THE JOURNAL OF PHYSICAL CHEMISTRY B

Article

Cite This: J. Phys. Chem. B 2018, 122, 11662−11676

pubs.acs.org/JPCB

Prediction of Misfolding-Specific Epitopes in SOD1 Using Collective Coordinates

Xubiao Peng,[†,∥] Neil R. Cashman,[‡] and Steven S. Plotkin*[,§]

[†]Department of Physics and Astronomy, University of British Columbia, Vancouver, British Columbia V6T 1Z1, Canada

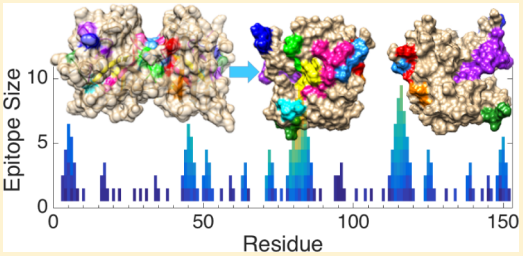
[‡]Brain Research Centre, University of British Columbia, Vancouver, British Columbia V6T 2B5, Canada

[§]Department of Physics and Astronomy, and Genome Sciences and Technology Program, University of British Columbia, Vancouver, British Columbia V6T 1Z1, Canada

[∥]Center for Quantum Technology Research, School of Physics, Beijing Institute of Technology, Haidian, Beijing 100081, China

Supporting Information

ABSTRACT: We introduce a global, collective coordinate bias into molecular dynamics simulations that partially unfolds a protein, in order to predict misfolding-specific epitopes based on the regions that locally unfold. Several metrics are used to measure local disorder, including solvent exposed surface area (SASA), native contacts (Q), and root mean squared fluctuations (RMSF). The method is applied to Cu, Zn superoxide dismutase (SOD1). For this protein, the processes of monomerization, metal loss, and conformational unfolding due to microenvironmental stresses are all separately taken into account. Several misfolding-specific epitopes are predicted, and consensus epitopes are calculated. These predicted epitopes are consistent with the "lower-resolution" peptide sequences used to raise disease-specific antibodies, but the epitopes derived from collective coordinates contain shorter, more refined sequences for the key residues constituting the epitope.



INTRODUCTION

Many misfolded proteins implicated in both neurodegenerative and systemic amyloid-related diseases appear to exhibit aggregated fibrils with a significant degree of native structure,[1] including transthyretin,[2] β2-microglobulin,[3] and superoxide dismutase (SOD1).[4,5] This suggests that local, rather than global, protein unfolding may play a significant role in the protein aggregates implicated in the propagation of these diseases.

Similarly, antibodies have been developed to target epitopes selectively exposed on non-native misfolded forms of several proteins that are misfolded in disease, including PrPc,[6] SOD1,[7] and TTR.[8] These antibodies have been shown to suppress fibril formation[8] and block cell-to-cell propagation of misfolded protein.[9] Antibodies positively selective to Aβ oligomers and negatively selective against Aβ fibrils have also been recently developed[10] and are currently being developed as Alzheimer's disease therapeutics.

Force fields parametrized quantum mechanically (e.g., CHARMM or AMBER)[11−13] are now sufficiently accurate to reproduce experimental folded protein structures *de novo* (i.e., to fold proteins).[14−16] The force fields used to fold proteins that are parametrized by quantum chemical methods tend to be most accurate near or around the native structure. By examining partial structural perturbations from the native ensemble, rather than global unfolding events, the known force fields would then be applied well within their range of validity. These force fields have also been extended to accurately model unfolded peptides, including intrinsically disordered proteins.[17]

Here we propose an algorithmic approach to predict protein regions that are most prone to disorder due to local unfolding. The general hypothesis is that weakly stable regions in the native fold may constitute target epitopes that are preferentially exposed in misfolded species. Statistical thermodynamic models of the low-energy excitations involved in local unfolding of proteins have been developed.[18−22] Modeling the unfolding problem, they treat the problem inverse to predicting folding nuclei.[23−26] Locally unfolded regions may result from significant cooperativity,[27] rendering some modeling approximations (using contiguous sequences for example) too severe. We avoid such approximations here through direct molecular dynamic simulation using collective coordinate biases. The predicted weakly stable regions may be utilized as candidate misfolding-specific epitopes (MSEs); these epitopes can be distinguished from the same regions in the native fold by their conformational properties. By properly scaffolding the epitopes,[28] antibodies may be raised to bind to them selectively in misfolded species. Sophisticated methods of epitope scaffolding have been successful in constructing effective antigenic targets for rational vaccine design.[29−32] Misfolding-selective antibodies

Special Issue: William A. Eaton Festschrift

Received: August 7, 2018
Published: October 16, 2018

DOI: 10.1021/acs.jpcb.8b07680
J. Phys. Chem. B 2018, 122, 11662−11676

have distinct advantages in targeting pathogenic forms of protein, while sparing healthy and functional forms of the protein.

The outline of this paper is as follows. We first describe a "collective coordinates" method to predict weakly stable, unfolding prone regions within a protein. Next, we apply our method specifically to the protein superoxide dismutase 1 (SOD1), a protein whose misfolding is implicated in familial forms of ALS. The predicted epitopes depend on the reference structure, whether dimeric or monomeric, and whether metalated or apo. We summarize the epitope predictions by calculating "consensus" misfolding-specific epitopes. We validate the method by comparison with experimental measures of solvent exposure and fluctuation dynamics, as well as by comparison of predicted and experimentally observed MSEs. We then make a comparison between the collective coordinates method and simple equilibrium measurements. Last, we compare the results for two different salts, KCl and NaCl.

COMPUTATIONAL METHODS

Input Structural Model. To predict epitopes exposed upon partial unfolding of the protein, we wish to predict likely local unfolding events, where a protein region deviates structurally from a putative "native" structural conformation. We ask the following: If a structured protein is *globally* challenged, perhaps by some anomalous environmental cue, such that in response it begins to unfold or misfold, which regions of the protein are most likely to unfold? To answer this question, we employ a technique known as collective coordinate biasing, described further below.

Figure 1 depicts the computational procedure for obtaining candidate unfolding-specific epitopes. The inputs to the prediction are both a structural model of the protein and an atomic force field, such as the CHARMM force field mentioned above. A structural model may be either an experimentally determined set of nuclear coordinates deposited on the protein data bank (PDB, www.rcsb.org,[33] or it may be a set of coordinates determined by computational structure prediction algorithms. In any case, the structural model may be a biologically functional structure, or it may be a misfolded, aggregated, or fibril structure. The structural model may also consist of one chain of the protein, or of many chains of protein, as is the case for fibril structures.

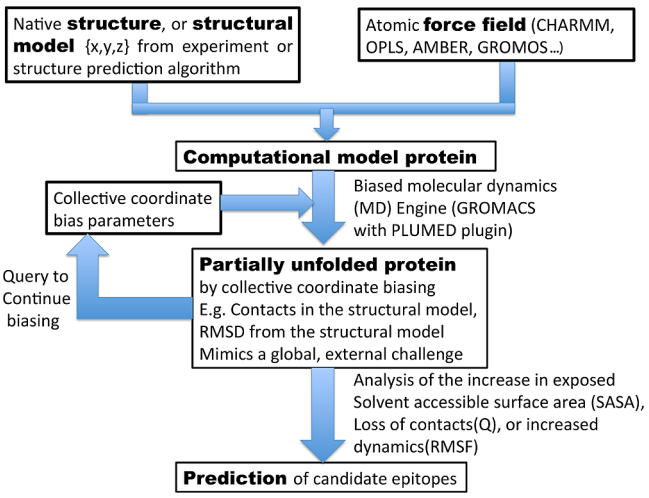


Figure 1. Flowchart of the steps in the prediction algorithm;[34] see text.

Protein System Choice. We choose superoxide dismutase (SOD1) as a protein to which we apply our method. SOD1 is a protein whose misfolding and propagation are implicated in some autosomal dominant forms of familial ALS.[35,36] No structure(s) of the misfolded, propagative species of SOD1 has been experimentally determined, however, and there is evidence in both ALS and other protein misfolding diseases to support the fact that such a species is conformationally plastic, i.e., polymorphic.[37,38] In the absence of such structural models, we can seek to predict MSEs by looking at regions of weak thermodynamic stability in native SOD1, with the assumption that regions weakly stable in native protein will also not have a differential preference for stability in misfolded protein, and will thus tend to be exposed to solvent and accessible for antibody binding. We use the terms misfolding-specific epitope (MSE) and unfolding-specific epitope interchangeably in this paper.

Mature SOD1 is a homodimer wherein each monomer contains a disulfide bond and binds a Cu and Zn ion (see Figure 2). Lack of formation or loss of dimer formation,[39,40] lack/loss of metals,[41] and lack/loss of the disulfide bond[42] are thought to be precursors on the pathway to pathogenesis.[43]

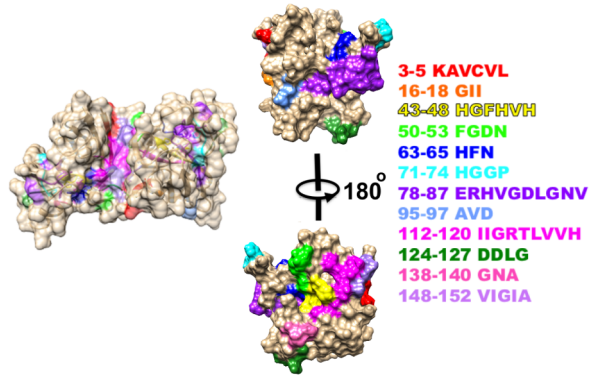


Figure 2. Predicted misfolding-specific epitopes (MSEs) from ΔSASA for E,E(SH) using the holo dimer as a reference state. The left side shows the epitopes in the context of the holo dimer. The right side shows the partially unfolded E,E(SH) structure from two different viewpoints. The color-coded MSEs are listed in the figure legend.

Most WT human SOD1 found in the CNS of hSOD1 transgenic mice is disulfide-bonded and inactive, suggesting a preponderance of either the E,E(SS) or E,Zn(SS) form.[44] As well, a reduced disulfide bond would not likely be stable under the oxidative conditions that would favor observed oxidative modifications of His and Trp residues, which induce aggregation of mutant SOD1.[45,46] On the basis of the cocrystal structure with SOD1's copper chaperone CCS,[47] as well biochemical analyses,[48] the intramolecular disulfide bond in SOD1 is reduced prior to Cu loading. Therefore, premature apo forms of SOD1 are likely reduced, and thus predominantly monomeric.[49] In our analysis below, we consider both E,E(SS) and E,E(SH) forms of SOD1 for the purpose of MSE prediction, as outlined in Table 1 below.

Modeling Native Protein. To model E,E(SS) SOD1, we started with the corresponding NMR structure (PDB 1RK7)[50] of a "pseudo-WT" obligate E,E(SS) SOD1 monomer containing five mutations, which allowed us to obtain an equilibrated native ensemble in reasonable simulation time. The mutations include

DOI: 10.1021/acs.jpcb.8b07680
J. Phys. Chem. B 2018, 122, 11662−11676

Table 1. Biased States and Reference States Corresponding to Disorder-Inducing Processes with Shorthand Used in the Text

biased state	reference state	notation	process
E,E(SS) stressed monomer	E,E(SS) monomer	$EE(SS)_{mon} \rightarrow EE(SS)^*$	partial unfolding
E,E(SH) stressed monomer	E,E(SS) monomer	$EE(SS)_{mon} \rightarrow EE(SH)^*$	SS reduction, partial unfolding
E,E(SS) stressed monomer	Cu,Zn(SS) monomer	$CuZn(SS)_{mon} \rightarrow EE(SS)^*$	metal loss, partial unfolding
E,E(SH) stressed monomer	Cu,Zn(SS) monomer	$CuZn(SS)_{mon} \rightarrow EE(SH)^*$	metal loss, SS reduction, partial unfolding
E,E(SS) stressed monomer	Cu,Zn(SS) dimer	$CuZn(SS)_{dim} \rightarrow EE(SS)^*$	metal loss, monomerization, partial unfolding
E,E(SH) stressed monomer	Cu,Zn(SS) dimer	$CuZn(SS)_{dim} \rightarrow EE(SH)^*$	metal loss, monomerization, SS reduction, partial unfolding

C6A and C111S to ablate intermolecular disulfide bonds, F50E and G51E to disrupt the native homodimer, and E133Q to increase the functional activity of the monomeric species.[51] To study the WT SOD1 E,E(SS) monomer using PDB 1RK7, we first mutate the above residues back to their WT identities using the software SCWRL.[52] This back-mutated structure is then equilibrated and biased by the collective coordinates algorithm to be 35% unfolded, as described below.

To model the native ensemble for E,E(SH) SOD1, the above back-mutated structure of WT E,E(SS) SOD1 is modified by reducing the disulfide bond in Gromacs, and the system is then allowed to equilibrate for 600 ns.

