Early steps in thermal unfolding of superoxide dismutase 1 are similar to the conformational changes associated with the ALS-associated A4V mutation

Tom Schmidlin1, Ken Ploeger2, Amanda L. Jonsson1 and Valerie Daggett1,2,3

1Department of Bioengineering, University of Washington, Seattle, WA 98195-5013 and 2Department of Biochemistry, University of Washington, Seattle, WA 98195-5013

3To whom correspondence should be addressed.
E-mail: daggett@u.washington.edu

Received May 16, 2013; revised May 16, 2013; accepted May 20, 2013

There are over 100 mutations in Cu/Zn superoxide dismutase (SOD1) that result in a subset of familial amyotrophic lateral sclerosis (fALS) cases. The hypothesis that dissociation of the dimer, misfolding of the monomer and subsequent aggregation of mutant SOD1 leads to fALS has been gaining support as an explanation for how these disparate missense mutations cause the same disease. These forms are only responsible for a fraction of the ALS cases; however, the rest are sporadic. Starting with a folded apo monomer, the species considered most likely to be involved in misfolding, we used high-temperature all-atom molecular dynamics simulations to explore the events of the wild-type protein unfolding through the denatured state. All simulations showed early loss of structure along the β5–β6 edge of the β-sandwich, supporting earlier findings of instability in this region. Transition state structures identified from the simulations are in good agreement with experiment, providing detailed, validated molecular models for this elusive state. Furthermore, we compare the process of thermal unfolding investigated here to that of the lethal A4V mutant-induced unfolding at physiological temperature and find that the pathways are very similar.

Keywords: ALS/misfolding/molecular dynamics/SOD1/unfolding

Introduction

Cu/Zn superoxide dismutase (SOD1) catalyzes the dismutation of superoxide, a byproduct of cellular respiration, to O2 and H2O2 (McCord and Fridovich, 1969). The protein is a 35 kDa cytoplasmic homodimer that is ubiquitous. Each monomer consists of 153 residues with eight antiparallel β strands (Tainer et al., 1983) in an immuoglobulin-like fold. The two sheets, which we call ‘A’ and ‘B’, consist of strands β1, 2, 3, 6 and β5, 4, 7, 8, respectively (Fig. 1A). Three long loops follow β4, β6 and β7: loop IV (residues 49–82), loop VI (102–115) and the electrostatic (ES) loop (121–142). A disulfide bond anchors loop IV to sheet A, supporting the channel into the active site between loop IV and the ES loop.

The charged residues of the ES loop have been suggested to be crucial for guiding the substrate into the active site (Getzoff et al., 1992; Banci et al., 1993; Fishet et al., 1994).

There is interest in SOD1 due to its association with in amyotrophic lateral sclerosis (ALS) (Rosen et al., 1993), a fatal motor neurodegenerative disease. Of the ALS cases, ~10% are familial, and, of these, ~20% have been linked to SOD1 (Cleveland and Rothstein, 2001; Wood et al., 2003). The method of pathogenesis of SOD1-related ALS is unknown, but several hypotheses have been proposed. One model involves mutant SOD1 misfolding to form a toxic species (Rakhit and Chakrabartty, 2006), which has been supported by many studies finding mutations decrease stability (Lindberg et al., 2002; Rodriguez et al., 2002; DiDonato et al., 2003; Shaw and Valentine, 2007; Banci et al., 2008), leading to amyloidogenic precursors to aggregation (Nordlund and Oliveberg, 2008). The misfolded protein hypothesis can explain the common phenotype of SOD1-associated ALS cases, as well as the over 100 sequentially and structurally diverse point mutations linked to the disease (ALSoD, Lill et al., 2011).

It is thought that the monomeric apo form (lacking the copper and zinc ions) represents the start of the pathway to a toxic species due to its low stability. The absence of metal binding decreases stability significantly (Assfalg et al., 2003; Banci et al., 2003; Lynch et al., 2004; Svensson et al., 2006; Bruns and Kopito, 2007; Strange et al., 2007; Ding and Dokholyan, 2008; Kayatekin et al., 2008), and there is a correlation between the severity of ALS-associated mutations and the mutant monomer’s instability (Lindberg et al., 2002). While most studies of the apo monomer have focused on ALS-associated mutations, studies examining the folding behavior of the wild-type protein have also been informative. Protein engineering (Nordlund and Oliveberg, 2006) and NMR studies (Assfalg et al., 2003; Banci et al., 2003) point to instability caused by the aggregation propensity of the β5–β6 edge, yet a more detailed model is desirable to better understand this process.

