doi:10.1016/j.jmb.2007.04.039 J. Mol. Biol. (2007) 370, 356–371

JMB

Available online at www.sciencedirect.com

ScienceDirect

ELSEVIER

Stable Intermediate States and High Energy Barriers in the Unfolding of GFP

Jie-rong Huang†, Timothy D. Craggs†, John Christodoulou and Sophie E. Jackson*

Chemistry Department Lensfield Road, University of Cambridge, Cambridge CB2 1EW, UK

We present a study of the denaturation of a truncated, cycle3 variant of green fluorescent protein (GFP). Chemical denaturation is used to unfold the protein, with changes in structure being monitored by the green fluorescence, tyrosine fluorescence and far-UV circular dichroism. The results show that the denaturation behaviour of GFP is complex compared to many small proteins: equilibrium is established only very slowly, over the time course of weeks, suggesting that there are high folding/unfolding energy barriers. Unfolding kinetics confirm that the rates of unfolding at low concentrations of denaturant are very low, consistent with the slow establishment of the equilibrium. In addition, we find that GFP significantly populates an intermediate state under equilibrium conditions, which is compact and stable with respect to the unfolded state (m_{IU}=4.6 kcal mol^{-1} M^{-1} and ΔG_{IU}=12.5 kcal mol^{-1}).

The global and local stability of GFP was probed further by measuring the hydrogen/deuterium (H/D) NMR exchange rates of more than 157 assigned amide protons. Analysis at two different values of pH showed that amide protons within the β-barrel structure exchange at the EX2 limit, consequently, free energies of exchange could be calculated and compared to those obtained from the denaturation-curve studies providing further support for the three-state model and the existence of a stable intermediate state. Analysis reveals that amide protons in β-strands 7, 8, 9 and 10 have, on average, higher exchange rates than others in the β-barrel, suggesting that there is greater flexibility in this region of the protein. Forty or so amide protons were found which do not undergo significant exchange even after several months and these are clustered into a core region encompassing most of the β-strands, at least at one end of the barrel structure. It is likely that these residues play an important role in stabilizing the structure of the intermediate state. The intermediate state observed in the chemical denaturation studies described here, is similar to that observed at pH 4 in other studies.

*Corresponding author

Keywords: EX1; two-state model; three-state model; NMR; protein folding

Introduction

Green fluorescent protein (GFP) from the jellyfish *Aequorea victoria* is one of the most important proteins currently used in biological and medical research having been extensively engineered for use as a marker of gene expression and protein localization, as an indicator of protein–protein interactions and as a biosensor.[1] Its widespread use results from its unique spectroscopic properties, the 238 residue protein undergoing an autocatalytic post-translational cyclization and oxidation of the polypeptide chain around residues Ser65, Tyr66 and Gly67, to form an extended conjugated π-system, the chromophore, which emits green fluorescence.[2] No cofactors are necessary for either the formation

† J.-r. H. and T. D. C. contributed equally to the work.

Abbreviations used: GFP, green fluorescent protein; H/D, hydrogen/deuterium; GdmCl, guanidinium chloride.

E-mail address of the corresponding author: sej13@cam.ac.uk

or the function of the chromophore,[3] which is embedded in the interior of the protein surrounded by an 11-stranded β-barrel[4,5] (Figure 1). GFP is remarkable for both its structural stability, and high fluorescence quantum yield, the latter a result of the fact that in the native state the chromophore is rigid and shielded from bulk solvent. Upon denaturation, the GFP chromophore remains chemically intact but fluorescence is lost with the destruction of tertiary structure. The green fluorescence is therefore a sensitive probe of the folding of the protein.[6]

In all cases, GFP needs to fold efficiently in order to function in the myriad of biological assays and experiments in which it is used, and inefficient folding is known to limit its use in some applications.[1] Despite this, relatively little is known about the folding of this protein either *in vitro* or *in vivo*. Recent studies by Kuwajima and co-workers have provided the most detailed information to date.[7,8] Their studies have focused on the folding pathway of GFP from the acid-denatured state and they have proposed a model in which GFP folds through several intermediate states. Although, this paper represents a significant contribution to our understanding of the folding of GFP, further studies using complementary techniques and probes are clearly necessary in order to provide a more complete description of the folding pathway of this large, complex and important protein.

A complete assignment of the NMR resonances for the backbone (^{13}C, ^{15}N and ^{1}H) of GFP has been published independently by ourselves and others[9,10] thus enabling the use of hydrogen/deuterium (H/D) exchange techniques to probe the stability and folding of the protein under equilibrium conditions.[11] These techniques have been used extensively on a number of proteins and have provided valuable information on different aspects of the folding energy landscapes. H/D exchange can inform on partially structured states (potential high energy intermediates on a folding pathway),[12–14] on global and local stability,[15–18] on residual structure in the denatured state,[19–24] as well as on cooperatively unfolding regions of proteins.[11,14] Although H/D exchange results have been reported for GFP, the conditions used did not allow a quantitative analysis of the results in terms of the global and local stability of the protein.[25]

Here, we have applied both optical spectroscopy and H/D exchange NMR experiments to study the global and local stability of GFP. Fluorescence, far-UV CD and NMR measurements are made under the same conditions at different pH values and temperatures, enabling not only a comparison between the two probes, but also the establishment of the H/D exchange regime (EX1/EX2). At the EX1 limit, where the intrinsic exchange constants are high in comparison to the closing rates, then the amide exchange rates are determined by the opening rates, which, for amide groups which only exchange on global unfolding, correspond to the unfolding rate.[15] In contrast, at the EX2 limit, where the intrinsic exchange constants are small in comparison with the closing rates, then the exchange rate observed is determined by the ratio of the opening and closing rate constants (the equilibrium between open and closed states) and the intrinsic exchange rate constant which depends

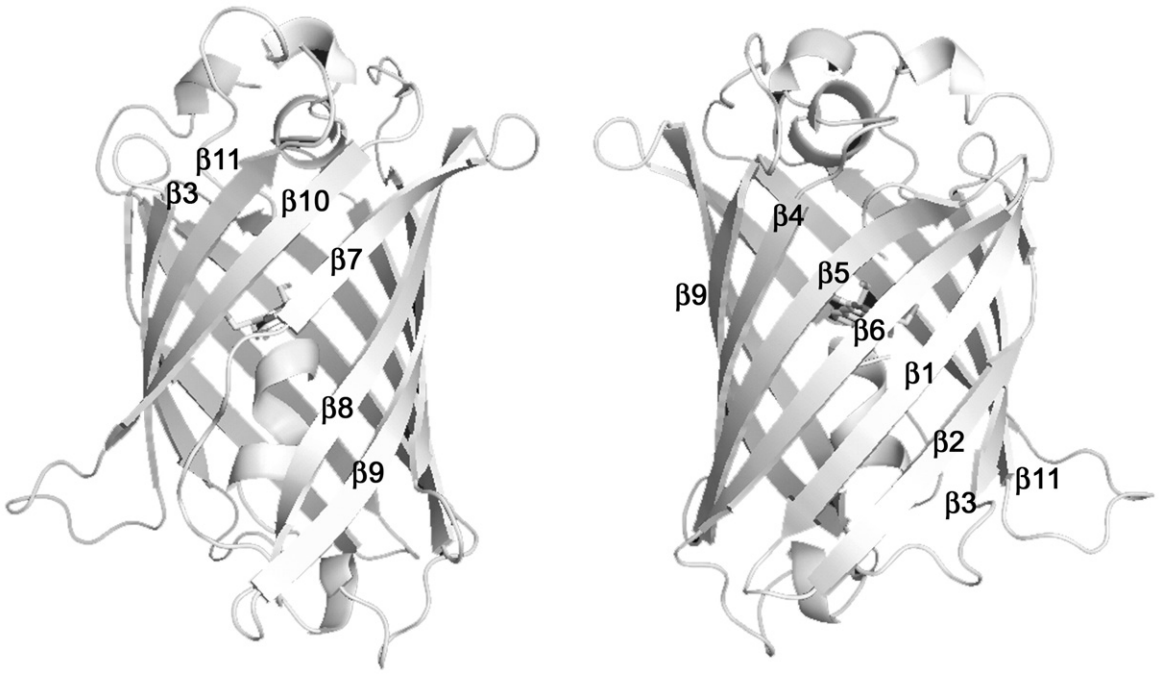

Figure 1. Schematic representations for the structure of GFPuv (PDB code: 1B9C) drawn by PyMol (DeLano Scientific LLC), viewed from two opposite sides. The chromophore is shown in stick mode. Each β-strand is numbered from the N to the C terminus.

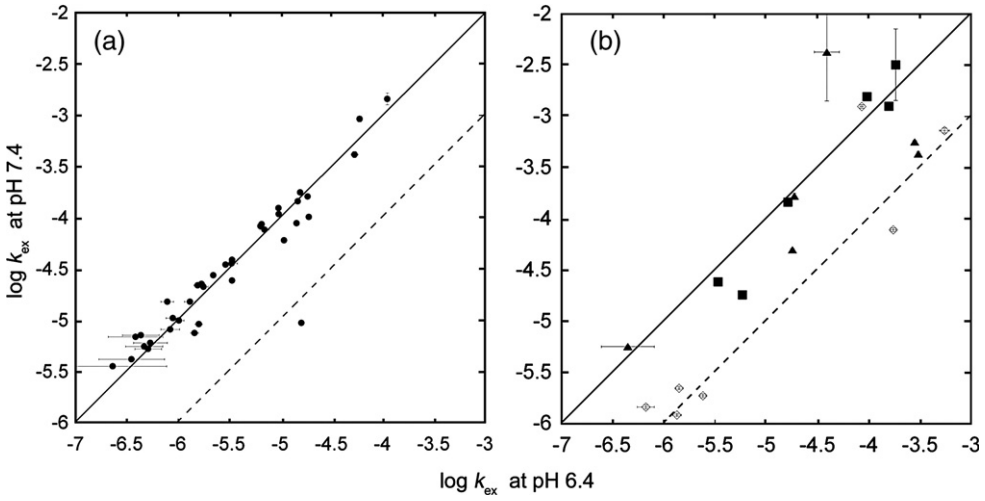

Figure 2. The correlation in exchange rate constants measured at pH 6.4 and pH 7.4 for amide protons of GFP: (a) Residues within the β-sheet structure (filled circles); (b) residues in regions of random coil (filled squares for those outside the barrel and filled triangles for those inside the barrel) or α-helices (open diamonds). The continuous line represents the EX2 limit, the broken line represents the EX1 limit.

upon pH.[15] Therefore, under EX2 conditions, thermodynamic data on the transition between open (unfolded) or closed (folded) forms of the protein can be obtained. Establishment of the exchange regime for GFP enables a full quantitative analysis of the exchange rates measured, thus leading to a complete description of the folding of GFP under equilibrium conditions. In both sets of experiments, a rigorous quantitative analysis is used to reveal the presence of, at least, one intermediate state. The properties of this intermediate state are discussed. In addition to the equilibrium experiments, the unfolding kinetics of GFP over a wide range of chemical denaturant concentrations has been measured to establish that there are high-energy barriers in the unfolding reaction consistent with the equilibrium results.

