JMB

Folding of Cu, Zn Superoxide Dismutase and Familial Amyotrophic Lateral Sclerosis

Sagar D. Khare[1], Feng Ding[2] and Nikolay V. Dokholyan[1]*

[1]*Department of Biochemistry and Biophysics, University of North Carolina at Chapel Hill School of Medicine, Chapel Hill NC 27599, USA*

[2]*Center for Polymer Studies Department of Physics, Boston University, Boston, MA 02215 USA*

*Corresponding author

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.

© 2003 Elsevier Ltd. All rights reserved.

Keywords: Cu, Zn superoxide dismutase; familial amyotrophic lateral sclerosis; molecular dynamics simulations; aggregation; misfolding

Introduction

Amyotrophic lateral sclerosis (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 Cu, Zn superoxide dismutase (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 acid residues 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.[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 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

---

Abbreviations used: ALS, amyotrophic lateral sclerosis; SOD1, Cu, Zn superoxide dismutase.

E-mail address of the corresponding author: dokh@med.unc.edu

associated FALS mutants.[24–30] Lepock *et al.*[27] and Rodriguez *et al.*[30] measured the stability of mutants in differential scanning calorimetry 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 affected significantly.[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 cysteine residues, 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 cysteine residues, show partial and full reversibility, respectively, upon denaturation. However, SOD1 from *P. leiognathi* also denatures irreversibly despite having no free cysteine residues.[36] Clearly, the presence of free cysteine residues is one of the crucial factors (but not the only factor) responsible for irreversible denaturation. One plausible reason for the effect of free cysteine residues 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 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. 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 acid residues)[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.

Results

We perform DMD simulations using the unscaled Gō model, to determine the thermodynamic properties of a model of SOD1 (see Methods). The temperature dependence of average potential energy is shown in Figure 1(a). 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 (≤ 2.83 Å) root-mean-square deviation (RMSD) from the native state (Figure 1(b)), 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 RMSD 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 acid residues is approximately three times that in the native state.
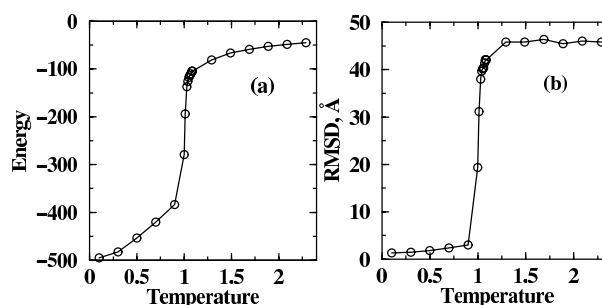