Molecular dynamics (MD) simulations are useful for adding details to the broad interpretations made from experiment. We have found that high-temperature (498 K) unfolding simulations have produced good agreement with experimental results for many proteins (Daggett et al., 1998; Fulton et al., 1999; Alonso and Daggett, 2000; Gianni et al., 2003; Jemth et al., 2004; Ferguson et al., 2005; Jemth et al., 2005; White et al., 2005; Petrovich et al., 2006; Scott et al., 2006; Sharpe et al., 2007), allowing us to view the unfolding pathway on a computationally feasible timescale. Here, we present five high-temperature unfolding simulations and one native state simulation of an apo, wild-type SOD1 monomer that agree well with Φ-value transition state (TS) analysis (Nordlund and Oliveberg, 2006). MD shows that the β5–β6 edge is the first to lose structure, followed by sheet A strands,
and then sheet B strands and the β1–β8 edge. The A4V SOD1 mutant, which results in one of the fastest declines among human patients, also loses structure at the β5–β6 edge in MD simulations at physiological temperature and neutral pH (Schmidlin et al., 2009).

**Methods**

**MD simulations**

The starting structure for all wild-type simulations was the 1.8 Å crystal structure of human SOD1 (1HL5—Chain A) (Strange et al., 2003), with the Cu^{2+} and Zn^{2+} ions removed. As discussed previously (Schmidlin et al., 2009), 1HL5 is preferred over the apo 1HL4 structure (Strange et al., 2003) because the 1HL4 structure is missing several residues and heavy atoms, and requires extensive model building to be useable for simulation.

We performed six independent simulations: two at 498 K for 33 ns each (498—1 and 498—2), three at 498 K for 5 ns each (498—3, 498—4, 498—5) and one at 298 K for 21 ns. In the 298 K simulation the disulfide bond (C57-C146) was left intact since it is normally buried and the simulation temperature is well below the T_m for metal-binding mutant SOD1 (Rodríguez et al., 2002). In the 498 K simulations this bond was removed prior to energy minimization because unfolding would likely expose it to the reducing environment of the cytosol, and removing this stabilizing bond allows for more complete unfolding of the protein. Additionally, while there is evidence that aberrant disulfide bonding involving the C57 and C146 residues is not necessary for aggregates to form (Karch and Borchelt, 2008), mutant protein with the bond reduced has been found to be a ‘major component of insoluble SOD1 aggregates’ (Karch et al., 2009). There is further evidence that protein disulfide isomerase is a risk factor in the disease (Kwok et al., 2013).

Minimization was preceded by the addition of hydrogen atoms to the SOD1 crystal structure (metal ions previously removed). This was followed by 1000 steps of steepest descent energy minimization. The protein was then solvated in a periodic box of flexible 3-center (F3C) water molecules (Levitt et al., 1997) that extended at least 10 Å away from any protein atom. Water densities were set at 0.829 g/cm^3 (Haar et al., 1984) and 0.997 g/cm^3 (Kell, 1967) for the 498 and 298 K simulations, respectively. Next, another 1000 steps of steepest descent minimization was performed on the entire system and was followed by dynamics of the solvent for 1 ps. After this, the system was heated to the desired temperatures (298 and 498 K) in 0.1° increments and was followed by minimization of the solvent for 500 steps. The protein was then minimized for another 500 steps.

Simulations were performed using in *lucem* molecular mechanics (ilmm) (Beck et al., 2000–2013) with previously described protocols (Beck and Daggett, 2004). These included a force field (Levitt et al., 1995), a 2 fs timestep and use of the microcanonical ensemble (NVE, constant number of particles, volume and total energy). Force-shifted 8 Å (498 K simulations) and 10 Å (298 K simulation) non-bonded cut-offs were used with periodic boundary conditions, and the interaction pair list was updated every three steps. We have reported previously (Beck et al., 2005) that the process of unfolding is not influenced by the difference in 8 Å versus 10 Å cut-off. Structures were saved every 1 ps for all simulations. We also saved structures at 0.2 ps granularity during the 0–2 ns time period in the 498 K simulations to have finer resolution when analyzing the TS.

**Transition state assignment**

A TS ensemble for each 498 K simulation was identified using the conformational clustering method (Li and Daggett, 1994; 1998). In brief, the Ca-RMSD was calculated for each structure to every other structure in the same simulation, resulting in a symmetrical matrix. Multidimensional scaling reduced the matrix to 3D, which then could be visualized as a 3D projection. Structures that were similar, with low Ca-RMSD to each other, appeared as clusters, and the TS ensemble was defined as the 5 ps window before the exit from the first extended native-like cluster. At 0.2 ps granularity, this resulted
in a total of 26 structures for each TS ensemble. The rationale for this approach has been described previously (Li and Daggett, 1994).

**Structure index**
The structure index, S-value (S), is useful as a semi-quantitative summary measure of structure and for comparing to experimentally determined Φ-values (Daggett et al., 1996). An S-value is comprised of two terms multiplied together, S2 and S3, which measure secondary and tertiary structure, respectively. For a given residue, an S-value close to 1, like Φ, suggests native-like structure in the TS, and values close to 0 suggest that the residue is unstructured in the TS. S- and Φ-values can be compared to determine the agreement between simulation and experiment, and S-values have been shown to agree well with Φ-values for many proteins (Daggett et al., 1996, 1998; Li and Daggett, 1998; Fulton et al., 1999; Gianni et al., 2003; Day and Daggett, 2005).