Results and Discussion

H/D exchange NMR experiments

The H/D exchange rates of 157 amide protons in GFP were measured at pH 6.4 and pH 7.4 at 37 °C over a period of several months by recording successive ^{15}N, ^{1}H heteronuclear single quantum coherence (HSQC) spectra. Amide protons were found to exchange with a very wide range of rates; the fastest protons having exchanged within the dead time of the experiment, whereas the slowest amides did not exchange significantly even after several months.

Characterising the exchange regime: EX1 or EX2 mechanism?

In order to interpret the measured amide exchange rate constants, and to be able to compare the H/D exchange data with those obtained from other experiments, it is important to know whether amide exchange is at the EX1 or EX2 limit. Several methods can be used to identify EX1 and EX2 mechanisms, including the dependence of k_{ex} on pH, which is a reliable and widely used method.[26–28] A plot of log k_{ex} *versus* pH for a given amide hydrogen is diagnostic: at the EX1 limit, k_{ex} does not change with pH and is equal to the opening rate constant, k_{op}. In contrast, at the EX2 limit, k_{ex} varies with pH because it is a function of the intrinsic exchange rate constants, k_{int}, which depends on pH. Hydrogen exchange is primarily a base-catalyzed reaction,[15] so k_{int} increases by one order of magnitude per pH unit. Therefore, at the EX2 limit, log k_{ex} would increase with pH in a linear fashion with a slope of one. One disadvantage of this method, however, is that it assumes the stability of the native state does not vary significantly with pH.[29]

Figure 3. (a) The network of hydrogen bonds in the β-barrel structure. A double line between two residues on two antiparallel β-strands represents two hydrogen bonds. A single-line represents one hydrogen bond. The measured exchange rate constants for the amide groups are classed as follows. Red, very slow (the half-life is longer than one month); yellow, ΔG_{HX} >9.0 kcal mol^{-1}; green, 7.0 kcal mol^{-1}<ΔG_{HX} <9.0 kcal mol^{-1}; blue, ΔG_{HX} <7.0 kcal mol^{-1}; dark blue, very fast (amide hydrogen exchanged within 20 min); grey, not assigned; white, overlapped peaks in the HSQC spectrum. (b) Amides in regions of random coil (circles) or α-helices (rectangles). Red, very slow; dark blue, very fast; yellow, exchanged within a month; white, overlapped peaks; grey, not assigned. Three-dimensional representation of the position of very slow exchanging residues. Red balls represent VS residues in β-strands. Yellow balls represent those in α-helices. The left and right Figures are from the same view point as used in Figure 1. The central Figure is from the top of the β-barrel.

An alternative approach is used here which is an extension of the method described above, and involves comparing the k_{ex} values of different protons at just two different pH values.[14,30–32] We measured the H/D exchange rate at two different pH values , pH7.4 and pH6.4 and by plotting log $k_{ex}^{pH\ 6.4}$

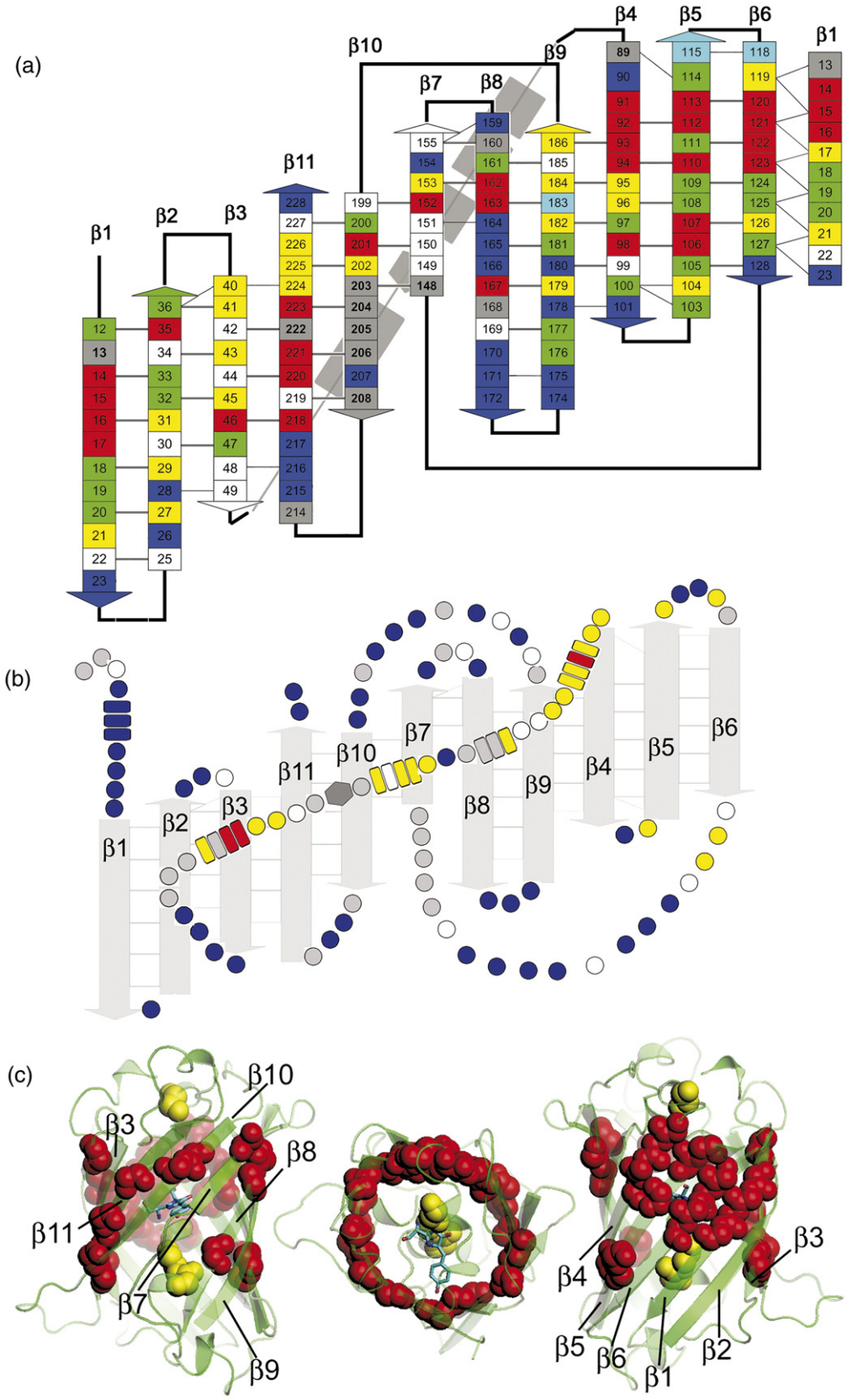

Figure 3 (*legend on previous page*)

versus $\log k_{ex}^{pH\,7.4}$, we can distinguish between the two limits. At the EX2 limit, we should obtain a straight line intersecting with the *y*-axis at one and with a gradient of one. At the EX1 limit, we should obtain a straight line with a gradient of one, which would intersect the *y*-axis at zero, as the exchange rate constants are independent of pH under these conditions.

Figure 2(a) shows the correlation in exchange rate constants measured at pH 6.4 and pH 7.4 for GFP amides in the β-sheet structure. It is clear from this plot that almost all the residues within the β-sheet exchange at the EX2 limit. The only exception is Gly104, however, on close inspection of the crystal structure[33] (Protein Data Bank accession code: 1B9C), it can be seen that its amide NH group is not involved in hydrogen-bond formation in the β-sheet. Only the carbonyl oxygen of Gly104 is H-bonded to the amide hydrogen of Phe100, while the carbonyl oxygen of Phe100 is hydrogen-bonded to the amide hydrogen of Asp103.

The correlation in exchange rate constants measured at pH 6.4 and pH 7.4 for amides in regions of coil or α-helix in GFP is shown in Figure 2(b). In comparison with the β-sheet residues, there are significantly fewer data points for coil and helical regions; this is largely because many of these are exposed on the surface of the protein and undergo extremely rapid exchange which cannot be measured. Data are coded to show: (i) residues in α-helical regions; (ii) residues in coil buried inside the β-barrel; and (iii) other residues. Many of the α-helical residues (shown by open diamonds in Figure 2(b)) exchange at the EX1 limit, although these are residues at the end of α-helices. In contrast, the residues in the middle of helices (Thr59, Leu60, Lys85) exchange extremely slowly and accurate rate constants for these residues cannot be calculated. In comparison with the data on amides involved in β-sheet formation that clearly show EX2 behaviour (Figure 2(a)), the results for amide protons in regions of coil or α-helix are somewhat ambiguous. Although, in theory, kinetic information can be obtained from amides which follow an EX1 mechanism,[29] we focus here on the amide groups which are known to exchange at the EX2 limit in order to undertake a thermodynamic analysis. All the following discussion is, therefore, based on the β-sheet amides, which form the barrel structure of GFP.

Calculating ΔG_{HX} from H/D exchange data

Because most of the residues in the 11-stranded β-barrel structure exchange at the EX2 limit, equation (3) can be used to calculate a free energy associated with the exchange process, ΔG_{HX}, for each residue. The calculated values of ΔG_{HX} of residues within the β-sheets are between 6–11 kcal mol^{-1}. These are the limits of ΔG_{HX}, which we can accurately calculate under the conditions used. If ΔG_{HX} is less than 6 kcal mol^{-1}, the half-life would be less than 3 min, and significantly shorter than the time it takes to acquire the first spectrum (ca 20 min). If ΔG_{HX} is greater than 12 kcal mol^{-1} the half-life is longer than 36 days and the precision of ΔG_{HX} from HX experiments is less reliable because of the tendency of GFP to aggregate at high temperature and the very high concentrations required in the NMR experiments over very long periods. Accordingly, we define two classes of exchange: (1) very fast (VF), where complete exchange occurs before the first spectrum is acquired; and (2) very slow (VS), where the intensity of the cross-peak is still greater than 50% its original value after one month of exchange, i.e. the half-life is longer than one month. For all the other amides measured, the exchange rate constants are converted to values of ΔG_{HX} which are then categorized into three classes: (i) ΔG_{HX} less than 7.0 kcal mol^{-1}; (ii) ΔG_{HX} between 7.0 and 9.0 kcal mol^{-1}; and (iii) ΔG_{HX} greater than 9.0 kcal mol^{-1}. In addition to VS and VF, we therefore have three additional classes of exchange, i.e. a total of five classes. These are mapped onto the structure of GFP (Figure 3(a)) and color-coded according to class. A wide range of values of ΔG_{HX} are observed as has been reported for other proteins; hydrogen exchange is caused by both local fluctuations and global/sub-global unfolding events in the absence of denaturant.[12] All the very slowly exchanging amide hydrogen atoms are likely to exchange *via* a mechanism which is dominated by the global unfolding event.

