Propensity to Form Amyloid Fibrils Is Encoded as Excitations in the Free Energy Landscape of Monomeric Proteins

Pavel I. Zhuravlev[1], **Govardhan Reddy**[2], **John E. Straub**[3] and **D. Thirumalai**[1]

1 - *Biophysics Program, Institute for Physical Science and Technology, Department of Chemistry and Biochemistry, University of Maryland, College Park, MD 20742, USA*
2 - *Solid State and Structural Chemistry Unit, Indian Institute of Science, Bangalore 560 012, India*
3 - *Department of Chemistry, Boston University, 590 Commonwealth Avenue, Boston, MA 02215–2521, USA*

Correspondence to D. Thirumalai: *thirum@umd.edu*
http://dx.doi.org/10.1016/j.jmb.2014.05.007
Edited by G. Hummer

Abstract

Protein aggregation, linked to many of diseases, is initiated when monomers access rogue conformations that are poised to form amyloid fibrils. We show, using simulations of src SH3 domain, that mechanical force enhances the population of the aggregation-prone (N^*) states, which are rarely populated under force free native conditions but are encoded in the spectrum of native fluctuations. The folding phase diagrams of SH3 as a function of denaturant concentration ($[C]$), mechanical force (f), and temperature exhibit an apparent two-state behavior, without revealing the presence of the elusive N^* states. Interestingly, the phase boundaries separating the folded and unfolded states at all $[C]$ and f fall on a master curve, which can be quantitatively described using an analogy to superconductors in a magnetic field. The free energy profiles as a function of the molecular extension (R), which are accessible in pulling experiments, (R), reveal the presence of a native-like N^* with a disordered solvent-exposed amino-terminal β-strand. The structure of the N^* state is identical with that found in Fyn SH3 by NMR dispersion experiments. We show that the timescale for fibril formation can be estimated from the population of the N^* state, determined by the free energy gap separating the native structure and the N^* state, a finding that can be used to assess fibril forming tendencies of proteins. The structures of the N^* state are used to show that oligomer formation and likely route to fibrils occur by a domain-swap mechanism in SH3 domain.

Introduction

The continuing effort to understand how proteins fold [1–5] is amply justified by the irrefutable link among misfolding [6,7], aggregation, and a growing list of diseases [8,9]. It is now firmly established that all proteins, regardless of their role in causing diseases, form amyloid fibrils under suitable conditions [9,10]. In most instances, newly synthesized proteins do not aggregate but rather fold, carry out the intended functions, and are subsequently degraded. The potential deleterious effects of interactions between misfolded structures leading to fibrils have made it urgent to understand the characteristics of proteins that harbor propensities to aggregate. Based on high-resolution crystal structures of fibrils of small peptides, sequence-based methods have been introduced to identify motifs that harbor amyloidogenic tendencies [11–13]. In addition to sequence, it is likely that the structures sampled by the monomer under native conditions encode not only the structures in the fibril state but also the rate of fibril formation. Thus, in order to decipher the aggregation-prone states, we need a complete structural characterization not of only the native state but also of higher free energy excitations that are amyloidogenic.

Since the cascade of events driving a monomer to a fibril depends both on the protein sequence and external conditions, there are multiple scenarios within the standard nucleation growth theory of protein aggregation. Regardless of the scenario, we theorize that protein aggregation is initiated if the aggregation-prone ensemble of structures (denoted

as the N^* state) is transiently populated due to thermal fluctuations or denaturation stress [14–16]. Thus, it is likely that both the tendency of a protein to form aggregates and, more interestingly, the rate of fibril formation are determined by the population of the N^* state. It also implies that the excitations in the spectrum of monomer conformations themselves are harbingers of protein aggregation. Consequently, identifying and revealing the stability and structure of the N^* state should yield quantitative insights into protein aggregation.

Because N^* is typically an excitation around the lowest free energy state (for a vast majority of proteins with a possible exception of mammalian prions [14]), it is only sparsely populated and thus is hard to detect experimentally. High structural and temporal resolutions are required to characterize structure and the extent of population of the N^* state. Techniques such as hydrogen exchange [17] and NMR spectroscopy [18] are successful in these endeavors. Here, we show that the elusive N^* state can be more readily identified using single-molecule force spectroscopy (SMFS) [19,20] because application of mechanical force (f) can enhance the population of high free energy states and slow down the dynamics. The only accessible experimental observable in the SMFS is the time-dependent change in the end-to-end distance (R), from which the f-dependent free energy profile can be determined. We show that $F(R)$ can reveal the presence of the elusive N^* state, the population of which is enhanced by the application of force (R being the variable conjugate to f), thus establishing the potential applicability of the SMFS to characterize excited states in the free energy spectra of monomers.

To reveal the nature of the N^* ensemble, we performed simulations of a coarse-grained model of the SH3 domain from the *Gallus gallus* src tyrosine kinase (PDB code 1SRL) at various values of constant force applied to the ends of the protein. We determined the folding phase diagrams in the (f,T) and ($f,[C]$) planes {[C] is the concentration of the denaturant guanidinium chloride (GdmCl)}. The phase diagrams show that SH3 folding (triggered by changing f, [C], or T) can be approximately characterized by a two-state model. The phase boundaries in the (f,T) and ($f,[C]$) planes collapse onto a universal curve, which we quantify using an analogy to superconductors in a magnetic field. The f-dependent $F(R)$ values reveal the presence of a state N^*, which becomes prominent as f increases. Structurally, the N^* state corresponds to melting of the β-sheet formed by the N- and C-terminal strands. Surprisingly, an identical structure, with very low population, which has the same native fold as src SH3, has been recently identified in Fyn SH3 [18]. Force-unfolding kinetics shows that N^* state becomes populated ahead of global unfolding. Most importantly, we establish that the structure of N^* allows us to determine the

mechanism of oligomer formation, which in turn naturally suggests a route to fibril formation. For the src SH3 protein, this occurs by a domain-swap mechanism. Our work also shows that estimates of fibril formation times can be made using the population of N^*, thus establishing a direct link between the entire folding landscape and the propensity to aggregate.

Results

SH3 stability as a function of perturbations

The native structure of SH3 domains consists of five anti-parallel β-strands packed to form two perpendicular β-sheets. The strands are connected by the RT-loop, n-src loop, and a short distal loop (Fig. 1a). The N-terminal β-strand (β4 in Fig. 1a) participates in the hydrophobic core along with the strands β1, β2, and β3. In order to quantify the equilibrium response of SH3 to mechanical force (f), we performed multiple replica exchange low-friction Langevin dynamics simulations (see Methods) by applying a fixed constant f to the ends of the protein with and without denaturant.

To discriminate between the folded and unfolded states that are populated in the equilibrium simulation trajectories, we use the order parameter χ (structural overlap function)

$$\chi(\{r\}) = \frac{1}{M}\sum\nolimits_{(i,j)} \Theta\left(\left|\left|r_i - r_j\right| - \left|r_i^0 - r_j^0\right|\right| - \Delta\right), \qquad (1)$$

where the sum is over the native contacts [as pairs of beads (i,j)], M is the number of native contacts, $\Theta(x)$ is the Heaviside function, $\Delta = 2$ Å is the tolerance in the definition of a contact, and $r_{i,j}$ and $r_{i,j}^0$, respectively, are the coordinates of the beads in a given conformation $\{r\}$ and the native state. The histogram of χ exhibits a bimodal behavior, implying that SH3 folds in an apparent two-state manner (see Fig. S1). The value of $\chi = \chi_c$ separating the native basin of attraction (NBA) from the unfolded ensemble is obtained from analyzing the thermodynamics near the transition point, giving $\chi_c = 0.65$. The fraction of proteins in the NBA $f_{\text{NBA}} = \langle\Theta(\chi_c - \chi)\rangle$, with the averaging being over the Hamiltonian, which includes the transfer energy contribution due to the presence of denaturant (see Methods).

