Biophysical Journal
Article

Biophysical Society

Soft Vibrational Modes Predict Breaking Events during Force-Induced Protein Unfolding

Mona Habibi,[1] Steven S. Plotkin,[1,*] and Jörg Rottler[1,2]
[1]Department of Physics and Astronomy and [2]Quantum Matter Institute, University of British Columbia, Vancouver, Canada

ABSTRACT We investigate the correlation between soft vibrational modes and unfolding events in simulated force spectroscopy of proteins. Unfolding trajectories are obtained from molecular dynamics simulations of a Gō model of a monomer of a mutant of superoxide dismutase 1 protein containing all heavy atoms in the protein, and a normal mode analysis is performed based on the anisotropic network model. We show that a softness map constructed from the superposition of the amplitudes of localized soft modes correlates with unfolding events at different stages of the unfolding process. Soft residues are up to eight times more likely to undergo disruption of native structure than the average amino acid. The memory of the softness map is retained for extensions of up to several nanometers, but decorrelates more rapidly during force drops.

INTRODUCTION

Understanding the molecular mechanism behind deformation of protein structures can provide insight into the pathways of protein unfolding and the progression of diseases associated with protein misfolding (1). Identifying which parts of the protein unfold first along various unfolding pathways may lead to new therapeutic avenues, because easily disrupted regions of proteins may be prone to partake in the aberrant intermolecular interactions involved in the propagation of misfolding diseases. Single-molecule force spectroscopy has become a routine tool to reveal the mechanically induced unfolding pathways, including intermediate states (2,3). Such experiments are fruitfully complemented by molecular dynamics simulations, where one can observe directly all the conformational changes during unfolding in atomic detail (4–14). In this way, rupture events identified via force extension curves can be associated with specific regions for proteins that fold or unfold through intermediate states.

An immediate question that arises is whether it is possible to predict weak, unfolding prone regions in a protein exclusively from its structure and interaction energies. Ultimately, the unfolding dynamics is governed by a complex free energy landscape, where transitions with low barriers are more likely to be sampled. In thermal equilibrium, the protein mainly explores conformations near free energy minima. The protein motions occur preferentially along directions that require little energy to excite. These motions are dictated by low-frequency normal modes or "soft modes." Many studies have characterized soft modes and their relevance to the movement of proteins based on elastic network models (15). Normal mode analysis reveals the fluctuation spectrum, and has proved to be a useful tool for identifying structurally stabilizing elements (16), collective motion (17,18), conformational transitions (19,20), the mechanical properties of viral capsids (21), protein unfolding (22), and to compute the spatially dependent internal dielectric function (23). Anisotropic network models (ANMs) (24,25) have had significant success in elucidating the structure and function of complex bimolecular assemblies, for instance ligand binding (26), collective motion (27), the function of ion channels (28), and allostery (29). These models often (but not always) ignore the atomic details and specific sequence of amino acids, but they generally contain information on the overall tertiary structure and topology of contacts between residues.

In a broad class of amorphous solids whose energy landscapes share many characteristics with proteins, soft vibrational modes have recently been used to identify so-called soft spots, which are regions in the disordered packing of particles that are more likely to rearrange when the material undergoes a shear deformation (30–34). These soft spots can be found from a superposition of the amplitudes of the low energy, quasilocalized modes in the vibrational spectrum. Sound waves appear to scatter off flow defects more strongly than other regions. Such studies have provided

Submitted August 4, 2017, and accepted for publication November 27, 2017.
*Correspondence: steve@phas.ubc.ca
Editor: Nancy Forde.
https://doi.org/10.1016/j.bpj.2017.11.3781

strong clues for a connection between structure and dynamics in glassy systems. The existence of soft spots also has implications for the emergence of thermally activated hopping dynamics below the glass transition (35–38) and the localization of shear into bands, which can lead to global structural failure in the system (39–42).

In the context of the mechanical properties of proteins, normal mode analysis also provides a promising platform. Eyal and Bahar (43) elucidated the response of green fluorescent protein (GFP), ubiquitin, and E2lip3 with the ANM to various localized perturbations. Dietz and Rief (44) employed a closely related approach to explore the anisotropic deformation of GFP. Su et al. (45) simulated complete mechanical unfolding of CI2 and barnase based on a Gaussian network model. In this particular study (45), when starting from the native state, the contact of the residue pair that exhibits the largest fluctuation amplitude in the normal mode spectrum is broken, and the process is repeated until all contacts are broken. The unfolding pathway obtained by this procedure agreed well with the unfolding pathway obtained directly through molecular dynamics simulations. Sułkowska et al. (46) simulated mechanical unfolding of a domain of titin (Ig-I27) and GFP with a coarse-grained C_α Gō-model, and analyzed intermediate states of the unfolding trajectory also with a Gaussian network model. They report that those residues that exhibit a particularly large amplitude in the slowest modes tend to break in the next steps of the unfolding process.

The purpose of this article is to further this body of previous results by exploring quantitatively how the correlation between soft modes and rearrangements can be used to predict which parts of a protein are more susceptible to disruption by mechanical perturbation. To this end, we compare the force-induced unfolding events obtained directly from molecular dynamics simulations to the unfolding events predicted by structural analysis of the protein based on the ANM (25,47) at several different stages of unfolding. As an example, we choose a mutant of the protein superoxide dismutase 1 (SOD1), which has a primarily β-barrel native structure with two long, metal-binding loops between strands 4 and 5 and strands 7 and 8 (see Fig. 1 *A*). Mutations of SOD1 have been shown to be associated with familial amyotrophic lateral sclerosis (48). Many of these mutations have been observed to show weakened folding stability. Single-molecule force spectroscopy experiments have been used to provide insight on the unfolding mechanism of SOD1 and have revealed several intermediates in the unfolding process (6,7). Therefore, there is a strong need to gain further insight into the unfolding pathways of SOD1 with the help of atomistic and coarse-grained simulations, and to understand correlations between structure and stability of this protein whose misfolds are implicated in neurodegenerative disease. Here, we study the mechanical unfolding of

