Sequence and Structural Determinants of Cu, Zn Superoxide Dismutase Aggregation

Sagar D. Khare,[1] Kyle C. Wilcox,[1] Peng Gong,[2] and Nikolay V. Dokholyan[1]*

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

[2]*Department of Biomedical Engineering, School of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina*

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: (1) We identify sequence fragments in SOD1 that have a high aggregation propensity, using only the sequence of SOD1, and (2) 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 N- and C-termini, and two crossover 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 intermolecular 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. Proteins 2005;61:617–632.**

Key words: protein aggregation; amyotrophic lateral sclerosis; Cu, Zn superoxide dismutase; molecular dynamics simulation; protein misfolding; domain-swapping

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] Experimental evidence suggests that specific, contiguous sequence fragments in proteins may be 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 (Thirumalai et al.[6] and references therein). The contribution, if any, of the identified peptides to the aggregation of the whole protein is determined by their context in the intact protein, but local structural propensities may also be important. For example, mutations found to diminish (or enhance) the aggregation of isolated helical peptide fragments of human procarboxypeptidase by enhancing helix propensities (comprised of local interactions) also diminish (or enhance) the aggregation of the entire polypeptide chain.[7] Interestingly, the procarboxypeptidase mutations that diminish aggregation do not affect the overall stability of the protein under the conditions of aggregation, showing that aggregation propensity of the intact protein was modulated by controlling the local structural propensities of sequence without perturbing its overall stability. Furthermore, the insertion of fragments (that aggregate in isolation) of human β2-microglobulin and the PI-SH3 domain into their respective nonaggregating homologs, mouse β2-microglobulin, and the α-spectrin-SH3 domain causes the engineered homologs to readily aggregate. Although the effect of the insertion of aggregating fragments on the overall stability of the protein was not determined, these results suggest that identification and engineering of sequence fragments that aggregate in isolation may be a strategy to modulate the aggregation of the intact protein.[8,9]

In the aggregated state, amyloidogenic proteins have been found to be arranged as β-strands in sheets.[6,10] In aggregates formed by a fragment of the Aβ-peptide, the amino acid sequence is stacked in parallel β-strands and is on average in exact register.[11] Similarly, for a 12-mer peptide designed to form amyloid fibrils, in-register paral-

Grant sponsor: Muscular Dystrophy Association; Grant number: MDA3720. Grant sponsor: March of Dimes Birth Defect Foundation; Grant number: 5-FY03-155. Grant sponsor: UNC/IBM Junior Investigator Award.

*Correspondence to: Nikolay V. Dokholyan, Department of Biochemistry and Biophysics, School of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599. E-mail: dokh@med.unc.edu

Received 21 January 2005; Accepted 3 May 2005

Published online 8 September 2005 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/prot.20629

lel β-strand topologies were observed in crystal structures.[12] Antiparallel β-strand topologies have also been observed for peptide aggregates, and changes in amphiphilicity of the aggregating peptide were found to lead to a change from antiparallel to parallel topology for peptides derived from Aβ-peptide.[13] Therefore, both parallel and antiparallel topologies can be adopted by peptides depending on their sequence and environmental conditions. Thus, in a previous study, aggregation-prone, short peptide sequences were successfully designed by estimating the fitness of the sequence in antiparallel β-sheet conformations.[14] However, for larger, intact proteins, considerable heterogeneity may exist in the alignment of amino acids from different polypeptide chains.

It is widely believed that protein aggregation requires partial or complete unfolding of the native state.[6,15] Unfolding may result in the exposure of amyloidogenic sequence fragments, which in turn may lead 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.[1,16] 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 sequences from different polypeptide chains, 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, here we address the following questions: (1) Which, if any, sequence fragments in a protein have high amyloidogenicity in isolation, and (2) 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[17]—a β-barrel composed of two β-sheets, each composed of four β-stands and connected by two crossover loops [Fig. 1(a)]. The 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.[18–20] 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.[21–23] The first step in the SOD1 aggregation possibly occurs via the non-native dimerization of apo-monomers.

The domain-swapping mechanism[24–26] was suggested as a plausible mechanism for the formation of aberrant dimers. Three-dimensional (3D) domain swapping is an event by which a monomer exchanges part of its structure with other, identical monomers to form an oligomer where each subunit has a similar structure to the monomer. Domain swapping, initially proposed as a mechanism of functional regulation, has also been proposed to lead to misfolding and aggregation.[26–28] Although there is little direct evidence for domain swapping as a mechanism for aggregation and amyloid formation, several experimental (Rousseau et al.[27] and references therein) and computational[29–31] 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.[32] 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.[26,33,34] Furthermore, domain-swapped forms of both the human prion protein and the amyloidogenic human cystatin C[29,35] have been crystallized. Computer simulation studies[29–31] using simplified native-structure-based Gō-models[36,37] 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,[31,38] two types of topologies have been detected: "closed" domain-swapped dimers, which are observed in X-ray crystal structures,[39] and more "open" oligomers, which can be propagated to form fibrils. Structural features of the computationally obtained fibrils agree with the X-ray diffraction patterns obtained in experiments.[40] Simulations of domain swapping, therefore, capture the process of misfolding and interchain interactions, and provide an avenue to investigate the early events in the amyloidogenesis of proteins. Although little is known experimentally about the structural process of SOD1 aggregation itself, domain swapping has been suggested as a mechanism for the aggregation of β2-microglobulin, a structural homolog of SOD1.[9] Therefore, domain swapping allows a description of the initial steps in SOD1 fibril formation.

To evaluate the 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 the two termini and two β-strands, and to the two crossover loops in SOD1 [Fig. 1(a)]—indicating that the high aggregation propensity of SOD1 may be due to both amyloidogenic sequence "hot spots" and their structural context. Our results suggest that aggregation of proteins may be mediated by the presence of both self-associating sequence fragments and the presence of these fragments in specific aggregation-prone structural elements of the protein.