To model Cu,Zn(SS) SOD1, we must reparameterize histidine 63, which bridges the Cu and Zn ions in the native structure and is doubly deprotonated.[53] There is no entry for such a histidine in the Charmm force field. To parametrize this side chain, we employ an optimization scheme. The quantum-chemical potential energy of Cu^{2+}; Zn^{2+}; all atoms in the histidine rings of the metal-coordinating amino acids H46, H48, H63, H71, H80, H120; and all atoms in the side chain of D83 is calculated in Gaussian 09.[54] The quantum-chemical potential energies of the apo protein system, and the isolated Cu^{2+} and Zn^{2+} in their native positions, are subtracted from this to obtain the interaction energy between the protein and the metals. This interaction potential energy is then compared with the classical interaction potential energy between protein and metals (for the PDB conformation) using the CHARMM22* force field. The partial charges of the H63 side chain ring are then allowed to vary and are relaxed to minimize the difference between the classical CHARMM22* interaction potential energy (electrostatic plus van der Waals) and the Gaussian-derived interaction potential energy. Hydrogen partial charges are constrained to be within $\pm 0.5e$, and other heavy atoms are constrained to be within $\pm 2e$; however, no atom charge reached these constrained limits during the relaxation process. This procedure gives the parameters in Supporting Information Table S1.

With these parameters for H63, the Cu and Zn are observed to remain coordinated inside SOD1 for 100 ns of equilibration, by which time the RMSD of the protein has converged. The native contact maps for Cu,Zn(SS) monomer and dimer are then obtained from the corresponding equilibrium ensemble.

Collective Coordinates. We perform simulations using the publicly available software packages GROMACS[55] and PLUMED[56] to implement a global unfolding bias in a collective coordinate (see Figure 1). Simulations were performed using the CHARMM22* force-field parameters[57] with the TIP3P water model,[58] and a salt concentration of 100 mM KCl. Intracellular salt concentrations actively favor KCl over NaCl, and the ion pair formation strengths and protein solvation properties of K^+ and Na^+ have been shown to differ.[59,60] We have thus also run simulation predictions using 100 mM NaCl, and we compare the results below.

The predictions based on the method are essentially as accurate as the force fields used. As mentioned above, distributed computing[61] or custom supercomputers[15] can now fold proteins using these force fields, which supports their accuracy.

A collective coordinate can be any function of the atomic positions or energies that applies a globally destabilizing influence to a protein under consideration, thereby inducing loss of native structure. The collective coordinate calculation for any conformation generally utilizes a scalar to measure the global degree of native structure present for that conformation. In the fully native structure, the collective coordinate might have a value of unity, while in a globally unfolded, random coil structure, the collective coordinate would then have a value at or near zero.

The collective coordinate is dynamically decreased from its value in the native ensemble, which induces the protein to adopt updated structures with progressively smaller values of the collective coordinate. Some examples of global collective coordinates are as follows:

(1) the fraction or number of native contacts of the updated structure, defined through the number of pairs of heavy (non-hydrogen) atoms that are within a cutoff distance of each other in the experimentally determined native structure;
(2) the total solvent-accessible surface area (SASA) of the updated structure, relative to the SASA of the native structure
(3) the structural overlap function, which is a nonlocal measure that matches distances between pairs of atoms in an arbitrary structure with the corresponding distances between those pairs of atoms in the native structure;[62]
(4) the root mean squared deviation (RMSD) of an updated structure relative to the native structural model;
(5) the radius of gyration of an updated structure relative to the radius of gyration of the native structure;
(6) the number of backbone hydrogen bonds in the updated structure from among the backbone hydrogen bonds in the native structure;
(7) the total intraprotein potential energy of the system
(8) the generalized Euclidean distance from the native structure;[63]
(9) combinations or compound functions of one or more of the above collective coordinates.

In general we refer to the global collective coordinates used to bias the MD simulation as a scalar Q.

The advantage of the collective coordinate method is as follows. The system is dynamically biased from the folded ensemble to an ensemble that is, say, 35% unfolded (and thus 65% folded). It is then constrained, via an effective harmonic potential, to equilibrate in this partially disordered ensemble. Because the bias toward structures with 35% disorder is global, we do not specify where the protein may choose to become

DOI: 10.1021/acs.jpcb.8b07680
J. Phys. Chem. B 2018, 122, 11662−11676

disordered to satisfy this constraint. The region(s) of disorder are chosen "by the protein" based on a complex interplay between its intrinsic energy function or force field, and the entropy change upon disordering those parts of the protein. The regions or "hot-spots" of the protein that are prone to becoming disordered constitute predictions of the method. The protein regions that are prone to disorder are intended to serve as epitopes to which therapeutic agents may be raised.

Collective coordinates have been used previously in several applications. Collective coordinate biasing was used to more accurately model a folding intermediate of barstar, by matching experimental mass spectrometry data from fast photochemical oxidation.[64] Collective coordinate biasing on multiple order parameters, including contact number and parallel and antiparallel β sheet content, was applied to sample the conformation space for a system of 18 chains each of 8 valines.[65,66] Collective coordinate biasing has been used along with metadynamics,[67,68] as well as to obtain the free energy landscape of EGFR kinase.[69]

In the application of this paper, the native structure corresponds to a protein structure obtained from the PDB and has a set of native contacts defined by all pairs of heavy (non-hydrogen) atoms within 4.8 Å of each other that are in different amino acids labeled by primary sequence residue index α, β, that satisfy $|\alpha - \beta| \geq 3$. For systems with multiple chains, one would include both interchain and intrachain contacts in counting the total number of native contacts in the native contact list. A given PDB structure for a protein with primary sequence of length roughly 100 amino acids typically has about 1700 native contacts when using the above definition. In the thermally equilibrated folded structural ensemble obtained from molecular dynamics simulations, the mean number of total contacts is somewhat smaller than the number in the PDB structure (e.g., about 1600), since some of the more weakly stable contacts tend to be broken simply due to thermal fluctuations. We will now quantify this equilibrium thermal average number of contacts, since it will be used as the reference structural ensemble in which to count native contacts.

Let us choose to take the number of native contacts as the collective coordinate used in a biasing potential function to partially unfold a protein. Even in the presence of non-native contacts, native contacts have been shown to determine the folding mechanism in both coarse-grained[70] and all-atom[71] simulations. In practice, the native contacts to consider for the biasing simulation are generated from the last 150 ns of an equilibrated ensemble starting from the PDB structure. We sample every 20 ps to obtain 7500 frames for the native ensemble. All contacts appearing with frequency ≥5% in this ensemble are included in a native contact list. This contact list will include some contacts not present in the PDB contact list, and it will have lost some other contacts that were initially present in the PDB structure.[72]

Biasing in molecular dynamics is most easily done when the collective coordinate is a continuous function of the atomic positions. We thus define a contact function $Q_{ij}(r_{ij})$ for each heavy-atom pair i, j in the list of native contacts. $Q_{ij}(r_{ij})$ is a function of the distance between the atom pair i, j, as follows:

$$Q_{ij}(r_{ij}) = \frac{1 - \left(\frac{r_{ij}}{r_0}\right)^n}{1 - \left(\frac{r_{ij}}{r_0}\right)^m} \quad (1)$$

Here, r_{ij} is the distance between atoms i and j in any arbitrary structure, and we take $r_0 = 4.8$ Å, $n = 6$, and $m = 12$. This function vs distance r_{ij} is plotted in the inset of Figure 3B. Since any

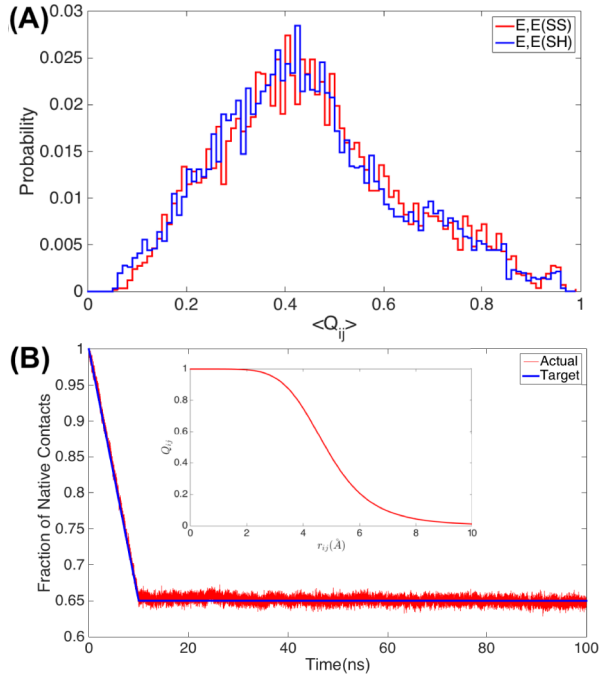


Figure 3. (A) Distribution of thermal-averaged probabilities of native contacts, $\langle Q_{ij}(r_{ij})\rangle_{nat}$, in the equilibrium native ensemble of E,E(SS) and E,E(SH) SOD1. (B) (Inset) Plot of Q_{ij} in eq 1 as a function of distance r_{ij}. (Main panel) Plot of $Q(t)$ given in eq 2 (red curve subject to thermal fluctuations) and the target $Q_c(t)$ (piecewise-smooth blue curve) vs time for a typical biasing simulation of E,E(SS)-SOD1 monomer.

contact (again defined by heavy atom pairs within a 4.8 Å) with probability ≥5% in the native ensemble is included in the native contact list, different native contacts (i, j) will have a distribution of equilibrium values of $\langle Q_{ij}(r_{ij})\rangle_{nat}$, as shown in Figure 3A.

There are many other functions that have a similar form as shown in the inset in Figure 3B. Any such function that goes from 1 to 0 as r goes from 0 to ∞ with a characteristic length scale of r_0 will work for this purpose. We choose the above parameters for r_0, n, and m to characterize a continuous function with the approximate range of physical interactions in the protein. Note that the set of contacts $\{Q_{ij}\}$ is used here simply as an order parameter, while the actual internal energy function of the protein governs the Boltzmann occupancies of structures with a given value of this order parameter.

As mentioned above, we use a continuous function $Q_{ij}(r_{ij})$ with well-defined derivative to weight contacts (rather than a Heaviside or discrete step function), because we must be able to apply a biasing force to atoms involved in native contacts as a continuous function of r_{ij} during the molecular dynamics simulation. The collective coordinate Q for any structure characterized by the set of pairwise distances $\{r_{ij}\}$ between heavy atoms is then defined by the equation

$$Q = \frac{\sum_{ij}^{N} Q_{ij}(r_{ij})}{\sum_{ij}^{N} \langle Q_{ij}(r_{ij})\rangle_{nat}} \quad (2)$$

In eq 2, Q_{ij} is given as in eq 1, the sum $\sum_{ij}^{N}$ is over the native contact list generated from the native equilibrium ensemble, and

DOI: 10.1021/acs.jpcb.8b07680
J. Phys. Chem. B 2018, 122, 11662−11676

the quantity in the denominator is the thermal average of the sum of the Q_{ij} values in the native equilibrium ensemble. This is the integral under either of the histograms in Figure 3A. For example, for E,E(SS) SOD1 equilibrated starting from PDB structure 1RK7, $\sum_{ij}^{N}\langle Q_{ij}(r_{ij})\rangle_{nat} = 2626$. The numerator of eq 2 is the sum of Q_{ij} in an arbitrary structure; Q in eq 2 is typically a number between zero and unity.

The distribution of $\langle Q_{ij}(r_{ij})\rangle_{nat}$ is very similar for E,E(SS) and E,E(SH) SOD1 (Figure 3A). There are only about 17 more contacts in the E,E(SS) native ensemble. This modest difference is likely a consequence of the CHARMM22* potential. We expect the mean contact probability, $\frac{1}{N}\sum_{ij}^{N}\langle Q_{ij}(r_{ij})\rangle_{nat}$, to be larger in the E,E(SS) native ensemble, and indeed, this is the case but again only modestly: 0.45 in E,E(SS) vs 0.44 in E,E(SH).

Implementation of the Collective Coordinate Method. Each system was initially equilibrated for 600 ns; during the last 50 ns, $\sum_{ij}^{N}\langle Q_{ij}\rangle$ was measured to obtain the denominator in eq 2 (this quantity remained stable and converged over this time window). To produce a partially disordered protein ensemble, a global bias is implemented to partially unfold the protein, as a time-dependent potential of the form

$$V(Q, t) = \frac{1}{2}k(Q - Q_c(t))^2 \tag{3}$$

where $Q_c(t)$ is a linearly decreasing function of time, which starts from a value corresponding to the folded equilibrium ensemble, and then linearly decreases with time at a predetermined rate as described below.

The center of the biasing potential, $Q_c(t)$, is moved from 1 to 0.65 over 10 ns, during which time the amount of structure initially present is systematically reduced to about 65% of the original value. Afterward, the bias is held fixed at $Q_c = 0.65$, and the system is equilibrated for another 90 ns while sampling configurations every 20 ps, yielding a biased, partially folded ensemble consisting of 4500 configurations. A typical unfolding trajectory of the target value of collective coordinate $Q_c(t)$ as a function of time along with the system's value of collective coordinate Q as a function of time are both shown in Figure 3A.