There are about 40 peaks in the HSQC spectra, which exchange in the VS category (Figure 4). The ΔG_{HX} of these very slowly exchanging cross-peaks can be estimated to be greater than 12 kcal mol^{-1}, and the largest ΔG_{HX} can be regarded as corresponding to global unfolding.[34] The exchange rate constants of these very slowly exchanging amide hydrogen atoms can be estimated to be in the order of 10^{-8}~10^{-9} s^{-1} (see Figure 5). In these cases, the ΔG_{HX} for global unfolding can be estimated to be between 14 kcal mol^{-1} and 15.5 kcal mol^{-1}.

An H/D exchange study of GFP has been reported by Holak and co-workers.[25] In their NMR study, measurements of H/D exchange were used to provide information on the conformational flexibility in the protein on the micro-to millisecond time-scale and these were compared with ^{15}N-relaxation measurements which report on pico-to nanosecond time-scale fluctuations. Whereas, the latter showed that the β-barrel of GFP is rigid on short time-scales, the H/D exchange showed that there is increased flexibility in β-strands 3 and 7–10 over longer time-scales. In our study, we repeated the H/D exchange measurements on GFP under controlled conditions, which allow us to extend the analysis of the exchange rates and enable us to calculate quantitative values of stability in terms of ΔG_{HX}.

Both our study and that of Holak and co-workers,[25] show that the region of β-barrel encompassing strands β7 to β10 has, on average, higher exchange rates than the rest of the β-barrel structure. However, even within these four strands there are some amides, which exchange only very slowly

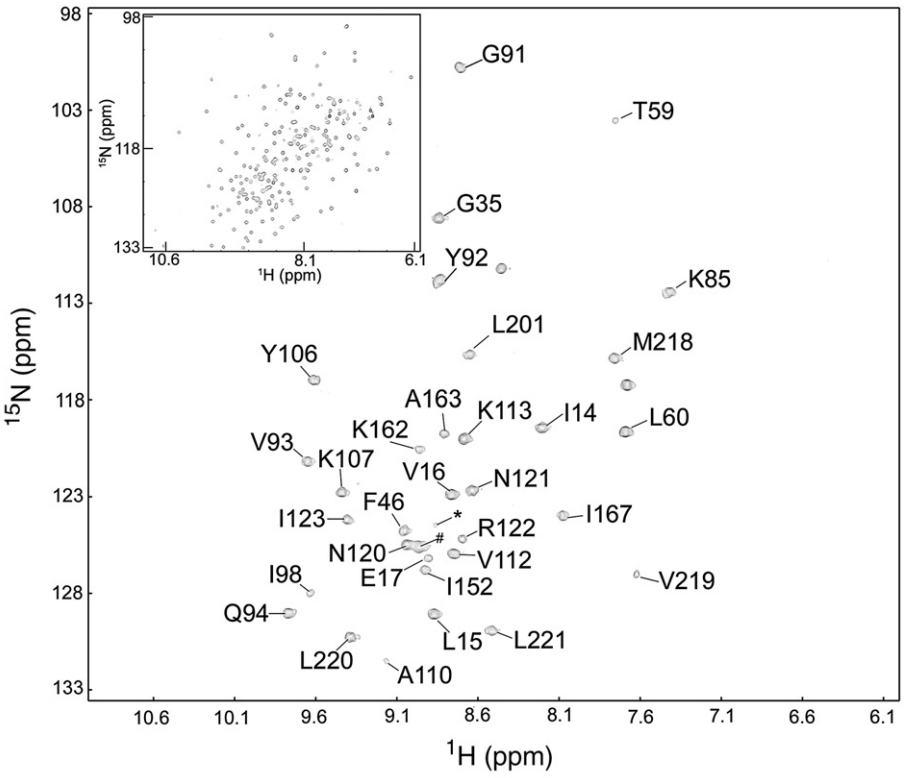

Figure 4. The HSQC spectrum after one month of H/D exchange. Hatch mark (#) are used to indicate the cluster of peaks of Leu42, Leu44, and Cys48; the asterisk (*) is the overlapped peak of H169 and I188. The inset shows the HSQC spectrum of GFP before H/D exchange.

(Figure 3(a)). It is, therefore, not completely clear whether the antiparallel four-stranded β-sheet formed by β-strands β7 to β10 is a cooperatively unfolding unit. In addition to the region between β7 and β10, we find that there is no very slowly exchanging amide between β-strands 2 and 3, suggesting that these strands might also be able to move apart from each other without losing their interactions with the rest of the β-barrel.

A three-dimensional representation of GFP labelled with the most slowly exchanging residues is shown in Figure 3(c). In general, these residues are clustered together primarily on one face of the β-barrel; however, at one end of the barrel they do form a ring-like structure. It is likely that these residues exchange only very slowly because they remain involved in hydrogen-bonded secondary structure in the intermediate state (see the next section).

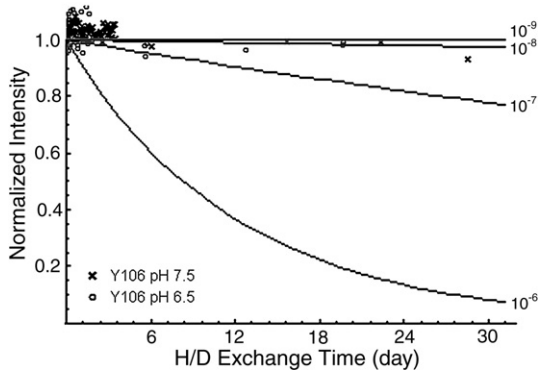

Figure 5. Exchange data for the very slowly exchanging amide group of Tyr106. The continuous curves represent the data expected for exchange rate constants of 10^{-6}–10^{-9} s^{-1}. This allows an estimation of the exchange rate constant for Tyr106 of between 10^{-8}–10^{-9} s^{-1}.

Chemical denaturation and fluorescence measurements

Guanidinium chloride (GdmCl) was used to induce the unfolding of GFP and denaturation curves were monitored by green fluorescence, a sensitive probe of the native state,[6–8] over a wide range of equilibration times. In contrast to chemical denaturation studies on small monomeric proteins, which reach equilibrium rapidly, we observe that GFP reaches equilibrium only very slowly (Figure 6). Unusual though this may seem, it is not unprecedented and has been reported in a number of studies on the unfolding of fluorescent proteins.[35,36] In these two studies, the stability of oligomeric fluorescent proteins (FPs) such as DsRed were compared with EGFP and stability changes over six to nine days reported.

Despite the fact that the system is not fully equilibrated, it is possible to fit each denaturation curve to various models of folding and obtain apparent thermodynamic parameters. It should be

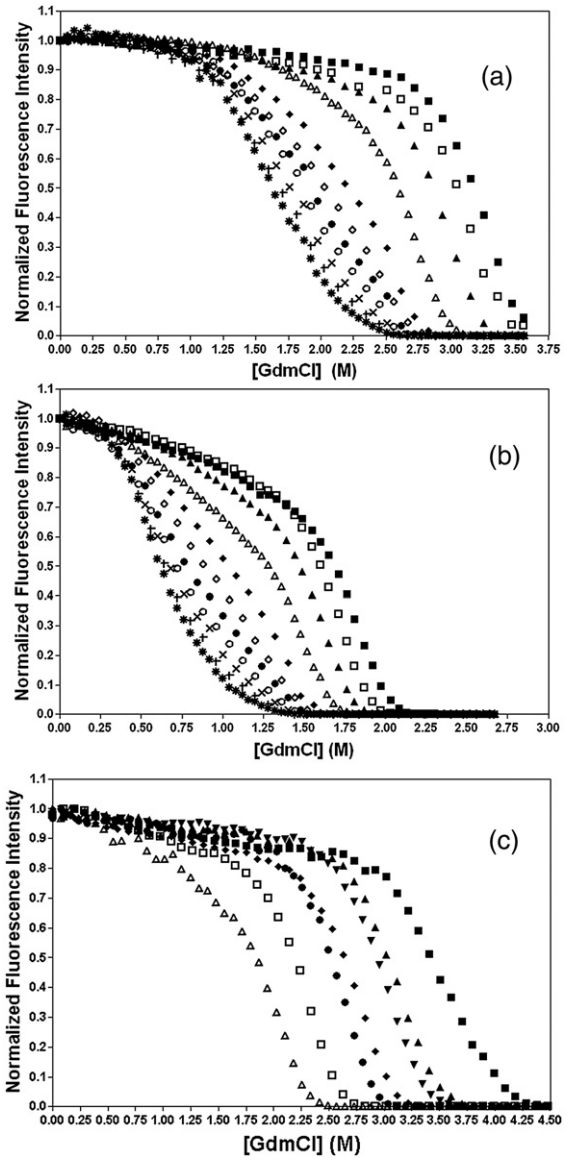

Figure 6. Chemical denaturation curves monitored by green fluorescence. (a) Green fluorescence was measured as a function of GdmCl concentration at pH 7.4 at 37 °C over a wide range of equilibration times: 14 h (filled squares), 24 h (open squares), 48 h (filled triangles), one week (open triangles), two weeks (filled diamonds), three weeks (open diamonds), four weeks (filled circles), five weeks (open circles), six weeks (crosses), seven weeks (plus signs), and eight weeks (asterisks). (b) Green fluorescence was measured as a function of GdmCl concentration at pH 6.4 at 37 °C over a wide range of equilibration times. The symbols are the same as for (a) except the first three time points are: 24 h (filled squares), 48 h (open squares), and 96 h (filled triangles). (c) Green fluorescence was measured as a function of GdmCl concentration at pH 6.0 at 25 °C over a wide range of equilibration times: 3 h (filled squares), 12 h (filled triangles), 24 h (reverse filled triangles), 48 h (filled diamonds), five days (filled circles), 13 days (open squares), and 44 days (open triangles).

stressed that the analysis does not assume that the system is at equilibrium; we are merely fitting the curves to equations that measure the relative populations of native, intermediate and denatured states.

The denaturation curves obtained at pH 7.4, 37 °C (Figure 6(a)) were first analyzed with a two-state model and the dataset at each time point fit to equation (4) to generate values for $[D]^{app}_{50\%}$ and m_{NU} as a function of denaturation time. The results are shown in Figure 7(a) and (b). In this case, $[D]^{app}_{50\%}$ decreases exponentially with equilibration time; however, the m_{NU} values vary significantly with time. The entire dataset was also fit globally allowing $[D]^{app}_{50\%}$ to vary with equilibration time, but sharing the same m_{NU} for all time points. The results for $[D]^{app}_{50\%}$ did not differ significantly from the original individual fitting (Figure 7(c)). An estimate of the apparent free energy of unfolding, ΔG^{app}_{NU}, after eight to nine weeks of equilibration, was obtained by fitting the data in Figure 7(c) to a single-exponential decay to obtain a value $[D]^{app}_{50\%}$ at eight to nineweeks, and then using the relationship $\Delta G^{app}_{NU} = m_{NU}[D]^{app}_{50\%}$ to give a value of 4.0(±0.3) kcal mol^{-1}. This value is considerably smaller than the estimated free energies for global unfolding obtained from the H/D exchange data.

