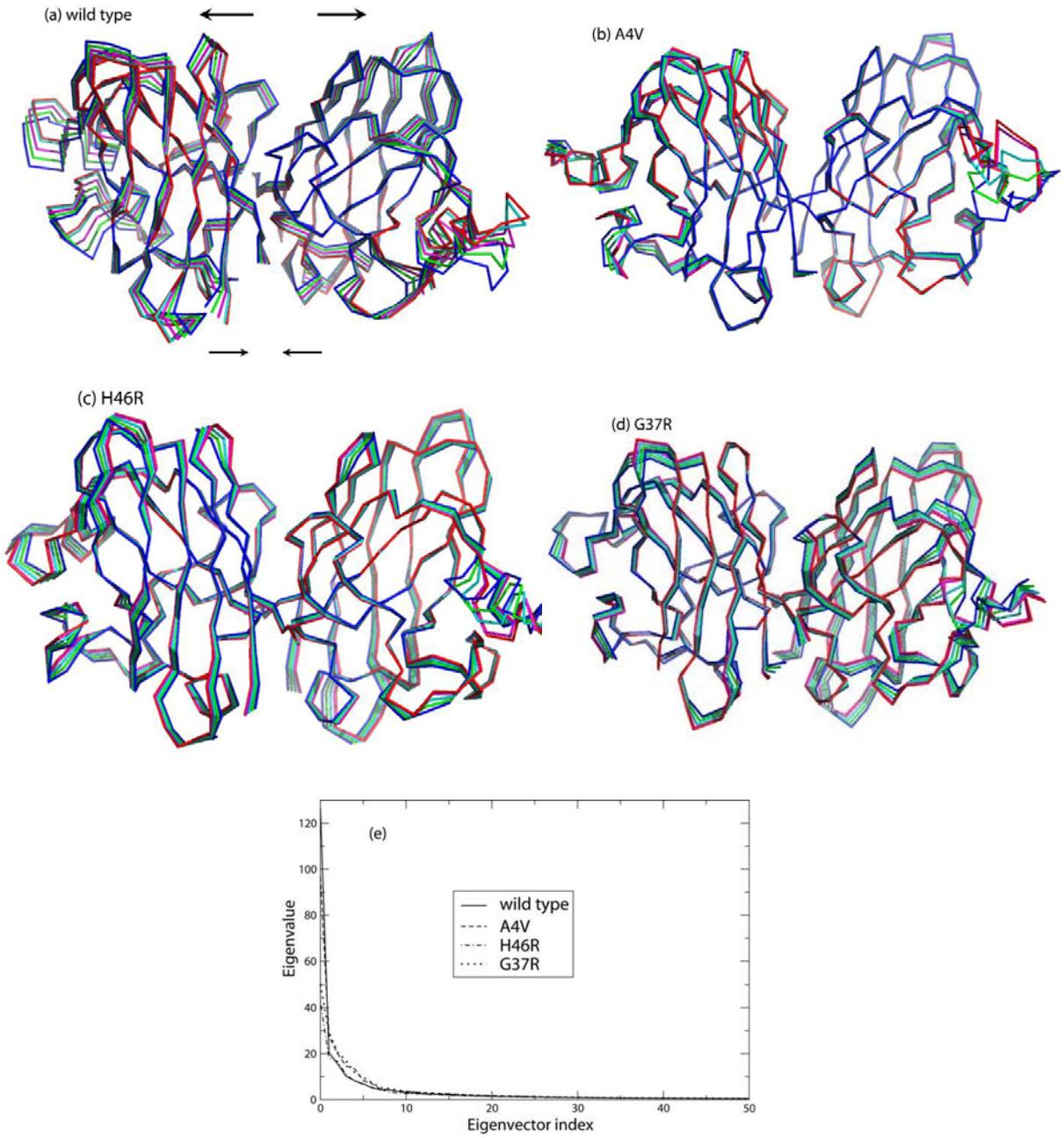


Figure 6.3 Essential dynamics analysis of dimers. The first eigenmode obtained from essential dynamics analysis projected on the structures of (a) wild type (b) A4V (c) H46R, and (d) G37R dimers. In the wild type a global breathing motion of the monomers is indicated by arrows. (e) The magnitude of eigenvalues corresponding to the first 50 eigenvectors. A majority (~75%) of the overall motion is captured by the first 5-10 eigenvectors.

