



Local Unfolding of Cu, Zn Superoxide Dismutase Monomer Determines the Morphology of Fibrillar Aggregates

Feng Ding¹, Yoshiaki Furukawa², Nobuyuki Nukina³
and Nikolay V. Dokholyan^{1*}

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

²Department of Chemistry, Keio University, Yokohama, Kanagawa, Japan

³Laboratory for Structural Neuropathology, RIKEN Brain Science Institute, Wako, Saitama, Japan

Received 19 November 2011;
accepted 14 December 2011

Edited by D. Case

Keywords:

SOD1 misfolding and aggregation;
fibrillar aggregate;
aggregation building block;
molecular dynamics;
multiscale modeling

Aggregation of Cu, Zn superoxide dismutase (SOD1) is often found in amyotrophic lateral sclerosis patients. The fibrillar aggregates formed by wild type and various disease-associated mutants have recently been found to have distinct cores and morphologies. Previous computational and experimental studies of wild-type SOD1 suggest that the apo-monomer, highly aggregation prone, displays substantial local unfolding dynamics. The residual folded structure of locally unfolded apoSOD1 corresponds to peptide segments forming the aggregation core as identified by a combination of proteolysis and mass spectroscopy. Therefore, we hypothesize that the destabilization of apoSOD1 caused by various mutations leads to distinct local unfolding dynamics. The partially unfolded structure, exposing the hydrophobic core and backbone hydrogen bond donors and acceptors, is prone to aggregate. The peptide segments in the residual folded structures form the “building block” for aggregation, which in turn determines the morphology of the aggregates. To test this hypothesis, we apply a multiscale simulation approach to study the aggregation of three typical SOD1 variants: wild type, G37R, and I149T. Each of these SOD1 variants has distinct peptide segments forming the core structure and features different aggregate morphologies. We perform atomistic molecular dynamics simulations to study the conformational dynamics of apoSOD1 monomer and coarse-grained molecular dynamics simulations to study the aggregation of partially unfolded SOD1 monomers. Our computational studies of monomer local unfolding and the aggregation of different SOD1 variants are consistent with experiments, supporting the hypothesis of the formation of aggregation “building blocks” via apo-monomer local unfolding as the mechanism of SOD1 fibrillar aggregation.

© 2011 Elsevier Ltd. All rights reserved.

Introduction

The misfolding and aggregation of Cu, Zn superoxide dismutase (SOD1) is associated with amyotrophic lateral sclerosis (ALS).^{1,2} More than 100 mutations in SOD1 have been identified in familial ALS patients. Both wild-type and mutant

*Corresponding author. E-mail address: dokh@unc.edu.

Abbreviations used: SOD1, superoxide dismutase; ALS, amyotrophic lateral sclerosis; DMD, discrete molecular dynamics; WHAM, weighted histogram analysis method; RMSF, root-mean-square fluctuation.

SOD1 can form insoluble fibrillar aggregates with a common cross- β amyloid structure,^{3,4} as observed in many other amyloidogenic proteins with distinct primary, secondary, and tertiary structures.⁵ Many of the research efforts in SOD1 misfolding have been focused on finding a general mechanism for how mutations promote SOD1 misfolding and aggregation, under the assumption of a common “mutation-independent” aggregation pathway and similar aggregate structures.⁶ However, a recent study of SOD1 aggregates formed by wild type and various mutants revealed distinct fibrillar core compositions and aggregate morphologies.⁷ Accordingly, phenotypic heterogeneity has been reported in familial ALS patients with different SOD1 mutations.⁸ Increasing evidence suggests that the structures and morphologies of protein aggregates affect their respective disease phenotypes and that polymorphism in protein aggregates associates with phenotypic heterogeneity.^{9–12} Hence, uncovering the molecular mechanism governing the formation of polymorphic amyloid aggregates is important for gaining an understanding of ALS phenotypic heterogeneity.

SOD1 forms a stable dimer in solution, with each SOD1 monomer binding one copper and one zinc ion and forming one intra-monomer disulfide bond. Various biochemical and biophysical studies have suggested that wild-type SOD1 dimer is exceptionally stable because of the coordination of metal ions.¹³ Mounting experimental and computational evidence suggest that apoSOD1 monomer is the most aggregation prone species.^{14–19} The loss of the coordinated metal ions destabilizes the protein with a significant population unfolded at physiological conditions.¹⁴ Although apoSOD1 is native like in the crystal structure,²⁰ the protein in solution features significant structural disorder and conformational flexibility.²¹ Both experimental^{22,23} and computational¹⁹ studies suggest that the apoSOD1 monomer features frequent local unfolding. In addition to the two long loops, the metal-coordinated (in the native state) strands 4, 5, and 7 feature a high level of local unfolding, and the N-terminal β -sheet is the most stable structural element (Fig. 1a). Interestingly, the same regions that are stable in the locally unfolded apoSOD1 also correspond to those regions that participate in the fibrillar core of wild-type SOD1 aggregates, having been identified as proteolysis-resistant peptides.⁷ This observation is consistent with the generic aggregation mechanism proposed earlier,^{24,25} where the residual structural elements in the partially unfolded protein interact with each other and serve as “building blocks” for the formation of fibrillar amyloid aggregates. Recently, local unfolding induced by mutations has also been found to play an important role in the aggregation of γ -crystallin in human cataracts.²⁶ Therefore, we hypothesize that the

various disease-causative mutations in SOD1 have different impacts on apoSOD1 conformational dynamics, which in turn lead to distinct patterns of local unfolding and thus the varied morphologies of the resulted aggregates.

To test this hypothesis, we apply a multiscale molecular dynamics approach to study the local unfolding of SOD1 monomer and the aggregation dynamics of multiple monomers. In the previous experimental study of SOD1 aggregates, Furukawa *et al.* discovered that three major regions comprise the fibril aggregate cores, including the N-terminal β -sheet (strands 1–3), the middle strand 6, and the C-terminal strand 8 (Fig. 1b). The N-terminal strands are observed in the aggregates formed by all SOD1 variants. As a result, there are only three possible combinations of core-forming peptide patterns: all three segments, the N-terminal sheet plus the middle strand, and the N-terminal sheet plus the C-terminal strand. Therefore, beside the wild-type SOD1 whose aggregation core is composed of all three segments, we also include two mutants, G37R and I149T, having representative aggregation core compositions.⁷ In the core of G37R and I149T, the C-terminal strand and the middle strand 6 are observed, respectively, in addition to the N-terminal strands (Fig. 1b). In order to probe the monomer conformational dynamics, we perform atomistic discrete molecular dynamics (DMD) simulations.¹⁹ In the atomistic simulations, we find that different mutations indeed result in different patterns of local unfolding and that the residual structures in the locally unfolded states are consistent with the core-forming segments, supporting the “building block” aggregation mechanism.²⁴ We further perform coarse-grained DMD simulations to study SOD1 monomer aggregation for each of the three SOD1 variants. To promote the formation of partially unfolded structures, we develop a hybrid structure-based interaction model where the interactions between the core-forming residues are enhanced. The reconstructed model structure of the amyloid aggregates is consistent with the experimentally observed morphology. Therefore, our multiscale simulations suggest a molecular mechanism of mutation-dependent SOD1 aggregation polymorphism.

Results and Discussion

We use a multiscale molecular dynamics approach to study the misfolding and aggregation of apoSOD1. We use the atomistic DMD simulations²⁷ to sample the conformational dynamics of the apoSOD1 monomer, the timescale of which is approximately microseconds to milliseconds. For monomer aggregation, which spans hours and days, 172