Although the fluorescence data at pH 7.4 at 37 °C, and that at pH 6.0 measured at 25 °C (Figure 6(c)), fit reasonably well to a two-state model, there is clear evidence from the data obtained at pH 6.4 and 37 °C (Figure 6(b)), where two transitions can be seen, that the unfolding of GFP is, at least, a three-state process. In fact, even at pH 7.4, a slight discrepancy in the two-state fit to the data can be observed (Figure 7(d)), and the data are more consistent with a three-state model (Figure 7(e)). The presence of an intermediate state can also explain why the apparent free energies from the fluorescence and NMR experiments do not agree.

Soulages has calculated the potential error in estimates of values for ΔG_{NU} and m_{NU} due to the presence of undetected intermediates from fitting data to a two-state model.[37] Two cases were considered: (1) native and intermediate states are indistinguishable by the observed unfolding property, for example, fluorescence; and (2) intermediate and unfolded states are indistinguishable. Each of these cases underestimate ΔG_{NU} by different amounts and, therefore, the ΔG_{NU} of GFP calculated as described above from the fluorescence data using the two-state model is likely to underestimate the real value. Mayne and Englander[38] have also shown from their studies on cytochrome *c* that the presence of partially folded intermediates populated within the transition region of denaturation curves can affect results quite dramatically, and the denaturation curves were, therefore, re-analyzed with a three-state model, allowing ΔG^{app}_{NI} and ΔG^{app}_{IU} to vary with equilibration time, but sharing m_{NI} and m_{IU} between all datasets obtained under the same experimental conditions. Plots of $[D]^{NIapp}_{50\%}$ and $[D]^{IUapp}_{50\%}$ against equilibration time are shown in Figure 7(f). An estimate of ΔG^{app}_{NI} and ΔG^{app}_{IU} after eight to nine weeks equilibration time was obtained by fitting these curves to a single-exponential process, and using the equations $\Delta G^{app}_{NI} = m_{NI}$

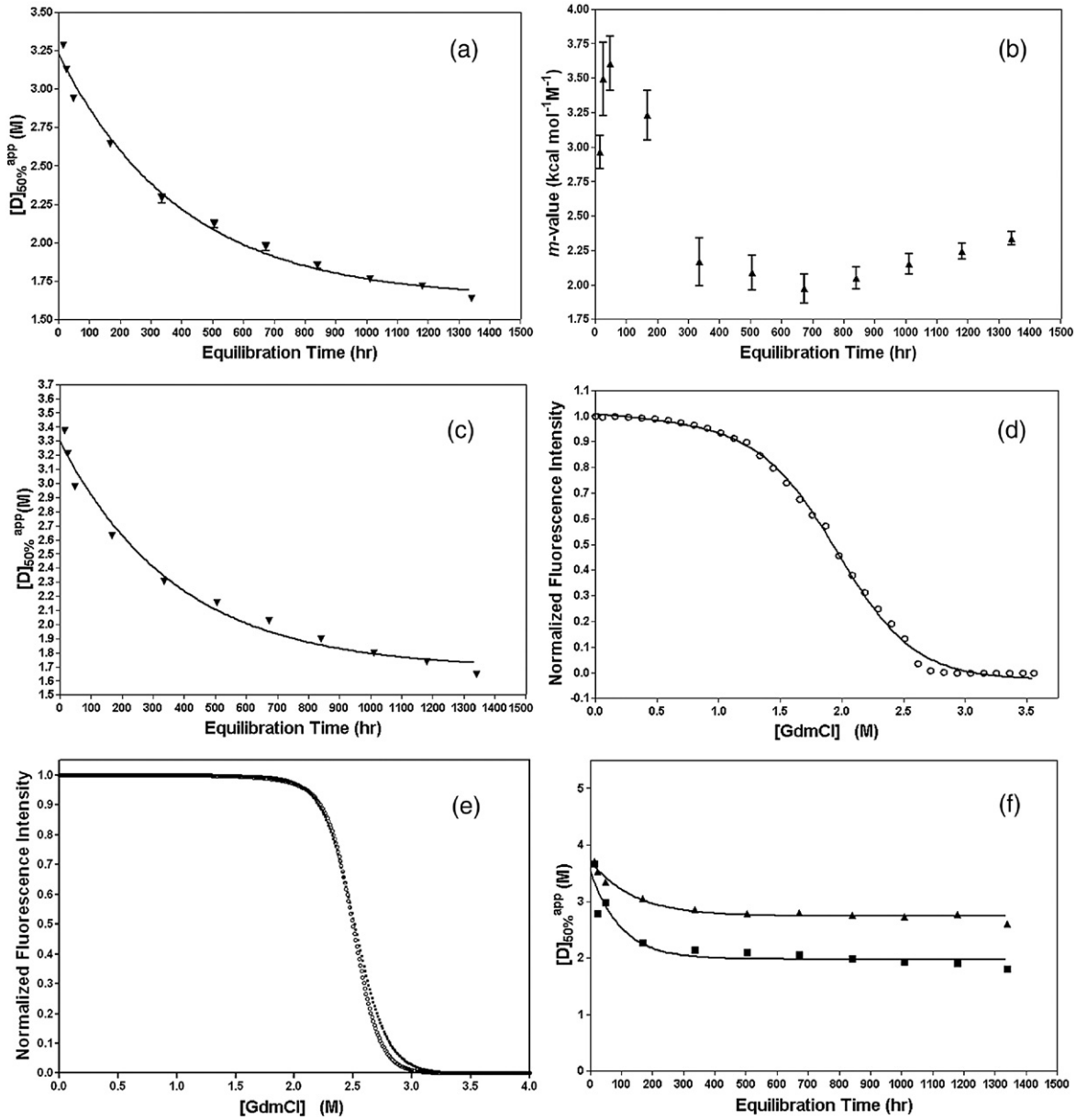

Figure 7. Analysis of the chemical denaturation curves shown in Figure 6. (a) $[D]_{50\%}^{app}$ *versus* equilibration time. The $[D]_{50\%}^{app}$ was calculated from the best fit of the denaturation curves at pH 7.4 and 37 °C to a two-state model. The continuous line is the best fit of the data to an equation describing a single-exponential process. (b) m_{NU} *versus* equilibration time. The m_{NU} was calculated from the best fit of the denaturation curves at pH 7.4 and 37 °C to a two-state model. (c) $[D]_{50\%}^{app}$ *versus* equilibration time. The $[D]_{50\%}^{app}$ was calculated from the global fit (sharing *m*-values) of the denaturation curves at pH 7.4 and 37 °C to a two-state model. (d) Single denaturation curve for GFP monitored by green fluorescence at pH 7.4 and 37 °C after four weeks equilibration time showing the raw data (open circles) and the best fit to a two-state model (continuous line). (e) Simulated denaturation curves from the two-state model (filled squares) and a three-state model (open circles). The parameters for generating these curves are: two-state: m_{NU} = 4.5 kcal mol^{-1} M^{-1}; ΔG_{NU} = 11.3 kcal mol^{-1}. Three-state: m_{NI} = 2 kcal mol^{-1} M^{-1}; m_{IU} = 4 kcal mol^{-1} M^{-1}; ΔG_{NI} = 6 kcal mol^{-1}; ΔG_{IU} = 9 kcal mol^{-1}. This Figure shows the small deviation between the experimental data observed and the fit when a two-state model is used. (f) $[D]_{50\%}^{app}$ *versus* equilibration time. $[D]^{NIapp}_{50\%}$ (filled squares) and $[D]^{IUapp}_{50\%}$ (filled triangles) are from the global fit of data acquired at pH 7.4 and 37 °C to a three-state model sharing m_{NI} and m_{IU} between all datasets. The continuous lines are single-exponential fits to these two datasets.

$[D]^{NIapp}_{50\%}$ and $\Delta G^{app}_{IU} = m_{IU}[D]^{IUapp}_{50\%}$ to yield values of 3.7(±0.3) kcal mol^{-1} and 12.5(±0.4) kcal mol^{-1}, respectively. An estimate of the overall stability of the native state of GFP with respect to the denatured state is, therefore, approximately 16 kcal mol^{-1}. These values are much more consistent with the exchange free energies calculated from the H/D exchange data. The apparent thermodynamic parameters obtained from these fits are listed in Table 1. This analysis was also undertaken for the data obtained at pH 6.0 and 25 °C and we observed very similar results (data not shown).

Table 1. Apparent thermodynamic parameters from GdmCl denaturation curves

		m_{NU} (kcal mol^{-1} M^{-1})	$[D]_{50\%}$ (M)[a]	ΔG_{NU} (kcal mol^{-1})[a]	
2-state[b]	No constraint	No trend	1.64±0.06	3.54±0.40	
	Shared *m*-value	2.38±0.04	1.68±0.08	4.00±0.26	
		m_{NI} (kcal mol^{-1} M^{-1})	m_{IU} (kcal mol^{-1} M^{-1})	ΔG_{NI} (kcal mol^{-1})[a]	ΔG_{IU} (kcal mol^{-1})[a]
3-state[b]	Shared *m*-values	1.89±0.06	4.56±0.08	3.7±0.3	12.5±0.4
3-state[c]	Shared *m*-values	2.60±0.07	4.24±0.31	6.0±0.2	10.5±0.4

[a] These values were obtained from fitting the plots of $[D]^{app}_{50\%}$ *versus* equilibration time to a single-exponential decay and then using the relationship: $\Delta G = m[D]_{50\%}$ to obtain estimates of the thermodynamic unfolding parameters at equilibrium.
[b] At 37 °C and pH 7.4.
[c] At 25 °C and pH 6.0.

The denaturation curves measured at pH 6.4 and 37 °C (Figure 6(b)) clearly will not fit a two-state mechanism, and fitting to a three-state model was undertaken. In this case, however, even global fits to a three-state model proved difficult (data not shown). One reason for this is that another intermediate state may be populated under these conditions.

From fitting of the fluorescence data to both two and three-state models, and from a comparison of the free energies of unfolding obtained from the H/D exchange and chemical denaturation data, it is apparent that the equilibrium unfolding of GFP is three-state. In order to obtain further information on the unfolding transitions and the intermediate state, we measured the unfolding curves using two other optical probes, far-UV circular dichroism (CD) and tyrosine fluorescence (Figure 8). Due to limitations with protein sample, only nine time points were taken with the tyrosine fluorescence, and three with far-UV CD, but both show similar behaviour to that observed with green fluorescence, i.e. a slow equilibration of the protein in GdmCl. A comparison of the normalized data measured using the three different optical probes under the same experimental conditions and at the same equilibration times are shown in Figure 8(c)–(e). It is clear from these data that the transition monitored by green fluorescence is consistently at slightly lower concentrations of denaturant compared with the transitions monitored by tyrosine fluorescence and far-UV CD, consistent with a three-state model in which an intermediate state is populated, which has lower green fluorescence than the native state but which still has considerable structure (as indicated by tyrosine fluorescence, which probes tertiary structure throughout the protein, and far-UV CD, which probes secondary structure). However, the two transitions (native to intermediate and intermediate to denatured) occur at similar denaturant concentrations and are overlapping. We did not try to fit the tyrosine or far-UV CD data to a three-state model as the signal-to-noise was considerable with these probes compared with the green fluorescence.

