

# Revealing the bifurcation in the unfolding pathways of GFP by using single-molecule experiments and simulations

Moritz Mickler\*, Ruxandra I. Dima†, Hendrik Dietz\*, Changbong Hyeon<sup>‡§</sup>, D. Thirumalai<sup>¶||</sup>, and Matthias Rief<sup>\*||\*\*</sup>

\*Physik Department E22, Technische Universität München, James-Frank-Strasse, D-85748 Garching, Germany; †Department of Chemistry, University of Cincinnati, Cincinnati, OH 45221; ‡Center for Theoretical Biological Physics, University of California at San Diego, La Jolla, CA 92093; §Department of Chemistry, Chung-Ang University, Seoul 156-756, Republic of Korea; ¶Biophysics Program, Institute for Physical Science and Technology, University of Maryland, College Park, MD 20741; and \*\*Munich Center for Integrated Protein Science CiPSM, 81377 Munich, Germany

Edited by José N. Onuchic, University of California at San Diego, La Jolla, CA, and approved October 25, 2007 (received for review June 11, 2007)

**Nanomanipulation of biomolecules by using single-molecule methods and computer simulations has made it possible to visualize the energy landscape of biomolecules and the structures that are sampled during the folding process. We use simulations and single-molecule force spectroscopy to map the complex energy landscape of GFP that is used as a marker in cell biology and biotechnology. By engineering internal disulfide bonds at selected positions in the GFP structure, mechanical unfolding routes are precisely controlled, thus allowing us to infer features of the energy landscape of the wild-type GFP. To elucidate the structures of the unfolding pathways and reveal the multiple unfolding routes, the experimental results are complemented with simulations of a self-organized polymer (SOP) model of GFP. The SOP representation of proteins, which is a coarse-grained description of biomolecules, allows us to perform forced-induced simulations at loading rates and time scales that closely match those used in atomic force microscopy experiments. By using the combined approach, we show that forced unfolding of GFP involves a bifurcation in the pathways to the stretched state. After detachment of an N-terminal  $\alpha$ -helix, unfolding proceeds along two distinct pathways. In the dominant pathway, unfolding starts from the detachment of the primary N-terminal  $\beta$ -strand, while in the minor pathway rupture of the last, C-terminal  $\beta$ -strand initiates the unfolding process. The combined approach has allowed us to map the features of the complex energy landscape of GFP including a characterization of the structures, albeit at a coarse-grained level, of the three metastable intermediates.**

AFM experiments | coarse-grained simulations | cross-link mutants | pathway bifurcation | plasticity of energy landscape

**P**rotein structures, which are astounding examples of self-organization in living systems, reach their folded states by navigating through a rugged energy landscape. Considerable progress has been made in understanding the folding mechanisms of small, single-domain proteins by using a combination of theory and experiments (1–5). Folding of many of these proteins can be approximately described as being two-state-like, that is, their energy landscape does not exhibit pronounced local minima corresponding to partially folded or misfolded structures. However, the folding energy landscapes of larger proteins, with complex topology, can be difficult to characterize because of the presence of multiple metastable intermediate structures (6, 7). A detailed characterization of the structures of the intermediate states and the associated energetics is a challenge for the experimentalist. In part, the difficulty arises because complex proteins generally fold slowly and tend to aggregate in bulk experiments (8), a problem that can be avoided in mechanical unfolding of single proteins. Indeed, single-molecule mechanical methods have recently provided new possibilities for directly probing the energy landscapes of proteins and RNA because they can monitor the population of partially folded states (9–16). In addition, these experiments can explore regions of the

energy landscape that are inaccessible in conventional experiments, thus providing a complete picture of the folding process (17–20). It is also possible to extract kinetic and some structural features of the intermediates by measuring the contour length increases (10–12) of the detaching secondary structural elements as force is varied. However, the description of the structures of the transiently populated intermediates requires a combined experimental and modeling approach.

Here, we use molecular simulations and experiments that use internal disulfide bond engineering to control and unravel the complex unfolding pathways of GFP. In principle, the force-induced unfolding pathways can be obtained by using all-atom molecular dynamics simulations (21). Because of severe restrictions in the achievable simulation time scales such studies are forced to use loading rates ( $r_f$  values) that are several orders of magnitude greater than experiments (22–24). We circumvent this problem to predict the forced-unfolding pathways of GFP, a 238-residue protein with a barrel-like architecture, using a coarse-grained self-organized polymer model (SOP) (25, 26) for which simulations can be performed at  $r_f$  values that are comparable to those used in single-molecule atomic force microscopy (AFM) experiments (25). The response of cross-link mutants (27–29) to force in AFM experiments can be used to direct the flux of molecules along a specific pathway that permits us to probe the existence of multiple unfolding routes. The combination of simulations and experiments on the cycle 3 GFP and two cross-link mutants show that, after the initial rupture of the N-terminal  $\alpha$ -helix, unfolding of the wild-type GFP occurs by a bifurcation in the unfolding pathways. Unraveling in the major pathway occurs from the N terminus with three populated intermediates, whereas rupture in the minor pathway begins from the C terminus with two detectable intermediates. The flux between the pathways and the nature of pathways can be altered by appropriate cross-links. Our work shows that, even though the native topology is the key determinant of the force-induced unfolding mechanism (30, 31), the presence of many kinetic barriers leading to multiple unfolding pathways is linked to a subtle balance between energetics that stabilize the local structural elements and chain connectivity that determines the global fold.

Author contributions: M.M. and R.I.D. contributed equally; R.I.D., D.T., and M.R. designed research; M.M., R.I.D., H.D., and C.H. performed research; M.M., R.I.D., H.D., C.H., D.T., and M.R. analyzed data; and R.I.D., D.T., and M.R. wrote the paper.