The titration curves, plotting f_{NBA} as a function of [C] (Fig. 1b) at different temperatures with $f = 7$ pN, show that the midpoint of the transition, C_{m}, calculated using $f_{\text{NBA}}([C_{\text{m}}]; f) = 0.5$, decreases sharply as T increases. Interestingly, the results in Fig. 1b show that, at $f \neq 0$, SH3 globally folds and unfolds reversibly in an apparent two-state manner, just as in ensemble experiments at a fixed T [21–25].
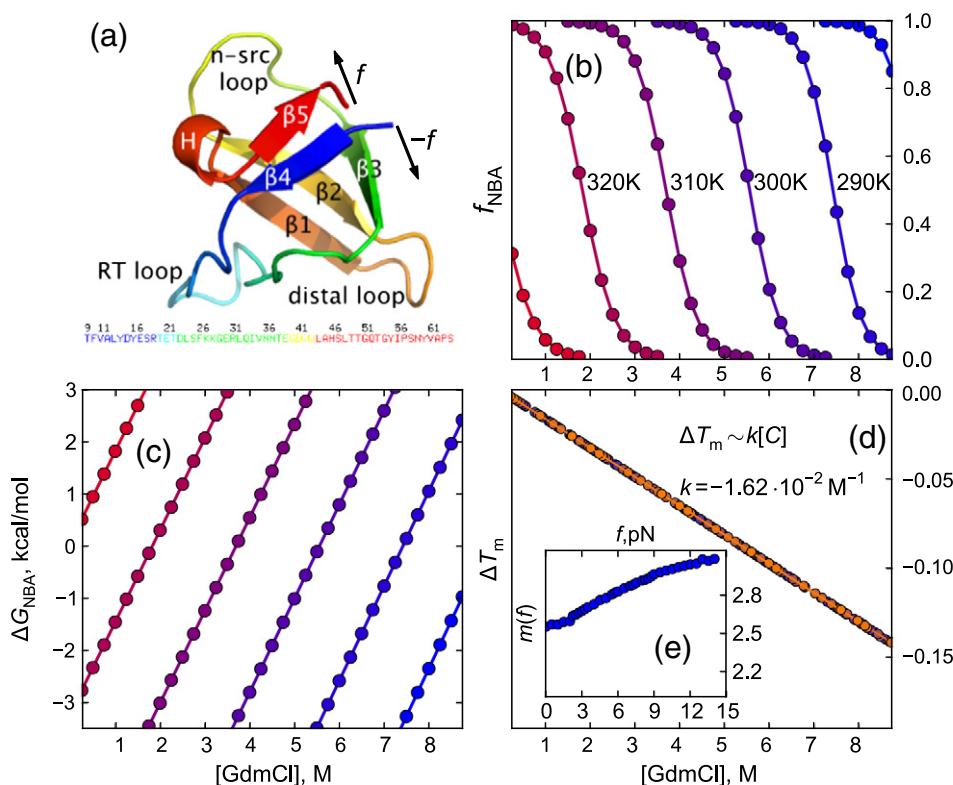
Fig. 1. Force-dependent thermodynamics as a function of temperature and denaturant concentration: (a) Crystal structure and sequence of the src SH3 domain (1SRL). (b) Guanidinium chloride titration curves at $f = 7$ pN. (c) Protein stabilities as functions of [GdmCl] in the two-state picture, $\Delta G_{NBA} = -k_B T \ln \frac{f_{NBA}}{(1-f_{NBA})}$. (d) Reduced melting temperatures $\Delta T_m = (T_m([C]) - T_m([0]))/T_m([0])$ at different forces from 0 pN (blue) to 14 pN (orange). The data for all forces fall on a single line as explained in the main text; [C] denotes the concentration of GdmCl and ΔT_m does not depend on f. (e, inset) $m(f)$ (plotted in kcal/mol/M at $T = 310$ K) depends weakly on f, with all of the dependence coming from the unfolded ensemble.

With the use of the results in Fig. 1b, the computed $\Delta G_{NBA}([C], T, f) = -k_B T \ln[f_{NBA}/(1 - f_{NBA})]$ at a fixed T with $f = 7$ pN is shown in Fig. 1c. At a fixed T, the dependence on [C] is given by $\Delta G_{NBA}([C], T, f) = \Delta G_{NBA}([0], T, f) + m(f)[C]$. Surprisingly, $m(f)$ is only weakly dependent on force because mechanical force does not perturb the intrinsic forces that determine the stability of proteins. Indeed, to the first approximation, m should be proportional to the solvent-accessible surface area (SASA), or $m \sim R_g^{2/3}$, and R_g does not depend on f while the protein is folded. From this perspective, use of f is a natural way to perturb the protein as opposed to T and [C], which invariably alter the interactions involving proteins, the solvent, and denaturants. All of the $m(f)$ arises from the increase in the SASAs of the unfolded ensemble (see Fig. S2), which grows as [C] approaches C_m from below. Linear response to [C] is also indicated in the dependence of $T_m([C]) = T_m([0], f) - \gamma(f)[C]$ at various f values (Fig. 1d). By the reasoning given above, we expect the reduced temperature $\Delta T_m = (T_m([C]) - T_m([0]))/T_m([0]) = k[C]$ to be proportional to [C] at all the forces, with k

being independent of f. The melting temperature without the denaturant $T_m([0], f)$ depends on f because the force changes the protein stability, but since it does not alter the interactions between the residues themselves, k should not depend on f. The thick continuous line in Fig. 1d, showing superposition of results at different f values, indeed satisfies the expected "scaling" behavior with $k = -0.016$ M^{-1} for all the f values.

Phase diagrams

From the titration curves at multiple forces and a fixed temperature ($T = 310$ K), we constructed the force-denaturant (f,[C]) phase diagram. Figure 2a shows that the critical (melting) force $f_m([C])$ needed to unfold SH3 increases as [C] decreases, reflecting the [C]-dependent stability of the native state. The boundary separating the NBA and unfolded basin of attraction (UBA) is relatively sharp, implying that SH3 behaves as a two-state folder in the (f,[C]) plane. The phase boundaries at other temperatures are shown in Fig. 2b. Similarly, the phase diagram as
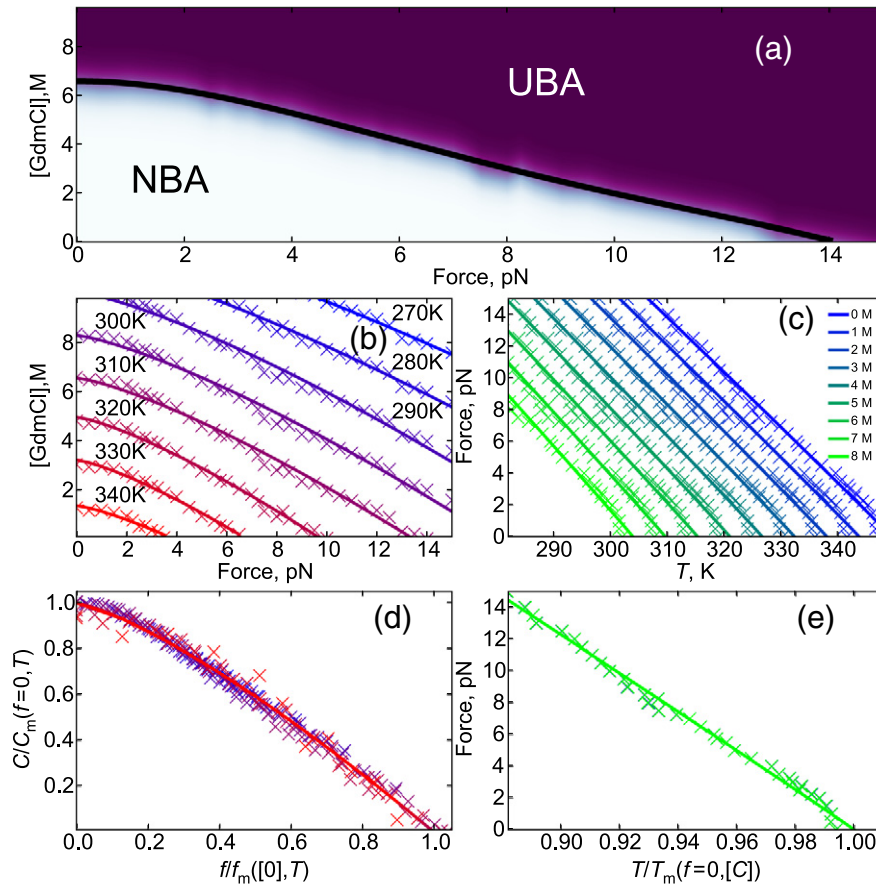
Fig. 2. Phase diagram of the src SH3 domain: (a) Force-denaturant phase diagram of the src SH3 domain at T = 310 K. A fraction of conformations inside the NBA is color-coded from white (f_{NBA} = 1) to purple (f_{NBA} = 0). (b) Boundaries between the folded and unfolded regions at different temperatures. (c) Boundaries between the folded and unfolded regions at different GdmCl concentrations. (d) (f,C) diagrams collapsed onto a master curve $C/C_0(T) = 1 − (f/f_0(T))^\beta$, β = 1.25 ± 0.1. (e) (f,T) diagrams collapsed onto a master curve $f/f_0([C]) = 1 − (T/T_0([C]))^\alpha$.

a function of (f,T) calculated at different [C] (see Fig. 2c) also shows a two-state behavior. Not surprisingly, the melting temperature and the critical (melting) force decrease as [C] increases (Fig. 2b). For a fixed T, the phase boundary can be quantified using

$$C_m(f) \sim C_0 \left(1 - \left(\frac{f}{f_0}\right)^\beta\right) \qquad (2)$$

where $C_m(f,T)$ is melting concentration, and $f_0(T)$ is the melting force at [C] = 0. The parameters $f_0(T)$, $C_0(T)$, and β should depend on the protein. From fitting, we find β = 1.25 ± 0.1, estimating error using jackknife [26]. A similar fit [$f_m/f_0([C]) = 1 − (T/T_m)^\alpha$] can be used to determine the family of curves at various values of [C] [27] (Fig. 2c and e). The continuous lines in Fig. 2d and e show that all curves

(given in Fig. 2b and c) collapse onto a master plot using the scaling functions mentioned above. The collapsed phase boundary in the restricted range of (f,T) values in Fig. 2e appears linear, but we expect the master curve to be nonlinear (α different from unity), which will be observable if the range of temperature is increased. However, such low T values may not be physically relevant. Nevertheless, the finding that the phase boundary for various values of [C] and f collapse onto a single curve is surprising and is amenable to experimental scrutiny.

