

Denaturant-dependent folding of GFP

Govardhan Reddy^a, Zhenxing Liu^b, and D. Thirumalai^{a,1}

^aBiophysics Program, Institute for Physical Science and Technology and Department of Chemistry and Biochemistry, University of Maryland, College Park, MD 20742; and ^bDepartment of Physics, Beijing Normal University, Beijing 100875, China

Edited by Peter G. Wolynes, Rice University, Houston, TX, and approved June 12, 2012 (received for review March 31, 2012)

We use molecular simulations using a coarse-grained model to map the folding landscape of Green Fluorescent Protein (GFP), which is extensively used as a marker in cell biology and biotechnology. Thermal and Guanidinium chloride (GdmCl) induced unfolding of a variant of GFP, without the chromophore, occurs in an apparent two-state manner. The calculated midpoint of the equilibrium folding in GdmCl, taken into account using the Molecular Transfer Model (MTM), is in excellent agreement with the experiments. The melting temperatures decrease linearly as the concentrations of GdmCl and urea are increased. The structural features of rarely populated equilibrium intermediates, visible only in free energy profiles projected along a few order parameters, are remarkably similar to those identified in a number of ensemble experiments in GFP with the chromophore. The excellent agreement between simulations and experiments show that the equilibrium intermediates are stabilized by the chromophore. Folding kinetics, upon temperature quench, show that GFP first collapses and populates an ensemble of compact structures. Despite the seeming simplicity of the equilibrium folding, flux to the native state flows through multiple channels and can be described by the kinetic partitioning mechanism. Detailed analysis of the folding trajectories show that both equilibrium and several kinetic intermediates, including misfolded structures, are sampled during folding. Interestingly, the intermediates characterized in the simulations coincide with those identified in single molecule pulling experiments. Our predictions, amenable to experimental tests, show that MTM is a practical way to simulate the effect of denaturants on the folding of large proteins.

complex folding landscape of GFP | equilibrium and kinetic pathways | multiple folding routes | self-organized polymer model | multicanonical simulation

Our understanding of the folding mechanisms of small single domain proteins has advanced significantly over the last twenty years thanks to theoretical developments (1–7), simulations (8–13), and advances in experimental methods (14–19). Indeed, it can be justly stated that simulations of structure-based models are remarkably successful in predicting the folding mechanisms in the presence (20–22) and absence of denaturants (23–25). More refined questions, such as the relationship between pathway diversity and the symmetry of the underlying native structures (26), transition path times for crossing free energy barriers (27), the nature of the transition state ensemble (28–31) have also been addressed using theory and experiments. In contrast, much less is known about how large proteins with number, N , of amino acids exceeding ≈ 200 with complex β -sheet fold.

Folding mechanisms of Green Fluorescent Protein (GFP), with predominantly β -sheet native structures stabilized by contacts between residues that are well separated along the sequence, have been extensively investigated (32–40). However, the details of the folding kinetics including the structures in the states that are visited along the folding routes are unexplored. It has been difficult to describe quantitatively the folding of GFP because of the extraordinarily long equilibration times, which has made it non-trivial to obtain consistent values of even thermodynamic quantities using different experimental probes. For example, the number of intermediates and their stabilities, and the

associated denaturant-dependent properties (m -values) are not unambiguously known for many of the variants of GFP. Because of the presence of a visual chromophore, that is shielded from water by the barrel structure (Fig. 1A), GFP has been used as a marker in a number of biotechnology and biophysical applications ranging from protein-protein interactions to gene expression (41). Consequently, it is important to map in detail the folding landscapes of FPs (32), which could aid in the development of new GFP tools for use in *in vivo* localization studies.

The folded structure of GFP has eleven β -strands that are closed to form the characteristic barrel structure (Figs. 1A and 1B). The chromophore, surrounded by the barrel and the capping helices, causes hysteresis (denaturant-induced equilibrium folding and unfolding do not coincide even on time scales that exceed hours) in the folding of a number of variants of GFP (34, 35). Mutants that inhibit chromophore formation fold without hysteresis, and could form the basis of describing the *de novo* folding of the complex barrel structure.

Here, we use simulations of a coarse-grained (CG) Self-Organized Polymer model (42) with side chains (SOP-SC) (21) of GFP, with $N = 230$. Our simulations are performed for a chromophore less variant of GFP, Citrine, which differs from the wild type sequence by point mutations at four locations, S65G/V68L/Q69M/S72A/T203Y (36). Thermal denaturation and unfolding induced by Guanidinium chloride (GdmCl) (simulated using the Molecular Transfer Model (MTM)) (20) show that equilibrium folding can be approximately described by a two-state model. Folding trajectories probing equilibrium fluctuations and free energy profiles point to the presence of intermediates. We show, in agreement with experiments (33, 38), using a variety of probes of the structures (distributions of the radius of gyration, R_g , solvent accessible surface area, and structural overlap function, χ) that in the intermediate, the N-terminus is folded whereas the C-terminus region comprising of strands β_7 to β_{10} are flexible and disordered. Upon initiating folding the time-dependent decrease in $\langle R_g(t) \rangle$, obtained using Brownian dynamics simulations (21, 43), shows that compaction occurs in at least two major stages. In accord with the predictions of the kinetic partitioning mechanism (KPM) (44) Citrine folds by at least four distinct routes. In the dominant pathway, folding commences from preformed local β -sheet structures and occurs in an almost all-or-none manner. Stable local structures associate by diffusion-collision mechanism (45) resulting in the folded state. Besides providing testable predictions, such as the linear decrease in the melting temperature as a function of GdmCl and urea concentrations, our work shows that the complexity of GFP folding can be fairly accurately captured using the SOP-SC model in conjunction with MTM.

Author contributions: G.R. and D.T. designed research; G.R., Z.L., and D.T. performed research; G.R. and D.T. contributed new reagents/analytic tools; G.R. and D.T. analyzed data; and G.R. and D.T. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

¹To whom correspondence should be addressed. E-mail: thirum@umd.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1201808109/-DCSupplemental.