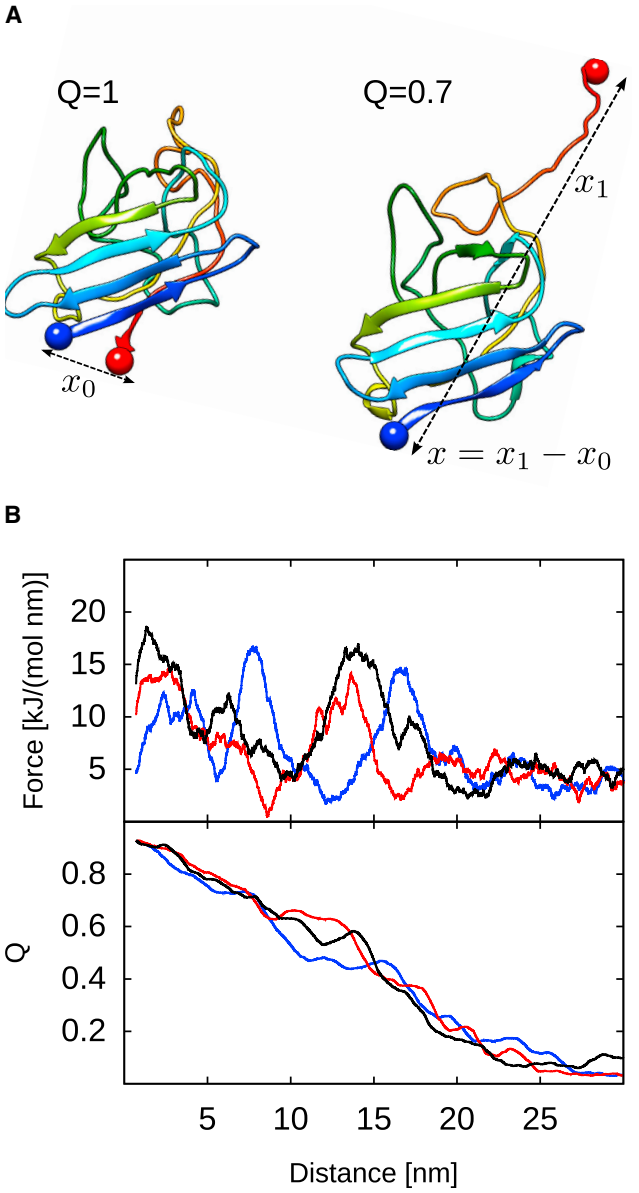

FIGURE 1 (*A*) Snapshots of the protein in the native state ($Q = 1$) and during pulling at $Q = 0.7$. (*B*) Force extension curves and Q (see Eq. 1) versus distance x (*dashed double-headed arrows* in *A*) for three different runs with fast pulling speed of 1 m/s. To see this figure in color, go online.

an apo, disulfide-reduced monomeric form of SOD1 with E133Q/C6A/C111S/F50E/G51E mutations (Protein Data Bank (PDB): 1RK7, model 1) (49). This laboratory engineered protein with partially restored enzymatic activity has been the focus of several previous experimental investigations (49–54).

In the next section, we describe the protein model, the force spectroscopy simulation method, details of measurements, and an overview on the soft mode analysis. Then, we present various metrics that quantify the correlation between residue breaking events and softness, as revealed through an ANM.

METHODS

Force spectroscopy simulations

For the molecular dynamics simulations, we employ a heavy atom-Gō (55) model that has been shown previously to reproduce unfolding pathways similar to full, all-atom, explicit solvent simulations (6,7). In the heavy atom-Gō model, all nonhydrogen atoms are present and the interaction between these atoms are defined by 6–12 Lennard-Jones interactions (55,56) based on the native state of the protein. Only atom pairs that are within a 6 Å cutoff distance in the native structure have attractive interactions. All other pairs of atoms are defined to be nonnative and are given short-range repulsive interactions (55). Harmonic bond, angle, and dihedral potentials are used to model covalently bonded atoms (55).

To simulate the force-induced unfolding of the protein, we tether both termini of the protein with harmonic potentials, and the C-terminus residue is then moved along the vector from C- to N-terminus with constant velocity. The stiffness of the tethers is set to 1000 $\mathrm{kJ/(mol \cdot nm^2)}$. We repeated the force spectroscopy simulations with two physical velocities of 1 m/s (fast pulling) and 0.1 m/s (slow pulling) 1000 and 40 times, respectively. The molecular dynamics simulations were carried out with Gromacs-4.5 (57), and the input files were generated from the Protein Data Bank structure using the SMOG web server (58). The time step was set at 2 fs and a Langevin thermostat with time constant of 1 ps was employed. The simulations were performed at ~90% of the folding temperature for the model, which was determined in a separate set of replica exchange molecular dynamics simulations (7). The initial configuration of the pulling simulations was obtained after 1 ns equilibration at the desired temperature.

Measurements

For all configurations obtained from the force spectroscopy simulations, we measured the end-to-end distance x between N- and C termini of the protein, the total fraction Q of native contacts, the fraction Q_i of native contacts for each residue, and the softness S_i of each residue in each configuration. The softness is calculated based on the 20 slowest normal modes within the remaining sequence-contiguous folded segment of the protein in each configuration (see below for details).

Native contacts

For configuration X, the fraction $Q(X)$ of native contacts is defined as

$$Q(X) = \frac{1}{|M|} \sum_{(i,j)\in M} \frac{1}{1+\exp\left[\beta^0\left(r_{ij}(X) - \lambda r_{ij}^0\right)\right]}, \qquad (1)$$

where $r_{ij}(X)$ is the distance between moieties i and j, r_{ij}^0 is the distance between heavy atoms i and j in the native state conformation and always less than a cutoff distance $r_{\mathrm{cut}} = 0.48$ nm, and M is the set of all pairs of native contacts (i,j) belonging to the native state. Amino acids that have a native contact must be separated by four or more residues in the primary sequence, $\beta^0 = 50\ \mathrm{nm}^{-1}$ is a smoothing parameter, and the factor $\lambda = 1.8$ takes into account fluctuations of the contacts (59).

