The extremely slow-exchanging core and acid-denatured state of green fluorescent protein

Jie-rong Huang,[1,2] Shang-Te Danny Hsu,[1] John Christodoulou,[1,3] and Sophie E. Jackson[1]

[1]Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, United Kingdom
[2]Division of Structure Biology, Biozentrum, University of Basel, Klingelbergstrasse 50, 4056 Basel, Switzerland
[3]Institute of Structural Molecular Biology, Department of Biochemistry and Molecular Biology, University College London; and School of Crystallography, Birkbeck College, United Kingdom

(Received 19 May 2008; accepted 7 July 2008; corrected 14 October 2008; published online 15 September 2008)

Green fluorescent protein (GFP) is a large protein with a complex eleven-stranded β-barrel structure. Previous studies have shown that it has a complex energy landscape for folding on which there are several intermediate states and a denatured state with significant residual structure. Here, we use two different types of H/D exchange measurement and nuclear magnetic resonance (NMR) techniques to probe the energy landscape for folding of GFP in further detail. H/D exchange experiments were performed over a wide range of conditions including different concentrations of denaturant. Results show that the penetration model dominates the exchange mechanism, consistent with the known stability and slow unfolding kinetics of GFP. H/D exchange experiments at high *p*H establish that there is an extremely slow-exchanging superstable core of amide protons in GFP that are clustered and located in β-strands 1, 2, 4, 5, and 6. These residues form part of a mini-β-sheet which we propose constitutes a folding nucleus. Using a pulsed-labeling strategy, the acid-denatured state has been investigated and the residual structure observed in earlier studies shown to locate to β-strands 1 and 3. There is some evidence that this residual structure is stabilized by a localized hydrophobic collapse of the polypeptide chain.
[DOI: 10.2976/1.2976660]

CORRESPONDENCE
Sophie E. Jackson:
sej13@cam.ac.uk

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 (Tsien, 1998). 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 and rigidly immobilized conjugated π system, the chromophore, which emits green fluorescence (Zimmer, 2002). No cofactors are necessary for either the formation or the function of the chromophore (Reid and Flynn, 1997), which is embedded in the interior of the protein surrounded by an 11-stranded β-barrel (Ormo *et al.*, 1996; Yang *et al.*, 1996) (Fig. 1).

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 its inefficient folding is known to limit its use in some applications (Tsien, 1998). Over the past 5 years, a number of studies have been published on the folding and stability of GFP and related fluorescent proteins, see Jackson *et al.* (2006) for a recent review. Kuwajima and co-workers have studied the complex refolding kinetics of GFP from an acid-denatured state and proposed a model of folding in which the protein folds through parallel pathways and

HFSP Journal © HFSP Publishing $25.00
Vol. 2, No. 6, December 2008, 378–387 http://hfspj.aip.org

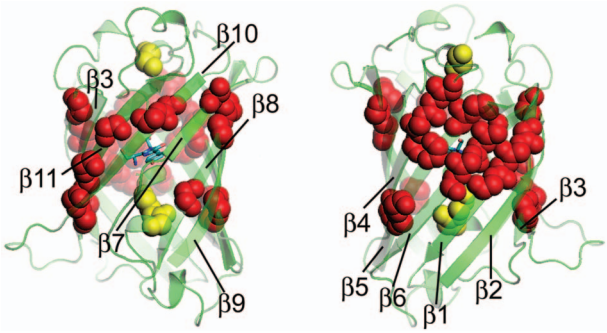

Figure 1. Schematic representations for the structure of GFPuv viewed from two opposite sides drawn by PyMol (DeLano Scientific LLC). The chromophore is shown in stick mode. Each β-strand is numbered from the N to the C terminus. The very slowly H/D exchanging residues at *p*H 7.4, 37 °C are represented as balls (red balls: residues on β-sheets; yellow balls: residues inside the barrel). The side containing β10, β7, β8, β9 has less very slow-exchanging protons compared to the side containing β4, β5, β6, β1, β2.

several intermediate states (Fukuda *et al.*, 2000; Enoki *et al.*, 2004). Recently, they have used small-angle x-ray scattering along with chromophore and tryptophan fluorescence studies to show that the two kinetic intermediates observed during refolding from *p*H 2.0 to 7.5 are closely related to an intermediate populated under equilibrium conditions at *p*H 4.0, which has molten-globule-like properties (Enoki *et al.*, 2006). In previous studies, we have combined H/D exchange nuclear magnetic resonance (NMR) techniques with fluorescence measurements of chemically induced denaturation to establish that there is at least one stable and highly structured intermediate state populated under equilibrium conditions (Huang *et al.*, 2007). Recently, studies on the superfolder GFP (sfGFP) showed that hysteresis observed in the unfolding and refolding equilibrium curves was due to the trapping of GFP in a native-like intermediate state (Andrews *et al.*, 2007). This rough-energy landscape of sfGFP was attributed to the novel backbone cyclization of GFP which is the first step in chromophore formation (Andrews *et al.*, 2007). Clearly, GFP can populate a number of intermediate states depending upon conditions. It is the aim of this study to gain further information on the nature of these intermediate states and the folding energy landscape for this large β-barrel protein.