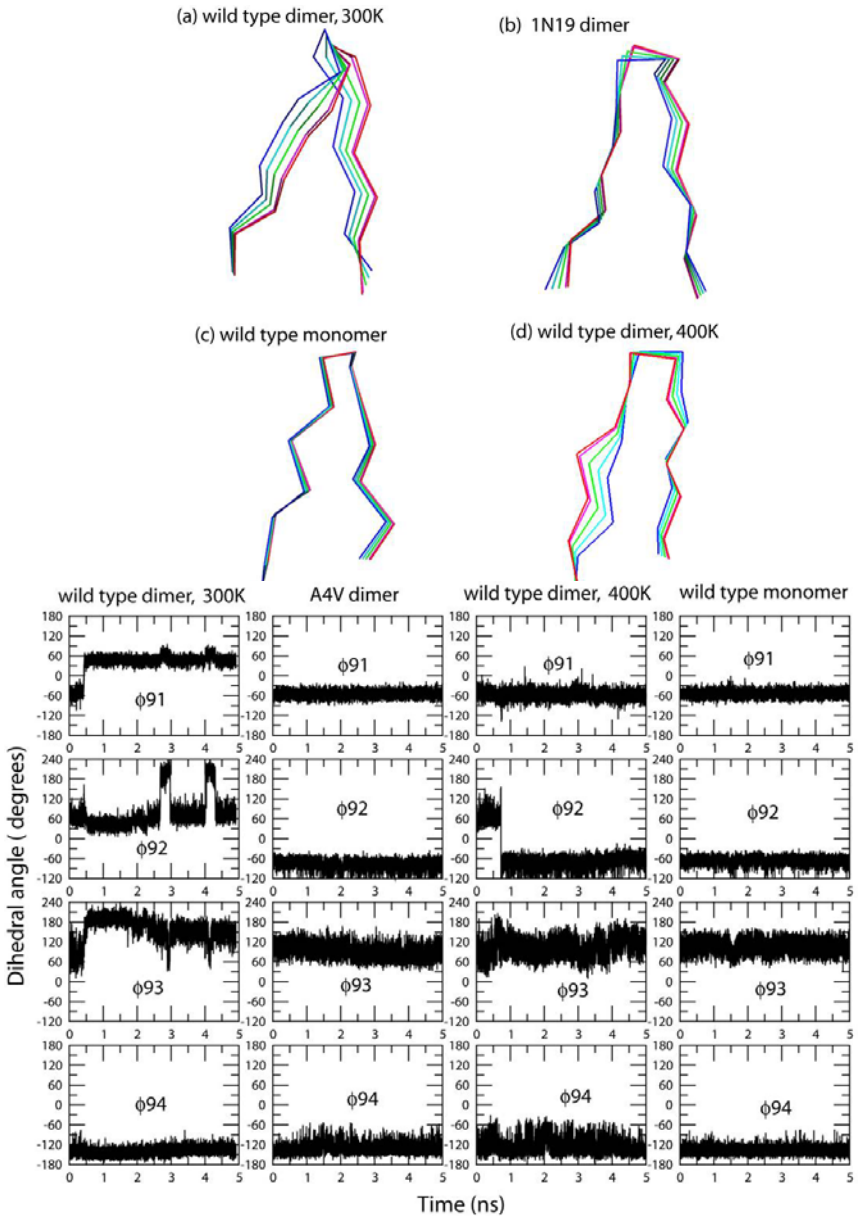
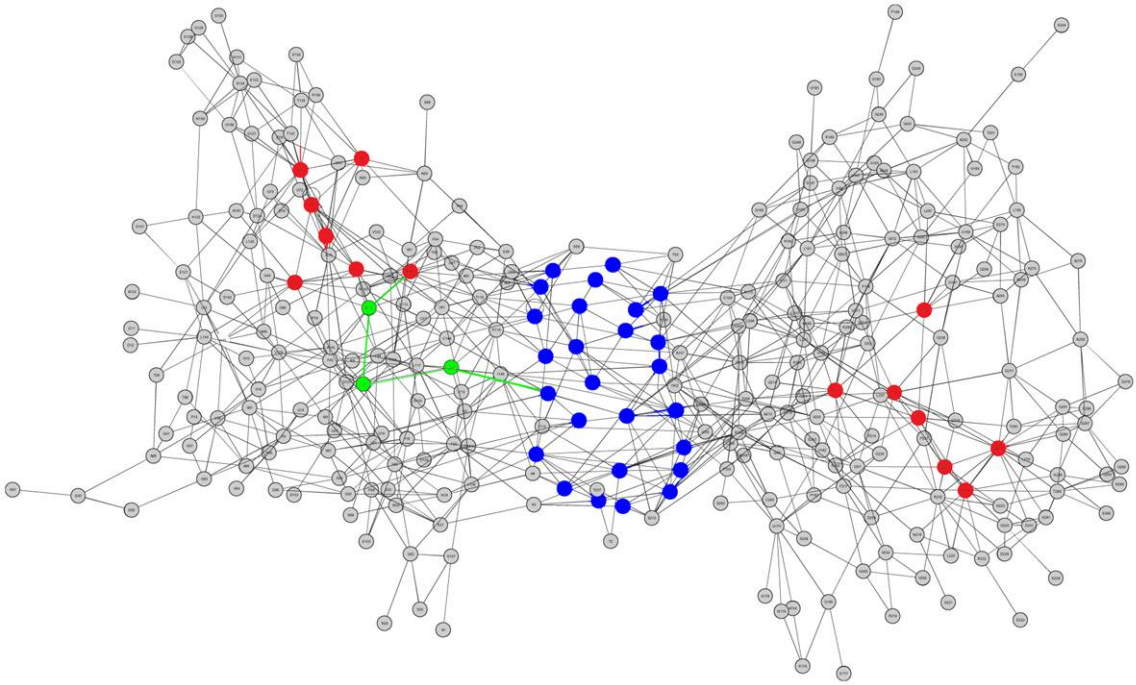
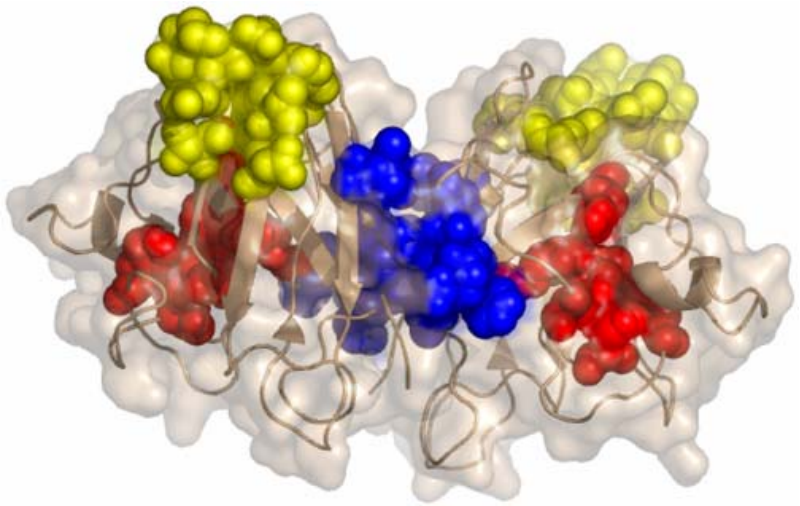


