Protein folding guides disulfide bond formation

Meng Qin[a,b], Wei Wang[a,1], and D. Thirumalai[b,1]

[a]National Laboratory of Solid State Microstructure, Department of Physics, and Collaborative Innovation Center of Advanced Microstructures, Nanjing University, Nanjing 210093, China; and [b]Biophysics Program, Institute for Physical Science and Technology, University of Maryland, College Park, MD 20742

Edited by Harold A. Scheraga, Cornell University, Ithaca, NY, and approved June 26, 2015 (received for review February 25, 2015)

The Anfinsen principle that the protein sequence uniquely determines its structure is based on experiments on oxidative refolding of a protein with disulfide bonds. The problem of how protein folding drives disulfide bond formation is poorly understood. Here, we have solved this long-standing problem by creating a general method for implementing the chemistry of disulfide bond formation and rupture in coarse-grained molecular simulations. As a case study, we investigate the oxidative folding of bovine pancreatic trypsin inhibitor (BPTI). After confirming the experimental findings that the multiple routes to the folded state contain a network of states dominated by native disulfides, we show that the entropically unfavorable native single disulfide [14–38] between Cys_{14} and Cys_{38} forms only after polypeptide chain collapse and complete structuring of the central core of the protein containing an antiparallel β-sheet. Subsequent assembly, resulting in native two-disulfide bonds and the folded state, involves substantial unfolding of the protein and transient population of nonnative structures. The rate of [14–38] formation increases as the β-sheet stability increases. The flux to the native state, through a network of kinetically connected native-like intermediates, changes dramatically by altering the redox conditions. Disulfide bond formation between Cys residues not present in the native state are relevant only on the time scale of collapse of BPTI. The finding that formation of specific collapsed native-like structures guides efficient folding is applicable to a broad class of single-domain proteins, including enzyme-catalyzed disulfide proteins.

disulfide proteins | native-like interactions | enzyme-catalyzed folding | early collapse | nonnative interactions

The landmark discovery that the information to fold a protein is fully contained in the primary amino acid sequence was based on oxidative refolding experiments on disulfide bond formation in ribonuclease A (RNase A) (1, 2). Anfinsen showed that the initially unfolded protein, generated by reducing the disulfide (S–S) bonds in the native state of RNase A, folds reversibly under oxidizing conditions by correctly reforming the four native S–S bonds (among 105 possibilities) between the eight cysteine (Cys) residues. Besides being central to the enunciation of the principles of protein folding, many secretary proteins, whose misfolding is linked to a number of diseases, contain S–S bonds (3). Although biophysical aspects of such proteins are not as well studied as those without S–S bonds, understanding the link between conformational folding coupled to disulfide bond formation (4–7) is important and challenging both from a chemical and biophysical perspective (8).

The formation of S–S bonds and their identities during folding can be monitored by quenching the oxidative process at various stages of the folding reaction (9). By arresting the reaction, it is possible to characterize the accumulated intermediates in terms of already formed S–S bonds (10). However, the relationship between protein folding and disulfide bond formation is nontrivial to establish because this requires separate reporters for disulfide bond formation and organization of the rest of the polypeptide chains. Even if the reaction can be arrested rapidly, the conformations of the intermediates are difficult to determine using experiments alone, although single molecule pulling experiments hold exceptional promise (7). Thus, well-calibrated computations are needed to decipher the precise relationship between conformational folding and S–S bond formation (11–14).

Here, we investigate the coupling between conformational folding and disulfide bond formation by creating a novel way to mimic the effect of disulfide bond formation and rupture in coarse-grained (CG) molecular simulations, which have proven useful in a number of applications (15–18). As a case study, we use the 58-residue bovine pancreatic trypsin inhibitor (BPTI) with three S–S bonds in the native state to illustrate the key structural changes that occur during the folding reaction. The pioneering experiments of Creighton (9) seemed to indicate that nonnative disulfide species (19–22) are obligatory for productive folding to occur (for a thoughtful analysis, see ref. 23). Subsequently, using acid quench technique (by lowering pH, resulting in slowing down of the thiol disulfide exchange reaction) and a superior way of separating the intermediates Weissman and Kim (24) found that only native single and multiple disulfide bonds accumulate during the folding process. A plausible resolution of these contradictory findings was provided using theoretical studies (5) and simulations using lattice models (12) showing that nonnative intermediates are formed only on the time scale of the global collapse of the polypeptide chain. On longer times, only native species (S–S bonds found only in the folded state) dominate, as surmised by Weissman and Kim (24).

The experimental studies could not resolve whether disulfide bond formation drives protein folding or vice versa, and has remained a major unsolved problem in protein folding. To solve this problem, we created a novel computational method to mimic disulfide bond formation and rupture within the context of a C_α representation of polypeptide chain by building on the demonstration by Scheraga and coworkers (6) that the formation or disruption of S–S bonds in these proteins can occur only if a few structurally important criteria (proximity of Cys residues, orientation, and accessibility of thiol groups to oxidative agents, see Figs. S1–S3) are met. We incorporated this physical insight in our model and simulated the oxidative folding of BPTI. Our results quantitatively capture the relative importance of all single and two disulfide intermediates that direct folding of BPTI. The initial rapid formation of single disulfide

Significance

Anfinsen inferred the principles of protein folding by studying a protein containing four disulfide bonds in the native state. However, how protein folding drives disulfide bond formation is poorly understood despite the role such proteins play in variety of extracellular and intracellular functions. We developed a method to mimic the complex chemistry of disulfide bond formation in molecular simulations, which is used to decipher the mechanism of folding of bovine pancreatic trypsin inhibitor. The flux to the native state is through a complex network of native-like states. Crucially, we establish that it is the conformational folding of the protein that directs disulfide bond formation. Testable predictions and implications for enzyme-catalyzed folding of disulfide proteins in vivo are outlined.

Author contributions: M.Q., W.W., and D.T. designed research; M.Q. and D.T. performed research; M.Q., W.W., and D.T. contributed new reagents/analytic tools; M.Q. and D.T. analyzed data; and D.T. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence may be addressed. Email: thirum@umd.edu or wangwei@nju.edu.cn.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1503909112/-/DCSupplemental.