The behavior of phase boundaries here is analogous to that found in superconductors, which exhibits a sharp phase boundary in the (H,T) plane described by the identical power law [28]. The analogy to superconductor in a magnetic field is appropriate because the transition is likely to be first order in both cases. As the magnetic field (H) in the superconductor, in our case, force enters linearly

into the Hamiltonian but is coupled to microscopic coordinates in a complicated fashion, leading to non-trivial response to *f* here or *H* in superconductor. Similarly, our Hamiltonian is a linear function of [*C*], which is coupled to a very complicated function of coordinates.

N^* state is visible in the presence of force

The global and experimentally accessible quantities in Fig. 1 show that, upon various perturbations (*T*, [*C*], and *f*), SH3 folds in an apparent two-state manner. A more nuanced picture emerges when the folding landscape is examined using a free energy profile *F*(*R*), where *R* is the extension of SH3 conjugate to *f*. Although not straightforward, *F*(*R*) can be precisely calculated [29] using folding trajectories generated in laser optical trap experiments [30,31]. The calculated free energy profile $F(R) = -k_B T \ln P(R)$ where k_B is the Boltzmann constant and *P*(*R*) is the distribution of *R* exhibits a minimum at *R* = 1.15 nm, which is distinct from the one at *R* = 0.68 nm corresponding to the NBA (Figs. 3 and 4). The fine structure in *F*(*R*) is subtle and is not noticeable in other quantities such as the distribution of the radius of gyration (see Figs. S3–S6). The *R* = 1.15 nm peak in *P*(*R*) does not correspond to the globally unfolded state at $R \approx 10$ nm. Barely noticeable at *f* = 0, the second minimum in *F*(*R*) becomes increasingly pronounced as *f* grows (Fig. 4). In other words, mechanical force reveals an elusive state in the *F*(*R*) profiles.

Structural characteristics of N^*

In order to characterize the structure of the N^* state corresponding to the peak at *R* = 1.15 nm (Fig. 4), we calculated the individual overlap parameters using
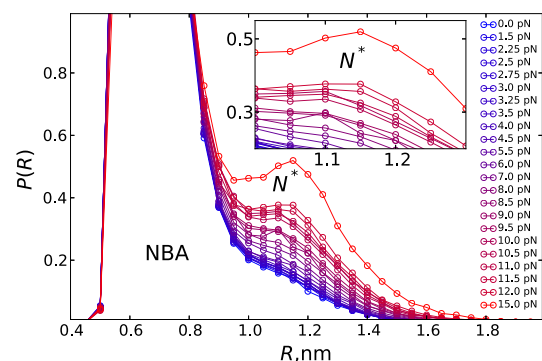

Fig. 4. Distributions of the end-to-end distance, *R*, at different values of the stretching force at a fixed *T* = 310 K. The smaller peak corresponds to the aggregation-prone state N^*, which becomes more prominent as the magnitude of the force grows.

$Q_{\beta1,\beta2}$, $Q_{\beta2,\beta3}$, and $Q_{\beta4,\beta5}$ [32] for each pair of β-strands that are part of a β-sheet using

$$Q(\{\mathbf{r}\}) = 1/M \sum \exp\left[-4\left(|\mathbf{r}_i - \mathbf{r}_j| - |\mathbf{r}_i^0 - \mathbf{r}_j^0|\right)^2\right], \quad (3)$$

where the sum is over the contacts defining the $Q(\{\mathbf{r}\})$ (e.g., for $Q_{\beta1,\beta2}$, these would be pairs of beads, one from β1 and the other one from β2 over all the beads in β1 and β2); *M* is the number of such contacts, $\mathbf{r}_{i,j}$ is the coordinates of the beads in the conformation for which the *Q* is calculated ($\{\mathbf{r}\}$), and $\mathbf{r}_{i,j}^0$ is the corresponding coordinates in the native state.

Two-dimensional histogram, in terms of the order parameters $(R, Q_{\beta4,\beta5})$ in Fig. 5a, shows that the second peak in the *P*(*R*) profiles (Fig. 4) corresponds to a smaller value of $Q_{\beta4,\beta5}$ compared to that found in the NBA. However, $Q_{\beta1,\beta2}$ and $Q_{\beta2,\beta3}$ retain their values in the native state (the higher peak) (Fig. 5b).
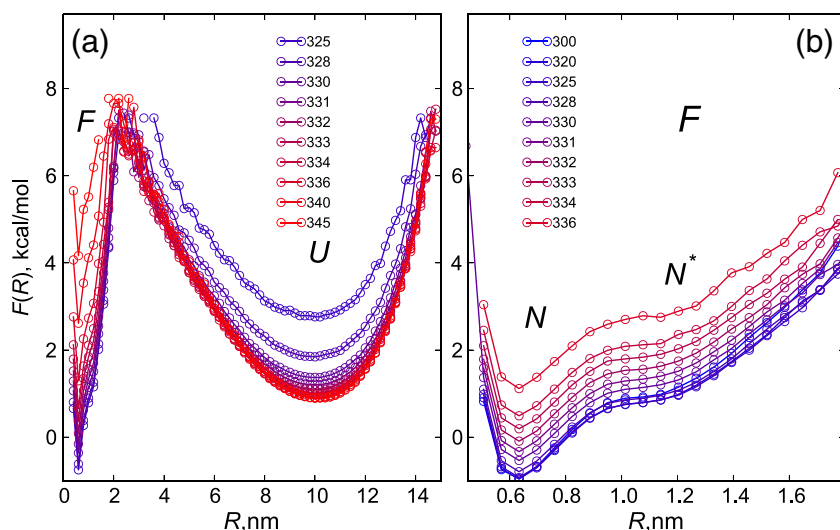

Fig. 3. Free energy profiles at *f* = 7.5 pN. (a) *F*(*R*) at various temperatures, showing that globally, SH3 even under force unfolds in an apparent two-state manner. The transition state location $\Delta x_{N \to U}^\ddagger \approx 2$ nm is insensitive to temperature. (b) Fine structure in *F*(*R*) shows the emergence of the N^* state, which is prominent at high forces. Population of the N^* state is between 3% (at 300 K) and 6% (at 340 K).
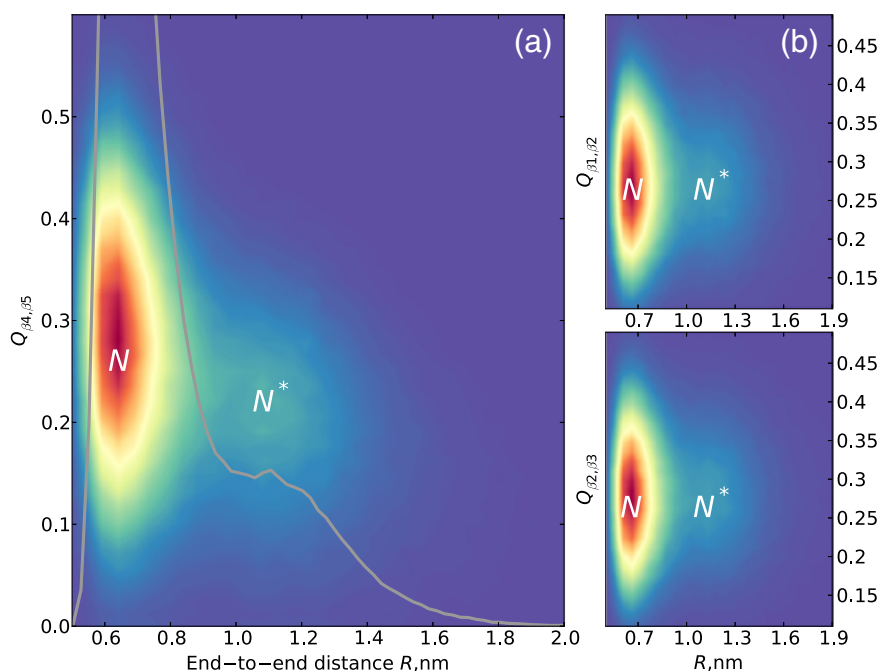
Fig. 5. Structural characteristics of the N^* state. Histograms of the conformations in the folded state at $f = 10$ pN and $T = 350$ K. The second, lower peak in $P(R)$ correlates with the melting of the β4-β5 sheet as shown by the fraction of contacts between the residues of those strands $Q_{\beta4,\beta5}$ (a). The y-axis on the one-dimensional plot shows the probability density, while for the two-dimensional plot, it is $Q_{\beta4,\beta5}$, and the probability density is shown by color from blue (low) to red (high). $Q_{\beta1,\beta2}$ and $Q_{\beta2,\beta3}$ (b) remain the same in the N^* state as in N.

