Published in final edited form as:

J Mol Biol. 2007 October 19; 373(2): 476–490. doi:10.1016/j.jmb.2007.07.071.

The rough energy landscape of superfolder GFP is linked to the chromophore

Benjamin T. Andrews¹, Andrea R. Schoenfish¹, Melinda Roy¹, Geoffrey Waldo², and Patricia A. Jennings^{1,*}

¹ Department of Chemistry and Biochemistry, University of California, San Diego, La Jolla, CA, 92093-0375

² Los Alamos National Laboratory, Los Alamos, NM, 87545

Abstract

Many GFP variants have been developed for use as fluorescent tags, and recently a superfolder GFP (sfGFP) has been developed as a robust folding reporter. This new variant shows increased stability and improved folding kinetics, as well as 100% recovery of native protein after denaturation. Here, we characterize sfGFP, and find that this variant exhibits hysteresis as unfolding and refolding equilibrium titration curves are non-coincident even after equilibration for more than eight half-lives as estimated from kinetic unfolding and refolding studies. This hysteresis is attributed to trapping in a native-like intermediate state. Mutational studies directed towards inhibiting chromophore formation indicate that the novel backbone cyclization is responsible for the hysteresis observed in equilibrium titrations of sfGFP. Slow equilibration and the presence of intermediates imply a rough landscape. However, *de novo* folding in the absence of the chromophore is dominated by a smoother energy landscape than that sampled during unfolding and refolding of the post-translationally modified polypeptide.

Keywords

β -barrel; hysteresis; protein folding; free energy landscape; post-translational modification

INTRODUCTION

Green Fluorescent Protein (GFP) has become a common label for many *in vivo* and *in vitro* applications due to GFP's ability to fold and form a visually fluorescent chromophore through autocatalytic cyclization and dehydration/oxidation reactions¹. GFP has been used as a reporter in folding^{2; 3}, protein-protein interactions^{4; 5; 6; 7}, and gene translation^{8; 9; 10}. However, the rate of GFP folding and chromophore formation limits the temporal resolution of these techniques, and their common use in high-throughput applications¹¹.

Studies of GFP have shown that folding is slow and that oxidation of the chromophore is the limiting step in fluorescence maturation¹². As the refolding of mature GFP is not fully reversible and is prone to aggregation¹³, variants of GFP have been developed with better folding and fluorescence properties- namely Cycle3 GFP (F99S M153T V163A)¹⁴, and

*Corresponding author, E-mail: E-mail: pajennin@ucsd.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

GFPmut2 (S65A V68L S72A)¹⁵. Folding improvements in Cycle3 GFP are attributed to avoiding aggregation traps in folding¹³; still, folding is very slow ($t_{1/2} = 4.5$ min) and limited by a proline isomerization^{13; 16}. Single molecule and solution studies of GFPmut2 show a lower stability and slower kinetics than Cycle3^{17; 18}, probably as it was optimized for *in vivo* fluorescent use rather than for stability and folding kinetics. While Cycle3 GFP has improved folding, the recovery of fluorescence after guanidine, urea, or acid denaturation is still not fully reversible, with a reported 80% rescue of native protein¹³. Thus, improving refolding efficiency to 100% would create a better reporter for a variety of applications^{3; 19; 20}.

Recently, a “superfolder” GFP (sfGFP) variant has been developed to minimize aggregation and speed up folding. This is accomplished by fusing it to the poorly folding protein ferritin, which also improves ferritin folding and recovery²¹. The sfGFP includes the Cycle3 mutations F99S, M153T, and V163A, the enhanced GFP mutations F64L and S65T, as well as six other mutations realized through directed evolution (S30R, Y39N, N105T, Y145F, I171V, A206V). This variant also contains the original Q80R mutation from PCR cloning²². This variant exhibits 100% fluorescence recovery after refolding from the urea denatured state, as well as faster refolding kinetics compared to cycle3 GFP when compared at a single denaturant concentration. The robustness of sfGFP folding is attributed to reduced misfolding and aggregation. These results are consistent with the design strategy of sfGFP, generating variants by selecting bright fluorescent colonies while expressing an aggregation-prone sfGFP-ferritin fusion protein²¹. Much of the folding improvement is attributed to the S30R mutation, which extends an internal ion-pair network within the interior of sfGFP²¹.

In this study, we characterized the full folding landscape of sfGFP using stopped flow and manual-mixing fluorescence kinetics, as well as fluorescent and circular dichroism (CD) equilibrium techniques. We find that sfGFP folds ten-fold faster than cycle3 GFP over a range of denaturant concentrations, and exhibits apparent hysteresis in equilibrium unfolding and refolding experiments, suggestive of a non-optimized energy landscape for this protein. However, while the funneled organization of the energy landscape dominates the kinetics of folding for highly evolved and/or well designed systems, GFP may be a special case because of its unusual chromophore. That is, the *de novo* folding pathway of GFP proceeds prior to chromophore formation while the refolding of GFP proceeds in the presence of the already cyclized chromophore. This landscape is not subject to the same evolutionary pressure as is folding *in vivo*. Consistent with this hypothesis, the apparent hysteresis is ameliorated in the presence of mutations that impede chromophore formation. One-dimensional ¹H NMR spectra show that sfGFP trapped in refolding has the same overall topology as sfGFP, while displaying subtle structural perturbations in the novel central helix containing the backbone cyclized chromophore. Taken together, our data provide the first explanation for the long-held observation of the inability to observe true equilibration in GFP folding based on unique structural characteristics of GFP, compared to other proteins.

RESULTS

The sfGFP was developed for robust folding by directed evolution and contains twelve mutations with respect to the wild-type protein²¹. These mutations include the mutation Q80R from PCR cloning, Cycle3 mutations F99S, M153T, and V163A, the enhanced GFP mutations F64L and S65T, as well as six other mutations (S30R, Y39N, N105T, Y145F, I171V, A206V), as reported previously²¹. In the current study we characterize the folding landscape of sfGFP.

Fluorescence and Circular Dichroism Spectra of Native & Denatured sfGFP

We used a combination of techniques to characterize the folding of sfGFP. The characteristic optical spectra of the native and denatured sfGFP are shown in Figure 1. The fluorescence

emission spectrum obtained upon excitation of the chromophore at 450 nm is similar to wtGFP (Figure 1a), with a peak at 508 nm, and a red-shifted, small shoulder arising from the protonated chromophore¹. Protein unfolding is accompanied by a complete loss of fluorescence emission at 508 nm. The fluorescence emission spectra of a second probe, the single tryptophan (Y74) of sfGFP, upon excitation at 295 nm for the native and denatured states are given in Figure 1b. The native emission spectrum has a fluorescence maximum at 328 nm. This is consistent with sequestration of the tryptophan side chain from solvent and packing into the protein core. The emission spectrum of the denatured protein is red shifted relative to that observed for the native protein, with a fluorescence maximum at 355 nm, and is consistent with the exposure of the tryptophan side-chain to aqueous solution upon unfolding²³. A third spectroscopic probe, the far-UV circular dichroism (CD) spectrum of native sfGFP (Figure 1c), exhibits a minimum at 217 nm with a shoulder at 222 nm, consistent with the small fraction of α -helix within the large β -barrel of sfGFP. Denaturation results in a loss of the characteristic CD spectrum of the native protein.

Equilibrium Transition Curves

Previous equilibrium unfolding of wild-type and mutant GFPs indicated a high stability of the native protein, but problems with protein aggregation resulted in a maximal recovery of $\approx 80\%$ of native signal during refolding²¹. Preliminary studies with sfGFP indicated that refolding and unfolding are fully reversible as 100% of native signal is recovered under strongly refolding conditions. Curves of the fraction of unfolded protein (F_{unfolded}) as a function of final denaturant concentration for the unfolding transition after 24, 96, and 192 hours of equilibration are shown in Figure 2 as detected by changes in chromophore fluorescence as a function of final denaturant concentration. The unfolding transition is unchanged between 96 and 192 hours, indicating that an “equilibrium” has been achieved.

In order to confirm the reversibility of the folding transition, we performed a *refolding* equilibrium analysis and the results of this study are plotted with the equilibrium *unfolding* data in Figure 3a. The unfolding transition fits to a three-state model; a two-state model (shown, Figure 3a, cyan, dotted line) was inadequate to fit the data. However, the refolding transition fits to a two-state model. Interestingly, while 100% of the fluorescent amplitude and CD are recovered under strongly refolding conditions ($\leq 1.7\text{M}$ Guanidine Hydrochloride), the unfolding and refolding transition curves are non-coincident (see Discussion). These results are consistent with observations from prior urea equilibrium studies, as the equilibrium refolding transition was observed at approximately 4.5 M urea at 25°C only after denaturation in 9M urea at 95°C²¹.