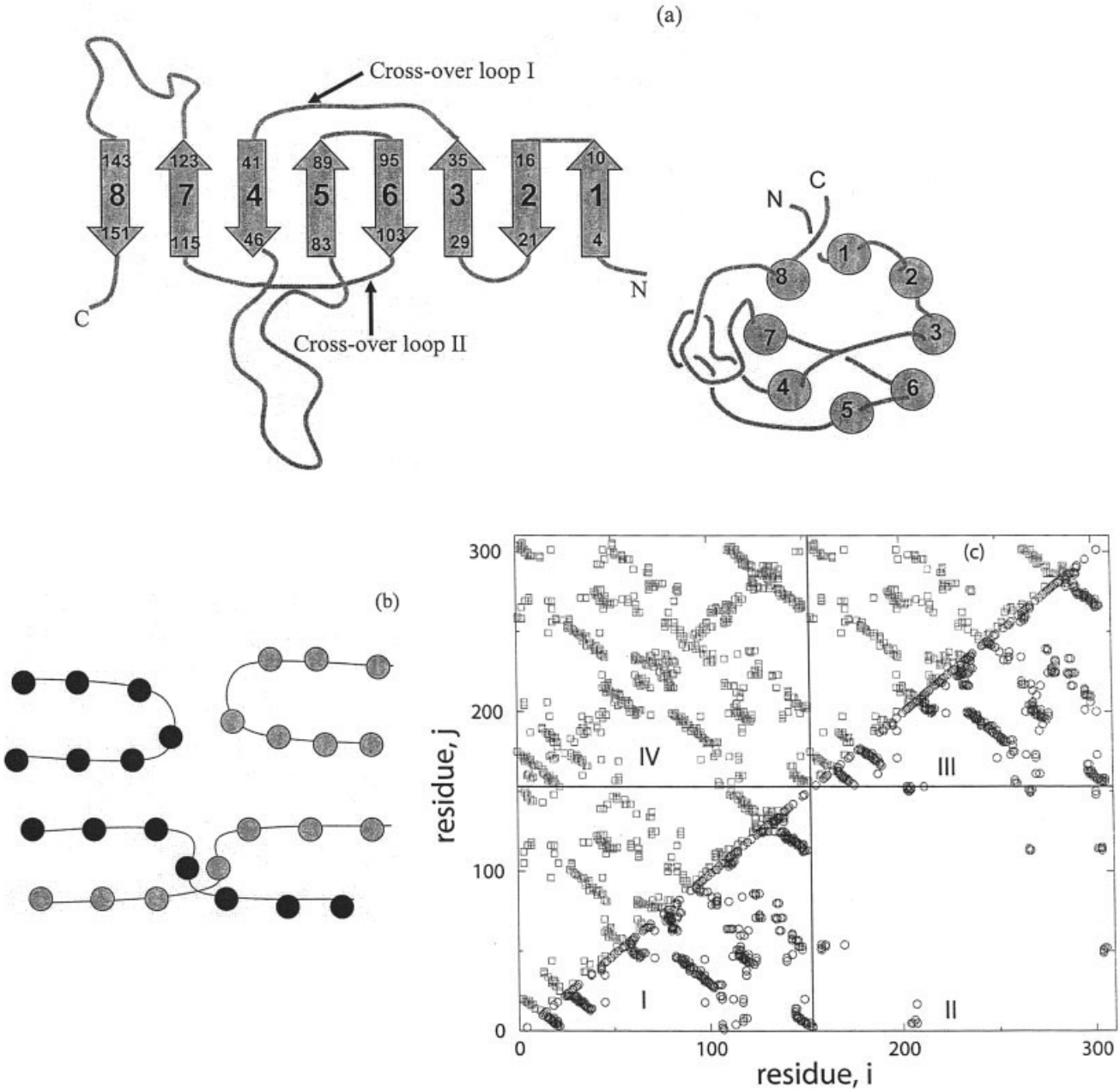


Fig. 1. SOD1 topology. (**a**) A schematic of SOD1. The regions identified as amyloidogenic are the β-strands 4 and 7 and the two crossover loops that connect the two β-sheets in the SOD1 barrel. (**b**) A schematic of domain-swapped interactions and (**c**) the domain-swapped contact map used for simulations. In the contact map, regions I and III correspond to the intramonomer 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.

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 antiparallel), and β-strand tetramer (parallel and antiparallel). For constructing the dimer and tetramer, we mount the sequence of each fragment on idealized parallel and antiparallel β-sheet structures using the package SCWRL. These template structures are constructed using the packages MOE (Schröedinger, Inc.) and have ideal parallel or antiparallel 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 an extended strand conformation generated using SYBYL.

Molecular Dynamics Simulations and Free Energy Calculations

For each fragment, we perform all-atom molecular dynamics (MD) simulations of each conformation and calcu-

late the conformational free energy. For each MD trajectory, the simulation time is 496 ps in explicit solvent [simple point charge (SPC) water model[41]] using the SigmaX2.2 MD program. Conformational free energy is calculated using the explicit solvent/implicit solvent (ES/IS) method[42] for every snapshot. Following the procedure for free energy calculations described in Urbanc et al.,[43] each MD simulation consists of relaxation of the peptide(s) and water, followed by a production run of 496 ps, where snapshots are collected at intervals of 1 ps. The conformational free energy is

$$G = \langle E \rangle - TS_{conf} + \langle \Delta G_{solv} \rangle, \quad (1)$$

where $\langle \ldots \rangle$ 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.[44]

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:

$$\Delta G_D^i = (\bar{G}_D^i - 2 \times \bar{G}_M^i) \pm \sqrt{\sigma^2(G_D^i) + 4 \times \sigma^2(G_M^i)}, \quad (2)$$

where $\bar{G}_D^i$, $\bar{G}_M^i$, $\sigma^2(G_D^i)$, and $\sigma^2(G_M^i)$ 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

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

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.,[45] to model the folding and misfolding of the SOD1 dimer. We assign pairwise, square-well interaction potentials between beads in a simplified polypeptide model according to contacts formed in the native state; that is, 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 Data Bank (PDB) 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 intermonomer 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

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

where ϵ is the strength of a contact, Δ_{ij}, $\Delta_{ik}^{\mathrm{dim}}$, 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 intermonomer overlap, or the effective concentration of the protein. When $\alpha_{\text{domain-swap}} > 1$, interactions between residues from different monomers are stronger than intramonomer interactions, resulting in a tendency of each chain to penetrate the pervaded volume of the other chain, rather than forming intrachain contacts. This scenario corresponds to a polypeptide concentration in the semidilute regimen.[46] 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 interprotein interactions may differ from the intraprotein 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 $\alpha_{\text{domain-swap}}$.

RESULTS

Sequence Determinants of SOD1 Amyloidogenesis

Since the formation of β-sheets is a necessary condition for aggregation,[1,14] to identify amyloidogenic sequence fragments in SOD1, we use idealized dimer and tetramer β-strand templates, in both parallel and antiparallel 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 antiparallel dimers and tetramers, respectively, we obtain the free energies of dimerization and tetramerization into these parallel and antiparallel 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)] for tetramerization. Although for a given residue the magnitude of the amyloidogenicity is larger in the tet-

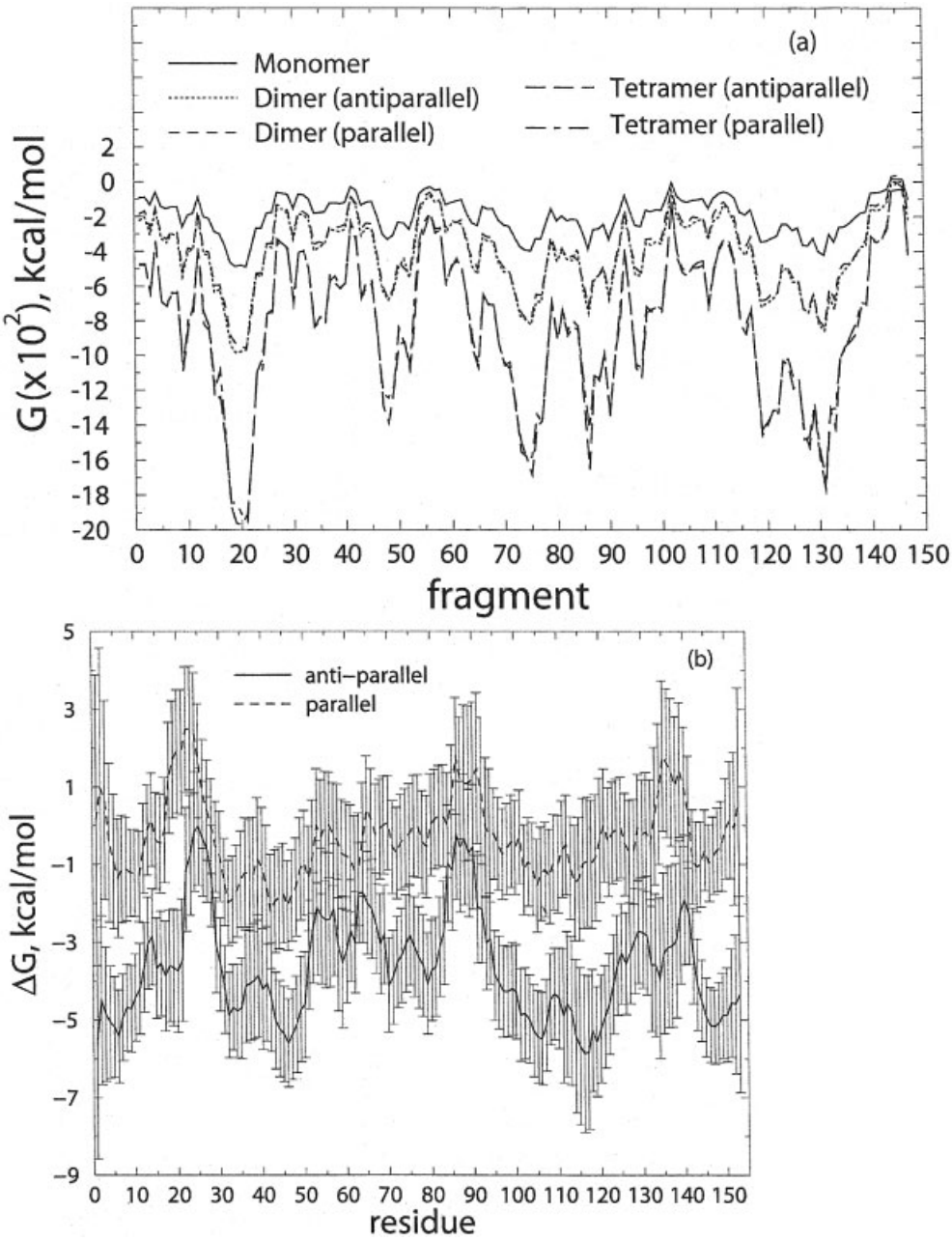
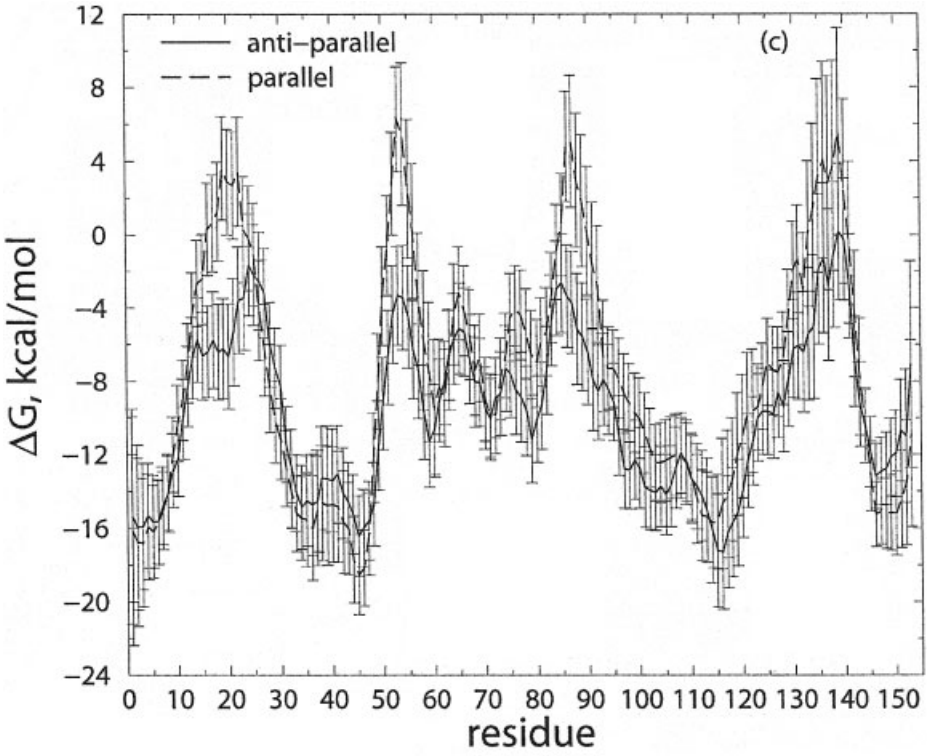