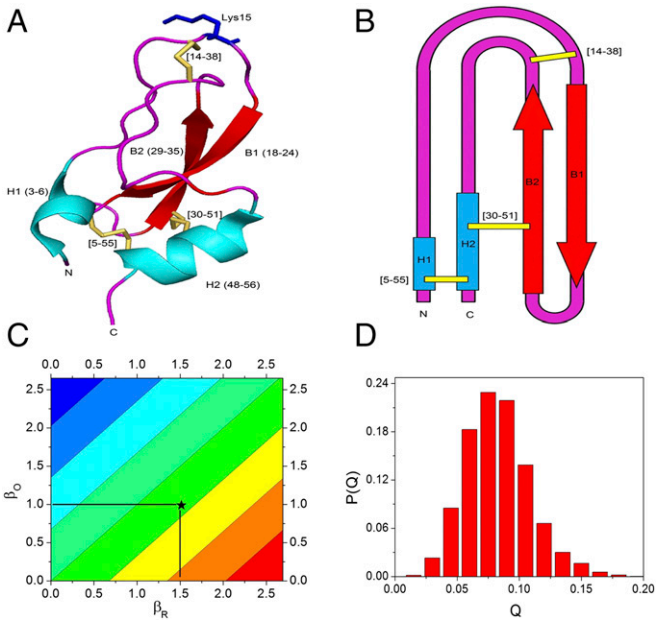

Fig. 1. (*A*) Ribbon diagram of the native structure of the 58 residue BPTI containing three disulfide bonds (marked in yellow) between residues Cys_5 and Cys_{55} [5–55], Cys_{14} and Cys_{38} [14–38], and Cys_{30} and Cys_{51} [30–51], respectively. The antiparallel β-sheet is in red. (*B*) Simplified representation of the secondary tructure of BPTI and the three native disulfide bonds in BPTI. (*C*) Variables β_O and β_R mimicking the redox conditions. Small β_O (β_R) represent strongly oxidizing (reducing) condition. The star with $\beta_O = 1.0$ and $\beta_R = 1.5$ is used in most of the simulations. These values are a mixture of mildly oxidizing and reducing condition. (*D*) Distribution of fraction of native contacts obtained from high-temperature simulations.

intermediates (in particular [14–38], an intermediate with disulfide bond between Cys_{14} and Cys_{38}), occurs only after substantial compaction of BPTI and complete structuring of the central antiparallel β-sheet shown in Fig. 1*A*. Formation of two-disulfide intermediates and the species N_{SH}^{SH} ([5–55, 30–51]) that is poised to fold rapidly to the folded state **N** requires substantial unfolding of BPTI. Loop formation dictated by entropic considerations and forces that drive chain compaction place the Cys residues in proximity to enable S–S bond formation, thus directing BPTI to the folded state (5). Our work also provides a general framework to simulate oxidative folding of disulfide-containing proteins, and firmly establishes that early formation of specifically collapsed structures results in efficient folding of single domain proteins.

Results

The native state of BPTI has three disulfide bonds between Cys_5 and Cys_{55}, Cys_{14} and Cys_{38}, and Cys_{30} and Cys_{51} (Fig. 1*A*). The 5–55 disulfide bond is located between the two terminal helices [helix α_1 (from Asp3 to Glu7) near the N terminus and α_2 (Ala48 to Gly56) in C terminus]. The maximally solvent-exposed [14–38] disulfide bond is near the terminus of the β-hairpin (formed between the β_1-strand from Ile18 to Asn24, and β_2-strand from Leu29 to Tyr35). The 30–51 disulfide bond bridges the β_1-strand and α_2 (Fig. 1*B*). When the three native disulfide bonds are reduced BPTI unfolds, resulting in the **R** state. The two-disulfide species [5–55, 14–38] is designated as **N***, [14–38, 30–51] is termed **N**′, and [5–55, 30–51] is $\mathbf{N}_{SH}^{SH}$. The native BPTI [5–55, 14–38, 30–51] is labeled **N**. In this standard description, the states of BPTI are described solely in terms of the [S–S] bonds without consideration of the conformations of the rest of the polypeptide chain.

Folding Network in Terms of Intermediates. To validate our simulations by direct comparisons with experiments, we map the folding pathways of BPTI in terms of the accumulated S–S bond intermediates. The flux map in Fig. 2, obtained using 2,000 folding trajectories, quantifies the flow toward the native state from any given intermediate in the folding network. In the earliest stage of folding, nearly 84% of the reduced BPTI rapidly forms the [14–38] intermediate (Fig. 2). The populations of [30–51] and [5–55] in the initial stages of BPTI folding are considerably less under the simulated redox conditions (Fig. 1*C*). The finding that the metastable [14–38] is the first native disulfide to form rapidly before further rearrangement agrees with theoretical predictions (5) and subsequent experimental validation (25, 26). As folding progresses, the kinetically unstable [14–38] rearranges to form the more stable [30–51] and [5–55] (Fig. 2). In our simulations, which mimic mildly oxidizing conditions, the transition from [14–38] to [5–55] and [30–51] involves transient population of compact **R**′ ensemble [(not the same as **R** ensemble) structures devoid of S–S bonds but with persistent secondary structures (see below)].

Formation of Two-Disulfide Species and the Native State. The single disulfide species rearrange with varying probabilities to form intermediates containing two native S–S bonds. Based on the cumulative analysis of all of the oxidizing events, the predominant species are [14–38, 30–51] (**N**′) and [5–55, 14–38] (**N***) rather than the productive [5–55, 30–51] ($\mathbf{N}_{SH}^{SH}$). The relative equilibrium fluxes show that the likelihood of populating **N*** from [5–55] is nearly twice as large as the formation of $\mathbf{N}_{SH}^{SH}$. Similarly, **N**′ formation from [30–51] is nearly six times more likely than $\mathbf{N}_{SH}^{SH}$. Because among the single native disulfide species [30–51] forms with the highest probability after the population of [14–38] decays, it follows that formation of intermediates with two S–S bonds occurs by conformational rearrangement involving [30–51] (25).