In an effort to test whether an irreversible process occurred during unfolding to which our probes were insensitive, we performed additional studies to check for non-coincidence of the unfolding and refolding transitions. **First**, we varied the protein concentration over an order of magnitude, and found identical behavior (data not shown) suggesting that higher order association was not responsible for this result. **Second**, we performed equilibrium denaturation on protein that had *previously been subjected* to the unfolding-refolding cycle. The unfolding transition curve for the unfolded-refolded protein is *identical* to the original unfolding curve obtained for the native sfGFP (which had never been previously subjected to unfolding data not shown). **Third**, we monitored the unfolding and refolding equilibrium transitions by tryptophan fluorescence as well as far-UV CD. The results of these studies are identical to the transitions observed by chromophore fluorescence, showing that the results are independent of the spectroscopic method used. **Fourth**, we compared the unfolding of protein over time which had previously been subjected to the unfolding-refolding cycle to that which had never been unfolded. The decay in fluorescent amplitude for either the native or refolded sfGFP exhibits the same rate and amplitude in their kinetic behavior. Therefore, there is no loss of

stability in the refolded protein (Figure 3b). Taken together, these data indicate that the apparent hysteresis between equilibrium unfolding and refolding transitions is an intrinsic property of the system, and is independent of experimental setup (see Discussion).

Unfolding Kinetics

The kinetic rates of sfGFP protein unfolding were monitored by both manual-mixing and stopped-flow fluorescence techniques. Unfolding was initiated by rapid dilution of native protein into varying final Gdn-HCl concentrations and the unfolding rates (τ_{u1}) were determined from fitting the time-dependent change in the fluorescence intensity (see Methods). The observed relaxation times, τ_{u1} , are plotted as a function of final denaturant concentration and the resulting chevron plot²⁴ is presented in Figure 4. Manual mixing and stopped flow results exhibit two relaxation phases at 7.0M Gdn-HCl and higher, and three phases at 6.5M Gdn-HCl and lower. The τ_{u3} phase shows weak dependence on the final denaturant concentration, and the τ_{u2} relaxation time is independent of the final Gdn-HCl concentration between 5 and 6.5 M Gdn-HCl (Figure 3). The τ_{u1} phase shows the strong denaturant dependence expected for a refolding reaction. Above 6.5 M Gdn-HCl, τ_{u2} and τ_{u1} coalesce. Extrapolation of the unfolding denaturant dependence of the τ_{u2} phase to 0M Gdn-HCl indicates a half life of 28 years for unfolding at room temperature.

Refolding Kinetics

The refolding of sfGFP was followed by manual mixing and stopped-flow fluorescence techniques. The time-dependent change within the first 200 seconds in the fluorescence intensity for a typical refolding reaction is presented in Figure 5. The first 400ms of the reaction (Figure 5, inset) shows a lag phase, which is evidence of an on-route, obligatory intermediate^{25; 26}, which has been reported for cycle3 GFP previously^{16; 27}. Three relaxation phases adequately describe the refolding process (see Figure 4) and, unlike the unfolding kinetics, each phase shows a strong denaturant dependence. The kinetic rates were at least an order of magnitude faster than observed for wtGFP and cycle3 GFP¹³. The time constants for the lag phase were estimated from the length of the lag phase versus guanidine, and are shown as filled triangles in Figure 4. The symbols are different to represent the estimation of the time constants, as opposed to fitting from direct kinetic data. The increase in the persistence of the lag phase may be attributed to the stabilization of the intermediate compared to the native state. Interestingly, no observable folding kinetics could be recorded between 1.75 and 5M Gdn-HCl for either the unfolding or refolding reactions by any techniques. The denaturant dependence of the unfolding reactions and its associated amplitudes would suggest that neither kinetic rates nor amplitude concerns should account for the inability to measure unfolding between 4.5 and 5M Gdn-HCl (See Discussion). Extrapolation of the denaturant dependence of the refolding relaxation times to 0M denaturant indicates half-lives of 390ms for τ_{R1} , 79ms for τ_{R2} , and 1.1ms for τ_{R3} . Corresponding time constants (directly from the Figure 4) are 520 ms, 140 ms, and 16 ms, respectively. Fitting of the overall apparent rate from the known folding and unfolding rates ($k_{app} = k_u + k_f$) gives a maximum half-life of 245 hours, or 10.2 days. However, as described above, there is no data in the hysteresis zone, and these numbers are an approximation based on known knowns. The unique β -barrel surrounding an α -helix fold of GFP has extremely high values of absolute contact order (29.5) and 214 sequence-distant pairs, leading to slow folding kinetics²⁸.

Double Jump Kinetics

The refolding of sfGFP shows three relaxation phases, all of which are Gdn-HCl dependent (Figure 4). Multiple phases are evidence of proline-isomerization limiting folding²⁹. Unfolding relaxation kinetics has a phase (τ_{u2}), which has no guanidine dependence; potential evidence of a proline isomerization becoming a rate-limiting step in unfolding at intermediate

guanidine concentrations. The presence of cis and trans prolyl-peptide bonds in the native state that give different isoforms with differing stabilities would also give rise to these results²⁹. According to the original proline isomerization hypothesis, the protein must have each of its proline imide bonds in the correct native state isomerization before refolding. Protein molecules which contain incorrect isomerization states must wait for the correct imide bond state before refolding. More recent studies have shown that incorrect isomers may fold, and the coupling of folding and isomerization will yield different rates for the folding process^{30; 31; 32}.

In order to test for the role of proline isomerization in the folding kinetics of sfGFP, a series of “double jump” or interrupted refolding experiments were carried out using manual mixing. This experiment involves quickly unfolding the protein (relative to proline isomerization) at high denaturant concentration. After a variable delay time, t , we dilute out the denaturant and measure the total intensity change and associated rates for refolding to the native state. If proline isomerization is occurring, there will be a progressive change in the amplitude of the observed refolding phases as a function of the delay time prior to refolding. The experiment involves a double jump of 0M to 7.4M to 0.22M Gdn-HCl, with the time spent in 7.4M recorded as the delay time. The fraction of the total amplitude as a function of the delay time is given in Figure 6. The data indicates a shifting population of the fast pathway (τ_2) to the slow pathway (τ_1) over the aging time. This is consistent with a proline isomerization occurring after unfolding.

Cyclophilin A

The amplitude shift of refolding phases during interrupted refolding experiments is consistent with proline isomerization limiting the folding reaction. However, any spectroscopically hidden phase occurring within the unfolded ensemble with slow kinetics will give these results. To further test if proline isomerization limits folding, we repeated the double jump experiments in the presence of Cyclophilin A (CycPA), a protein with peptidyl-proline isomerase activity. In this experiment, we again perform the double jump from 0M to 7.4M to 0.22M Gdn-HCl with time spent in 7.4M Gdn-HCl recorded as the aging time. However, the refolding reaction is performed by dilution of the 7.4M Gdn-HCl sfGFP solution into buffer containing CycPA, with a final concentration of 0.22M Gdn-HCl after a 600s aging time. The presence of CycPA increases the population folding via the fast (τ_2) route by 10% (Figure 6). The shifting of the phases from slow (τ_1) to fast (τ_2) over the maturation time, and the recovery of the fast phase in the presence of Cyclophilin A provides evidence of a proline isomerization causing multiple phases in folding and unfolding. Although CycPA doesn't give a full recovery of folding to the fast phase, refolding is being performed at 0.22M Gdn-HCl, which lowers the activity of CycPA by destabilization.

Probing the Center of the Chevron

Interestingly, the guanidine-HCl concentrations at which no detectable kinetics in the unfolding and refolding reaction are observed (see Figure 4) corresponds to the region of difference in the unfolding and refolding curves presented in Figure 3a. In an effort to test if the inability to measure kinetics in the center of the chevron is a result of a change in mechanism or rate determining step in the unfolding and refolding reactions, we performed a series of kinetic experiments. These experiments are similar to double jump experiments with an infinite aging time. In these experiments we varied the initial conditions by equilibrating samples within the transition regions of the denaturation (D) and renaturation (R) curves (Figure 7-and Methods and Materials) prior to kinetic analysis. We then monitored the kinetics of a jump from the initial conditions within the transition zone to final conditions within the hysteresis zone (Figure 7).