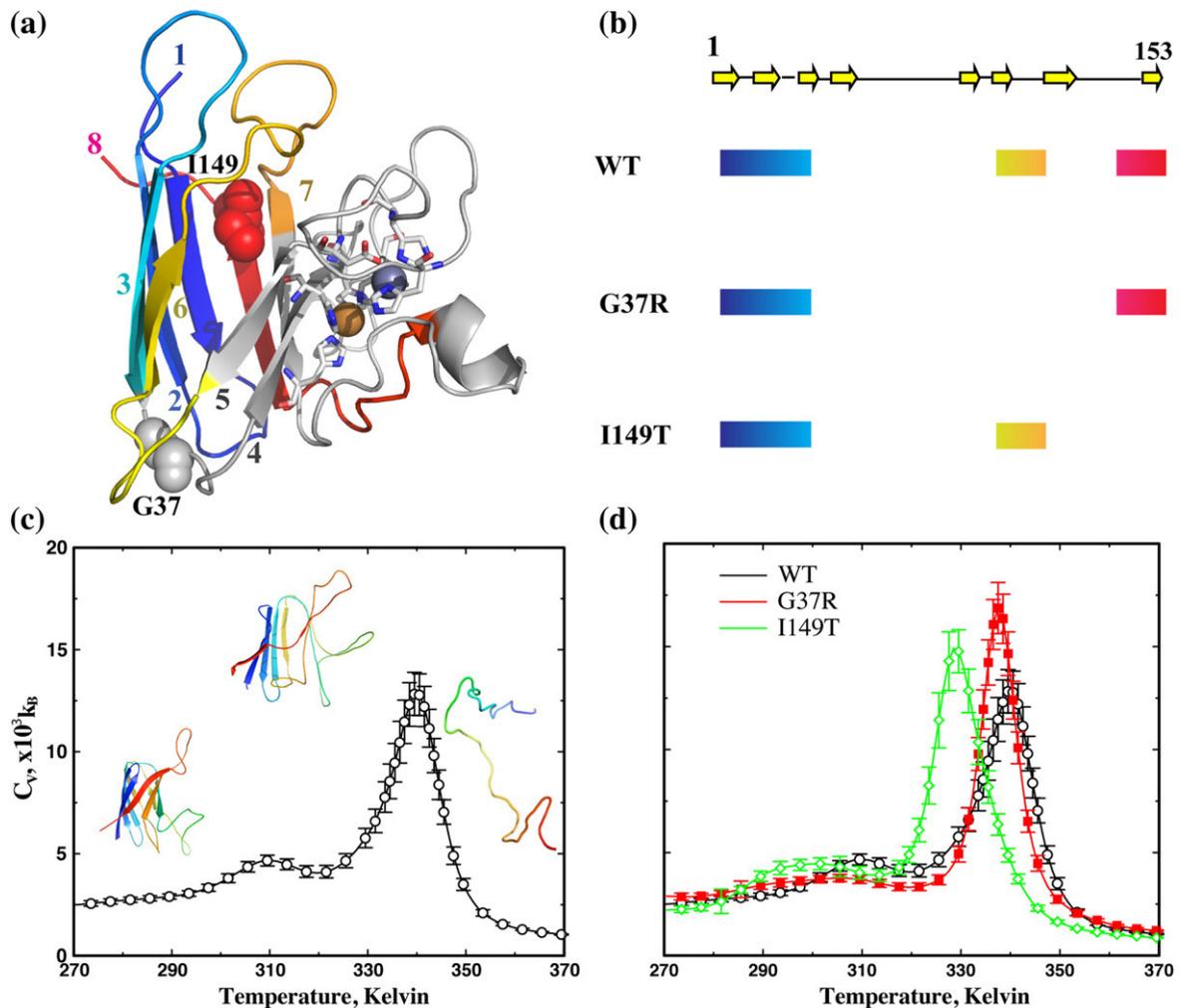


Fig. 1. The thermodynamics of apoSOD1 monomer. (a) In SOD1 monomer, the two metal ions (spheres) are coordinated by residues (sticks) in strands 4, 5, and 7. The amyloid core of wild-type (WT) SOD1 is composed of the N-terminal strands 1–3 (blue to cyan), the middle strand 6 (yellow), and the C-terminal strand 8 (red). Other regions are colored gray. The two mutated residues in this study, G37 and I149, are shown in sphere. (b) The regions forming the amyloid core of three typical SOD1 variants: WT, G37R, and I149T. The arrows illustrate the strands along the primary sequence. The thick lines highlight the regions found in the proteolysis-resistant amyloid core for each of the three SOD1 variants. (c) The specific heat of the WT apoSOD1 as a function of temperature. The specific heat values and corresponding statistical uncertainties (shown as error bars) are computed using WHAM analysis of the replica exchange simulation trajectories (Materials and Methods). At temperatures below the low-temperature peak, the protein is native like with the β -barrel intact. At temperatures higher than the high-temperature peak, the protein is unfolded, without persistent secondary and tertiary structures. At the intermediate temperature between the two peaks, the protein is partially unfolded. (d) The specific heat curves for all three variants feature similar local unfolding and global unfolding thermodynamics.

173 we use coarse-grained DMD simulations²⁴ with
 174 experimental constraints to enhance sampling of
 175 protein aggregation.

176 Conformational dynamics of apoSOD1 177 monomer

178 We perform all-atom DMD simulations²⁷ to
 179 sample the conformational dynamics of apoSOD1
 180 monomers. The protein is in united-atom represen-

tation, with all heavy atoms and polar hydrogen
 181 explicitly modeled. The simulation is performed
 182 with implicit solvent, and interatomic interactions
 183 are modeled by a physical force field adapted from
 184 Medusa,²⁸ which includes van der Waals,
 185 solvation,²⁹ and hydrogen bonding potentials. In
 186 addition to the previous version of the all-atom DMD
 187 force field,²⁷ we also introduce screened electrostatic
 188 interactions between charged residues (Materials
 189 and Methods). Atomistic DMD simulations allow
 190

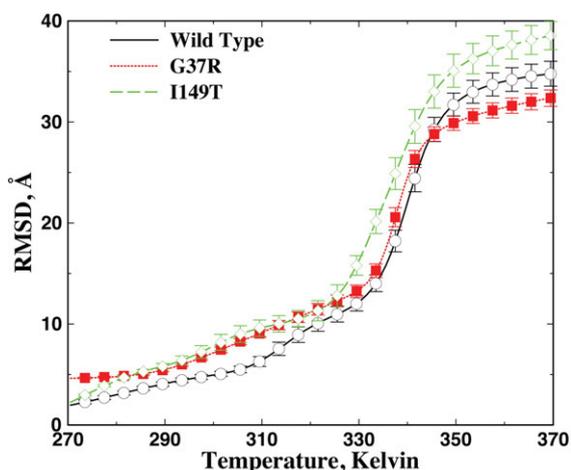


Fig. 2. Average RMSD as a function of temperature. Average RMSD is computed using WHAM analysis. Based on the self-consistently determined density of states $\rho(E)$ and the conditional probability to observe a structure with an RMSD of R at the given energy E , $P(R|E)$, the average RMSD and corresponding error bar can be computed accordingly (Materials and Methods).

the density of states is self-consistently determined from overlapping energy histograms for the replica exchange simulations.³⁴ To reduce structural relaxation artifacts from the starting crystallography structure, we exclude the first quarter of each simulation trajectory from the WHAM analysis. As a simple approach to test convergence of the simulations, we divide the rest of the simulations into halves, and then compute and compare the histogram of potential energies for each of the half. We find that the histograms from both halves are close to each other, suggesting the convergence of the simulations (Fig. S1). Given the density of states, thermodynamic parameters such as specific heat and average root-mean-square deviation (RMSD) from the starting crystal structure can be calculated (details in Materials and Methods).

The specific heat of wild-type apoSOD1 monomer features a major peak at $T_f \sim 330$ K (Fig. 1c), which corresponds to the global unfolding or melting transition of the protein. Above the transition temperature $T > T_f$, the protein is random coil like with large RMSD (Fig. 2). At lower temperatures, the protein is folded or at least partially folded with lower RMSD. Interestingly, a minor peak appears at a lower temperature, $T_p \sim 310$ K, which corresponds to a partial unfolding transition. Such a weak structural transition has also been observed in earlier computational studies, although the global unfolding transition was the main focus of these studies, using the melting temperature as the measure of thermal stability.¹⁹ In an earlier work by Zhou and Karplus, a similar transition was termed surface-molten to solid transition.³⁵ The lower value of the local unfolding temperature as compared to the global unfolding temperature suggests that the apoprotein undergoes significant local unfolding before it becomes globally unfolded (Fig. 2), which is consistent with the NMR study of the wild-type apoSOD1 monomer in solution.²¹

We similarly compute the specific heat of the two SOD1 mutants (Fig. 1d). Both mutants have a major and a minor peak in the specific heat, similar to that of the wild type. Because of the vast conformational space of SOD1, we do not expect DMD simulations to reach folding/unfolding equilibrium as observed in simulations of small fast-folding proteins.²⁷ For example, the melting temperature of wild-type apoSOD1 monomer in DMD simulations (~ 330 K) is higher than that measured in experiments (~ 315 K).³⁶ However, starting from the crystal structure, we expect that our microsecond-long replica exchange simulations would be able to sufficiently sample conformations in the folded and partially folded states (e.g., the convergence of simulations illustrated in Fig. S1 and Fig. 2), and the computed unfolding temperatures can be used as a qualitative measure of a protein's thermostability. Compared to the wild type, the I149T mutant has a

efficient sampling of the large-scale conformational dynamics of proteins and protein complexes.^{19,30,31}

In order to efficiently sample the conformational dynamics of apoSOD1, we utilize replica exchange³² DMD simulations, which allow enhanced sampling of the conformational space. We allocate 12 replicas for each SOD1 variant, with each replica running at different temperatures from low to high (Materials and Methods). Periodically, replicas with neighboring temperature values exchange their simulation temperatures stochastically. A temporarily trapped state in a replica can be rescued by simulating at a higher temperature, thereby enhancing the sampling efficiency of DMD simulations. For each simulation, we start from the crystal structure conformation, with metal ions and the disulfide bond removed. The structures for wild type (Protein Data Bank ID: 1SPD) and G37R (Protein Data Bank ID: 1AZV) are known. The starting structure of I149T is modeled by amino acid substitution and rotamer optimization using Eris.³³ In the simulations, cysteine residues do not form disulfide bond mimicking a reducing condition. The total simulation time of each replica is 100 ns. Therefore, we perform a total of 1.2 μ s of DMD simulations for each SOD1 variant (Materials and Methods).