Fig. 1 *A* shows two structures of the protein: one in the native state (when all native contacts are present and $Q = 1$) and one when the protein is partially unfolded ($Q = 0.7$). SOD1 has a β-barrel structure with two long loops dressing the core. Three force-extension curves and the fraction of native contacts versus distance for three trajectories are depicted in Fig. 1 *B*. As we pull the C-terminus, the protein unfolds while exhibiting several force peaks, and the number of native contacts decreases. Each unfolding event (Q-drop) corresponds to a force drop.

The value of Q_i is defined as the number of native contacts between residue i and all other residues, divided by the total number of native contacts that residue i has in the Protein Data Bank structure. During the pulling simulations, we monitor Q_i for all residues in all configurations. Residue i is considered unfolded, with native contacts broken, once it loses more than 40% of its native contacts ($Q_i < 0.6$). The high threshold of $Q_i < 0.6$ is chosen to make sure that the end residues are fairly folded and not loose, otherwise their softness (see below for definition) would be very large. Fig. 2 illustrates the order in which residues break during the course of unfolding by plotting the value of Q at breaking, Q_{break}, as a function of residue index. Since Q decreases monotonically with extension (see Fig. 1), residues with a high value of Q_{break} break early in the unfolding process, whereas those with a small value break last. The residues that break first in order of Q are those in strand $\beta 8$ and loops VII and IV, because they have the largest Q-values at breaking. The most stable core of the protein is made up of β-strands 2, 3, and 6. Results for slow pulling are little changed, which suggests a weak rate dependence.

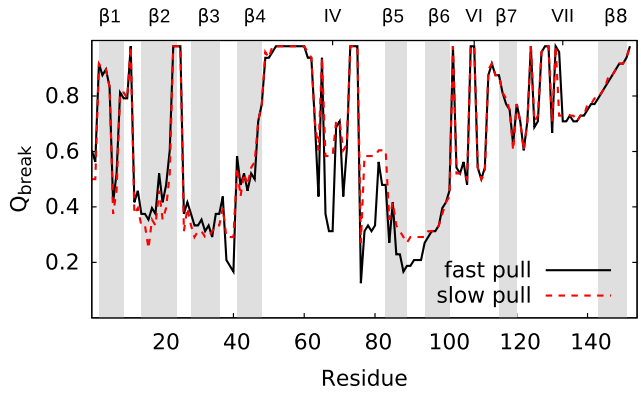

FIGURE 2 Loss of native structure during mechanical unfolding. The contacts involving residue i are defined as broken when Q_i drops below 0.6. Data is averaged over all 1000 (fast) or 40 (slow) pulling simulations. To see this figure in color, go online.

Soft modes

In this article, the Prody package (24,47) is used to calculate the softness S_i of residue i based on the ANM (17,25,47). In the ANM, each residue is reduced to a single node at the position of its C_α-atom that interacts with other nodes in the structure of the protein with a harmonic potential. The total potential energy of the system in any given configuration is the sum of these harmonic potentials,

$$V_{\mathrm{ANM}} = \frac{\gamma}{2} \sum_{i,j<i} \mathbf{\Gamma}_{ij} \left(R_{ij} - R_{ij}^0\right)^2, \qquad (2)$$

where γ is the spring constant, and R_{ij} and R_{ij}^0 are the respective instantaneous and equilibrium distances between nodes i and j. $\mathbf{\Gamma}_{ij}$ are the elements of the Kirchhoff matrix $\mathbf{\Gamma}$, in which $\mathbf{\Gamma}_{ij} = 1$ if nodes i and j are within 15 Å in the structure, otherwise $\mathbf{\Gamma}_{ij} = 0$. The Hessian matrix $\mathcal{H}$ for this network of N nodes is the second derivative of V_{ANM} with respect to the residue positions. It is formed by 3×3 block matrices:

$$H_{ij} = \frac{\gamma \mathbf{\Gamma}_{ij}}{\left(R_{ij}^0\right)^2} \begin{bmatrix} X_{ij}X_{ij} & X_{ij}Y_{ij} & X_{ij}Z_{ij} \\ Y_{ij}X_{ij} & Y_{ij}Y_{ij} & Y_{ij}Z_{ij} \\ Z_{ij}X_{ij} & Z_{ij}Y_{ij} & Z_{ij}Z_{ij} \end{bmatrix}, \qquad (3)$$

where X_{ij}, Y_{ij} and Z_{ij} are the components of the distance vector separation vector R_{ij}^0. $\mathcal{H}$ has $3N - 6$ nonzero eigenvalues λ_k and corresponding

eigenvectors $\mathbf{u_k}$. The eigenvectors describe the vibrational direction and the relative amplitude in the different modes,

$$\mathcal{H}^{-1} = \sum_{k=1}^{3N-6} \frac{1}{\lambda_k} \mathbf{u_k} \mathbf{u_k}^T. \quad (4)$$

For a single mode k, the square of the normal mode amplitudes multiplied by the inverse of the eigenvalue of the mode yields the squared fluctuation in that mode, $\langle (\Delta R_i)^2 \rangle_k \propto \lambda_k^{-1} [u_k]_i^2$. We define the softness S_i of residue i as the sum of the squared fluctuations for a given set of normal modes (47), which makes our softness proportional to a thermal Debye-Waller or crystallographic B-factor. In this study, the softness is calculated for the slowest (softest) 20 normal modes. We find that including more than 20 modes no longer changes the softness appreciably.

Note that S_i is only calculated for the C_α-atoms of residues in the folded segment of the protein in configuration X. The folded segment is defined as a contiguous sequence of n residues with residue index $i \le j \le i+n$, where $Q_i(X) > 0.6$ and $Q_{i+n}(X) > 0.6$. The loops IV and VII (residue indices 51–84 and 126–139) are unfolded in the early stage of the pulling. However, since these loops are in between folded segments of the protein, they are considered part of the folded segment, even though they have $Q_i < 0.6$.