**Native state flexibility**
The average Cα flexibility was calculated using principal components as described by Teodoro et al. (2003), and implemented and employed by Benson and Daggett (2008). Following a 2-ns equilibrium period, we aligned the structure at every picosecond to the starting structure by using a rigid least-squares fitting of Cα atoms and holding the protein’s center of mass at the origin. The mean position for each heavy atom was calculated by subtracting the mean position from the center of mass at the origin. The mean position for each heavy atom was considered the ‘primary axis of flexibility’ and the variance term. The direction of the first principal component was assigned by hydrogen bonding using our DSSP (Kabsch and Sander, 1983) analysis module (Souras and Daggett, 2008). Images were prepared using UCSF Chimera (Pettersen et al., 2004).

**Analysis methods**
Analyses were performed at 1 ps granularity using ielm. Additional analysis in the TSs was performed at 0.2 ps to improve sampling within these ensembles. Atoms were considered to be in contact if the C–C distance was ≤5.4 Å or any other heavy-atom distance was ≤4.6 Å. Secondary structure was assigned by hydrogen bonding using our DSSP (Kabsch and Sander, 1983) analysis module (Souras and Daggett, 2008). Images were prepared using UCSF Chimera (Pettersen et al., 2004).

**Results and discussion**
In the following results we begin with the native state and then describe the protein unfolding process and pivotal conformational states populated along the way, noting similarities and differences between the different independent trajectories.

**The monomeric, apo native state of SOD1**
The Cα-RMSD of the 298 K simulation increased gradually over the first 2 ns to 2.5 Å from the starting structure. From 2 to 21 ns, the Cα-RMSD ranged from 2 to 3 Å. Thus, we used the structures from 2 to 21 ns to represent the apo native state ensemble. The average Cα-RMSD from the starting structure over this time period was 2.6 Å (±0.1 Å).

The major structural change observed involved the loss of secondary structure from the C-terminal end of the β6 strand (the bottom as shown in Fig. 1), which continued upwards towards the β5–β6 turn, beginning after 5 ns and as the simulation progressed. Similar loss of β structure in β6 has been noted in other simulations (Schmidlin et al., 2009). Some fraying of the last one or two residues on each end was seen in other strands, but no other widespread secondary structure loss was observed.

The average Cα flexibility was analyzed to further investigate movement in the native state ensemble. Within loop IV, residues 60–65 and 70–71 were less flexible than the surrounding residues (Fig. 1B and C), suggesting that these regions served as an anchor, limiting the flexibility of the rest of loop IV. Residues 63 and 71 bind the structurally important zinc atom, with H63 also binding the catalytically active copper ion. These anchors are apparent as local minima in the flexibility graph between β4 and β5 (Fig. 1B). A similar low flexibility region was present in the helix of the ES loop, residues 134–136. Sheet A (β1, 2, 3 and 6) was more flexible than sheet B (β5, 4, 7 and 8), and both sheets showed increasing flexibility towards the β5–β6 edge (Fig. 1B and C). The most flexible regions in the entire protein were β6 and the loop between β1 and β2.

**Pre-transition state events in the thermal unfolding of SOD1**
Simulations 498-1, 498-3 and 498-5 unfolded in a similar manner, which will be referred to as the dominant unfolding pathway. Simulation 498-1 is discussed in detail below, with the minor differences in 498-3 and 498-5 noted. The other simulations (498-2 and 498-4) were considered alternate pathways, although the broad trends presented in this section were similar across all five simulations.

Unfolding began with the straightening of the topmost portion of loop IV away from sheet B and the start of main chain hydrogen bond loss. The movement in loop IV was associated with the loss of hydrophobic contacts between residues 57–63 and β7, occurring from 50 to 70 ps. Similar movement of loop IV is suggested in NMR structures of the apo monomer (Banci et al., 2003; Hornberg et al., 2007), hinting that this movement may be a feature of the apo monomers rather than being linked to unfolding per se.

Concurrent with the loss of hydrophobic contacts, main chain hydrogen bonds were lost in sheet A and between β5 and neighboring strands (Fig. 2). Of the main-chain hydrogen bonds in sheet A, those connecting β1 and β2 were most persistent (Fig. 2). An experimental study tracking the early events of thermal unfolding using hydrogen/deuterium exchange (Durazo et al., 2009) supports our observation of earlier loss of structure between β2 and β3 than between β1 and β2.