ApoSOD1 features frequent local unfolding before global unfolding

Based on replica exchange simulations, we use the weighted histogram analysis method (WHAM; Materials and Methods) to compute the thermodynamics of apoSOD1 unfolding. In WHAM analysis,

282 lower melting temperature and thus a weaker
 283 thermal stability, while G37R has a similar or
 284 slightly higher melting temperature. Interestingly,
 285 the two mutants have lower local unfolding
 286 temperatures (the low-temperature peak in Fig. 1d)
 287 than that of the wild type, suggesting that mutations
 288 enhance the local unfolding of apoSOD1.

289 *Mutations lead to different local unfolding dynamics*

290 In order to characterize the conformational dyn-
 291 amics of different SOD1 variants in the partially
 292 unfolded states, we reconstruct the conformational
 293 ensemble from the replica exchange simulation
 294 trajectories (see *Materials and Methods*) based on
 295 the RMSD ranges identified from WHAM calcula-
 296 tions (Fig. 2). Given the partially unfolded structural
 297 ensemble for each SOD1 variant, we compute the

298 root-mean-square fluctuation (RMSF) of each residue
 299 around its average positions (Fig. 3a and c). A
 300 larger RMSF value denotes higher conformational
 301 flexibility of the residue in the locally unfolded state.
 302 We also compute the average RMSD of each residue
 303 from the corresponding native structure (Fig. 3b and
 304 d). In the RMSD per residue calculation, we exclude
 305 the two major loops, residues 51–80 and 121–141,
 306 which are coordinated by metal ions in the crystal
 307 structure and are highly disordered without metals
 308 bound. For the wild type, the three N-terminal
 309 strands, the middle strand near residue 100, and the
 310 C-terminal strand all feature low RMSD and RMSF
 311 values. These results are consistent with previous
 312 experimental²² and computational^{6,19} studies of
 313 local unfolding in the apoSOD1 monomer. These
 314 same segments were also found to participate in the
 315 fibrillar core by wild-type apoSOD1.⁷

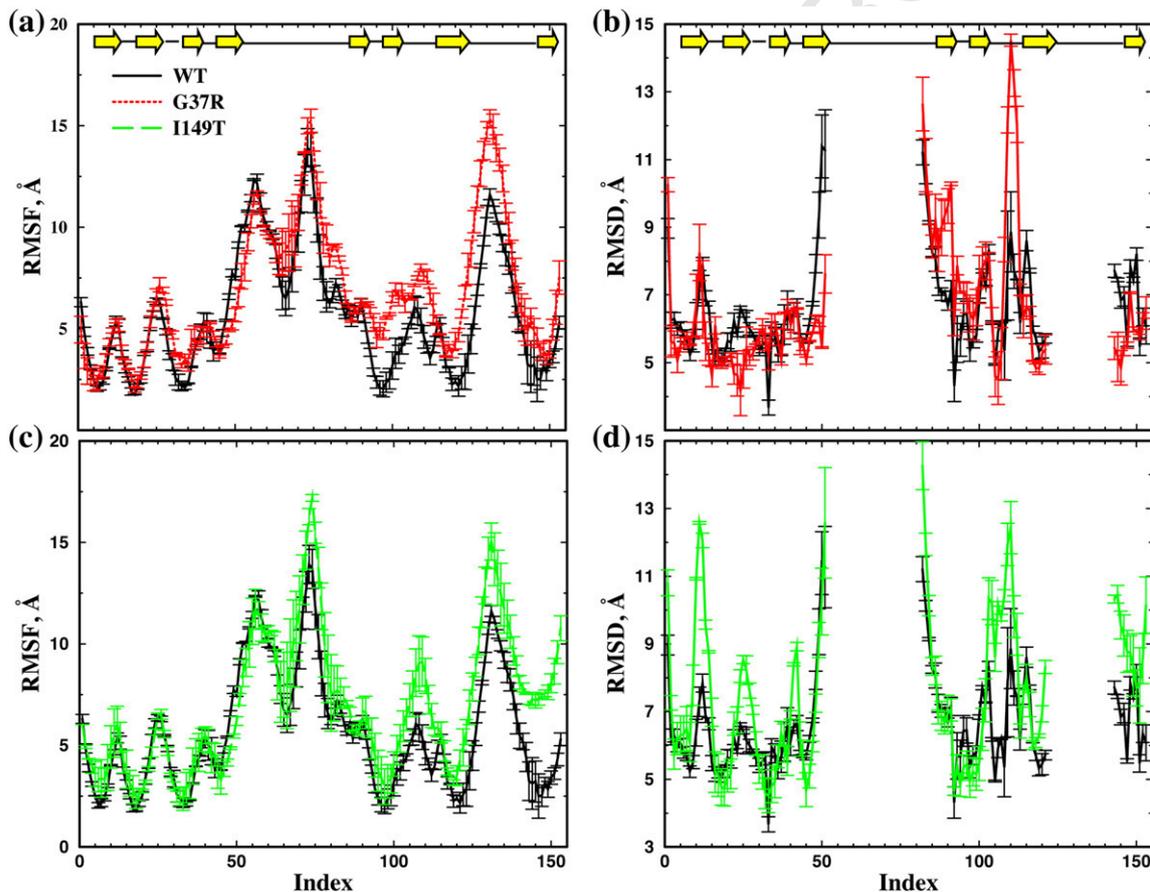


Fig. 3. The local unfolding conformational dynamics of apoSOD1 monomer. The locally unfolded conformational ensemble for each SOD1 variant is reconstructed from the simulation trajectories (*Materials and Methods*). The corresponding conformational dynamic parameters, including RMSF (a and c) and RMSD (b and d) per residue, are computed based on the reconstructed structural ensemble. In order to estimate the error bars, we split the simulation trajectory into halves and compute the dynamic parameters independently. To avoid overlapping lines, we compare the RMSF and RMSD between WT and G37R (a and c) and between WT and I149T (b and d) separately. The average RMSD per residue is computed with respect to the native states. Due to their large conformational flexibility, we do not include the loops in the RMSD calculation.

316 G37R displays larger RMSD and RMSF in the
 317 middle strand near residue 100 compared to the
 318 wild type, while the N- and C-terminal strands of
 319 G37R have low RMSD and RMSF similar to that

of the wild type (Fig. 3a and b). In the I149T
 320 mutant, the C-terminal strand has larger RMSF
 321 and RMSD values than those of the wild type.
 322 The N-terminal and middle strands of I149T
 323