The final step is the rearrangement of the two-disulfide species to **N**. The flux map in Fig. 2 shows that only a very small fraction of **N*** and **N**′ states is kinetically connected to **N**. Because the population of **N*** is less than **N**′, we conclude that **N*** is a kinetic trap. Our simulations show that these dead-end species must

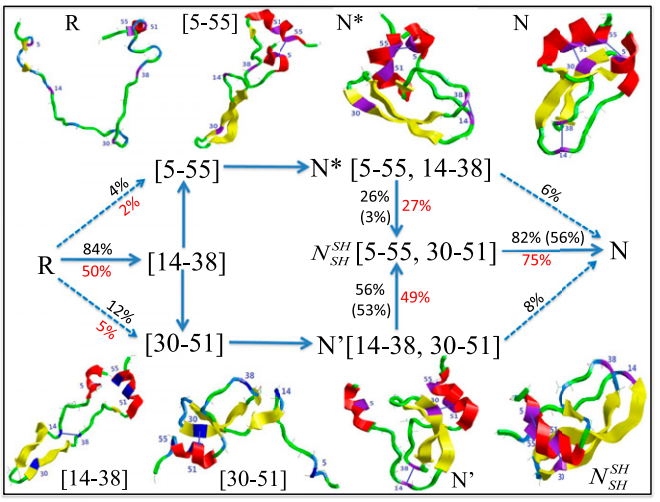

Fig. 2. The folding pathway, represented as a network of native-like states, connecting the fully unfolded state (**R**) to the folded (**N**) state with three disulfide bonds. For $\beta_O = 1.0$ and $\beta_R = 1.5$, the early event produces predominantly [14–38]. Subsequently, there is a bifurcation in the pathway with this intermediate rearranging to [5–55] and [30–51]. The percentages indicate the dominant route to **N** from **R**. Representative structures of all of the relevant states are shown. The numbers in parentheses were obtained from simulations with $\beta_O = 2.0$ and $\beta_R = 4.0$, which mimics the redox conditions used by Weissman and Kim (24). In red are the fluxes through the various native intermediates to the native state obtained from simulations that consider native and nonnative disulfide bond formation. The qualitative agreement between the two simulations is striking.

undergo rearrangement by reduction of [14–38], and subsequent formation of the productive two disulfide species, $\mathbf{N}_{SH}^{SH}$. In this process nonnative species are transiently populated in some of the folding trajectories (see below). The simulated folding pathway is in very good agreement with experiments, setting the stage for us to elucidate how protein folding controls disulfide bond formation.

Flux Map Depends on Redox Conditions. The range of redox conditions is approximately mimicked by varying the $[\beta_O, \beta_R]$ values. One way in which redox conditions are altered in in vitro experiments is by changing the buffer concentrations of glutathione disulfide (GSSH) and glutathione (GSH) (27). To assess the effect of changing redox conditions on BPTI folding we performed simulations with $\beta_O = 2.0$ and $\beta_R = 4.0$, which mimics the conditions used by Weissman and Kim (24). The large value of β_R, corresponding to weak reducing conditions, means that the probability of an already formed [S–S] bond has very small probability of undergoing reduction (see *SI Text* for details). It is in this sense that the $[\beta_O, \beta_R]$ values used in these simulations qualitatively mimic the Weissman and Kim (24) conditions. The percentage of molecules that reach **N** is now considerably less, and the transition from $\mathbf{N}^*$ is so slow that only a very small (the values are given in parenthesis in Fig. 2) fraction of $\mathbf{N}^*$ reaches the $\mathbf{N}_{SH}^{SH}$. Thus, $\mathbf{N}^*$ is effectively a dead-end kinetic trap. The overall findings are in excellent agreement with experiments (24).

Interestingly, the rate of formation and the fraction of **R** ensemble that reaches [14–38] do not depend on the redox conditions. To establish the robustness of this finding, we varied β_O with β_R fixed at 1.5 (Fig. 1*C*). Neither the yield nor the time ($\tau_{[14\text{-}38]}$) for forming [14–38] changes appreciably as β_O changes (Fig. 3*A*). Thus, [14–38] formation depends predominantly on the statistics of loop formation and compaction of BPTI resulting in the near complete ordering of the β-hairpin that brings Cys_{14} and Cys_{38} in proximity (5).

Kinetics of Single Disulfide Bond Formation. We quantify the kinetics of formation of single disulfide species using the time-dependent changes in the population, $P_\alpha(t) = \int_0^t P_{fp}^\alpha(s)ds$, where α refers to [14–38], [30–51], or [5–55], and $P_{fp}^\alpha(s)$ is the distribution of first passage time for the formation of the α^{th} species. We calculated $P_{fp}^\alpha(s)$ by determining the first time the α^{th} species is reached in each folding trajectory. The results in Fig. 3*B* for $P_u^\alpha(t) = 1 - P_\alpha(t)$ show that although [14–38] forms rapidly with substantial probability (Fig. 2) (relative to [5–55] or [30–51]), it decays also rapidly rearranging to [5–55] and [30–51]. Closer inspection of $P_u^\alpha(t)$ for [5–55] and [30–51] shows a delay in the decay that is absent in $P_{[14-38]}(t)$, implying that only after [14–38] forms does one observe population of the other two native single disulfide intermediates. Because [14–38] is exposed to solvent, it is vulnerable to further oxidation or reduction. In contrast, both [5–55] and [30–51] are buried (Fig. 1*A*), and hence once they are formed they are not as vulnerable to further rearrangement as [14–38]. Consequently, [14–38] is unstable and rearranges to the more stable [30–51] and [5–55] involving substantial unfolding (Fig. S4) through structures that are more compact than reduced BPTI (Fig. S5).