Fig. 2. Sequence propensities for aggregation. (**a**) The free energy of the 147 overlapping heptamer sequence fragments of SOD1, mounted on various template backbones. We use the average contribution of each residue to the free energy of all fragments containing the residue to evaluate the free energy of oligomerization of the residue. (**b**) The free energy of dimerization for each residue. (**c**) The free energy of tetramerization for each residue. (**d**) Sequence profile for aggregation obtained using the TANGO program. The inset shows the region from residue 28 to 36 at a higher magnification.

ramer compared to the dimer, we obtain similar trends in the free energy profiles for parallel, antiparallel, 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 crossover 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



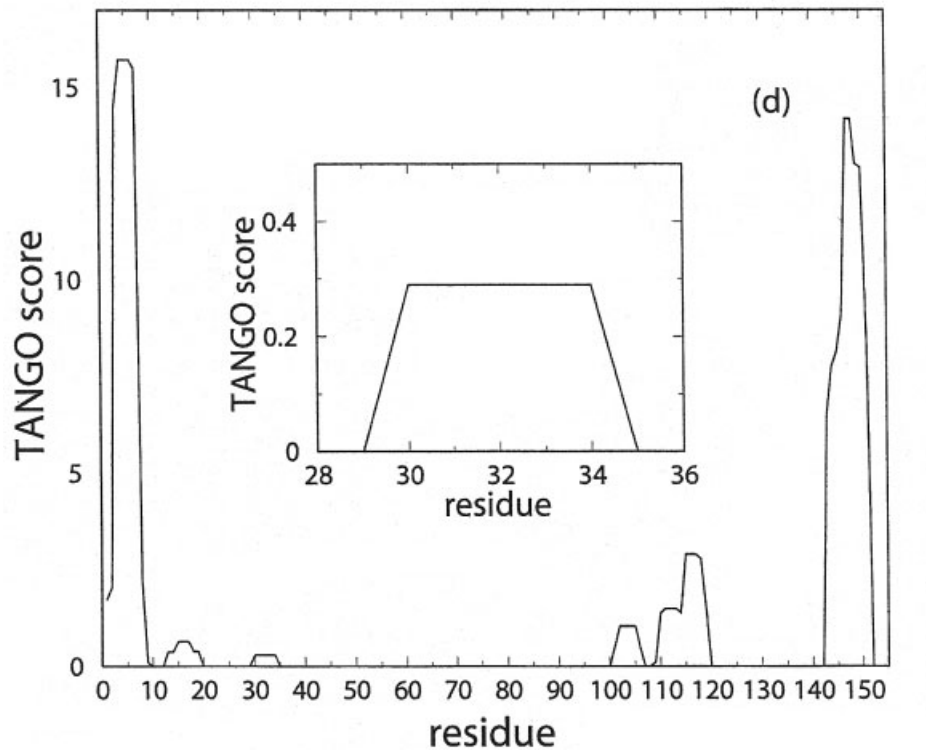


Figure 2. (Continued.)