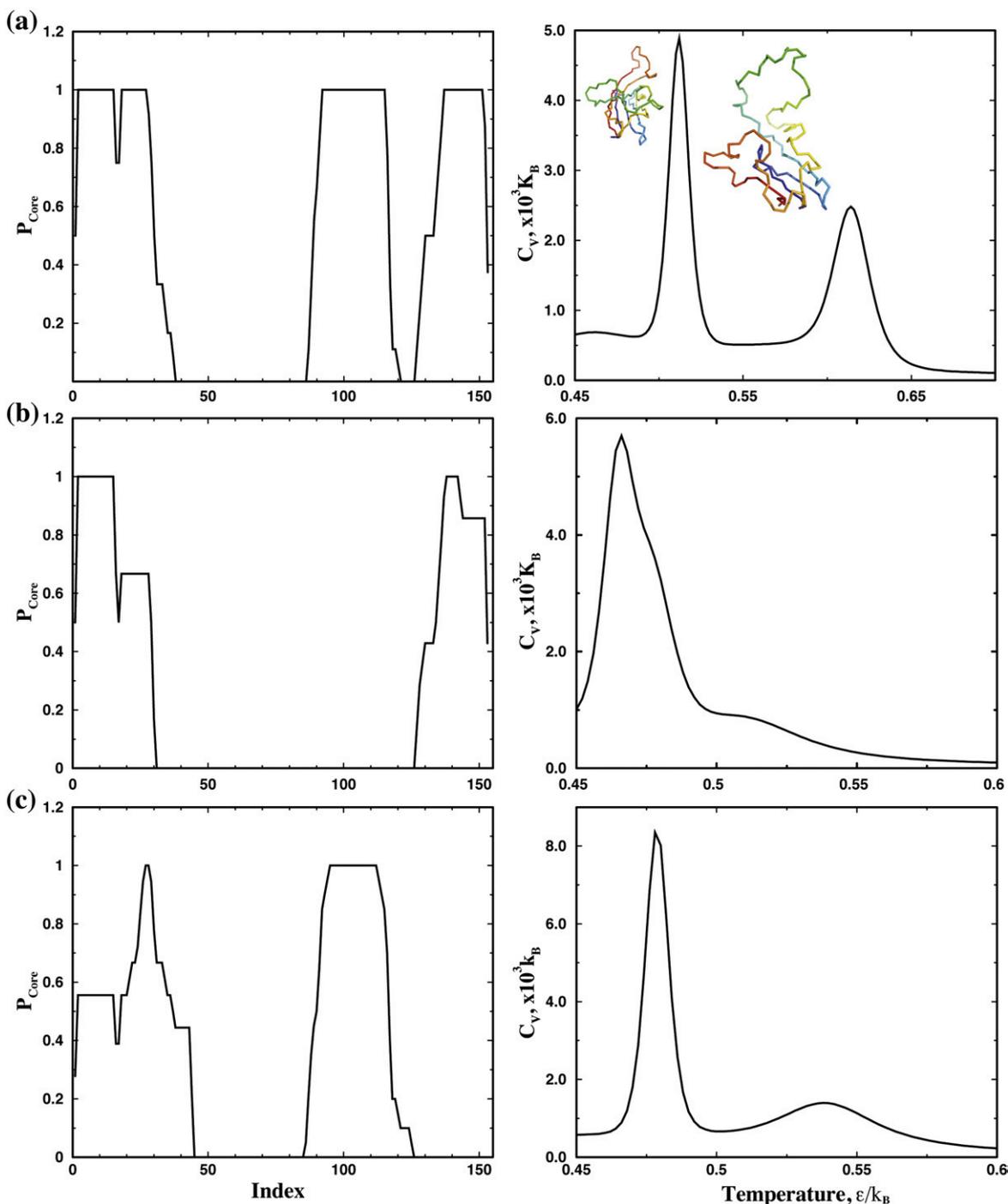


Fig. 4. Thermodynamics of coarse-grained SOD1 monomer: (a) WT, (B) G37R, and (c) I149T. On the left column, the probability of each residue to form the amyloid core (P_{Core}) is computed from the experimentally identified, proteolysis-resistant peptides in the amyloid core (Materials and Methods). On the right column, the specific heat of the coarse-grained SOD1 monomer is computed from replica exchange simulations of monomer folding. Due to the experimentally based bias potential, the protein features a stable partially folded intermediate state.

324 feature low RMSD and RMSF as is found in the
 325 wild type, although the RMSD of the turns in
 326 I149T is large (Fig. 3c and d). These regions with
 327 low RMSD and RMSF values in mutant SOD1
 328 correspond to the core-forming peptides in the
 329 respective aggregates⁷ (Fig. 1). Therefore, there is
 330 a correlation between the computationally identi-
 331 fied stable, native-like regions in the locally
 332 unfolded apoSOD1 monomer and the proteoly-

333 sis-resistant peptide segments in the correspond-
 334 ing amyloid core. Taken together, these results
 335 suggest that mutations affect the local unfolding
 336 dynamics and result in different patterns of local
 337 unfolding. The peptide segments in the well-
 338 defined residual structure in the locally unfolded
 339 apoSOD1 can serve as a “building block” for fibril
 340 aggregates, which interact with each other to form
 341 the fibrillar core.

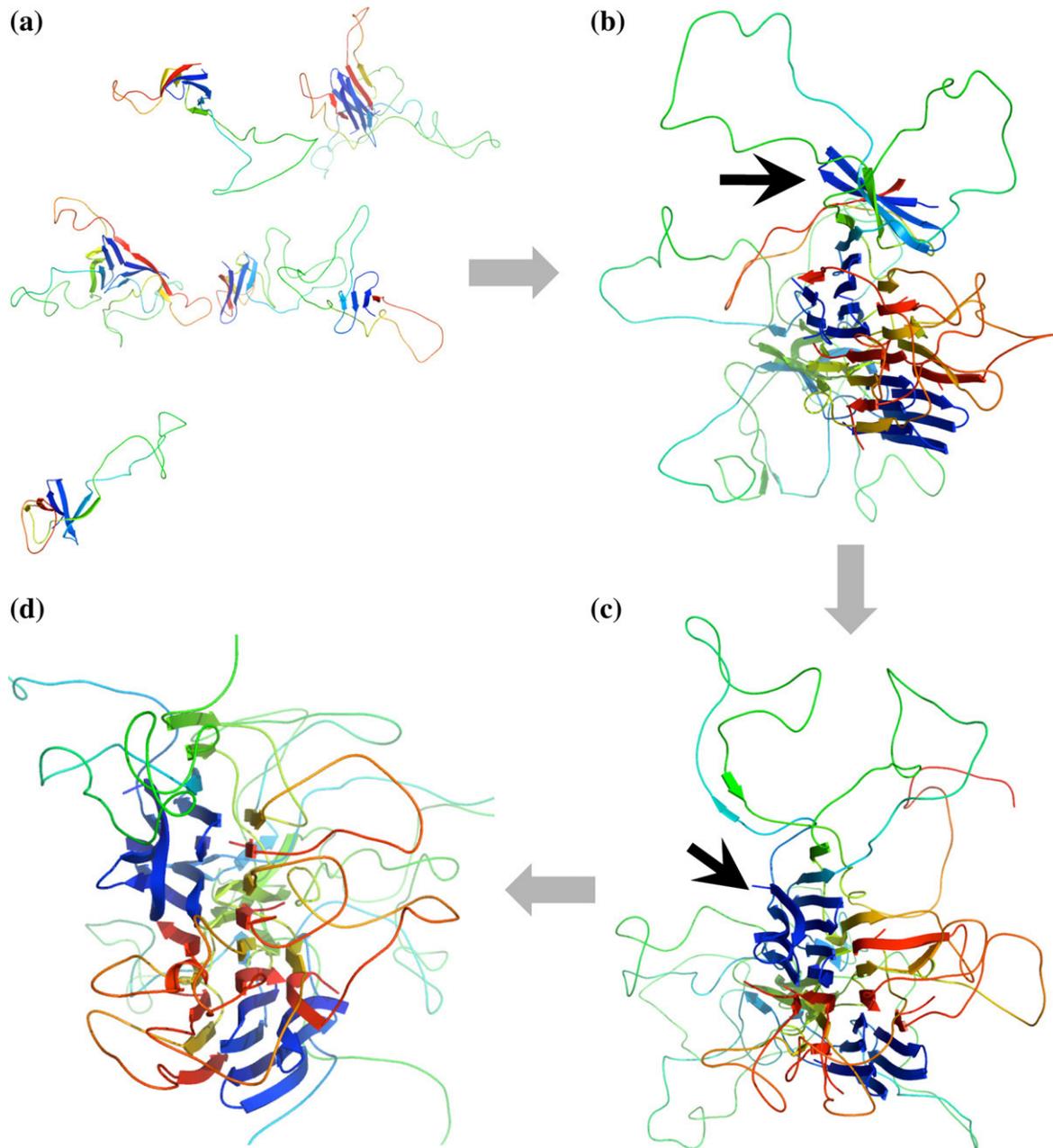


Fig. 5. The aggregation process of WT SOD1 monomers. (a) The initial configuration of eight SOD1 monomers in the simulation box. The monomer concentration is as high as 1 mM. The cartoon representation is assigned and illustrated by using MOLSCRIPT⁴⁰ and PyMOL, respectively. (b) One monomer (indicated by the arrow) associates with the end of the nascent amyloid-like aggregate and forms the cross- β core. (c) The associated monomer is incorporated into the aggregate after structural rearrangement. (d) The resulting stable aggregate structure with well-formed cross- β core.