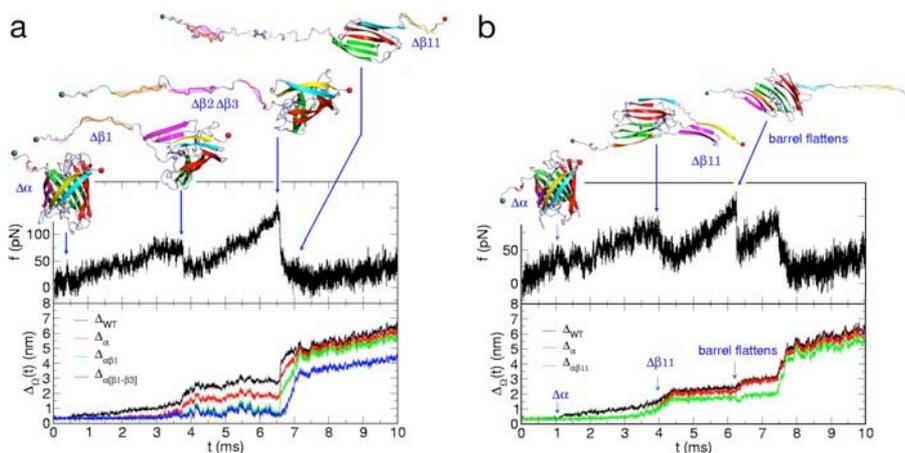
The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

||To whom correspondence may be addressed. E-mail: thirum@umd.edu or mrief@ph.tum.de.

This article contains supporting information online at [www.pnas.org/cgi/content/full/0705458104/DC1](http://www.pnas.org/cgi/content/full/0705458104/DC1).

© 2007 by The National Academy of Sciences of the USA



**Fig. 1.** Time-dependent changes in the force and root-mean-square deviations [ $\Delta\Omega(t)$ ]. (a) Changes in the force for cycle 3 GFP for molecules that reach state by the major pathway. (Upper) The structures of the three metastable intermediates along the unfolding pathway are shown and the secondary structural elements that detach are labeled. (Lower)  $\Delta\Omega(t)$  is shown for different structural elements  $\Omega$  ( $\Omega$  for black, red, blue, and green correspond to WT,  $\alpha$ ,  $\alpha\beta1$ ,  $\alpha[\beta1 - \beta3]$ , respectively). (b) Same as a, except the forces and the structures are for molecules that unfold along the minor pathway. (Lower)  $\Delta\Omega(t)$  is shown for various fragments. The green curve corresponds to  $\Delta_{\alpha\beta11}(t)$ .

## Results

### Forced Unfolding of Cycle 3 GFP (Wild Type) Occurs by Bifurcation of Pathways.

To unravel the complexity of the unfolding pathways we performed simulations by using the SOP model (see *Methods*) at the pulling speed,  $v = 2.5 \mu\text{m/s}$ . From simulations of a number of trajectories we find that there are two unfolding pathways. The time-dependent changes in the force curves for both pathways show multiple peaks (Fig. 1 *a* and *b*) with similar force values. To assign the structural changes that accompany the changes in force, we have determined the dynamical changes in the root-mean-square distances (RMSDs) with respect to the intact Protein Data Bank (PDB) structure [PDB ID code 1gfl (32)] and with various secondary structures removed. By comparing  $\Delta_{\text{WT}}(t)$  and  $\Delta_{\alpha}(t)$  we can infer the time when H1 rips from the intact structure by the jump in  $\Delta_{\text{WT}}(t)$ , while  $\Delta_{\alpha}(t)$  would remain flat. Here,  $\Delta_{\text{WT}}(t)$  is the RMSD in relation to the cycle 3 GFP structure and  $\Delta_{\alpha}(t)$  is the RMSD with respect to the folded structure with the  $\alpha$ -helix removed (see *Methods* and [Supporting Information \(SI\) Text](#) for details). In the major pathway, H1 opens at  $t \approx 0.6$  ms (Fig. 1*a*) that is indicated by a small abrupt increase in  $\Delta_{\text{WT}}$ . Similarly, comparison of  $\Delta_{\alpha}$  and  $\Delta_{\alpha\beta1}$  (computed with respect to a structure in which  $\alpha$  and the N-terminal  $\beta1$  are removed) shows that (Fig. 1*a*) there is an increase in  $\Delta_{\alpha}$  at  $t \approx 3.7$  ms, which corresponds to unraveling of strand  $\beta1$ . The increase in  $\Delta_{\alpha\beta1}$  in relation  $\Delta_{\alpha\beta1/\beta2/\beta3}$  shows that at  $t \approx 6.6$  ms the substructures  $\beta2/\beta3$  unravel. At times  $\geq 7$  ms, the RMSD of the WT and the various substructures increases greatly and is accompanied by deformation of the barrel. In our earlier study (25), we showed that, at this juncture,  $>60\%$  of the hydrophobic residues are exposed to the solvent. As a result the lifetimes of additional intermediates, which may be resolvable in simulations, would be too short for experimental detection (see *SI Text* for additional explanation). Thus, the simulations predict that there should be three detectable peaks that correspond to intermediates GFP $\Delta\alpha$ , GFP $\Delta\alpha\beta1$ , and GFP $\Delta\alpha\beta1\Delta[\beta2\beta3]$ , where the square brackets indicate that the strands  $\beta2$  and  $\beta3$  rip almost simultaneously.

The unfolding pathway in  $\approx 28\%$  of the molecules is drastically different (Fig. 1*b*). Because the magnitude of forces at which the first two peaks rip and the lengths of the detached secondary structural elements ( $\beta1$  and  $\beta11$ ) in the two pathways are similar (see Fig. 1 *a* and *b*), the structures are difficult to resolve experimentally (11). By following the analysis described above we have used the simulations to identify the secondary structural elements that unravel at the two discrete peaks. The unraveling of H1 occurs at  $t \approx 1$  ms. Comparison of  $\Delta_{\alpha}$  and  $\Delta_{\alpha\beta11}$  shows (Fig. 1*b*) that the C terminus strand  $\beta11$  rips at  $t \approx 4$  ms. Because the barrel is totally destabilized after  $\beta11$  unravels, no additional intermediates should be detectable. Thus, in the minor pathway