Figure 1. SOD1 thermodynamic properties. Dependence on temperature of (a) average energy, (b) RMSD in DMD simulations with unscaled Gō potential. The sigmoidal curves in both (a) and (b) show that SOD1 undergoes a collapse transition. Similar curves are obtained for DMD simulations with the scaled Gō potential.
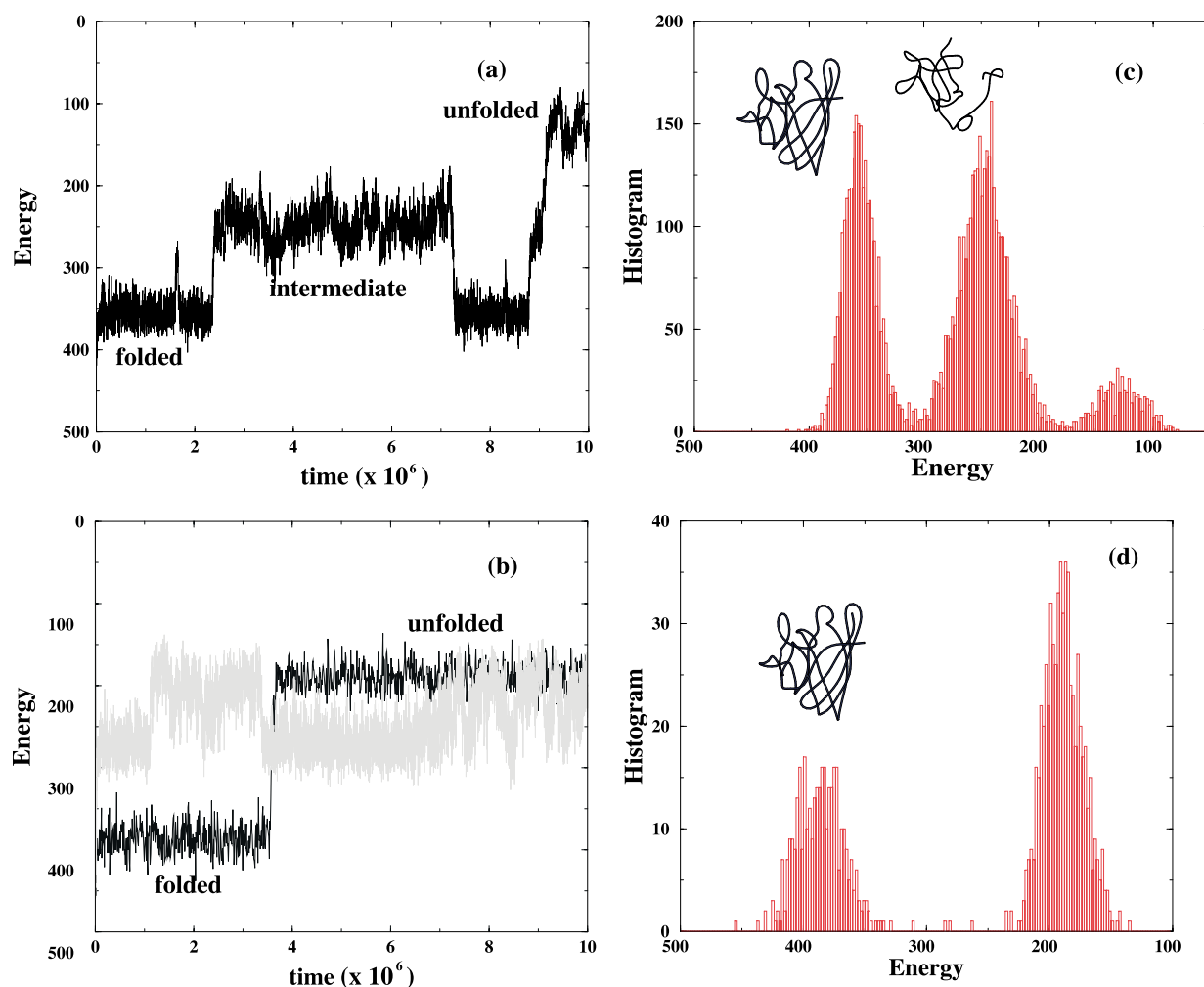
Figure 2. Typical DMD trajectories near T_F for (a) unscaled and (b) scaled Gō models. One trajectory is shown out of four for the unscaled Gō model and five for the scaled Gō model. In (b), a control trajectory with randomly chosen contacts (see the text) strengthened is shown in grey. Histograms of energy values near T_F, for (c) unscaled and (d) scaled Gō models, averaged over four and five 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]