342 **Aggregation of apoSOD1 monomers**

343 The aggregation of SOD1 is a long timescale
 344 process, which takes days, weeks, and months
 345 under the quiescent condition³⁷ and hours under
 346 the agitated condition.⁷ Modeling this process with
 347 high-resolution DMD simulations is computationally
 348 challenging. Instead, we use a coarse-grained
 349 two-bead protein model in DMD simulations²⁴ to
 350 study the aggregation of apoSOD1 monomers,
 351 where a structure-based interaction potential for
 352 aggregation^{24,38} is used (Materials and Methods). To
 353 promote the formation of partially unfolded SOD1
 354 as observed in both computation and experiments,⁷
 355 we assign stronger attractive interactions between
 356 residues forming the amyloid core. The core-forming
 357 residues are identified from proteolysis and
 358 mass spectroscopy of the aggregates of SOD1
 359 variants (Materials and Methods; Fig. 4).

360 *Experimental constraints promote partial unfolding*
361 *in SOD1 monomers*

362 We first characterize the coarse-grained monomer
 363 folding dynamics of the three SOD1 variants. For
 364 each variant, we perform replica exchange DMD
 365 simulations of the monomer and compute the
 366 specific heat using WHAM (Materials and Methods).
 367 Due to the two types of interactions (structure-based
 368 interactions and experimentally derived core in-
 369 teractions), the coarse-grained SOD1 monomers
 370 feature three-state folding dynamics with two
 371 distinct peaks in specific heat, which corresponds
 372 to folded intermediate and intermediate unfolded
 373 transitions, respectively³⁹ (Fig. 4a). Strong attrac-
 374 tions between the experimentally determined core-
 375 forming residues stabilize the partially unfolded
 376 intermediates, the structures of which are consistent
 377 with the experimentally derived input constraints
 378 with the core-forming segments folded (Fig. 4).
 379 Although the specific heat plots between coarse-
 380 grained and atomistic simulations of corresponding
 381 proteins are different (Fig. 1 and Fig. S1), which is
 382 expected due to different models and also different
 383 types of interaction potentials, they all display two
 384 peaks featuring a partially unfolded intermediate. To
 385 model the aggregation of SOD1 via the association of
 386 the partially unfolded intermediates, we perform
 387 equilibrium simulations of multiple SOD1 mono-

mers at the average temperature between the two 388
 transitions. 389

390 *Formation of amyloid-like SOD1 aggregates*

391 For each SOD1 variant, we perform DMD 391
 simulations with eight SOD1 monomers in a cubic 392
 box with dimensions of 227 Å, corresponding to a 393
 high concentration of approximately 1 mM. At a 394
 temperature that promotes local unfolding, each 395
 isolated monomer remains in the intermediate state. 396
 For example, in wild-type SOD1 monomers, the N- 397
 terminal strands (blue strands in Fig. 5a), the central 398
 strand 6 (yellow strand in Fig. 5a), and the C- 399
 terminal strand (red strand in Fig. 5a) are all folded. 400
 The monomers associate with each other in simula- 401
 tions and form amyloid-like oligomers (Figs. 5 and 402
 6), where the residual strands in each monomer 403
 assemble into β -sheets via inter- and intra-monomer 404
 hydrogen bonds. These β -sheets face each other to 405
 form high-order “cross- β ” structures as shown by 406
 the computed fibrillar diffraction pattern (Fig. 6). 407
 There are strands from the same protein incorporat- 408
 ed into neighboring sheets, which further stabilize 409
 the aggregates in addition to the side-chain-side- 410
 chain interactions. Our aggregation simulations of 411
 three SOD1 variants demonstrate the formation of 412
 amyloid-like aggregates by association of the parti- 413
 ally unfolded SOD1 monomers. The amyloid-like 414
 aggregate structure of SOD1 is not a simple stacking 415
 of the residual folded structure but requires a major 416
 rearrangement of each monomer. 417

418 The ends of each fibril core expose unsatisfied 418
 hydrogen bond donors and acceptors, which allow 419
 for further fibril growth. For instance, in Fig. 5b–d, 420
 we illustrate one monomer incorporation event from 421
 the wild-type simulation. An SOD1 monomer 422
 initially associates with one end of the amyloid- 423
 like aggregate by diffusion (Fig. 5b) and undergoes 424
 structural rearrangement in order to be incorporated 425
 into the ordered aggregate (Fig. 5c and d). In the 426
 final aggregate structures, we notice that both wild 427
 type (Fig. 5d) and I149T mutant (Fig. 6c) form a 428
 single-core aggregate, while G37R forms an aggre- 429
 gate with two cores during the course of DMD 430
 simulations (Fig. 6b). This result is consistent with 431
 the experimentally observed morphology of G37R 432
 fibrillar aggregate, which is much thinner, branched, 433
 and less ordered compared to the fibrils formed by 434

Fig. 6. The fibrillar aggregates of SOD1: (a) WT, (b) G37R, and (c) I149T. The first and the second columns correspond to the aggregates formed in simulations. Two views are shown by a 90° rotation along the axis of the amyloid fibril. The aggregate of G37R contains two cores. The computed fibril diffraction patterns of the aggregates²⁴ feature the typical “cross- β ” characteristics (third column). The peak along the fibril axis corresponds to the hydrogen bonds between strands, and the peak perpendicular to the axis corresponds to the separation between the adjacent β -sheet. The fourth column corresponds to the electron microscopic images of aggregates formed *in vitro*. The SOD1 fibrils were prepared by shaking disulfide-reduced apoSOD1 at 37 °C, 1200 rpm for 50 h (Materials and Methods). A bar in each panel represents 0.1 μ m.

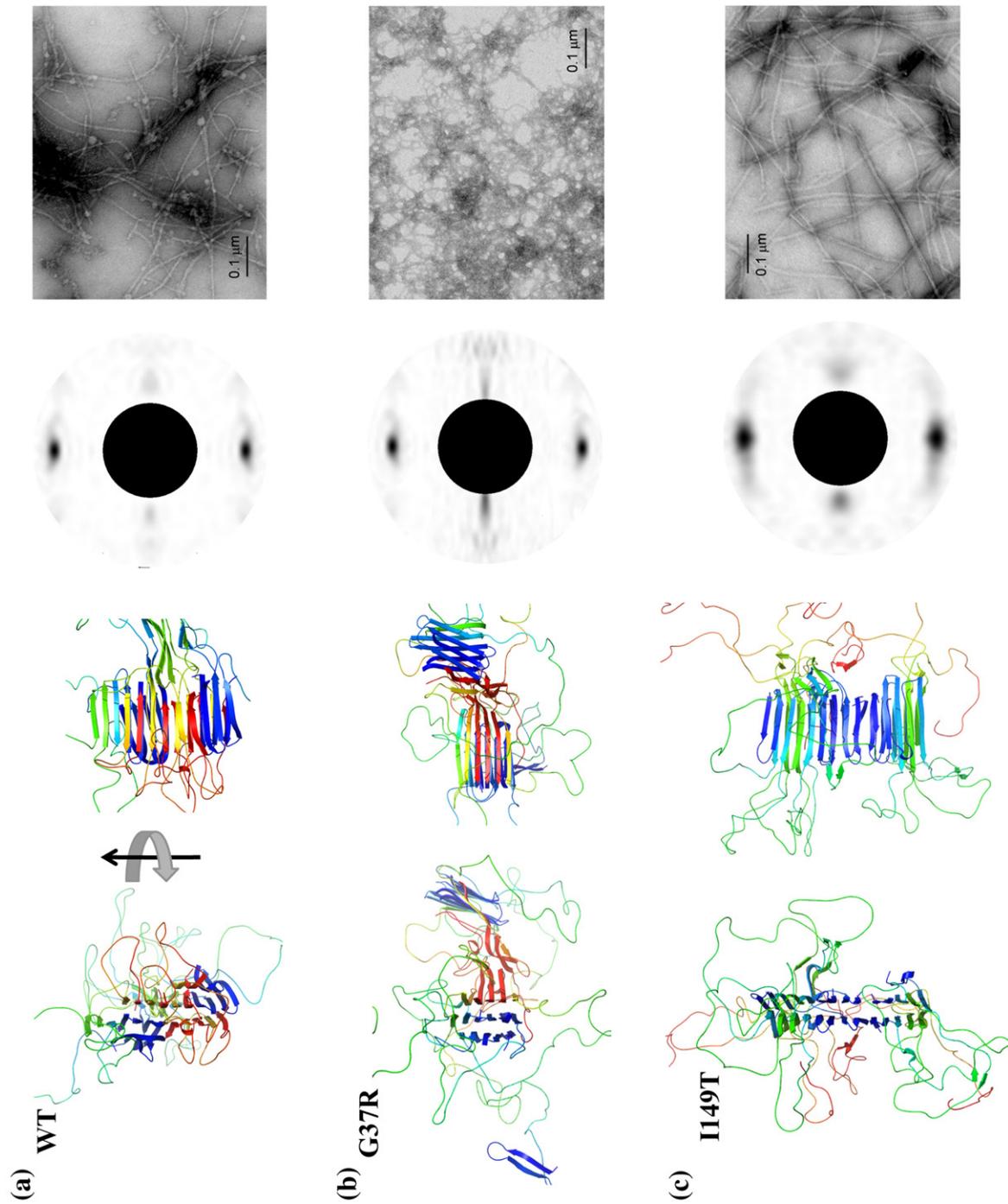


Fig. 6 (legend on previous page)

435 wild type and I149T (Fig. 6). Therefore, we postulate
 436 that the abundance of flexible loops in partially
 437 unfolded G37R results in the formation of many
 438 smaller β -rich aggregates and inhibits the formation
 439 of the long and discrete fibrillar aggregates observed
 440 in wild type and I149T.