In order to ensure that the slow decrease in green fluorescence signal, tyrosine fluorescence and far-UV CD signal observed over time was not due to either the chemical modification of the chromophore or degradation of the protein sample over the long incubation periods of the denaturation experiments, absorbance and fluorescence spectra and SDS–PAGE were recorded for samples of GFP that had been incubated for extended periods of time (up to two years at room temperature in either 3 or 6 M GdmCl in PBS buffer, with 0.1 mM TCEP). Neither experiment showed any evidence of degradation of the protein or chromophore (data not shown), thus providing strong evidence that the transitions we observe are due to the slow unfolding of the protein.

In addition, the reversibility of GFP unfolding in GdmCl and refolding into low concentrations of denaturant was measured by NMR and fluorescence spectroscopy. A sample of GFP was incubated in high concentrations of GdmCl overnight to ensure complete unfolding and then refolded into 0.1 M GdmCl. A 1-D ^{1}H spectrum of the refolded GFP was acquired and compared with a spectrum of a native sample of GFP under identical buffer conditions and protein concentrations. The results, shown in the Supplementary Data, establish that GFP can be reversible refolded with >95% efficiency under these conditions. An emission spectrum of this refolded GFP sample confirmed this and showed that more than 90% of the GFP refolds under these conditions to a state which has the same spectral properties as the native protein (see Supplementary Data).

Nature of the intermediate state – *m*-value calculations

In order to get a better understanding of the intermediate state, in particular to get some estimate of how structured this state is compared to the native state, we calculated theoretical *m*-values for several possible intermediate structures. This was achieved by using SASA calculations for both folded and unfolded regions of GFP,[39–41] combined with empirical data linking changes in SASA with *m*-values[42] as described in Materials and Methods. We were guided in the generation of possible intermediate structures by the H/D exchange results, assuming that regions which showed increased levels of exchange compared to average values, are less stable and are more likely to be unfolded in an intermediate state. We generated potential models of intermediates in which the following elements of secondary structure are unfolded: $I_{7–10}$ (β7 to β10

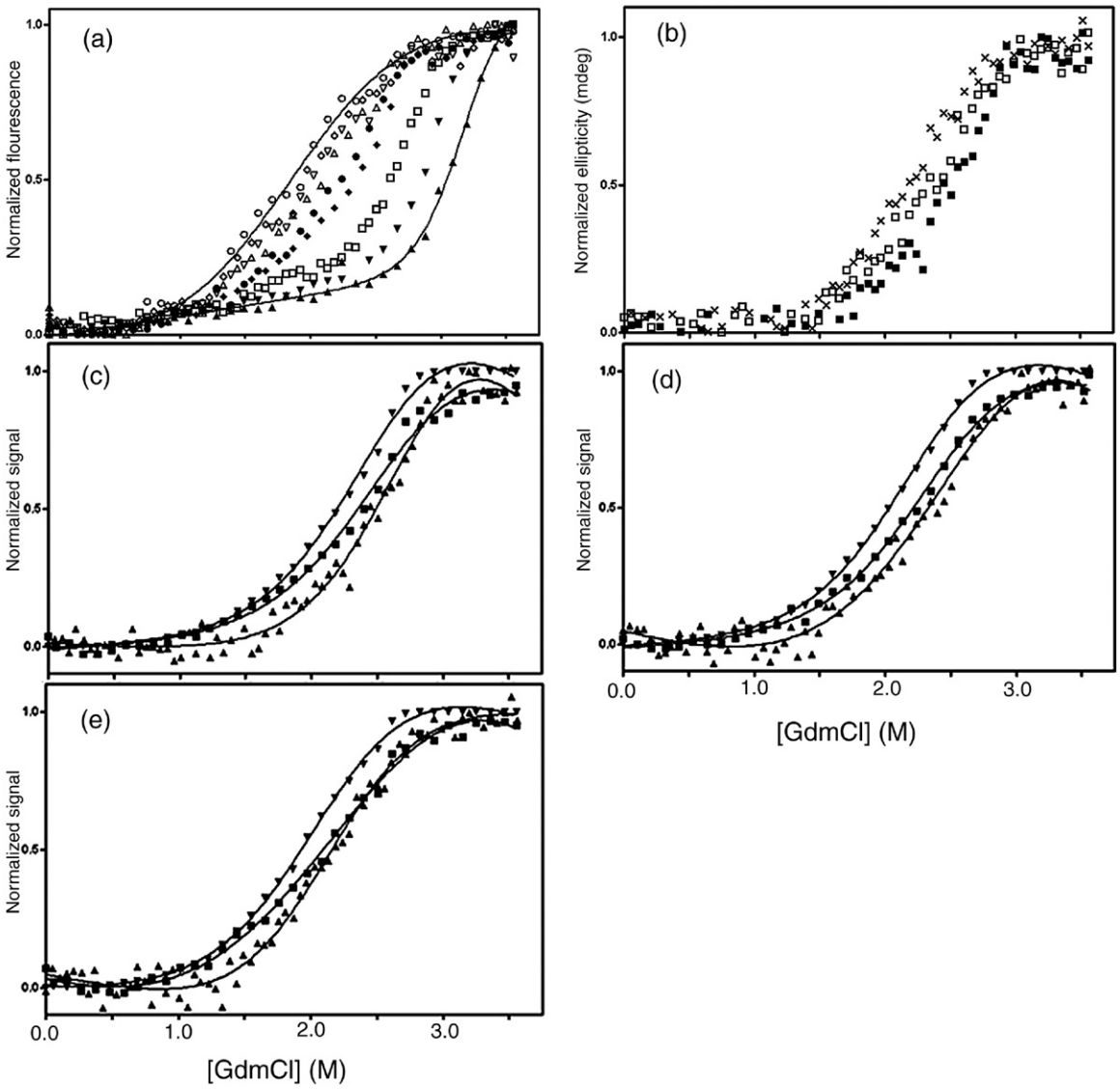

Figure 8. (a) The denaturation curves of tyrosine fluorescence with different equilibration times. 24 h (filled triangles), 48 h (reverse filled triangles),one week (open squares), two weeks (filled diamonds), three weeks (filled circles), four weeks (open triangles), five weeks (open diamonds), six weeks (reverse open triangles) and seven weeks (open circles). The 48 h and seven week data were fitted with two-state model (continuous lines). (b) The denaturation curves followed by far-UV CD after equilibration times of two weeks (filled squares), three weeks (open squares), and four weeks (crosses). (c) Comparison of the denaturation curves after two weeks equilibration with different probes. Green fluorescence (reverse triangles), tyrosine fluorescence (squares) and far-UV circular dichroism (triangles). (d) Comparison of the denaturation curves after three weeks equilibration with different probes. Green fluorescence (reverse triangles), tyrosine fluorescence (squares) and far-UV circular dichroism (triangles). (e) Comparison of the denaturation curves after four weeks equilibration with different probes. Green fluorescence (reverse triangles), tyrosine fluorescence (squares) and far-UV circular dichroism (triangles). As GFP is monomeric under all the experimental conditions studied, there is no protein concentration dependence to the unfolding curves. Therefore, the green and tyrosine fluorescence and far-UV CD data can be compared directly despite the fact that data are acquired at different protein concentrations.

are unfolded), I_{7-9} (β7 to β9 are unfolded) or I_{7-8} (β7 to β8 are unfolded). The results are shown in Table 2. The theoretical m_{NU} value for complete unfolding of GFP in GdmCl predicted in this way is 6.32 kcal mol^{-1} M^{-1}, very similar to the experimental value measured (6.45 kcal mol^{-1} M^{-1}; Table 1), thus giving validity to the method. The experimental value for m_{IU}, 4.56 kcal mol^{-1} M^{-1}, is most similar to the theoretical m_{IU} value calculated for the I_{7-8} intermediate (4.69 kcal mol^{-1} M^{-1}), whereas the experimental value for m_{NI}, 1.89 kcal mol^{-1} M^{-1}, lies between the theoretical m_{NI} values calculated for I_{7-8} and I_{7-9} (1.63 kcal mol^{-1} M^{-1} and 2.22 kcal mol^{-1} M^{-1}, respectively). Although these calculations are only estimates, and do not tell us which elements of secondary/tertiary structure are unfolded in the intermediate state, they do provide a good guide for the extent of structure. The results suggest that at

Table 2. Change in solvent accessible surface area and calculation of theoretical *m*-values

	U (Å^2)	N (Å^2)		$\Delta SASA_{NU}$ (Å^2)	m_{NU} (kcal mol^{-1} M^{-1})
	32,346	9764		22,582	6.32
	I (Å^2)	$\Delta SASA_{NI}$ (Å^2)	m_{NI} (kcal mol^{-1} M^{-1})	$\Delta SASA_{IU}$ (Å^2)	m_{IU} (kcal mol^{-1} M^{-1})
I_{7-10}	19,274	9510	2.66	13,072	3.66
I_{7-9}	17,694	7930	2.22	14,652	4.10
I_{7-8}	15,600	5835	1.63	16,747	4.69

U and N represent the fully unfolded and native states, respectively. I represents the intermediate state. Three possible intermediates are considered here, I_{7-10}, I_{7-9}, and I_{7-8}. $\Delta SASA_{NU}$ is the change in SASA between N and U, i.e. (the SASA of U) – (the SASA of N). The same rule applies to N and I, I and U. All *m*-values in this Table were calculated using equation (6).

least two or perhaps three β-strands are unfolded and coil in the intermediate state consistent with the H/D exchange results which show that β-strands 7–10 have increased flexibility. In addition, Helms *et al.*[43] used a 1 ns molecular dynamic simulation of GFP to show that the β-barrel is disrupted only between β7 and β8 strands over these timescales. The results suggest that not all four strands are unstructured in the intermediate, and it is therefore possible that only two to three strands are completely displaced, or that all four of these strands are partially displaced. There is some evidence for this from a careful inspection of the H/D exchange data which show that there are very slowly exchanging amide protons in these strands at one end of the β-barrel (Figure 3(c)).