Collapse and Complete Ordering of β-Hairpin Precede Formation of Disulfide Bond. The connection between S–S bond formation and folding is best captured by analyzing a typical folding trajectory shown in Fig. 4*A* (Movie S1). Another example is given in Fig. S6. A number of inferences can be drawn from Fig. 4*A*: (*i*) Fig. 4*A* shows that before the formation of a single S–S bond, BPTI is compact with R_g decreasing from 24 Å in **R** state to less than 15 Å; (*ii*) upon compaction, Cys_{14} and Cys_{38} are in proximity with correct orientation resulting in the formation of the S–S bond. As reflected in Fig. 4*A*, [14–38] forms ahead of [5–55] and [30–51]. Most importantly, before formation of [14–38], the antiparallel β-hairpin is fully structured and the terminal α-helices adopt native-like structures. The rest of BPTI is disordered. We substantiate this finding quantitatively by a plotting $d_{[14-38]}$ as a function of the fraction of the native contacts, Q_β, between the two β-strands. The near perfect anticorrelation between these quantities (Fig. 4*B*) shows that as Q_β increases, the distance between residues 14 and 38 decreases,

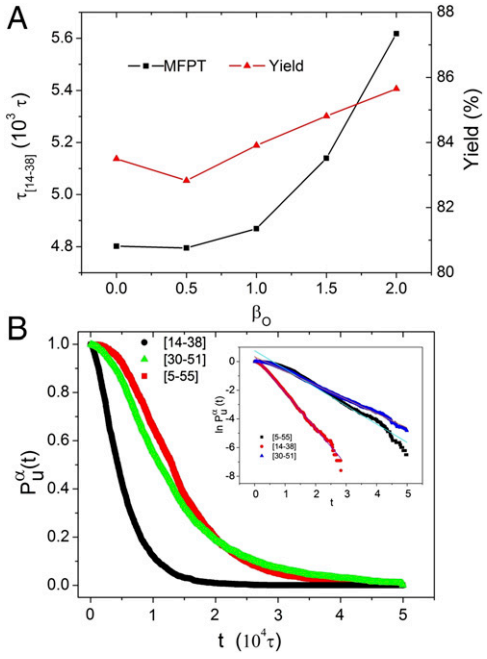

Fig. 3. (*A*) Dependence of the mean first passage time (black line) for forming [14–38] as the oxidizing condition is changed from being strong (small β_O) to weak (large β_O). The red line gives the yield of [14–38]. (*B*) Time-dependent decay, $P_u^\alpha(t)$, of the three native single disulfide species. (*Inset*) $lnP_u^\alpha(t)$; the lines are linear fits.

thus facilitating the formation of [14–38]. Even at $d_{[14-38]} \approx 12$ Å, a value that is too large for [14–38] to form, Q_β adopts near native value (Fig. 4*B*), implying that the core hairpin formation is complete before [14–38] formation starts. (*iii*) When $d_{[14-38]} \approx 5.7$ Å, which is close enough to form a disulfide bond, the distribution of Q_β is peaked at a value that exceeds 0.8 (Fig. 4*C*), implying complete ordering of the hairpin. (*iv*) As time progresses, [14–38] undergoes oxidation and reduction multiple times under the conditions of simulations (Fig. 4*A* and Fig. S6) because it is exposed to the solvent. Although kinetically favored, the low stability of [14–38] results in its disruption as folding proceeds. Thus, collapse of the polypeptide chain and complete structuring of the β-hairpin followed by [14–38] formation are key events that direct protein folding.

Rearrangement of $\mathbf{N}^*$ and $\mathbf{N}'$ to $\mathbf{N}_{SH}^{SH}$ Involves Nonnative Species. Although predominantly native-like interactions drive disulfide bond formation, experiments (24) showed that rearrangement of $\mathbf{N}'$ ([30–51, 14–38]) to $\mathbf{N}_{SH}^{SH}$ involves transient population of nonnative species. In several of the folding trajectories, such as the one in Fig. 5*A*, we find that in the conversion process between two native-like disulfide bond intermediates nonnative species, [30–51, 5–38] and [30–51, 5–14], are transiently populated. These were precisely the ones identified in experiments (see figure 6*A* in ref. 24). Interestingly, Fig. 5*A* also shows that the protein unfolds substantially because establishing the disulfide bond between 5 and 55 requires exposure to oxidation agents (28). Similar conclusions can be drawn from Fig. 5*B*, which shows the dynamics of conformational changes in the $\mathbf{N}^* \rightarrow \mathbf{N}_{SH}^{SH}$ transition. Note that in this case BPTI unfolds to a much greater degree (compare the scales in Fig. 5 *A* and *B*) than in the transition from $\mathbf{N}' \rightarrow \mathbf{N}_{SH}^{SH}$. The snapshots of the structures show that the key secondary structural elements are intact, which means that the transiently populated nonnative species are compact (Fig. S7).

Stability of β-Hairpin Correlates with $\tau_{[14-38]}$. Our simulations reveal that β-hairpin is fully formed long before the distance between residues 14 and 38 is close enough to establish S–S bond (Fig. 4*B*). It stands to reason that if interactions favoring β-hairpin formation