Consequently, the second peak in $P(R)$ must reflect a state in which the β4-β5 sheet is melted (or disordered) with the rest of the native structure remaining intact. Because the N^* structure is native-like in SH3, we surmise that it is a native substate, which is hidden at $f = 0$. This assumption is corroborated by kinetic simulations showing that N^* is frequently visited during the native dynamics. We identify the ensemble of conformations belonging to the second peak as the N^* state that is aggregation prone (see below for comparison to experiments).

Force-induced unfolding kinetics

In order to determine if remnants of the N^* state arise in the force-unfolding kinetics of SH3 domain, we performed Brownian dynamics simulations at high friction as described in Methods at $T = 340$, 350, and 360 K and at $f = 10$ pN. The observed unfolding times spanned a broad range of time-scales (from hundreds of microseconds to over 10 ms) (see Table S1 showing the unfolding times for individual trajectories). The sequence of events during unfolding was always the following. First, the β4-β5 ruptures, populating the N^* state. Following this event, in about 2–3 μs, the β1-β2-β3 sheet rips, taking a few more microseconds (Fig. 6b and c). In this process, contacts between either β2-β3 or β1-β2

can break first. Both β2-β3 and β1-β2 sheets continue to transiently and partially reform in the unfolded state. Thus, breaking of the β4-β5 sheet is not by itself rate limiting in the global unfolding of SH3 under these conditions. The sheet ruptures and reforms (Fig. 6) on the order of hundreds of microseconds (in agreement with the reported accessibility of the N^* on the scale of milliseconds in the NMR experiment for Fyn SH3 [18]), but only when it stays melted long enough for the other β-sheet to follow (about 2 μs) does the protein globally unfold. The kinetic simulations show that N^* is the same intermediate identified in equilibrium free energy profiles. Thus, in the folding landscape of SH3, the N^* state is indeed a folding intermediate forming after the major folding barrier is crossed leading to formation of N (formation of the β1-β2-β3 sheet). Since we observe unfolding rather folding and N^* is visited multiple times before global unfolding, it is natural to think of it as a native substate; N^* is a native excitation around the folded state and is visited during native dynamics even when the protein is thermodynamically in the folded state. After the initial event, there is a bifurcation in the unfolding pathways; in some of the trajectories, β1-β2 ruptures first (Fig. 6b for example), while in others, β2-β3 unfolds first (Fig. 6c). Thus, the kinetic trajectories also show a two-state behavior, with two possible unfolding pathways.
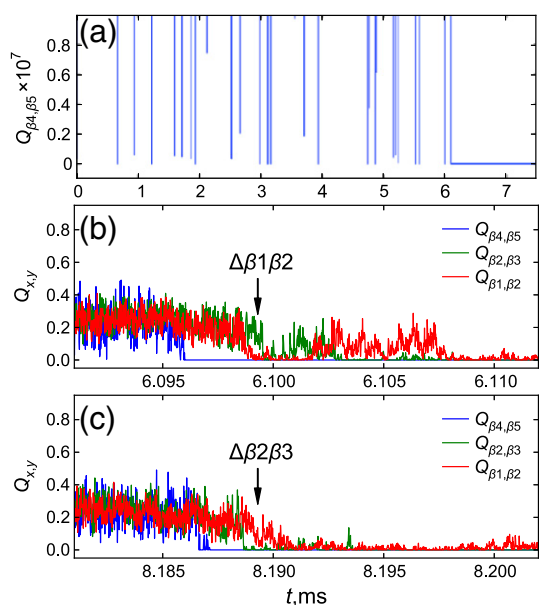
Fig. 6. Highlights of unfolding kinetics under stretching (trajectories at $f = 10$ pN, $T = 350$ K, $[C] = 0$ are shown). (a) The β4-β5 sheet melts and reforms on the order of 10^{-5} s before completely unfolding. (b and c) When β4-β5 stays unfolded long enough so that contacts in the other β-sheet start breaking, the β1-β2-β3 sheet also melts completing the unfolding. Contacts between β1 and β2 and those between β2 and β3 continue to transiently and partially reform even after reaching the unfolded state. β1-β2 breaks before the β2-β3 in trajectory shown in (b). The reverse sequence occurs in the trajectory shown in (c).

Dimerization

Our main hypothesis is that the excited N^* state is prone to aggregation. In order to test this hypothesis, we have simulated the process of dimerization, starting from the two monomers in N^* conformations, which were generated in the monomer simulations with one pair of β4 and β5 in contact and the other pair on the opposite sides (Fig. 7a). We performed overdamped Brownian dynamics simulations at $T = 277$ K, controlling the concentration of the protein by constraining the distance between the centers of mass to 15 Å (~ R_g of a monomer) [33]. On a relatively short timescale, we observed dimerization, through the stages presented in simulation snapshots in Fig. 7b–d: partial unfolding of RT-loops and the remaining β4 finding the β5 of the other molecule, forming the domain-swapped dimer (see Supplementary Movie 1). Thus, the most probable route to aggregation in this protein is through domain-swap mechanism, which was already established in a previous study [34]. Here, we explicitly show that domain-swapped structures form readily by accessing a high energy N^* state.

To investigate whether this process of swapping the β4 can provide a pathway to aggregation of multiple molecules, we also simulated addition of a third

monomer to the dimer. We started from a configuration where dimer is not formed with one of the sticky β4 strands being unstructured (Fig. 7c). With the same conditions for temperature and concentration as for dimer formation, we observed formation of a trimer (Fig. 7e and Supplementary Movie 2).

Discussion

N^* state as precursor to oligomerization

Theoretical arguments along with molecular dynamics simulations of peptides and proteins [14,35–37] have shown that aggregation is initiated when the protein (at least transiently) populates a high free energy N^* state. In the last few years, there have been several experiments on proteins of different lengths, with no sequence or structural similarity, in which the N^* states have been experimentally characterized [12,18,38]. The closest example that is very similar to the N^* state studied here arises in *G. gallus* Fyn SH3 domain [18]. Remarkably, Kay *et al*. identified using relaxation dispersion NMR experiments a state with low population in A39V/N53P/V55L mutant Fyn SH3 under native conditions [18] that is identical with our finding for the structurally similar src SH3 domain. The backbone structure of the folding intermediate determined for Fyn SH3 is similar to the native state everywhere except in regions adjacent to the N- and C-termini. A detailed comparison is provided in Fig. S7. Of particular note is that the C-terminal β5 strand is disordered in the Fyn SH3 intermediate. The melting of the β4-β5 sheet leaves the N-terminal region exposed, which can trigger oligomer formation. They further corroborated the aggregation propensity of the region by preparing a truncation mutant, which is a mimic of the intermediate (or the N^* state), and established that the resulting mutant forms amyloid-like fibrils with high β-strand content.

Interpreting the results of the NMR experiments on Fyn SH3 mutant using our findings, we conclude that the folding intermediate of Fyn SH3 and the native substate identified in the src SH3 from the tyrosine kinase are the N^* states. In other words, they are the ensembles of conformations in the spectrum of native excitations that each monomer must populate in the course of oligomerization and fibril formation. Although not linked to any known disease, SH3 domains are known to aggregate in *in vitro* experiments. In many proteins containing the SH3 domain, for instance, the growth factor receptor-bound protein 2 [39] or phosphatidylinositol 3′-kinase [40], SH3 is located at the N-terminus of the protein. It is likely that the mechanism of aggregation by exposure of the sticky N-terminal β-strand of SH3 might remain operational *in vivo* as well.
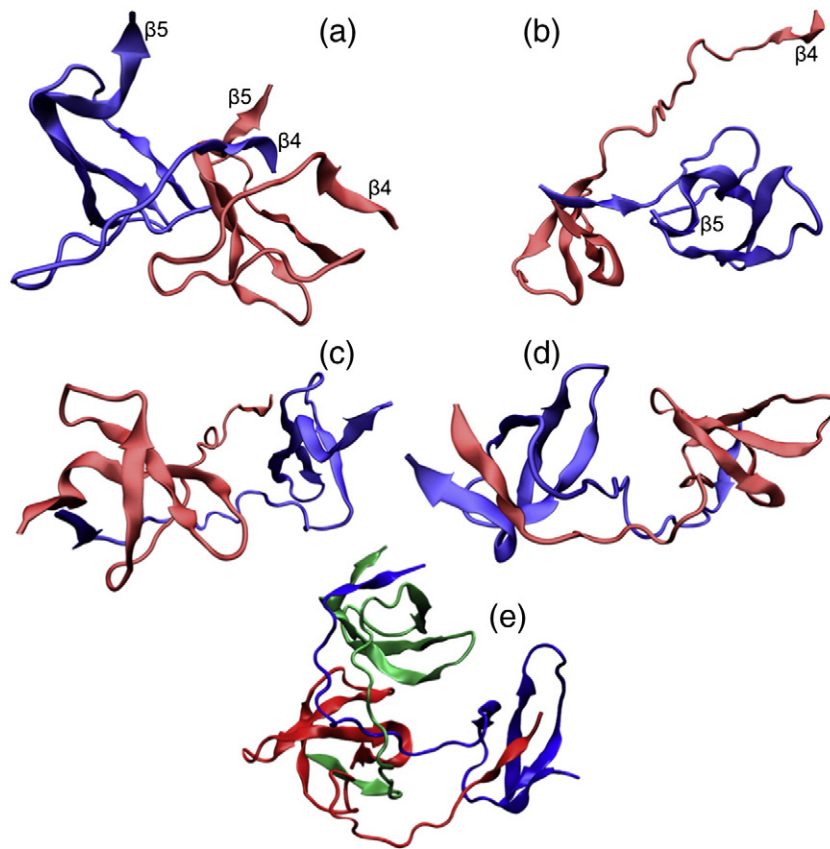
Fig. 7. Stages of dimerization through domain swapping. (a) Starting configuration, corresponding to two monomers in N^* conformation. (b) RT-loop of the red molecule unfolded with sticky β4 on its end; the blue RT-loop starts to unfold. (c) RT-loops unfolded. Red β4 continues diffusional motion around the blue monomer. (d) Red β4 finds forms a β-sheet by associating with the blue β5 in a domain-swapped dimer. (e) Trimer of the src SH3 domain formed during simulations of three SH3 molecules.