RESULTS AND DISCUSSION

Softness of residues

In Fig. 3 *A*, we show the mean softness $\langle S_i(x) \rangle$ for the residues in the most frequently folded segment of the protein in the native state and at extensions $x = 4, 8, 12$ nm. For the largest extension, the protein is about half-unfolded as $Q = 0.5$ (see Fig. 1). We do not consider larger extensions $Q < 0.5$ because the remaining native protein fragments become too structurally diverse for the ANM analysis to be meaningfully applied. Our previous study showed that as Q drops from 1 to 0.5, SOD1 follows a single unfolding pathway (7). $\langle S_i \rangle$ is calculated by averaging the modes obtained from the still folded part of the protein over a subset of all available configurations that include the dominant pathway at distances between x and $x + \delta x$, where $\delta x = 0.02$ nm. As the protein extends (unfolds), the average softness of the protein structure increases. We name the residues with the higher values of softness (at the peak positions) as soft regions, whereas the rigid regions of the proteins are segments with lower values of S_i. According to Fig. 3 *A*, the peaks in softness that define the soft regions at each extension are approximately conserved as a function of extension.

In Fig. 3 *B*, the C_α-atoms of the native structure along with the eigenvectors of the first three slowest modes at this state are illustrated. In the soft regions (*red colored regions* of the protein backbone), localized modes with large amplitudes appear.

In Fig. 4, we show configurations of the folded segment of the protein in the native state and at extensions $x = 4, 8, 12$ nm. The structures are color-coded based on the mean softness. Rigid residues with $\langle S_i(x) \rangle$ smaller than the median softness of all residues in the folded segment are shown in blue, and residues with softness in the top 20% percentile are shown in red. We observe that residues with larger softness will break soon thereafter, e.g., between extensions of 8 and 12 nm, where strand $\beta 1$ loses all its secondary structure at $x = 12$ nm. The two loops are soft regions, and contact residues in these loops disorder early in the pulling simulations, as can be seen in Fig. 2.

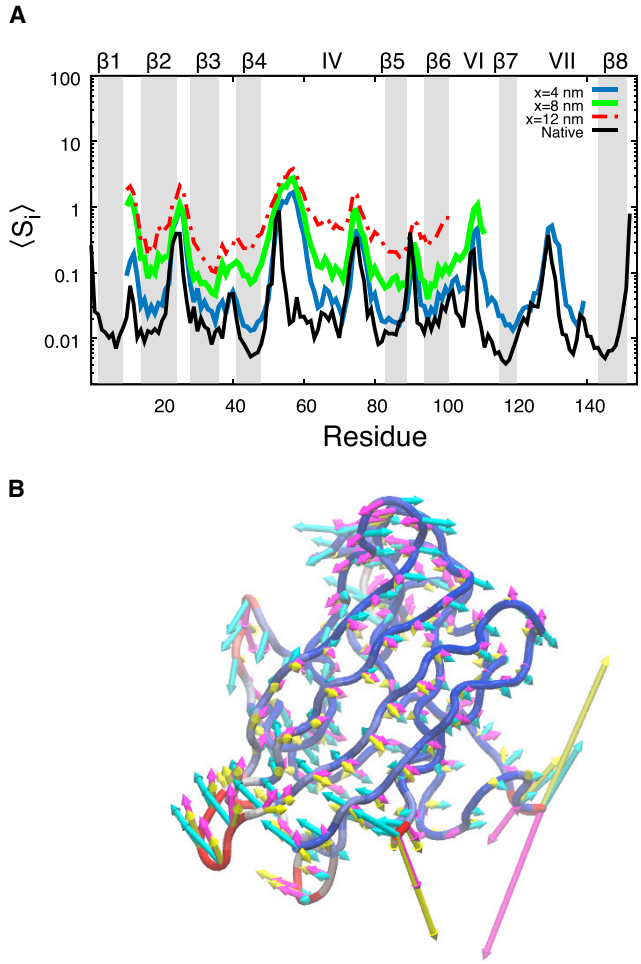

FIGURE 3 (*A*) Mean softness of residues in the folded segment of the protein versus residue index at different extensions for the fast pulling simulations. The contiguous region where S_i is calculated decreases as the extension increases. (*B*) Three slowest modes shown as arrows on the C_α-backbone of the native structure in cyan, magenta, and yellow, respectively. In the backbone structure, residues with softness more than median are shown in red. To see this figure in color, go online.

Next, we explore quantitatively the correlation between breaking events and softness through three different metrics that have proven useful in an earlier study on sheared glassy materials (33). In the first approach, we look at the probability of breaking events as a function of softness (Fig. 5). We also discuss the success rate of breaking event prediction (Fig. 6). Finally, we explore how far in the future the softness map can predict breaking events (Fig. 7).

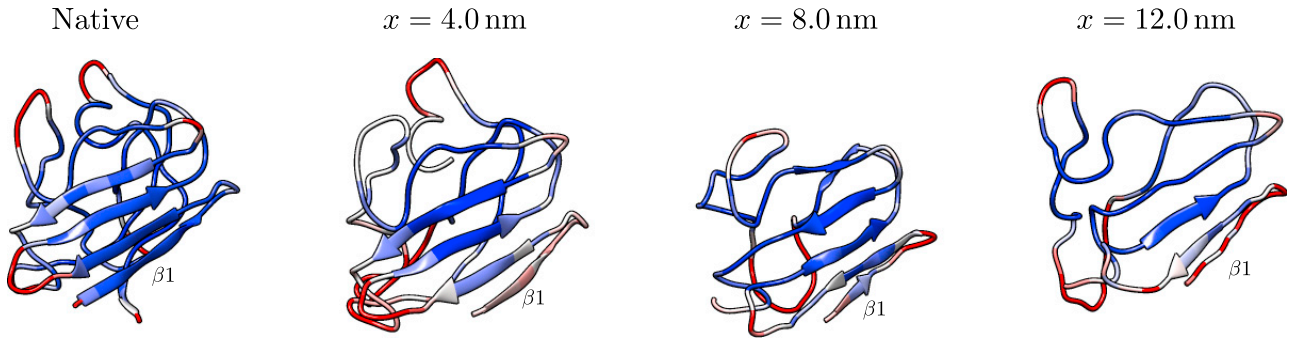

FIGURE 4 Structure of the folded segment of the protein of one run at different extensions for fast pulling. Residues with $\langle S_i \rangle$ in the top 20% percentile are shown in red, white represents the median softness, and blue shows residues with softness less than the median. To see this figure in color, go online.