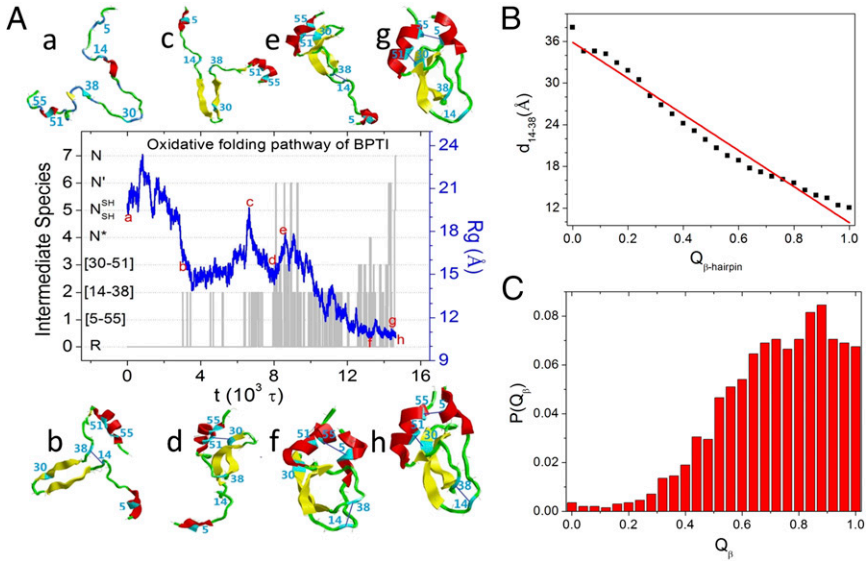

Fig. 4. (*A*) An example of one of 2,000 oxidative folding trajectories showing the route to the folded state from a fully reduced starting conformation with low *Q*. The blue curve shows the decrease in the radius of gyration, R_g (scale is on the right). The gray lines show formation of various disulfide species labeled on the left. Snapshots (*a*–*h*) show some of the conformations sampled in the trajectory. (*B*) Plot of the distance between the distance between residues 14 and 38, d_{14-38}, as a function of the fraction, Q_β, of contacts in the β-hairpin shows that the hairpin (formed between β-strands from Ile18 to Asn24, and from Leu29 to Tyr35) is fully structured before [14–38] formation. (*C*) Distribution of Q_β when d_{14-38} = 5.7 Å (the distance at which [S–S] bond forms) for the first time in a folding trajectory.

are strengthened, then $\tau_{[14-38]}$ should decrease. To test this prediction, we stabilized the β-hairpin by increasing λ (Eqs. **S1** and **S2**), involving only the residues in the β-hairpin. We find that $\tau_{[14-38]}$ decreases as λ increases (Fig. 6), thus demonstrating the correlation between $\tau_{[14-38]}$ and the β-hairpin stability. The yield of [14–38] (red line in Fig. 6) also increases as the hairpin stability increases. We repeated the simulations by creating a pseudo mutant in which only [14–38] can form, which is realized by setting $K_s = 0$ (Eq. **1**) for all other Cys residues. The changes in the rate of [14–38] folding essentially mirrors that for the WT. These simulations predict a twofold increase in rate if the hairpin is stabilized.

The finding that an increase in the stability of β-hairpin (Fig. 1*A*) results in a decrease in the folding time of [14–38] is in quantitative accord with experiments (29). Mutation of a number of bulky hydrophobic residues in the β-strands (residues 18–24 and 29–35), several of them being distant from 14 or 38, by Ala in a pseudo mutant (residues 5, 30, 51, and 55 were replace by Ala) showed retardation of [14–38] formation (figure 4 in ref. 29). The maximum decrease in the rate is about a factor of 2, with most substitutions showing a more modest decrease relative to the pseudo WT. Remarkably, our simulations capture the extent of rate changes found in experiments quantitatively, thus establishing the crucial role the β-hairpin plays in enabling [14–38] to form. More recently, experiments have further illustrated the link between hydrophobic interactions and disulfide bond formation in controlling fibrillar and globular aggregate formation in egg white lysozyme (30, 31).

Factors Determining Protein Collapse and [14–38] Formation. To understand the importance of the initial formation of [14–38], which subsequently leads to native-like states **N*** and **N′**, neither of which contains [14–38], we performed a series of simulations. Consider a variant of BPTI in which S–S bond formation between Cys residues 30, 51, and 55 and interactions between nonbonded residues are all repulsive (obtained by setting $\varepsilon_1 = 0$ in Eq. **S2**). This model, referred to as type I, is a polymer mutant of BPTI serves as a reference system for assessing the role of entropy in forming the single disulfide species. In this limit, we expect that the relative probability of forming the single disulfide intermediates should be proportional to the probability (see figure 1 in ref. 5) of loop formation, $P(l) \approx {}^{1}/_{\theta_3}\,(1 - exp(-l/l_0))$, where l is the number of bonds separating two residues, l_0 (roughly, 2-3) is the persistence length of the polypeptide chain, and $\theta_3 \approx 2.2$ (32). The theory predicts that the ratio of the probability formation of [5–14] ($l = 9$) to [14–38] ($l = 24$) should be roughly $(24/9)^{\theta_3} \approx 9$. Based on the theory for loop formation kinetics (33–35), we predict a similar ratio for time scales for forming such contacts.

The type II model is the same as type I except that only the interactions between residues in the β-hairpin are purely repulsive. All other nonbonded interactions retain the values as in the WT. The type III model is a pseudo mutant of the WT created by preventing the S–S bond formation between Cys residues 30, 51, and 55. We achieve this by setting K_s to zero in Eq. **1** for these three residues. Such mutants can be created in experiments by Ala substitutions (29). In all these simulations, $\beta_O = 1.0$ and $\beta_R = 1.5$.

From 3,000 trajectories for the type I model, the mean time, $\tau_{[5-14]}$, for forming the most probable [5–14] is $\approx 6{,}000\ \tau$, whereas [14–38] forms in $\tau_{[14-38]} \approx 8.7 \times 10^4 \tau$ (Fig. S8*A*). The ratio is ~15, which agrees well with predictions based on polymer theory. The

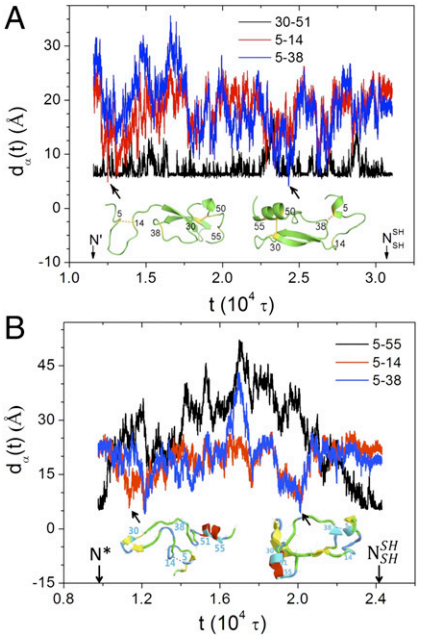

Fig. 5. (*A*) Illustration of the dynamics of rearrangement from N′ to N^{SH}_{SH} using distance between the different Cys residues. The colors are illustrated in the figure. In this transition, BPTI samples both the nonnative species [30–51, 5–14] (structure on the left) and [30–51, 5–38] (conformation on the right). (*B*) Same as *A* except this trajectory describes the **N*** → $\mathbf{N}^{SH}_{SH}$ transition.