β2-Microglobulin and acylphosphatase populate the N^* under native conditions

Two other examples are also worth pointing out. In one of the earliest experiments, Radford *et al.* showed that aggregation of β2-microglobulin into amyloid-like fibrils occurs from a native-like (N^*) intermediate that has low population at equilibrium [38,41]. Despite the large barrier created by the proline frozen in the *cis* conformation, it was suggested that the N^* state is part of the native-stateensemble separated from the NBA by a large enough free energy gap that its population is low under native conditions. Finally, using NMR experiments and molecular dynamics with hydrogen/deuterium exchange data as restraints, authors reconstructed a free energy profile using RMSD of the C^α as the reaction coordinate for acylphosphatase was generated in the absence and presence of trifluoroethanol. These results suggest that a native-like N^* state becomes visible at non-zero concentration of trifluoroethanol. Just as in SH3 domains, interactions between two critical β-strands in acylphosphatase are disrupted exposing them to the solvent, thus making them aggregation prone [42]. Taken together, these results not only show that aggregation by populating N^* state is a generic mechanism but also reinforce earlier prediction that sequences of most natural proteins with predominantly β-sheet secondary structure have evolved so that intermolecular interactions between edge strands (β5 in SH3 for example) are unfavorable [43,41].

Population of N^* and fibril formation rates

In our previous work [44], we showed that the timescale of fibril formation is related to P_{N^*} (expressed in terms of percentage), the probability of populating the N^* state as

$$\tau_{fib} \approx \lambda \tau_F \exp(-C P_{N^*}(T)), \tag{4}$$

where τ_F is the folding time, $\lambda \approx 10^8$ and $C \approx 1$ [44]. It is tempting to estimate τ_{fib} for SH3 domain using the experimentally measured $\tau_F \approx 1.7 \times 10^{-2}$ s. The

value of $P_{N^*}(T)$ is 2% for Fyn SH3 domain [18], which coincides with the calculated value for src SH3 domain (see Figs. S8 and S9). With the use of τ_F and the estimate for $P_{N^*}(T)$, the abovementioned equation yields τ_{fib} on the order of 2–3 days, which is consistent with the measurements reported for Fyn SH3 [18].

A few implications for amyloidogenesis follow from the importance of P_{N^*}. (i) The fibril formation timescale should decrease dramatically as P_{N^*} increases. (ii) Since P_{N^*} (at a fixed temperature) depends on the free energy gap between N^* and N states, it follows that only those proteins for which there is a reasonable probability of being the N^* state under physiological conditions would aggregate in relevant timescales. (iii) The temperature dependence of τ_{fib} [Eq. (4)] is highly non-trivial because $P_{N^*}(T)$ would be a maximum at some optimal temperature and is expected to decrease at low temperatures. (iv) It is the free energy gap between N^* and the NBA, rather than the global stability, that determines the propensity of a protein to aggregate.

In our simulations, force was applied to the ends of the protein. However, in simulations that we performed with applying force to the residues 9 (N-terminal end) and 59 (before the β5-sheet in sequence), as in reported SMFS experiments with SH3 [19], the N^* is also accentuated by force, showing some robustness of the SMFS as a tool for highlighting the elusive states. We also note that although not easily detectable, N^* is present as the excitation in the native ensemble even at $f = 0$; thus, many of the implications for fibril formation do not depend on force (including the pulling points). Thus, the detectability of N^* to some extent does not depend on the pulling direction.

From N^* to oligomers to fibrils

Our simulations of the process of oligomer formation starting from N^* structures support the finding that N^* is aggregation prone. The ease of assembly of dimer and trimer by domain-swap mechanism allows for a route to aggregation in SH3 domain. Our trimer simulation shows that when the dimer is not fully formed, the sticky β4 dangling on the unfolded RT-loop associated not with counterpart in the dimer but with a third molecule in N^* state, leading to the formation of a trimer. In the formed dimer (or trimer), the contacts between β4 and β5 break once in a while, allowing for the sticky β4 strand of another molecule to come into proximity, leading to the growth of higher-order structures. This domain-swapping (or rather β-strand swapping) mechanism can include an arbitrary number of molecules: $β4_Aβ5_A$ monomer, $β4_Aβ5_B + β4_Bβ5_A$ dimer, $β4_Aβ5_B + β4_Bβ5_C + β4_Cβ5_A$ trimer, $β4_Aβ5_B + β4_Bβ5_C + β4_Cβ5_D + β4_Dβ5_A$ tetramer, and so on (Fig. 8). The strand exchange mechanism shows that, as the molecular weight of higher-order structures increases, the higher is the probability that

the β4 strand at the end of the assembly will dangle on the RT-loop rather than form contacts with β5 because it becomes harder to accommodate the RT-loops within the oligomer of growing size.

The domain-swapping mechanism of src SH3 aggregation has been reported more than a decade ago [34]. Our main message here is the importance of native fluctuations of a monomer leading to the population of the aggregation-prone state N^* and how this state can be gleaned by application of mechanical force. Our oligomerization simulations serve merely as an illustration of the principle that force can be used to access the N^* states. Subsequent growth of oligomers can occur by domain-swap mechanism as in SH3 or by other complex scenarios.

Rationale for using the coarse-grained models

In all computer simulations, it is important, but seldom practiced, to ascertain the extent to which the models capture reality. Our view is that the efficacy of the model can only be assessed by the ability to provide new insights and predict the outcomes of experiments. We have indeed produced a number of novel testable predictions using coarse-grained models. This level of falsifiable predictions cannot be currently achieved using all-atom simulations. We
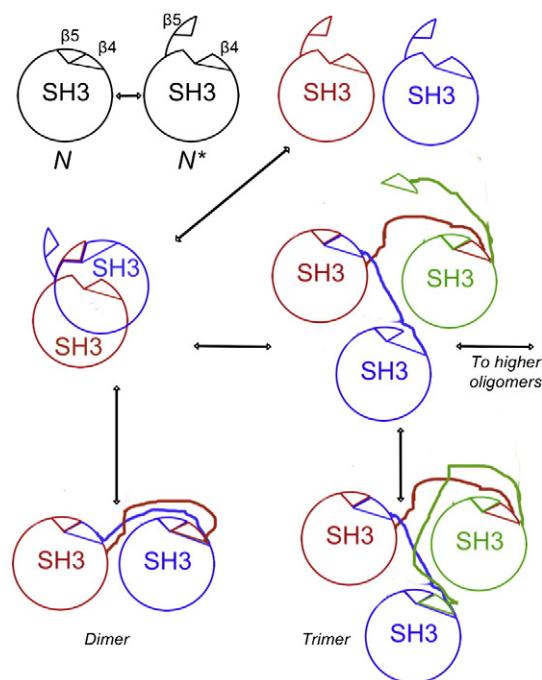

Fig. 8. The pathway to aggregation through the N^* state. First, if two molecules in N^* state come close together and one of the sticky β4 makes contacts with the other molecule, the RT-loops will unfold. Eventually, the other β4 can come into contact with β5 of another molecule forming a domain-swapped dimer. This process can be repeated resulting in fibril formation.

have shown in a number of studies that coarse-grained model used here has been remarkably successful in reproducing and predicting the force-stretching single-molecule experimental data in proteins [29,45–47] and RNA [46,48] including predictions that have been later validated quantitatively. The SOP-SC (*s*elf-*o*rganized *p*olymer model with *s*ide *c*hains) model used in this work is a more refined version of the self-organized polymer model. It has reproduced experimental results for the denaturant-induced unfolding [49,50]. The choice of the SOP-SC model, therefore, is very adequate for studying the questions we have answered in this work.

Conclusions

We have established, using *in silico* force spectroscopy of the src SH3 domain, that response of native proteins to mechanical force provides quantitative insights into protein aggregation. The force-denaturant phase diagram for the SH3 domain shows that the boundary separating the folded and unfolded states is universal [Eq. (2)] implying that determination of C_0, f_0, and β is sufficient to predict the phase diagram at any temperature.