Side chains in β2–β3, β5–β4 and β6–β5 began to lose contacts from 70 to 100 ps (Fig. 2). For β2–β3 residues, only the solvent exposed (Q15-N26, N19-W32, E21-W32 and E21-K30) side chain contacts were lost prior to the TS. The alternate unfolding simulations (498-3 and 498-5) also lost these β2–β3 side chain contacts before others were lost, but not until after the TS. Similarly, main chain hydrogen bonds were completely lost between β2–β3, β5–β4 and β6–β5 (Fig. 2) prior to the TS. In sheet A, strands β1–β2 and β3–β6 had persistent hydrogen bonds lasting past the TS. Between β1 and β2, the longest lasting were on the N-terminal half of the strand. Conversely, hydrogen bonds between β3 and β6 showed greater persistence towards the top of the strands.
Simulation 498-3 showed differences in main chain hydro-
gen bond loss. Strands $\beta_1-\beta_2$ were the first to lose hydrogen bonds, followed by $\beta_2-\beta_3$ and $\beta_3-\beta_6$ loss. However, as in 498-1, the longest lasting hydrogen bonds were between $\beta_1$ and $\beta_2$. Persistence between $\beta_3$ and $\beta_6$ differed slightly with the longest lasting hydrogen bond at the bottom of the strands, V31-I99, rather than at the top.

By the 156 ps TS structure (Fig. 3A), the $\beta_5-\beta_6$ edge opened, sheet A was losing main chain hydrogen bonds and loop IV moved upwards. Long-lasting hydrogen bonds existed near the bottom of $\beta_1-\beta_2$, and some hydrogen bonds between strands $\beta_3$ and $\beta_6$ remained intact. Local unfolding of the $\beta_5-\beta_6$ edge has been suggested as a potential site for aggregation by previous experimental (Nordlund and Oliveberg, 2006) and MD studies (Khare and Dokholyan, 2006; Strange et al., 2007; Ding and Dokholyan, 2008; Schmidlin et al., 2009), and the inherent instability suggested by our simulations could lead to edge-to-edge aggregation.
Potential TS structures were identified using conformational clustering (Li and Daggett, 1994, 1998). The average Ca-RMSD of the TS ensemble to the starting structure ranged from 3.6 to 6.5 Å. Representative structures from each TS ensemble are given in Fig. 4A.

Sheet A contained varying amounts of native structure in the various TS ensembles. Typically the β1 side of the sheet retained more structure than the β6 side, while the two native salt bridges (K3-E21 and K30-E100) were retained, holding together the bottom of the sheet. Simulation 498-4 was the exception, lacking the K3-E21 salt bridge, and as a result the N-terminus swings away from the rest of sheet A.

Other regions were more consistent. The β5–β6 edge opened, with β5 pulling away from sheet B, resulting in an average increase of 30% in solvent accessible surface area (SASA) of the protein, corresponding to a β Tanford value of βT ≈ 0.7. This value compares favorably with the experimental value of 0.67 for the apo C6A/C611A/F50E/G51E (pseudowild-type, or pWT) monomer (Nordlund and Oliveberg, 2006). pWT-SOD1 has been used in several studies (Lindberg et al., 2004, 2005; Byström et al., 2010; Hwang

Fig. 4 Transition state structures. (A) Representative structures from each transition state ensemble. Transition states were the following times: 156–161 ps (498-1), 108–113 ps (498-3), 103–108 ps (498-5), 207–212 ps (498-2) and 218–223 (498-4). (B) Comparison of average S-value and φ-value. Average S-value was calculated by equally weighting the S-value for each simulation. (C) Average S-value and φ-value colored onto the transition state structure from 498-1. Residues with values from 0 to 0.3 were colored red and considered unstructured, those with values 0.3–0.6 were colored magenta and considered moderately structured, and those with values >0.6 were colored blue and considered highly structured. Correlations plotted are relative to the experimental phi-values.
et al., 2010) to probe apo monomeric SOD1 folding/unfolding behavior because it yields a stable monomer. We ran five additional simulations of this pWT monomer to ensure it behaves the same as WT, and we observed the same steps of unfolding and a similar TS ensemble, including the same $\beta_T$ value.

Another similarity between the various simulations was the highly structured region formed by sheet B strands $\beta_4$, $\beta_7$ and $\beta_8$. Nearly all native contacts were maintained between these strands (Fig. 5), leading to a native-like sheet in the TS. However, $\beta_4$ was shortened compared to the starting structure. This was due to $\beta_5$ pulling away and losing contact with $\beta_4$.

Structure indices ($S$-values) were calculated for every residue over each TS ensemble. These serve as a semi-quantitative reflection of structure on a residue-by-residue basis and can be compared with protein engineering generated $\Phi$-values. While there have not been $\Phi$-values generated for wild-type SOD1, they have been determined for pWT-SOD1 (Nordlund and Oliveberg, 2006). The correlation between $S$-values from individual WT simulations and the experimental $\Phi$-values ranged from 0.61 to 0.68. When the TS ensembles were pooled to improve sampling, the correlation improved to 0.73. The TS ensembles from the dominant pathway agreed better with experiment ($R = 0.74$) than did those from the alternate pathway ($R = 0.68$). The pWT controls were also in good agreement with experiment, with $R = 0.71$ for the five pooled TS ensembles.