In a typical trajectory (Figure 2(a)) at the temperature corresponding to the mid-point of the abrupt change in potential energy (Figure 1(a)) 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 (Figure 2(c)), 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 pairwise 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 pairwise contributions (see Methods) that are plotted in the lower triangle in Figure 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 Figure 4 and are involved in contacts between oppositely
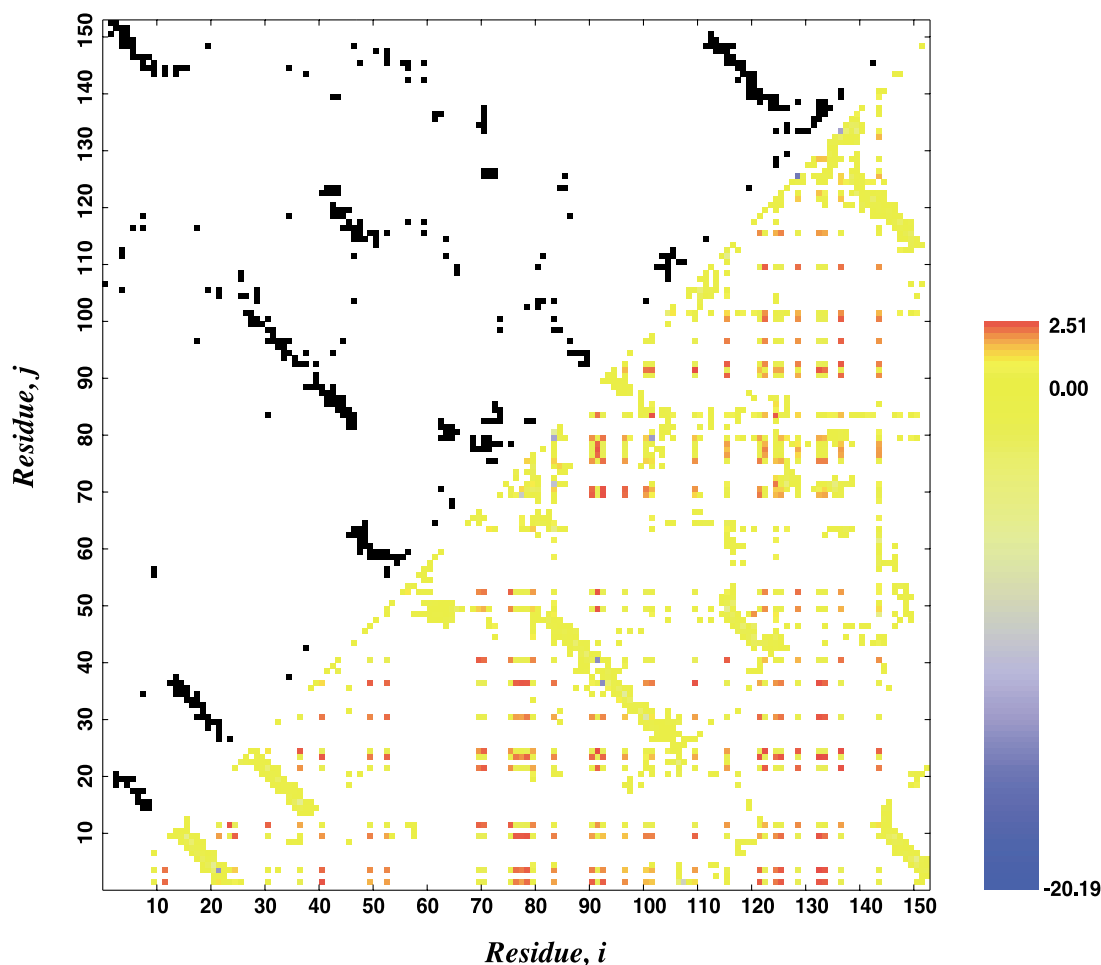
Figure 3. Contact map (upper triangle) and CHARMM generated energy map (lower triangle) of the SOD1 monomer. The color shading in the energy map is assigned according to the interaction energy values; red indicates repulsive interactions and blue indicates attractive interactions. Energy values are in kcal/mol.

charged amino acid pairs on the surface of the protein. Out of the 16 identified interactions, six have been found to be directly disrupted in patients with FALS as a result of point mutations (listed in Table 1) in the residues making these contacts†. In most cases, the mutations correspond to the change of a charged amino acid for an oppositely charged

Table 1. Mutations in the identified kinetically important residues also found in patients with FALS

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

In the first column, six of the 16 contacts that we identify are shown. The second column lists mutations that disrupt these contacts and are found in FALS patients.

† http://www.alsod.org

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 acid residues 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, RMSD and R_g (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 (Figure 2(b)) that the intermediate state is no longer populated and the protein only exists in either the folded or the unfolded state (Figure 2(b) 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 16 contacts and scale them by the same procedure. In such control simulations, we do not observe such effect on the folding kinetics (Figure 2(b)), thus demonstrating that the effect of strengthening the identified contacts is specific. 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 cysteine residues. 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 $T = 2.3$ $(\gg T_F)$, and cooling the system to low temperature $T = 0.4$ $(\ll T_F)$ slowly, while allowing the possibility of interactions between any of the four cysteine residues to form improper disulfide bond(s). We model disulfide bonds by an attractive potential $(b = 2, \gamma = -1$ in equation (2)). As a control, we repeat the annealing procedure without these non-native disulfide bonds $(b = 1, \gamma = 1$ in equation (2)). We observe that out of 20 trajectories with improper disulfide interactions, four 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.