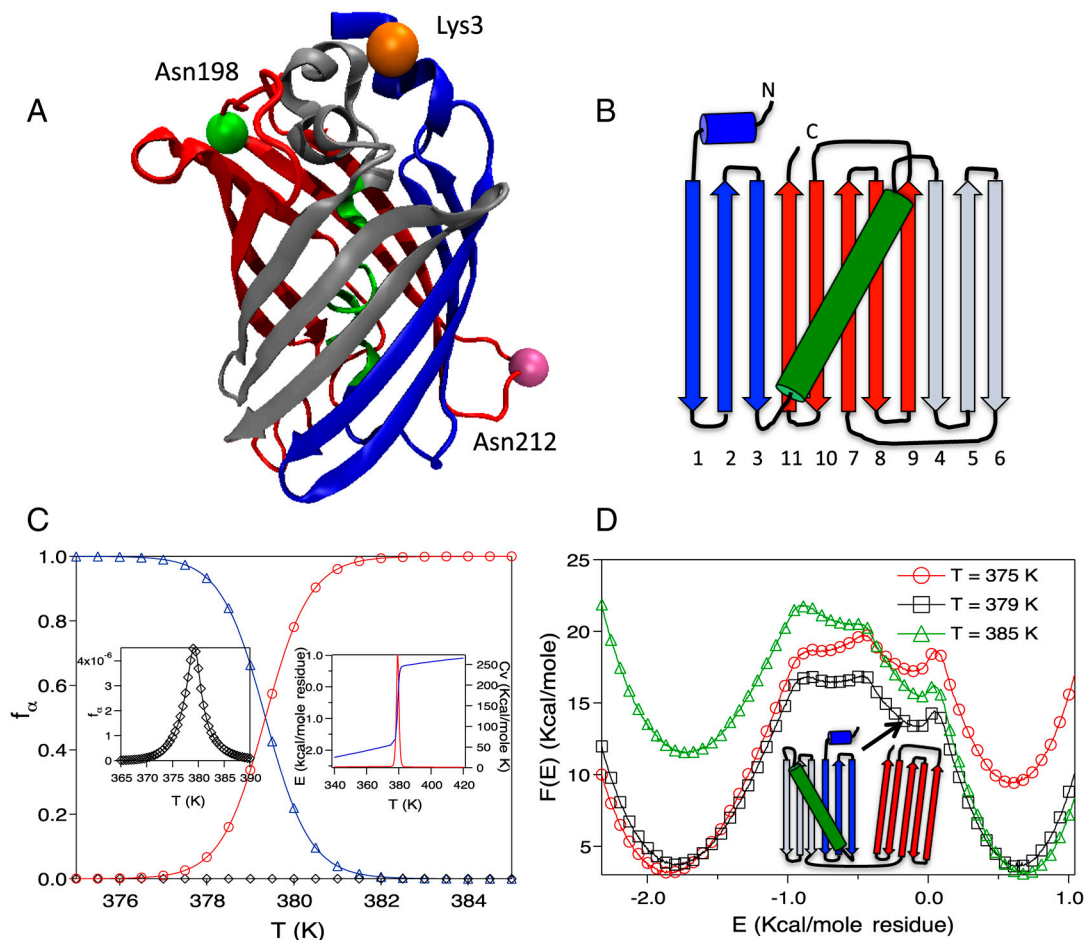


Fig. 1. GFP structure and thermal folding. (A) Ribbon diagram of Citrine (Yellow version of GFP, PDB ID: 1HUY). Residues marked in orange, green and mauve are Lys3, Asn198, and Asn212, respectively. (B) Splay representation of GFP. The N-terminus β -strands are represented in blue, the kinked helix at the center of the β -strand barrel is in green, the three β -strands in the center, which form local contacts are in silver, and the C-terminus β -strands are in red. (C) Fraction of GFP in the NBA, INT, and UBA as a function of temperature, T , are shown in triangles, diamonds, and circles respectively. The left inset shows $f_{INT}(T)$ as a function of T . The right inset shows energy, E , and the heat capacity, C_v as a function of T . (D) Free energy of GFP as a function of E for temperatures above and below the melting temperature ($T = 379$ K). The structure of one of the intermediates is shown in splay representation.

Results

Thermal Denaturation: We used multicanonical (MC) molecular dynamics simulations (46, 47) so that thermodynamic properties of Citrine can be computed accurately (see *SI Text*, and *Movie S1* for details). Dependence of the total energy $\langle E(T) \rangle$ as a function of temperature, T , and the associated heat capacity $C_v(T) (= \frac{\langle E^2 \rangle - \langle E \rangle^2}{k_B T^2})$ both indicate that folding occurs cooperatively (inset in Fig. 1C). The melting temperature, corresponding to the peak in $C_v(T)$ in the absence of denaturants is, $T_m = 379$ K (see right inset in Fig. 1C). The structural overlap function, χ , is used to distinguish between the Native Basin of Attraction (NBA), equilibrium intermediates (I_{EQL}), and UBA, the unfolded basin of attraction (*Methods*). The fraction of molecules in the Native Basin of Attraction, $f_{NBA}(T)$, as a function of T shows that the folding transition is cooperative (Fig. 1C) suggesting that a two-state description is adequate. The value of T_m computed using $P(T_m) = 0.5$ yields $T_m = 379$ K, which coincides with the peak in the heat capacity (Fig. 1C). The T_m value obtained in our simulations are in very good agreement with calorimetry measurements (48, 49), thus validating the SOP-SC model.