If the protein follows two separate routes in unfolding and refolding, the expectation is that varying the initial conditions will allow us to probe the kinetics within the transition region. The observed rates will allow us to probe the guanidine dependence within the center of the chevron plot (Figure 4) and the endpoint of the reaction will track along the hysteresis curve backwards; i.e. the refolding transition will follow from the center of the curve, along the modeled line to a fully unfolded state within the hysteresis area. Conversely, the unfolding transition will follow from the center of the curve to the fully refolded state (Figure 7a). However, if the hysteresis is caused by trapping of the protein in a broad, flat barrier region of the folding landscape at intermediate denaturant concentrations, there will be no relaxation, as this represents a highly glassy state. The resultant fluorescence intensity will not change as we probe the center of the glassy hysteresis zone; i.e. the result will appear to be a minor hysteresis curve, as seen in magnetic systems³³ (Figure 7b). The results of our studies are plotted in Figure 7c as black filled circles and squares for refolding and unfolding reactions, respectively. The data are consistent and resemble a minor hysteresis curve, with no observable kinetics (Figure 7c). This data provides evidence of a glassy transition or trapped state in the folding landscape resulting in the observed hysteretic phenomenon.

Equilibrium Transitions Revisited

Interestingly, we observe significant shifts in the unfolding and refolding reactions between 24 and 96 hours, however, we see no change between 96 and 192 hours (Figure 2). However, the extrapolated kinetic results suggest that the equilibrium curves should coincide after five half-lives, or 2 months. In Figure 8, we present the results of monitoring the equilibrium reactions after 58 and 102 days. Importantly, the refolding reaction shows no change in the transition curve observed between 96 hours and 106 days. Conversely, while the unfolding transition curve shows no change between 96 and 192 hours, a shift in the transition curve to lower denaturant concentrations occurs after 58 and 102 days, similar to results seen by Sophie Jackson^{34; 35}. Despite the shift to lower denaturant concentrations, the observed transition curve retains three-state behavior and does not approach the observed refolding curve. If folding was just extremely slow, and insufficient equilibration times were the explanation for the observed hysteresis, the refolding curve would have also shown change over time. Our results indicate no change over time for refolding, and demonstrate that a separate process controls *un*folding at intermediate guanidine concentrations.

Equilibrium Transitions of sfGFP mutants R96A and Y74M/M88Y

Solution studies have shown that the GFP chromophore not only can undergo isomerization, but also is held in a strained configuration by the protein core that surrounds the central helix³⁶. GFP folds into a native barrel structure prior to chromophore formation, and the α -helix backbone kinking which accompanies chromophore formation requires that the protein be folded into the barrel structure. In an effort to understand if the chromophore of sfGFP contributed to the observed hysteresis, we wished to compare the folding of the sfGFP protein which had yet to undergo chromophore formation. Previously, the R96A GFP mutant was discovered to slow chromophore formation from minutes, to months³⁷. We mutated this residue in sfGFP and monitored the equilibrium unfolding and refolding of the protein by the change in tryptophan fluorescence as a function of added denaturant concentration (Figure 9a). The equilibrium unfolding and refolding transitions of the R96A sfGFP are superimposable, and show no evidence of hysteresis.

We have shown that like cycle3 GFP, proline isomerization slows the folding kinetics of sfGFP. Scanning mutagenesis in our laboratory revealed that mutation of the N-terminal residues in two X-Pro peptide sequences Y74P75 and M88P89 to Y74M/M88Y also hindered chromophore formation. The equilibrium refolding and unfolding transitions for this variant also removed the apparent hysteresis (Figure 9b). These data support the hypothesis that the

hysteresis seen in folding is coupled to the conformation of the novel chromophore. Interestingly, the cooperativity of the folding/unfolding transitions for the double mutant has decreased from that observed for R96A. These two mutations map to the helical cap of the barrel which is tightly pinned to the barrel in the sfGFP (see Discussion). Destabilization of cap/barrel interactions may alter the solvent accessible surface in the double mutant protein by facilitating solvent penetration into the core.

NMR Spectra of sfGFP and trapped sfGFP

We have shown that abolishing the chromophore from the structure of sfGFP abolishes hysteresis seen in folding. We acquired 1-D spectra of sfGFP and trapped sfGFP as described for Figure 7 (Figure 10). Upon dilution or desalting of guanidine from the 1.8M sample, fluorescence is maintained at the same level as that observed in the trapped species described in Figure 7 (data not shown).

One-dimensional ^1H spectra have similar chemical shift dispersion, showing that overall structure is the same between sfGFP in both native and “trapped” states (Figure 10a). However, upon close inspection, overlaid spectra show subtle changes in both the amide and aliphatic regions of the spectra. Changes around 0 ppm (Figure 10b) and 6.5 (Figure 10c) ppm are expanded for clarity. Subtle changes also appear at 8.2, 2.3, 7.4, 7.9, and 8.8 ppm.

DISCUSSION

Hysteresis and Protein Folding

Superfolder GFP was generated by directed evolution and demonstrates 100% fluorescence recovery in refolding experiments, a significant improvement for molecular reporting²¹. Interestingly, equilibrium unfolding and refolding experiments are non-coincident (Figure 2a), and at first glance, might suggest lack of complete reversibility in folding and/or aggregation. However, experiments with sfGFP protein which has been unfolded and refolded generate the same equilibrium and kinetic unfolding curves as protein which has never been unfolded (Figure 2b). In addition, there is no evidence of protein concentration dependence, as equilibrium curves were the same over a ten-fold protein concentration.

The non-coincidence of the sfGFP unfolding and refolding transitions shows apparent hysteresis. Hysteresis is a property in which a system does not immediately respond to the forces applied to it, and may arise from a bifurcation in the energy landscape. This, in turn, leads to a bistable system where the observed equilibrium is dependent not only on the final conditions, but also the initial (historical) conditions³³. Common manifestations of hysteresis are lag effects (not necessarily temporal) between an applied force and the resultant final (not necessarily equilibrium) state, bistability, as well as memory effects^{38; 39; 40}. Hysteresis was originally described in magnetic systems where an applied magnetic field will magnetize a material; once removed, the material will retain the magnetism. This property is exploited in magnetic storage, such as magnetic tapes and hard drives.

Hysteresis has been seen in protein folding, although it is typically limited to multimeric^{41; 42; 43; 44} and modular repeat proteins such as seen in the unfolding/refolding of titin^{45; 46}, a molecule composed of 21 repeat domains. Different activation energies for folding and unfolding between domains may cause hysteresis curves if unfolding is controlled by domain-by-domain transitions while refolding is a global process. Interestingly, denaturant-induced unfolding and refolding of trimeric type III collagen also shows hysteresis in which proline cis-trans isomerization may be a factor due to a proline rich region⁴⁷. A mechanism which explained the hysteresis observed in collagen folding has been proposed. Similar to titin, hysteresis is thought to be caused in type III collagen refolding due to slow annealing from

loop rearrangement, compared to very cooperative denaturation; unfolding occurs on a local scale, while refolding requires a global process⁴⁸. While hysteresis is rare in single domain, monomeric proteins, it has been observed in single molecule folding experiments where ensemble averaging is removed, as demonstrated in RNaseH⁴⁹, and GFP.

Single molecule studies of GFP unfolding shows the presence of intermediates in mechanical unfolding attributed to the detachment of single β -strands from the native structure⁵⁰. Further experiments on GFP show that when mechanically unfolded from linkages besides the N or C terminals, there are multiple pathways of unfolding, which share similar stability⁵¹. Multiple pathways are one result of hysteresis, although unfolding mechanisms described may be different under mechanical stress versus chemical denaturation.

Slow Folding & Unfolding Kinetics Do Not Explain Hysteresis in sfGFP

A simple explanation of the apparent hysteresis is that the slow folding and unfolding kinetics of sfGFP are not conducive to reaching folding equilibrium. Although equilibrium titrations appear to be static after 96 hours (Figure 2), predicted kinetics within the center of the chevron plot (Figure 4) estimate equilibrium to be reached after two months. Importantly the refolding reaction maintains two-state behavior and shows no change in the observed transition after 3 months (Figure 8). Conversely, the unfolding transitions shift over at least 3 months and maintains three-state behavior, consistent with the presence of a native-like intermediate (N*) seen in cycle3 GFP unfolding³⁵. If insufficient equilibration time were responsible for the observed hysteresis effect, we would have seen changes in both the unfolding and refolding transitions with time. However, the fact that only the *unfolding* transition continues to shift with time indicates that a different barrier controls the unfolding and refolding reactions. These barriers must also be large as recently described for the irreversibility in the assembly of the SNARE fusion machines^{52; 53}, and in GFP³⁵. In native-state hydrogen exchange studies fluctuations within the native state allow population of high energy species and labeling of interior residues,^{54; 55}. Similarly, fluctuations in the native state of sfGFP under destabilizing conditions allow for the transient population of fully unfolded protein, which can only return to a lower energy state via the landscape used for the refolding reaction.