The most structured region of the TS was $\beta_7$, with $S > 0.60$ for residues 117–120 (Fig. 4B). Other high $S$-values in $\beta_1$, $\beta_2$ and $\beta_8$ formed an open ring-like structure around the $\beta_1$–$\beta_8$ edge through the $C$-terminus side of $\beta_8$ and through the middle of $\beta_1$ and $\beta_2$ (Fig. 4C). Residues 44–48 of $\beta_4$, which formed contacts with $\beta_7$ in the starting structure, had slightly lower $S$-values and averaged 0.47 ($\pm 0.08$). The $\beta_5$–$\beta_6$ edge and the loops had the lowest $S$-values.

In one case (V119) the WT $S$-values disagree greatly with the experimental $\Phi$-values (Nordlund and Oliveberg, 2006) (Fig. 4B). In fact, when V119 is omitted the correlation between the $S$- and $\Phi$-values increases to 0.84 for WT and 0.81 for pWT. For V119 the $S$-value suggests the mutant protein is highly structured at this site, while the $\Phi$-value suggests it is only moderately structured. Our simulations showed this residue, toward the end of $\beta_7$, was in a region that was highly structured in all the TS structures. In addition, both the $S$- and $\Phi$-values of nearby L117 (also within $\beta_7$) are high, suggesting that this region, which directly contacts V119, is highly structured in the TS.

Our folding nucleus, as defined by $S$-values, agrees well with the experimentally determined nucleus (Nordlund and Oliveberg, 2006), except for the differences noted above. Residues from four strands ($\beta_1$, $\beta_2$, $\beta_4$ and $\beta_7$) are prominent in both nuclei, but the MD folding nucleus adds the C-terminal half of $\beta_8$ (residues 147, 148, 150 and 151) and does not include $\beta_3$. Disagreement over the inclusion of the C-terminal half of $\beta_8$ arises because in the experimental study the structure of $\beta_8$ was inferred from a single residue: I149. Low $\Phi$- and $S$-values were found for this residue. However, the surrounding residues (148, 150 and 151) had an average $S$-value of 0.59 ($\pm 0.05$) and formed a structured region around the $\beta_1$–$\beta_8$ edge. Strand $\beta_3$ (residues 29–36) was not included in our folding nucleus because it was less structured than nearby strands $\beta_1$ and $\beta_2$ (Fig. 4B and C). The average $S$-value in $\beta_3$ (0.37 ± 0.10) was less than in $\beta_1$ (0.54 ± 0.20) and $\beta_2$ (0.48 ± 0.14). Also the lowest values in $\beta_3$ (residues 32–34) were aligned with the highest structured residues in $\beta_1$ and $\beta_2$, contrary to the expectation if these strands formed a continuous structure.

These minor differences highlight the advantage of combining experimental and theoretical studies. Experiments are...
needed to validate theory, and theory can fill in the details of experiments. Illustrating this point is the agreement of Φ and S at I149, but different conclusions as to the inclusion of β8 in the folding nucleus. Based on high neighboring S-values, we included part of β8 in the nucleus of the TS—data points that were unavailable by experiment.

Post-transition state unfolding events

After the TS, unfolding progressed with continued loss of structure in sheet A. In 498-1, side chain contacts were lost early between β3–β6 and β1–β2 (Fig. 2). By 210 ps, 50 ps after starting side chain loss, β3–β6 had lost all side chain contacts other than a persistent salt bridge, K30-E100. This salt bridge lasted until 1.6 ns, past the unfolding of all strands. The order differed in 498-5. The first β3–β6 hydrophobic side chain contact was lost before the TS, but all other side chain contacts persisted longer than the β1–β2 and β2–β3 contacts.

Unfolding of sheet B and the β1–β8 edge started with main chain hydrogen bond loss, and by 280 ps, the interstrand hydrogen bonds between β4 and β7 were completely lost (Fig. 2). Shortly afterwards, at 300 ps, first β4–β7 side chain interactions were lost (Fig. 2). In the other unfolding simulations on the dominant pathway (498-3 and 498-5), the order of these events differed. Both simulations began to lose β1–β8 main chain hydrogen bonds before sheet B did, but hydrophobic side chain residues remained in contact. Also in 498–3, sheet B main chain hydrogen bonds were broken much later (900 ps) but were lost quickly and cooperatively.

The next step of unfolding in the dominant pathway was the movement of loop IV away from sheet B and towards β5. During 300–370 ps, the loop pivoted around a persistent salt bridge (E40-R115) near β4 and β7 and broke its hydrophobic side chain contacts with β8. This movement is illustrated in the 340 ps structure of 498-1 (Fig. 3A).