Of particular importance is the demonstration that mechanical force can be used to elucidate a fine structure in the organization of the native state. In particular, force highlights the aggregation-prone N^* state in the native ensemble of SH3, by increasing its stability. The structure of the N^* state in src SH3 domain is identical with an aggregation-prone folding intermediate recently found in the Fyn SH3 domain in a relaxation dispersion NMR experiment [18]. Simulations of kinetics of unfolding under force shows that the N^* state has to be formed and linger for a substantial time period prior to global unfolding. Our work shows that force spectroscopy in combination with simulations carried out under the same conditions as experiments is a viable technique to characterize important substates in the spectrum of native excitations even if they have low populations, as is the case in SH3 domain.

A by-product of our study is that force (or more generally stress) can enhance the probability of oligomer and fibril formation. This implies that proteins operating against load (motor proteins for example) have evolved to minimize the tendency to sample N^* states by ensuring that the free energy gap separating the N^* and the native state is large.

It follows from our work that aggregation propensity of folded proteins under native conditions is determined by the accessibility of the N^* state, which in SH3 domains involves partial unfolding of the folded state. Thus, it is the free energy gap separating the NBA and N^* state ($\Delta G_{\text{NBA},N^*}$), rather than the global stability, which plays a critical role in protein aggregation. A corollary of this prediction is that mutations that increase (decrease) $\Delta G_{\text{NBA},N^*}$ would decrease (increase) aggregation probability. Protein aggregation, therefore, depends not only on sequence but also on $\Delta G_{\text{NBA},N^*}$, which can be manipulated by changing external conditions.

Our work also shows that by knowing the structure of N^* state it might be possible to construct structural models for oligomers, as we have performed for the src SH3 domain. The generality of this approach requires further work.

Methods

Model

Our simulations were performed using the SOP-SC [49] for the protein. In the SOP-SC model, each amino acid is represented using two interaction centers. The energy function in the SOP-SC representation of a polypeptide chain is taken to be

$$\mathcal{H} = U_{\text{LJ}}^{\text{nat}} + U_{\text{SS}}^{\text{nnat}} + U_{\text{SS}}^{\text{neib}} + U_{\text{FENE}}. \tag{5}$$

The $U_{\text{LJ}}^{\text{nat}}$ potential is based on the native structure (PDB structure in our case), which assigns an attraction between the residues that are in contact in the native state and are at least two residues apart along the chain. The form of $U_{\text{LJ}}^{\text{nat}}$ is

$$U_{\text{LJ}}^{\text{nat}} = \sum_{(j-i)>2} \epsilon_{\text{bb}} f_{\text{LJ}}\left(\boldsymbol{r}_i^{\text{bb}}, \boldsymbol{r}_j^{\text{bb}}\right) + + \sum_{(j-i)>2} \epsilon_{\text{ss}}\left|_{ij} - 0.7\right| f_{\text{LJ}}\left(\boldsymbol{r}_i^{\text{ss}}, \boldsymbol{r}_j^{\text{ss}}\right) + \\ + \sum_{|j-i|>2} \epsilon_{\text{bs}} f_{\text{LJ}}\left(\boldsymbol{r}_i^{\text{bs}}, \boldsymbol{r}_j^{\text{bs}}\right), \tag{6}$$

where the sums are over the residues. We use the Lennard-Jones function for the beads that are in contact in the native state (with the minimum at the native distance):

$$f_{\text{LJ}}(\boldsymbol{r}_1, \boldsymbol{r}_2) = \left[\left(\frac{r_{12}^0}{|\boldsymbol{r}_1 - \boldsymbol{r}_2|}\right)^{12} - 2\left(\frac{r_{12}^0}{|\boldsymbol{r}_1 - \boldsymbol{r}_2|}\right)^6\right]\Delta\left(r_{12}^0\right), \tag{7}$$

where $\Delta = 1$ if the native distance in the PDB $r_{12}^0 < R_c$ (we use $R_c = 8$ Å) and $\Delta = 0$ otherwise. We use the Betancourt–Thirumalai statistical potential for ϵ_{ij} [51]. Superscripts b and s stand for backbone bead and side chain bead, respectively.

Non-native interactions are modeled by soft sphere interactions:

$$U_{\text{SS}}^{\text{nnat}} = \sum_{(j-i)>2} \varepsilon_l \left(\frac{\sigma^{\text{bb}}}{r_{ij}^{\text{bb}}}\right)^6 \left(1 - \Delta\left(r_{ij}^{\text{bb},0}\right)\right) + \\ + \sum_{(j-i)>2} \varepsilon_l \left(\frac{\sigma_{ij}^{\text{ss}}}{r_{ij}^{\text{ss}}}\right)^6 \left(1 - \Delta\left(r_{ij}^{\text{ss},0}\right)\right) + \tag{8} \\ + \sum_{|j-i|>2} \varepsilon_l \left(\frac{\sigma_{ij}^{\text{bs}}}{r_{ij}^{\text{bs}}}\right)^6 \left(1 - \Delta\left(r_{ij}^{\text{bs},0}\right)\right),$$

where $\sigma^{bb} = 3.8$ Å is the average distance between neighboring C^{α} atoms and $\sigma_{ij}^{\alpha\beta}$ is the sum of van der Waals radii of corresponding beads (backbone or side chain depending on the amino acid).

Repulsive interactions between neighboring residues stabilize local secondary structure:

$$U_{SS}^{neib} = \sum_{(j-i)\leq 2} \varepsilon_l \left(\frac{\sigma^{bb}}{r_{ij}^{bb}}\right)^6 + \varepsilon_l \left(\frac{\sigma_{ij}^{ss}}{r_{ij}^{ss}}\right)^6 + \sum_{|j-i|\leq 2} \varepsilon_l \left(\frac{\sigma_{ij}^{bs}}{r_{ij}^{bs}}\right)^6. \quad (9)$$

The values of all σ and ε are the same as used in Ref. [49].

The chain connectivity is described by U_{FENE} or finite extensible nonlinear elastic potential:

$$U_{FENE} = \sum_{bonds} \frac{k}{2} R_0^2 \log\left(1 - \frac{l-l_0}{R_0^2}\right), \quad (10)$$

where the summation is over all bonds (whether between two backbone beads or between a backbone and a side-chain bead), l is the distance between the two bonded beads, and l_0 is the length of the bond in the native state. We used spring constant $k = 20$ kcal/mol/Å^2 and $R_0 = 2$ Å.

Model for denaturant

The molecular transfer model [49,52,53] is a phenomenological way to account for the effect of denaturants, where the effective interactions involving denaturant represent the free energy of transfer of a residue or backbone from pure water to the aqueous solution of the denaturant with concentration [C]. The transfer free energy term is a sum of terms proportional to the SASAs of individual residues (backbone and side chain):

$$\mathcal{H}_{[C]}(\boldsymbol{r}) = \mathcal{H}_{[0]}(\boldsymbol{r}) + \Delta G(\boldsymbol{r}, [C]) =$$
$$= \mathcal{H}_{[0]}(\boldsymbol{r}) + \sum_k \delta g(k, [C]) \frac{\alpha_k(\boldsymbol{r})}{\alpha_{Gly-k-Gly}}, \quad (11)$$

where the sum is over all backbone and side-chain interaction centers, $\delta g(k, [C]) = m_k[C] + b_k$ is the transfer free energy of the bead k, and m_k and b_k are different for the backbone and side chains and depend on the residue identity. The SASA is $\alpha_k(\boldsymbol{r})$ for bead k in the protein conformation r, and $\alpha_{Gly-k-Gly}$ is the SASA of the bead k in the tripeptide Gly-k-Gly. We used the same values of m_k, b_k, and $\alpha_{Gly-k-Gly}$ as in Ref. [49]. We used the procedure described in Ref. [54] to calculate the SASA.

Calculating SASA is computationally expensive. It is possible to avoid calculating $\Delta G(\boldsymbol{r}, [C])$ at every simulation step, if only equilibrium properties are of interest. If $\mathcal{H}_{[C]}$ is the full Hamiltonian, $\mathcal{H}_0$, and $-\Delta G(\boldsymbol{r}, [C])$ is the restriction potential similar to the one used in umbrella sampling simulation, it is possible to perform converged simulations using the Hamiltonian $\mathcal{H}_0$. The properties of the system under the unperturbed Hamiltonian $\mathcal{H}_0 = \mathcal{H}_{[C]}$ can then be calculated using the weighted histogram analysis method (WHAM) equations [55].

Multichain interactions

The non-bonded interactions between beads on different chains were the same as if they were on the same chain. That is, if two residues, for example, V61 and V11, form native contact, then V61 and V11 have the same native interaction (with εs and σs for valine–valine) even if they belong to different chains. Two residues that interact non-natively (i.e., via soft sphere repulsion) also behave the same way on different chains. Similar model for multichain interaction has been employed by Ding *et al.* [34] for studying of aggregation of src SH3; that is, the native interactions were the same between the residues whether on the same or different chains. The differences were that Ding *et al.* used pure Go model and added hydrogen bonds between the backbone atoms of different molecules.