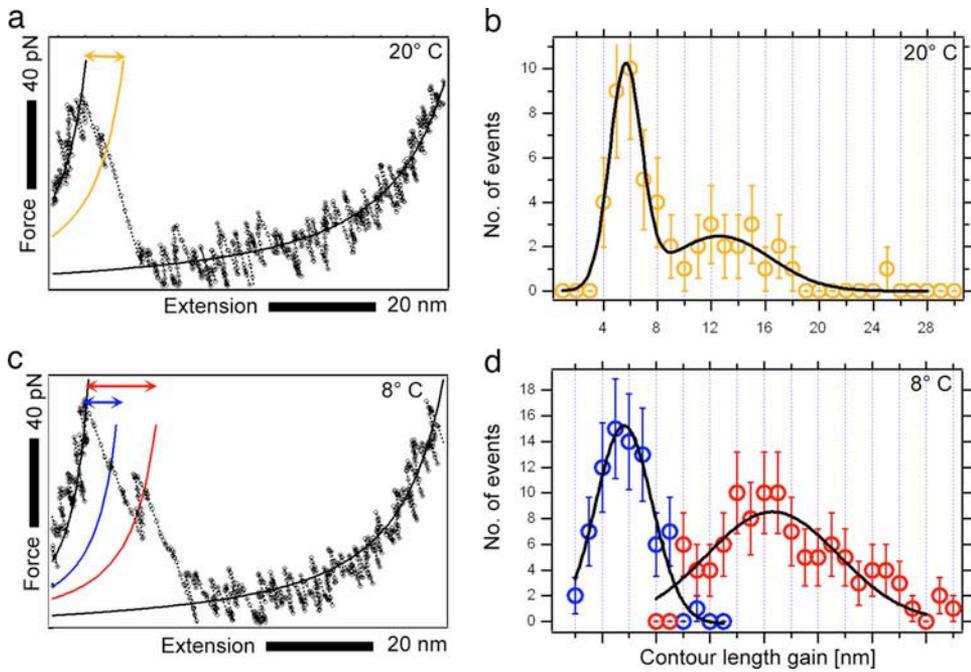
we predict that there should be only two discrete metastable intermediates, namely, GFP $\Delta\alpha$  and GFP $\Delta\alpha\beta11$ .

To validate the theoretical predictions we performed mechanical unfolding experiments using AFM (see *Methods*). Sample traces of GFP pulled at its termini at  $20^\circ\text{C}$  are shown in Fig. 2*a*. Recently, we showed (11) that GFP, when loaded at the termini, unfolds by the rupture of the small N-terminal  $\alpha$ -helix, resulting in the formation of GFP $\Delta\alpha$ . As in the simulations, the major unfolding peak proceeds from GFP $\Delta\alpha$ . During its relaxation phase, the AFM cantilever exhibits intermediate dwell levels that last for time spans ranging from hundreds of microseconds to a few milliseconds. A histogram of the length gains from the major peak to the intermediate levels shows a clear maximum at extension  $\approx 6.5$  nm (Fig. 2*b*). The second intermediate state (GFP $\Delta\alpha\beta$ ) lacks an additional full  $\beta$ -strand, whereas the folded portion of the intermediate presumably retains the native structure. The simulations show that it is the N terminus strand  $\beta1$  that unravels in the major pathway. Furthermore, by combining the theoretical predictions and experiments we can surmise that on unfolding of the  $\alpha$ -helix and  $\beta1$  the rest of the structure is intact (Fig. 1*a* and [SI Fig. 7](#)).

The histogram in Fig. 2*b* is skewed toward larger lengths that may be indicative of additional intermediates in the unfolding pathway. To resolve the presence of a third possible intermediate, we performed additional experiments at a lower temperature ( $8^\circ\text{C}$ ; see Fig. 2*c*). The presence of a distinct new population at extension peaked around 16.3 nm is clearly evident (Fig. 2*d*). Although the experiments show that a third intermediate is populated, we can assert only by comparison with SOP simulations that it should correspond to GFP $\Delta\alpha\beta1\Delta[\beta2\beta3]$  (Fig. 1*a*).

### Cross-Link Mutants Control Unfolding Pathways.

Can the details of the unfolding pathways predicted by the simulations also be observed in single-molecule mechanical experiments? In particular, the prediction that mechanical unfolding of GFP occurs by pathway bifurcation requires experiments that preferentially block one of the pathways. Earlier studies have used inserts of 5-aa residues into the loops of Ig domains to achieve such information (33, 34). The 2-nm gain in length on amino acid insertion makes it possible to identify the  $\beta$ -strands that are detached in an unfolding intermediate. However, large inserts can significantly compromise folding mechanism and the structure of the protein under investigation. Inserts of 5 alanine residues into the loops flanking the proximal and distal  $\beta$ -sheets of GFP failed to produce functional GFP protein in the current study. To circumvent these difficulties we used disulfide-bond engineering to control the unfolding pathways within a single protein domain. We engineered two different mutants with disulfide bonds to cross-link the N-terminal  $\beta$ -sheet (GFP $\beta_{\text{N-lock}}$ ) and the last C-terminal  $\beta$ -sheet (GFP $\beta_{\text{C-lock}}$ ) to the rest of the GFP structure (see Fig. 3 *a* and *d*). These two mutants should allow us