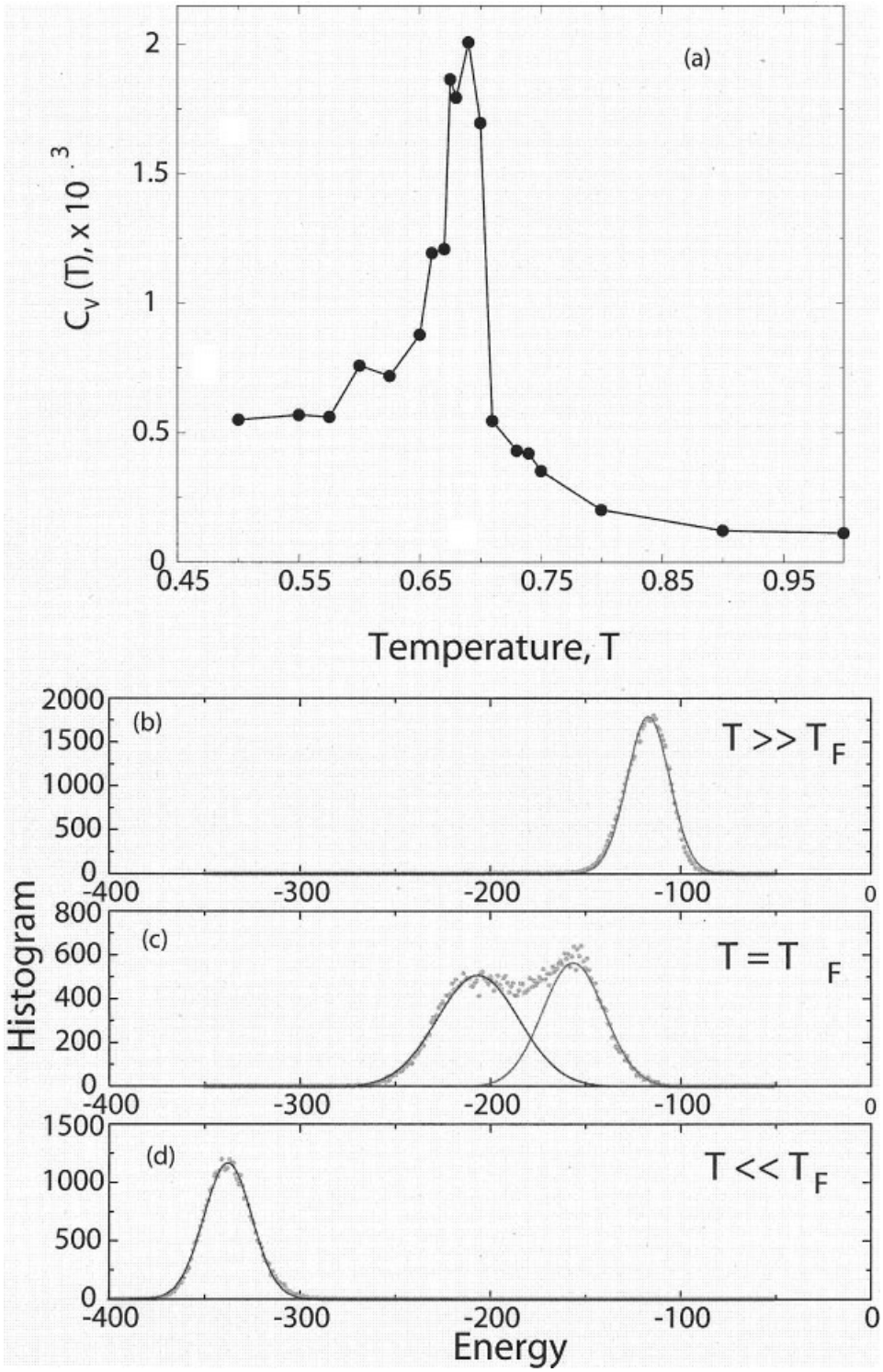


Fig. 3. Folding of SOD1 monomer and dimer. (**a–d**) Folding of the SOD1 monomer. (**a**) The heat capacity as a function of temperature shows a single peak at the folding transition temperature, T_F. The histograms of populations at (**b**) $T \ll T_F$, (**c**) $T = T_F$, and (**d**) $T > T_F$ show 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.

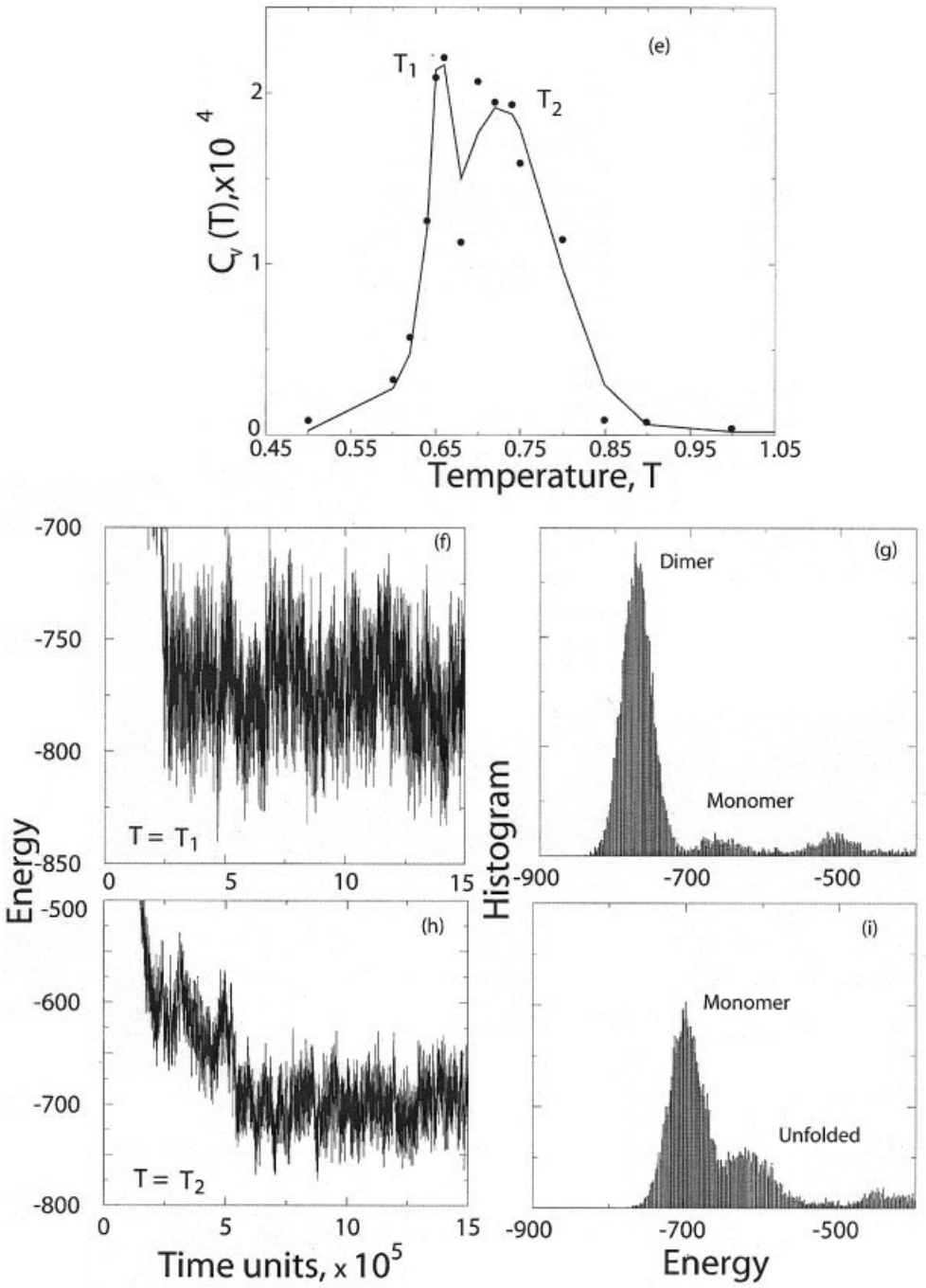


Figure 3. (Continued.)

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 antiparallel β-strands.

As a control, we obtain a SOD1 sequence profile for aggregation using the TANGO[47] server developed by Serrano and colleagues. We find that the "β-aggregation propensity" prediction from TANGO [Fig. 2(d)] is in good qualitative agreement with the aggregation profile obtained by our free energy method. The prediction from TANGO identifies the N- and the C-termini, and the residue sequence 100–120 identical to our prediction, and the region 35–45 identified by us is also amyloidogenic (although weakly so) according to the TANGO prediction [Fig. 2(d), inset].

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[48–53] and, therefore, the native topology of SOD1 determines the specific substructures 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,[54] simplified native-structure-based models such as the Gō model are used to study folding. In the Gō model,[36,37] 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 cutoff 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.[55] For modeling SOD1 folding, we use a coarse-grained representation of the SOD1 polypeptide chain developed by Ding et al.[56] (six-bead model, four backbone and two 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[45] (see Methods section). Starting from a stretched conformation, we perform equilibrium simulations at constant temperatures, T, ranging from $T = 0.1$ to 1.0 [Fig. 3 (a through d)]. We find that the six-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 [Fig. 3(a)] at the folding transition temperature, T_F. An MD trajectory at $T = T_F$ consists of two distinct populations—folded and unfolded—at equilibrium with each other, characteristic of a two-state protein [Fig. 3(c)]. At $T < T_F$ [Fig. 3(b)] and $T > T_F$ [Fig. 3(d)], on the other hand, only the folded and the unfolded states, respectively, are populated.

Next, we model the folding of the SOD1 dimer [Fig. 3(e through i)]. In addition to the intramonomeric contacts of the two monomers, the contact map of the dimer contains contacts corresponding to the native dimer interface [square II in Fig. 1(c)]. 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(e)]. The MD trajectories at T_1 consist of two distinct populations, corresponding to the dimer and two folded monomers, respectively [Fig. 3(f and g)], whereas at T_2, the two populations correspond to the folded monomers and unfolded monomers respectively [Fig. 3(h and i)]. At low temperatures ($T \ll T_1$), the native dimer is formed, whereas 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[21,22,57]:

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

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 (Fig. 3), we study the misfolding of the SOD1 dimer. In Eq. (4), when $\alpha_{\text{domain-swap}} > 1$, the relatively higher strength of intermonomer domain-swapped interactions (see Methods section) compared to the intramonomer interactions mimics an increase in the protein concentration. At high concentrations, native structure formation within a monomer competes with the formation of domain-swapped[24,25] topologies [Fig. 1(b and c)]. Further, nonspecific 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, $\alpha = 0.5$, 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 intermonomer contacts formed by each residue at the temperatures $T = 0.65$ to $T = 0.75$ (Fig. 4). 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, first we analyze intermonomer interactions in 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 $\alpha = 0.5$, the folding of the dimer is nativelike, as evidenced by frequency profile [Fig. 4(a)] where residues that are involved in the native dimer interface form intermonomer contacts with a high frequency at both $T = 0.65$ (black bars) and $T = 0.75$ (gray bars). However, other residues in the vicinity of the dimer interface residues, such as 3, 55, and 107–110 also form intermonomer 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 $\alpha = 1$, 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 [Fig. 4(b)]. The residues that form intermonomer contacts include the native dimer residues and several other residues, especially at the N- and the C-termini, and the