Simulations

We used LAMMPS (*L*arge-scale *A*tomic/*M*olecular *M*assively *P*arallel *S*imulator) [56] to perform the simulations, enhancing it for the SOP-SC model. Langevin dynamics is performed as the numerical solution of the Langevin equation for every protein bead:

$$m\dot{\boldsymbol{v}} = -\zeta \boldsymbol{v} + \boldsymbol{F}_c + \boldsymbol{F}_r(t), \quad (12)$$

where m is the mass of the bead, $\boldsymbol{v}(t) = \dot{\boldsymbol{r}}$; F_c is the conservative forces acting on the bead (interactions with the other bead), $\boldsymbol{F}_c = \partial\mathcal{H}/\partial\boldsymbol{r}$; F_r is the white noise random force with $\langle F_r^i(t)\rangle = 0$ and $\langle F_r^i(t)F_r^j(t')\rangle = 2\zeta kT\delta_{ij}\delta(t-t')$. We integrated the Langevin equations using Verlet algorithm as implemented in LAMMPS [56,57]. The timescale in Eq. (12) is defined by ζ/m. The relevant timescale for low-friction limit where inertial term is significant in SOP-SC model is $\tau_L = \sqrt{m\sigma^2/\varepsilon_l} = 2$ ps. For low-friction limit (when simulating thermodynamics), we use $\zeta = 0.05$ m/τ_L. We used LAMMPS facilities for the replica exchange simulations. We made random exchange attempts at 100 τ_L intervals, with the same intervals between saving the trajectory snapshots for analysis (to calculate the transfer energies). For the overdamped limit (when simulating kinetics), where the inertial term in the Langevin equation is negligible, the relevant timescale is $\tau_H = \zeta\sigma^2/kT$. For high friction, we use $\zeta = 50m/\tau_L$, yielding $\tau_H \approx 150$ ps [49,58].

For calculating phase diagrams, we have performed replica exchange simulations at low friction and at the following fixed values of f: 0, 0.5, 1, 1.5, 2, 2.25, 2.5, 2.75, 3, 3.25, 3.5, 4, 4.5, 5, 5.5, 5.75, 6, 6.5, 7, 7.25, 7.5, 8, 8.25, 8.5, 9, 9.5, 10, 10.5, 11, 11.5, 12, 12.5, 13, 13.5, 14, and 15 pN, at [C] = 0. We used the sampled conformations at various temperatures to infer the distribution of NBA and denatured states at particular values of C and T for every f, using WHAM with the transfer energy in place of biasing potential as detailed in the subsection below. From individual replicas, we calculated distributions and free energy profiles as a function of end-to-end distance R for various values of f (at fixed T and [C] = 0) and T (for fixed f and [C] = 0).

For calculating unfolding kinetics, we ran 16 trajectories at 3 different temperatures, as detailed in the Table S1.

To illustrate dimerization and path to fibril formation, we started the simulation with two molecules in N^* conformation. We fixed the distance between their centers of mass at ~ 15 Å ($\sim R_g$) with a parabolic potential (as a way to simulate high concentration) and ran Brownian dynamics at $T = 277$ K until observing the formation of a dimer, just after about 100 µs. The whole process is shown in Supplementary Movie 1.

For the trimer formation, we started from a snapshot from the dimerization trajectory and added a third molecule in N^* state. We constrained the three distances between three centers of mass to the same ~ 15 Å. We observed the formation of trimer (i.e., the last $\beta 4$ finding its spot on another molecule next to its $\beta 5$) after about 1 ms. Supplementary Movie 2 shows ~ 50 µs preceding the formation of a trimer.

WHAM for equilibrium properties

For simulations with the Hamiltonian

$$\mathcal{H}(\mathbf{r}) = \mathcal{H}_0(\mathbf{r}) + \lambda V_{rest}(\mathbf{r}), \tag{13}$$

where V_{rest} is the restriction potential with a coupling constant λ, WHAM can be used to calculate equilibrium properties of the system with Hamiltonian $\mathcal{H}_0(\mathbf{r})$. In case of MTM, Eq. (13) becomes

$$\mathcal{H}_{[0]}(\mathbf{r}) = \mathcal{H}_{[C]}(\mathbf{r}) - \Delta G(\mathbf{r}, [C]). \tag{14}$$

The unnormalized probability density of an observable ξ from M simulations performed using $\mathcal{H}_{[0]}$ with ith simulation carried out at temperature T_i is [55]:

$$P(\xi, T_i, [C]) = \frac{\sum_{k=1}^{M} N_k(\xi) \exp(-\mathcal{H}_{[C]}/k_B T_i)}{\sum_{m=1}^{M} n_m \exp(f_m - \mathcal{H}_{[0]}/k_B T_m)}, \tag{15}$$

where $N_k(\xi)$ is the histogram of ξ in kth simulation and n_m is the number of snapshots (conformations) analyzed for mth simulation. The free energy of the ith simulation f_i is

$$f_i = -\log \sum_{\xi} P(\xi, T_i, [0]), \tag{16}$$

which can be found self-consistently using

$$f_i = -\log \sum_{k=1}^{M} \sum_{t=1}^{n_k} \frac{\exp(-\mathcal{H}_{[0]}(k,t)/k_B T_i)}{\sum_{m=1}^{M} n_m \exp(f_m - \mathcal{H}_{[0]}(k,t)/k_B T_m)}. \tag{17}$$

In Eq. (17), $\mathcal{H}_{[0]}(k, t)$ is the potential energy of the tth snapshot from the kth simulation (in the absence of denaturants).

After calculating the f_i, the averages for the MTM Hamiltonian (11) can be obtained using (15):

$$\langle \xi([C_i], T_i) \rangle =$$
$$= \frac{1}{Z([C_i], T_i)} \times \sum_{k=1}^{M} \sum_{t=1}^{n_k} \frac{\xi(k,t) \exp(-\mathcal{H}_{[C_i]}(k,t)/k_B T_i)}{\sum_{m=1}^{M} n_m \exp(f_m - \mathcal{H}_{[0]}(k,t)/k_B T_m)}, \tag{18}$$

with the partition function

$$Z([C_i], T_i) = \sum_{k=1}^{M} \sum_{t=1}^{n_k} \frac{\exp(-\mathcal{H}_{[C_i]}(k,t)/k_B T_i)}{\sum_{m=1}^{M} n_m \exp(f_m - \mathcal{H}_{[0]}(k,t)/k_B T_m)}, \tag{19}$$

where $\mathcal{H}_{[C_i]}(k,t) = \mathcal{H}_{[0]}(k,t) + \Delta G(k,t,[C_i])$, and $\xi(k, t)$, $\mathcal{H}_{[0]}(k, t)$, and $\Delta G(k, t, [C_i])$ are the values of ξ, potential energy, and transfer energy to the solution of concentration $[C_i]$ in the tth snapshot of kth simulation.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.jmb.2014.05.007.

References

[1] Schuler B, Eaton WA. Protein folding studied by single-molecule FRET. Curr Opin Struct Biol 2008;18(1):16–26.

[2] Lindorff-Larsen K, Piana S, Dror RO, Shaw DE. How fast-folding proteins fold. Science 2011;334(6055):517–20.

[3] Onuchic JN, Wolynes PG. Theory of protein folding. Curr Opin Struct Biol 2004;14(1):70–5.

[4] Shakhnovich E. Protein folding thermodynamics and dynamics: where physics, chemistry, and biology meet. Chem Rev 2006;106(5):1559–88.

[5] Thirumalai D, O'Brien EP, Morrison G, Hyeon C. Theoretical perspectives on protein folding. Ann Rev Biophys 2010;39(39):159–83.

[6] Shea J-E, Urbanc B. Insights into Aβ aggregation: a molecular dynamics perspective. Curr Top Med Chem 2012;12(22):2596–610.

[7] Straub JE, Thirumalai D. Principles governing oligomer formation in amyloidogenic peptides. Curr Opin Struct Biol 2010;20(2):187–95.

[8] Aguzzi A, O'Connor T. Protein aggregation diseases: pathogenicity and therapeutic perspectives. Nat Rev Drug Discov 2010;9(3):237–48.

[9] Chiti F, Dobson CM. Protein misfolding, functional amyloid, and human disease. Annu Rev Biochem 2006;75:333–66.

[10] Bucciantini M, Giannoni E, Chiti F, Baroni F, Formigli L, Zurdo J, et al. Inherent toxicity of aggregates implies a common mechanism for protein misfolding diseases. Nature 2002;416(6880):507–11.

[11] Goldschmidt L, Teng PK, Riek R, Eisenberg D. Identifying the amylome, proteins capable of forming amyloid-like fibrils. Proc Natl Acad Sci U S A 2010;107(8):3487–92.

[12] De Simone A, Dhulesia A, Soldi G, Vendruscolo M, Hsu S-TD, Chiti F, et al. Experimental free energy surfaces reveal the mechanisms of maintenance of protein solubility. Proc Natl Acad Sci U S A 2011;108(52):21057–62.

[13] Tartaglia GG, Pawar AP, Campioni S, Dobson CM, Chiti F, Vendruscolo M. Prediction of aggregation-prone regions in structured proteins. J Mol Biol 2008;380(2):425–36.

[14] Thirumalai D, Klimov DK, Dima RI. Emerging ideas on the molecular basis of protein and peptide aggregation. Curr Opin Struct Biol 2003;13(2):146–59.