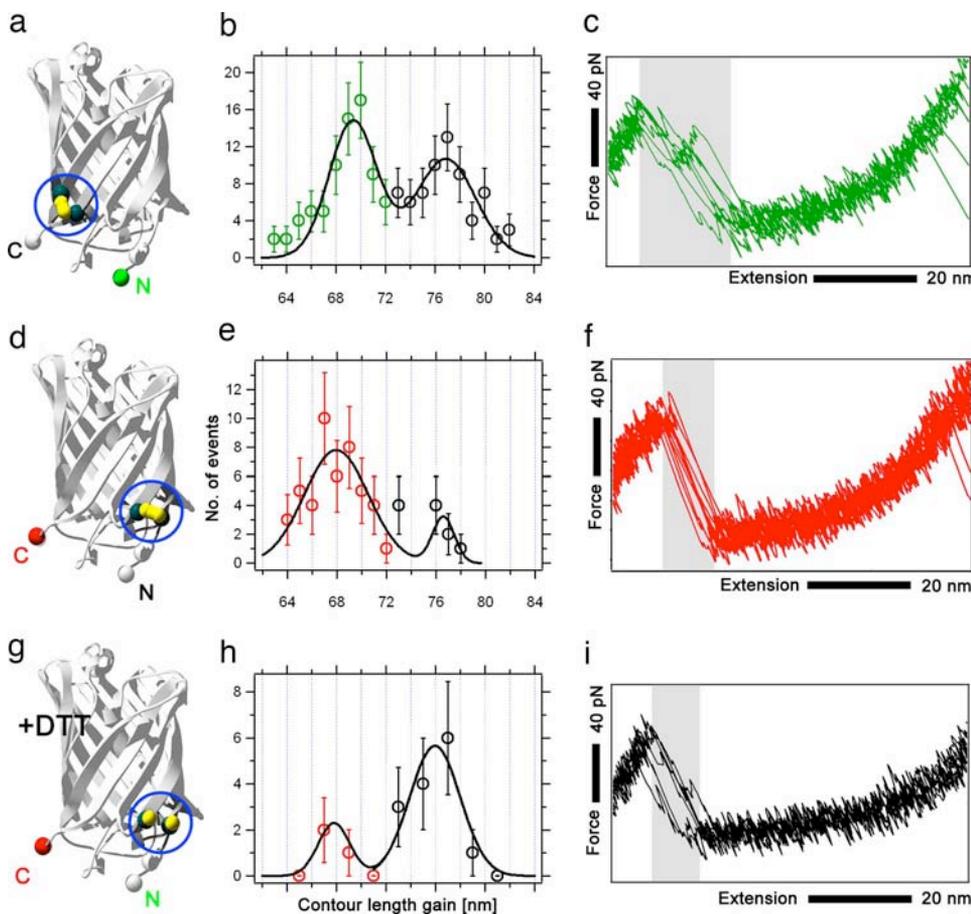


**Fig. 2.** Mechanical unfolding of wild-type GFP $\Delta\alpha$  at two temperatures. (a) Typical unfolding pattern of GFP $\Delta\alpha$  at 20°C. Solid lines show worm-like chain fits to the data. (b and d) Histogram of the length gains observed for partial unfolding of GFP $\Delta\alpha$  into subsequent intermediate states (see arrows in a and c). (c) Typical unfolding pattern of GFP $\Delta\alpha$  at 8°C.

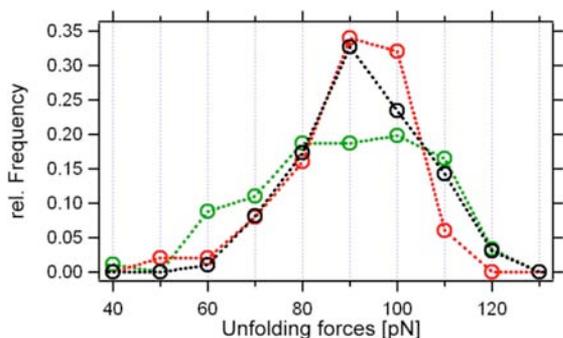
to distinguish whether unfolding of the GFP barrel proceeds by detachment of N- or C-terminal  $\beta$ -strands. The mutant that suppresses the major unfolding pathway should greatly affect the unfolding pattern as well as the unfolding force, whereas the other mutant should provide an essentially unaltered unfolding sequence.

Both mutants show absorption and fluorescence emission spectra that are indistinguishable from the wild type (SI Fig. 8). This is a strong indication that the cross-links do not alter the GFP structure, and hence the underlying energy landscape.

A histogram of the total unfolding length obtained for the



**Fig. 3.** Internal cysteine-engineering to alter the unfolding pathway of GFP $\Delta\alpha$ . (a) Scheme of GFP $\beta_{C-lock}$ . (b) Histogram of the length gain on unfolding of single GFP $\beta_{C-lock}$  molecules from GFP $\Delta\alpha$  to the fully unfolded state. Green section indicates the population where the disulfide bond has been successfully formed. (c) Overlay of several experimental force-extension traces for unfolding of GFP $\beta_{C-lock}$ , with the disulfide bond formed. The population of intermediate states, manifested by dwell levels in the cantilever relaxation phase (gray-shaded zone), is clearly visible. (d) Scheme of GFP $\beta_{N-lock}$ . (e) Histogram of the length gain on unfolding of single GFP $\beta_{N-lock}$  molecules from GFP $\Delta\alpha$  to the fully unfolded state. Red section indicates the population where the disulfide bond has been successfully formed. (f) Overlay of several experimental force-extension traces for unfolding of GFP $\beta_{N-lock}$ , with the disulfide bond formed. Dwell levels in the cantilever relaxation phase (gray-shaded zone) are consistently absent. (g) Scheme of GFP $\beta_{N-lock}$  under reducing conditions. (h) Histogram of the length gain on unfolding of single GFP $\beta_{N-lock}$  molecules from GFP $\Delta\alpha$  to the fully unfolded state as observed in experiments under reducing buffer conditions (20 mM DTT). Red section indicates the population where the disulfide bonds persist even under the action of DTT. (i) Overlay of several experimental force-extension traces for unfolding of reduced GFP $\beta_{N-lock}$ . Dwell levels in the cantilever relaxation phase are again consistently observed.



**Fig. 4.** Histogram of the forces required to induce unfolding of wild-type GFP $\Delta\alpha$  (black line), GFP $\beta_{C-lock}$  (red line), and GFP $\beta_{N-lock}$  (green line).