Does Hysteresis in Protein Folding Arise From a Rough Energy Landscape?

Hysteresis can arise from a bifurcation in the energy landscape. This, in conjunction with large kinetic barriers, would constitute a rough energy landscape, as described in protein folding theory. Multiple energy wells (intermediates) are observed often in protein folding, so why is hysteresis such a rare observation?

Random heteropolymers sample many configurations with little bias and the resulting landscape is described as frustrated. As these heteropolymers could be easily trapped in glassy transitions, hysteresis may be seen. The ground state of a random heteropolymer arises from competition between conflicting energy contributions and the resulting landscape is described as frustrated. If proteins behaved like random heteropeptides, even a minor change in sequence could cause an alternate structure to become the new ground state. Thus, the energy landscape for a random sequence would be rugged and the dynamics glassy^{56; 57}. This type of landscape would not be conducive to reliable protein fold and function. However, a rough landscape may show hysteresis, as these heteropolymers could be easily trapped in glassy transitions.

Proteins have evolved to robustly fold so that spurious single-site mutations rarely affect the global fold. In this case, interactions are not in conflict as described above for the random heteropolymer, but are supportive and cooperatively lead to a low-energy structure. Thus, stable protein folds have mutually supportive, cooperative interactions in the native structure said to be “consistent” or “minimally frustrated”. Energy decreases, on average, as sampled

structures approach the native state and the energy landscape is “funneled”^{56; 57}. This funneling predicts simple kinetics for small single-domain proteins, in agreement with experimental results⁵⁸. Hysteresis is thus a rare occurrence. Therefore, does the novel global fold of GFP lead to the hysteresis observed in folding studies?

Structural Overview of GFP: A Bent Helix & A Twisted Chromophore

The crystal structure reveals a unique 11 stranded β -barrel surrounding a central α -helix, which contains the chromophore (Figure 11)⁵⁹. GFP must be folded into a barrel structure prior to *de novo* chromophore formation¹². The mechanism of chromophore formation in GFP has been proposed to be facilitated by mechanical stress induced by the β -barrel scaffold in the immature native protein. Tension which concentrates force on the central α -helix is proposed to create an energetically unfavorable “kink”. Subsequent chromophore formation could reduce the strain in the α -helix by accommodating the kink with new stabilizing bonds forming the chromophore⁶⁰. The central α -helix may be positioned in the central core of the protein by a hydrophobic patch of conserved amino acids Phe27, Leu60, and Leu125⁶¹. It is possible that a combination of hydrophobicity and mechanical stress is able to strain the helix enough to distort it, setting the stage for chromophore maturation through catalysis by E222 and R96.

Chromophore formation involves a novel, irreversible backbone rearrangement of the tripeptide SYG (residues 65–67) and a cyclization/dehydration and oxidation reaction, to create the para-hydroxybenzylideneimidazolinone chromophore (Figure 11, green)¹². The actual order of the reactions is still under debate^{37; 62; 63; 64}. Once the chromophore is formed, the protein matrix provides steric restraints, keeping it in its native, twisted state^{36; 65}. In simulated chromophore compounds free of the β -barrel though, the chromophore is planar, as expected in unfolded GFP^{65; 66}. This non-planarity as well as *cis-trans* isomerization seen in the chromophore of native GFP is also seen in crystal structures of kindling fluorescent proteins⁶⁷. Enoki et.al.¹⁶ report a “breathing” mechanism in the early folding kinetics of cycle3 GFP between the tryptophan and chromophore fluorescence that is consistent with the *cis-trans* proline isomerization observed in NMR studies of the native state⁶⁸. The breathing mechanism is described as the tryptophan residue being buried in a native-like structure, released, and then reburied again and quenched by the chromophore, possibly due to proline and chromophore isomerizations hindering folding. Our NMR results show a similar overall fold, with subtle changes (Figure 10), possibly related to isomerizations in prolyl-peptide bonds and the chromophore.

Hysteresis is Abolished in the Absence of Chromophore Formation

The structure of GFP has two characteristics that may lead to a rough landscape. First, the chromophore is slightly twisted in the native form, whereas it is planar in the unfolded form (see above). Second, the central helix shows many deviations from optimal geometry. Our results indicate that the sequence variant, R96A, has impaired chromophore formation and no hysteresis (Figure 9a). In addition, we found through scanning mutagenesis that mutation of residues N-terminal to the prolines in the central helix (Figure 11), not only abolished efficient chromophore formation, but also removed the hysteretic effect (Figure 9b). Interestingly, in X-Pro dipeptides, the X residue can alter the *cis-trans* ratio of the X-Pro bond in peptides and unfolded states of proteins⁶⁹. Thus, these results suggest that these two prolines may also be responsible for the proline isomerization limitation observed in the kinetics of folding of GFP. Prior to this study, only the chemical steps in the maturation of the chromophore had been identified. Our results may pinpoint structural requirements on the helix capping regions for efficient chromophore formation.

Inspection of the crystal structure leads us to speculate that the observed alteration in maturation in the Y74M/M88Y sfGFP reflects a synergy between proline isomerization state at the ends

of the helix and efficient chromophore formation. Taken together, our results indicate that the hysteretic effect in sfGFP is linked to the formation of the chromophore, and removal of this feature generates a less-frustrated energy landscape for this protein.

The Functional Energy Landscape of sfGFP: Frustration Introduced by the Chromophore

The funneled organization of the energy landscape dominates the kinetics of folding for highly evolved and/or well designed proteins^{70; 71}. Proteins, however, *do not* have one landscape for folding and a different one for function. There is some speculation that functional residues are a hindrance to folding^{72; 73; 74; 75}. Interestingly, our recent studies on the β -trefoil protein interleukin-1 β indicate that the residues within and contacting the functionally significant β -bulge may introduce “frustration” into the folding of this protein^{76; 77}. Therefore “frustration” associated with conserving specific functional sites may alter not only the folding trajectories, but the transitions observed in native proteins. Thus it appears that the native basin of the energy landscape dictates structural changes related to function^{78; 79}.

The results presented herein indicate that the energy landscape of GFP is particularly rough, leading to hysteresis. However, the landscape can be made smoother by hindering chromophore formation. Other post-translational modifications such as glycosylation⁸⁰, deamidation⁸¹, and phosphorylation⁸² also appear to change the landscape and native basins of these proteins. Therefore, unlike the case observed for random heteropolymers, a specific structural feature of GFP leads to frustration in the energy landscape. More precisely, a cyclization in the polypeptide chain frustrates the landscape, as seen in simulation of small polypeptides⁸³. The *de novo* folding pathway of GFP proceeds *prior to* chromophore maturation while the *refolding* of GFP proceeds in the presence of the already cyclized chromophore. As chromophore formation requires the folding of the barrel, the *de novo* folding pathway should lack hysteretic behavior and have the smooth energy landscape more typical of 11-stranded β -barrel protein folds.. Consistent with this hypothesis, the apparent hysteresis is ameliorated in the presence of mutations that impede chromophore formation.

METHODS & MATERIALS

sfGFP Expression & Purification

Recombinant superfolder GFP (sfGFP) was subcloned into the kanamycin resistant pET28(a+) vector and the resultant plasmid transformed into BL21 (DE3) *Escherichia coli* cells for expression. Mutations were introduced using primer-directed mutagenesis (R96A and Y74M/M88Y) into sfGFP. Expression and purification were performed similarly to previous methods²¹, with some changes. Cells were grown in Luria Broth at 30°C to an OD₆₀₀ of 0.6, and protein expression was induced with 1mM IPTG for 5 hours at 25°C. Cell pellets were re-suspended in buffer (pH 7.9) containing 0.5M NaCl, 20mM Tris-HCl, 5mM imidazole, 0.1mM EDTA, 1mM DTT, 10mg/mL PMSF and sonicated on ice (Sonic Dismembrator 550, Fischer Scientific) with 30 s pulses separated by 30 s breaks for a total of 20 minutes.