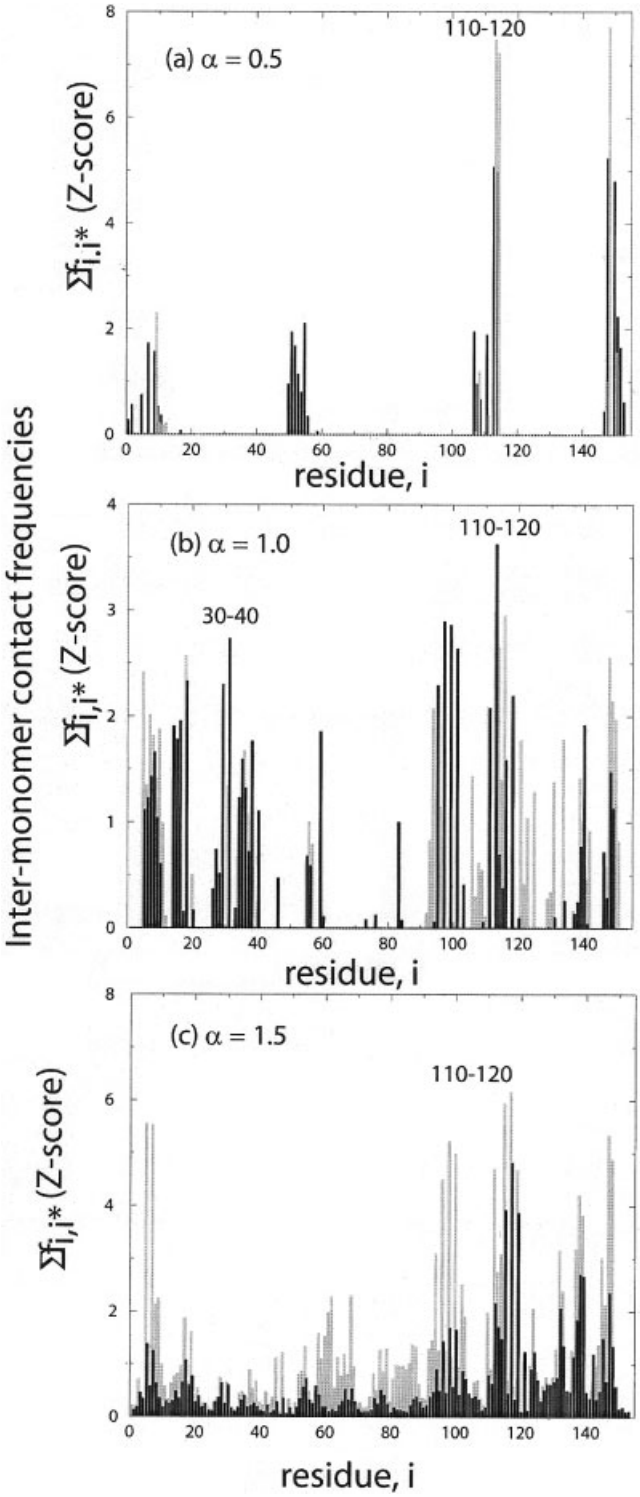


Fig. 4. Intermonomer interactions during misfolding. The cumulative frequency of intermonomer contacts at (**a**) $\alpha = 0.5$, (**b**) $\alpha = 1$, and (**c**) $\alpha = 1.5$ at $T = 0.65$ (black) and $T = 0.75$ (gray). As the strength of domain-swapped interactions increases, the native dimer interface is lost and several competing dimeric structures are populated.

regions 25–40 and 100–120. The formation of intermonomer 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 $\alpha = 1.5$ and $T = 0.65$, the native dimer interface is not seen as evidenced by a disappearance of intermonomer contacts by native interface residues [Fig. 4(c)]. Instead, a new set of residues, 55–65, 90–100, and 110–120, is observed to form the most intermonomer 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 alternative dimer interfaces.

To investigate the underlying structural interactions responsible for alternative dimer formation in our models, we analyze in detail dimer topology as a function of α-values, and T. We classify the observed topologies according to the elements of secondary structure forming intermonomer contacts in a complete phase diagram [Fig. 5(a)]. We find that at $\alpha \leq 0.5$, the folding of the dimer is nativelike in the temperature range $T = 0.2$ to $T = 1.0$, and no non-native dimer topologies are formed. The intermonomer contact frequency map in Figure 5(b) thus resembles the native contact map [square III in Fig. 1(c)]. Similarly, at $\alpha = 1$ and $T < 0.5$, the native dimer is formed. In the range $0.5 \leq T \leq 0.6$, dimers with a small number of non-native intermonomer contacts are formed [modified dimer MDI, Fig. 5(c)]. Interestingly, several of these contacts correspond to the identical sequences from the two monomers interacting with each other (e.g., residue sequences 39–44, 70–74, 85–92, and 123–135 self-associate and interact with each other) [Fig. 5(c)]. For $\alpha = 1$ and $0.6 \leq T \leq 0.8$ [Fig. 5(f)], and for $\alpha = 1.5$ and $0.2 \leq T \leq 0.9$ [Fig. 5(d and e)], extensive self-association of the following fragments occurs during the formation of structures with varying degrees of domain-swapped interactions (modified dimer MDII): the N-terminal, 21–40, 87–100, 90–115, 120–130, and the C-terminal [Fig. 5(d through f)]. This finding suggests that for model SOD1 domain swapping, self-association of certain sequences allows the formation of domain-swapping interactions. The self-associated fragments observed in simulations (with no explicit sequence information) overlap with those identified by sequence analysis [N- and C-termini, residues 35–45 and 100–120, Fig. 2(c)], indicating a synergy between structural and sequence elements in SOD1 with a high propensity to self-associate.

An examination of the structures with the varying degrees of domain swapping 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 some of these topologies. These observations are in agreement with recent computational and experimental studies[58,59] 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 Figure 5(g and h). Each of these two non-native dimeric states persists throughout the entire simulation time of 1.5×10^5 time