It is interesting to compare the intermediate we observe during the GdmCl unfolding of GFP with that populated at pH 4.[8] This acid-induced intermediate has been studied using a range of biophysical techniques including small-angle X-ray scattering measurements and fluorescence spectroscopy by the Kuwajima group.[8] They conclude that the low pH intermediate is a molten-globule like state in which the hydrophobic core around Trp57 is organized but with much reduced levels of quenching of this tryptophan by the GFP chromophore, the chromophore itself showing little fluorescence. The SAXS data indicated a compact and globular structure more expanded than the native state. Given the current level of structural resolution we have for the GdmCl-induced intermediate state, the results suggest that the two stable intermediates populated under different experimental conditions may have similar structures: a local unfolding around β-strands 7–10 would result in an increase in the overall dimensions of the protein, loss of the green fluorescence as the chromophore is exposed to solvent, but with Trp57 remaining largely buried and in a hydrophobic environment. It should be noted, however, that higher resolution structural data on the GdmCl-induced intermediate state is needed to confirm this idea.

Kinetics studies on the unfolding of GFP

The results from the chemical denaturation studies described above, suggest that there are high-energy barriers to the unfolding/folding reactions of GFP which lead to a rather slow equilibration of the system compared with that observed for many small proteins. To investigate this further, the unfolding kinetics of GFP as a function of denaturant concentration were measured under identical con-

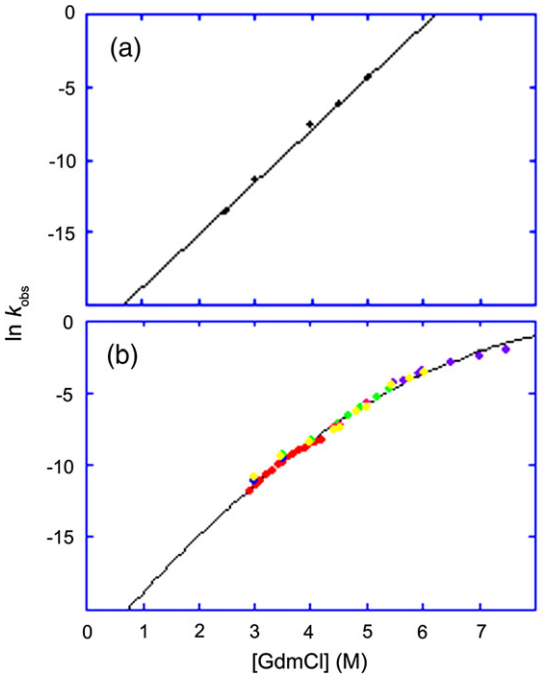

Figure 9. Unfolding kinetics of GFP. (a) Unfolding rate constants measured at pH 7.4 and 37 °C as a function of denaturant concentration, measured using green fluorescence. The continuous line is the best fit of the data to a straight line. The half-life for the unfolding of GFP at low concentrations of denaturant (2 M GdmCl) is estimated to be approximately 30 days. (b) Unfolding rate constants measured at pH 6.0 and 25 °C as a function of denaturant concentration, using different probes and instrumentation. Values calculated from the GdmCl denaturation curves shown in Figure 6 by plotting green fluorescence as a function of time for a particular concentration of denaturant are shown in red; kinetic unfolding experiments using manual mixing and a fluorimeter to measure green fluorescence are shown in blue and green; unfolding rate constants measured using a stopped-flow apparatus and green fluorescence are shown in purple; and unfolding rate constants measured using manual mixing and a fluorimeter but using tyrosine fluorescence to probe folding are shown in yellow. The continuous line shows the best fit of all the data to a second-order polynomial.

ditions to the fluorescence and NMR experiments. The results obtained at pH 7.4 and 37 °C are shown in Figure 9(a). At high concentrations of denaturant, e.g. 5 M GdmCl, the rate of unfolding is reasonably fast with a half-life of about 30–40 s. However, as the concentration of denaturant decreases the rate of unfolding decreases rapidly, the slope of the plot being 3.6 M^{-1}, considerably steeper than for many small proteins. Extrapolating the unfolding rate constant to low concentrations of denaturant, e.g. 1–2 M, which is approximately the midpoint of denaturation curves after several weeks equilibration, results in unfolding half lives on the order of weeks to months. These results are, therefore, consistent with the chemical denaturation curves, which reach equilibrium only very slowly.

To investigate the unfolding kinetics further, experiments were also undertaken at pH 6.0 and 25 °C (Figure 9(b)), using a combination of different probes and methods. Rate constants obtained using green fluorescence to probe folding were in complete agreement with those calculated from measurements of tyrosine fluorescence (Figure 9(b)). As there are ten tyrosine residues in GFP, which are located in different regions of the structure, tyrosine fluorescence is a good probe of the tertiary structure of the protein. Measurements were made over an extended range of GdmCl concentrations by employing a variety of methods. At high concentrations of denaturant, rates were measured using a stopped-flow apparatus, whilst those at lower concentrations of denaturant were measured using manual mixing and a fluorimeter. Finally, rate constants were calculated at very low concentrations of denaturant by analysis of the data in Figure 6. The rate constants calculated by these different methods and probes were remarkably consistent giving rise to a smooth curve (Figure 9(b)). In contrast to the data acquired at pH 7.4 and 37 °C that fit well to a linear equation, data at pH 6.0 and 25 °C showed significant curvature and was fit to a second-order polynomial equation. This apparent difference in behaviour at 25 °C and 37 °C could simply arise from the fact that data at 25 °C was acquired over a much larger range of denaturant concentrations. Curvature in the unfolding limbs of the so-called chevron plots has now been observed for many proteins and has been attributed to movement of the transition state (Hammond behaviour)[44] or changes in the rate-limiting step due to the presence of a high-energy intermediate.[45] As we have strong evidence from the denaturation curves and H/D exchange NMR data, that there is a stable intermediate state present during the unfolding of GFP, we propose that the curvature observed in the kinetic plots is due to the presence of this intermediate.

Conclusion

By combining measurements of the denaturant-induced unfolding of GFP monitored by optical spectroscopies with measurements of the H/D exchange rates of 157 (nearly two-thirds) of GFPs amide protons, we have three strong pieces of evidence to suggest that there is a stable intermediate state in the unfolding of GFP populated under equilibrium conditions. This includes the fits of the green fluorescence chemical denaturation data measured under three different experimental conditions to two and three-state models, a comparison of the free energies of unfolding obtained from H/D exchange experiments and chemical denaturation studies, and a comparison of the chemical denaturation of GFP with different optical probes.

The intermediate state we observe is compact and stable with respect to the denatured state with a m_{IU} value of 4.6 kcal mol^{-1} M^{-1} and ΔG_{IU} value of 12.5 kcal mol^{-1}. It is interesting to note that these values are comparable to the stabilities and *m*-values measured for the native state of many small monomeric proteins.[46] The intermediate retains considerable secondary and tertiary structure; however, it has a reduced green fluorescence signal. From an analysis of the H/D exchange results and the equilibrium unfolding *m*-values, we get some information on the nature of the intermediate state. From this, we can propose a possible structure in which some of the β-strands in the region between β7–β10 have been displaced and in which there is access of solvent to the green chromophore. It is interesting to note that it has been suggested that chromophore formation takes place in a partially structured intermediate state, which may be similar to that we observe in our equilibrium experiments reported here.

In addition, we have established that this large and structurally complex protein is extremely slow to unfold at low concentrations of denaturant. This gives rise to a somewhat unusual behaviour where the chemical denaturation unfolding curves change with equilibration time.

Materials and Methods

Protein expression and purification

The gene encoding GFPuv (Clontech) was cloned into a modified pRSET vector (Invitrogen) without a hexahistidine tag. A stop codon was introduced at residue 230 using polymerase chain reaction (PCR) to produce a truncated form of GFPuv.[10] The resulting plasmid (trGFPuv) was fully sequenced. The wild-type GFP referred to throughout this paper is a pseudo-wild-type GFP corresponding to the trGFPuv construct described above.

Single colonies of transformed *Escherichia coli* cells (C41) harboring trGFPuv were picked from TYE ampicillin plate and were used to inoculate 5 ml of 2×TY media containing 0.1 mg/ml of ampicillin and were grown overnight in a shaker at 37 °C. This overnight culture was used to inoculate 1 l 2×TY media containing 0.1 mg/ml of ampicillin. The growth cultures were left 37 °C on a shaker until the cell density at $A_{600\ nm}$ was 0.7–0.8. The temperature was reduced to 25 °C and the cells were induced with 1 ml of 1 M IPTG and left shaking overnight at 25 °C. The fluorescent green cells were harvested by centrifugation (SLC 4000 Sorvall rotor, 15 min at 8000 rpm at 4 °C) and resuspended into 50 mM Tris (pH 8.0). The

cells were lysed by sonication on ice for 4 min (15 s pulse on and 45 s pulse off) power cycles at power level 8 with a 9 mm probe using a sonicator (Misonix Inc). The lysate was centrifuged at 18,000 rpm (SS34 Sorvall rotor) at 4 °C for 45 min. The supernatant was then pooled and loaded onto a Q-Sepharose (Pharmacia) column (1.5 cm by 10 cm) at 2 ml min^{-1} until the green protein was visibly bound to the top of the column. After six column volume washes of 50 mM Tris (pH 8.0), trGFPuv was eluted with a gradient using 50 mM Tris (pH 8.0), 0.5 M NaCl. The green fractions were pooled and concentrated (Vivaspins 20; Vivascience). The concentrated protein (<10 ml) was gel filtrated on a Superdex G75 column (Amersham Biosciences), which was pre-equilibrated with phosphate-buffered saline (PBS) containing 137 mM NaCl, 2.7 mM KCl, 10 mM phosphate (pH 7.4), and a flow rate of 3 ml min^{-1}. The purity of the GFP eluted was confirmed by SDS–PAGE. For NMR studies, a uniformly ^{15}N-labeled sample was prepared using the same procedure as described above except the 1 l 2×TY overnight growth media were replaced by minimal media,[47] in which the only source of nitrogen was $^{15}NH_4Cl$.

Reagents

Ultra pure guanidinium chloride (GdmCl) was purchased from ICN Biomedicals, Inc. ^2H_2O (99.96%) and $^{15}NH_4Cl$ were from Cambridge Isotope Laboratories, Inc. All other chemicals were of analytical grade and purchased form Sigma, BDH, or Melford Laboratories. Millipore-filtered, double-deionised water was used throughout.

NMR acquisition and H/D exchange experiments

All NMR spectra were acquired on a Bruker AVANCE 700 MHz spectrometer equipped with 5 mm ^1H, ^{13}C, and ^{15}N triple-resonance cryogenic probe heads. Phase-sensitive HSQC) with decoupling during acquisition and 3-9-19 pulse sequence water suppression with gradients[48–50] was used to record the 2-D ^1H-^{15}N spectra. All HSQC spectra were acquired with 1024 (t_2) and 256 (t_1) complex points. All spectra were acquired at 37 °C.

H/D exchange experiments were initiated by adding ^2H_2O (99.96%) to lyophilized protein, which had been prepared at the required pH and buffer conditions. The concentration of GFP was about 400 μM (samples tended to aggregate at concentrations above 600 μM), and the samples were stored at 37 °C between measurements.