Probability of breaking event versus softness

To quantify how the probability $P(S(x))$ of a residue breaking depends on residue softness, we first compute the softness $S_i(x)$ of a given residue i by averaging over several frames in a given run from extension x to $x + \delta x$, where $\delta x = 0.02$ nm. This averaging procedure is employed to reduce thermal noise. Then, we compute the value of $Q_i(x)$ again for several frames between extension $x + \delta x$ and $x + 2\delta x$. A residue is defined as broken if it has lost more than 40% of its contacts ($Q_i < 0.6$). The breaking probability of that residue is then simply the fraction of all frames in which that residue is broken, i.e., the fraction of all frames where $Q < 0.6$. Frames are grouped by extension x into bins of width 1 nm, and residues already broken at smaller extensions are not considered as breaking events at x. We then average the breaking probability over all available runs for four different extensions of unfolding.

As more residue contacts break, the average softness of the protein increases (see Fig. 3 *A*). To compare the softness at different extensions x, we normalize the softness by the average softness $\langle S(x) \rangle$ of all residues in the folded segment of the configurations corresponding to distance x. Only folded residues contribute to that average. Similarly, we normalize the breaking probability $P(S(x))$ by the average breaking probability at that extension.

As can be seen in Fig. 5, the breaking probability always strongly correlates with softness. For hard regions ($S/\langle S(x) \rangle < 1$), the average probability is very low. For residues with S more than the average softness of the protein, the probability always lies above the average breaking probability and has an overall increasing trend. The slope is largest in the native state, for which the breaking probability can be enhanced up to eightfold relative to the average value in fast pulling. For slow pulling, the trends are generally comparable, but we cannot explore larger values of softness due to limited statistics.

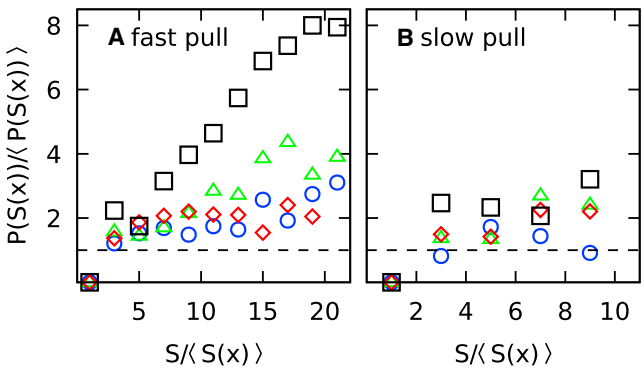

FIGURE 5 Breaking probability as a function of softness for extensions $x = 0, 4, 8, 12$ nm (*black* (□), *blue* (△), *green* (∘), and *red* (◇)). The color code is the same as in Fig. 3. The breaking probability is rescaled by the average breaking probability of the residues at that extension. Likewise, softness is rescaled by the average softness at that extension: (*A*) fast pulling (1.0 m/s) and (*B*) slow pulling (0.1 m/s). Statistical error bars are of order the symbol size. To see this figure in color, go online.

Success rate of breaking event prediction

We calculate the correlation between softness and residue breaking events in terms of a predictive success rate $\Theta(f)$,

$$\Theta(f) = \left\langle \left(\sum_{i=1}^{N} S_i^{(b)}(f) \times \Delta q_i \right) \Big/ \left(\sum_{i=1}^{N} \Delta q_i \right) \right\rangle, \quad (5)$$

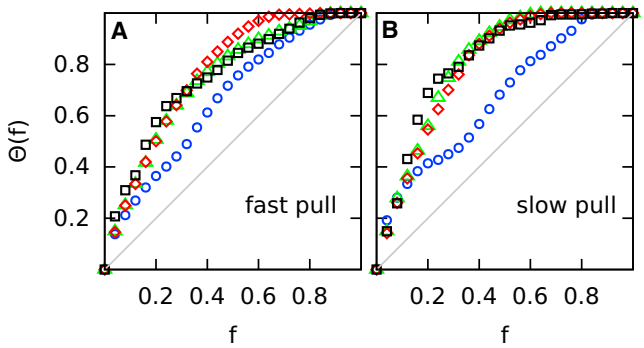

FIGURE 6 Predictive success rate of breaking events as a function of the fraction of the softest residues of the protein for fast (*A*) and slow (*B*) pulling at extensions $x = 0, 4, 8, 12$ nm (*black* (□), *blue* (△), *green* (∘), and *red* (◇)). The gray line indicates the success rate expected for randomly chosen residues. Error bars are of order the symbol size. To see this figure in color, go online.

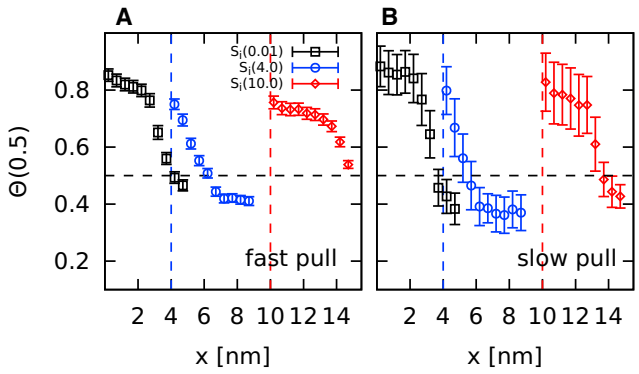

FIGURE 7 Evolution of predictive success rate $\Theta(0.5)$ at constant softness fraction of 50% for fast (*A*) and slow (*B*) pulling. $S_i(x)$ is calculated at x, and Q_i is monitored at subsequent extensions (see text). Vertical dashed lines (*blue* and *red*) show the points where the softness is calculated, and the horizontal dashed line indicates the decorrelated value. To see this figure in color, go online.

where N is the number of folded residues and the average is taken over all configurations at a particular extension. Here, we first rank-order the softness values and then binarize the softness such that a fraction f of the residues with largest softness are assigned a value of $S_i^{(b)} = 1$, and all other residues have a softness of 0. Similarly, $\Delta q_i = 1$ if residue has on average less than 60% native contacts and is considered broken, otherwise $\Delta q_i = 0$. $\Theta(f)$ equals one if all broken residues are also soft residues. Fig. 6 shows the average predictive success rate as as function of the fraction f of soft residues for different extensions. Also shown (*solid diagonal gray line*) is the function $\Theta(f)$ on which all data would fall if breaking occurred at random locations.