The two state nature of thermal unfolding obtained from Fig. 1C hides the presence of equilibrium intermediates, which are evident in the MC folding trajectory (see Fig. S1 in the *SI Text*). The time dependent changes in E and χ in Fig. S1 show that at least two equilibrium states (hereafter referred collectively as I_{EQL}) are populated. The dips in the plots of free energy profiles, $F(E)$ as a function of E , above and below T_m correspond to population of intermediates with χ values in the range specified in *Methods* (Fig. 1D). The inset in Fig. 1C shows that $f_{INT}(T)$ is very small, which explains the apparent two-state folding of Citrine in

the absence of the chromophore (see also (34)). The finding that I_{EQL} are populated during the folding process (see also Fig. 2) provides indirect support to the analysis used in experimental studies, which found that a three-state description provides a quantitative fit to denaturant unfolding of Citrine in the presence of chromophore. Thus, the intermediate is, in all likelihood, stabilized by the chromophore.

Denaturant Induced Unfolding: To make a direct comparison with ensemble experiments (33, 34) we simulated the effects of Guanidinium chloride (GdmCl) using the MTM (see *Methods* and *SI Text* for details). Following our previous work (21), we choose a simulation temperature, T_s , at which the free energy differences between the NBA and the unfolded state, $\Delta G_{NU} (= G_N(T_s) - G_U(T_s))$ calculated from simulations agree with the measured value of $\Delta G_{NU}^E = -16.5$ kcal/mole for R96A mutant (34) at denaturant concentration, $[C] = 0$. We chose R96A as a reference GFP for fixing T_s because it exhibits a clear equilibrium two-state transition (34). Since the absolute values of the effective interaction energies in CG models cannot be determined, we used T_s as an adjustable parameter to get accurate estimate of ΔG_{NU} . The choice of T_s , which is the only parameter that is fixed in our simulations, amounts to choosing the overall free energy scale (21). The dependence of $f_{NBA}[C]$ and $f_{UBA}[C]$ on $[C]$ at $T_s = 368.2$ K also shows that Citrine folds and unfolds reversibly in an apparent two state manner (Fig. 2A). However, the right inset in Fig. 2A shows signs for the presence of an intermediate.

The midpoint of the folding transition $f([C_m]) = 0.5$ yields $C_m = 1.3$ M. The simulated equilibrium titration curves can be fit using a two state model from which the dependence of $\Delta G_{NU}([C])$ on $[C]$ can be calculated using $\Delta G_{NU}([C]) = -RT_s \ln(\frac{f_{NBA}([C])}{1-f_{NBA}([C])})$. The linear fit, $\Delta G_{NU}([C]) = G_{NU}^A([0]) +$

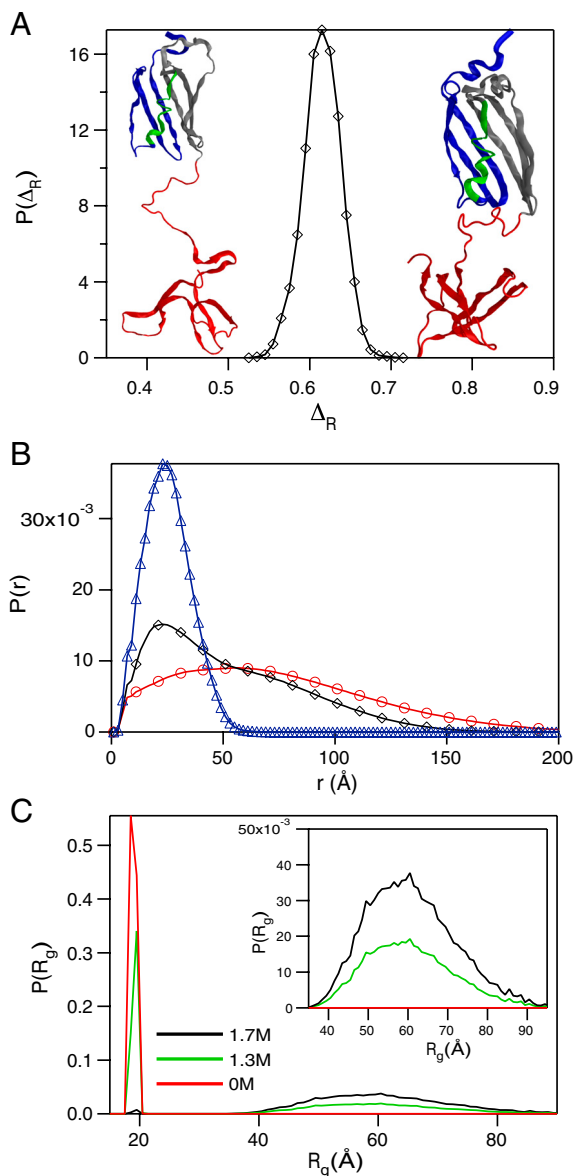


Fig. 3. (A) Distribution of $P(\Delta_R)$ where Δ_R is the relative accessible surface area (see text for details) in one of the kinetic intermediates with attributes similar to that found under equilibrium. The typical structure sampled at equilibrium is on the left whereas the structure on the right represents a conformation obtained in the kinetic simulation in the E_{QL} pathway. (B) Distribution, $P(r)$, as a function of r , the distance between a pair of sites for NBA (triangles), INT (diamonds), and UBA (circles) computed using conformations sampled at equilibrium with $[C] = 0$. (C) Distribution of R_g as a function of GdmCl concentration. The midpoint of the transition is $C_m = 1.3$ M. The inset shows $P(R_g)$ at large R_g .

Collapse and Folding Kinetics: We used Brownian dynamics simulations to generate 80 folding trajectories starting from an initial ensemble of equilibrated structures at a high temperature ($T_H > 420$ K) and reducing the temperature to $T = 300$ K to initiate folding. From the distribution of first passage times, $P_{FP}(s)$, we calculated the probability, $P_U(t) = \int_0^t P_{FP}(s) ds$, that Citrine has not folded at time, t (inset in Fig. 4A). The folding time obtained from the exponential decay of $P_U(t) \approx \exp(-t/\tau_F)$ is $\tau_F \approx 5$ ms. Although there is a wide range for τ_F reported in experiments the smallest value for τ_F for Citrine from recent experiments is ≈ 1 s (36). Thus, our estimate of τ_F is at least a hundred fold less than experimental values, which could arise

from neglect of solvent and the coarse-grained representation of the proteins.