Subsequent unfolding events similarly showed the β5–β6 edge region losing structure first, then sheet A, and finally sheet B and the edge region β1–β8. By 410 ps, the β5–β6 edge had lost hydrophobic side chain contacts, which is illustrated by the opening of the β5–β6 turn in the 560 ps structure in Fig. 3A. Also shown in this structure is the loss of the remaining main chain hydrogen bonds between β1 and β2 (previously lost at 550 ps). Sheet B strands β7–β8 had lost some hydrogen bonds, yet side chains remained in contact.

The structure at 710 ps (Fig. 3A) shows the loss of remaining sheet A side chain contacts, yet sheet B maintained most side chain contacts. Main chain hydrogen bonds between β1 and β8 had been lost at 620 ps (Fig. 2), but side-chain contacts kept this edge together. In simulation 498-1 there was a greater prevalence of non-native three-residue helices throughout the protein in the later stages of unfolding, beginning around 600 ps and forming transiently thereafter (Supplementary data, Fig. S1).

β4, β7 and β8 were the last strands to lose side-chain contacts in sheet B, as well as the β1–β8 edge. The 810 ps structure in Fig. 3A shows the partial loss of side chain contacts between sheet B strands, but they remained in contact in the core of the structure. Over time, interactions between F45 and B7 residues increase as illustrated in the 810 ps structure. Similarly, L8-A145 holds β1 and β8 together near the top of the strands because other hydrophobic side chain contacts were lost.

As sheet B lost side chain contacts and began to separate, the ES loop expanded from inside the loop and progressed towards the opening, corresponding to loss of hydrophobic side chain contacts. The 810 ps structure in Fig. 3A shows late stages of ES loop opening.

Summary of the early unfolding events in the dominant unfolding pathway

The early unfolding events affected the protein in a consistent order. First the β5–β6 edge, including neighboring strands β4 and β3, was disrupted. This was followed by slow unfolding of sheet A strands, with β3 separating first. Then, sheet B and the β1–β8 edge were lost, generally beginning with β4 separating from the other strands. Finally, as strands β7 and β8 separated, the ES loop opened.

Native state flexibility analysis predicted the early unfolding events reasonably well. A previous large-scale study comparing native state flexibility with early unfolding events showed that the most flexible regions are the first to lose structure during thermal unfolding (Benson and Daggett, 2008), a finding that is supported by the SOD1 data presented here. The most flexible strand was β6, which is part of the β5–β6 edge that lost structure first. Sheet A strands were the next most flexible, decreasing in order from β3 to β1. Sheet B strands β4, β7 and β8 had very low flexibility (Fig. 1C) and similarly were the last to be lost. However, β5 was much less flexible than would be expected by the early loss of structure in this strand, illustrating that although flexibility is a reasonable predictor of early unfolding events, certain regions move differently during unfolding than in native state.

Variations in the earliest unfolding events

Two simulations, 498-2 and 498-4, unfolded via an alternate pathway. The differences are shown in the structures given in Fig. 3B, from 498-4. Both alternate simulations lost the β1–β8 edge very early. The 100 ps structure in Fig. 3B shows this edge fully separated while the β5–β6 edge remained mostly intact. Moving to the TS structure at 218 ps (Fig. 3B), loss of the β5–β6 edge caused sheets A and B to separate. Compared with the TS structures from the dominant pathway (Fig. 4A), the major difference was the alternate pathway’s two-sheet structure versus the dominant pathway’s half-barrel (around the β1–β8 edge). However, aside from the β1–β8 edge, the unfolding events were similar.

Particular to 498-4 was the jutting out of the N-terminus from sheet A (Fig. 4A). By the TS, the K3-E21 salt bridge near the N-terminus between β1 and β2 was broken allowing the N-terminus to swing freely. Following the TS, β1 separated from sheet A (Fig. 3B). For comparison, β3 separated from the front sheet first in the dominant unfolding pathway (Fig. 3A). This difference was only seen in 498-4.

The denatured state

The denatured state was reached after further unfolding. The denatured state was defined as the ensemble of structures from 1.5 to 33 ns of the two longer simulations (498-1 and 498-2). This definition was based on the stabilization of properties after 1.5 ns, including SASA, total contacts and radius of gyration. Interestingly, the Co-RMSD in 498-2 continued to increase over 6–16 ns, from 13.4 to 18.5 Å, although the other properties remained stable. This increase was associated with the separation of β1 and β8 as the N-terminus flipped away.
from the C-terminus. Similar movement was not seen in 498-1, as the N- and C-termini remained close.

Structures within the denatured state were heterogeneous between the two simulations and within the individual simulations. The Cα-RMSD was >10 Å between any structure in 498-1 compared with any structure in 498-2, and it was over 15 Å for most of the alignments. The major structural difference between the ensembles was the association of the N- and C-termini in 498-1 and the separation of the termini in 498-2 (Fig. 6A), as mentioned above. The snapshots in Fig. 6A also show the wide range of conformations adopted by each individual ensemble. A representative structure for the ensemble was selected by determining which structure had the lowest Cα-RMSD when compared to the rest of the ensemble. These intra-ensemble calculations gave 10.5 Å (± 2.1 Å) and 9.6 Å (± 2.3 Å) for 498-1 and 498-2, respectively. The slightly lower value for 498-2 was due to tighter clustering in the center of the structures (Fig. 6A).