The lysate was cleared by centrifugation at 14,000g for 30 minutes. The soluble sfGFP was bound and eluted from a Ni-NTA resin (Novagen) according to manufacturer’s protocol. sfGFP was purified further on a Sephacryl S-40(Pharmacia Biotech) size-exclusion column to remove His-rich proline isomerase in 25mM NaPO₄ and 100mM NaCl at pH 6.5. Fractions containing sfGFP were analyzed by SDS-PAGE for purity and the purified protein was collected and concentrated and stored in 25mM NaPO₄ and 100mM NaCl at pH 6.5. Yield was determined by a theoretical extinction coefficient ($\epsilon=31519 \text{ M}^{-1}\text{cm}^{-1}$).

Stability Measurements: Guanidine-Induced Equilibrium Unfolding & Refolding: Fluorescence & CD Spectroscopy

Chromophore fluorescence spectra were obtained using the Fluoromax-2 spectrofluorimeter (Spex Edison, NJ) with an excitation wavelength of 450 nm and the emission measured at 508 nm unless otherwise noted. Circular Dichroism data were collected on the 60-DS spectrophotometer (Aviv Instruments). Equilibrium unfolding experiments were prepared with protein samples (0.1 mg/mL in 12mM Tris-HCl at pH 6.8 and varying Gdn-HCl concentrations ranging from 0 to 7.5M, at a temperature of 22 °C. Experiments were equilibrated for 96 hours, unless noted otherwise. Refolding curves were prepared by equilibrating sfGFP in 7.6 M Gdn-HCl and subsequently refolded by dilution to varying final Gdn-HCl concentrations and equilibrated for 96 hours. Unfolding transition curves using previously refolded sfGFP were prepared by equilibration in 6.4M Gdn-HCl, subsequent refolding to 1M Gdn-HCl overnight (20 hours, approximately 1000 half-lives), concentration of the refolded protein, and subsequent dilution to varying final Gdn-HCl concentrations with a protein concentration of 0.1 mg/mL sfGFP. The fraction folded by chromophore fluorescence was monitored at 508nm, after excitation at 450nm (3nm slit width). The fraction folded using tryptophan fluorescence was determined from the average wavelength (equation 1), after excitation at 295nm (5nm slit width)

$$\frac{\sum_{\lambda=311-420nm} \lambda_i \times I_{\lambda_i}}{\sum_{\lambda=311-420nm} I_{\lambda_i}} \quad (1)$$

Manual Mixing sfGFP Unfolding and Refolding Kinetics

Measurement of the unfolding relaxation times, τ_u , which is dependent on the final Gdn-HCl concentration were performed by manual-mixing fluorescence spectroscopy, using a Fluoromax-2 spectrofluorimeter (Spex), with the excitation wavelength set to 450 nm and the emission intensity monitored at 508 nm. The proteins were unfolded over a range of denaturant concentrations between 5.0 and 8.0M in 12.5mM Tris-HCl at pH 6.8, 25°C. The time-dependent decay in the observed fluorescent signal was adequately fit by a double exponential equation. The dead-time for the manual mixing experiments was ~10 sec. For refolding experiments, the protein was unfolded in 8M Gdn-HCl at a protein concentration of 2.7 mg/ml and allowed to equilibrate. Refolding was performed by dilution of the unfolded protein with the appropriate amount of buffer and Gdn-HCl to obtain the desired final Gdn-HCl concentrations of 0.2–1.2M. The final protein concentration in refolding experiments was 0.1mg/ml unless stated otherwise.

Interrupted Refolding/Double Jump Kinetics by Manual Mixing

The double jump kinetics were performed by rapidly mixing a concentrated stock protein solution (33 mgs/ml) with Gdn HCl solution to a final denaturant concentration of 7.4 M using manual mixing techniques under conditions previously described for kinetic analysis. After a variable delay (aging) time, 30 μ l of the protein/7.4 M GdnHCl solution was injected into a cuvette containing 970 μ l of buffer solution, thermostated at 25°C. The solution was rapidly mixed and the refolding of the protein monitored as described above. The final guanidine concentration during refolding was 0.22 M.

Stopped-Flow Unfolding & Refolding Kinetics

Stopped-flow fluorescence studies were performed with an Applied Photophysics Pi*-180 stopped flow instrument (Applied Photophysics, London). Data were collected in logarithmic mode with an excitation wavelength of 450nm and emission collected using a 500 nm cut off filter with a dead time of 5ms. Typical refolding spectra were averaged over a minimum of five runs. Refolding experiments were initiated by a 1:11 dilution of 1.1mg/mL unfolded sfGFP (equilibrated in 7.73 M Gdn-HCl in buffer containing 12.5 mM Tris (pH=6.8)) into varying final concentrations of Gdn-HCl. The final concentration of sfGFP was held constant at 0.1 mg/mL, unless otherwise noted in the manuscript. Unfolding experiments were initiated by a 1:11 dilution of 1.1 mg/mL sfGFP with 12.5 mM Tris (pH=6.8) into varying final concentrations of Gdn-HCl and 12.5mM Tris (pH=6.8).

Kinetic Analysis of the Hysteresis Zone

Kinetic studies were performed as described in the manual mixing section of this manuscript. However, the initial conditions for the individual kinetic experiments correspond to the protein equilibrated at conditions of 4.5 M (unfolding) and 1.8M (refolding) Gdn-HCl, respectively. The initial samples were prepared as described under the stability measurements section. We chose the initial conditions to correspond to points on the unfolding and refolding equilibrium transition.

Data Fitting

The kinetic data were fit to a series of exponentials:

$$A(t) = \sum A_i \exp(-t/\tau_i) + A \quad (2)$$

Where $A(t)$ is the optical signal at time t , A is the optical signal at infinite time, A_i is the signal change associated with each phase (i), and τ_i is the relaxation time for each individual phase. By combining the stopped-flow and manual-mixing traces the data was globally fit using in-house software. The number and magnitude of the exponential phases was determined as previously described²⁴.

Equilibrium Transition Curve Fitting

All equilibrium data was fit to either a two-state or three-state model. Both models assume a linear dependence of ΔG_{unf} on denaturant concentration (m). The two-state model is:

$$S = \frac{S_N + S_U \times K_{N \leftrightarrow U}}{1 + K_{N \leftrightarrow U}} \quad (3)$$

Where

$$K_{N \leftrightarrow U} = \exp\left(\frac{-\Delta G_{N \leftrightarrow U}^{H_2O} + (m_{N \leftrightarrow U} \times [D])}{RT}\right) \quad (4)$$

Where $[D]$ is the guanidine concentration, m is the linear dependence of $\Delta G_{N \leftrightarrow U}$ on denaturant concentration, and $\Delta G_{N \leftrightarrow U}^{H_2O}$ is the free energy of unfolding at 0M denaturant, and S_N and S_U are the signal of the native and unfolded states, respectively.

The three-state model is similar to the two-state model:

$$S = \frac{K_{N \leftrightarrow I}(Z + K_{I \leftrightarrow U})}{1 + K_{N \leftrightarrow I}(1 + K_{I \leftrightarrow U})} \quad (5)$$

Where $K_{N \leftrightarrow I}$ and $K_{I \leftrightarrow U}$ are as described above, and Z is a signal reweighting for the intermediate signal with the native and denatured signal:

$$Z = \frac{S_I - S_N}{S_U - S_N} \quad (6)$$

Data was fit through a non-linear least squares minimization using in-house scripts and the curve-fitting tool in Matlab.

NMR Experiments

All NMR experiments were performed at 315 °K on a Bruker Avance800 equipped with a triple-resonance triple-axis gradient probe. Protein samples were prepared as described above, with the addition of an additional quick-spin column step, to remove additional guanidine from a trapped species. Trapped species were prepared as described above by unfolding sfGFP at 6M guanidine-HCl for 10 minutes, and then jumped to 1.8 M guanidine-HCl for 96 hours. The trapped species corresponds to as sample in the transition of the refolding curve, shown as black boxes in Figure 7. Final conditions for each sample were 30mM phosphate buffer at pH 6.8, and 100mM NaCl, 95% H₂O/5% D₂O. Data were processed using Felix 2004 software (Accelrys, San Diego, CA). One-dimensional ¹H spectra were acquired for sfGFP, R96A sfGFP, and trapped sfGFP (described above). Spectra were acquired with a 20 ppm spectral width using a 1-D with Watergate suppression^{84; 85}.

Acknowledgments

We would like to thank the members of the laboratory for discussions of the work, including Melinda Roy for thought provoking and insightful discussion. Work was funded by the Molecular Biophysics Training Grant (GM 08326) (BTA), LANL (GW/PAJ) and NIH grants (GM 54038 and DK 54441)(PAJ).