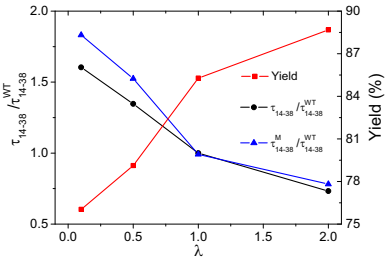

Fig. 6. Time for forming [14–38] relative to the wild-type as function of λ (defined in Eqs. S1 and S2). The stability of the β-hairpin increases (decreases) as λ increases (decreases). The relative times are given for two variants. One of them (black line) is for the WT (λ = 1.0) and the other blue is for a pseudo mutant, in which disulfide bonds other than [14–38] cannot form. The red curves show the yield of [14–38] for the WT.

least probable [5–38] forms about 26 times slower than [5–14], which is also in rough accord with theory.

We generated 4,218 trajectories for the type II model in which the formation of the central β-hairpin is prevented. In this case, the three single disulfide bonds form in comparable times (Fig. S8*B*) with $\tau_{[5-14]} \approx 2.5 \times 10^4\tau$, $\tau_{[14-38]} \approx 1.9 \times 10^4\tau$, and $\tau_{[5-38]} \approx 4.8 \times 10^4\tau$, suggesting that the presence of helices (Fig. 1*A*) and favorable tertiary interactions results in BPTI collapse enabling the three one-disulfide bonds to form albeit without discrimination.

For the type III pseudo WT model (only disulfide bonds involving 5, 14, and 38 can form), we created 2,249 trajectories. In all these trajectories, [14–38] forms, whereas in only a minority of trajectories the formation of [5–14] and [5–38] is observed; this is surprising because the mean folding time of the WT is less than the maximum time for which the simulations are run for this model. Thus, the calculated $\tau_{[5-14]} \approx 0.8 \times 10^4\tau$ is a lower bound. In sharp contrast, we find that [14–38] forms with unit probability in all of the trajectories with $\tau_{[14-38]} \approx 5{,}400\ \tau$ (Fig. S8*C*). With the caveat that the estimates for $\tau_{[5-14]}$ and $\tau_{[5-38]}$ are lower bounds, our simulations show that [14–38] forms considerably faster (with substantial yield) than the nonnative one disulfide intermediates (Fig. S8*C*).

The results for the WT and the three mutant simulations show that the rapid formation requires not only collapse of the protein but complete formation of the central β-hairpin and the helices. The very slow kinetics associated with [5–14] formation in the type III model also rules out any significant role such nonnative species play in the folding of BPTI. In addition, the present work shows in no uncertain terms the inevitability of protein collapse and [14–38] formation in guiding the folding of BPTI. We should emphasize that despite the important kinetic role that [14–38] plays in the wild-type BPTI folding, mutants (Cys to Ser) lacking these two cysteines can also reach the native state with altered kinetics (36). We performed simulations with $E_{SS} = 0$ (Eq. **1**) for only [14–38]. In accord with experiments, we find that this in silico mutant does reach the native by different pathways (compared with the WT) through a simpler network of states (Fig. S9).

Effect of Nonnative Disulfide Bond Formation. To investigate how nonnative (NN) disulfide bonds affect the folding pathways, we performed simulations allowing for [S–S] bond formation between all of the Cys residues. In these simulations the criteria for rupture and formation of [S–S] bonds are identical regardless of whether native or nonnative disulfide bonds are involved. We obtained the flux diagram from 2,000 folding trajectories, and the results are shown in Fig. 2; additional consequences for the initial steps in the folding are shown in Fig. S10. There are two important lessons that come from these simulations. (*i*) The differences between these simulations and the ones based on the Go model arises only in the early stages of BPTI folding. In simulations with NN disulfide bond formation ∼43% of molecules form a mixture of various single NN disulfide species, such as [5–14] and [5–38]. Nearly 50% reach the native [14–38] state slightly ahead of all other NN single disulfide species (Fig. 2). All of the single disulfide species (native and nonnative) are present only after substantial collapse of BPTI (Fig. S10). These findings are in excellent agreement with theoretical predictions (5). Interestingly, the percentages of molecules that form various single disulfide species (Fig. 2) are in close agreement with experiments (26), which is remarkable given the simplicity of the model. (*ii*) There is virtually no difference between results obtained using the Go model and the one with NN disulfide bond formation in the flux through two disulfide species (Fig. 2). Thus, the major conclusions reached based on the Go model simulations remain valid when NN disulfide bond formation is allowed.

Discussion

Nonnative Interactions. Our earlier works (5) showed that on times exceeding τ_c, the population of nonnative species decrease, whereas those of native intermediates increase. The present simulations (using models with and without nonnative disulfide bonds) show that, before the first native disulfide intermediate ([14–38]) forms, BPTI adopts compact conformation, which allows us to focus on the network of connected states involving native-like native intermediates. Because the initial collapse is nonspecific, it follows that nonnative intermediates are likely formed stochastically (5, 12) (*SI Text*), limited only by topological restriction due to chain connectivity and stability, and do not direct folding. Beyond the collapse stage the dominant native interactions facilitate the formation of the folded state. In the process of rearrangement from $\mathbf{N}^*$ and $\mathbf{N}'$, compact intermediates containing nonnative species (Fig. 5) (24) are transiently populated. Lattice model simulations show that one or two nonnative contacts are likely to be part of the TSE even in the folding of proteins without disulfide bonds (37, 38). Recent analyses of atomically detailed simulations and theoretical arguments also suggest that in the folding of small proteins only native-like interactions dominate (39, 40).

Predictions. We have made testable predictions. (*i*) If the stability of the β-hairpin is compromised by suitable mutations, then the formation of the crucial kinetic intermediate [14–38] is impeded. As a result nonnative intermediates (in particular, [5–14]) compete with the formation of [14–38], thus derailing efficient folding. (*ii*) For the wild-type BPTI, we predict that the yields and the relative rates of native single disulfide intermediates do not depend on the redox conditions (concentrations of GSSH and GSH, e.g.). However, the flux through other states can be dramatically altered, as shown in Fig. 2. (*iii*) If the central β-hairpin is destabilized, then the nonnative intermediates [5–14] and [5–38] and the native [14–38] form equally efficiently even though the probability of forming [5–14] is greater than [14–38]. Only upon destabilizing the two peripheral helices and the β-hairpin, the population of [5–14] greatly exceeds that of [14–38], as expected from polymer theory (5, 33).