[15] Tarus B, Straub JE, Thirumalai D. Dynamics of Asp23-Lys28 salt-bridge formation in Abeta10-35 monomers. J Am Chem Soc 2006;128(50):16159–68.

[16] Massi F, Straub JE. Energy landscape theory for Alzheimer's amyloid beta-peptide fibril elongation. Proteins 2001;42(2):217–29.

[17] Hu W, Walters BT, Kan Z-Y, Mayne L, Rosen LE, Marquesee S, et al. Stepwise protein folding at near amino acid resolution by hydrogen exchange and mass spectrometry. Proc Natl Acad Sci U S A 2013;110(19):7684–9.

[18] Neudecker P, Robustelli P, Cavalli A, Walsh P, Lundstrom P, Zarrine-Afsar A, et al. Structure of an intermediate state in protein folding and aggregation. Science 2012;336(6079):362–6.

[19] Jagannathan B, Elms PJ, Bustamante C, Marquesee S. Direct observation of a force-induced switch in the anisotropic mechanical unfolding pathway of a protein. Proc Natl Acad Sci U S A 2012;109(44):17820–5.

[20] Jagannathan B, Marquesee S. Protein folding and unfolding under force. Biopolymers 2013;99(11):860–9. http://dx.doi.org/10.1002/bip.22321.

[21] Fernández-Escamilla AM, Cheung MS, Vega MC, Wilmanns M, Onuchic JN, Serrano L. Solvation in protein folding analysis: combination of theoretical and experimental approaches. Proc Natl Acad Sci U S A 2004;101(9):2834–9.

[22] Grantcharova VP, Baker D. Folding dynamics of the src SH3 domain. Biochemistry 1997;36(50):15685–92.

[23] Grantcharova VP, Riddle DS, Baker D. Long-range order in the src SH3 folding transition state. Proc Natl Acad Sci U S A 2000;97(13):7084–9.

[24] Riddle DS, Grantcharova VP, Santiago JV, Alm E, Ruczinski I, Baker D. Experiment and theory highlight role of native state topology in SH3 folding. Nat Struct Biol 1999;6(11):1016–24.

[25] Martínez JC, Serrano L. The folding transition state between SH3 domains is conformationally restricted and evolutionarily conserved. Nat Struct Biol 1999;6(11):1010–6.

[26] Efron B. The Jackknife, the Bootstrap, and Other Resampling Plans. Philadelphia, PA: Society for Industrial and Applied Mathematics; 1982.

[27] Klimov DK, Thirumalai D. Stretching single-domain proteins: phase diagram and kinetics of force-induced unfolding. Proc Natl Acad Sci U S A 1999;96(11):6166–70.

[28] Keefe PD. Quantum mechanics and the second law of thermodynamics: an insight gleaned from magnetic hysteresis in the first order phase transition of an isolated mesoscopic-size type I superconductor. Phys Scr 2012;2012(T151):014029.

[29] Hinczewski M, Gebhardt JCM, Rief M, Thirumalai D. From mechanical folding trajectories to intrinsic energy landscapes of biopolymers. Proc Natl Acad Sci U S A 2013;110(12):4500–5.

[30] Greenleaf WJ, Block SM. Single-molecule, motion-based DNA sequencing using RNA polymerase. Science 2006;313(5788):801.

[31] Stigler J, Ziegler F, Gieseke A, Gebhardt JCM, Rief M. The complex folding network of single calmodulin molecules. Science 2011;334(6055):512–6.

[32] Hardin C, Eastwood MP, Luthey-Schulten Z, Wolynes PG. Associative memory hamiltonians for structure prediction without homology: alpha-helical proteins. Proc Natl Acad Sci U S A 2000;97(26):14235–40.

[33] Dima RI, Thirumalai D. Exploring protein aggregation and self-propagation using lattice models: phase diagram and kinetics. Protein Sci Publ Protein Soc 2002;11(5):1036–49.

[34] Ding F, Dokholyan NV, Buldyrev SV, Stanley HE, Shakhnovich EI. Molecular dynamics simulation of the SH3 domain aggregation suggests a generic amyloidogenesis mechanism. J Mol Biol 2002;324(4):851–7.

[35] De Simone A, Kitchen C, Kwan AH, Sunde M, Dobson CM, Frenkel D. Intrinsic disorder modulates protein self-assembly and aggregation. Proc Natl Acad Sci U S A 2012;109(18):6951–6.

[36] Krobath H, Estacio SG, Faisca PFN, Shakhnovich EI. Identification of a conserved aggregation-prone intermediate state in the folding pathways of Spc-SH3 amyloidogenic variants. J Mol Biol 2012;422(5):705–22.

[37] Rosenman DJ, Connors CR, Chen W, Wang C, García AE. Aβ monomers transiently sample oligomer and fibril-like configurations: ensemble characterization using a combined MD/NMR approach. J Mol Biol 2013;425(18):3338–59.

[38] Eichner T, Radford SE. A generic mechanism of beta2-microglobulin amyloid assembly at neutral pH involving a specific proline switch. J Mol Biol 2009;386(5):1312–26.

[39] Lowenstein EJ, Daly RJ, Batzer AG, Li W, Margolis B, Lammers R, et al. The SH2 and SH3 domain-containing protein GRB2 links receptor tyrosine kinases to ras signaling. Cell 1992;70(3):431–42.

[40] Chen JK, Schreiber SL. SH3 domain-mediated dimerization of an n-terminal fragment of the phosphatidylinositol 3-kinase p85 subunit. Bioorg Med Chem Lett 1994;4(14):1755–60.

[41] Jahn TR, Parker MJ, Homans SW, Radford SE. Amyloid formation under physiological conditions proceeds via a native-like folding intermediate. Nat Struct Mol Biol 2006;13(3):195–201.

[42] Bemporad F, Chiti F. "Native-like aggregation" of the acylphosphatase from *Sulfolobus solfataricus* and its biological implications. FEBS Lett 2009;583(16):2630–8.

[43] Richardson JS, Richardson DC. Natural beta-sheet proteins use negative design to avoid edge-to-edge aggregation. Proc Natl Acad Sci U S A 2002;99(5):2754–9.

[44] Li MS, Co NT, Reddy G, Hu C-K, Straub JE, Thirumalai D. Factors governing fibrillogenesis of polypeptide chains revealed by lattice models. Phys Rev Lett 2010;105(21):218101.

[45] Mickler M, Dima RI, Dietz H, Hyeon C, Thirumalai D, Rief M. Revealing the bifurcation in the unfolding pathways of GFP by using single-molecule experiments and simulations. Proc Natl Acad Sci U S A 2007;104(51):20268–73.

[46] Hyeon C, Dima RI, Thirumalai D. Pathways and kinetic barriers in mechanical unfolding and refolding of RNA and proteins. Structure (London, England: 1993) 2006;14(11):1633–45.

[47] Hyeon C, Thirumalai D. Capturing the essence of folding and functions of biomolecules using coarse-grained models. Nat Commun 2011;2:487.

[48] Lin J-C, Thirumalai D. Relative stability of helices determines the folding landscape of adenine riboswitch aptamers. J Am Chem Soc 2008;130(43):14080–1.

[49] Liu Z, Reddy G, O'Brien EP, Thirumalai D. Collapse kinetics and chevron plots from simulations of denaturant-dependent folding of globular proteins. Proc Natl Acad Sci U S A 2011;108(19):7787–92.

[50] Reddy G, Liu Z, Thirumalai D. Denaturant-dependent folding of GFP. Proc Natl Acad Sci U S A 2012;109(44):17832–8.

[51] Betancourt MR, Thirumalai D. Pair potentials for protein folding: choice of reference states and sensitivity of predicted native states to variations in the interaction schemes. Protein Sci Publ Protein Soc 1999;8(2):361–9.

[52] O'Brien EP, Ziv G, Haran G, Brooks BR, Thirumalai D. Effects of denaturants and osmolytes on proteins are accurately predicted by the molecular transfer model. Proc Natl Acad Sci U S A 2008;105(36):13403–8.

[53] Liu Z, Reddy G, Thirumalai D. Theory of the molecular transfer model for proteins with applications to the folding of the src-SH3 domain. J Phys Chem B 2012;116(23):6707–16.

[54] Hayryan S, Hu C-K, Skrivánek J, Hayryan E, Pokorný I. A new analytical method for computing solvent-accessible surface area of macromolecules and its gradients. J Comput Chem 2005;26(4):334–43.

[55] Kumar S, Rosenberg JM, Bouzida D, Swendsen RH, Kollman PA. The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. J Comput Chem 1992;13(8):1011–21.

[56] Plimpton S. Fast parallel algorithms for short-range molecular dynamics. J Comput Phys 1995;117:1–19.

[57] Verlet L. Computer "experiments" on classical fluids. I. Thermodynamical properties of Lennard-Jones molecules. Phys Rev 1967;159:98–103.

[58] Veitshans T, Klimov D, Thirumalai D. Protein folding kinetics: timescales, pathways and energy landscapes in terms of sequence-dependent properties. Fold Des 1997;2(1):1–22.