The potential $V(Q, t)$ in eq 3 is implemented by adding it to the total energy of the system. The system will try to minimize its free energy, but it will take time to do so; this is one reason for the lag in Figure 3B. The other reason for the lag between the red and blue curves in Figure 3B is because there is a nonzero residual force present when the system is perturbed from the native structure, which drives the system toward the native structure, and so results in a new equilibrium value of Q that is slightly higher than Q_c in the presence of the potential V.

If the rate of decrease of Q_c is too rapid, the values of Q characterizing the system will substantially deviate from the value of the target Q_c, and the perturbation on the system due to $V(Q, t)$ will induce a highly nonequilibrium unfolding process. We wish to maintain a quasi-equilibrium (adiabatic) process as the protein unfolds. The rate of decrease for $Q_c(t)$ is thus determined by the condition that the actual Q is not too much different from the target Q_c. In practice, we found that differences of $\lesssim$0.5% on average could be achieved by rather rapid rates of about $dQ/dt \leq 10^8\ s^{-1}$, i.e., a Q that decreased from 1 to 0.65 over ~10 ns. The above lag between Q and Q_c also depends on the "spring constant" in eq 3, described further below. A quasi-static (adiabatic) perturbation yields an unfolding process that is governed primarily by the interactions within the system, rather than the response to perturbing forces that may be much larger than the stabilizing forces in the system.

In our simulations of SOD1, we used 1.11×10^5 kJ/mol for the "spring constant" k in eq 3. This value of spring constant k gave small values of the lag in Q during protein unfolding as described above. It also results in a mean force due to variations in the collective coordinate that for any given atom is actually rather small. To see this, note that the variation of Q in Figure 3 for 10 ns < t < 100 ns is about $\Delta Q \approx 0.005$, so the "restoring force" per atom due to changes in Q is approximately $k\Delta Q/N_{atom}$ or about 0.5 kJ/mol $\approx 0.25k_BT$ for SOD1 (note the collective force due to variations in Q rather than distance has units of kJ/mol).

Identification of Local Unfolding-Specific Epitopes. As the protein is globally biased to unfold, it does not unfold homogeneously, but rather it spontaneously unfolds locally in more weakly stable regions. It is these weakly stable regions that we are interested in predicting. Under the hypothesis that weakly stable regions in the native structure are likely to be exposed in misfolded forms of the protein, these regions constitute candidate MSEs, which may be exploited for diagnostic or therapeutic applications.

For a given protein structure, we perform a number of independent biasing simulations (typically 10) to ensure that protein regions that are observed to be exposed in a given biasing simulation are indeed consistently exposed, and not the result of a rare random fluctuation in a particular simulation. We thus consider regions of the protein for which a significant fraction f of the simulations show an increase in exposure upon biasing. In practice, we implemented 10 repeated simulations in total, and we take $f = 0.8$, corresponding to *at least* 8 of 10 simulations displaying the epitope.

We identify the locally unfolded regions using several methods:

(1) By comparing the change in solvent-accessible surface area (SASA), between the biased ensemble of structures and the initial folded equilibrium ensemble of structures before biasing. In this calculation we consider the side chain surface area for every residue except glycine, for which we use the total residue surface area (which amounts to the backbone surface area for glycine). The surface area buried by metals must also be included for the holo reference states. A region with significantly increased surface exposure is identified as an MSE.

(2) By comparing the change in folded-ensemble native contacts (Q) between the biased ensemble of structures and the initial equilibrium ensemble of structures before biasing. Like SASA above, native contacts are considered as a function of amino acid sequence. A region with significantly decreased number of native contacts is identified as an MSE.

(3) By comparing the change in root mean-squared fluctuations (RMSF) between the biased and initial ensembles. A region with significantly increased RMSF is identified as an MSE.

Method 1 (SASA) is a natural choice for monitoring the exposure of an epitope and consequent accessibility to antibody binding. Method 2 (Q) is also a natural choice because the order parameter used to identify MSEs is then the same as the one used to bias the system. Method 3 (RMSF) is a natural order parameter to measure the increase in dynamics of particular regions of the protein. Locally unfolded regions predicted with

DOI: 10.1021/acs.jpcb.8b07680
J. Phys. Chem. B 2018, 122, 11662−11676

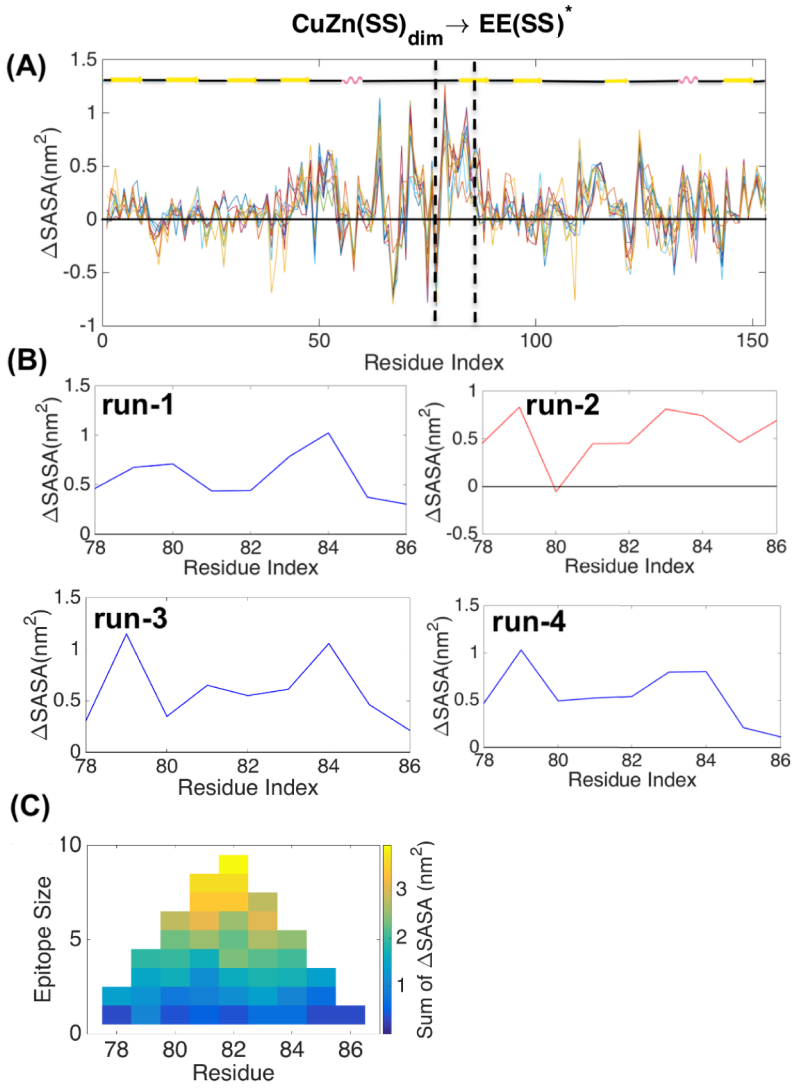


Figure 4. (A) Change in solvent-accessible surface area, ΔSASA, upon global bias to $Q = 0.65$ for the $CuZn(SS)_{dim} \rightarrow EE(SS)^*$ transition (see Table 1). ΔSASA is plotted vs residue index. Each of the 10 curves represents one simulation. (B) Illustration of the method used to identify MSEs. Four of 10 simulations of the 9 amino acid segment containing residues 78−86, a predicted MSE delineated by dashed lines in panel A, are magnified here for illustration. Simulations where all residues in the segment satisfy ΔSASA ≥ 0 are shown as blue lines (runs 1, 3, and 4), while those not satisfying the criterion are shown in red lines (run 2). (C) The "fireplot" of total SASA change within MSE 78−86 upon biasing to $Q = 0.65$ (see text).

the above metrics may be compared with experimental results measuring hydrogen exchange,[73] discussed further below.

The MSE prediction depends significantly on the reference state to which the stressed protein is being compared. For SOD1, several alternatives may be chosen for both the biased species and the reference species, with different physical interpretations resulting from each choice. Here we consider the cases given in Table 1, which also introduces a shorthand notation for each of the transformations used to predict misfolding-specific epitopes (MSEs), wherein stressed states are denoted with asterisks. The newly exposed MSEs predicted from the $CuZn(SS)_{mon} \rightarrow EE(SH)^*$ monomer, for example, result from the processes of metal loss, disulfide reduction, and partial local unfolding, but not monomerization.

RESULTS AND DISCUSSION

Identifying Misfolding-Specific Epitopes. Figure 4A shows a plot of the change in solvent-accessible surface area (ΔSASA) of the partially unfolded E,E(SS) SOD1 monomer, compared to a monomer in the native Cu,Zn(SS) dimer. ΔSASA is plotted vs residue index. The partially unfolded E,E(SS) monomer was biased to $Q = 0.65$, as described in the Computational Methods section. There are 10 curves in this plot, each of which corresponds to a separate biasing simulation for this monomer.

The N-terminal 38 residues show little change upon transformation from $CuZn(SS)_{mon} \rightarrow EE(SS)^*$; however, regions from residues 39−153 show significant positive or negative changes, superimposed on a general upward shift. The general upward shift in the mean is about 0.11 nm^2, while the standard deviation about this mean shift, due to differential solvent exposure of different protein regions, is 0.28 nm^2. The standard deviation simulation-to-simulation for the same residue is, on average, only 0.01 nm^2.

Figure 4B focuses on an exemplary segment from Figure 4A, residues 78−86, delimited by the vertical dashed lines, which was predicted as an MSE. Four of the 10 simulations are shown. For this potential MSE, 3 of the 4 runs shown (1, 3, and 4) satisfy the MSE selection criterion of ΔSASA > 0 for *all* residues contained in the MSE, while run 2 fails the selection criterion for

DOI: 10.1021/acs.jpcb.8b07680
J. Phys. Chem. B 2018, 122, 11662−11676

residue 80. In practice, we account for stochastic fluctuations by employing a criterion where ΔSASA > 0 must be satisfied for at least 8 of 10 simulation runs for all residues in an MSE of a given length in order to be selected as a prediction. The criterion is examined for the largest MSEs first; the MSE size is then progressively decreased in increments of one amino acid down to a length of 3 residues.

Figure 4C shows a "fireplot" of the increase in SASA upon biasing to $Q_c = 0.65$, specifically for the group of 9 amino acids spanning residues 78−86. The x-axis indicates the central residue index of a predicted MSE. For even residue-length epitopes, the residue on the x-axis is taken by convention to be the residue just left of center. The y-axis of Figure 4C indicates the sequence length of the MSE, i.e., the number of residues of the potential candidate MSE. For example, there is one epitope of length 9 that satisfied ΔSASA > 0 for all residues in at least 8 of 10 simulations, centered at residues 82, corresponding to amino acids 78−86.

The color coding in each rectangle shown in Figure 4C gives the change in surface area of the group of underlying residues characterized by the length on the y-axis and center position on the x-axis. The mean value over runs of ΔSASA for contiguous strings of amino acids within a given region is color-coded; i.e., for a given predicted MSE, the color coding is given by $(1/N_{\text{runs}})\sum_1^{N_{\text{runs}}}\sum_{\alpha=1}^{L}\Delta\text{SASA}_\alpha$, where the inner sum is over L residues contained within the epitope, and the outer average is over simulation runs. For example, ⟨ΔSASA⟩ for MSE $(x, y) = (82,9)$ is 3.92 nm^2, whereas ⟨ΔSASA⟩ of the region $(x, y) = (81, 6)$ subsumed by the MSE is 3.11 nm^2.

Misfolding-Specific Epitope Predictions. Figure 5 and Table 2 give the MSE predictions based on ΔSASA exposure, using the criterion that at least 3 contiguous residues must show increased SASA for at least 8/10 simulations. Figure 5A shows the MSE predictions for the stressed E,E(SH) monomer, using a monomer from the Cu,Zn(SS) SOD1 dimer as a reference state. Peaks in the fireplot correspond to the size and location of predicted epitopes. There are 12 epitopes of length ≥ 3 predicted; the largest epitope is 10 residues along, centered between residues G82 and D83 (epitope 78−87 in Table 2), exposing 3.9 nm^2 of new surface area. The 12 predicted epitopes sequences are listed in Table 2.

Figure 5B shows a fireplot giving the MSE predictions using CuZn(SS)$_{\text{mon}}$ SOD1 as the reference state. Only seven epitopes are now predicted (Table 2). Four epitopes, 3−8, 16−18, 50−53, and 148−152, involve exposure of the dimer interface. These are specific to the dimeric reference state. The latter C-terminal MSE is consistent with a SOD1 exposed dimer interface (SEDI) epitope consisting of residues 145−151, to which antibodies have been raised to selectively identify misfolded SOD1.[74] There also are two additional predicted MSEs that are specific to the dimer reference state: 95−97 and 138−140. These are not in the dimer interface but are allosterically sequestered from solvent by dimerization. We found no epitopes present in the CuZn(SS)$_{\text{mon}}$ reference that were not also present in the CuZn(SS)$_{\text{dim}}$ reference.