The time-dependent changes in $\langle R_g(t) \rangle$ during the folding process shows that collapse occurs in two stages (Fig. 4A). The decay of $\langle R_g(t) \rangle$ can be fit using $\langle R_g(t) \rangle = a_1 \exp(-t/\tau_1) + a_2 \exp(-t/\tau_2)$, where $a_1 = (23.4 \pm 0.2)$ Å, $a_2 = (22.5 \pm 0.2)$ Å, $\tau_1 = 3.6$ ms and $\tau_2 = 43.47$ ms (Fig. 4A). The decrease in the value of $R_g(t)$ in each stage is roughly the same (≈ 22 Å). However, there is a separation in the time scales in the two stages implying that distinctly different ensembles of structures are sampled in two stages. The structures reached at the end of the first stage correspond to minimum energy compact structures (50) that guide the folding of Citrine. The predicted multistage compaction of GFP, similar to that found in other proteins (51), can be tested using time-resolved SAXS experiments.

Parallel Pathways and Kinetic Intermediates: By analyzing the 80 folding trajectories using E (see *Methods*) as the progress variable for the folding reaction (Fig. S2) we find that Citrine folds along four pathways. For reasons explained below, we classify the structures populated in three of the pathways as kinetic intermediates and the intermediate in the fourth has most of the hallmarks of I_{EQL} . Representative trajectories one from each pathway, displaying energy as a function of t , show that in each pathway folding occurs in stages (Fig. 4B). The nature of structures and their lifetimes vary greatly depending on the pathway. In the dominant pathway KIN1, through which $\approx 50\%$ of the flux to the native state is channeled, Citrine folds in nearly a two state manner (Fig. 4B). In this pathway, β -strands at the N-terminus, the core of the protein, and the C-terminal strands form rapidly in a single step. The folding units in the long lived meta-stable states collide and consolidate (see *Movie S1* in the *SI Text*) as envisioned in the diffusion-collision model (45).

In the second kinetic pathway, KIN2, representing $\approx 16\%$ of the trajectories, folding occurs through an intermediate in which tertiary contacts between strands β_3 and β_{11} and the interface between β_1 and β_6 are not formed (Fig. 4C). The structural properties of the intermediate in KIN2, such as the distributions of R_g and χ , differ from I_{EQL} , and hence we conclude that it is observed only during the process of folding. Similarly, in the third kinetic pathway, KIN3, through which about 10% of the flux to the native state flows, the intermediates shown in Fig. 4D are observed. The intermediate shown in Fig. 4D, has a kinked helix interacting with the N-terminal β -sheets. The splay diagram shows further that long-range interactions involving β_3 and β_{11} as well as β_1 and β_6 are disrupted.

Equilibrium Folding Intermediate: In about 15% of the trajectories an intermediate whose structural characteristics almost coincide with I_{EQL} is populated. The folding trajectory (Fig. 4B) and the $[\chi, R_g]$ plot in Fig. 4D show that the intermediate has a long lifetime. There are four lines of evidence which show that the kinetic and equilibrium intermediates are structurally similar. First, the $\langle R_g \rangle$ obtained using the conformations sampled with folded N-terminal strands (the C-terminal strands are fluid-like) at equilibrium is ~ 14 Å whereas those obtained from kinetic trajectories is ~ 13 Å. Second, the typical conformations sampled at equilibrium and during the folding process are similar (compare the structures in Fig. 3A). In both the structures the strands β_1 through β_6 are formed whereas the C-terminal is more flexible. Third, the values of χ for the conformations corresponding to the plateau in E as a function of t (purple line in Fig. 4B) is in the range $0.7 < \chi \leq 0.895$, which coincides with the estimates obtained using equilibrium trajectories (see also the shaded green region in Fig. 2B). Fourth, we calculated the distribution, $P(\Delta_R)$, of $\Delta_R = \frac{(A_U - A_I)}{(A_U - A_N)}$, where A_U , A_I and A_N are the solvent accessible surface area (SASA) for the unfolded, I_{EQL} , and the native states