441 In summary, we performed multiscale molecular
 442 dynamics simulations describing monomer confor-
 443 mational dynamics and oligomer formation of
 444 apoSOD1 in wild type and two mutants, G37R
 445 and I149T, each of which has distinct aggregation
 446 morphology and peptide segments forming the
 447 fibrillar core. Our simulation results suggested a
 448 generic SOD1 aggregation mechanism. After the
 449 loss of the stabilizing metal ions and disulfide bond,
 450 apoSOD1 monomer undergoes significant local
 451 unfolding before global unfolding. Mutations affect
 452 the conformational dynamics and result in distinct
 453 local unfolding patterns. The residual structure of
 454 the partially unfolded protein exhibits an exposed
 455 hydrophobic core and unsatisfied hydrogen bond
 456 donors and acceptors, which is prone to aggregate.
 457 At aggregation-favoring conditions, including high
 458 concentration and temperatures that promote local
 459 unfolding, the partially unfolded apoSOD1 mono-
 460 mers associate into amyloid-like aggregates with
 461 “cross- β ” characteristics. The aggregate structure is
 462 not formed by a simple stacking of the persistent
 463 residual structure of the monomer but, rather,
 464 requires structural rearrangement to form the
 465 ordered structure. We also found a correlation
 466 between the aggregate structures in simulations
 467 and the mesoscopic aggregate morphologies ob-
 468 served in experiments. Mutant G37R, which forms
 469 less ordered fibrillar aggregates *in vitro*, is found to
 470 form two smaller amyloid-like cores due to the large
 471 portion of unstructured segments inhibiting the
 472 formation of a single ordered “cross- β ” core as
 473 observed in the two other variants (Fig. 5). There-
 474 fore, mutations affect the residual structure of the
 475 locally unfolded apoSOD1, where the peptide
 476 segments serve as the “building block” of SOD1
 477 aggregation.^{24,25} The structured and unstructured
 478 regions of the partially folded state determine the
 479 morphology of the aggregates.

480 Materials and Methods

481 Atomistic DMD simulations

482 DMD is a special type of molecular dynamics simulation
 483 where pairwise interaction potentials are modeled with
 484 discontinuous functions.⁴¹ The algorithm for DMD can be
 485 found in Refs. 42 and 43. We use an atomistic DMD force
 486 field introduced in Ref. 27 to study apoSOD1 monomer
 487 dynamics. Briefly, we use the united-atom model to
 488 represent the protein, where all heavy atoms and polar
 489 hydrogen atoms are explicitly modeled. The bonded

interactions include covalent bonds, bond angles, and
 490 dihedrals. We include van der Waals, solvation, and
 491 environment-dependent hydrogen bonding interactions in
 492 the nonbonded interactions. The solvation energy is
 493 modeled using the Lazaridis–Karplus implicit solvation
 494 model with the fully solvated conformation as the
 495 reference state.²⁹ The hydrogen bond interaction is
 496 modeled using a reaction algorithm.⁴⁴ In addition to the
 497 previous version of the atomistic DMD force field,²⁷ we
 498 also add electrostatic interactions between charged resi-
 499 dues, including the basic and acidic residues. We assign
 500 integer charges to the central atoms of charged groups: CZ
 501 for arginine, NZ for lysine, CG for aspartic acid, and CD for
 502 glutamic acid. We use the Debye–Hückel approximation to
 503 model the screened charge–charge interactions. The Debye
 504 length is set at approximately 10 Å by assuming water
 505 relative permittivity of 80 and a monovalent electrolyte
 506 concentration of 0.1 mM. We discretize the continuous
 507 electrostatic interaction potential with an interaction range
 508 of 30 Å, where the screened potential approaches 0 (Fig. 7).
 509 We use the constant volume DMD simulations with period
 510 boundary conditions and control the simulation temper-
 511 ature using the Anderson thermostat.⁴⁵ 512

513 Replica exchange simulations and WHAM analysis

514 We use the replica exchange method to perform
 515 simulations of multiple copies of the same system in
 516 parallel at various temperatures. At given time intervals,
 517 replicas with neighboring temperatures exchange temper-
 518 ature values according to a Metropolis-based stochastic
 519 algorithm. We set the temperature exchange interval as
 520 50 ps. Exchange between replicas increases the sampling
 521 efficiency in that energetic barriers can be overcome more
 522 easily and, in less time, with exposure to higher

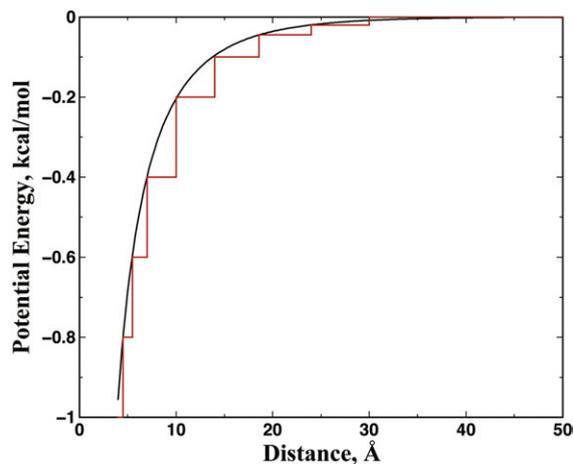


Fig. 7. Screened Debye–Hückel potential function between two opposite single charges. The continuous potential has a Debye length of ~ 10 Å, assuming the relative permittivity of water of 80 and a monovalent salt concentration of 0.1 mM. The step function in red is the discretized step function utilized in DMD simulations. For two atoms of the same charge, we change the sign of the potential to indicate a repulsive.

523 temperatures. In our simulations, we allocate 12 replicas
524 with temperatures of 270, 282, 294, 303, 312, 321, 330, 339,
525 345, 352.5, 365, and 377.5 K for the simulations of
526 apoSOD1 monomers.

527 We perform WHAM analysis using the trajectories from
528 the replica exchange simulations. The WHAM method
529 utilizes multiple simulation trajectories with overlapping
530 sampling along the reaction coordinates to self-consistently
531 compute the density of states $\rho(E)$ by combining histograms
532 from different simulation trajectories.³⁴ Given the density of
533 states, the folding specific heat (C_v) can be computed at
534 various temperatures according to the partition function
535 $Z = \int \rho(E) \exp(-E/k_B T) dE$. Here, k_B is the Boltzmann
536 constant. To compute the average RMSD as a function of
537 temperature, we compute the conditional probability $P(A|E)$
538 by observing a structure with an RMSD of A at the given
539 energy E , evaluated from all simulation trajectories. The
540 average RMSD as a function of temperature can be computed
541 as $\langle A(T) \rangle = 1/Z \int A \cdot P(A|E) \rho(E) \exp(-E/k_B T) dE$. We
542 also estimate the error bars as statistical uncertainty⁴⁶ in the
543 WHAM estimation of specific heat and average RMSD. The
544 temporal correlation in sequentially generated configurations
545 is obtained by autocorrelation analysis.

546 Reconstruction of the locally unfolded states

547 We use an RMSD range $[D_{\min}, D_{\max}]$ to identify the
548 locally unfolded structures from the trajectories. In order to
549 identify the cutoff values, we compute average RMSD as a
550 function of temperature using WHAM analysis (Fig. 2).
551 From the specific heat, we identify the transition temper-
552 ature at low temperatures corresponding to local unfold-
553 ing, T_1 . We assign the average RMSD at T_1 as D_{\min} for each
554 of the three SOD1 variants. For all three SOD1 variants, the
555 melting or global unfolding starts from a state with an
556 average RMSD of approximately 10 Å (Fig. 2). Therefore,
557 we use a D_{\max} of 10 Å for all three variants.