Figure 5C shows a fireplot of the MSE predictions using monomeric E,E(SS) SOD1 as the reference state. Only 2 epitopes are predicted, and these arise solely from local unfolding, cf., Table 1. Epitope 46−49 overlaps with a longer MSE predicted using holo monomer or dimer as a reference (Table 2). Epitope 58−62, indicated by an arrow in Figure 5C, is specific to the E,E(SS) reference state. Evidently, this epitope is more exposed in the holo state than the apo state. Structural

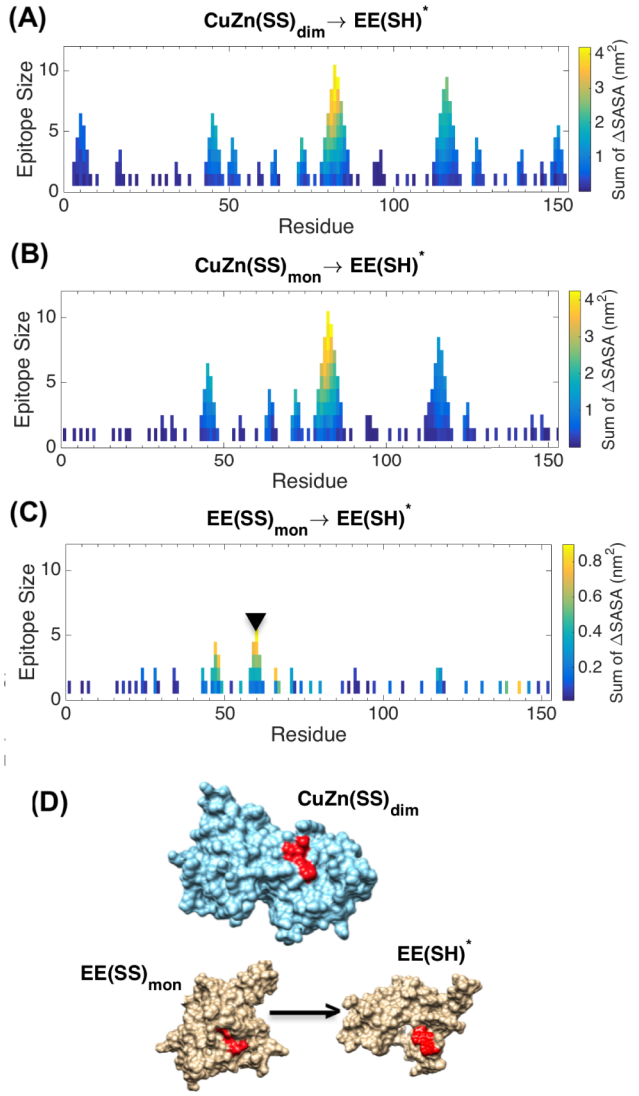


Figure 5. Fireplots giving the MSEs for SOD1 E,E(SH) monomer, as predicted from ΔSASA relative to different reference states: (A) a monomer in the context of the Cu,Zn(SS) SOD1 dimer, (B) isolated Cu,Zn(SS) monomer, and (C) E,E(SS) monomer. The novel MSE (58−62), which is not present for the metalated reference states, is marked with an arrow. (D) Centroid configurations of the CuZn(SS) dimer and EE(SS) monomer as indicated. The novel MSE (58−62) predicted in panel C is shown in red and is more exposed in the dimer. A snapshot showing exposure of the epitope from EE(SS) monomer to the stressed E,E(SH) monomer is also shown.

comparison of this epitope in the holo dimer and apo monomer indeed shows that this is the case (Figure 5D): ΔSASA = −1.0 nm^2 in going from the holo dimer to the apo monomer.

The epitopes 46−49 and 58−62 are not due to disulfide bond reduction, because EE(SS)$_{\text{mon}}$ → EE(SS)* also predicts them (Table 2). As well, comparison of CuZn(SS)$_{\text{mon}}$ → EE(SH)* and CuZn(SS)$_{\text{mon}}$ → EE(SS)* in Table 2 also shows significant overlap between the predicted MSEs, indicating a minimal role of disulfide bond reduction in determining the protein regions of increased solvent exposure due to stress.

On the other hand, comparison in Table 2 of CuZn(SS)$_{\text{mon}}$ → EE(SS)* and EE(SS)$_{\text{mon}}$ → EE(SS)* shows that there are substantially more MSEs exposed when metals are lost in the process of stress. Consistent with this, several of the epitopes obtained from the CuZn(SS)$_{\text{mon}}$ → EE(SS)* transition contain

DOI: 10.1021/acs.jpcb.8b07680
J. Phys. Chem. B 2018, 122, 11662−11676

Table 2. Epitopes Predicted from ΔSASA[a]

	$CuZn(SS)_{dim}$		$CuZn(SS)_{mon}$		$EE(SS)_{mon}$	
	residue	sequence	residue	sequence	residue	sequence
EE(SH)*	3−8	KAVCVL				
	16−18	GII				
	43−48	HGFHVH	43−48	HGFHVH	46−49	HVHE
	50−53	FGDN				
					58−62	TSAGP
	63−65	HFN	63−66	HFNP		
	71−74	HGGP	71−74	HGGP		
	78−87	ERHVGDLGNV	78−87	ERHVGDLGNV		
	95−97	ADV				
	112−120	IIGRTLVVH	112−114	IIG		
			113−120	IGRTLVVH		
	124−127	DDLG	124−126	DDL		
	138−140	GNA				
	148−152	VIGIA				
EE(SS)*	2−8	TKAVCVL				
	16−18	GII				
	43−48	HGFHVH	43−48	HGFHVH	46−49	HVHE
	50−54	FGDNT				
					58−62	TSAGP
	62−64	PHF	62−64	PHF		
	71−74	HGGP	71−74	HGGP		
	78−86	ERHVGDLGN	78−86	ERHVGDLGN		
	104−106	ISL	104−106	ISL		
	112−120	IIGRTLVVH	111−113	CII		
			112−114	IIG		
			113−120	IGRTLVVH		
	124−127	DDLG				
	144−146	LAC				
	148−152	VIGIA				

[a]Stressed states, either E,E(SS) or E,E(SH) monomer, are listed left. Reference states are listed at the top.

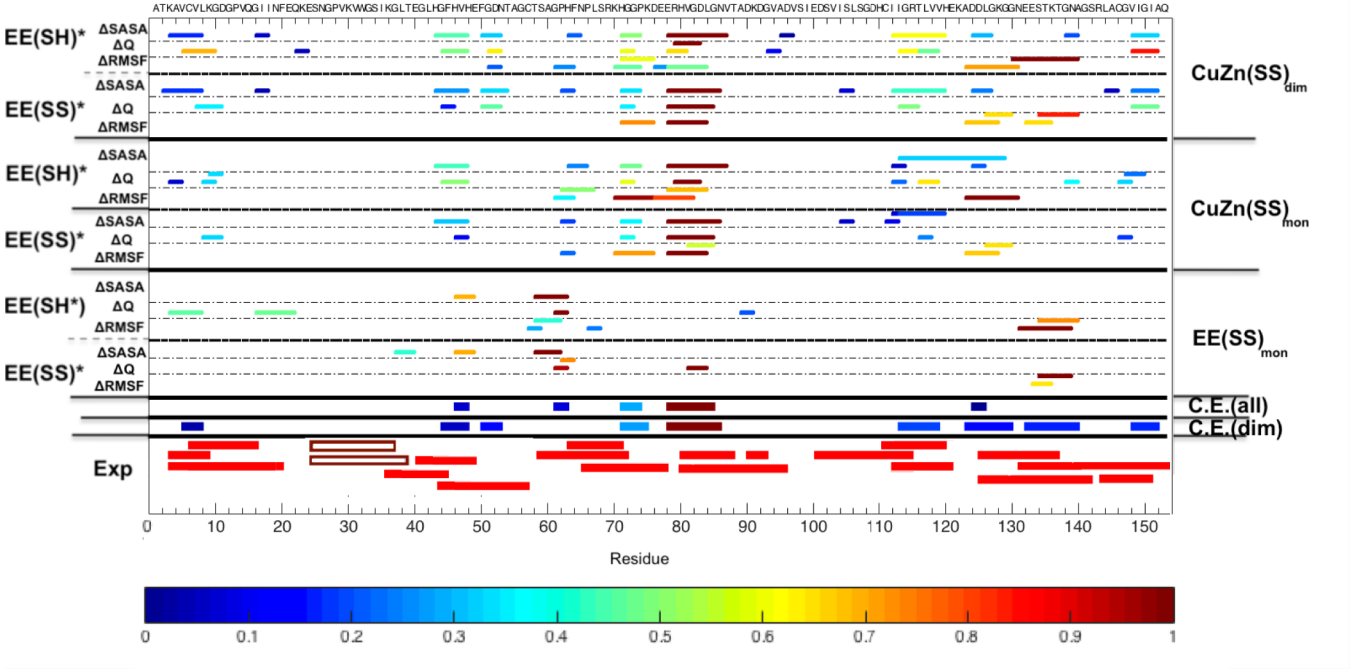


Figure 6. MSE predictions for EE(SS)* and EE(SH)* SOD1 with different metrics (ΔSASA, ΔQ, and ΔRMSF). Initial states are denoted on the right of the figure; stressed states and prediction metric are denoted on the left of the figure. C.E.(all) = consensus epitopes using all data (see text). C.E.(dim) = consensus epitopes using the $CuZn(SS)_{dim}$ reference. Exp = experimentally observed epitopes for known misfolding-specific antibodies. Open rectangles = epitopes for human peptide-specific antibodies (rather than MSEs; see text).

DOI: 10.1021/acs.jpcb.8b07680
J. Phys. Chem. B 2018, 122, 11662−11676

residues that coordinate metals, namely, epitopes (residues): 43−48 (H46, H48), 62−64 (H63), 71−74 (H71), 78−86 (H80, D83), and 113−120 (H120).

As was the case for stressed EE(SH) SOD1, comparison of $CuZn(SS)_{dim} \rightarrow EE(SS)^*$ and $CuZn(SS)_{mon} \rightarrow EE(SS)^*$ shows exposure of cryptic epitopes involving the dimer interface, which are nearly the same as those for $CuZn(SS)_{dim} \rightarrow EE(SH)^*$.

We also performed the MSE predictions for the E,E(SS) monomer using other metrics besides the increase in solvent exposure. These include the loss of local native contacts (ΔQ) and the increase in root-mean-square fluctuation (ΔRMSF). Analogous criteria are applied to these order parameters as ΔSASA: For native contacts Q, contiguous stretches of residues had to show a consistent decrease in 8/10 simulations to be considered a predicted MSE; for RMSF, contiguous stretches had to show a consistent increase above a threshold in 8/10 simulations.

Because ΔRMSF is positive for nearly all residues upon biasing, we require a threshold to apply to ΔRMSF to predict highly dynamic regions. The most natural criterion is to examine those residues that have a ΔRMSF greater than the average $\langle\Delta\text{RMSF}\rangle$. An MSE is then a contiguous stretch of residues having $\Delta\text{RMSF} > \langle\Delta\text{RMSF}\rangle$ in 8/10 simulations. The prediction results from ΔQ and ΔRMSF are listed in Tables S2 and S3, respectively.

A synthesis of the predicted MSEs using all 3 unfolding metrics ΔSASA, ΔQ, and ΔRMSF, for both stressed states EE(SS)* and EE(SH)*, and for all reference states, is given in Figure 6. Each row is defined by the stressed state and the unfolding metric (left labels), and the reference state (right labels). Predicted epitopes for each case are shown as colored bars; the color indicates the value of the metric normalized to its maximum value for each row.

The 3 different unfolding metrics address different but possibly correlated aspects of unfolding. We can calculate the significance of the Pearson correlation between the categorical prediction of an epitope (i.e., 0 or 1) across any pair of unfolding metrics (see Table 3). A correlation is calculated for two unfolding metrics by a simple binary comparison of whether or not each metric predicts an epitope for a given residue, and then by summing over residue index. The degree of correlation differs depending on the reference and stressed states. The prediction metrics generally correlate well with each other, but occasionally they differ; e.g. the comparison between Q and RMSF for $CuZn(SS)_{dim} \rightarrow EE(SH)^*_{mon}$ does not correlate. This may indicate that the disordered parts of the stressed protein are more misfolded than unfolded, since loss of native contacts does not correlate with increase in dynamics.