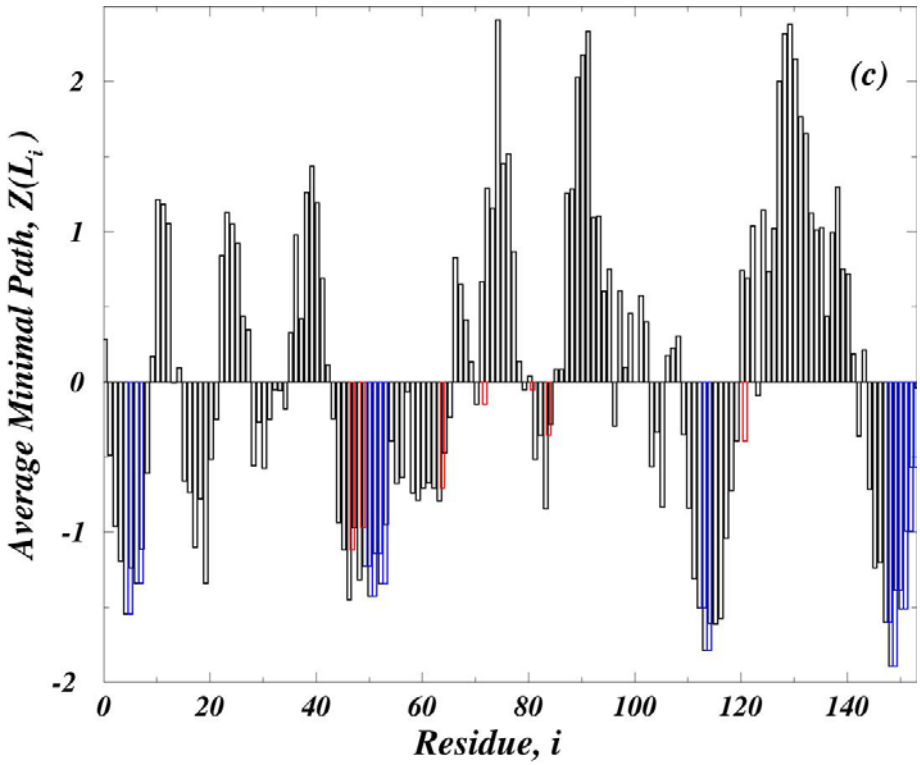
Figure 6.4 Motions of the β5-β6 loop in the first eigenmode of (a) wild type dimer at 300K (b) A4V dimer (c) wild type monomer, and (d) wild type dimer at 400K. The motion of the H46R and G37R is similar to the motions observed in (b) and (c) (data not shown). (e) The timecourse of the dihedral angle Φ for the residues 90-94 in the β5-β6 loop. Consistent with the higher flexibility of the loop observed in (a), the residues 90-93 undergo flipping motions in the wild type dimer at 300K, but are rigidified in all mutant dimers, monomer and at 400K (data for H46R and G37 dimers not shown).