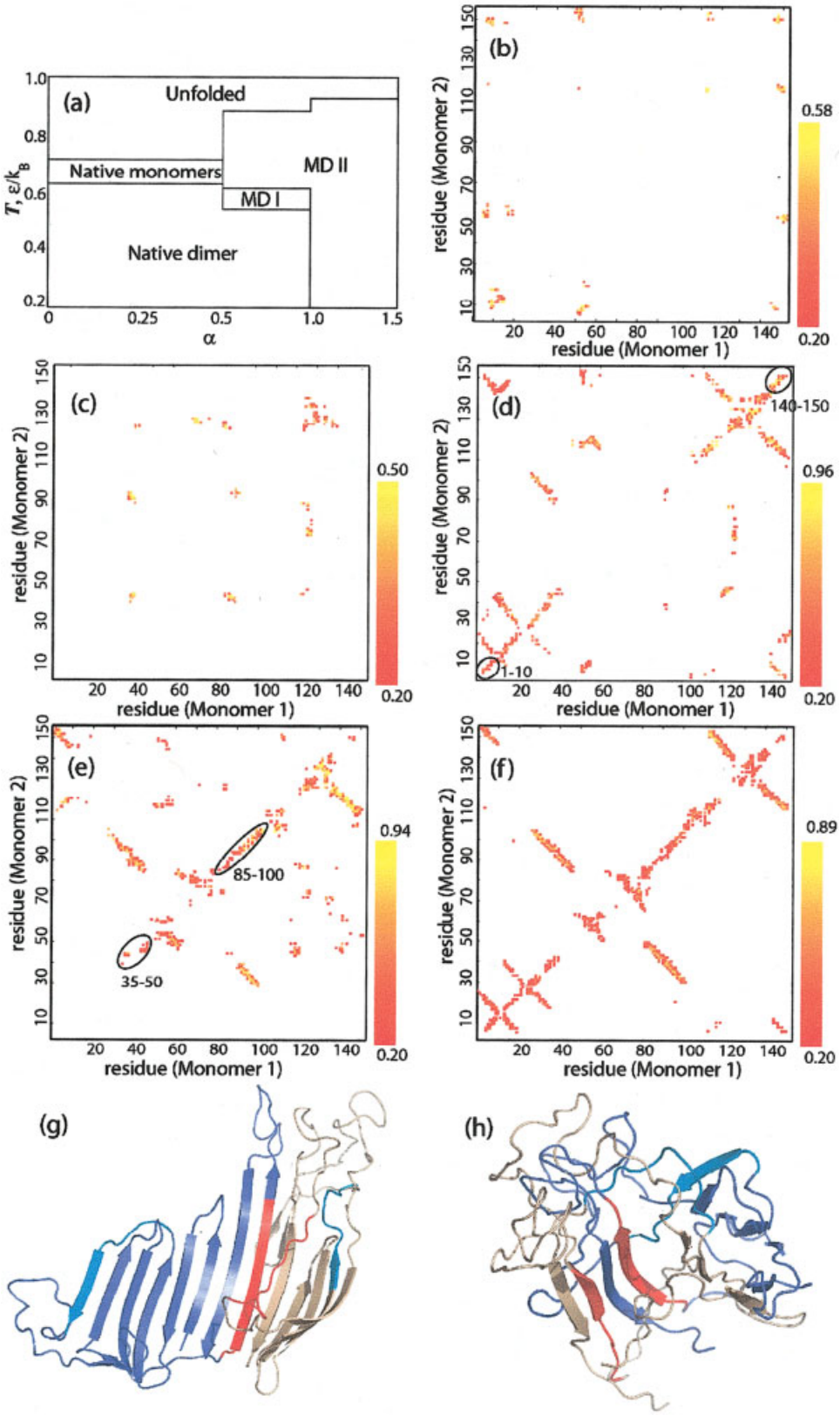

(a)
Unfolded
Native monomers
MD II
MD I
Native dimer
T, ε/k_B
α
(b)
(c)
(d)
140-150
1-10
(e)
85-100
35-50
(f)
residue (Monomer 1)
residue (Monomer 2)
0.58
0.50
0.96
0.94
0.89
0.20
(g)
(h)

Figure 5.

units once it is formed. Figure 5(g) shows a domain-swapping interaction to form a nativelike β-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. Figure 5(h) shows a non-native SOD1 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.

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 and Serrano[14] 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 antiparallel conformations. Although protein and peptide aggregation involves conversion to β-sheets, whether the β-sheets are parallel or antiparallel 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, antiparallel topologies are preferably formed.[60] In recent study of the dynamics of peptide dimerization, Hwang et al.[58] found that the β-strand peptide dimers exhibited both parallel and antiparallel β-sheets, with an overall preference for antiparallel arrangements. Since both parallel and antiparallel topologies are likely in the early stages of aggregation, we used both topologies for constructing the templates, and also found that antiparallel β-sheets have lower free energies for a given sequence [Fig. 2(b and c)]. We postulate that the lower free energy of the antiparallel structures is due to less strained hydrogen bonds in the antiparallel 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 nonlinearly 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 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.[43,61]

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.[62] In protein modeling algorithms, for example, the ROSETTA program developed by Baker and coworkers,[63,64] 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.[14,65,66] 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 frag-

Fig. 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 intermonomer 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 intermonomer 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 dimers are shown in (**g**) and (**h**). These structures are formed by in-register interactions between residues 110 and 120 (red) or residues 35 and 45 (blue). We propose that there are multiple mechanisms by which these modified dimers can further propagate the aggregate structures and eventually form fibrils.

ments 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 obtained from our all-atom MD simulations of sequence fragments is simply a reflection of some intrinsic physicochemical property of the fragments, such as β-strand propensity and hydrophobicity, we compare the amyloidogenicity profile with these physicochemical 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.,[66,67] in which average hydrophobicity, β-sheet propensity, and net charge are calculated for each sequence and a linear regression fit to the aggregation rate is obtained. Following Chiti et al., we assume that the free energy of multimer formation calculated in our MD simulations is

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

where p_{β}, $p_{hydrophob}$, and $p_{\pm}$ are the β-sheet propensity, hydrophobicity, and net charge, respectively, 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. 6(a through c)]. Second, for each template structure, namely, the dimer and the tetramer with parallel and antiparallel 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. 6(d through 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 coworkers, 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.[68]

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 induced by domain swapping. We use no explicit sequence information during the identification of structural regions that are likely to self-associate, and vice versa. Since we find overlapping 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[21] that a minimal mechanism for the aggregation of SOD1 is

$$D \rightleftharpoons 2M_{apo} + metals \overset{k_{agg}}{\rightarrow} A, \quad (7)$$

where D is the native dimer, M_{apo} is the apo-monomer, and A is the aggregate. The mechanism in Eq. (7) is minimal because the second reaction, $M_{apo} \rightarrow A$, is likely to be a multistep 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.[25,38] The domain swapping of two monomers leads to the formation of structures that 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 intermonomer contacts[30,38]). Therefore, we study the misfolding of SOD1 dimer using domain swapping and find that the formation of domain-swapped structures occurs with the self-association of specific elements of the SOD1 structure. These self-associated sequences identified by domain-swapping simulations with no explicit sequence information include all the regions identified by sequence analysis, suggesting that SOD1 domain-swapping-induced misfolding allows, under certain conditions, the self-association of fragments with a propensity to self-associate.

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 $\alpha_{domain\text{-}swap}$, with which we control the relative of strengths of intra- and interchain interactions. Under a range of these destabilizing conditions, we observe the preferential self-association of residues fragments N- and C-termini, 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.[38] 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 environ-