N-terminal cross-link (GFP $\beta_{N-lock}$ ) shows a major peak at 68.4 nm that is reduced by  $\approx 8.6$  nm compared with the 77 nm of the wild type (see Fig. 3e). This reduction in length is the expected value for the shortening of the GFP polypeptide chain due to the cross-link between the amino acids 11 and 36. The length histogram also exhibits a minor peak at the full-length gain of 77 nm. This peak represents a minor fraction of un-cross-linked protein. Hence,  $>86\%$  of the GFP molecules have successfully formed a cross-link. The major unfolding peak of the cross-linked GFP $\beta_{N-lock}$  shows a clean single-step transition from GFP $\Delta\alpha$  to the fully unfolded GFP (Fig. 3f). Even experiments performed at 8°C, which greatly enhanced the occurrence of intermediate states in the wild type (Fig. 2c), did not show any signs of intermediates in GFP $\beta_{N-lock}$ . Addition of DTT to the buffer (Fig. 3g) almost fully restored the length gain and unfolding intermediates as observed in the wild-type protein (see Fig. 3e, h, and i). Surprisingly, the force distribution of the cross-linked mutant is almost identical to the wild-type protein (Fig. 4).

Unfolding length distributions of the C-terminal cross-link mutant (GFP $\beta_{C-lock}$ ) exhibit a bimodal distribution at both the expected shortened length of 70 nm as well as the full length of 77 nm (Fig. 3b). Obviously, formation of the cross-link occurs only with  $\approx 50\%$  probability. The unfolding pattern of the short population with successfully formed disulfide bond exhibits both unfolding intermediates present also in the wild type (Fig. 2c). Again, the force distribution is almost identical to the one observed for the wild type (Fig. 4). Interestingly, experimental observation frequency of the unfolding intermediates GFP $\Delta\alpha\Delta\beta$  and GFP $\Delta\alpha\Delta\beta1\Delta\beta2$  seems to be increased in GFP $\beta_{C-lock}$  compared with the wild type (see Table 1).

**GFP $\beta_{N-lock}$  Mutant Unfolds from a Single Populated Intermediate.** The disulfide bond between the two CYS residues was mimicked by using a covalent bond that in the SOP model is represented by using the FENE potential (first term in Eq. 1). We generated 30 trajectories for this mutant, each starting from the same PDB conformation. The force-extension curves (FECs) show that there are several peaks. However, the instability of the intermediates beyond the first peak would escape experimental detection (see below). The structure of the first peak is assigned by comparing the

**Table 1. Frequency of observing dwell levels (detection limit  $\geq 100 \mu s$ ) in the cantilever relaxation after GFP $\Delta\alpha$  unfolding, reflecting population of the GFP $\Delta\alpha\Delta\beta$  intermediate state**

	Intermediate, %	No intermediate, %
Wild-type GFP	40.8	59.2
GFP $\beta_{C-lock}$	72.7	27.3
GFP $\beta_{N-lock}$	0	100

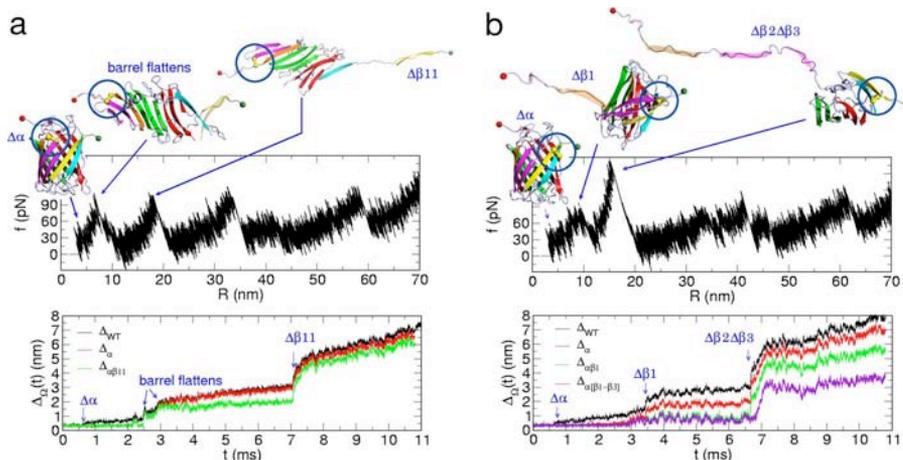
time-dependent variations in  $\Delta W_T$  and  $\Delta\alpha$  (Fig. 5a) which show that H1 unravels at  $t \approx 1.6$  ms. After the rupture of H1 the barrel opens, increasing the potential solvent accessibility of the core hydrophobic residues that, in turn, leads to complete unfolding of GFP. The prediction that the GFP $\beta_{N-lock}$  unfolds with only one populated intermediate is consistent with experimental results (Fig. 3f) that show only one peak in the FEC. Along the minor pathway in the unfolding of cycle 3 GFP there is an additional intermediate corresponding to unraveling of  $\beta11$  that is absent in the GFP $\beta_{N-lock}$  cross-link. The reason for the absence of the second intermediate is that, in the GFP $\beta_{N-lock}$  mutant, the barrel is destabilized prior to the unfolding of  $\beta11$ , whereas, in cycle 3 GFP, the barrel flattens after  $\beta11$  rips (see SI Fig. 7).

**Unfolding Route of the GFP $\beta_{C-lock}$  Mutant Coincides with the Major Pathway of Cycle 3 GFP.** In all of the 30 trajectories generated for the GFP $\beta_{C-lock}$  disulfide-bonded mutant, there are three detectable intermediates in the FECs as well as in the time dependence of the force. The structures of the intermediates can be unambiguously determined by comparing the time-dependent changes in RMSD with respect to the intact conformation and structures with selected missing secondary structural elements. Analysis of the RMSD plots (Fig. 5) shows that the order of unfolding coincides with the major pathway in cycle 3 GFP. The results are not only in accord with experiments on the GFP $\beta_{C-lock}$  mutant, but also confirm that in the majority of the cycle 3 GFP molecules unfolding is initiated from the N terminus.