For all extensions, the soft residues are more likely to break than randomly chosen ones. We observe the maximum difference between $\Theta(f)$ and the random line at ~50% fractional coverage, where up to 90% of the breaking events are predicted at extensions $x = 0, 8$ and 12 nm. Again, there is very little difference between fast and slow pulling. A notably smaller predictive success rate is observed at $x = 4$ nm (*blue symbols*), which lie much closer to the random line.

Correlation with future breaking events

The analysis so far has established a robust correlation between softness and breaking probability in a short time interval immediately after the softness was calculated. How long does this correlation last? In Fig. 7, we show the evolution of the success rate of $\Theta(f = 0.5)$ for the still folded residues at three different extensions as a function of the distance between softness and breaking event measurements. The softness is measured at extension $x_0 = 0, 4,$ and 10 nm, and new breaking events are monitored during the subsequent unfolding trajectory. The horizontal black dashed line shows the success rate based on randomly chosen residues. The correlation between softness and the breaking event slowly decays, decorrelating to the value for randomly chosen residues after ~4 nm of further extension for the reference states of $x = 0$ and $x = 10$ nm. When starting from extension $x = 4$ nm by contrast, we observe a more rapid decorrelation over ~2 nm. This shorter decorrelation length is qualitatively consistent with the lower predictive success rate observed for the same extension in Fig. 5. Again these results are insensitive to the pulling rate.

To understand the rapid decorrelation at extension 4 nm, we look at the averaged force extension curves and $\langle Q \rangle$ curves versus distance for both fast (*dark blue*) and slow pulling (*pink*) rate in Fig. 8. In general, $\langle Q \rangle$ decreases very slowly or stays constant when the force ramps up. For these extensions $(x = 0 - 3, 10 - 13$ nm), the correlation between softness and breaking events decays slowly and indeed $\langle Q \rangle$ changes little. By contrast, for extensions $4 - 8$ nm, multiple force drops occur and $\langle Q \rangle$ decreases much more quickly for both fast and slow pulling rates. These results indicate that the memory of the softness map is erased faster by major unfolding events. From our data, we conclude that softness and residue breaking become uncorrelated after Q has changed by more that 0.1.

CONCLUSIONS

We quantified the correlation between residue contact breaking events and the softness (or B-values) calculated

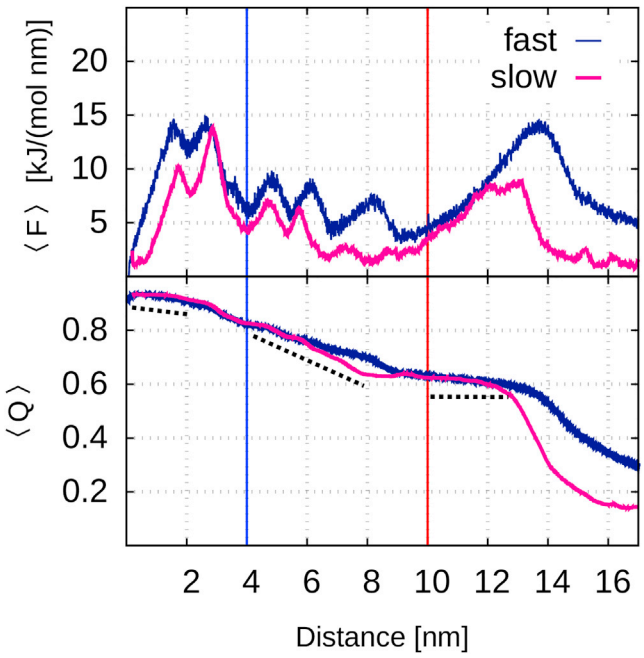

FIGURE 8 Force extension curves and $\langle Q \rangle$ versus distance for fast and slow pulling averaged over all runs are shown in navy and pink, respectively. Vertical lines show the points where softness is calculated. Dotted lines indicate the approximate slopes of the $\langle Q \rangle$ versus distance curve in the three regimes as a guide to the eye. To see this figure in color, go online.

using a normal mode analysis based on the ANM (24,25). Here we explored these correlations for the native state as well as several partially unfolded stages of a monomer of SOD1 with statistical metrics so far not applied to proteins. We monitored the residues that break in a short window right after the softness was calculated. We showed that the probability to observe a breaking event increases monotonically as a function of softness and eventually reaches a plateau. The trend between breaking probability and softness is largest for the native state. In the rigid regions (small softness), the average breaking probability is very low. We also showed that 50% of the soft residues at any extension are responsible for up to 90% of the breaking events in a short window after the softness measurement. We then explored how far into the future the softness can predict breaking events. We found that the strength of the prediction of breaking events slowly decreases as the distance between softness computations and breaking events increases. More rapid decorrelation is observed when the protein undergoes larger unfolding events signaled by a force drop. In general, the correlation vanishes after Q has decreased by $\sim$0.1 after the value of Q where the softness was calculated. Reducing the pulling rate by one order of magnitude does not change the results appreciably.

It is interesting to compare our results for soft mode rupture correlations to a previous analysis in plastically deformed glassy solids (33). In this system, the probability for particles to rearrange increased about sevenfold at the softest regions that were identified with the same procedure. Similarly, the top 30% of the softest particles predicted $\sim$70% of all rearrangements, and more than 50% of the material had to undergo rearrangements to erase the softness-plasticity correlation. This analysis shows that the protein ANM soft modes are performing similarly when viewed through these metrics. Even the half-unfolded protein still exhibits much stronger than random correlations. It is remarkable that a fairly simple network model is able to capture these trends.

An intriguing direction for future research appears to be to generate unfolding events by displacing residues along the directions given by the soft modes and couple with short re-equilibration intervals, similiar to the recently proposed hybrid co-molecular dynamics simulation algorithm (60). In this way, unfolding simulations at much slower pulling speeds could be realized. The timescale discrepancy between simulated force spectroscopy and experimental assays remains one of the biggest obstacles on the way to achieving quantitative agreement between in silico and in vitro experiments.