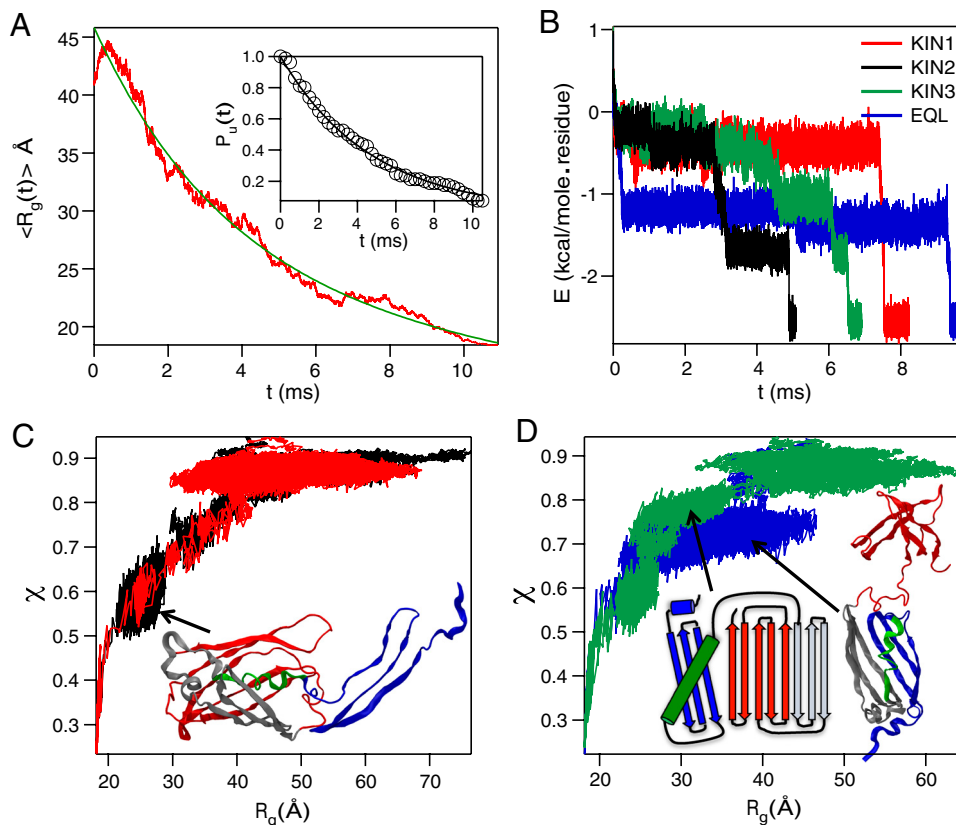


Fig. 4. GFP folding kinetics. (A) Time (t) dependent decrease in $\langle R_g(t) \rangle$. The solid line in green is a two exponential fit to the data. The inset shows the fraction of trajectories unfolded, $P_U(t)$ as function of t , with the solid black line being an exponential fit. (B) Total energy E as a function of t for folding along four pathways. A representative trajectory for each pathway is shown. (C) Dynamic profiles generated using the folding trajectories in terms of χ and R_g for *KIN1* and *KIN2* in red and black respectively. The ribbon diagram of the kinetic intermediate in *KIN2* shows that $\beta_1 - \beta_3$ sheets are not packed onto the rest of the GFP barrel structure. (D) Plot of χ as function of R_g for folding pathways *KIN3* and *EQL* are in green and blue, respectively. Interactions involving the helix and $\beta_1 - \beta_3$ in the first intermediate sampled in *KIN3* with the rest of the barrel are absent. The second intermediate in *KIN3* is similar to the one observed for *KIN2*. In the intermediate populated in the *EQL* pathway, the C-terminal β -sheets do not interact with the ordered N-terminal strands $\beta_1 - \beta_6$.

respectively using the conformations sampled during the kinetic simulations (plateau region in the purple curve in Fig. 4B). From $P(\Delta_R)$ in Fig. 3A we find that the mean values $\langle \Delta_R \rangle \approx 0.62$, which compares well with the experimental value $\langle \Delta_R \rangle = 0.59$ estimated for the equilibrium intermediate in the Citrine folding (36). Thus, we surmise that the structures sampled along the *EQL* pathway (Fig. 4B) are similar to those observed under equilibrium conditions.

Relation Between Collapse, Secondary Structure Formation and Folding: It has been shown that for efficient folding, collapse and folding occur nearly simultaneously (52), where as for larger complex proteins collapse precedes by folding. In Fig. 4C and 4D we show the conformations sampled in all the pathways. Each conformation is represented by χ and R_g . We conclude from the results in Figs. 4C and 4D that in all the four pathways Citrine undergoes compaction (reduction in R_g) before folding (decrease in χ). These findings show that the search for the NBA occurs among the ensemble of minimum energy compact structures (50). To explore the link between secondary structure formation and collapse we plot the fraction, f_{ss} , of secondary structure acquired as the polypeptide chain collapses. Fig. S3 shows that in all the pathways the value of f_{ss} is relatively small even after substantial collapse, which also implies that collapse generates a fluid-like globule. Consolidation of structure occurs only after reduction in chain dimension.

Topological Traps: In 6 out of the 80 folding trajectories topological traps (44) give rise to Citrine misfolding. In five of the trajectories, Citrine is kinetically trapped in a native-like structure in which the loop connecting β_9 and β_{10} (Fig. 5B) is in an incorrect position, thus disrupting the interactions between β -strands 4 and 9. In the other misfolded structure (35), the helix is outside the barrel and hence does not pack in the center as observed in the native structure.

Discussion

Intermediates in GFP Folding: Compelling evidence for postulating the presence of equilibrium intermediates is the non-coincidence of free energies extracted using fluorescence and NMR measurements even though visually the decrease in fluorescence as a function of GdmCl suggests that a two-state description is sufficient (33). More recently, single molecule fluorescence experiments have presented clear evidence for three state equilibrium folding of Citrine with the chromophore (36). The presence of high energy equilibrium intermediates in R96A is only evident in free energy profiles computed as a function of energy (Fig. 1D), radius of gyration (Fig. 2B), and the fraction of native contacts (Fig. S1B). The structural characteristics of I_{EQL} support the finding based on equilibrium H/D exchange experiments showing order in the N-terminus and flexibility in the C-terminal region spanning β_7 to β_{10} (33). It is worth pointing out that one of the intermediates populated in single molecule pulling experiments corresponds to unfolding in the C-terminus domain with β_1 through β_6 remaining intact (53). Thus, both experiments and simulations point to the presence of I_{EQL} . The observation that I_{EQL} in our simulations is a high energy intermediate occurring with negligible probability whereas the ones characterized in experiments are stable suggests that the intermediates in GFP folding are stabilized by the chromophore.

Comparing Simulation and Experimental Folding Pathways: We have found evidence for kinetic intermediates that direct folding to the native state. The range of kinetic intermediates observed in our simulations have previously been identified in single molecule pulling experiments and simulations (39, 40, 42, 53). When GFP is stretched by applying mechanical force, f , between residues 3 and 212 (Fig. 1), GFP unfolds along two routes. In the first one, unfolding is in an all-or-none process where as in the second pathway unfolding occurs by populating an intermediate (53). The increase in the contour length upon unfolding GFP from the folded to the intermediate state is consistent with formation of