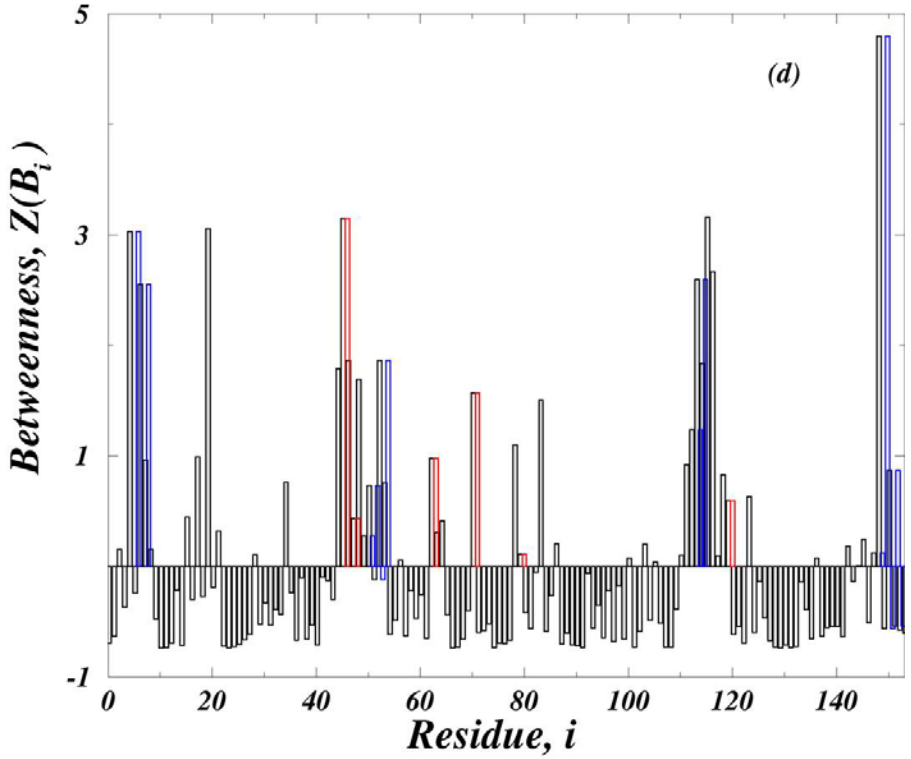
(a)



(b)




(c)
Average Minimal Path, Z(L_i)
2
1
0
-1
-2
0
20
40
60
80
100
120
140
Residue, i