ACKNOWLEDGMENTS

S.S.P. and J.R. acknowledge WestGrid (https://www.westgrid.ca) and Compute Canada/Calcul Canada (https://www.computecanada.ca/home/) for providing computing resources.

J.R. thanks the Natural Sciences and Engineering Research Council of Canada for financial support. S.S.P. is supported by Alberta Prion Research Institute Research Team Program grant PTM13007, and the Canadian Institutes of Health Research Transitional Operating grant 2682.

REFERENCES

1. Zhuravlev, P. I., G. Reddy, …, D. Thirumalai. 2014. Propensity to form amyloid fibrils is encoded as excitations in the free energy landscape of monomeric proteins. *J. Mol. Biol.* 426:2653–2666.
2. Mojumdar, S. S., D. R. Dee, …, M. T. Woodside. 2016. Multiple intermediates in the folding of superoxide dismutase 1 revealed by single molecule force spectroscopy. *Biophys. J.* 110:497a.
3. Moffitt, J. R., Y. R. Chemla, …, C. Bustamante. 2008. Recent advances in optical tweezers. *Annu. Rev. Biochem.* 77:205–228.
4. Lu, H., B. Isralewitz, …, K. Schulten. 1998. Unfolding of titin immunoglobulin domains by steered molecular dynamics simulation. *Biophys. J.* 75:662–671.
5. West, D. K., D. J. Brockwell, …, E. Paci. 2006. Mechanical resistance of proteins explained using simple molecular models. *Biophys. J.* 90:287–297.
6. Habibi, M., J. Rottler, and S. S. Plotkin. 2016. As simple as possible, but not simpler: exploring the fidelity of coarse-grained protein models for simulated force spectroscopy. *PLoS Comput. Biol.* 12:e1005211.
7. Habibi, M., J. Rottler, and S. S. Plotkin. 2017. The unfolding mechanism of monomeric SOD1 by simulated force spectroscopy. *Biochim Biophys Acta.* 1865:1631–1642.
8. Kouza, M., C.-K. Hu, …, A. Kolinski. 2013. A structure-based model fails to probe the mechanical unfolding pathways of the titin I27 domain. *J. Chem. Phys.* 139:065103.
9. Kouza, M., C.-K. Hu, …, M. S. Li. 2009. Protein mechanical unfolding: importance of non-native interactions. *J. Chem. Phys.* 131:215103.
10. Sotomayor, M., and K. Schulten. 2007. Single-molecule experiments in vitro and in silico. *Science.* 316:1144–1148.
11. Brockwell, D. J., E. Paci, …, S. E. Radford. 2003. Pulling geometry defines the mechanical resistance of a β-sheet protein. *Nat. Struct. Biol.* 10:731–737.
12. Carrion-Vazquez, M., H. Li, …, J. M. Fernandez. 2003. The mechanical stability of ubiquitin is linkage dependent. *Nat. Struct. Biol.* 10:738–743.
13. Mücksch, C., and H. M. Urbassek. 2016. Accelerating steered molecular dynamics: toward smaller velocities in forced unfolding simulations. *J. Chem. Theory Comput.* 12:1380–1384.
14. Zheng, W. 2014. All-atom and coarse-grained simulations of the forced unfolding pathways of the SNARE complex. *Proteins.* 82:1376–1386.
15. Chennubhotla, C., A. J. Rader, …, I. Bahar. 2005. Elastic network models for understanding biomolecular machinery: from enzymes to supramolecular assemblies. *Phys. Biol.* 2:S173–S180.
16. Rader, A. J., G. Anderson, …, J. Klein-Seetharaman. 2004. Identification of core amino acids stabilizing rhodopsin. *Proc. Natl. Acad. Sci. USA.* 101:7246–7251.
17. Tama, F., and Y.-H. Sanejouand. 2001. Conformational change of proteins arising from normal mode calculations. *Protein Eng.* 14:1–6.
18. Yang, L., G. Song, and R. L. Jernigan. 2007. How well can we understand large-scale protein motions using normal modes of elastic network models? *Biophys. J.* 93:920–929.
19. Delarue, M., and Y.-H. Sanejouand. 2002. Simplified normal mode analysis of conformational transitions in DNA-dependent polymerases: the elastic network model. *J. Mol. Biol.* 320:1011–1024.
20. Zheng, W., B. R. Brooks, and G. Hummer. 2007. Protein conformational transitions explored by mixed elastic network models. *Proteins.* 69:43–57.
21. Tama, F., and C. L. Brooks, 3rd. 2005. Diversity and identity of mechanical properties of icosahedral viral capsids studied with elastic network normal mode analysis. *J. Mol. Biol.* 345:299–314.