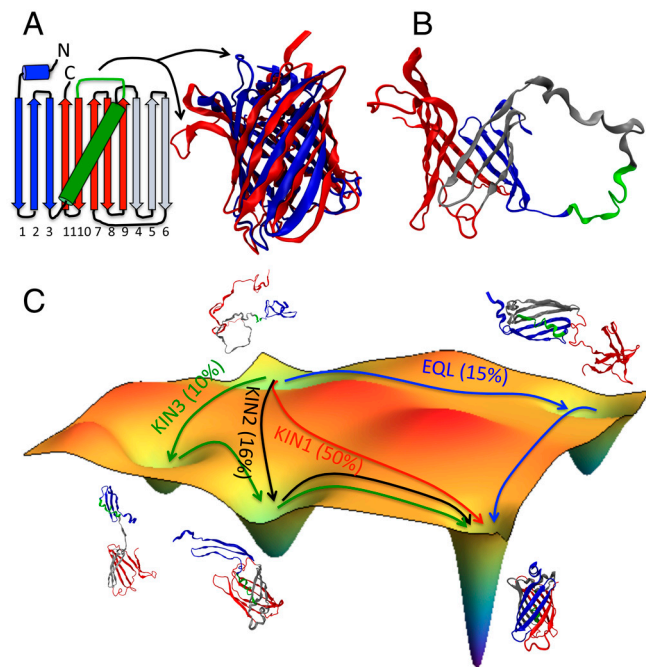


Fig. 5. Topological traps and model for GFP folding. (A) The wild type GFP structure in blue is superimposed onto the misfolded structure in red. Misfolding of the loop connecting β_9 and β_{10} causes a topological trap. The arrows show the correct (folded) and incorrect (misfolded) conformation of the loop. (B) In the second misfolded structure the barrel involving the 11 β -strands form with no contact between the central α -helix and rest of the structure. (C) Folding landscape and network of connected states based on simulations. The flux through the four channels and the routes from the UBA to the NBA are indicated by arrows. Because topological traps (structures in (A) and (B)) are dead ends in the folding process they do not reach the NBA on the time scale of our simulations. Thus, the total flux to the NBA from UBA is less than 100%. Representative structures sampled along the four pathways as well as an unfolded conformation are shown.

a structure in which strands (1–6) are intact and the rest are unfolded. Our simulations and previous ensemble experiments (33, 37, 38) have identified such an intermediate as occurring both in equilibrium and during the folding process.

When mechanical force is applied (53) to the residues 3 and 198, GFP unfolds by populating a different intermediate in which the 3 N-terminus β -strands (β_1 , β_2 , β_3) have ruptured away from the rest of the barrel. In addition, when f is applied to the ends of the molecule the dominant unfolding pathway also occurs through such an intermediate (40, 42). We showed (42) that upon initiating folding by reducing the mechanical force from a high value to $f = 0$, a long-lived metastable intermediate similar to that observed in KIN2 and KIN3 is populated before the barrel structure forms. We find here that in the process of refolding, upon temperature quench, the intermediate with loss of interactions between strands (1–3) and rest of the barrel is populated in pathways KIN2 and KIN3.

Single molecule fluorescence experiments have also shown that Citrine unfolds by parallel pathways. In one of the pathways, the folded state is reached directly, whereas an intermediate with a low FRET is populated in the other (36). It is likely, that these low FRET states are an ensemble of conformations containing a mixture of equilibrium and kinetic intermediates. Taken together it is clear that folding GFP must occur through a number of distinct intermediates, which become visible only by using different

techniques. By combining all the results from simulations and experiments, we propose a model (Fig. 5B) for GFP folding that involves a complex network of connected states through which the flux to the native flows. The resulting multiple pathways include equilibrium and kinetic intermediates as well as misfolded structures. Detailed comparison between our simulations and experiments show that a number of different experimental probes are needed to quantitatively map the folding landscape of GFP, and presumably other proteins with complex topology.

Collapse and Folding: Since it was first demonstrated theoretically that folding and collapse transitions are intimately linked for proteins that reach the native state efficiently (54, 55) there has been considerable interest in validating this concept. For wild type GFP it is unmistakable that folding is preceded by populating compact intermediate species. Both SAXS experiments (38) as well as our simulations show that a compact state, with $\langle R_g \rangle$ intermediate between the values in the folded and unfolded states, is present and is likely to be on-pathway to the NBA. The kinetic folding trajectories clearly show (Fig. 4 C and D) that GFP first collapses before the formation of secondary and tertiary interactions. Our finding is strikingly similar to the order in which structure is acquired in monellin with β -sheet topology (51) as well as in GFP upon force quench (42).

Methods

SOP-Side Chain (SOP-SC) Model: In the SOP-SC model, which uses residue-dependent interaction between SCs (56), each residue is represented by two interaction centers one for the backbone atoms and the other corresponds to the side chain atoms. The SOP-SC model is constructed using the crystal structure of Citrine, an improved yellow version of GFP (57), with the Protein Data Bank PDB ID: 1HUJ. Citrine has 5 mutations compared to the wild type GFP (S65G, V68L, Q69M, S72A, T203Y). The residues labeled 0 and 1A in 1HUJ are deleted. In Citrine the chromophore, labeled residue 66 in 1HUJ, involves residues GLY65, TYR66, and GLY67. In our simulations the chromophore is disabled by eliminating the chemical reaction involving the three chromophore residues. The functional form of SOP-SC model and the parameters of the CG force field are given in Table S1 in the *SI Text*.

Molecular Transfer Model (MTM): The energy of transferring a specific protein conformation from water to the denaturant solution at concentration $[C]$ is written as $\Delta G_{tr}([C]) = \sum_{i=1}^{2N} \delta g_{tr,i}([C]) (\alpha_i / \alpha_{Gly-i-Gly})$ where the summation includes both backbone and side chain atoms, $\delta g_{tr,i}([C])$ is the experimentally measured transfer free energy of i , α_i is the solvent accessible surface area (SASA) of the bead i , $\alpha_{Gly-i-Gly}$ is the solvent accessible surface area of the bead in the tripeptide Gly- i -Gly. The radii of amino acid backbone and sidechain are given in Table S2. The transfer free energies $\delta g_{tr,i}([C])$ for the backbone and side chains, taken from experiments (20, 22, 58), are listed in table S3 in ref. (21). The values for $\alpha_{Gly-i-Gly}$ are listed in table S4 in ref. (21). The thermodynamic properties of a protein at $[C] \neq 0$ are obtained using the procedure described earlier (20, 22).