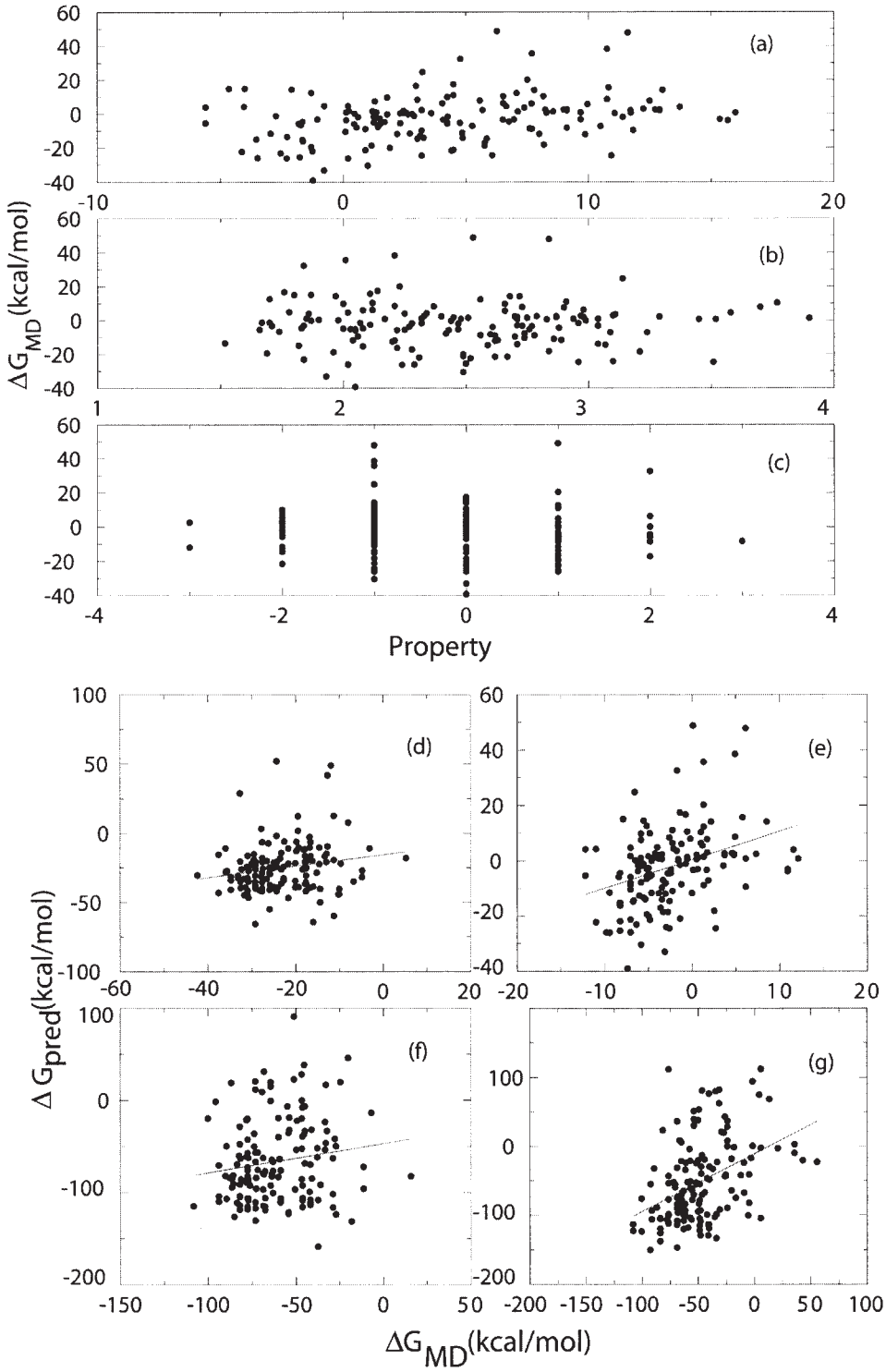


Fig. 6. The correlation between the free energies of dimerization into antiparallel β-strands obtained in MD simulations with physicochemical properties of the fragment. Correlations of free-energy of association with (**a**) the average hydrophobicity, (**b**) β-sheet propensity, and (**c**) net charge 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**) antiparallel dimeric, (**f**) parallel dimeric, (**g**) antiparallel tetrameric, and (**h**) parallel tetrameric template structures. We conclude that the free energy of oligomerization calculated in the MD simulation is not sufficiently explained by these physicochemical properties of the sequence.

mental conditions, such as denaturants, pH, and temperature, that have been used to generate the aggregates.[23,69–71]

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 interchain contacts.

ACKNOWLEDGMENT

We thank F. Ding for assistance with simulations and for helpful discussions.

REFERENCES

1. Dobson CM. Protein folding and misfolding. Nature 2003;426:884–890.
2. Dobson CM. Protein chemistry. In the footsteps of alchemists. Science 2004;304:1259–1262.
3. Cleveland DW, Rothstein JD. From Charcot to Lou Gehrig: deciphering selective motor neuron death in ALS. Nat Rev Neurosci 2001;2:806–819.
4. Reches M, Porat Y, Gazit E. Amyloid fibril formation by pentapeptide and tetrapeptide fragments of human calcitonin. J Biol Chem 2002;277:35475–35480.
5. Tenidis K, Waldner M, Bernhagen J, Fischle W, Bergmann M, Weber M, Merkle ML, Voelter W, Brunner H, Kapurniotu A. Identification of a penta- and hexapeptide of islet amyloid polypeptide (IAPP) with amyloidogenic and cytotoxic properties. J Mol Biol 2000;295:1055–1071.
6. Thirumalai D, Klimov DK, Dima RI. Emerging ideas on the molecular basis of protein and peptide aggregation. Curr Opin Struct Biol 2003;13:146–159.
7. Villegas V, Zurdo J, Filimonov VV, Aviles FX, Dobson CM, Serrano L. Protein engineering as a strategy to avoid formation of amyloid fibrils. Protein Sci 2000;9:1700–1708.
8. Ventura S, Zurdo J, Narayanan S, Parreno M, Mangues R, Reif B, Chiti F, Giannoni E, Dobson CM, Aviles FX, Serrano L. Short amino acid stretches can mediate amyloid formation in globular proteins: the Src homology 3 (SH3) case. Proc Natl Acad Sci USA 2004;101:7258–7263.
9. Ivanova MI, Sawaya MR, Gingery M, Attinger A, Eisenberg D. An amyloid-forming segment of beta 2-microglobulin suggests a molecular model for the fibril. Proc Natl Acad Sci USA 2004;101:10584–10589.
10. Tycko R. Progress towards a molecular-level structural understanding of amyloid fibrils. Curr Opin Struct Biol 2004;14:96–103.
11. Benzinger TLS, Gregory DM, Burkoth TS, Miller-Auer H, Lynn DG, Botto RE, Meredith SC. Propagating structure of Alzheimer's beta-amyloid(10-35) is parallel beta-sheet with residues in exact register. Proc Natl Acad Sci USA 1998;95:13407–13412.
12. Makin OS, Atkins E, Sikorski P, Johansson J, Serpell LC. Molecular basis for amyloid fibril formation and stability. Proc Natl Acad Sci USA 2005;102:315–320.
13. Gordon DJ, Balbach JJ, Tycko R, Meredith SC. Increasing the amphiphilicity of an amyloidogenic peptide changes the beta-sheet structure in the fibrils from antiparallel to parallel. Biophys J 2004;86:428–434.
14. de la Paz ML, Serrano L. Sequence determinants of amyloid fibril formation. Proc Natl Acad Sci USA 2004;101:87–92.
15. Rochet JC, Lansbury PT. Amyloid fibrillogenesis: themes and variations. Curr Opin Struct Biol 2000;10:60–68.
16. Richardson JS, Richardson DC. Natural beta-sheet proteins use negative design to avoid edge-to-edge aggregation. Proc Natl Acad Sci USA 2002;99:2754–2759.
17. Tainer JA, Getzoff ED, Beem KM, Richardson JS, Richardson DC. Determination and analysis of the 2A structure of copper, zinc superoxide-dismutase. J Mol Biol 1982;160:181–217.
18. Rodriguez JA, Valentine JS, Eggers DK, Roe JA, Tiwari A, Brown RH Jr, Hayward LJ. Familial amyotrophic lateral sclerosis-associated mutations decrease the thermal stability of distinctly metallated species of human copper/zinc superoxide dismutase. J Biol Chem 2002;277:15932–15937.
19. Lindberg MJ, Tibell L, Oliveberg M. Common denominator of Cu/Zn superoxide dismutase mutants associated with amyotrophic lateral sclerosis: decreased stability of the apo state. Proc Natl Acad Sci USA 2002;99:16607–16612.
20. Johnston JA, Dalton MJ, Gurney ME, Kopito RR. 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 2000;97:12571–12576.
21. Khare SD, Caplow M, Dokholyan NV. Determination of the rate and equilibrium constants for a multi-step reaction for aggregation of superoxide dismutase in ALS. Proc Natl Acad Sci USA 2004;101:15094–15099.
22. Ray SS, Nowak RJ, Strokovich K, Brown RH Jr, Walz T, Lansbury PT Jr. An intersubunit disulfide bond prevents in vitro aggregation of a superoxide dismutase-1 mutant linked to familial amytrophic lateral sclerosis. Biochemistry 2004;43:4899–4905.
23. Rakhit R, Crow JP, Lepock JR, Kondejewski LH, Cashman NR, Chakrabartty A. Monomeric Cu,Zn-superoxide dismutase is a common misfolding intermediate in the oxidation models of sporadic and familial amyotrophic lateral sclerosis. J Biol Chem 2004;279:15499–15504.
24. Bennett MJ, Choe S, Eisenberg D. Domain swapping—entangling alliances between proteins. Proc Natl Acad Sci USA 1994;91:3127–3131.
25. Schlunegger MP, Bennett MJ, Eisenberg D. Oligomer formation by 3D domain swapping: a model for protein assembly and misassembly. Adv Protein Chem 1997;50:61–122.
26. Liu YS, Gotte G, Libonati M, Eisenberg D. A domain-swapped RNase A dimer with implications for amyloid formation. Nat Struct Biol 2001;8:211–214.
27. Rousseau F, Schymkowitz JWH, Itzhaki LS. The unfolding story of three-dimensional domain swapping. Structure 2003;11:243–251.
28. Liu Y, Eisenberg D. 3D domain swapping: as domains continue to swap. Protein Sci 2002;11:1285–1299.
29. Yang S, Levine H, Onuchic JN. Protein oligomerization through domain swapping: Role of inter-molecular interactions and protein concentration. J Mol Biol. Forthcoming.
30. Yang SC, Cheung MS, Onuchic JN, Levine H. Molecular dynamics simulations on domain swapping. Biophys J 2004;86:267A–268A.
31. Yang SC, Cho SS, Levy Y, Cheung MS, Levine H, Wolynes PG, Onuchic JN. Domain swapping is a consequence of minimal frustration. Proc Natl Acad Sci USA 2004;101:13786–13791.
32. Rousseau F, Schymkowitz JWH, Wilkinson HR, Itzhaki LS. Three-dimensional domain swapping in p13suc1 occurs in the unfolded state and is controlled by conserved proline residues. Proc Natl Acad Sci USA 2001;98:5596–5601.
33. Liu YS, Hart PJ, Schlunegger MP, Eisenberg D. The crystal structure of a 3D domain-swapped dimer of RNase A at a 2.1-angstrom resolution. Proc Natl Acad Sci USA 1998;95:3437–3442.
34. Liu YS, Gotte G, Libonati M, Eisenberg D. Structures of the two 3D domain-swapped RNase A trimers. Protein Sci 2002;11:371–380.
35. 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. Nat Struct Biol 2001;8:316–320.
36. Go N, Abe H. Noninteracting local-structure model of folding and unfolding transition in globular proteins: I. Formulation. Biopolymers 1981;20:991–1011.
37. Abe H, Go N. Noninteracting local-structure model of folding and unfolding transition in globular proteins: II. Application to two-dimensional lattice proteins. Biopolymers 1981;20:1013–1031.
38. Ding F, Dokholyan NV, Buldyrev SV, Stanley HE, Shakhnovich EI. Molecular dynamics simulation of the SH3 domain aggregation suggests a generic amyloidogenesis mechanism. J Mol Biol 2002;324:851–857.
39. Kishan KVR, Scita G, Wong WT, DiFiore PP, Newcomer ME. The SH3 domain of Eps8 exists as a novel intertwined dimer. Nat Struct Biol 1997;4:739–743.