Analysis of H/D exchange data and the EX1/EX2 limit

Exchangeable amide hydrogen atoms (NH) which are involved in hydrogen-bonded structure can exchange with solvent hydrogen atoms only when they are transiently exposed to solvent in some kind of closed to open reaction.[15,16] The Linderstrøm-Lang Model has been applied to structurally protected hydrogen atoms in H/D exchange experiments[51,52]:

$$NH_{closed} \underset{k_{cl}}{\overset{k_{op}}{\rightleftharpoons}} NH_{open} \xrightarrow{k_{int}} ND_{open} \qquad (1)$$

where k_{op} and k_{cl} are the opening (unfolding) and closing (folding) rate constants, respectively. The intrinsic rate constant for exchange, k_{int}, depends on the residue type and various conditions (pH, temperature, neighboring amino acids, and isotope effects) and can be estimated on the basis of model compound data.[53] The exchange rate constant, k_{ex}, can be determined from Scheme (1):

$$k_{ex} = \frac{k_{op} \times k_{int}}{k_{op} + k_{cl} + k_{int}} \qquad (2)$$

Two limits of equation (2), so-called EX1 and EX2 limits, have been derived and described elsewhere.[15,16] At the EX2 limit, ΔG_{HX} can be calculated from the measured exchange rate constants using the following equation:

$$\Delta G_{HX} = -RT \ln K_{op} = RT \ln \frac{k_{int}}{k_{ex}} \qquad (3)$$

where K_{op} is the equilibrium opening constant and is equal to k_{op}/k_{cl}.

When calculating the intrinsic exchange rates of residues from model compound data,[53] we used pD values of 7.8 and 6.8 (0.4 pH unit above the measured pH in ^2H_2O).

Fluorescence measurements

Fluorescence measurements were taken with a SLM-Aminco Bowman series 2 luminescence spectrometer using a 1 cm pathlength cuvette. The excitation wavelength was 395 nm with a band pass of 4 nm for both excitation and emission. The largest difference in fluorescence between the native and unfolded states was observed at 507 nm, and emission at this wavelength was used in the subsequent analysis. For tyrosine fluorescence, the excitation wavelength was 276 nm and emission wavelength was 307 nm. The volume of each equilibrium sample was 900 μl with a final concentration of 1 μM GFP in PBS, 0.1 mM TCEP, and the concentration of GdmCl varied, typically from 0 to 3–4 M. Measurements were made at pH 6.4 and pH7.4, and the pH value of the PBS buffer was adjusted by titrating with HCl before dispensing. Prepared samples were stored in a 37 °C incubator before fluorescence measurements were made. For measurements at 25 °C, final conditions were 1 μM GFP, 50 mM Mes(pH 6.0), 0.1 mM TCEP with a range of GdmCl concentrations from 0–4 M.

Far-UV circular dichroism measuremens

Far-UV CD spectra were recorded using a Jasco J-720 spectropolarimeter with an emission band bass of 2 nm. Scans were taken between 210 nm and 250 nm at a scan rate of 1 nm s^{-1}. The largest difference between native and unfolded state of GFP was observed at 222 nm, which was used to monitor unfolding. The protein sample was 10 μM with various concentrations of GdmCl (0 to 4 M). Prepared samples were stored and measured at 37 °C. A 0.1 cm cell was used.

Two-state and multi-state model analysis

Equations for fitting fluorescence unfolding data to a two-state model have been derived and described in detail[54]:

$$F = F_N - (F_N - F_U)\frac{\exp\left(\frac{m([D]-[D]_{50\%})}{RT}\right)}{1+\exp\left(\frac{m([D]-[D]_{50\%})}{RT}\right)} \qquad (4)$$

where F is the observed intensity of optical property (e.g. fluorescence) and F_N and F_U are values of the fluorescence intensities of the native and unfolded forms of the protein, R is the gas constant, T is the absolute temperature, m is a constant related to the average fractional change in the degree of exposure of residues on unfolding, and [D] is the concentration of denaturant. $[D]_{50\%}$ is the midpoint of the unfolding transition and concentration of denaturant at which 50% of the sample is unfolded and 50% folded. In practice, the fluorescence data have sloping baselines. Equation (4) can be modified to take this into account by substituting, $F_N = \alpha_{N+}\ \beta_N[D]$, and $F_U = \alpha_{U+}\ \beta_U[D]$, where β_N and β_U are the slopes of the native and denatured baselines, respectively, and α_N and α_U are the intensity of optical properties of the native state and unfolded state, respectively.[55] The resulting equation is that used to fit the data to a two-state model.

For more complex models, including a three-state model, a single equation as a function of the concentration of denaturant can be derived.[56] However, using commercially available software, a combination of the equations described below was used for data analysis. For example, the following equations describing a three-state model were used for data analysis within the software package GraphPad Prism 4.0 (San Diego, CA):

$$K_{NI} = \exp\left(\frac{m_{NI}[D] - \Delta G_{NI}}{RT}\right),\ K_{IU} = \exp\left(\frac{m_{IU}[D] - \Delta G_{IU}}{RT}\right),$$
$$Y_N = \alpha_N + \beta_N[D],\ Fr_N = \frac{1}{1 + K_{NI} + K_{NI}K_{IU}},$$
$$Fr_I = \frac{K_{NI}}{1 + K_{NI} + K_{NI}K_{IU}},\ Fr_U = \frac{K_{NI}K_{IU}}{1 + K_{NI} + K_{NI}K_{IU}},$$
$$F = Y_N + Fr_I(Y_I - Y_N) + Fr_U(Y_U - Y_N) \qquad (5)$$

where Y_N, Y_I, and Y_U are intensities of native, intermediate and unfolded states, respectively. ΔG_{NI} is the difference in Gibbs' free energy of native and intermediate states, and ΔG_{IU} is the difference between intermediate and unfolded states. m_{NI} is a constant related to the average fractional change in degree of exposure of residues between native and intermediate states; the same as m_{IU} but between intermediate and unfolded states. In terms of statistical mechanics, Fr_N, Fr_I, and Fr_U are fractions of the partition function in a three-energy-state system, and the plot of fractional populations of different states *versus* denaturant concentration can be generated from these equations. Full datasets acquired under the same experimental conditions but at different equilibration times were globally fit to the three-state model; values for Y_N, Y_I, Y_U and m_{NI} and m_{IU} were shared between the different equilibration times, whilst ΔG_{NI} and ΔG_{NI} were allowed to vary. Although there are a large number of variables in the three-state model equation, which would make fitting to a single unfolding curve unreliable, the use of a global analysis allows an accurate determination of the apparent thermodynamic parameters.

Solvent-accessible surface area and *m*-value calculations

The solvent-accessible surface area (SASA) of native trGFPuv was calculated from its X-ray crystal structure[33] (PBD code: 1B9C) using the web-based program GETAREA.[39] The SASA of possible intermediate states were estimated using the following method. First, a partially folded structure of GFP was generated simply by deletion of different regions from the PDB file. In generating possible intermediate states we were guided by the H/D exchange results that showed that β-strands 7, 8, 9 and 10, have on average higher exchange rates than the rest of the β-barrel, and may, therefore, be a cooperatively unfolding unit (see Results and Discussion). Three possible intermediate states were considered: I_{7-10} (β7 to β10 are unfolded), I_{7-9} (β7 to β9 are unfolded) or I_{7-8} (β7 to β8 are unfolded). The SASA of the partially folded structures were then calculated using GETAREA as described above. Second, the SASA of the remaining unfolded, unstructured region was calculated using the Upper Bound Model, which is based on a hard-sphere simulation.[41] This method was also used for calculating the SASA of the full-length denatured state assuming that there is no significant residual structure.

The *m*-value of a protein is highly correlated to the change of SASA (ΔSASA) between the different states,[42] e.g. native and unfolded state, and the *m*-value for a protein without cross-links derived from GdmCl-induced denaturation curves follows the following relationship[42]:

$$m\text{-value} = \Delta\text{SASA} \times (0.28 \pm 0.03) \qquad (6)$$

Theoretical *m*-values were calculated using this equation and the ΔSASAs calculated above.

Kinetic measurements

The rate of unfolding of trGFPuv in GdmCl was measured using stopped-flow apparatus (Applied Photophysics). The unfolding buffer and protein solution were mixed in a 10:1 ratio giving final conditions of 1 μM GFP in 50 mM Mes (pH 6.0), 0.1 mM TCEP, and GdmCl concentrations in the range 4.0 M– 7.5 M for the experiments at 25 °C and 1 μM GFP in PBS (pH 7.4), 0.1 mM TCEP for experiments at 37 °C. Rapid mixing was achieved by the simultaneous injection of the solutions through a T-jet mixing chamber using two Hamilton syringes. The unfolding of the trGFPuv was monitored by fluorescence with an excitation wavelength of 395 nm and a 405 nm cut-off filter. At lower concentrations of GdmCl (2.5M–6.0 M), manual mixing was employed and the fluorescence change monitored using a Varian Cary Eclipse Fluorescence spectrophotometer. Changes in the green fluorescence were monitored by excitation at 395 nm and emission at 507 nm. Changes in the tyrosine fluorescence were monitored by excitation at 275 nm and emission at 303 nm thereby minimising the contribution of the single tryptophan residue to the spectra. Excitation and emission band passes were 4 nm. All kinetic traces were averaged over at least three runs and the data fitted well to an equation describing a single-exponential process.

Kinetic parameters from equilibrium data

Data were obtained at many different GdmCl concentrations for the purposes of analyzing the native-denatured state equilibrium. Fluorescence data obtained at different equilibration times but at the same GdmCl concentration were plotted, giving the change in fluorescence *versus* time. These traces were then fitted to an equation describing a single-exponential process.

Acknowledgements

The work was funded in part by the Welton Foundation. J.R.H. acknowledges financial support from the Cambridge Overseas Trust and J.C. from a Wellcome Trust International Prize Travelling Research Fellowship. T.D.C. was funded by the BBSRC. The authors thank the Biomolecular NMR Facility (Department of Chemistry, University of Cambridge), E. Coulstock, and A.L. Mallam for technical assistance.