Discussion

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

$$D \rightleftharpoons 2M \rightarrow U \qquad (1)$$

where D, M and U are SOD1 dimer, monomer and unfolded states, respectively. In order to under-

stand the effect of mutations on SOD1 folding, it is important to study the effect of mutation on both processes in equation (1). Here, we study the second process $(M \rightarrow U)$ 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
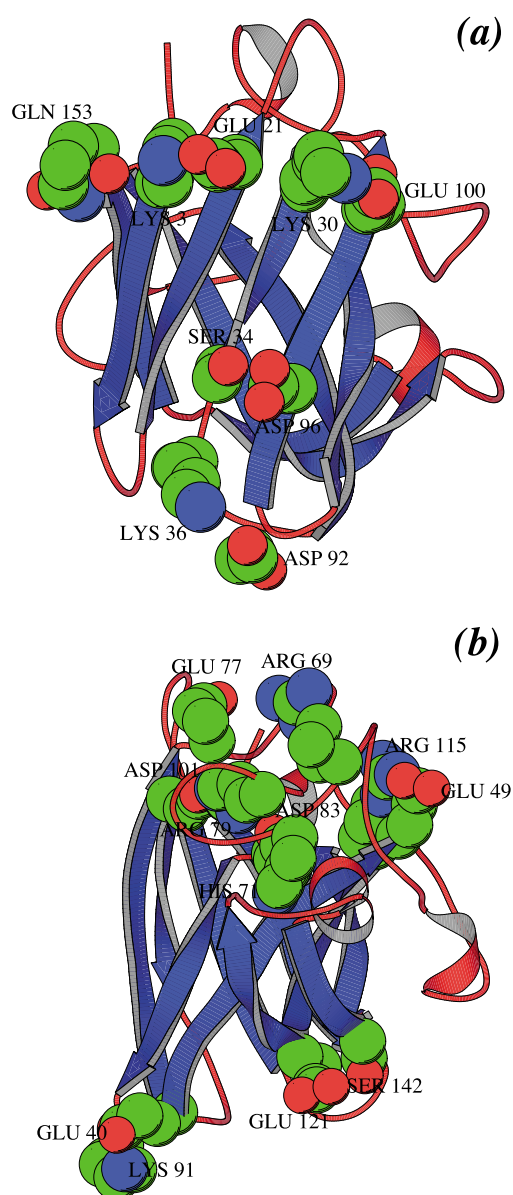

Figure 4. Two projections (a) and (b) of SOD1 monomer structure highlighting residues making the identified contacts important for the two-state folding of SOD1: Lys3-Glu21, Lys3-Glu153, Lys30-Glu100, Ser34-Asp96, Lys36-Asp92, Glu40-Lys91, Glu49-Arg115, Arg69-Glu77, His71-Asp83, Arg79-Asp83, Arg79-Asp101, His80-Asp83, Asp101-Val103, Glu121-Ser142, Asp125-Lys128, Glu133-Lys136. Figure generated using Kraulis' Molscript program and coordinates from PDB structure 1SPD.

residues are found on the surface of the protein and not in its hydrophobic core (Figure 4). By comparing the identified residues with a comprehensive list of mutations associated with FALS†, we find that, six out of the 16 contacts identified in our simulations are found to be implicated in the disease. In Table 1, 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) super-fold 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 resi-

Table 2. Comparison of kinetically important residues

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

Kinetically important 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.

dues in Tnf3 by performing a structural alignment of the two proteins using DALI[51] and list these residues in Table 2. 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 that 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 & 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 Figure 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
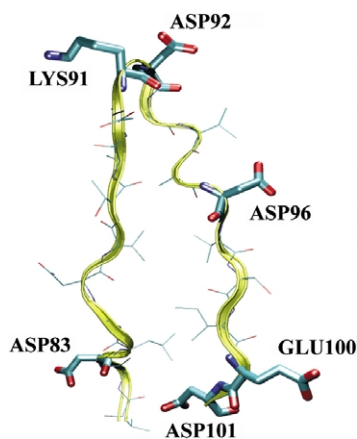
Figure 5. S5-S6 cleft. The residues in the cleft which are identified as crucial for making the folding kinetics two-state in our model, are highlighted. This cleft represents one of the edges of the β-barrel and is a possible site for initiation of non-native contact formation as also found,[53] leading to aggregation.