558 Coarse-grained DMD simulations

559 We use a two-bead protein model to study apoSOD1
560 aggregation.^{24,47} In the two-bead model, each amino acid
561 is represented by only the α -carbon (backbone) and β -
562 carbon (side chain). The bonded interactions between
563 neighboring atoms along the peptide chain are assigned so
564 as to mimic peptide geometry.⁴⁷ We use a structure-based
565 potential to model the side-chain-side-chain packing
566 interactions, where native interactions in the native state
567 as observed in the crystal structure are favored. Two
568 interacting residues can form either intra- or inter-
569 monomer contacts, in order to promote protein-protein
570 association.^{24,38} The attractions between β -carbons are
571 assigned with a hardcore distance of 3 Å and an
572 interaction range of 7.5 Å. We also model the backbone-
573 backbone hydrogen bond interaction as in Ref. 24.

574 To stabilize the partially unfolded state, where the
575 amyloid-core-forming residues remain folded, we assign a
576 strong attraction between the core residues. In the
577 proteolysis/mass spectroscopy study of aggregates,
578 there are overlapping peptides observed for each core-
579 forming region (e.g., the N-terminal region for wild-type
580 SOD1).⁷ For all of the *overlapping peptides* observed
581 experimentally, we assume that each peptide has an

equal chance to participate in the amyloid core. Therefore,
582 we can compute the probability for each residue in a given
583 region to be observed in the amyloid core, $P^{\text{Core}}(i)$ (Fig. 4).
584 By combing all regions, we obtain the core-forming
585 probability for all residues. We therefore introduce an
586 experimentally derived bias potential to the structure-
587 based potential:
588

$$E_{ij} = [\epsilon^{G^o} + (\epsilon^{\text{Core}} - \epsilon^{G^o}) P^{\text{Core}}(i) P^{\text{Core}}(j)] A_{ij}$$

589 Here, Δ_{ij} is the native interaction matrix, which equals 1 if
590 the two residues i and j are in contact in the native state
591 and 0 otherwise. ϵ^{G^o} and ϵ^{Core} are the energy scales for the
592 structure-based and experimentally determined bias
593 potentials, correspondingly. In our simulations, we assign
594 $\epsilon^{G^o} = 0.5 \epsilon_0$ and $\epsilon^{\text{Core}} = 1.0 \epsilon_0$ with ϵ_0 as the energy unit. The
595 hydrogen bond interaction strength is $2.0 \epsilon_0$.
596
597

598 Electron microscopic observation of *in vitro* 599 SOD1 fibrils

600 Preparation and observation of SOD1 fibrils (WT, G37R,
601 and I149T) were performed as described previously.⁷
602 Briefly, following constant agitation of 100 mM SOD1 at
603 37 °C, 1200 rpm for 50 h, we collected insoluble aggregates
604 by ultracentrifugation at 110,000g for 30 min. In order to
605 avoid shearing fibrillar structures during agitation, we
606 further performed a seeding reaction by adding 10 mM
607 SOD1 aggregates (monomer based) to 100 mM soluble
608 SOD1 and incubating the mixture at 37 °C for 3 days in the
609 absence of agitation. Insoluble pellets were again collected
610 by ultracentrifugation, resuspended in pure water, and
611 adsorbed on 400-mesh grids coated by a glow-charged
612 supporting membrane. After negative staining with 1%
613 uranyl acetate, images were obtained using an electron
614 microscope (1200EX, JEOL).
615

616 Supplementary materials related to this article can be
617 found online at doi:10.1016/j.jmb.2011.12.029

Acknowledgements

618 We thank Elizabeth A. Proctor and Rachel Redler
619 for helpful discussions and critical reading of the
620 manuscript. Calculations are performed on the
621 topsail high-performance computing cluster at the
622 University of North Carolina at Chapel Hill. This
623 work was supported by the National Institutes of
624 Health grant R01 GM080742.
625

References

- 626 Q4
627 1. Rosen, D. R., Siddique, T., Patterson, D., Figlewicz,
628 D. A., Sapp, P., Hentati, A. *et al.* (1993). Mutations in
629 Cu/Zn superoxide dismutase gene are associated
630 with familial amyotrophic lateral sclerosis. *Nature*,
631 **362**, 59–62.
- 632 2. Brijn, L. I., Miller, T. M. & Cleveland, D. W. (2004).
633 Unraveling the mechanisms involved in motor neuron
634 degeneration in ALS. *Annu. Rev. Neurosci.* **27**, 723–749.