(d)
Betweenness, Z(B_i)
5
3
1
-1
0
20
40
60
80
100
120
140
Residue, i

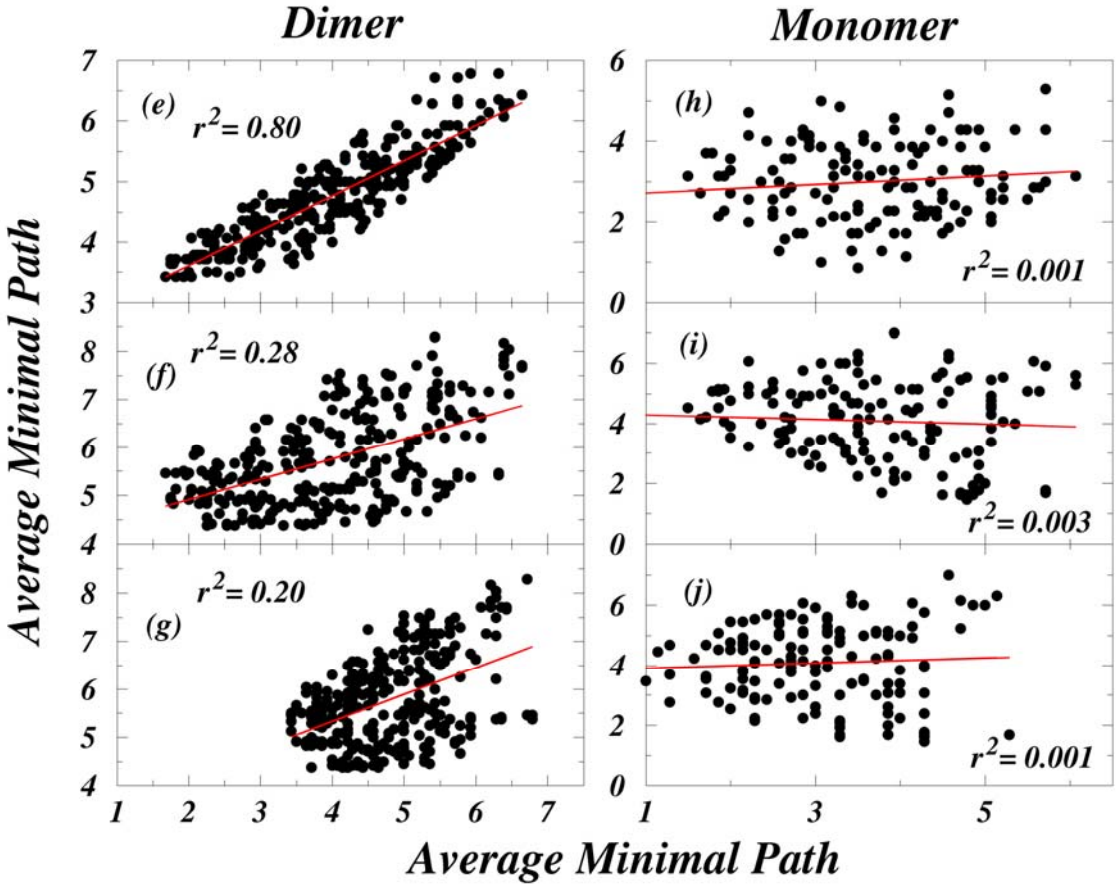


Figure 6.5 The network of interactions in SOD1. (a) A representation of the graph of the SOD1 dimer. Each node corresponds to a residue, and an edge exists between the nodes if they form a van der Waal's contact. The dimer interface residues are colored blue and the residues in the metal-binding sites are colored red. An example minimal path from L8 to the dimer interface and to the metal-binding sites is shown in green. (b) A representation of the dimer interface residues (blue), metal-coordinating residues (red) and residues constituting an unrelated surface (yellow; see (e) and (h)). (c) and (d). The distribution of normalized average minimal paths from a given residue to all other residues in the protein and betweenness of each residue. The dimer interface residues are blue, metal coordinating residues are red. Both of these sets of residues are highly connected to the rest of the protein. (e)-(g) and (h)-(j). The correlation between average minimal paths from all residues in SOD1 dimer (or monomer) to the residues on the interface and the metal-binding sites (top panels). For a dimer, the metal-coordinating residues of each monomer are considered to compose the metal-binding site. As control, minimal paths to another surface of SOD1 (b) shows no correlation with paths to the dimer interface (middle panels) or the metal-binding sites (bottom panels).