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 & Hecht[54] have shown that strategically designed mutations in β-strand overhangs disfavor aggregation of normally aggregating proteins, further supporting our hypothesis.

Since the residues modulating the folding kinetics of SOD1 are involved in salt bridges, a prediction made here is that the presence of salt will induce the formation of folding intermediates and/or aggregation. The residues identified by us may be the "structural gatekeeper" residues which prevent inter-polypeptide association, similar to the charged residues observed in the S6 protein by Otzen *et al.*[55] Indirect evidence for salt-mediated effects on folding kinetics comes from the observation that salt modulates the aggregation of SOD1 as shown by a recent report[56] in which apo-SOD1 aggregates were observed at low pH and high (150 mM) salt concentrations. Studies of salt induced amyloid-like protofibril formation[57] have revealed that varying salt concentrations could induce different morphologies in the final aggregates, indicating that the folding pathways may have been perturbed significantly. Thus, we expect that salt should affect the folding kinetics and aggregation propensity of apo-SOD1.

We show that the presence of oxidizable cysteine residues 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 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 apoenzyme is, therefore, not a likely mechanism of SOD1 misfolding/aggrega-

tion. 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 cysteine residues, 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.

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.

Methods

The unscaled Gō model

We use the discrete molecular dynamics (DMD) algorithm[43,58,59] to study the folding thermodynamics of SOD1. DMD has recently been used to study the folding kinetics[45,46,60,61] and aggregation of proteins.[20,62] We model the SOD1 chain by the beads-on-a-string model developed by Ding *et al.*[46] with beads corresponding to all C^α and C^β atoms and constraints between neighboring beads to mimic real protein flexibility. We study the folding of one monomer derived from a crystal structure of SOD1 (PDB accession code: 1SPD). We use the Gō potential[38] to model the interaction energy, V_{ij}, between C^β atoms (C^α for Gly) i and j of SOD1:

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

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 C^β atoms (C^α for glycine) in the native state, respectively. We assign attractive interaction for native contacts ($\gamma = -1$), defined to exist between pairs of residues whose C^β atoms (C^α 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 C^β atoms (C^α for glycine) which are farther than 7.5 Å in the native state.

The depth of the attractive square-well u_{ij} in equation (2) determines the strength of the attractive interaction between two amino acid residues. 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 the 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:[63]

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

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[64] to construct an all atom model of the SOD1 monomer in the native state, starting from its crystal structure. Recently, higher resolution crystal structures of apo-SOD1 have appeared in the literature,[65] 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:[63]

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

We compute $E_{ij}^{\text{non-bonded}}$ 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.[66] We approximate the solvent as a continuous dielectric medium with a dielectric constant ($\epsilon = 80$) of

water and the protein as a continuous medium with dielectric constant ($\epsilon = 1$) of vacuum. We compute the solvation energy by solving the Poisson–Boltzmann equation using the finite difference method[67] available in CHARMM. Following Dominy *et al.*,[68] 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:

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

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,[69–71] 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:

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

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 Figure 3.

Paci *et al.*[63] have tested the validity of Gō models and shown that around the native states of proteins, the Gō approximation corresponds closely to a molecular mechanics solvent shielded empirical energy decomposition (EEF1), similar in spirit to our decomposition procedure described above. The approach presented here therefore uses the close correspondence between the Gō and all-atom models near the native state to introduce sequence-specificity in the Gō model of SOD1 folding by scaling contact energies approximately according to their decomposed values. Other groups have also incorporated sequence information into the Gō model. Karanicolas *et al.*[72] and Clementi *et al.*[73] have shown that incorporation of sequence-specific scaling in the Gō model leads to correct discrimination between the folding mechanisms of protein G and protein L, two structurally similar proteins with distinct folding pathways. Karanicolas *et al.* use the Miyazawa–Jernigan interaction matrix as the scaling factor for contacts between residues represented by their C^α-atoms, to introduce sequence specificity, whereas Clementi *et al.* used an all-atom model to incorporate sequence effects.