Abbreviations used

GFP	green fluorescent protein
sfGFP	superfolder green fluorescent protein
EDTA	ethylenediaminetetraacetic acid
DTT	dithiothreitol
PMSF	Phenylmethylsulphonylfluoride
Gdn-HCl	guanidinium hydrochloride
CD	

circular dichroism

YFP

yellow fluorescent protein

CycPA

Cyclophilin A

References

1. Tsien RY. The green fluorescent protein. *Annu Rev Biochem* 1998;67:509–44. [PubMed: 9759496]
2. Waldo GS. Improving protein folding efficiency by directed evolution using the GFP folding reporter. *Methods Mol Biol* 2003;230:343–59. [PubMed: 12824594]
3. Waldo GS, Standish BM, Berendzen J, Terwilliger TC. Rapid protein-folding assay using green fluorescent protein. *Nat Biotechnol* 1999;17:691–5. [PubMed: 10404163]
4. Cabantous S, Pedelacq JD, Mark BL, Naranjo C, Terwilliger TC, Waldo GS. Recent advances in GFP folding reporter and split-GFP solubility reporter technologies. Application to improving the folding and solubility of recalcitrant proteins from *Mycobacterium tuberculosis*. *J Struct Funct Genomics* 2005;6:113–9. [PubMed: 16211507]
5. Cabantous S, Terwilliger TC, Waldo GS. Protein tagging and detection with engineered self-assembling fragments of green fluorescent protein. *Nat Biotechnol* 2005;23:102–7. [PubMed: 15580262]
6. Wilson CG, Magliery TJ, Regan L. Detecting protein-protein interactions with GFP-fragment reassembly. *Nat Methods* 2004;1:255–62. [PubMed: 16145770]
7. Magliery TJ, Wilson CG, Pan W, Mishler D, Ghosh I, Hamilton AD, Regan L. Detecting protein-protein interactions with a green fluorescent protein fragment reassembly trap: scope and mechanism. *J Am Chem Soc* 2005;127:146–57. [PubMed: 15631464]
8. Subramanian S, Srienc F. Quantitative analysis of transient gene expression in mammalian cells using the green fluorescent protein. *J Biotechnol* 1996;49:137–51. [PubMed: 8879169]
9. Blokpoel MC, O'Toole R, Smeulders MJ, Williams HD. Development and application of unstable GFP variants to kinetic studies of mycobacterial gene expression. *J Microbiol Methods* 2003;54:203–11. [PubMed: 12782376]
10. Kalir S, McClure J, Pabbaraju K, Southward C, Ronen M, Leibler S, Surette MG, Alon U. Ordering genes in a flagella pathway by analysis of expression kinetics from living bacteria. *Science* 2001;292:2080–3. [PubMed: 11408658]
11. Lu C, Bentley WE, Rao G. A high-throughput approach to promoter study using green fluorescent protein. *Biotechnol Prog* 2004;20:1634–40. [PubMed: 15575693]
12. Reid BG, Flynn GC. Chromophore formation in green fluorescent protein. *Biochemistry* 1997;36:6786–91. [PubMed: 9184161]
13. Fukuda H, Arai M, Kuwajima K. Folding of green fluorescent protein and the cycle3 mutant. *Biochemistry* 2000;39:12025–32. [PubMed: 11009617]
14. Crameri A, Whitehorn EA, Tate E, Stemmer WP. Improved green fluorescent protein by molecular evolution using DNA shuffling. *Nat Biotechnol* 1996;14:315–9. [PubMed: 9630892]
15. Cormack BP, Valdivia RH, Falkow S. FACS-optimized mutants of the green fluorescent protein (GFP). *Gene* 1996;173:33–8. [PubMed: 8707053]
16. Enoki S, Saeki K, Maki K, Kuwajima K. Acid denaturation and refolding of green fluorescent protein. *Biochemistry* 2004;43:14238–48. [PubMed: 15518574]
17. Campanini B, Bologna S, Cannone F, Chirico G, Mozzarelli A, Bettati S. Unfolding of Green Fluorescent Protein mut2 in wet nanoporous silica gels. *Protein Sci* 2005;14:1125–33. [PubMed: 15802645]
18. Cannone F, Bologna S, Campanini B, Diaspro A, Bettati S, Mozzarelli A, Chirico G. Tracking unfolding and refolding of single GFPmut2 molecules. *Biophys J* 2005;89:2033–45. [PubMed: 15994904]

19. Waldo GS. Genetic screens and directed evolution for protein solubility. *Curr Opin Chem Biol* 2003;7:33–8. [PubMed: 12547424]
20. Pedelacq JD, Piltch E, Liong EC, Berendzen J, Kim CY, Rho BS, Park MS, Terwilliger TC, Waldo GS. Engineering soluble proteins for structural genomics. *Nat Biotechnol* 2002;20:927–32. [PubMed: 12205510]
21. Pedelacq JD, Cabantous S, Tran T, Terwilliger TC, Waldo GS. Engineering and characterization of a superfolder green fluorescent protein. *Nat Biotechnol* 2006;24:79–88. [PubMed: 16369541]
22. Chalfie M, Tu Y, Euskirchen G, Ward WW, Prasher DC. Green fluorescent protein as a marker for gene expression. *Science* 1994;263:802–5. [PubMed: 8303295]
23. Lakowicz, JR. *Principles of Fluorescence Spectroscopy*. Vol. 2. Kluwer Academic, Plenum Publishers; New York: 1999.
24. Matthews CR. Effect of point mutations on the folding of globular proteins. *Methods Enzymol* 1987;154:498–511. [PubMed: 3431461]
25. Heidary DK, Gross LA, Roy M, Jennings PA. Evidence for an obligatory intermediate in the folding of interleukin-1 beta. *Nat Struct Biol* 1997;4:725–31. [PubMed: 9303000]
26. Heidary DK, O'Neill JC Jr, Roy M, Jennings PA. An essential intermediate in the folding of dihydrofolate reductase. *Proc Natl Acad Sci U S A* 2000;97:5866–70. [PubMed: 10811909]
27. Enoki S, Maki K, Inobe T, Takahashi K, Kamagata K, Oroguchi T, Nakatani H, Tomoyori K, Kuwajima K. The equilibrium unfolding intermediate observed at pH 4 and its relationship with the kinetic folding intermediates in green fluorescent protein. *J Mol Biol* 2006;361:969–82. [PubMed: 16889795]
28. Kamagata K, Arai M, Kuwajima K. Unification of the folding mechanisms of non-two-state and two-state proteins. *J Mol Biol* 2004;339:951–65. [PubMed: 15165862]
29. Brandts JF, Halvorson HR, Brennan M. Consideration of the Possibility that the slow step in protein denaturation reactions is due to cis-trans isomerism of proline residues. *Biochemistry* 1975;14:4953–63. [PubMed: 241393]
30. Krieger F, Moglich A, Kiefhaber T. Effect of proline and glycine residues on dynamics and barriers of loop formation in polypeptide chains. *J Am Chem Soc* 2005;127:3346–52. [PubMed: 15755151]
31. Kiefhaber T, Quaa R, Hahn U, Schmid FX. Folding of ribonuclease T1. 1. Existence of multiple unfolded states created by proline isomerization. *Biochemistry* 1990;29:3053–61. [PubMed: 2110823]
32. Schultz DA, Schmid FX, Baldwin RL. Cis proline mutants of ribonuclease A. II. Elimination of the slow-folding forms by mutation. *Protein Sci* 1992;1:917–24. [PubMed: 1304376]
33. Ewing JA. On the Production of Transient Electric Currents in Iron and Steel Conductors by Twisting Them When Magnetised, or by Magnetising Them When Twisted. *Proc Royal Soc Lon* 1883;36:117–135.
34. Jackson SE, Craggs TD, Huang JR. Understanding the folding of GFP using biophysical techniques. *Expert Rev Proteomics* 2006;3:545–59. [PubMed: 17078767]
35. Huang JR, Craggs TD, Christodoulou J, Jackson SE. Stable Intermediate States and High Energy Barriers in the Unfolding of GFP. *J Mol Biol* 2007;370:356–71. [PubMed: 17512539]
36. Chen MC, Lambert CR, Urgitis JD, Zimmer M. Photoisomerization of green fluorescent protein and the dimensions of the chromophore cavity. *Chemical Physics* 2001;270:157–164.
37. Barondeau DP, Kassmann CJ, Tainer JA, Getzoff ED. Understanding GFP chromophore biosynthesis: controlling backbone cyclization and modifying post-translational chemistry. *Biochemistry* 2005;44:1960–70. [PubMed: 15697221]
38. Bertotti G. Energetic and thermodynamic aspects of hysteresis. *Physical Review Letters* 1996;76:1739–1742. [PubMed: 10060505]
39. Bertotti, G.; Mayergoyz, ID. *The Science of Hysteresis*. Vol. 1. Vol. 3. Academic Press; Boston: 2006.
40. Bertotti G, Mayergoyz ID, Basso V, Magni A. Functional integration approach to hysteresis. *Phys Rev E Stat Phys Plasmas Fluids Relat Interdiscip Topics* 1999;60:1428–40. [PubMed: 11969902]