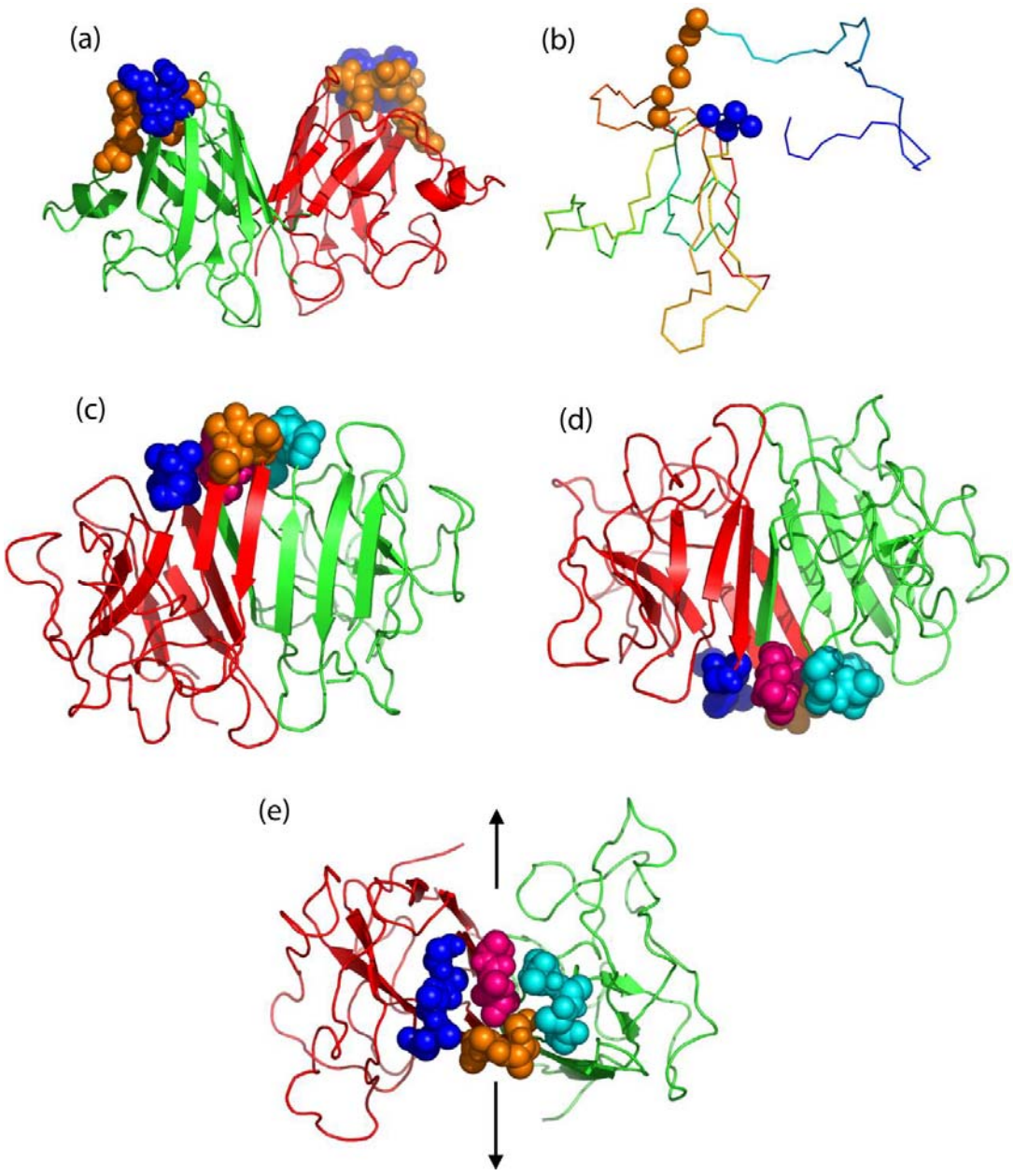


Figure 6.6 A model for the role of the β-plug region in SOD1 misfolding and aggregation. (a) the loops 37-42 (orange) and (90-93) comprising the β-plug region are highlighted on the native SOD1 dimer (b) A putative intermediate state observed previously in models of SOD1 folding, in which the β-plug is disrupted leading to partial unfolding of SOD1 monomer. (c), (d) and (e) front, back and top views, respectively, of a non-native dimer observed previously in models of non-native SOD1 dimers observed under destabilizing conditions. The β-plug regions of the two dimers interact leading to the formation of an alternative dimer interface in which there is extensive hydrogen bonding between β6 and β3 strands from different residues.

CHAPTER 7

CONCLUSIONS

7.1 Summary

The aggregation of proteins is implicated in neurodegeneration in several human diseases. Uncovering the mechanisms of aggregation may not only lead to the possibility of designing rational therapies aimed at inhibiting aggregation, but is also important for understanding the general principles that govern the formation of aggregates. This dissertation aimed at the elucidation of the mechanism of aggregation of the homodimeric metallo-enzyme Cu, Zn superoxide dismutase (SOD1) whose aggregation in motor neurons is implicated in the familial form of the neurodegenerative disease Amyotrophic Lateral Sclerosis (FALS). More than 100 point mutations in SOD1, distributed all over the sequence and three dimensional structure of the protein, are involved causally in FALS.

We showed (1) that *in vitro* SOD1 aggregation is a multi-step reaction that minimally consists of dimer dissociation, metal loss from the monomers, and oligomerization of the apo-monomers:

$$D \underset{k_{on}}{\overset{k_{off}}{\rightleftarrows}} 2M_{holo} \underset{k_m^+}{\overset{k_m^-}{\rightleftarrows}} 2M_{apo} \xrightarrow{k_{agg}} A$$

where D_{holo}, M_{holo}, M_{apo} and A are the holo-dimer, holo-monomer, apo-monomer and aggregate respectively. We measured the rate and equilibrium constants corresponding to each step under aggregation-promoting conditions (pH 3.5) and showed that under near-physiological conditions (pH 7.8), where aggregation is drastically reduced, dimer dissociation is less thermodynamically favorable and extremely slow. These results indicated that FALS-linked SOD1 aggregation occurs by a mutation-induced increase in dimer dissociation and/or metal-loss leading to an increase in apo-monomer formation. In accord with this expectation, we computationally estimated that 70 of 75 known FALS-associated point mutants in SOD1 decrease the overall dimer stability and/or increase dissociation propensity. Thus, we

concluded that mutant forms of SOD1 result in FALS because mutations lead to higher than normal concentrations of M_{apo}, thereby leading to a significantly increased probability for forming protein aggregates.