Concluding Remarks

To provide a theoretical description of how protein folding drives disulfide bond formation, we developed a novel method to mimic the chemistry of disulfide bond formation and rupture in simulations using coarse-grained models. The simulations reproduce the experimentally inferred parallel pathways, involving flux through a network of connected native-like states. Our work explains all of the key features inferred from experiments. (*i*) The global folding pathways found in simulations are in near quantitative agreement with experiments. The flux through the network of native intermediates, including the rate-limiting step involving the formation of $\mathbf{N}_{SH}^{SH}$, is in excellent agreement with experiments. (*ii*) In the process of conversion of native-like N′ and $\mathbf{N}^*$ to $\mathbf{N}_{SH}^{SH}$, compact nonnative species [30–51, 5–38] and [30–51, 5–14] are populated (24). (*iii*) The decrease in the rate of [14–38] formation, as the antiparallel β-sheet is destabilized (Fig. 6), supports mutation experiments reporting similar retardation in the rates (29). We also predict that the rate should increase upon stabilizing the β-hairpin. (*iv*) Inclusion of nonnative disulfide bonds improves quantitative agreement in the flux through native single disulfide species. However, qualitatively, all of the features are captured using a native-centric model.

The finding that polypeptide collapse should occur before disulfide bond formation is supported by single molecule pulling experiments (7), investigating the role protein disulfide isomerase (PDI) plays in catalyzing oxidative folding (41) of a broad class of disulfide proteins. Although derived in the context of disulfide-containing proteins, the link between collapse and folding is also applicable to the folding of globular proteins (42, 43).

Methods

Coarse-Grained Model. In our CG Go-like model (44), each residue is represented as a single reaction center located at the C_α position. We use a harmonic potential for the covalent disulfide bonds (14, 45, 46) given by

$$E_{SS} = K_s(r_\alpha - r_{\alpha 0})^2, \quad [1]$$

where $\alpha = 1,2,3$ labels the three disulfide bonds, r_α is the distance in the α^{th} [S–S] bond, and $r_{\alpha 0}$ is the equilibrium distance in the folded structure. The values of $r_{\alpha 0}$ ($\alpha = 1,2,3$) are 5.61, 5.67, and 6.26 (Å) for 5–55, 14–38, and 30–51, respectively, and $K_s = 20\ \varepsilon\text{Å}^{-2}$, where ε is the unit of energy. The energy function and the details of formation and rupture of [S–S] bonds are described in *SI Text*.

ACKNOWLEDGMENTS. We acknowledge Shaon Chakrabarty, Mauro Mugnai, and Pavel Zhuravlev for pertinent comments on the manuscript. This work was supported by National Science Foundation Grant CHE 13-61946 (to D.T.), National Natural Science Foundation of China Grants 11374148 and 11334004, and the 973 Program 2013CB834100.