41. Lai Z, McCulloch J, Lashuel HA, Kelly JW. Guanidine hydrochloride-induced denaturation and refolding of transthyretin exhibits a marked hysteresis: equilibria with high kinetic barriers. *Biochemistry* 1997;36:10230–9. [PubMed: 9254621]
42. Rumfeldt JA, Stathopoulos PB, Chakrabarty A, Lepock JR, Meiering EM. Mechanism and thermodynamics of guanidinium chloride-induced denaturation of ALS-associated mutant Cu,Zn superoxide dismutases. *J Mol Biol* 2006;355:106–23. [PubMed: 16307756]
43. Clark AC, Sinclair JF, Baldwin TO. Folding of bacterial luciferase involves a non-native heterodimeric intermediate in equilibrium with the native enzyme and the unfolded subunits. *J Biol Chem* 1993;268:10773–9. [PubMed: 8496144]
44. Benitez-Cardoza CG, Rojo-Dominguez A, Hernandez-Arana A. Temperature-induced denaturation and renaturation of triosephosphate isomerase from *Saccharomyces cerevisiae*: evidence of dimerization coupled to refolding of the thermally unfolded protein. *Biochemistry* 2001;40:9049–58. [PubMed: 11467968]
45. Minajeva A, Kulke M, Fernandez JM, Linke WA. Unfolding of titin domains explains the viscoelastic behavior of skeletal myofibrils. *Biophys J* 2001;80:1442–51. [PubMed: 11222304]
46. Kellermayer MS, Smith SB, Bustamante C, Granzier HL. Complete unfolding of the titin molecule under external force. *J Struct Biol* 1998;122:197–205. [PubMed: 9724621]
47. Davis JM, Bachinger HP. Hysteresis in the triple helix-coil transition of type III collagen. *J Biol Chem* 1993;268:25965–72. [PubMed: 8245030]
48. Engel J, Bachinger HP. Cooperative equilibrium transitions coupled with a slow annealing step explain the sharpness and hysteresis of collagen folding. *Matrix Biol* 2000;19:235–44. [PubMed: 10936448]
49. Cecconi C, Shank EA, Bustamante C, Marqusee S. Direct observation of the three-state folding of a single protein molecule. *Science* 2005;309:2057–2060. [PubMed: 16179479]
50. Dietz H, Rief M. Exploring the energy landscape of GFP by single-molecule mechanical experiments. *Proc Natl Acad Sci U S A* 2004;101:16192–7. [PubMed: 15531635]
51. Dietz H, Berkemeier F, Bertz M, Rief M. Anisotropic deformation response of single protein molecules. *Proc Natl Acad Sci U S A*. 2006
52. Fasshauer D, Antonin W, Subramaniam V, Jahn R. SNARE assembly and disassembly exhibit a pronounced hysteresis. *Nat Struct Biol* 2002;9:144–51. [PubMed: 11786917]
53. Barrick D, Hughson FM. Irreversible assembly of membrane fusion machines. *Nat Struct Biol* 2002;9:78–80. [PubMed: 11813007]
54. Chamberlain AK, Handel TM, Marqusee S. Detection of rare partially folded molecules in equilibrium with the native conformation of RNaseH. *Nat Struct Biol* 1996;3:782–7. [PubMed: 8784352]
55. Englander SW. Native-state HX. *Trends Biochem Sci* 1998;23:378. [PubMed: 9810223]author reply 379–81
56. Onuchic JN, Luthey-Schulten Z, Wolynes PG. Theory of protein folding: the energy landscape perspective. *Annu Rev Phys Chem* 1997;48:545–600. [PubMed: 9348663]
57. Bryngelson JD, Onuchic JN, Succi ND, Wolynes PG. Funnels, pathways, and the energy landscape of protein folding: a synthesis. *Proteins* 1995;21:167–95. [PubMed: 7784423]
58. Oliveberg M, Wolynes PG. The experimental survey of protein-folding energy landscapes. *Quarterly Reviews of Biophysics* 2005;38:245–288. [PubMed: 16780604]
59. Ormo M, Cubitt AB, Kallio K, Gross LA, Tsien RY, Remington SJ. Crystal structure of the *Aequorea victoria* green fluorescent protein. *Science* 1996;273:1392–5. [PubMed: 8703075]
60. Baronneau DP, Putnam CD, Kassmann CJ, Tainer JA, Getzoff ED. Mechanism and energetics of green fluorescent protein chromophore synthesis revealed by trapped intermediate structures. *Proc Natl Acad Sci U S A* 2003;100:12111–6. [PubMed: 14523232]
61. Seifert MH, Georgescu J, Ksiazek D, Smialowski P, Rehm T, Steipe B, Holak TA. Backbone dynamics of green fluorescent protein and the effect of histidine 148 substitution. *Biochemistry* 2003;42:2500–12. [PubMed: 12614144]
62. Cubitt AB, Heim R, Adams SR, Boyd AE, Gross LA, Tsien RY. Understanding, Improving and Using Green Fluorescent Proteins. *Trends in Biochemical Sciences* 1995;20:448–455. [PubMed: 8578587]

63. Rosenow MA, Huffman HA, Phail ME, Wachter RM. The crystal structure of the Y66L variant of green fluorescent protein supports a cyclization-oxidation-dehydration mechanism for chromophore maturation. *Biochemistry* 2004;43:4464–4472. [PubMed: 15078092]
64. Siegbahn PEM, Wirstam M, Zimmer M. Theoretical study of the mechanism of peptide ring formation in green fluorescent protein. *International Journal of Quantum Chemistry* 2000;81:169–186.
65. Maddalo SL, Zimmer M. The role of the protein matrix in green fluorescent protein fluorescence. *Photochem Photobiol* 2006;82:367–72. [PubMed: 16613487]
66. Patnaik SS, Trohalaki S, Pachter R. Molecular modeling of green fluorescent protein: structural effects of chromophore deprotonation. *Biopolymers* 2004;75:441–52. [PubMed: 15497152]
67. Henderson JN, Ai HW, Campbell RE, Remington SJ. Structural basis for reversible photobleaching of a green fluorescent protein homologue. *Proc Natl Acad Sci U S A* 2007;104:6672–7. [PubMed: 17420458]
68. Seifert MH, Ksiazek D, Azim MK, Smialowski P, Budisa N, Holak TA. Slow exchange in the chromophore of a green fluorescent protein variant. *J Am Chem Soc* 2002;124:7932–42. [PubMed: 12095337]
69. Reimer U, Scherer G, Drewello M, Kruber S, Schutkowski M, Fischer G. Side-chain effects on peptidyl-prolyl cis/trans isomerisation. *J Mol Biol* 1998;279:449–60. [PubMed: 9642049]
70. Leopold PE, Montal M, Onuchic JN. Protein Folding Funnels - a Kinetic Approach to the Sequence Structure Relationship. *Proceedings of the National Academy of Sciences of the United States of America* 1992;89:8721–8725. [PubMed: 1528885]
71. Onuchic JN, LutheySchulten Z, Wolynes PG. Theory of protein folding: The energy landscape perspective. *Annual Review of Physical Chemistry* 1997;48:545–600.
72. Kubelka J, Hofrichter J, Eaton WA. The protein folding ‘speed limit’. *Current Opinion in Structural Biology* 2004;14:76–88. [PubMed: 15102453]
73. Schreiber C, Buckle AM, Fersht AR. Stability and Function - 2 Constraints in the Evolution of Barstar and Other Proteins. *Structure* 1994;2:945–951. [PubMed: 7866746]
74. Shoichet BK, Baase WA, Kuroki R, Matthews BW. A Relationship between Protein Stability and Protein Function. *Proceedings of the National Academy of Sciences of the United States of America* 1995;92:452–456. [PubMed: 7831309]
75. Yang WY, Gruebele M. Folding at the speed limit. *Nature* 2003;423:193–197. [PubMed: 12736690]
76. Chavez LL, Gosavi S, Jennings PA, Onuchic JN. Multiple routes lead to the native state in the energy landscape of the beta-trefoil family. *Proceedings of the National Academy of Sciences of the United States of America* 2006;103:10254–10258. [PubMed: 16801558]
77. Gosavi S, Chavez LL, Jennings PA, Onuchic JN. Topological frustration and the folding of interleukin-1 beta. *Journal of Molecular Biology* 2006;357:986–996. [PubMed: 16469330]
78. Miyashita O, Onuchic JN, Wolynes PG. Nonlinear elasticity, proteinquakes, and the energy landscapes of functional transitions in proteins. *Proc Natl Acad Sci U S A* 2003;100:12570–5. [PubMed: 14566052]
79. Miyashita O, Wolynes PG, Onuchic JN. Simple energy landscape model for the kinetics of functional transitions in proteins. *J Phys Chem B* 2005;109:1959–69. [PubMed: 16851180]
80. Petrescu AJ, Wormald MR, Dwek RA. Structural aspects of glycomes with a focus on N-glycosylation and glycoprotein folding. *Curr Opin Struct Biol* 2006;16:600–7. [PubMed: 16938454]
81. Orru S, Vitagliano L, Esposito L, Mazzarella L, Marino G, Ruoppolo M. Effect of deamidation on folding of ribonuclease A. *Protein Sci* 2000;9:2577–82. [PubMed: 11206080]
82. Shen T, Zong C, Hamelberg D, McCammon JA, Wolynes PG. The folding energy landscape and phosphorylation: modeling the conformational switch of the NFAT regulatory domain. *Faseb J* 2005;19:1389–95. [PubMed: 16126906]
83. Levy Y, Becker OM. Energy landscapes of conformationally constrained peptides. *Journal of Chemical Physics* 2001;114:993–1009.
84. Piotto M, Saudek V, Sklenar V. Gradient-Tailored Excitation for Single-Quantum Nmr-Spectroscopy of Aqueous-Solutions. *Journal of Biomolecular Nmr* 1992;2:661–665. [PubMed: 1490109]