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. *Advan. Neurol.* **56**, 227–231.
5. Rost, D. R., Siddique, T., Patterson, D., Figlewicz, D. A., Sapp, P., Hentati, A. 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. et al. (1994). Motor neuron degeneration in mice that express a human Cu,Zn superoxide dismutase mutation. *Science*, **264**, 1772–1775.
7. Gaudette, M., 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. USA*, **97**, 12571–12576.
9. Bruijn, L. I., Houseweart, M. K., Kato, S., Anderson, K. L., Anderson, S. D., Ohama, E. et al. (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. Expt. Neurol.* **56**, 523–530.
11. Reaume, A. G., Elliott, J. L., Hoffman, E. K., Kowall, N. W., Ferrante, R. J., Siwek, D. F. et al. (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 neuro-degenerative 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. et al. (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. USA*, **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. et al. (1991). Thermo-stabilization 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. USA*, **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 dynamic fluorescence study. *Biochemistry*, **31**, 7224–7230.
30. Rodriguez, J. A., Valentine, J. S., Eggers, D. K., Roe, J. A., Tiwari, A., Brown, R. H., Jr & 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 Sci.* **7**, 545–555.

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. USA*, **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. USA*, **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 thermodynamics 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. *Fold. Des.* **3**, 577–587.

44. Dokholyan, N. V., Li, L., Ding, F. & Shakhnovich, E. I. (2002). Topological determinants of protein folding. *Proc. Natl Acad. Sci. USA*, **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. Opin. Struct. 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 β-sheet proteins use negative design to avoid edge-to-edge aggregation. *Proc. Natl Acad. Sci. USA*, **99**, 2754–2759.

53. Elam, J. S., Taylor, A. B., Strange, R. W., Antonyuk, S., Doucette, P. A., Rodriguez, J. A. *et al.* (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. USA*, **99**, 2760–2765.

55. Otzen, D. E., Kristensen, O. & Oliveberg, M. (2000). Designed protein tetramer zipped together with a hydrophobic alzheimer homology: a structural clue to amyloid assembly. *Proc. Natl Acad. Sci. USA*, **97**, 9907–9912.

56. DiDonato, M., Criag, L., Huff, M. E., Thayer, M. M., Cardoso, R. M. F., Kassmann, C. J. *et al.* (2003). ALS mutants of human superoxide dismutase form fibrous aggregates *via* framework destabilization. *J. Mol. Biol.* **332**, 601–615.

57. Fujiwara, S., Matsumoto, F. & Yonezawa, Y. (2003). Effects of salt concentration on the association of amyloid protofilaments of hen egg-white lysozyme studied by time-resolved neutron scattering. *J. Mol. Biol.* **331**, 21–28.

58. 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.

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

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

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

62. 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.

63. Paci, E., Vendruscolo, M. & Karplus, M. (2002). Native and non-native interactions along protein folding and unfolding pathways. *Proteins: Struct. Funct. Genet.* **47**, 379–392.

64. 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.

65. Strange, R. W., Antonyuk, S., Hough, M. A., Doucette, P. A., Rodriguez, J. A., Hart, P. J. *et al.* (2003). The structure of holo and metal-deficient wild-type human Cu,Zn superoxide dismutase and its relevance to familial amyotrophic lateral sclerosis. *J. Mol. Biol.* **328**, 877–891.

66. Sharp, K. A. (1998). Electrostatic interactions in macromolecules. *Curr. Opin. Struct. Biol.* **3**, R108–R111.

67. 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.

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

69. 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.

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

71. 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.

72. Karanicolas, J. & Brooks, C. L., III (2002). The origins of asymmetry in the folding transition states of protein L and protein G. *Protein Sci.* **11**, 2351–2361.

73. Clementi, C., Garcia, A. & Onuchic, J. N. (2003). Interplay among tertiary contacts, secondary structure formation and side-chain packing in the protein folding mechanism: an all-atom representation study. *J. Mol. Biol.* **326**, 933–954.

Edited by C. R. Matthews