## Discussion

The combined experimental and simulation approach of this study allows us to unravel in great detail the diverse unfolding paths and intermediate structures of GFP with complex architecture. Experiment (11) and simulations (25) show that the major mechanical unfolding event proceeds from an intermediate structure GFP $\Delta\alpha$  where the N-terminal  $\alpha$ -helix is detached from the GFP barrel. From here, the major path of unfolding occurs by two additional intermediate states. The SOP simulations predicted that unraveling of  $\beta1$  leading to GFP $\Delta\alpha\Delta\beta1$  is the second event in the major unfolding pathway. The data shown in Fig. 2 now prove that the detachment of the first  $\beta$ -strand occurs from the N terminus because cross-linking of this strand in GFP $\beta_{N-lock}$  completely eliminates the population of this intermediate (Fig. 3f). It is likely that the transition to the next intermediate state GFP $\Delta\alpha\Delta\beta1\Delta\beta2$  also involves detachment of two additional  $\beta$ -strands from the N terminus. However, this conclusion cannot be drawn from experimental data alone. Again, the SOP simulation results provide crucial information by corroborating the structure of the third intermediate GFP $\Delta\alpha\Delta\beta1\Delta\beta2$ . The SOP simulations show, in accord with experiments, that the detachment of the two  $\beta$ -strands from GFP $\Delta\alpha\Delta\beta1$  results in a length gain of 16.3 nm. Examination of the structures shows that the third intermediate does correspond to the unraveling of  $\beta2$  and  $\beta3$  with little perturbation of the rest of the structure (SI Figs. 7 and 8).

The most striking conclusion from the simulations is the prediction of a bifurcation between the major N-terminal pathway involving the two additional intermediates and a minor C-terminal pathway lacking additional unfolding intermediates. Bifurcation between two pathways is extremely difficult to predict because it requires very precise determination of the kinetic barriers. A mere difference of 1 kT in activation barrier height shifts the relative flux between the two competing pathways by almost a factor of 3. Do we find any evidence for such a bifurcation in the experimental data? The first piece of evidence is the striking similarity in the unfolding force distributions of the wild-type GFP, GFP $\beta_{N-lock}$ , and GFP $\beta_{C-lock}$  (Fig. 4). Obviously, a very similar force is required to unfold GFP both from the N terminus and from the C terminus. A bifurcation



**Fig. 5.** Simulation results for force-extension curves and  $\Delta\Omega(t)$  for cross-link mutants. (a) Force as a function of extension for GFP $\beta_{N-lock}$  mutant. Residues 11 and 36 are connected by a disulfide cross-link. The structural changes that accompany the unfolding process are shown by blue arrows. The bottom profile gives  $\Delta\Omega(t)$  as a function of  $t$  for  $\Omega =$  WT (black),  $\Omega = \alpha$  (red), and  $\Omega = \alpha\beta11$  (green). (b) Same as a, except the FEC and  $\Delta\Omega(t)$  are for the GFP $\beta_{C-lock}$  mutant in which residues 202 and 225 are covalently linked by a disulfide bond. In addition to the  $\Delta\Omega(t)$  shown in a, the purple gives  $\Delta\alpha[\beta1-\beta3](t)$ .

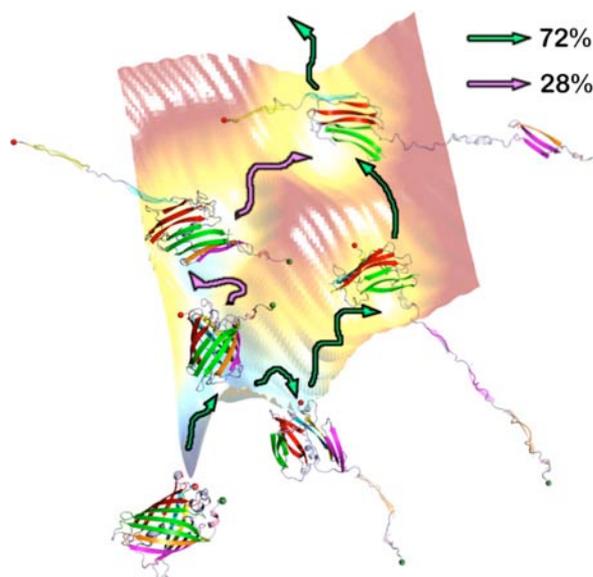
should also lead to a significant fraction of unfolding events in the GFP proceeding from the C terminus and, hence, lacking both the intermediates GFP $\Delta\alpha\Delta\beta1$  and GFP $\Delta\alpha\Delta\beta1\Delta2\beta$ . Analysis of the relative fraction of those curves exhibiting intermediates and those curves lacking intermediates for both GFP and GFP $\beta_{C-lock}$  is shown in Table 1. The fraction of curves exhibiting intermediates is significantly increased in GFP $\beta_{C-lock}$  compared with GFP. This observation is consistent with a bifurcation because the C-terminal cross-link mutant GFP $\beta_{C-lock}$  likely funnels the unfolding events into the N-terminal unfolding channel. The existence of two pathways in the mechanical unfolding of the wild-type GFP should manifest itself in a slight difference in unfolding force for the events unfolding by the two respective pathways. We pooled all data from the wild type and analyzed the unfolding forces for the two pathways separately. For the pathway with intermediate we find an average unfolding force of  $93.76 \pm 0.9$  pN ( $n = 194$ ) and for the pathway lacking the intermediate the average unfolding force is  $98.22 \pm 1.28$  pN ( $n = 107$ ). The error is given by the standard error of the mean. This finding further supports the existence of two unfolding pathways. The unfolding simulations of the cross-link mutants provide a microscopic basis for such an interpretation. The simulations of GFP $\beta_{C-lock}$  clearly show that its unfolding pathway coincides with the major unfolding pathway of cycle 3 GFP. Thus, blocking the minor pathway diverts the flux of molecules from the minor to the major unfolding route. Surprisingly, unfolding of GFP $\beta_{N-lock}$  proceeds by populating only GFP $\Delta\alpha$  because the barrel is entirely destabilized (SI Fig. 9) before the rupture of  $\beta11$ .