40. Sunde M, Serpell LC, Bartlam M, Fraser PE, Pepys MB, Blake CC. Common core structure of amyloid fibrils by synchrotron X-ray diffraction. J Mol Biol 1997;273:729–739.
41. Hermans J, Berendsen HJC, Vangunsteren WF, Postma JPM. A consistent empirical potential for water–protein interactions. Biopolymers 1984;23:1513–1518.
42. Vorobjev YN, Almagro JC, 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 1998;32:399–413.
43. Urbanc B, Cruz L, Ding F, Sammond D, Khare S, Buldyrev SV, Stanley HE, Dokholyan NV. Molecular dynamics simulations of amyloid beta dimer formation. Biophys J 2004;87:2310–2321.
44. Vorobjev YN, Hermans J. Free energies of protein decoys provide insight into determinants of protein stability. Protein Sci 2001;10: 2498–2506.
45. Khare SD, Ding F, Dokholyan NV. Folding of Cu, Zn superoxide dismutase and familial amyotrophic lateral sclerosis. J Mol Biol 2003;334:515–525.
46. Rubinstein M, Colby RH. Polymer physics. New York: Oxford University Press; 2003.
47. Fernandez-Escamilla AM, Rousseau F, Schymkowitz J, Serrano L. Prediction of sequence-dependent and mutational effects on the aggregation of peptides and proteins. Nat Biotechnol 2004;22:1302–1306.
48. Dokholyan NV, Li L, Ding F, Shakhnovich EI. Topological determinants of protein folding. Proc Natl Acad Sci USA 2002;99:8637–8641.
49. Dokholyan NV, Borreguero JM, Buldyrev SV, Ding F, Stanley HE, Shakhnovich EI. Identifying importance of amino acids for protein folding from crystal structures. Macromol Crystallogr D 2003;374: 618–640.
50. Pande VS, Grosberg AY, Tanaka T. Heteropolymer freezing and design: towards physical models of protein folding. Rev Modern Phys 2000:72:259–314.
51. Bryngelson JD, Onuchic JN, Socci ND, Wolynes PG. Funnels, pathways, and the energy landscape of protein-folding—a synthesis. Proteins 1995;21:167–195.
52. Shea JE, Onuchic JN, Brooks CL. Molecular dynamics study of the mechanism and thermodynamics of folding of the src-SH3 protein. Abstracts of Papers of the American Chemical Society 2001;221:U406–U407.
53. Plotkin SS, Onuchic JN. Understanding protein folding with energy landscape theory—Part I: Basic concepts. Quart Rev Biophys 2002;35:111–167.
54. Ding F, Dokholyan NV. Simple but predictive protein models. Trends Biotechnol. Forthcoming.
55. Dokholyan NV, Buldyrev SV, Stanley HE, Shakhnovich EI. Discrete molecular dynamics studies of the folding of a protein-like model. Fold Des 1998;3:577–587.
56. Ding F, Borreguero JM, Buldyrey SV, Stanley HE, Dokholyan NV. Mechanism for the alpha-helix to beta-hairpin transition. Proteins 2003;53:220–228.
57. Mei G, Rosato N, Silva N Jr, Rusch R, Gratton E, Savini I, Finazzi-Agro A. Denaturation of human Cu/Zn superoxide-dismutase by guanidine-hydrochloride—a dynamic fluorescence study. Biochemistry 1992;31:7224–7230.
58. Hwang W, Zhang SG, Kamm RD, Karplus M. Kinetic control of dimer structure formation in amyloid fibrillogenesis. Proc Natl Acad Sci USA 2004;101:12916–12921.
59. Tcherkasskaya O, Sanders W, Chynwat V, Davidson EA, Orser CS. The role of hydrophobic interactions in amyloidogenesis: example of prion-related polypeptides. J Biomol Struct Dyn 2003;21:353–365.
60. Tycko R. Solid-state nuclear magnetic resonance techniques for structural studies of amyloid fibrils. Methods Enzymol 2001;339: 390–413.
61. Urbanc B, Cruz L, Yun S, Buldyrev SV, Bitan G, Teplow DB, Stanley HE. In silico study of amyloid beta-protein folding and oligomerization. Proc Natl Acad Sci USA 2004;101:17345–17350.
62. Doyle R, Simons K, Qian H, Baker D. Local interactions and the optimization of protein folding. Proteins 1997;29:282–291.
63. Simons KT, Ruczinski I, Kooperberg C, Fox BA, Bystroff C, Baker D. Improved recognition of native-like protein structures using a combination of sequence-dependent and sequence-independent features of proteins. Proteins 1999;34:82–95.
64. Bystroff C, Baker D. Prediction of local structure in proteins using a library of sequence–structure motifs. J Mol Biol 1998;281:565–577.
65. Gazit E. A possible role for pi-stacking in the self-assembly of amyloid fibrils. FASEB J 2002;16:77–83.
66. Chiti F, Stefani M, Taddei N, Ramponi G, Dobson CM. Rationalization of the effects of mutations on peptide and protein aggregation rates. Nature 2003;424:805–808.
67. Chiti F, Calamai M, Taddei N, Stefani M, Ramponi G, Dobson CM. Studies of the aggregation of mutant proteins in vitro provide insights into the genetics of amyloid diseases. Proc Natl Acad Sci USA 2002;99:16419–16426.
68. 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. J Mol Biol 2004;342:345–353.
69. DiDonato M, Craig L, Huff ME, Thayer MM, Cardoso RM, Kassmann CJ, Lo TP, Bruns CK, Powers ET, Kelley JW, Getzoff ED, Tainer JA. ALS mutants of human superoxide dismutase form fibrous aggregates via framework destabilization. J Mol Biol 2003;332:601–615.
70. Stathopulos PB, Rumfeldt JA, Scholz GA, Irani RA, Frey HE, Hallewell RA, Lepock JR, Meiering EM. Cu/Zn superoxide dismutase mutants associated with amyotrophic lateral sclerosis show enhanced formation of aggregates in vitro. Proc Natl Acad Sci USA 2003;100:7021.
71. Rakhit R, Cunningham P, Furtos-Matei A, Dahan S, Qi XF, Crow JP, Cashman NR, Kondejewski LH, Chakrabatty A. Oxidation-induced misfolding and agagaregation of superoxide dismutase and its implications for amyotrophic lateral sclerosis. J Biol Chem 2002;277:47551–47556.