Despite the large structural variation between and within ensembles, the global physical properties were similar (Table I). Based on the Cα radius of gyration, our denatured state was 24% expanded compared with the native state. The maximum spanning radius, an indicator of compactness, indicates that the denatured ensemble is 37% expanded relative to the native state, which is in good agreement with the experimentally determined 40% (Assfalg et al., 2003).

In both denatured ensembles, non-native salt bridges and helices were present and tended to be between similar regions of the protein. The propensities to form these interactions may explain the similarities in the properties, and these interactions are described in greater detail below.

The most persistent long-range interactions in the ensembles were electrostatic interactions. Interestingly, the regions that interacted in both independently derived denatured state ensembles involved the same charged residues. Trends were seen in the order of formation of these contacts. A network of salt bridges formed near the bottom of loop IV prior to reaching the denatured state and was present for the majority of the denatured state. In addition to the native interactions R69-E78 and R69-E77, a non-native salt bridge formed between

Fig. 6 Representative structures from the denatured ensemble for 498-1 and 498-2. (A) Six structures from each denatured ensemble are shown. Times represented are evenly spaced—5, 10, 15, 20, 25 and 30 ns. Structures are colored the same as in Fig. 1, and are aligned on the mean structures given in (B). (B) Closer look at important contacts shared in the mean structures for the runs.
K70-D76. These residues hold the lower portion of loop IV in a native-like kink (Fig. 6B). Contact maps show that besides these intra-loop interactions, there were few other contacts between this loop and the rest of the protein (Supplementary data, Fig. S2A).

The non-native E40-K122 salt bridge also formed prior to the denatured state (before 1.5 ns) and was present in the majority of the structures in the ensemble. These residues are before B4 (E40) and in the beginning of loop VII (K122) in the native structure, and the resultant salt bridge holds the B4 and B7 strands close. Other nearby, and likely influenced, strands included B3 and to a lesser extent B6. This non-native interaction helped stabilize the denatured state and bring the distant B4 and B7 strands in proximity during folding.

As noted above, the salt bridges in sheet A (K3-E21 and K30-E100) lasted into the denatured state. Generally these interactions were lost early in the denatured state, with K3-E21 lasting longer than K30-E100. Loss of K30-E100 resulted in B3 and B6 losing their sheet-like parallel structure, although these strands continued to occasionally interact as the non-native interaction between E40-K122 kept these residues near each other. The K3-E21 salt bridge persisted longer, likely helped by the proximity in sequence and the single turn separating B1 and B2. This interaction lasted much longer in 498-2 because of additional stability added by the non-native K3-D109 interaction, which was absent in 498-1.

In the denatured ensembles, a second network of salt bridges formed between the B5–B6 turn and ES loop (Fig. 6B). Residues D90-K128 and K91-D125 formed non-native interactions within 2 ns of each other. These interactions limited the movement of B5, by holding the lower residues of B5 relatively fixed. The loop IV network contributed some to the stability of B5 as the nearby strand B6, further away from this additional network, moved more.

Lastly, a fluctuating non-native salt bridge between E49-R143 formed for multiple nanosecond stretches in both ensembles. This contact brought the beginning portion of strand B8 close to the ending portion of B4. Also because of the non-native interaction, E40-K122, which in turn influenced B4 and B1 and B6 and B7, were close when E49-R143 was present. Like the E40-K122 non-native salt bridge, the E49-R143 interaction greatly shortened the distance between strands that are adjacent in the native structure but are distant in sequence. Interaction between B7 and B8 was not seen in 498-1, but when present in other simulations this interaction greatly shortened the ES loop.

Non-native helices formed throughout the denatured state. The longest lived helices were seen in similar regions, near B1, B4, B5 and B7. To understand the effects of these non-native helices, we compared the long-range contacts the helices made in the 498-1 and 498-2 simulations. While identical contacts were rare when comparing the denatured ensembles, the non-native helices that formed near B1 and B4 showed similar contact patterns.

A non-native helix formed near B1 for two time periods for each of the denatured ensembles. For both periods in 498-2 and the 4.5–8 ns span in 498-1, this helix contacted loop VI. In 498-2 these regions were also in contact for the majority of the denatured ensemble, as shown by a contact map (Supplementary data, Fig. S2A). Specifically, a non-native hydrogen bond between T2 and D109 formed early in the denatured state and its persistence accounted for the contact pattern observed in Supplementary data, Fig. S2. In 498-1 this region shifted to contact the ES loop from its early contacts with loop VI (Supplementary data, Fig. S2B), explaining the difference in contact patterns between the early (4.5–8 ns) and late (27–30.5 ns) spans in which the non-native helix was present. K3 was responsible for the observed shifts as it exchanged a salt bridge with D109 in loop VI for E133 in the ES loop. The K3-E133 salt bridge was then exchanged for the K3-D124 salt bridge, which lasted until the end of the simulation. Following the formation of this more stable interaction, hydrogen bonds and hydrophobic contacts formed between nearby residues (Supplementary data, Fig. S3A). These long-lasting interactions in both denatured ensembles suggest that the residues in the non-native helix near B1 form interactions that stabilize the positively charged region near the N-terminus throughout the denatured state.