In a previous study we showed that the multiexponential lifetime distribution of GFP $\Delta\alpha\Delta\beta$  is also consistent with a simple single-pathway reaction scheme (11). Only combination of simulation and cross-link mutants allowed us to distinguish unambiguously between the single and bifurcating pathway picture. Taken together the present results show that during the transition to the stretched state induced by force, cycle 3 GFP explores a complex energy landscape (Fig. 6) in which there is a multiplicity of routes. Clearly, the fraction of molecules that traverse the two dominant pathways can be altered by mutations or force direction, which attests to the inherent plasticity of GFP and, in all likelihood, other proteins as well. Indeed, it is well recognized that the features of the energy landscapes of proteins that determine the folding pathways and kinetic barriers are controlled by a subtle balance between native topology and ener-

getics that stabilize the folded structure (3–5). The present and previous studies show that the energy landscape can presumably be manipulated by using cross-link mutants or by merely changing the points of force application that alters the direction along which strain propagates (16, 27, 35), and hence the unfolding routes in the rugged energy landscape (Fig. 6).

## Methods

**Self-organized Polymer (SOP) Model.** We use a topology-based model for GFP in which each amino acid is represented by using the  $C_\alpha$  atom. The total energy



**Fig. 6.** The complex energy landscape for GFP unfolding is constructed based on the results of simulations and experiments. Starting from the folded structure, which corresponds to the native basin of attraction, unfolding occurs by bifurcation in the pathway after the rupture of the  $\alpha$ -helix. The molecules that unfold by the dominant pathway are shown by green arrows and the purple arrows show the fate of GFP molecules that follow the minor pathway. The structures of the intermediates in the various basins are explicitly shown. The approximate fraction of the molecules along each pathway is indicated. These numbers can be altered by mutations (cross-link in this study) and by changing the force direction. Thus, the energy landscape is not only rugged, but also can be manipulated.

function for a conformation, specified in terms of the coordinates  $\{r_{ij}\}$  ( $i = 1, 2, \dots, N$ ), where  $N$  is the number of residues is

$$\begin{aligned}
 V_T &= V_{\text{FENE}} + V_{\text{NB}}^{\text{ATT}} + V_{\text{NB}}^{\text{REP}} \\
 &= - \sum_{i=1}^{N-1} \frac{k}{2} R_0^2 \log \left( 1 - \frac{(r_{i,i+1} - r_{i,i+1}^o)^2}{R_0^2} \right) \\
 &\quad + \sum_{i=1}^{N-3} \sum_{j=i+3}^N \varepsilon_h \left[ \left( \frac{r_{ij}^o}{r_{ij}} \right)^{12} - 2 \left( \frac{r_{ij}^o}{r_{ij}} \right)^6 \right] \Delta_{ij} \\
 &\quad + \sum_{i=1}^{N-2} \varepsilon_l \left( \frac{\sigma}{r_{i,i+2}} \right)^6 + \sum_{i=1}^{N-3} \sum_{j=i+3}^N \varepsilon_l \left( \frac{\sigma}{r_{ij}} \right)^6 (1 - \Delta_{ij}). \quad [1]
 \end{aligned}$$

The distance between two neighboring interaction sites  $i$  and  $i + 1$  is  $r_{i,i+1}$  and  $r_{ij}^o$  is its value in the native structure. The finite extensible nonlinear elastic (FENE) potential (first term in Eq. 1) describes the backbone chain connectivity. We use Lennard-Jones potential (second term in Eq. 1) to account for the interactions that stabilize the native state. If the noncovalently linked beads  $i$  and  $j$  for  $|i - j| > 2$  are within a cutoff distance  $R_C$  (i.e.,  $r_{ij} < R_C$ ), then  $\Delta_{ij} = 1$ . If  $r_{ij} > R_C$ , then  $\Delta_{ij} = 0$ . A uniform value for  $\varepsilon_h$ , which specifies the strength of the nonbonded interactions, is assumed. All nonnative interactions (third term in Eq. 1) are repulsive. To prevent interchain crossing we chose appropriate value of  $\sigma$  (see SI Table 2 for the values of the parameters in the energy function). The loading rates in this study, unlike in the standard steered molecular dynamics simulations, are comparable to the experimental values.

**Simulations.** We used Brownian dynamics simulations (36, 37) at  $T = 300$  K to generate the mechanical unfolding trajectories. We estimate the simulation time scale by using  $\tau_{\text{fl}} = \left( \frac{\zeta \varepsilon_h}{k_B T} \right) h \tau_L$ , where  $\tau_L = 3$  ps. We used  $\zeta = 50 \tau_L^{-1}$  in the overdamped limit, which approximately corresponds to the friction constant for a molecule in water. After the experimental setup, the C-terminal end is stretched at a constant pulling speed ( $v = 2.5 \mu\text{m/s}$ ) while keeping the N-terminal end fixed.

The cantilever spring constant is taken to be  $k_s = 35$  pN/nm, which is in the range of (1–100) pN/nm used in the AFM experiments.

#### Assigning Structures of Intermediates by Comparing Global and Partial RMSDs.

We compared the time-dependent changes in the global  $[\Delta_{\text{WT}}(t)]$  partial RMSDs  $[\Delta_{\Omega}(t)]$  to compute the time when a secondary structural element detaches from the folded GFP. Here,  $\Delta_{\text{WT}}(t)$  for a conformation at  $t$  is calculated with respect to the PDB structure of cycle 3 GFP, whereas  $\Delta_{\Omega}(t)$  is calculated with respect to the folded structure from which the secondary structural element  $\Omega$  is removed (see SI Text for details). By varying  $\Omega$  we can pinpoint the structure and the life-time of specific intermediates.

**Single-Protein Force Spectroscopy.** Single-molecule force measurements were performed on a custom-built atomic force microscope at room temperature. Gold-coated cantilevers (BioLevers, Olympus) with spring constant and resonance frequency of 6 pN/nm and 2 kHz (type B) were used. Temperature control was achieved by placing the experimental setup in a peltier-operated fridge (MediaMarkt). Protein solution was incubated on freshly evaporated gold surfaces. All force traces were collected at a pulling speed of  $2 \mu\text{m/s}$ . Unfolding forces were determined by taking the force peak value for each individual unfolding event in single-molecule traces.