To understand the mechanism of formation of oligomers and aggregates from the apo-monomer, we investigated the sequence and structural determinants of SOD1 aggregation (2). We identified sequence fragments in SOD1 that have a high self-association propensity, and in molecular dynamics simulations of the SOD1 dimer folding and misfolding we identified regions of the protein that have a high propensity to form intermolecular interactions. These regions correspond to the two cross-over loops and two β-strands in the Greek-key native fold of SOD1. These results indicated that the high aggregation propensity of mutant SOD1 may result from a synergy of two factors: the presence of highly amyloidogenic sequence fragments ("hot-spots"), and the presence of these fragments in regions of the protein that are structurally most likely to form inter-molecular contacts under destabilizing conditions. We concluded that the balance between the self-association of aggregation-prone sequences and the specific structural context of these sequences in the native state – a balance that is altered by mutation – determines the aggregation propensity of SOD1.

In simplified models of SOD1 monomer folding, we identified residues and structural motifs crucial for maintaining the fold fidelity of SOD1(3). Remarkably, we found that these residues are located on the surface of the protein and not in the hydrophobic core. In particular, the structural fidelity of a particular region – the β-plug region – of SOD1 was found to be critical for maintaining the two-state folding of the SOD1 monomer observed in experiments. Mutations in some of the identified residues in the β-plug region are found in patients with the

disease. A possible role of the β-plug region in aggregation, which is also supported by NMR and crystallographic studies(4, 5), was further underscored in our molecular dynamics studies of wild type and mutant SOD1. We found that in the wild type dimer, the β5-β6 hairpin, a part of the β-plug region, undergoes a significant twisting motion on the picosecond-nanosecond timescale, and maintains a network of hydrogen bonds that stabilizes the rest of the β-plug region. However, upon dimer dissociation, mutation or exposure to mildly destabilizing conditions (400 K), this motion is absent and the network of hydrogen bonds stabilizing the β-plug region is weakened. Along with the altered dynamics of the β-plug region, on the nanosecond timescale, the two monomers in the wild type dimer undergo a coupled "breathing" motion around an axis parallel to the dimer interface, but this coupling is significantly diminished in the mutants. Thus, we postulated that the loss of the global coupling mode in the mutants leads to a higher propensity to locally unfold in the β-plug region. The destabilized β-plug region may lead to the formation of non-native inter-molecular contacts, as evidenced in a crystal structure of SOD1 assemblies and in our models of non-native dimers, thus implicating the β-plug region as a crucial determinant of SOD1 aggregation mechanism.

7.2 Implications and future directions

The studies outlined in this dissertation have provided insights, at the level of the primary, secondary, tertiary and quaternary structure of SOD1 into its aggregation mechanism. These insights have several implications for research in the fields of SOD1 aggregation and FALS in particular, and protein aggregation in general.

A finding in our work is that the SOD1 aggregation is a multi-step process, involving dimer dissociation as a first step. This finding immediately suggests a therapeutic

strategy aimed at the design of small molecules that stabilize the SOD1 dimer. Ray et al. have recently followed this strategy and identified small molecules from a high-throughput screen that stabilize the native SOD1 dimer(6). They have shown that drug binding diminishes or abolishes aggregation of several mutants *in vitro*, and it is likely that these molecules may be good initial leads for drug candidates against aggregation in FALS. Thus, taken together, an important implication from these studies is that a biochemical/biophysical approach taken by us, and by Ray et al., is promising for therapeutic strategy development. A more detailed understanding of the mechanism of aggregation, and of the aggregate structures, may suggest other avenues and targets for more effective intervention. For example, another finding in this dissertation is the identification, in computational models at multiple scales, of the β-plug region of the SOD1 molecule as an important site for mutation-induced disruption of structure (Chapters 4, 5, and 6). The disruption of the β-plug by mutations is possibly a property common among all mutants, and therefore, represents another process that, like dimer dissociation, may be targeted for drug discovery.