- 635 3. Furukawa, Y., Kaneko, K., Yamanaka, K.,
636 O'Halloran, T. V. & Nukina, N. (2008). Com-
637 plete loss of post-translational modifications
638 triggers fibrillar aggregation of SOD1 in the
639 familial form of amyotrophic lateral sclerosis. *J.*
640 *Biol. Chem.* **283**, 24167–24176.
- 641 4. Oztug Durer, Z. A., Cohlberg, J. A., Dinh, P., Padua,
642 S., Ehrenclou, K., Downes, S. *et al.* (2009). Loss of metal
643 ions, disulfide reduction and mutations related to
644 familial ALS promote formation of amyloid-like
645 aggregates from superoxide dismutase. *PLoS One*, **4**,
646 e5004.
- 647 5. Jahn, T. R., Makin, O. S., Morris, K. L., Marshall, K. E.,
648 Tian, P., Sikorski, P. & Serpell, L. C. (2010). The
649 common architecture of cross- β amyloid. *J. Mol. Biol.*
650 **395**, 717–727.
- 651 6. Khare, S. D. & Dokholyan, N. V. (2006). Common
652 dynamical signatures of familial amyotrophic lateral
653 sclerosis-associated structurally diverse Cu, Zn super-
654 oxide dismutase mutants. *Proc. Natl Acad. Sci. U. S. A.*
655 **103**, 3147–3152.
- 656 7. Furukawa, Y., Kaneko, K., Yamanaka, K. & Nukina,
657 N. (2010). Mutation-dependent polymorphism of Cu,
658 Zn-superoxide dismutase aggregates in the familial
659 form of amyotrophic lateral sclerosis. *J. Biol. Chem.*
660 **285**, 22221–22231.
- 661 8. Andersen, P. M., Nilsson, P., Keranen, M. L., Forsgren,
662 L., Hagglund, J., Karlsborg, M. *et al.* (1997). Phenotypic
663 heterogeneity in motor neuron disease patients with
664 CuZn-superoxide dismutase mutations in Scandina-
665 via. *Brain*, **120**, 1723–1737.
- 666 9. Bruce, M. E. & Fraser, H. (1991). Scrapie strain
667 variation and its implications. *Curr. Top. Microbiol.*
668 *Immunol.* **172**, 125–138.
- 669 10. Tanaka, M., Chien, P., Naber, N., Cooke, R. &
670 Weissman, J. S. (2004). Conformational variations in
671 an infectious protein determine prion strain differ-
672 ences. *Nature*, **428**, 323–328.
- 673 11. Jones, E. M. & Surewicz, W. K. (2005). Fibril
674 conformation as the basis of species- and strain-
675 dependent seeding specificity of mammalian prion
676 amyloids. *Cell*, **121**, 63–72.
- 677 12. Nekooki-Machida, Y., Kurosawa, M., Nukina, N., Ito,
678 K., Oda, T. & Tanaka, M. (2009). Distinct conforma-
679 tions of *in vitro* and *in vivo* amyloids of huntingtin-
680 exon1 show different cytotoxicity. *Proc. Natl Acad. Sci.*
681 *USA*, **106**, 9679–9684.
- 682 13. Shaw, B. F. & Valentine, J. S. (2007). How do ALS-
683 associated mutations in superoxide dismutase 1
684 promote aggregation of the protein? *Trends Biochem.*
685 *Sci.* **32**, 78–85.
- 686 14. Kayatekin, C., Zitzewitz, J. A. & Matthews, C. R.
687 (2010). Disulfide-reduced ALS variants of Cu, Zn
688 superoxide dismutase exhibit increased populations
689 of unfolded species. *J. Mol. Biol.* **398**, 320–331.
- 690 15. Khare, S. D., Caplow, M. & Dokholyan, N. V. (2004).
691 The rate and equilibrium constants for a multistep
692 reaction sequence for the aggregation of superoxide
693 dismutase in amyotrophic lateral sclerosis. *Proc. Natl*
694 *Acad. Sci. USA*, **101**, 15094–15099.
- 695 16. Arnesano, F., Banci, L., Bertini, I., Martinelli, M.,
696 Furukawa, Y. & O'Halloran, T. V. (2004). The
697 unusually stable quaternary structure of human Cu,
698 Zn-superoxide dismutase 1 is controlled by both
699 metal occupancy and disulfide status. *J. Biol. Chem.* **699**
700 **279**, 47998–48003.
- 701 17. Hough, M. A., Grossmann, J. G., Antonyuk, S. V.,
702 Strange, R. W., Doucette, P. A., Rodriguez, J. A. *et al.*
703 (2004). Dimer destabilization in superoxide dismutase
704 may result in disease-causing properties: struc-
705 tures of motor neuron disease mutants. *Proc. Natl*
706 *Acad. Sci. USA*, **101**, 5976–5981.
- 707 18. Rakhit, R., Crow, J. P., Lepock, J. R., Kondejewski, L.
708 H., Cashman, N. R. & Chakrabartty, A. (2004).
709 Monomeric Cu,Zn-superoxide dismutase is a com-
710 mon misfolding intermediate in the oxidation models
711 of sporadic and familial amyotrophic lateral sclerosis.
712 *J. Biol. Chem.* **279**, 15499–15504.
- 713 19. Ding, F. & Dokholyan, N. V. (2008). Dynamical roles
714 of metal ions and the disulfide bond in Cu, Zn
715 superoxide dismutase folding and aggregation. *Proc.*
716 *Natl Acad. Sci. USA*, **105**, 19696–19701.
- 717 20. Elam, J. S., Taylor, A. B., Strange, R., Antonyuk, S.,
718 Doucette, P. A., Rodriguez, J. A. *et al.* (2003). Amyloid-
719 like filaments and water-filled nanotubes formed by
720 SOD1 mutant proteins linked to familial ALS. *Nat.*
721 *Struct. Biol.* **10**, 461–467.
- 722 21. Banci, L., Bertini, I., Cramaro, F., Del Conte, R. &
723 Viezzoli, M. S. (2003). Solution structure of Apo Cu,Zn
724 superoxide dismutase: role of metal ions in protein
725 folding. *Biochemistry*, **42**, 9543–9553.
- 726 22. Nordlund, A. & Oliveberg, M. (2006). Folding of
727 Cu/Zn superoxide dismutase suggests structural
728 hotspots for gain of neurotoxic function in ALS:
729 parallels to precursors in amyloid disease. *Proc. Natl*
730 *Acad. Sci. USA*, **103**, 10218–10223.
- 731 23. Shaw, B. F., Durazo, A., Nersissian, A. M., Whitelegge,
732 J. P., Faull, K. F. & Valentine, J. S. (2006). Local unfolding
733 in a destabilized, pathogenic variant of superoxide
734 dismutase 1 observed with H/D exchange and mass
735 spectrometry. *J. Biol. Chem.* **281**, 18167–18176.
- 736 24. Ding, F., Dokholyan, N. V., Buldyrev, S. V., Stanley,
737 H. E. & Shakhnovich, E. I. (2002). Molecular dynamics
738 simulation of the SH3 domain aggregation suggests a
739 generic amyloidogenesis mechanism. *J. Mol. Biol.* **324**,
740 851–857.
- 741 25. Sinha, N., Tsai, C. J. & Nussinov, R. (2001). A
742 proposed structural model for amyloid fibril elonga-
743 tion: domain swapping forms an interdigitating β -
744 structure polymer. *Protein Eng.* **14**, 93–103.
- 745 26. Das, P., King, J. A. & Zhou, R. (2011). Aggregation of
746 γ -crystallins associated with human cataracts via
747 domain swapping at the C-terminal β -strands. *Proc.*
748 *Natl Acad. Sci. USA*, **108**, 10514–10519.
- 749 27. Ding, F., Tsao, D., Nie, H. & Dokholyan, N. V. (2008).
750 *Ab initio* folding of proteins with all-atom discrete
751 molecular dynamics. *Structure*, **16**, 1010–1018.
- 752 28. Ding, F. & Dokholyan, N. V. (2006). Emergence of
753 protein fold families through rational design. *PLoS*
754 *Comput. Biol.* **2**, e85.
- 755 29. Lazaridis, T. & Karplus, M. (2000). Effective energy
756 functions for protein structure prediction. *Curr. Opin.*
757 *Struct. Biol.* **10**, 139–145.
- 758 30. Karginov, A. V., Ding, F., Kota, P., Dokholyan, N. V. &
759 Hahn, K. M. (2010). Engineered allosteric activation of
760 kinases in living cells. *Nat. Biotechnol.* **28**, 743–747.
- 761 31. Proctor, E. A., Ding, F. & Dokholyan, N. V. (2011).
762 Structural and thermodynamic effects of post-762

- 763 translational modifications in mutant and wild type
764 Cu, Zn superoxide dismutase. *J. Mol. Biol.* **408** ,
765 555–567.
- 766 32. Sugita, Y. & Okamoto, Y. (1999). Replica-exchange
767 molecular dynamics method for protein folding.
768 *Chem. Phys. Lett.* **314**, 141–151.
- 769 33. Yin, S., Ding, F. & Dokholyan, N. V. (2007). Eris: an
770 automated estimator of protein stability. *Nat. Methods*,
771 **4**, 466–467.
- 772 34. Kumar, S., Bouzida, D., Swendsen, R. H., Kollman, P. A.
773 & Rosenberg, J. M. (1992). The weighted histogram
774 analysis method for free-energy calculations on bio-
775 molecules. 1. The method. *J. Comput. Chem.* **13**,
776 1011–1021.
- 777 35. Zhou, Y. & Karplus, M. (1997). Folding thermody-
778 namics of a model three-helix-bundle protein. *Proc.*
779 *Natl Acad. Sci. USA*, **94**, 14429–14432.
- 780 36. Rodriguez, J. A., Shaw, B. F., Durazo, A., Sohn, S. H.,
781 Doucette, P. A., Nersissian, A. M. *et al.* (2005).
782 Destabilization of apoprotein is insufficient to
783 explain Cu,Zn-superoxide dismutase-linked ALS
784 pathogenesis. *Proc. Natl Acad. Sci. USA*, **102**,
785 10516–10521.
- 786 37. Vassall, K. A., Stubbs, H. R., Primmer, H. A., Tong, M.
787 S., Sullivan, S. M., Sobering, R. *et al.* (2011). Decreased
788 stability and increased formation of soluble aggre-
789 gates by immature superoxide dismutase do not
790 account for disease severity in ALS. *Proc. Natl Acad.*
791 *Sci. USA*, **108**, 2210–2215.
- 792 38. Yang, S., Cho, S. S., Levy, Y., Cheung, M. S.,
793 Levine, H., Wolynes, P. G. & Onuchic, J. N. (2004).
794 Domain swapping is a consequence of minimal
787 frustration. *Proc. Natl Acad. Sci. USA*, **101**, 795
13786–13791. 796
39. Khare, S. D., Ding, F. & Dokholyan, N. V. (2003). 797
Folding of Cu, Zn superoxide dismutase and familial 798
amyotrophic lateral sclerosis. *J. Mol. Biol.* **334** , 799
515–525. 800
40. Kraulis, P. J. (1991). MOLSCRIPT: a program to 801
produce both detailed and schematic plots of protein 802
structures. *J. Appl. Crystallogr.* **24**, 4. 803
41. Rapaport, D. C. (1978). Molecular dynamics simula- 804
tion of polymer chains with excluded volume. *J. Phys.* 805
A: Math. Gen. **11**, L213–L217. 806
42. Allen, M. P. & Tildersley, D. J. (1989). *Computer* 807
Simulation of Liquids. Clarendon Press, New York, NY. 808
43. Rapaport, D. C. (1997). *The Art of Molecular Dynamics* 809
Simulation. Cambridge University Press, Cambridge, 810
UK. 811
44. Ding, F., Borreguero, J. M., Buldyrev, S. V., Stanley, H. E. 812
& Dokholyan, N. V. (2003). Mechanism for the α -helix to 813
 β -hairpin transition. *Proteins*, **53**, 220–228. 814
45. Andersen, H. C. (1980). Molecular dynamics simula- 815
tions at constant pressure and/or temperature. *J.* 816
Chem. Phys. **72**, 2384–2393. 817
46. Chodera, J. D., Swope, W. C., Pitera, J. W., Seok, C. & 818
Dill, K. A. (2007). Use of the weighted histogram 819
analysis method for the analysis of simulated and 820
parallel tempering simulations. *J. Chem. Theory* 821
Comput. **3**, 26–41. 822
47. Ding, F., Dokholyan, N. V., Buldyrev, S. V., Stanley, 823
H. E. & Shakhnovich, E. I. (2002). Direct molecular 824
dynamics observation of protein folding transition 825
state ensemble. *Biophys. J.* **83**, 3525–3532. 826