Data Analysis: We use the structural overlap function (59) $\chi = \frac{2}{(M^2 - 5M + 6)} \sum_{i=1}^{M-3} \sum_{j=i+3}^M \Theta(\delta - |r_{ij} - r_{ij}^0|)$ to distinguish between the various populated states. Here, $M = 2N = 460$ is the number of interaction centers in the SOP-SC representation of GFP, r_{ij} is the distance between the beads i and j with r_{ij}^0 being the corresponding distance in the folded state, Θ is the Heaviside step function, and $\delta = 2 \text{ \AA}$. The conformation with $0 < \chi \leq 0.7$ are folded, an intermediate state has $0.7 < \chi \leq 0.895$, and conformations with $\chi > 0.895$ are unfolded (Fig. S1 in the *SI Text*).

ACKNOWLEDGMENTS. We are pleased to acknowledge useful discussions with Sophie Jackson. This work was supported by a grant from the National Science Foundation through grant CHE 09-14033. ZL acknowledges financial support from the National Natural Science Foundation of China under the grant no. 11104015.

- Thirumalai D, Hyeon C (2005) RNA and protein folding: Common themes and variations. *Biochemistry* 44:4957–4970.
- Onuchic J, LutheySchulten Z, Wolynes P (1997) Theory of protein folding: The energy landscape perspective. *Annu Rev Phys Chem* 48:545–600.

- Shakhnovich E (2006) Protein folding thermodynamics and dynamics: Where physics, chemistry, and biology meet. *Chem Rev* 106:1559–1588.
- Dill KA, Ozkan SB, Shell MS, Weikl TR (2008) The protein folding problem. *Annu Rev Biophys* 37:289–316.