22. Srivastava, A., and R. Granek. 2013. Cooperativity in thermal and force-induced protein unfolding: integration of crack propagation and network elasticity models. *Phys. Rev. Lett.* 110:138101.
23. Guest, W. C., N. R. Cashman, and S. S. Plotkin. 2011. A theory for the anisotropic and inhomogeneous dielectric properties of proteins. *Phys. Chem. Chem. Phys.* 13:6286–6295.
24. Eyal, E., L.-W. Yang, and I. Bahar. 2006. Anisotropic network model: systematic evaluation and a new web interface. *Bioinformatics.* 22:2619–2627.
25. Atilgan, A. R., S. R. Durell, …, I. Bahar. 2001. Anisotropy of fluctuation dynamics of proteins with an elastic network model. *Biophys. J.* 80:505–515.
26. Meireles, L., M. Gur, …, I. Bahar. 2011. Pre-existing soft modes of motion uniquely defined by native contact topology facilitate ligand binding to proteins. *Protein Sci.* 20:1645–1658.
27. Eyal, E., A. Dutta, and I. Bahar. 2011. Cooperative dynamics of proteins unraveled by network models. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* 1:426–439.
28. Shrivastava, I. H., and I. Bahar. 2006. Common mechanism of pore opening shared by five different potassium channels. *Biophys. J.* 90:3929–3940.
29. Xu, C., D. Tobi, and I. Bahar. 2003. Allosteric changes in protein structure computed by a simple mechanical model: hemoglobin T<–>R2 transition. *J. Mol. Biol.* 333:153–168.
30. Widmer-Cooper, A., H. Perry, …, D. R. Reichman. 2008. Irreversible reorganization in a supercooled liquid originates from localized soft modes. *Nat. Phys.* 4:711–715.
31. Tanguy, A., B. Mantisi, and M. Tsamados. 2010. Vibrational modes as a predictor for plasticity in a model glass. *EPL(Europhysics Letters).* 90:16004.
32. Manning, M. L., and A. J. Liu. 2011. Vibrational modes identify soft spots in a sheared disordered packing. *Phys. Rev. Lett.* 107:108302.
33. Smessaert, A., and J. Rottler. 2014. Structural relaxation in glassy polymers predicted by soft modes: a quantitative analysis. *Soft Matter.* 10:8533–8541.
34. Smessaert, A., and J. Rottler. 2015. Correlation between rearrangements and soft modes in polymer glasses during deformation and recovery. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 92:052308.
35. Sastry, S., P. G. Debenedetti, and F. H. Stillinger. 1998. Signatures of distinct dynamical regimes in the energy landscape of a glass-forming liquid. *Nature.* 393:554–557.
36. Wang, J., S. S. Plotkin, and P. G. Wolynes. 1997. Configurational diffusion on a locally connected correlated energy landscape; application to finite, random heteropolymers. *J. Phys. I.* 7:395–421.
37. Schweizer, K. S., and E. J. Saltzman. 2003. Entropic barriers, activated hopping, and the glass transition in colloidal suspensions. *J. Chem. Phys.* 119:1181–1196.
38. Bhattacharyya, S. M., B. Bagchi, and P. G. Wolynes. 2008. Facilitation, complexity growth, mode coupling, and activated dynamics in supercooled liquids. *Proc. Natl. Acad. Sci. USA.* 105:16077–16082.
39. Spaepen, F. 1977. A microscopic mechanism for steady state inhomogeneous flow in metallic glasses. *Acta Metall.* 25:407–415.
40. Langer, J. S. 2001. Microstructural shear localization in plastic deformation of amorphous solids. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 64:011504.
41. Dasgupta, R., H. G. E. Hentschel, and I. Procaccia. 2012. Microscopic mechanism of shear bands in amorphous solids. *Phys. Rev. Lett.* 109:255502.
42. Wisitsorasak, A., and P. G. Wolynes. 2017. Dynamical theory of shear bands in structural glasses. *Proc. Natl. Acad. Sci. USA.* 114:1287–1292.
43. Eyal, E., and I. Bahar. 2008. Toward a molecular understanding of the anisotropic response of proteins to external forces: insights from elastic network models. *Biophys. J.* 94:3424–3435.
44. Dietz, H., and M. Rief. 2008. Elastic bond network model for protein unfolding mechanics. *Phys. Rev. Lett.* 100:098101.
45. Su, J. G., C. H. Li, …, C. X. Wang. 2008. Protein unfolding behavior studied by elastic network model. *Biophys. J.* 94:4586–4596.
46. Sułkowska, J. I., A. Kloczkowski, …, R. L. Jernigan. 2008. Predicting the order in which contacts are broken during single molecule protein stretching experiments. *Proteins.* 71:45–60.
47. Bakan, A., L. M. Meireles, and I. Bahar. 2011. ProDy: protein dynamics inferred from theory and experiments. *Bioinformatics.* 27:1575–1577.
48. Rosen, D. R., T. Siddique, …, H. X. Deng. 1993. Mutations in Cu/Zn superoxide dismutase gene are associated with familial amyotrophic lateral sclerosis. *Nature.* 362:59–62.
49. Banci, L., I. Bertini, …, M. S. Viezzoli. 2003. Solution structure of Apo Cu,Zn superoxide dismutase: role of metal ions in protein folding. *Biochemistry.* 42:9543–9553.
50. Banci, L., I. Bertini, …, M. S. Viezzoli. 1995. Synthesis and characterization of a monomeric mutant Cu/Zn superoxide dismutase with partially reconstituted enzymic activity. *Eur. J. Biochem.* 234:855–860.
51. Getzoff, E. D., D. E. Cabelli, …, R. A. Hallewell. 1992. Faster superoxide dismutase mutants designed by enhancing electrostatic guidance. *Nature.* 358:347–351.
52. Ferraroni, M., W. Rypniewski, …, S. Mangani. 1999. The crystal structure of the monomeric human SOD mutant F50E/G51E/E133Q at atomic resolution. The enzyme mechanism revisited. *J. Mol. Biol.* 288:413–426.
53. Banci, L., I. Bertini, …, M. S. Viezzoli. 2000. Backbone dynamics of human Cu,Zn superoxide dismutase and of its monomeric F50E/G51E/E133Q mutant: the influence of dimerization on mobility and function. *Biochemistry.* 39:9108–9118.
54. Assfalg, M., L. Banci, …, P. R. Vasos. 2003. Superoxide dismutase folding/unfolding pathway: role of the metal ions in modulating structural and dynamical features. *J. Mol. Biol.* 330:145–158.
55. Whitford, P. C., J. K. Noel, …, J. N. Onuchic. 2009. An all-atom structure-based potential for proteins: bridging minimal models with all-atom empirical forcefields. *Proteins.* 75:430–441.
56. Noel, J. K., and J. N. Onuchic. 2012. The many faces of structure-based potentials: from protein folding landscapes to structural characterization of complex biomolecules. *In* Computational Modeling of Biological Systems. Springer, pp. 31–54.
57. Pronk, S., S. Páll, …, E. Lindahl. 2013. GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics.* 29:845–854.
58. Noel, J. K., P. C. Whitford, …, J. N. Onuchic. 2010. SMOG@ctbp: simplified deployment of structure-based models in GROMACS. *Nucleic Acids Res.* 38:W657–W661.
59. Best, R. B., G. Hummer, and W. A. Eaton. 2013. Native contacts determine protein folding mechanisms in atomistic simulations. *Proc. Natl. Acad. Sci. USA.* 110:17874–17879.
60. Gur, M., J. D. Madura, and I. Bahar. 2013. Global transitions of proteins explored by a multiscale hybrid methodology: application to adenylate kinase. *Biophys. J.* 105:1643–1652.