Table 3. Pearson Correlation (r, p) between the Predictions from Different Metrics

	EE(SH)*	EE(SS)*
$CuZn(SS)_{dim}$	(SASA:Q) = $(0.54, 4.8 \times 10^{-13})$	(SASA:Q) = $(0.53, 1.3 \times 10^{-12})$
	(SASA:RMSF) = $(0.22, 5.6 \times 10^{-3})$	(SASA:RMSF) = (0.10, 0.21)
	(Q:RMSF) = (0.04, 0.60)	(Q:RMSF) = (0.15, 0.06)
$CuZn(SS)_{mon}$	(SASA:Q) = $(0.43, 2.6 \times 10^{-8})$	(SASA:Q) = $(0.49, 1.0 \times 10^{-10})$
	(SASA:RMSF) = $(0.41, 1.4 \times 10^{-7})$	(SASA:RMSF) = $(0.38, 1.8 \times 10^{-6})$
	(Q:RMSF) = (0.04, 0.67)	(Q:RMSF) = $(0.33, 2.9 \times 10^{-5})$
$EE(SS)_{mon}$	(SASA:Q) = $(0.24, 3.4 \times 10^{-3})$	(SASA:Q) = $(0.37, 3.4 \times 10^{-6})$
	(SASA:RMSF) = $(0.27, 9.2 \times 10^{-4})$	(SASA:RMSF) = (−0.05, 0.50)
	(Q:RMSF) = (0.09, 0.25)	(Q:RMSF) = (−0.03, 0.70)

Consensus Misfolding-Specific Epitopes. Scanning across the primary sequence, some regions have a much higher tendency to contain MSEs than others. We can quantify the likelihood that a given residue is part of a predicted epitope, to obtain a "consensus map" of epitopes across the primary sequence. To this end, Figure 7 focuses on a the region in Figure 6 between residues 69 and 90, which contains two consensus epitopes.

For each residue index i, we sum over the $3 \times 6 = 18$ cases shown in Figures 6 and 7, to obtain a total number $\Omega_i = \sum_{\alpha=1}^{18} n_{i\alpha}$. If a case α contains an MSE at that position i, we add a number $n_{i\alpha}$ between 0 and 1, depending on the magnitude of the normalized value given to the epitope (represented by the color of the bars in Figures 6 and 7). If more than one overlapping epitope is predicted for residue i in a given case, we take the largest value. If a case does *not* contain an MSE at position i, we assign $n_{i\alpha} = -1/3$ for that case. The value $-1/3$ is chosen simply by convention so that no epitope prediction for *any* of the three metrics would have equal negative weight as a single epitope prediction with maximal positive weight. Plots of $n_{i\alpha}$ as a function of each case α are shown for two residue indices in Figure 7.

A plot of Ω_i vs residue index i is shown in Figure 7. Consensus epitopes are defined as the regions where $\Omega_i > 0$ and are shown color-coded in Figure 6. The indices and sequences are given in Table 4. Figure 6 and Table 4 also give the epitopes predicted by this method but using only the 6 cases having the most mature form of the protein, $CuZn(SS)_{dim}$, as the reference state.

COMPARISON WITH EXPERIMENTS

Experimental Measures of Dynamics and Solvent Exposure. Support for the prediction method may be obtained by comparing the prediction methods to experimental quantities. Figure 8A plots the simulated amino acid solvent exposure in the equilibrium ensemble of EE(SS) SOD1, vs residue index (blue curve). These numbers agree reasonably well with experimentally measured H/D exchange rates for E,E(SS) C6A/C111S/F50E/G51E SOD1[75] (red curve) (r = 0.42, $p = 4\times 10^{-8}$), with peaks for both exchange rate and SASA occurring in the same locations. On the other hand, the epitopes predicted on the basis of the additional process of stressing the monomer as a proxy for anomalous environmental conditions, $EE(SS)_{mon} \rightarrow EE(SS)^*$, would not have been obtained from equilibrium H/D exchange: Comparing the simulated ΔSASA of this process with H/D exchange rates, the mean correlation/significance and standard deviation from the 10 independent simulation runs described in the Computational Methods section is ($r = 0.01 \pm 0.09$, $p = 0.5 \pm 0.3$).

Further supporting computational predictions of native dynamics, simulated equilibrium dynamical fluctuations (RMSF) agree well with experimental measurements of the dynamics of E,E(SS) C6A/C111S/F50E/G51E/E133Q SOD1[50] (Figure 8B, $r = 0.55$, $p = 5 \times 10^{-11}$). The latter are obtained from the ratio of spectral density functions $J(\omega_H)/J(\omega_N)$, which measures dynamics that are fast compared with the tumbling rate. This correlation is substantially weakened when comparing ΔRMSF obtained from the $EE(SS)_{mon} \rightarrow EE(SS)^*$ transition with $J(\omega_H)/J(\omega_N)$: The mean correlation/signifi-

DOI: 10.1021/acs.jpcb.8b07680
J. Phys. Chem. B 2018, 122, 11662−11676

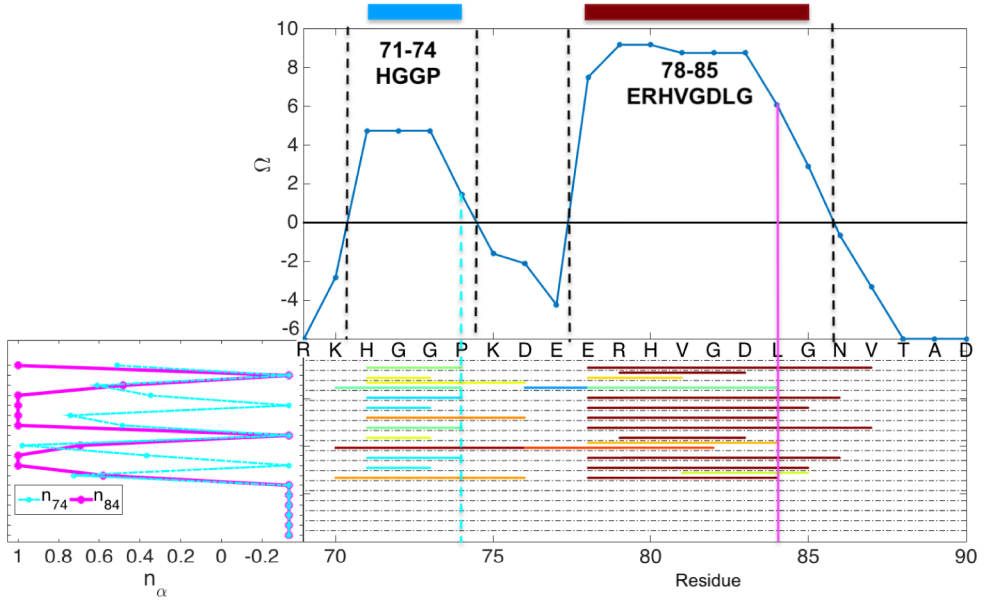


Figure 7. Closeup of residues 69−90, illustrating the method of finding consensus epitopes according to the predictions from different metrics (see text). The top figure shows the values of Ω_i for each residue i of this segment, where two consensus epitopes (71−74 and 78−85) are identified. The left panel shows the normalized metric value $n_{i,\alpha}$ at residues i = 74 and i = 84 as a function of each prediction metric.

Table 4. Consensus MSEs[a]

all references		$CuZn(SS)_{dim}$ reference	
residue	sequence	residue	sequence
		5−8	VCVL
46−48	HVH	44−48	GFHVH
		50−53	FGDN
61−63	GPH		
71−74	HGGP	71−75	HGGPK
78−85	ERHVGDLG	78−86	ERHVGDLGN
		113−119	IGRTLVV
124−126	DDL	123−130	ADDLGKGG
		132−140	EESTKTGNA
		148−152	VIGIA

[a]Left side of the table: consensus MSEs from all predictions, including all reference and stressed states. Right side of the table: consensus MSEs from all predictions having $CuZn(SS)_{dim}$ as the reference state.

cance and standard deviations from the 10 independent simulation runs are (r = 0.4 ± 0.1, p = (0.7 ± 2) × 10^{-3}).

Local unfolding of EE(SS) A4V mutant SOD1 has been investigated by H/D exchange-mass spectrometry.[76] Interestingly, a key destabilized region consists of residues 50−53 (FGDN), which is consistent with one of the consensus epitopes predicted by the collective coordinates method. On the other hand, we do not observe by any of our metrics local destabilization of the larger region containing residues 21−53 observed by Shaw et al.[76] or in other studies investigating mutant Cu,Zn-metalated variants.[77] It is likely that destabilizing mutations lengthen locally disordered regions, and possibly expose additional epitopes.

Experimentally Determined Disease-Specific Epitopes (DSEs). Several antibodies have been found to bind only to SOD1 when insoluble or in inclusions. The corresponding epitopes for these antibodies are said to be disease-specific epitopes (DSEs); a collection of DSEs from the literature are given in Table 5. The predicted *consensus* MSEs (Table 4) all have at least partial overlap with the experimental DSEs. The

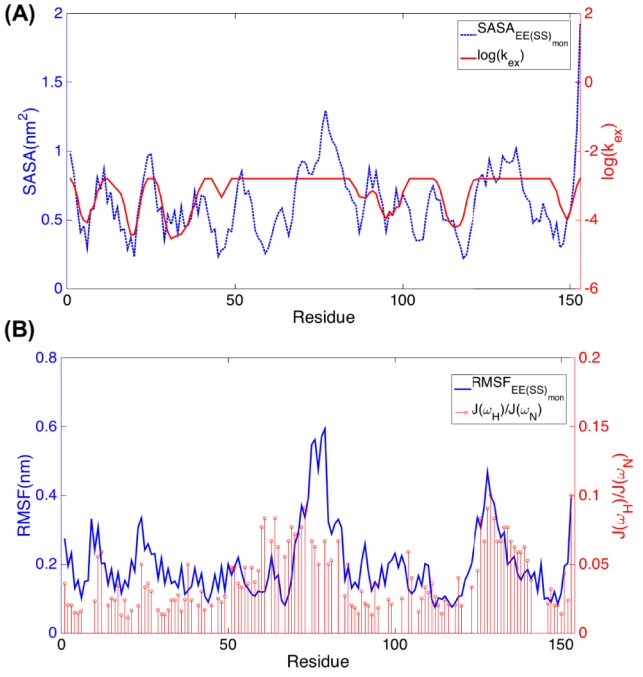


Figure 8. Equilibrium simulated quantities of exposure and dynamics for E,E(SS) SOD1 correlated with experimental measurements. (A) Simulated SASA of each residue (dashed blue line) and experimental H/D exchange rate (solid red line). (B) Simulated RMSF of each residue (blue line) and experimental ratio of spectral density functions $J(\omega_H)/J(\omega_N)$ (red lines).

predicted consensus MSEs using the $CuZn(SS)_{dim}$ reference overlap significantly with DSE1−DSE7 as well as the epitope 80−88 in ref 78. However, the epitopes we predict in Table 4 are generally more restricted in sequence than the experimental DSEs, which are often obtained by lower-resolution peptide mapping assays.

We note that the polyclonal antibodies with epitopes 24−36[79] and 24−39[80] in Table 5 were selected to distinguish the human

DOI: 10.1021/acs.jpcb.8b07680
J. Phys. Chem. B 2018, 122, 11662−11676

Table 5. Antibodies and/or Misfolding-Specific Epitopes in SOD1

antibody (epitope name)	residues	sequence	relevant refs
(DSE4)	3−9	KAVCVLK	Cashman et al. (2007)[82]
3−20 polyclonal Ra-ab	3−20	KAVCVLKGDGPVQGIINF	Jonsson et al. (2004)[80]
MS785(Derlin-1)	6−16	CVLKGDGPVQG	Fujisawa et al. (2012)[83]
antihuman polyclonal[a]	24−36	CESNGPVKVWGSIK	L.I. Bruijn et al. (1997)[79]
antihuman poly Ra-ab[a]	24−39	CESNGPVKVWGSIKGLT	Jonsson et al. (2004)[80]
5C6 (DSE5)	35−45	IKGLTEGLHGF	Cashman et al. (2007)[82]
HuMab $16_{L\text{-}40}$	40−47	EGLHGFHV[b]	Broering et al. (2013)[78]
USOD polyclonal	42−48	LHGFHVH[b]	Kerman et al. (2010)[84]
(DSE7)	41−48	GLHGFHVH[b]	Cashman et al. (2007)[82]
HuMab $3_{L\text{-}42}$	42−49	LHGFHVHE[b]	Broering et al. (2013)[78]
43−57 polyclonal Ra-ab	43−57	HGFHVHEFGDNTAGC	Jonsson et al. (2004)[80]
58−72 polyclonal Ra-ab	58−72	TSAGPHFNPLSRKHG	Jonsson et al. (2004)[80]
HuMab $37_{L\text{-}63}$	63−71	HFNPLSRKH	Broering et al. (2013)[78]
(DSE3)	65−78	NPLSRKHGGPKDEE	Cashman et al. (2007)[82]
HuMab $11_{L\text{-}80}$	80−88	HVGDLGNVT	Broering et al. (2013)[78]
80−96 polyclonal Ra-ab	80−96	HVGDLGNVTADKDGVAD	Jonsson et al. (2004)[80]
C4F6	90−93	DKDG	Ayers et al. (2014)[85]
100−115 polyclonal Ra-ab	100−115	EDSVISLSGDHCIIGR	Jonsson et al. (2004)[80]
(DSE6)	110−120	HCIIGRTLVVH	Cashman et al. (2007)[82]
HuMab $33_{L\text{-}112}$	112−121	IIGRTLVVHE	Broering et al. (2013)[78]
pAb-SOD1$^{125-137}$	125−137	DLGKGGNEESTKT	Vande Velde et al. (2008)[86]
3H1 (DSE2)	125−142	DLGKGGNEESTKTGNAGS	Cashman et al. (2007),[82] Grad et al. (2011),[81] (2014)[9]
131−153 polyclonal Ra-ab	131−153	NEESTKTGNAGSRLACGVIGIAQ	Jonsson et al. (2004),[80] Forsberg et al. (2010)[87]
SEDI polyclonal (DSE1)	143−151	RLACGVIGI	Cashman et al. (2007),[82] Rakhit et al. (2007)[74]

[a]tg mouse antihuman Abs (see note in text). [b]These epitopes are combined in Figure 6 as one epitope spanning 40−49.