Supplementary Data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmb.2007.04.039

References

1. Tsien, R. Y. (1998). The green fluorescent protein. *Annu. Rev. Biochem.* **67**, 509–544.
2. Zimmer, M. (2002). Green fluorescent protein (GFP): applications, structure, and related photophysical behavior. *Chem. Rev.* **102**, 759–781.
3. Reid, B. G. & Flynn, G. C. (1997). Chromophore formation in green fluorescent protein. *Biochemistry*, **36**, 6786–6791.
4. Ormo, M., Cubitt, A. B., Kallio, K., Gross, L. A., Tsien, R. Y. & Remington, S. J. (1996). Crystal structure of the *Aequorea victoria* green fluorescent protein. *Science*, **273**, 1392–1395.
5. Yang, F., Moss, L. G. & Phillips, G. N., Jr. (1996). The molecular structure of green fluorescent protein. *Nature Biotechnol.* **14**, 1246–1251.
6. Fukuda, H., Arai, M. & Kuwajima, K. (2000). Folding of green fluorescent protein and the cycle3 mutant. *Biochemistry*, **39**, 12025–12032.
7. Enoki, S., Saeki, K., Maki, K. & Kuwajima, K. (2004). Acid denaturation and refolding of green fluorescent protein. *Biochemistry*, **43**, 14238–14248.
8. Enoki, S., Maki, K., Inobe, T., Takahashi, K., Kamagata, K., Oroguchi, T. *et al.* (2006). The equilibrium unfolding intermediate observed at pH 4 and its relationship with the kinetic folding intermediates in green fluorescent protein. *J. Mol. Biol.* **361**, 969–982.
9. Georgescu, J., Rehm, T., Wiehler, J., Steipe, B. & Holak, T. A. (2003). Backbone H(N), N, C(alpha) and C(beta) assignment of the GFPuv mutant. *J. Biomol. NMR*, **25**, 161–162.
10. Khan, F., Stott, K. & Jackson, S. (2003). ^{1}H, ^{15}N and ^{13}C backbone assignment of the green fluorescent protein (GFP). *J. Biomol. NMR*, **26**, 281–282.
11. Bai, Y., Sosnick, T. R., Mayne, L. & Englander, S. W. (1995). Protein folding intermediates: native-state hydrogen exchange. *Science*, **269**, 192–197.
12. Englander, S. W. (2000). Protein folding intermediates and pathways studied by hydrogen exchange. *Annu. Rev. Biophys. Biomol. Struct.* **29**, 213–238.
13. Hoang, L., Bedard, S., Krishna, M. M., Lin, Y. & Englander, S. W. (2002). Cytochrome c folding pathway: kinetic native-state hydrogen exchange. *Proc. Natl Acad. Sci. USA*, **99**, 12173–12178.
14. Yan, S., Kennedy, S. D. & Koide, S. (2002). Thermodynamic and kinetic exploration of the energy landscape of *Borrelia burgdorferi* OspA by native-state hydrogen exchange. *J. Mol. Biol.* **323**, 363–375.
15. Dempsey, C. E. (2001). Hydrogen exchange in peptides and proteins using NMR-spectroscopy. *Prog. NMR Spectrosc.* **39**, 135–170.
16. Krishna, M. M., Hoang, L., Lin, Y. & Englander, S. W. (2004). Hydrogen exchange methods to study protein folding. *Methods*, **34**, 51–64.
17. Maity, H., Lim, W. K., Rumbley, J. N. & Englander, S. W. (2003). Protein hydrogen exchange mechanism: local fluctuations. *Protein Sci.* **12**, 153–160.
18. Krishna, M. M., Lin, Y., Mayne, L. & Englander, S. W. (2003). Intimate view of a kinetic protein folding intermediate: residue-resolved structure, interactions, stability, folding and unfolding rates, homogeneity. *J. Mol. Biol.* **334**, 501–513.
19. Jeng, M. F., Englander, S. W., Elove, G. A., Wand, A. J. & Roder, H. (1990). Structural description of acid-denatured cytochrome *c* by hydrogen exchange and 2D NMR. *Biochemistry*, **29**, 10433–10437.
20. Hughson, F. M., Wright, P. E. & Baldwin, R. L. (1990). Structural characterization of a partly folded apomyoglobin intermediate. *Science*, **249**, 1544–1548.
21. Dabora, J. M., Pelton, J. G. & Marqusee, S. (1996). Structure of the acid state of *Escherichia coli* ribonuclease HI. *Biochemistry*, **35**, 11951–11958.
22. Arai, M. & Kuwajima, K. (1996). Rapid formation of a molten globule intermediate in refolding of alpha-lactalbumin. *Fold. Des.* **1**, 275–287.
23. Chamberlain, A. K. & Marqusee, S. (1998). Molten globule unfolding monitored by hydrogen exchange in urea. *Biochemistry*, **37**, 1736–1742.
24. Forge, V., Wijesinha, R. T., Balbach, J., Brew, K., Robinson, C. V., Redfield, C. & Dobson, C. M. (1999). Rapid collapse and slow structural reorganisation during the refolding of bovine alpha-lactalbumin. *J. Mol. Biol.* **288**, 673–688.
25. Seifert, M. H., Georgescu, J., Ksiazek, D., Smialowski, P., Rehm, T., Steipe, B. & Holak, T. A. (2003). Backbone dynamics of green fluorescent protein and the effect of histidine 148 substitution. *Biochemistry*, **42**, 2500–2512.
26. Sivaraman, T., Arrington, C. B. & Robertson, A. D. (2001). Kinetics of unfolding and folding from amide hydrogen exchange in native ubiquitin. *Nature Struct. Biol.* **8**, 331–333.
27. Benitez-Cardoza, C. G., Stott, K., Hirshberg, M., Went, H. M., Woolfson, D. N. & Jackson, S. E. (2004). Exploring sequence/folding space: folding studies on multiple hydrophobic core mutants of ubiquitin. *Biochemistry*, **43**, 5195–5203.
28. Arrington, C. B. & Robertson, A. D. (1997). Microsecond protein folding kinetics from native-state hydrogen exchange. *Biochemistry*, **36**, 8686–8691.
29. Ferraro, D. M., Lazo, N. D. & Robertson, A. D. (2004). EX1 hydrogen exchange and protein folding. *Biochemistry*, **43**, 587–594.
30. Kiefhaber, T. & Baldwin, R. L. (1995). Kinetics of hydrogen bond breakage in the process of unfolding of ribonuclease A measured by pulsed hydrogen exchange. *Proc. Natl Acad. Sci. USA*, **92**, 2657–2661.
31. Pinheiro, T. J., Cheng, H., Seeholzer, S. H. & Roder, H. (2000). Direct evidence for the cooperative unfolding of cytochrome c in lipid membranes from H-(2)H exchange kinetics. *J. Mol. Biol.* **303**, 617–626.
32. Qu, Y. & Bolen, D. W. (2003). Hydrogen exchange kinetics of RNase A and the urea:TMAO paradigm. *Biochemistry*, **42**, 5837–5849.
33. Battistutta, R., Negro, A. & Zanotti, G. (2000). Crystal structure and refolding properties of the mutant

F99S/M153T/V163A of the green fluorescent protein. *Proteins: Struct. funt. Genet.* **41**, 429–437.
34. Pace, C. N. (1975). The stability of globular proteins. *CRC Crit. Rev. Biochem.* **3**, 1–43.
35. Stepanenko, O. V., Verkhusha, V. V., Kazakov, V. I., Shavlovsky, M. M., Kuznetsova, I. M., Uversky, V. N. & Turoverov, K. K. (2004). Comparative studies on the structure and stability of fluorescent proteins EGFP, zFP506, mRFP1, "dimer2", and DsRed1. *Biochemistry*, **43**, 14913–14923.
36. Verkhusha, V. V., Kuznetsova, I. M., Stepanenko, O. V., Zaraisky, A. G., Shavlovsky, M. M., Turoverov, K. K. & Uversky, V. N. (2003). High stability of *Discosoma* DsRed as compared to *Aequorea* EGFP. *Biochemistry*, **42**, 7879–7884.
37. Soulages, J. L. (1998). Chemical denaturation: potential impact of undetected intermediates in the free energy of unfolding and m-values obtained from a two-state assumption. *Biophys. J.* **75**, 484–492.
38. Mayne, L. & Englander, S. W. (2000). Two-state vs. multistate protein unfolding studied by optical melting and hydrogen exchange. *Protein Sci.* **9**, 1873–1877.
39. Fraczkiewicz, R. & Braun, W. (1998). Exact and efficient analytical calculation of the accessible surface areas and their gradients for macromolecules. *J. Computat. Chem.* **19**, 319–333.
40. Miller, S., Janin, J., Lesk, A. M. & Chothia, C. (1987). Interior and surface of monomeric proteins. *J .Mol. Biol.* **196**, 641–656.
41. Creamer, T. P., Srinivasan, R. & Rose, G. D. (1995). Modeling unfolded states of peptides and proteins. *Biochemistry*, **34**, 16245–16250.
42. Myers, J. K., Pace, C. N. & Scholtz, J. M. (1995). Denaturant m values and heat capacity changes: relation to changes in accessible surface areas of protein unfolding. *Protein Sci.* **4**, 2138–2148.
43. Helms, V., Straatsma, T. P. & McCammon, J. A. (1999). Internal dynamics of green fluorescent protein. *J. Phys. Chem. ser. B*, **103**, 3263–3269.
44. Oliveberg, M. (2001). Characterisation of the transition states for protein folding: towards a new level of mechanistic detail in protein engineering analysis. *Curr. Opin. Struct. Biol.* **11**, 94–100.
45. Sanchez, I. E. & Kiefhaber, T. (2003). Evidence for sequential barriers and obligatory intermediates in apparent two-state protein folding. *J. Mol. Biol.* **325**, 367–376.
46. Jackson, S. E. (1998). How do small single-domain proteins fold? *Fold Des.* **3**, R81–R91.
47. Marley, J., Lu, M. & Bracken, C. (2001). A method for efficient isotopic labeling of recombinant proteins. *J. Biomol. NMR*, **20**, 71–75.
48. Bodenhausen, G. & Ruben, D. J. (1980). Natural abundance N-15 NMR by enhanced heteronuclear spectroscopy. *Chem. Phys. Letters*, **69**, 185–189.
49. Piotto, M., Saudek, V. & Sklenar, V. (1992). Gradient-tailored excitation for single-quantum NMR-spectroscopy of aqueous-solutions. *J. Biomol. NMR*, **2**, 661–665.
50. Sklenar, V., Piotto, M., Leppik, R. & Saudek, V. (1993). Gradient-tailored water suppression for H-1-N-15 HSQC experiments optimized to retain full sensitivity. *J. Mag. Reson. ser. A*, **102**, 241–245.
51. Linderstrom-Lang, K. (1958). In (Neuberger, A., ed.), *Symposium of Protein Structure*. Methuen & Co. Ltd., London.
52. Hvidt, A. & Nielsen, S. O. (1966). Hydrogen exchange in proteins. *Adv. Protein Chem.* **21**, 287–386.
53. Bai, Y., Milne, J. S., Mayne, L. & Englander, S. W. (1993). Primary structure effects on peptide group hydrogen exchange. *Proteins: Struct. Funct. Genet.* **17**, 75–86.
54. Pace, C. N. (1986). Determination and analysis of urea and guanidine hydrochloride denaturation curves. *Methods Enzymo.* **131**, 266–280.
55. Santoro, M. M. & Bolen, D. W. (1988). Unfolding free energy changes determined by the linear extrapolation method. 1. Unfolding of phenylmethanesulfonyl alpha-chymotrypsin using different denaturants. *Biochemistry*, **27**, 8063–8068.
56. Barrick, D. & Baldwin, R. L. (1993). Three-state analysis of sperm whale apomyoglobin folding. *Biochemistry*, **32**, 3790–3796.

Edited by K. Kuwajima

(Received 6 November 2006; received in revised form 5 April 2007; accepted 16 April 2007)
Available online 20 April 2007