The non-native helix near B4 interacted with B5 and B6 in both denatured ensembles. One or more salt bridges formed prior to helix formation, followed by the addition of nearby hydrophobic interactions during the time when the helix was present (Supplementary data, Fig. S3B). The major difference between the ensembles was the duration of the salt bridges. In 498-1, a single salt bridge was formed 2 ns before and lasted 2.5 ns longer than the helix. In contrast, the two salt bridges in 498-2 formed prior to the denatured state (well before the

---

**Table I. Properties of the SOD conformational ensembles**

<table>
<thead>
<tr>
<th></th>
<th>Starting structureb</th>
<th>MD native statec</th>
<th>Denatured state</th>
<th>Averagee</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>498-1d</td>
<td>498-2d</td>
</tr>
<tr>
<td>#Total contacts</td>
<td>591</td>
<td>565 ± 8.5</td>
<td>389 ± 24.0</td>
<td>393 ± 23.3</td>
</tr>
<tr>
<td>#Native contacts</td>
<td>590</td>
<td>491 ± 9.9</td>
<td>159 ± 7.7</td>
<td>168 ± 6.8</td>
</tr>
<tr>
<td>Cα-RMSD (Å)</td>
<td>0.27</td>
<td>2.58 ± 0.12</td>
<td>18.4 ± 0.90</td>
<td>16.3 ± 2.12</td>
</tr>
<tr>
<td>Total SASA (Å²)</td>
<td>8166 ± 137</td>
<td>11.66 ± 631</td>
<td>11.825 ± 565</td>
<td>11.744</td>
</tr>
<tr>
<td>Radius of gyration (Å)</td>
<td>14.2</td>
<td>14.5 ± 0.08</td>
<td>17.6 ± 0.72</td>
<td>18.5 ± 1.10</td>
</tr>
<tr>
<td>Maximum spanning Distance (Å)</td>
<td>43.0</td>
<td>42.6 ± 0.70</td>
<td>53.5 ± 3.65</td>
<td>63.4 ± 7.94</td>
</tr>
<tr>
<td>B-strand content</td>
<td>37%</td>
<td>32%</td>
<td>0.2%</td>
<td>0.2%</td>
</tr>
<tr>
<td>Helical content</td>
<td>5%</td>
<td>4%</td>
<td>6.4%</td>
<td>10.0%</td>
</tr>
</tbody>
</table>

---

*aThe denatured state ensemble is defined as 1.5 – 33 ns of 498-1 and 498-2.*

*bThe average starting values from the 498 runs, 0 ns time points. These time points come after minimization and some changes to the structure have already occurred.*

*cThe values shown are the average and standard deviation over the native state ensemble (1.5–21 ns of 298-1).*

*dThe values shown are the average and standard deviation over the denatured state ensembles.*

*eThe average of the two denatured state ensembles.*

---

Unfolding of SODI
formation of the non-native helix) and lasted for most of the denatured ensemble. We note that this helix lasts much longer in 498-2 and may be related to salt bridge stability. Contacts made by the two other non-native helices (near β5 and β7) generally differed between the ensembles. As noted previously, the major difference between the ensembles was the relative position of the N- and C-termini. Structures in 498-2 were elongated compared to 498-1 (Fig. 6), contributing to the differences in contact patterns. Some similarity in contacts was seen; however, these were primarily in regions that were affected by the non-native salt bridges described previously.

Conclusions

Molecular dynamics simulations of the thermal denaturation of SOD1 protein show many similarities across the unfolding trajectories. Three of the five simulations unfold along a similar path, referred to as the dominant pathway, while the other two unfold along an alternate path. The alternate pathway still has commonalities with the dominant pathway. These unfolding pathways support the case for instability of the β5–β6 edge of the protein, which is normally protected from improper monomeric β-strand to β-strand association by a characteristic bend in these edge strands (Richardson and Richardson, 2002). The high temperature employed here provides information about the denatured state as well, and we find good agreement between the high temperature TS ensembles generated here and experimental data (Nordlund and Oliveberg, 2006).

The early stages of unfolding in these high temperature simulations are very similar to structural alterations in the protein induced by the lethal A4V mutation in simulations at physiological temperature (Schmidlin et al., 2009); however, the unfolding does not proceed to the TS in the mutant simulations (Fig. 7).