SOD1 sequence from that of the mouse and, in this sense, they are unrelated to the likelihood of that region to misfold. However, these antibodies (as well as other antibodies to a region overlapping with DSE2) were observed to stain inclusions in huG85R transgenic mice, so we include these epitopes here and in Figure 6. We also note that some variants of the experimental epitopes are further altered to contain oxidatively modified residues when selecting for antibody clones; for example, the cysteine (C) residue in an oxidized variant of DSE1 is oxidized to cysteine sulfinic acid or cysteic acid.[81]

Misfolding-Specific Epitopes from Stressed Proteins Are Distinct from Equilibrium Measures. As a control study to test the necessity of the collective coordinates method, we have performed the same method for predicting MSEs, but without stressing the protein using collective coordinates; e.g., we find the epitopes predicted from the transitions $EE(SS)_{mon} \rightarrow EE(SH)_{mon}$ and $EE(SS)_{mon} \rightarrow EE(SS)_{mon}$, but with no biasing potential applied to the protein in the final state. Epitopes are again found using the metrics ΔSASA, ΔQ, and ΔRMSF. We analyze the above two transitions because they directly address stressing vs nonstressing as opposed to the additional processes of monomerization and metal removal. The epitopes found this way are less robust than those found by stressing the protein; this mandates longer equilibration times for the reference state. We determined the equilibration time based on two conditions: (1) that the trivial transition $EE(SS)_{mon} \rightarrow EE(SS)_{mon}$ predicts no epitopes, and (2) that the epitopes predicted by the transition $EE(SS)_{mon} \rightarrow EE(SH)_{mon}$ have converged as a function of the equilibration time of the reference state $EE(SS)_{mon}$. Criterion 1 mandated an equilibration time $\gtrsim$500 ns; criterion 2 mandated an equilibration time $\gtrsim$700 ns.

The results are shown in Figure S1. The epitopes predicted from the control equilibration transformation are generally different from those predicted from the stressed transformation. No epitopes are predicted for the transformation $EE(SS)_{mon} \rightarrow EE(SH)_{mon}$ when ΔQ is used as a metric, which does not allow a correlation coefficient to be calculated. When ΔSASA is used as a metric, the correlation is insignificant: $(r, p) = (0.08, 0.3)$; when ΔRMSF is used as a metric, there is only a weak correlation (which however shows significance): $(r, p) = (0.2, 4 \times 10^{-3})$. The correlations are summarized in Table S4.

Comparison between Two Different Salts. Figure 9 compares the epitopes predicted for two different salts, KCl and NaCl. We see from the figure that the epitope predictions are nearly the same, but that the effect of NaCl is to further destabilize the protein resulting in larger disordered regions, particularly in the epitopes centered around residue 5 and residue 95. This destabilizing effect of NaCl vs KCl is behind the

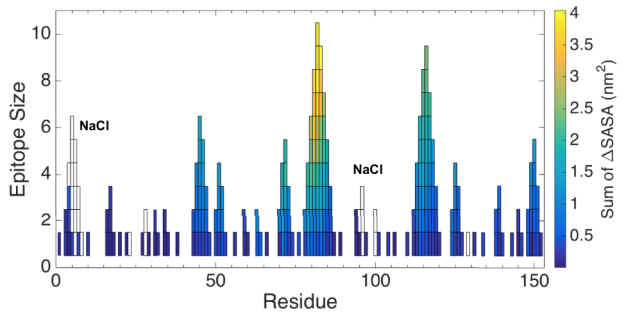


Figure 9. Fireplot of epitope predictions using ΔSASA for Cu,Zn-$(SS)_{dim} \rightarrow EE(SH)^*$, where the solution contains either 100 mM KCl (color) or 100 mM NaCl (open labeled boxes). The predictions are nearly identical except for the two labeled epitopes, which have increased in length.

DOI: 10.1021/acs.jpcb.8b07680
J. Phys. Chem. B 2018, 122, 11662−11676

"salting out" of proteins seen *in vitro*, and is proposed as one reason for cellular sodium/potassium pumps.[59,60]

CONCLUSION

Here, we have developed a simple method for computationally predicting misfolding-specific epitopes, based on computing locally unfolded regions of a protein. This method rests on the hypothesis that regions that are the most weakly stable in the context of the native structure are the most likely to be exposed in the ensemble of misfolded structures. The results of the computational predictions are consistent with experimentally derived disease-specific epitopes, but they are more refined in that they are restricted to a subset of the experimentally determined epitopes. We have also provided "consensus epitopes" obtained by averaging across some or all of our prediction metrics.

Antibodies selective for these unfolded epitopes may be raised by standard techniques involving conjugating peptides of the predicted epitopes (rather than the protein) to an immunogen (typically BSA or KLH) and injecting into mouse or rabbit, harvesting lymphocytes, and extracting monoclonal antibodies from hybridomas. By raising such antibodies to locally unfolded regions, we select conformationally against the native structure, which may be confirmed *a posteriori* by direct screening. Misfolding-selective antibodies, or their variants such as intrabodies or nanobodies, are useful therapeutically because they target pathological misfolded protein while sparing healthy protein whose function may be vital to the cell.

To properly scaffold the epitope for optimal selectivity, cyclic peptides, linear peptides, or other protein constructs that contain the epitope sequence may be used. Conformational distinction may be determined both computationally and by direct screening, typically using ELISA or SPR.

Other important extensions of the technique developed here are to apply this technique to multichain systems, using the aggregated, proto-fibrillar structure as the "native" state. This is particularly relevant for intrinsically disordered peptides that are known to form oligomers or aggregates, such as Aβ peptide, tau protein, or α-synuclein. In these cases, negative selection against such a fibril structure by identifying locally unfolded regions can resolve the problem of plaque binding, which has been problematic in Alzheimer's therapies and resulted in dose-limitation or withdrawal of patients from clinical trials due to edema or microhemorrhages observed in patient neuroimages.

ASSOCIATED CONTENT

Supporting Information

This material is available free of charge via the Internet at http://pubs.acs.org/. The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jpcb.8b07680.

> Reparameterized partial charges on HIS63 in the CHARMM22* force field, used for CuZn(SS) SOD1; predicted MSEs using the metrics ΔQ and ΔRMSF; MSE predictions using native equilibrated ensembles rather than stressed, partially unfolded ensembles; and correlation coefficients between the collective coordinates method and the equilibration method for different metrics and transitions (PDF)

AUTHOR INFORMATION

Corresponding Author

*E-mail: steve@phas.ubc.ca. Phone: 604-822-8813.

ORCID

Steven S. Plotkin: 0000-0001-8998-877X

Notes

The authors declare the following competing financial interest(s): ProMIS Neurosciences has an option to develop the Collective Coordinates technology described here, owned by the University of British Columbia. Dr. Plotkin and Dr. Cashman are Chief Physics Officer and Chief Scientific Officer, respectively, of ProMIS Neurosciences.

ACKNOWLEDGMENTS

This work was supported by Canadian Institutes of Health Research Transitional Operating Grant 2682, and the Alberta Prion Research Institute, Research Team Program Grant PTM13007. We also acknowledge WestGrid (www.westgrid.ca) and Compute Canada/Calcul Canada (www.computecanada.ca) for providing computing resources. The authors thank Will C. Guest and Eric Mills for their valuable feedback, particularly in the beginning stages of this project. The authors also thank Miguel Garcia for helpful advice on using the Gaussian quantum chemistry package.

REFERENCES

(1) Chiti, F.; Dobson, C. M. Amyloid formation by globular proteins under native conditions. *Nat. Chem. Biol.* **2009**, *5*, 15−22.

(2) Olofsson, A.; Ippel, J. H.; Wijmenga, S. S.; Lundgren, E.; Öhman, A. Probing solvent accessibility of transthyretin amyloid by solution NMR spectroscopy. *J. Biol. Chem.* **2004**, *279*, 5699−5707.

(3) Hoshino, M.; Katou, H.; Hagihara, Y.; Hasegawa, K.; Naiki, H.; Goto, Y. Mapping the core of the β2-microglobulin amyloid fibril by H/D exchange. *Nat. Struct. Biol.* **2002**, *9*, 332−336.

(4) Elam, J. S.; Taylor, A. B.; Strange, R.; Antonyuk, S.; Doucette, P. A.; Rodriguez, J. A.; Hasnain, S. S.; Hayward, L. J.; Valentine, J. S.; Yeates, T. O.; et al. Amyloid-like filaments and water-filled nanotubes formed by SOD1 mutant proteins linked to familial ALS. *Nat. Struct. Mol. Biol.* **2003**, *10*, 461−467.

(5) Nordlund, A.; Leinartaitė, L.; Saraboji, K.; Aisenbrey, C.; Gröbner, G.; Zetterström, P.; Danielsson, J.; Logan, D. T.; Oliveberg, M. Functional features cause misfolding of the ALS-provoking enzyme SOD1. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 9667−9672.

(6) Paramithiotis, E.; Pinard, M.; Lawton, T.; LaBoissiere, S.; Leathers, V. L.; Zou, W.-Q.; Estey, L. A.; Lamontagne, J.; Lehto, M. T.; Kondejewski, L. H.; et al. A prion protein epitope selective for the pathologically misfolded conformation. *Nat. Med.* **2003**, *9*, 893.

(7) Rakhit, R.; Robertson, J.; Velde, C. V.; Horne, P.; Ruth, D. M.; Griffin, J.; Cleveland, D. W.; Cashman, N. R.; Chakrabartty, A. An immunological epitope selective for pathological monomer-misfolded SOD1 in ALS. *Nat. Med.* **2007**, *13*, 754.

(8) Higaki, J. N.; Chakrabartty, A.; Galant, N. J.; Hadley, K. C.; Hammerson, B.; Nijjar, T.; Torres, R.; Tapia, J. R.; Salmans, J.; Barbour, R.; et al. Novel conformation-specific monoclonal antibodies against amyloidogenic forms of transthyretin. *Amyloid* **2016**, *23*, 86−97.

(9) Grad, L. I.; Yerbury, J. J.; Turner, B. J.; Guest, W. C.; Pokrishevsky, E.; O'Neill, M. A.; Yanai, A.; Silverman, J. M.; Zeineddine, R.; Corcoran, L.; et al. Intercellular propagated misfolding of wild-type Cu/Zn superoxide dismutase occurs via exosome-dependent and-independent mechanisms. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 3620−3625.

(10) Silverman, J.; Gibbs, E.; Peng, X.; Martens, K.; Balducci, C.; Wang, J.; Yousefi, M.; Cowan, C. M.; Lamour, G.; Louadi, S. A Rational Structured Epitope Defines a Distinct Subclass of Toxic Amyloid-beta Oligomers. *ACS Chem. Neurosci.* **2018**, *9*, 1591.