1. Anfinsen CB (1973) Principles that govern the folding of protein chains. *Science* 181(4096):223–230.
2. Anfinsen CB, Scheraga HA (1975) Experimental and theoretical aspects of protein folding. *Adv Protein Chem* 29:205–300.
3. Schröder M, Kaufman RJ (2005) The mammalian unfolded protein response. *Annu Rev Biochem* 74:739–789.
4. Welker E, Wedemeyer WJ, Narayan M, Scheraga HA (2001) Coupling of conformational folding and disulfide-bond reactions in oxidative folding of proteins. *Biochemistry* 40(31):9059–9064.
5. Camacho CJ, Thirumalai D (1995) Theoretical predictions of folding pathways by using the proximity rule, with applications to bovine pancreatic trypsin inhibitor. *Proc Natl Acad Sci USA* 92(5):1277–1281.
6. Wedemeyer WJ, Welker E, Narayan M, Scheraga HA (2000) Disulfide bonds and protein folding. *Biochemistry* 39(15):4207–4216.
7. Kosuri P, et al. (2012) Protein folding drives disulfide formation. *Cell* 151(4):794–806.
8. Fass D (2012) Disulfide bonding in protein biophysics. *Annu Rev Biophys* 41:63–79.
9. Creighton TE (1974) The single-disulphide intermediates in the refolding of reduced pancreatic trypsin inhibitor. *J Mol Biol* 87(3):603–624.
10. Chang JY (2011) Diverse pathways of oxidative folding of disulfide proteins: Underlying causes and folding models. *Biochemistry* 50(17):3414–3431.
11. Thirumalai D, Klimov DK, Dima RI (2002) Insights into specific problems in protein folding using simple concepts. *Computational Methods for Protein Folding*, Advances in Chemical Physics, ed Friesner RA (Wiley, New York), Vol 120, pp 35–76.
12. Camacho CJ, Thirumalai D (1995) Modeling the role of disulfide bonds in protein folding: Entropic barriers and pathways. *Proteins* 22(1):27–40.
13. Abkevich VI, Shakhnovich EI (2000) What can disulfide bonds tell us about protein energetics, function and folding: Simulations and bioninformatics analysis. *J Mol Biol* 300(4):975–985.
14. Czaplewski C, Oldziej S, Liwo A, Scheraga HA (2004) Prediction of the structures of proteins with the UNRES force field, including dynamic formation and breaking of disulfide bonds. *Protein Eng Des Sel* 17(1):29–36.
15. Tozzini V (2010) Minimalist models for proteins: A comparative analysis. *Q Rev Biophys* 43(3):333–371.
16. Hyeon C, Thirumalai D (2011) Capturing the essence of folding and functions of biomolecules using coarse-grained models. *Nat Commun* 2:487.
17. Whitford PC, Sanbonmatsu KY, Onuchic JN (2012) Biomolecular dynamics: Order-disorder transitions and energy landscapes. *Rep Prog Phys* 75(7):076601.
18. Noid WG (2013) Perspective: Coarse-grained models for biomolecular systems. *J Chem Phys* 139(9):090901.
19. Creighton TE (1977) Conformational restrictions on the pathway of folding and unfolding of the pancreatic trypsin inhibitor. *J Mol Biol* 113(2):275–293.
20. Creighton TE, Goldenberg DP (1984) Kinetic role of a meta-stable native-like two-disulphide species in the folding transition of bovine pancreatic trypsin inhibitor. *J Mol Biol* 179(3):497–526.
21. Darby NJ, Morin PE, Talbo G, Creighton TE (1995) Refolding of bovine pancreatic trypsin inhibitor via non-native disulphide intermediates. *J Mol Biol* 249(2):463–477.
22. Creighton TE (1992) The disulfide folding pathway of BPTI. *Science* 256(5053):111–114.
23. Goldenberg DP (1992) Native and non-native intermediates in the BPTI folding pathway. *Trends Biochem Sci* 17(7):257–261.
24. Weissman JS, Kim PS (1991) Reexamination of the folding of BPTI: Predominance of native intermediates. *Science* 253(5026):1386–1393.
25. Dadlez M, Kim PS (1995) A third native one-disulphide intermediate in the folding of bovine pancreatic trypsin inhibitor. *Nat Struct Biol* 2(8):674–679.
26. Bulaj G, Goldenberg DP (1999) Early events in the disulfide-coupled folding of BPTI. *Protein Sci* 8(9):1825–1842.
27. Kibria FM, Lees WJ (2008) Balancing conformational and oxidative kinetic traps during the folding of bovine pancreatic trypsin inhibitor (BPTI) with glutathione and glutathione disulfide. *J Am Chem Soc* 130(3):796–797.
28. Weissman JS, Kim PS (1995) A kinetic explanation for the rearrangement pathway of BPTI folding. *Nat Struct Biol* 2(12):1123–1130.
29. Dadlez M (1997) Hydrophobic interactions accelerate early stages of the folding of BPTI. *Biochemistry* 36(10):2788–2797.
30. Xie J, Qin M, Cao Y, Wang W (2011) Mechanistic insight of photo-induced aggregation of chicken egg white lysozyme: The interplay between hydrophobic interactions and formation of intermolecular disulfide bonds. *Proteins* 79(8):2505–2516.
31. Xie J-B, et al. (2012) Photoinduced fibrils formation of chicken egg white lysozyme under native conditions. *Proteins* 80(11):2501–2513.
32. des Cloizeaux J (1979) Short range correlation between elements of a long polymer in a good solvent. *J Phys (Paris)* 41:223–238.
33. Thirumalai D (1999) Time scales for the formation of the most probable tertiary contacts in proteins with applications to cytochrome *c*. *J Phys Chem B* 103:608–610.
34. Chang IJ, Lee JC, Winkler JR, Gray HB (2003) The protein-folding speed limit: Intra-chain diffusion times set by electron-transfer rates in denatured Ru(NH3)5(His-33)-Zn-cytochrome c. *Proc Natl Acad Sci USA* 100(7):3838–3840.
35. Hinczewski M, Tehver R, Thirumalai D (2013) Design principles governing the motility of myosin V. *Proc Natl Acad Sci USA* 110(43):E4059–E4068.
36. Goldenberg DP (1988) Kinetic analysis of the folding and unfolding of a mutant form of bovine pancreatic trypsin inhibitor lacking the cysteine-14 and -38 thiols. *Biochemistry* 27(7):2481–2489.
37. Klimov DK, Thirumalai D (2001) Multiple protein folding nuclei and the transition state ensemble in two-state proteins. *Proteins* 43(4):465–475.
38. Li L, Mirny LA, Shakhnovich EI (2000) Kinetics, thermodynamics and evolution of non-native interactions in a protein folding nucleus. *Nat Struct Biol* 7(4):336–342.
39. Best RB, Hummer G, Eaton WA (2013) Native contacts determine protein folding mechanisms in atomistic simulations. *Proc Natl Acad Sci USA* 110(44):17874–17879.
40. Wolynes PG, Onuchic JN, Thirumalai D (1995) Navigating the folding routes. *Science* 267(5204):1619–1620.
41. Weissman JS, Kim PS (1993) Efficient catalysis of disulphide bond rearrangements by protein disulphide isomerase. *Nature* 365(6442):185–188.
42. Camacho CJ, Thirumalai D (1993) Minimum energy compact structures of random sequences of heteropolymers. *Phys Rev Lett* 71(15):2505–2508.
43. Hofmann H, et al. (2012) Polymer scaling laws of unfolded and intrinsically disordered proteins quantified with single-molecule spectroscopy. *Proc Natl Acad Sci USA* 109(40):16155–16160.
44. Go N (1983) Theoretical studies of protein folding. *Annu Rev Biophys Bioeng* 12: 183–210.
45. Qin M, Zhang J, Wang W (2006) Effects of disulfide bonds on folding behavior and mechanism of the beta-sheet protein tendamistat. *Biophys J* 90(1):272–286.
46. Chinchio M, Czaplewski C, Liwo A, Oldziej S, Scheraga HA (2007) Dynamic formation and breaking of disulfide bonds in molecular dynamics simulations with the UNRES force field. *J Chem Theory Comput* 3(4):1236–1248.
47. Veitshans T, Klimov D, Thirumalai D (1997) Protein folding kinetics: Timescales, pathways and energy landscapes in terms of sequence-dependent properties. *Fold Des* 2(1):1–22.
48. Hwang C, Sinskey AJ, Lodish HF (1992) Oxidized redox state of glutathione in the endoplasmic reticulum. *Science* 257(5076):1496–1502.
49. van Mierlo CPM, Darby NJ, Neuhaus D, Creighton TE (1991) Two-dimensional 1H nuclear magnetic resonance study of the (5-55) single-disulphide folding intermediate of bovine pancreatic trypsin inhibitor. *J Mol Biol* 222(2):373–390.
50. van Mierlo CPM, Darby NJ, Neuhaus D, Creighton TE (1991) (14-38, 30-51) double-disulphide intermediate in folding of bovine pancreatic trypsin inhibitor: A two-dimensional 1H nuclear magnetic resonance study. *J Mol Biol* 222(2):353–371.