85. Sklenar V, Piotto M, Leppik R, Saudek V. Gradient-Tailored Water Suppression for H-1-N-15 Hsqc Experiments Optimized to Retain Full Sensitivity. *Journal of Magnetic Resonance Series A* 1993;102:241–245.

Figure 1. Native and Denatured Spectra of sfGFP

Red traces show native protein, blue traces show completely unfolded protein at 7.6M guanidine after 96 hours. (A) Chromophore fluorescence spectra excited at 450nm shows a complete loss of signal upon unfolding. (B) Tryptophan fluorescence spectra excited at 295 nm shows a red peak shift upon unfolding. (C) Circular Dichroism spectra is characteristic of a random coil upon unfolding. All samples were in 12.5 mM Tris-HCl, pH 6.8 and 25°C.

Figure 2. Guanidine Induced Equilibrium Unfolding Transitions of sfGFP at Multiple Times at pH 6.8 in 12.5mM Tris-HCl at 25°C after 24 hours (circles), 96 hours (squares), and 192 hours (diamonds). Lines show fits to guide the eye. After 96 hours, unfolding reactions appear to reach an apparent equilibrium.

Figure 3. Guanidine Induced Equilibrium Unfolding & Refolding Transition of sfGFP

(A) at pH 6.8 in 12.5mM Tris-HCl at 25°C after 96 hours. Folded state was measured by chromophore fluorescence at 508nm and excited at 450nm (slit width 3nm). Filled squares show unfolding transition, filled circles show refolding transition, and filled triangles show unfolding transition of sfGFP which has been previously unfolded and refolded. Red line shows the refolding transition fits to a two-state model. The unfolding transition fits to a three-state model (darker blue), as the light blue dotted line shows two-state fit to unfolding transition, which doesn't capture the difference in cooperativity between the beginning and end of the transition. The non-coincidence of these curves shows an apparent hysteresis. (B) Unfolding kinetic jump following chromophore fluorescence from 0.94M Guanidine HCl to 7.0 M in

12.5mM Tris-HCl, pH 6.8 and 25°C of both folded (blue) and refolded (red) sfGFP. There is no overall stability change between folded sfGFP, or that which has been through an unfolding-refolding cycle.

Figure 4. Superfolding GFP Chevron

at pH 6.8 in 20mM Tris-HCl at 25°C. Folded state was measured by fluorescence at 508nm and excited at 450nm (slit width 3nm). sfGFP folding and refolding residence time dependence on [Gdn-HCl]. Red lines and points show refolding, blue lines and points show unfolding. Circles, squares and diamonds show τ_1 , τ_2 , and τ_3 respectively. Lines are fit to focus on trends. The lag phase is shown in black, with filled triangles because it was calculated from the persistence of the lag phase, opposed to directly fit data from a kinetic model. Multiple phases are evidence of intermediates, and phases with low denaturant dependence (unfolding τ_2 and τ_3) may be linked with proline isomerization.

Figure 5. Refolding Kinetics
in 12.5 mM Tris-HCl at 25°C. Folded state was measured by chromophore fluorescence above 500nm, and excited at 450nm. Inset shows the first 400 ms; a lag phases in chromophore fluorescence is evident. This is evidence of an on-pathway, obligate intermediate.

Figure 6. Double Jump Kinetics

in 12.5mM Tris-HCl at pH 6.8 at 25°C. Folded state was measured by chromophore fluorescence at 508nm and excited at 450nm (slit width 3nm). sfGFP was unfolded for reported maturation time, and % amplitudes of the slow phase (square) and fast phases (circle) from a two-state fit was determined (blue). Fit phases are shown as trendlines (blue). CyclophilinA was added to the refolding buffer (0.038 mg/mL in final concentration) to determine if the slow phase was contributed from a proline isomerization (red). An amplitude shift from a fast phase to a slow phase is consistent with a proline isomerization. A reaction staying in the fast phase during the same experiment in the presence of cyclophilinA is strong evidence that multiple phases are caused by proline isomerization.

Figure 7. Probing Hysteresis

in 12.5 mM Tris-HCl at pH 6.8 at 25°C. Initial conditions were at black squares, and results could either show (a) multiple processes causing hysteresis, leading to multiple chevrons or (b) a highly glassy transition causing hysteresis, leading to intermediate hysteresis curves. Our results (c) are consistent with a highly glassy transition.

Figure 8. Long-time Guanidine Induced Equilibrium Folding and Refolding Transitions at pH 6.8 in 12.5 mM Tris-HCl at 25°C after 96 hours (4Days, circles), 58 days (squares), and 102 days (diamonds). Folded state was measured by chromophore fluorescence excited at 450nm, and observed at 508 nm (3nm slit width). After following transition curves for extremely long time scales, unfolding transitions collapse the hysteresis zone, while refolding transitions do not change.

Figure 9. Guanidine Induced Equilibrium Unfolding Transition of R96A (a) and M88Y/Y74M without chromophore (b) sfGFP at pH 6.8 in 12.5mM Tris-HCl at 25° C. R96A equilibrated for 96 hours. M88Y/Y74M equilibrated for 24 hours. Folded state was measured by change in average wavelength of tryptophan fluorescence emission from excitation at 295nm (slit width 3nm). Blue squares show unfolding transition and red circles show refolding transition. Black line shows two-state folding fit to all points. In the absence of the chromophore, hysteresis is removed.

Figure 10. 1-D ^1H 800 MHz Spectra of “Trapped” and Native sfGFP
 at pH 6.8 in 30mM phosphate, 100mM NaCl and 5% D_2O at 316 °K. Trapped species were prepared as described in Materials & Methods by unfolding sfGFP at 6M guanidine-HCl for 10 minutes, and then jumped to 1.8 M guanidine-HCl for 96 hours. The trapped species corresponds to as sample in the transition of the refolding curve, shown as a black boxes in Figure 7. (A) The chemical shift dispersion of the spectra indicate quite similar global folds, while certain areas show marked heterogeneity (B&C), showing evidence of trapped folding caused by structural heterogeneity. Boxes in A show regions expanded in B & C. Red traces are native sfGFP at 3 mg/mL and blue traces are trapped sfGFP at 3 mg/mL.

Figure 11. Structure of GFP

(a) GFP consists of an 11 stranded β -barrel surrounding an α -helix wherein the chromophore lies (in green). Proline residues discussed are in red. (b) The barrel is stripped away showing detail of the α -helix and the prolines. This representation was modeled with PyMol using the sfGFP crystal structure (2B3P). The prolines cap the ends of the central α -helix, which we hypothesize are required for helix structure and chromophore formation.