DOI: 10.1021/acs.jpcb.8b07680
J. Phys. Chem. B 2018, 122, 11662−11676

(11) Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; et al. CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J. Comput. Chem.* **2010**, *31*, 671−690.
(12) Bjelkmar, P.; Larsson, P.; Cuendet, M. A.; Hess, B.; Lindahl, E. Implementation of the CHARMM force field in GROMACS: Analysis of protein stability effects from correction maps, virtual interaction sites, and water models. *J. Chem. Theory Comput.* **2010**, *6*, 459−466.
(13) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.* **1995**, *117*, 5179−5197.
(14) Snow, C. D.; Nguyen, H.; Pande, V. S.; Gruebele, M. Absolute comparison of simulated and experimental protein-folding dynamics. *Nature* **2002**, *420*, 102−106.
(15) Lindorff-Larsen, K.; Piana, S.; Dror, R. O.; Shaw, D. E. How fast-folding proteins fold. *Science* **2011**, *334*, 517−520.
(16) Lindorff-Larsen, K.; Maragakis, P.; Piana, S.; Eastwood, M. P.; Dror, R. O.; Shaw, D. E. Systematic validation of protein force fields against experimental data. *PLoS One* **2012**, *7*, e32131.
(17) Huang, J.; Rauscher, S.; Nawrocki, G.; Ran, T.; Feig, M.; de Groot, B. L.; Grubmüller, H.; MacKerell, A. D., Jr CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nat. Methods* **2017**, *14*, 71.
(18) Plotkin, S. S.; Wang, J.; Wolynes, P. G. Correlated energy landscape model for finite, random heteropolymers. *Phys. Rev. E: Stat. Phys., Plasmas, Fluids, Relat. Interdiscip. Top.* **1996**, *53*, 6271.
(19) Plotkin, S. S.; Onuchic, J. N. Understanding protein folding with energy landscape theory Part II: Quantitative aspects. *Q. Rev. Biophys.* **2002**, *35*, 205−286.
(20) Hilser, V. J.; Freire, E. Structure-based calculation of the equilibrium folding pathway of proteins. Correlation with hydrogen exchange protection factors. *J. Mol. Biol.* **1996**, *262*, 756−772.
(21) Hilser, V. J.; García-Moreno E, B.; Oas, T. G.; Kapp, G.; Whitten, S. T. A statistical thermodynamic model of the protein ensemble. *Chem. Rev.* **2006**, *106*, 1545−1558.
(22) Cashman, N. R.; Plotkin, S. S.; Guest, W. C. Methods and systems for predicting misfolded protein epitopes. International Application No. PCT/CA2009/001413, 2010.
(23) Muñoz, V.; Eaton, W. A. A simple model for calculating the kinetics of protein folding from three-dimensional structures. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96*, 11311−11316.
(24) Alm, E.; Baker, D. Prediction of protein-folding mechanisms from free-energy landscapes derived from native structures. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96*, 11305−11310.
(25) Galzitskaya, O. V.; Finkelstein, A. V. A theoretical search for folding/unfolding nuclei in three-dimensional protein structures. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96*, 11299−11304.
(26) Guerois, R.; Serrano, L. The SH3-fold family: experimental evidence and prediction of variations in the folding pathways. *J. Mol. Biol.* **2000**, *304*, 967−982.
(27) Canet, D.; Last, A. M.; Tito, P.; Sunde, M.; Spencer, A.; Archer, D. B.; Redfield, C.; Robinson, C. V.; Dobson, C. M. Local cooperativity in the unfolding of an amyloidogenic variant of human lysozyme. *Nat. Struct. Biol.* **2002**, *9*, 308.
(28) Skerra, A. Engineered protein scaffolds for molecular recognition. *J. Mol. Recognit.* **2000**, *13*, 167−187.
(29) Correia, B. E.; Ban, Y.-E. A.; Holmes, M. A.; Xu, H.; Ellingson, K.; Kraft, Z.; Carrico, C.; Boni, E.; Sather, D. N.; Zenobia, C.; et al. Computational design of epitope-scaffolds allows induction of antibodies specific for a poorly immunogenic HIV vaccine epitope. *Structure* **2010**, *18*, 1116−1126.
(30) Ofek, G.; Guenaga, F. J.; Schief, W. R.; Skinner, J.; Baker, D.; Wyatt, R.; Kwong, P. D. Elicitation of structure-specific antibodies by epitope scaffolds. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 17880−17887.
(31) Azoitei, M. L.; Ban, Y.-E. A.; Julien, J.-P.; Bryson, S.; Schroeter, A.; Kalyuzhniy, O.; Porter, J. R.; Adachi, Y.; Baker, D.; Pai, E. F.; et al. Computational design of high-affinity epitope scaffolds by backbone grafting of a linear epitope. *J. Mol. Biol.* **2012**, *415*, 175−192.
(32) Correia, B. E.; Bates, J. T.; Loomis, R. J.; Baneyx, G.; Carrico, C.; Jardine, J. G.; Rupert, P.; Correnti, C.; Kalyuzhniy, O.; Vittal, V.; et al. Proof of principle for epitope-focused vaccine design. *Nature* **2014**, *507*, 201.
(33) Bernstein, F. C.; Koetzle, T. F.; Williams, G. J.; Meyer, E. F.; Brice, M. D.; Rodgers, J. R.; Kennard, O.; Shimanouchi, T.; Tasumi, M. The protein data bank. *Eur. J. Biochem.* **1977**, *80*, 319−324.
(34) Plotkin, S. Systems and Methods for Predicting Misfolded Protein Epitopes by Collective Coordinate Biasing. International Application No. PCT/CA2016/051306, 2016.
(35) Rosen, D. R.; Siddique, T.; Patterson, D.; Figlewicz, D. A.; Sapp, P.; Hentati, A.; Donaldson, D.; Goto, J.; O'Regan, J. P.; Deng, H.-X.; et al. Mutations in Cu/Zn superoxide dismutase gene are associated with familial amyotrophic lateral sclerosis. *Nature* **1993**, *362*, 59−62.
(36) Deng, H.-X.; Hentati, A.; Tainer, J. A.; Iqbal, Z.; Cayabyab, A.; Hung, W.-Y.; Getzoff, E. D.; Hu, P.; Herzfeldt, B.; Roos, R. P.; et al. Amyotrophic Lateral Sclerosis and Structural Defects in Cu, Zn Superoxide Dismutase. *Science* **1993**, *261*, 1047−1051.
(37) Neupane, K.; Solanki, A.; Sosova, I.; Belov, M.; Woodside, M. T. Diverse metastable structures formed by small oligomers of α-synuclein probed by force spectroscopy. *PLoS One* **2014**, *9*, e86495.
(38) Redler, R. L.; Fee, L.; Fay, J. M.; Caplow, M.; Dokholyan, N. V. Non-native soluble oligomers of Cu/Zn superoxide dismutase (SOD1) contain a conformational epitope linked to cytotoxicity in amyotrophic lateral sclerosis (ALS). *Biochemistry* **2014**, *53*, 2423−2432.
(39) Khare, S. D.; Caplow, M.; Dokholyan, N. V. The rate and equilibrium constants for a multistep reaction sequence for the aggregation of superoxide dismutase in amyotrophic lateral sclerosis. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 15094−15099.
(40) Das, A.; Plotkin, S. S. Mechanical probes of SOD1 predict systematic trends in metal and dimer affinity of ALS-associated mutants. *J. Mol. Biol.* **2013**, *425*, 850−874.
(41) Lindberg, M. J.; Tibell, L.; Oliveberg, M. Common denominator of Cu/Zn superoxide dismutase mutants associated with amyotrophic lateral sclerosis: decreased stability of the apo state. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 16607−16612.
(42) Niwa, J.-i.; Yamada, S.-i.; Ishigaki, S.; Sone, J.; Takahashi, M.; Katsuno, M.; Tanaka, F.; Doyu, M.; Sobue, G. Disulfide bond mediates aggregation, toxicity, and ubiquitylation of familial amyotrophic lateral sclerosis-linked mutant SOD1. *J. Biol. Chem.* **2007**, *282*, 28087−28095.
(43) Hörnberg, A.; Logan, D. T.; Marklund, S. L.; Oliveberg, M. The coupling between disulphide status, metallation and dimer interface strength in Cu/Zn superoxide dismutase. *J. Mol. Biol.* **2007**, *365*, 333−342.
(44) Jonsson, P. A.; Graffmo, K. S.; Andersen, P. M.; Brännström, T.; Lindberg, M.; Oliveberg, M.; Marklund, S. L. Disulphide-reduced superoxide dismutase-1 in CNS of transgenic amyotrophic lateral sclerosis models. *Brain* **2006**, *129*, 451−464.
(45) Rakhit, R.; Cunningham, P.; Furtos-Matei, A.; Dahan, S.; Qi, X.-F.; Crow, J. P.; Cashman, N. R.; Kondejewski, L. H.; Chakrabartty, A. Oxidation-induced misfolding and aggregation of superoxide dismutase and its implications for amyotrophic lateral sclerosis. *J. Biol. Chem.* **2002**, *277*, 47551−47556.
(46) Zhang, H.; Andrekopoulos, C.; Joseph, J.; Chandran, K.; Karoui, H.; Crow, J. P.; Kalyanaraman, B. Bicarbonate-dependent peroxidase activity of human Cu, Zn-Superoxide Dismutase induces covalent aggregation of protein intermediacy of tryptophan-derived oxidation products. *J. Biol. Chem.* **2003**, *278*, 24078−24089.
(47) Lamb, A. L.; Torres, A. S.; O'Halloran, T. V.; Rosenzweig, A. C. Heterodimeric structure of superoxide dismutase in complex with its metallochaperone. *Nat. Struct. Biol.* **2001**, *8*, 751.
(48) Furukawa, Y.; Torres, A. S.; O'Halloran, T. V. Oxygen-induced maturation of SOD1: a key role for disulfide formation by the copper chaperone CCS. *EMBO J.* **2004**, *23*, 2872−2881.

DOI: 10.1021/acs.jpcb.8b07680
J. Phys. Chem. B 2018, 122, 11662−11676

(49) Lindberg, M. J.; Normark, J.; Holmgren, A.; Oliveberg, M. Folding of human superoxide dismutase: disulfide reduction prevents dimerization and produces marginally stable monomers. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 15893−15898.
(50) Banci, L.; Bertini, I.; Cramaro, F.; Del Conte, R.; Viezzoli, M. S. Solution structure of Apo Cu, Zn superoxide dismutase: role of metal ions in protein folding. *Biochemistry* **2003**, *42*, 9543−9553.
(51) Banci, L.; Bertini, I.; Cramaro, F.; Del Conte, R.; Viezzoli, M. S. Solution structure of Apo Cu, Zn superoxide dismutase: role of metal ions in protein folding. *Biochemistry* **2003**, *42*, 9543−9553.
(52) Krivov, G. G.; Shapovalov, M. V.; Dunbrack, R. L. Improved prediction of protein side-chain conformations with SCWRL4. *Proteins: Struct., Funct., Genet.* **2009**, *77*, 778−795.
(53) Banci, L.; Bertini, I.; Cramaro, F.; Del Conte, R.; Viezzoli, M. S. The solution structure of reduced dimeric copper zinc superoxide dismutase: the structural effects of dimerization. *Eur. J. Biochem.* **2002**, *269*, 1905−1915.
(54) Frisch, M.; Trucks, G.; Schlegel, H.; Scuseria, G.; Robb, M.; Cheeseman, J.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G.; et al. *Gaussian 09, Revision A.02*; Gaussian Inc.: Wallingford, CT, 2009.
(55) Pronk, S.; Páll, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; van der Spoel, D.; et al. GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* **2013**, *29*, btt055.
(56) Bonomi, M.; Branduardi, D.; Bussi, G.; Camilloni, C.; Provasi, D.; Raiteri, P.; Donadio, D.; Marinelli, F.; Pietrucci, F.; Broglia, R. A.; et al. PLUMED: A portable plugin for free-energy calculations with molecular dynamics. *Comput. Phys. Commun.* **2009**, *180*, 1961−1972.
(57) Piana, S.; Lindorff-Larsen, K.; Shaw, D. E. How robust are protein folding simulations with respect to force field parameterization? *Biophys. J.* **2011**, *100*, L47−L49.
(58) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926−935.
(59) Collins, K. D. Charge density-dependent strength of hydration and biological structure. *Biophys. J.* **1997**, *72*, 65−76.
(60) Vrbka, L.; Vondrášek, J.; Jagoda-Cwiklik, B.; Vácha, R.; Jungwirth, P. Quantification and rationalization of the higher affinity of sodium over potassium to protein surfaces. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 15440−15444.
(61) Pande, V. S.; Baker, I.; Chapman, J.; Elmer, S. P.; Khaliq, S.; Larson, S. M.; Rhee, Y. M.; Shirts, M. R.; Snow, C. D.; Sorin, E. J.; et al. Atomistic protein folding simulations on the submillisecond time scale using worldwide distributed computing. *Biopolymers* **2003**, *68*, 91−109.
(62) Camacho, C. J.; Thirumalai, D. Kinetics and thermodynamics of folding in model proteins. *Proc. Natl. Acad. Sci. U. S. A.* **1993**, *90*, 6369−6372.
(63) Das, A.; Sin, B.; Mohazab, A.; Plotkin, S. Unfolded protein ensembles, folding trajectories, and refolding rate prediction. *J. Chem. Phys.* **2013**, *139*, 121925.
(64) Heinkel, F.; Gsponer, J. Determination of protein folding intermediate structures consistent with data from oxidative footprinting mass spectrometry. *J. Mol. Biol.* **2016**, *428*, 365−371.
(65) Baftizadeh, F.; Biarnes, X.; Pietrucci, F.; Affinito, F.; Laio, A. Multidimensional view of amyloid fibril nucleation in atomistic detail. *J. Am. Chem. Soc.* **2012**, *134*, 3886−3894.
(66) Fiorin, G.; Klein, M. L.; Hénin, J. Using collective variables to drive molecular dynamics simulations. *Mol. Phys.* **2013**, *111*, 3345−3362.
(67) Leone, V.; Marinelli, F.; Carloni, P.; Parrinello, M. Targeting biomolecular flexibility with metadynamics. *Curr. Opin. Struct. Biol.* **2010**, *20*, 148−154.
(68) Provasi, D.; Filizola, M. Putative active states of a prototypic g-protein-coupled receptor from biased molecular dynamics. *Biophys. J.* **2010**, *98*, 2347−2355.
(69) Sutto, L.; Gervasio, F. L. Effects of oncogenic mutations on the conformational free-energy landscape of EGFR kinase. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 10616−10621.
(70) Clementi, C.; Plotkin, S. S. The effects of nonnative interactions on protein folding rates: theory and simulation. *Protein Sci.* **2004**, *13*, 1750−1766.
(71) Best, R. B.; Hummer, G.; Eaton, W. A. Native contacts determine protein folding mechanisms in atomistic simulations. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 17874.
(72) Habibi, M.; Röttler, J.; Plotkin, S. S. As Simple As Possible, but Not Simpler: Exploring the Fidelity of Coarse-Grained Protein Models for Simulated Force Spectroscopy. *PLoS Comput. Biol.* **2016**, *12*, e1005211.
(73) Bai, Y.; Sosnick, T. R.; Mayne, L.; Englander, S. W. Protein folding intermediates: native-state hydrogen exchange. *Science* **1995**, *269*, 192−197.
(74) Rakhit, R.; Robertson, J.; Velde, C. V.; Horne, P.; Ruth, D. M.; Griffin, J.; Cleveland, D. W.; Cashman, N. R.; Chakrabartty, A. An immunological epitope selective for pathological monomer-misfolded SOD1 in ALS. *Nat. Med.* **2007**, *13*, 754−759.
(75) Danielsson, J.; Kurnik, M.; Lang, L.; Oliveberg, M. Cutting off the functional loops from the homo-dimeric enzyme superoxide dismutase 1 (SOD1) monomeric β-barrels. *J. Biol. Chem.* **2011**, *286*, 33070.
(76) Shaw, B. F.; Durazo, A.; Nersissian, A. M.; Whitelegge, J. P.; Faull, K. F.; Valentine, J. S. Local unfolding in a destabilized, pathogenic variant of superoxide dismutase 1 observed with H/D exchange and mass spectrometry. *J. Biol. Chem.* **2006**, *281*, 18167−18176.
(77) Assfalg, M.; Banci, L.; Bertini, I.; Turano, P.; Vasos, P. R. Superoxide dismutase folding/unfolding pathway: role of the metal ions in modulating structural and dynamical features. *J. Mol. Biol.* **2003**, *330*, 145−158.
(78) Broering, T. J.; Wang, H.; Boatright, N. K.; Wang, Y.; Baptista, K.; Shayan, G.; Garrity, K. A.; Kayatekin, C.; Bosco, D. A.; Matthews, C. R.; et al. Identification of human monoclonal antibodies specific for human SOD1 recognizing distinct epitopes and forms of SOD1. *PLoS One* **2013**, *8*, e61210.
(79) Bruijn, L.; Becher, M.; Lee, M.; Anderson, K.; Jenkins, N.; Copeland, N.; Sisodia, S.; Rothstein, J.; Borchelt, D.; Price, D.; et al. ALS-linked SOD1 mutant G85R mediates damage to astrocytes and promotes rapidly progressive disease with SOD1-containing inclusions. *Neuron* **1997**, *18*, 327−338.
(80) Jonsson, P. A.; Ernhill, K.; Andersen, P. M.; Bergemalm, D.; Brännström, T.; Gredal, O.; Nilsson, P.; Marklund, S. L. Minute quantities of misfolded mutant superoxide dismutase-1 cause amyotrophic lateral sclerosis. *Brain* **2004**, *127*, 73−88.
(81) Grad, L. I.; Guest, W. C.; Yanai, A.; Pokrishevsky, E.; O'Neill, M. A.; Gibbs, E.; Semenchenko, V.; Yousefi, M.; Wishart, D. S.; Plotkin, S. S.; et al. Intermolecular transmission of superoxide dismutase 1 misfolding in living cells. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 16398−16403.
(82) Cashman, N. R.; Chakrabartty, A.; Rakhit, R.; Osterman, J. B. Methods and compositions to treat and detect misfolded-SOD1 mediated diseases. International Application No. PCT/CA2007/000346, 2007.
(83) Fujisawa, T.; Homma, K.; Yamaguchi, N.; Kadowaki, H.; Tsuburaya, N.; Naguro, I.; Matsuzawa, A.; Takeda, K.; Takahashi, Y.; Goto, J.; et al. A novel monoclonal antibody reveals a conformational alteration shared by amyotrophic lateral sclerosis-linked SOD1 mutants. *Ann. Neurol.* **2012**, *72*, 739−749.
(84) Kerman, A.; Liu, H.-N.; Croul, S.; Bilbao, J.; Rogaeva, E.; Zinman, L.; Robertson, J.; Chakrabartty, A. Amyotrophic lateral sclerosis is a non-amyloid disease in which extensive misfolding of SOD1 is unique to the familial form. *Acta Neuropathol.* **2010**, *119*, 335−344.
(85) Ayers, J. I.; Xu, G.; Pletnikova, O.; Troncoso, J. C.; Hart, P. J.; Borchelt, D. R. Conformational specificity of the C4F6 SOD1 antibody; low frequency of reactivity in sporadic ALS cases. *Acta neuropathologica communications* **2014**, *2*, 1.
(86) Vande Velde, C.; Miller, T. M.; Cashman, N. R.; Cleveland, D. W. Selective association of misfolded ALS-linked mutant SOD1 with the cytoplasmic face of mitochondria. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 4022−4027.