5. Munoz V, Eaton W (1999) A simple model for calculating the kinetics of protein folding from three-dimensional structures. *Proc Natl Acad Sci USA* 96:11311–11316.
6. Kubelka J, Henry ER, Cellmer T, Hofrichter J, Eaton WA (2008) Chemical, physical, and theoretical kinetics of an ultrafast folding protein. *Proc Natl Acad Sci USA* 105:18655–18662.
7. Thirumalai D, O'Brien EP, Morrison G, Hyeon C (2010) Theoretical Perspectives on Protein Folding. *Annu Rev Biophys* 39:159–183.
8. Fersht A, Daggett V (2002) Protein folding and unfolding at atomic resolution. *Cell* 108:573–582.
9. Shaw DE, et al. (2010) Atomic-Level Characterization of the Structural Dynamics of Proteins. *Science* 330:341–346.
10. Voelz VA, Singh VR, Wedemeyer WJ, Lapidus LJ, Pande VS (2010) Unfolded-state dynamics and structure of protein L characterized by simulation and experiment. *J Am Chem Soc* 132:4702–4709.
11. Hyeon C, Thirumalai D (2011) Capturing the essence of folding and functions of biomolecules using coarse-grained models. *Nat Commun* 2:487.
12. Whitford PC, et al. (2009) An all-atom structure-based potential for proteins: Bridging minimal models with all-atom empirical forcefields. *Proteins* 75:430–441.
13. Zhang Z, Chan HS (2010) Competition between native topology and nonnative interactions in simple and complex folding kinetics of natural and designed proteins. *Proc Natl Acad Sci USA* 107:2920–2925.
14. Schuler B, Eaton WA (2008) Protein folding studied by single-molecule FRET. *Curr Opin Struct Biol* 18:16–26.
15. Nickson AA, Clarke J (2010) What lessons can be learned from studying the folding of homologous proteins? *Methods* 52:38–50.
16. Borgia A, Williams PM, Clarke J (2008) Single-molecule studies of protein folding. *Annu Rev Biochem* 77:101–125.
17. Bartlett AL, Radford SE (2009) An expanding arsenal of experimental methods yields an explosion of insights into protein folding mechanisms. *Nat Struct Mol Biol* 16:582–588.
18. Shank EA, Ceconi C, Dill JW, Marqusee S, Bustamante C (2010) The folding cooperativity of a protein is controlled by its chain topology. *Nature* 465:637–U134.
19. Gebhardt JCM, Borschloegla T, Rief M (2010) Full distance-resolved folding energy landscape of one single protein molecule. *Proc Natl Acad Sci USA* 107:2013–2018.
20. O'Brien E, Ziv G, Haran G, Brooks B, Thirumalai D (2008) Effects of denaturants and osmolytes on proteins are accurately predicted by the molecular transfer model. *Proc Natl Acad Sci USA* 105:13403–13408.
21. Liu Z, Reddy G, O'Brien EP, Thirumalai D (2011) Collapse kinetics and chevron plots from simulations of denaturant-dependent folding of globular proteins. *Proc Natl Acad Sci USA* 108:7787–7792.
22. O'Brien E, Brooks B, Thirumalai D (2009) Molecular origin of constant m -values, denatured state collapse, and residue-dependent transition midpoints in globular proteins. *Biochemistry* 48:3743–3754.
23. Oliveberg M, Wolynes PG (2005) The experimental survey of protein-folding energy landscapes. *Q Rev Biophys* 38:245–288.
24. Klimov D, Thirumalai D (2000) Native topology determines force-induced unfolding pathways in globular proteins. *Proc Natl Acad Sci USA* 97:7254–7259.
25. Fernandez-Escamilla A, et al. (2004) Solvation in protein folding analysis: Combination of theoretical and experimental approaches. *Proc Natl Acad Sci USA* 101:2834–2839.
26. Klimov DK, Thirumalai D (2005) Symmetric connectivity of secondary structure elements enhances the diversity of folding pathways. *J Mol Biol* 353:1171–1186.
27. Chung HS, Louis JM, Eaton WA (2009) Experimental determination of upper bound for transition path times in protein folding from single-molecule photon-by-photon trajectories. *Proc Natl Acad Sci USA* 106:11837–11844.
28. Klimov D, Thirumalai D (2002) Stiffness of the distal loop restricts the structural heterogeneity of the transition state ensemble in SH3 domains. *J Mol Biol* 317:721–737.
29. Ding F, Guo W, Dokholyan N, Shakhnovich E, Shea J (2005) Reconstruction of the src-SH3 protein domain transition state ensemble using multiscale molecular dynamics simulations. *J Mol Biol* 350:1035–1050.
30. Onuchic J, Socci N, LutheySchulten Z, Wolynes P (1996) Protein folding funnels: The nature of the transition state ensemble. *Folding Des* 1:441–450.
31. Guo Z, Thirumalai D (1997) The Nucleation-collapse mechanism in protein folding: Evidence for the non-uniqueness of the folding nucleus. *Folding Des* 2:277–341.
32. Hsu STD, Blaser G, Jackson SE (2009) The folding, stability and conformational dynamics of beta-barrel fluorescent proteins. *Chem Soc Rev* 38:2951–2965.
33. Huang J, Craggs T, Christodoulou J, Jackson S (2007) Stable intermediate states and high energy barriers in the unfolding of GFP. *J Mol Biol* 370:356–371.
34. Andrews B, Schoenfish A, Roy M, Waldo G, Jennings P (2007) The rough energy landscape of Superfolder GFP is linked to chromophore. *J Mol Biol* 373:476–490.
35. Andrews B, Gosavi S, Finke J, Onuchic J, Jennings P (2008) The dual-basin landscape in GFP folding. *Proc Natl Acad Sci USA* 105:12283–12288.
36. Orte A, Craggs T, White S, Jackson S, Klenerman D (2008) Evidence of an intermediate and parallel pathways in protein unfolding from single-molecule fluorescence. *J Am Chem Soc* 130:7898–7907.
37. Enoki S, Saeki K, Maki K, Kuwajima K (2004) Acid denaturation and refolding of Green Fluorescent Protein. *Biochemistry* 43:14238–14248.
38. Enoki S, et al. (2006) The equilibrium unfolding intermediate observed at pH 4 and its relationship with the kinetic folding intermediates in green fluorescent protein. *J Mol Biol* 361:969–982.
39. Dietz H, Rief M (2006) Exploring the energy landscape of GFP by single-molecule mechanical experiments. *Proc Natl Acad Sci USA* 101:16192–16197.
40. Mickler M, et al. (2007) Revealing the bifurcation in the unfolding pathways of GFP by using single-molecule experiments and simulations. *Proc Natl Acad Sci USA* 104:20268–20273.
41. Tsien R (1998) The green fluorescent protein. *Annu Rev Biochem* 67:509–544.
42. Hyeon C, Dima R, Thirumalai D (2006) Pathways and kinetic barriers in mechanical unfolding and refolding of RNA and proteins. *Structure* 14:1633–1645.
43. Veitshans T, Klimov D, Thirumalai D (1996) Protein folding kinetics: timescales, pathways and energy landscapes in terms of sequence-dependent properties. *Folding Des* 2:1–22.
44. Guo Z, Thirumalai D (1995) Kinetics of protein-folding—nucleation mechanism, time scales, and pathways. *Biopolymers* 36:83–102.
45. Karplus M, Weaver D (1994) Protein-folding dynamics—the diffusion-collision model and experimental data. *Protein Sci* 3:650–668.
46. Okamoto Y, Hansmann U (1995) Thermodynamics of helix-coil transitions studied by multicanonical algorithms. *J Phys Chem* 99:11276–11287.
47. Hansmann U, Okamoto Y, Eisenmenger F (1996) Molecular dynamics, Langevin and hybrid Monte Carlo simulations in a multicanonical ensemble. *Chem Phys Lett* 259:321–330.
48. Nagy A, Malnasi-Csizmadia A, Somogyi B, Lorinczy D (2004) Thermal stability of chemically denatured green fluorescent protein (GFP)—A preliminary study. *Thermochim Acta* 410:161–163.
49. Melnik TN, Povarnitsyna TV, Glukhov AS, Uversky VN, Melnik BS (2011) Sequential melting of two hydrophobic clusters within the green fluorescent protein GFP-cycle3. *Biochemistry* 50:7735–7744.
50. Camacho C, Thirumalai D (1993) Minimum energy compact structures of random sequences of heteropolymers. *Phys Rev Lett* 71:2505–2508.
51. Kimura T, et al. (2005) Specific collapse followed by slow hydrogen-bond formation of beta-sheet in the folding of single-chain monellin. *Proc Natl Acad Sci USA* 102:2748–2753.
52. Klimov D, Thirumalai D (1996) Factors governing the foldability of proteins. *Proteins* 26:411–441.
53. Bertz M, Kunfermann A, Rief M (2008) Navigating the folding energy landscape of green fluorescent protein. *Angew Chem Int Ed* 47:8192–8195.
54. Guo ZY, Thirumalai D, Honeycutt JD (1992) Folding kinetics of proteins—A model study. *J Chem Phys* 97:525–535.
55. Camacho CJ, Thirumalai D (1993) Kinetics and thermodynamics of folding in model proteins. *Proc Natl Acad Sci USA* 90:6369–6372.
56. Betancourt M, Thirumalai D (1999) Pair potentials for protein folding: choice of reference states and sensitivity of predicted native states to variations in the interaction schemes. *Protein Sci* 8:361–369.
57. Griesbeck O, Baird G, Campbell R, Zacharias D, Tsien R (2001) Reducing the environmental sensitivity of yellow fluorescent protein. mechanism and applications. *J Biol Chem* 276:29188–29194.
58. Auton M, Bolen D (2004) Additive transfer free energies of the peptide backbone unit that are independent of the model compound and the choice of concentration scale. *Biochemistry* 43:1329–1342.
59. Guo Z, Thirumalai D (1996) Kinetics and thermodynamics of folding of a de Novo designed four-helix bundle protein. *J Mol Biol* 263:323–343.