For the open problem of understanding the major driving forces and developing general frameworks for understanding protein aggregation (Chapter 1), the methods developed and used here for identification of aggregation-prone regions within a protein can easily be extended to other protein sequences. In particular, the combination of sequence-based and structure-based approach for determination of aggregation-prone regions underscores the importance of considering the structural context while interpreting the aggregation propensities of isolated peptide fragments from protein sequences. Our mechanistic studies using multiple-scale computational models also support the crucial role of edge-strands (7) in aggregation, the β-plug region identified in our models at multiple time

and length scales is the site of contact between edge-strands in the SOD1 barrel. The exposure of hydrogen bonding donors and acceptors at the edge-strands leads to putatively "domain-swapped" structures and the exposure of putatively aggregation-prone sequences in SOD1. These findings imply that several driving forces are simultaneously co-operatively involved in the misfolding and aggregation of protein chains.

By themselves and in conjunction with other developments in the field, our studies suggest several lines of enquiry that may help detail, at atomic resolution, both the molecular mechanism of SOD1 aggregation and the structure of SOD1 aggregates. First, the rate and equilibrium constants for the aggregation of mutant SOD1 under identical conditions as the wild type can be measured to evaluate the effect of mutations on individual steps of the aggregation pathway. Using the rates and equilibrium constants for various mutants, structure-activity relationships between the site of mutation and the effect on rates can be obtained.

Second, the putatively aggregation-prone regions of the SOD1 identified in our study (Chapter 4) can be synthesized as peptide fragments, and their aggregation behavior can be studied to evaluate their self-association propensity. If it is found that the identified peptide fragments aggregate in isolation, it will suggest that the aggregation of SOD1 is driven by the self-association of specific sub-sequences in the protein. If the above scenario is viable, these hot-spot peptides can be incubated with SOD1 under aggregation-promoting conditions. If it is found that the peptides have an ability to diminish or abolish aggregation, a therapeutic strategy to prevent aggregation may then involve the design of peptides or small molecules structurally analogous to these aggregation hot-spots, such that the designed drugs can competitively bind unfolded SOD1 and prevent further propagation of the aggregate.

Third, partial proteolysis or H/D exchange of aggregation reaction mixtures combined with mass spectrometry can be used to select from the leads generated for structures of non-native dimers (Chapter 4) by looking for consistency between solvent-exposed sites identified in these experiments and in the models. To refine the models, solvent exposed sites obtained by the proteolysis-mass spectrometric approach can be used as constraints to guide the computational approach used to generate putative oligomeric structures. This combination of experimental techniques and computational models, if successful, can have wide applicability in uncovering the aggregation mechanism of other proteins. Fourth, recent evidence Cleveland and co-workers(8) shows that association of misfolded and/or oligomerized SOD1 with mitochondria may be involved in the motor-neuron specific origin of toxicity in ALS. We have developed techniques (Chapter 2) that can be used to isolate aggregate species of various sizes *in vitro*. With the possibility of isolating mitochondria from cell cultures, these aggregates can be used to enquire about the molecular basis of the SOD1-mitochondrion interaction. In particular, this experimental system can be used to test the hypothesis that there is a size-range of oligomers that interacts aberrantly with the mitochondria, thus causing cytotoxicity.

Thus, it is hoped that the insights gathered from the studies in this dissertation will provide avenues to delineate at atomic resolution the mechanism of protein aggregation, and to probe the molecular basis of cytotoxicity, thereby aid the development of rational therapeutic strategies against debilitating neurodegenerative diseases.

7.3 References

1. Khare, S. D., Caplow, M. & Dokholyan, N. V. (2004) *Proceedings of the National Academy of Sciences of the United States of America* **101**, 15094-15099.

2. Khare, S. D., Wilcox, K. C., Gong, P. & Dokholyan, N. V. (2005) *Proteins-Structure Function and Bioinformatics* **61**, 617-632.

3. Khare, S. D., Ding, F. & Dokholyan, N. V. (2003) *Journal of Molecular Biology* **334**, 515-525.

4. Shipp, E. L., Cantini, F., Bertini, I., Valentine, J. S. & Banci, L. (2003) *Biochemistry* **42**, 1890-1899.

5. Hart, P. J., Liu, H. B., Pellegrini, M., Nersissian, A. M., Gralla, E. B., Valentine, J. S. & Eisenberg, D. (1998) *Protein Sci* **7**, 545-555.

6. Ray, S. S., Nowak, R. J., Brown, R. H. & Lansbury, P. T. (2005) *Proceedings of the National Academy of Sciences of the United States of America* **102**, 3639-3644.

7. Richardson, J. S. & Richardson, D. C. (2002) *Proceedings of the National Academy of Sciences of the United States of America* **99**, 2754-2759.

8. Liu, J., Lillo, C., Jonsson, P. A., Velde, C. V., Ward, C. M., Miller, T. M., Subramaniam, J. R., Rothstein, J. D., Marklund, S., Andersen, P. M. *et al.* (2004) *Neuron* **43**, 5-17.