DOI: 10.1021/acs.jpcb.8b07680
J. Phys. Chem. B 2018, 122, 11662−11676

(87) Forsberg, K.; Jonsson, P. A.; Andersen, P. M.; Bergemalm, D.; Graffmo, K. S.; Hultdin, M.; Jacobsson, J.; Rosquist, R.; Marklund, S. L.; Brännström, T. Novel antibodies reveal inclusions containing non-native SOD1 in sporadic ALS patients. *PLoS One* **2010**, *5*, e11552.

DOI: 10.1021/acs.jpcb.8b07680
J. Phys. Chem. B 2018, 122, 11662−11676

Supplementary Material for: Prediction of Misfolding-Specific Epitopes in SOD1 Using Collective Coordinates

Xubiao Peng,[†,§] Neil R. Cashman,[‡] and Steven S. Plotkin[*,¶]

†*Department of Physics and Astronomy, University of British Columbia, Vancouver, British Columbia V6T1Z1, Canada*

‡*Brain Research Centre, University of British Columbia, Vancouver, Canada*

¶*Department of Physics and Astronomy; Genome Sciences and Technology, University of British Columbia, Vancouver, British Columbia V6T1Z1, Canada*

§*Center for Quantum Technology Research, School of Physics, Beijing Institute of Technology, Haidian, Beijing, 100081,China*

E-mail: steve@phas.ubc.ca

Phone: 604-822-8813

This supporting information has following contents,

1) The re-parameterized partial changes on HIS63 in CHARMM22* forcefield (Table S1).

2) The predicted MSEs using metric ΔQ (Table S2).

3) The predicted MSEs using metric ΔRMSF (Table S3).

4) The MSE prediction summary for the equilibrated control group (Fig S1)

5) The correlation coefficients for predictions between collective coordinates method and pure equilibration method for different metrics (Table S4).

Table S1: Partial charges for each atom on the doubly-deprotonated Histidine 63 in CuZn(SS) SOD1, using the CHARMM22* forcefield re-parameterized based on Gaussian quantum-classical potential energy matching. Atoms with modified partial charges are denoted by a double-lined box with bold text.

Atom Name	Atom Type	Partial Charge
N	NH1	-0.47
HN	H	0.31
CA	CT1	0.07
HA	HB	0.09
CB	CT2	-0.18
HB1	HA	0.09
HB2	HA	0.09
ND1	**NR1**	**-1.933**
CG	**CPH1**	**1.49**
CE1	**CPH2**	**0.24**
HE1	**HR1**	**-0.092**
NE2	**NR2**	**-1.523**
CD2	**CPH1**	**0.638**
HD2	**HR3**	**0.18**
C	C	0.51
O	O	-0.51

Table S2: Epitopes predicted from ΔQ. Stressed states—either EE(SS) or EE(SH) monomer—are listed left. Reference states are listed at the top.

	$CuZn(SS)_{dim}$		$CuZn(SS)_{mon}$		$EE(SS)_{mon}$	
	Residue	Sequence	Residue	Sequence	Residue	Sequence
EE(SH)*	-	-	3-5	KAV	-	-
	5-10	VCVLKG	8-10	LKG	8-10	LKG
	-	-	9-11	KGD	-	-
	22-24	QKE	-	-	-	-
	44-48	GFHVH	44-48	GFHVH	-	-
	51-53	GDN	-	-	-	-
	-	-	-	-	61-63	GPH
	71-73	HGG	71-73	HGG	-	-
	78-81	ERHV	-	-	-	-
	79-83	RHVGD	79-83	RHVGD	-	-
	93-95	GVA	-	-	-	-
	113-116	IGRT	112-114	IIG	-	-
	116-119	TLVV	116-119	TLVV	-	-
	-	-	138-140	GNA	-	-
	-	-	146-148	CGV	-	-
	148-152	VIGIA	147-150	GVIG	-	-
EE(SS)*	7-11	VLKGD	8-11	LKGD	-	-
	44-46	GFH	-	-	-	-
	-		46-48	HVH	-	-
	50-53	FGDN	-	-	-	-
	-	-	-	-	61-63	GPH
	-	-	-	-	-	-
	71-73	HGG	71-73	HGG	-	-
	78-85	ERHVGDLG	78-85	ERHVGDLG	-	-
	113-116	IGRT	-	-	-	-
	-	-	116-118	TLV	-	-
	124-127	DDLG	-	-	-	-
	-	-	146-148	CGV	-	-
	148-152	VIGIA	-	-	-	-

Table S3: Epitopes predicted from ΔRMSF. Stressed states—either EE(SS)* or EE(SH)* monomer—are listed left. Reference states are listed at the top.

	$CuZn(SS)_{dim}$		$CuZn(SS)_{mon}$		$EE(SS)_{mon}$	
	Residue	Sequence	Residue	Sequence	Residue	Sequence
EE(SH)*	51-53	GDN	-	-	52-55	DNTA
	61-64	GPHF	61-64	GPHF	57-62	CTSAGP
	-	-	62-67	PHFNPL	66-68	PLS
	70-74	KHGGP	70-76	KHGGPKD	-	-
	71-76	HGGPKD			-	-
	76-78	DEE	76-82	DEERHVG	-	-
	78-84	ERHVGDL	78-84	ERHVGDL	-	-
	123-131	ADDLGKGGN	123-131	ADDLGKGGN	129-139	GGNEESTKTGN
	130-140	GNEESTKTGNA	-	-	131-140	NEESTKTGNA
EE(SS)*	-	-	62-64	PHF	-	-
	71-76	HGGPKD	70-76	KHGGPKD	-	-
	78-84	ERHVGDL	78-84	ERHVGDL	-	-
	-	-	81-85	VGDLG	-	-
	123-128	ADDLGK	123-128	ADDLGK	-	-
	126-130	LGKGG	126-130	LGKGG	-	-
	132-136	EESTK	-	-	132-136	EESTK
	134-140	STKTGNA	-	-	134-139	STKTGN

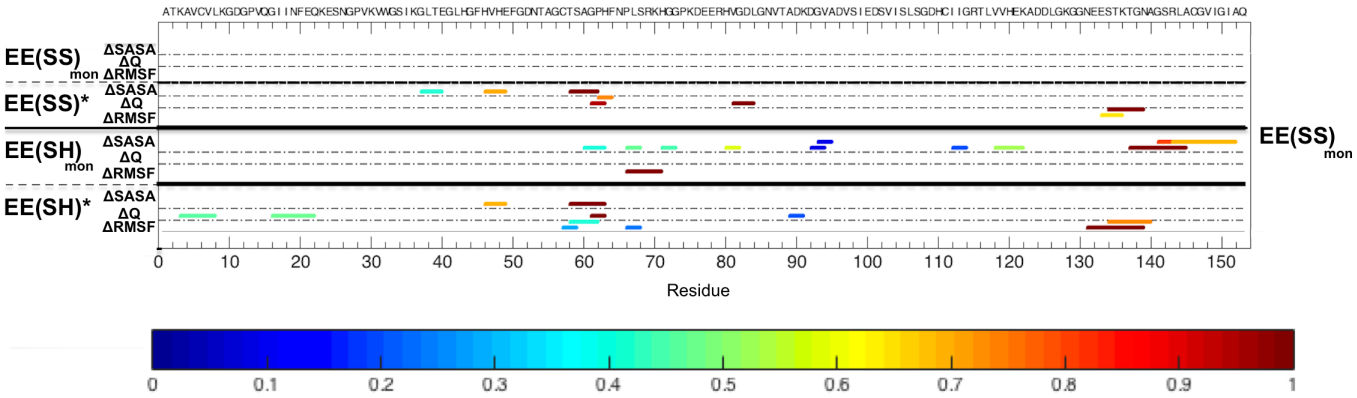


Figure S1: MSE predictions using native equilibrated ensembles rather than stressed, partially-unfolded ensembles (c.f. Fig 6 in the main text).

Table S4: The correlation (r,p) of the predicted epitopes between from collective coordinates stressing simulation and from the corresponding equilibration simulations for different metrics and reference states. All entries containing (-,-) have no predictions for the equilibrium ensembles, and so (r,p) are undefined.

	Δ SASA	Δ Q	Δ RMSF
$\mathrm{EE(SS)_{mon}} \rightarrow \mathrm{EE(SH)^{*}/EE(SH)_{mon}}$	(0.078, 0.33)	(-,-)	(0.23,4e-3)
$\mathrm{EE(SS)_{mon}} \rightarrow \mathrm{EE(SS)^{*}/EE(SS)_{mon}}$	(-, -)	(-,-)	(-,-)