**Contour Length Measurements.** Worm-like chain force-extension traces were calculated by using the interpolation formula reported by Bustamante *et al.* (38) and fit to individual peaks in the experimental force-extension traces by varying the contour length with a fixed persistence length (= 0.5 nm).

**Internal Cysteine Engineering.** We investigated cycle 3 GFP with an additional point mutation S2G. The DdFLN (1–5)-GFP fusion protein was as described in ref. 11. Pairwise point mutations of wild-type GFP residues to cysteines (residues 202 and 225 for GFP $_{\beta\text{-lock}}$  and residues 11 and 36 for GFP $_{\beta\text{N-lock}}$ ) were performed with a QuikChange multisite-directed mutagenesis kit (Stratagene). Purification of the His $_6$ -tagged proteins was done with Ni-NTA affinity chromatography at 4°C. All GFP double-cysteine mutants showed typical, bright GFP fluorescence, indicating the presence of the native GFP structure. Absorption and emission spectra are shown in the SI Fig. 8. Proteins were stored in PBS-buffer, at pH = 7.4. Internal disulfide bonds formed spontaneously without further treatment.

**ACKNOWLEDGMENTS.** This work was supported in part by National Science Foundation Grant CHE-05-14056 (to D.T.) and by Deutsche Forschungsgemeinschaft Grant RI 99013-1 (to M.R.).

- Eaton WA, Munoz V, Hagen SJ, Jas GS, Lapidus LJ, Henry ER, Hofrichter J (2000) *Annu Rev Biophys Biomol Struct* 29:327–359.
- Fersht AR, Daggett V (2002) *Cell* 108:573–582.
- Onuchic JN, Wolynes PG (2004) *Curr Opin Struct Biol* 14:70–75.
- Thirumalai D, Hyeon C (2005) *Biochemistry* 44:4957–4970.
- Shakhnovich E (2006) *Chem Rev* 106:1559–1588.
- Baldwin RL, Rose GD (1999) *Trends Biochem Sci* 24:77–83.
- Enoki S, Saeki K, Maki K, Kuwajima K (2004) *Biochemistry* 43:14238–14248.
- Zimmer M (2002) *Chem Rev* 102:759–781.
- Forman JR, Clarke J (2007) *Curr Opin Struct Biol* 17:58–66.
- Oesterhelt F, Oesterhelt D, Pfeiffer M, Engel A, Gaub HE, Müller DJ (2000) *Science* 288:143–146.
- Dietz H, Rief M (2004) *Proc Natl Acad Sci USA* 101:16192–16197.
- Onoa B, Dumont S, Liphardt J, Smith SB, Tinoco I, Jr, Bustamante C (2003) *Science* 299:1892–1895.
- Li PTX, Bustamante C, Tinoco I, Jr (2007) *Proc Natl Acad Sci USA* 104:7039–7044.
- Perez-Jimenez R, Garcia-Manyes S, Ainaravaru S, Fernandez JM (2006) *J Biol Chem* 281:40010–40014.
- Ng SP, Rounsevell RW, A, S, Geierhaas CD, Williams PM, Paci E, Clarke J (2005) *J Mol Biol* 350:776–789.
- Brockwell DJ, Paci E, Zinober RC, Beddard GS, Olmsted PD, Smith DA, Perham RN, Radford SE (2003) *Nat Struct Biol* 10:731–737.
- Hyeon C, Thirumalai D (2006) *Biophys J* 90:3410–3427.
- Fernandez JM, Li H (2004) *Science* 303:1674–1678.
- Li MS, Hu CK, Klimov DK, Thirumalai D (2006) *Proc Natl Acad Sci USA* 103:93–98.
- Walther KA, Gräter F, Dougan L, Badilla CL, Berne BJ, Fernandez JM (2007) *Proc Natl Acad Sci USA* 104:7916–7921.
- Isralewitz B, Gao M, Schulten K (2001) *Curr Opin Struct Biol* 11:224–230.
- Lu H, Schulten K (2000) *Biophys J* 79:51–65.
- Gräter F, Shen JH, Jiang HL, Gautel M, Grubmüller H (2005) *Biophys J* 88:790–804.
- Paci E, Karplus M (2000) *Proc Natl Acad Sci USA* 97:6521–6526.
- Hyeon C, Dima RI, Thirumalai D (2006) *Structure (London)* 14:1633–1645.
- Hyeon C, Thirumalai D (2007) *Biophys J* 92:731–743.
- Dietz H, Berkemeier F, Bertz M, Rief M (2006) *Proc Natl Acad Sci USA* 103:12724–12728.
- Ainaravaru SR, Bruijic J, Huang HH, Wiita AP, Lu H, Li L, Walther KA, Carrion-Vazquez M, Li H, Fernandez JM (2007) *Biophys J* 92:225–233.
- Sharma D, Perisic Q, Peng Q, Cio Y, Lam C, Lu H, Li H (2007) *Proc Natl Acad Sci USA*, in press.
- Klimov DK, Thirumalai D (2000) *Proc Natl Acad Sci USA* 97:7254–7259.
- Klimov DK, Thirumalai D (1999) *Proc Natl Acad Sci USA* 96:6166–6170.
- Yang F, Moss LG, Phillips GN, Jr (1996) *Nat Biotechnol* 14:1246–1251.
- Schwaiger I, Kardinal A, Schleicher M, Noegel AA, Rief M (2004) *Nat Struct Mol Biol* 11:81–85.
- Carrion-Vazquez M, Oberhauser AF, Fowler SB, Marszalek PE, Broedel SE, Clarke J, Fernandez JM (1999) *Proc Natl Acad Sci USA* 96:3694–3699.
- Carrion-Vazquez M, Li H, Lu H, Marszalek PE, Oberhauser AF, Fernandez JM (2003) *Nat Struct Biol* 10:738–743.
- Ermak DL, McCammon JA (1978) *J Chem Phys* 69:1352–1369.
- Veitshans T, Klimov D, Thirumalai D (1996) *Folding Des* 2:1–22.
- Bustamante C, Marko JF, Siggia ED, Smith S (1994) *Science* 265:1599–1600.