In our previous H/D exchange studies of GFP, around 40 amide protons were identified that were so highly protected from exchange that accurate rate constants could not be obtained under the conditions used (*p*H 7.4, 37 °C) (Huang *et al.*, 2007). Many of these residues are clustered on one side of the β-barrel structure of GFP located in β-strands 1, 2, 4, 5, and 6, with considerably fewer located on the opposite face in β-strands 3, 7, 8, 9,10, and 11 (Fig. 1). Here, we investigate the exchange behaviour of amide protons in GFP in further detail. In addition to studying the denaturant concentration dependence of the exchange rates of those amide groups which have measurable exchange rates at *p*H 7.4, 37 °C, we measured the exchange rates of the 40 very slow-exchanging amide protons over a wide range of conditions which promote faster exchange. The results show that there is a superstable core within the structure of GFP, in which amide protons are protected from H/D exchange even at high *p*H, high temperature, and high concentrations of denaturant.

Denatured states containing hydrophobic clusters and/or well-defined secondary structure have been observed for an increasing number of proteins (Pace *et al.*, 1992; Shortle *et al.*, 1992; Shortle and Ackerman, 2001; Klein-Seetharaman *et al.*, 2002; Robic *et al.*, 2003; Wirmer *et al.*, 2004; Religa *et al.*, 2005). Fersht and co-workers have suggested that, under some conditions, the denatured states of proteins may be similar to early folding intermediates (Oliveberg and Fersht, 1996; Religa *et al.*, 2005). Studies on the properties of denatured states have therefore become an important area in the study of protein folding pathways. From far-ultra violet (UV) circular dichroism (CD), it is clear that there is residual secondary structure in the acid-denatured state of GFP (Enoki *et al.*, 2004) results supported by small-angle x-ray scattering studies (Enoki *et al.*, 2006) and ^{19}F-NMR/photo-CIDNP experiments (Khan *et al.*, 2006). Other than the fact that residual structure is present in the denatured state of GFP at low *p*H, little is known about which regions of the protein are involved. Here, we have employed a H/D exchange approach for probing those nonexchanged protons in the acid-denatured state that are associated with residual structure. Ten residues are observed which are protected from exchange in the acid-denatured state—revealing important information on the structure in this state. A significant number of these residues are in regions which are identified as forming part of the superstable core of GFP. The implications of these results for the folding mechanism of GFP are discussed.

RESULTS AND DISCUSSION

Mechanisms of H/D exchange in proteins

There are two dominant mechanisms by which amide protons in proteins can exchange with solvent. One involves a global or subglobal unfolding event (Englander and Kallenbach, 1983) and the other involves penetration of a solvent molecule into the protein structure and into proximity of the site of the amide group undergoing exchange (Miller and Dill, 1995; Dempsey, 2001). In the latter case, the so-called penetration model, hydrogen exchange between the protein and bulk solvent, is induced by the redistribution of interior hydrogen bonds or from the random association of pre-existing interior cavities (Nakanishi *et al.*, 1973; Karplus and McCammon, 1981; Wagner and Wuthrich, 1982; Otting *et al.*, 1991; Mayo and Baldwin, 1993). Thus, solvent

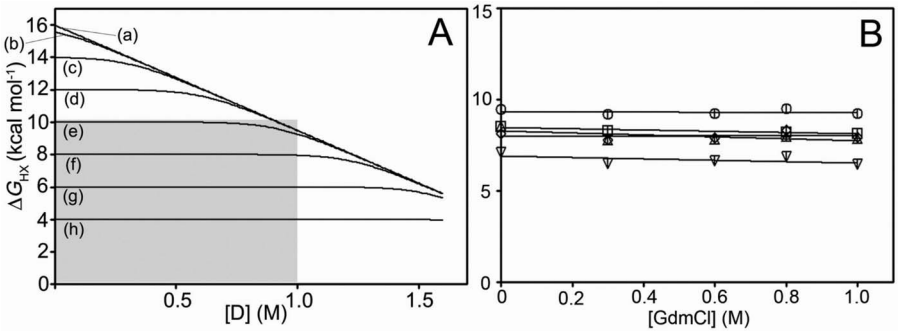

Figure 2. Comparison of calculated and observed exchange energies. (A) Estimation of ΔG_{HX} versus denaturant concentration for different exchange mechanisms. All curves are generated by Eq. (1), with m=6.5 kcal mol^{-1} M^{-1}, ΔG^0_{HX}=16 kcal mol^{-1}, and T=310 K. Lines (a)–(h) are generated using different values of ΔG_{pt} from 16 to 4 kcal mol^{-1}. (B) Measurable H/D exchange peaks at pH 7.4, 37 °C as a function of denaturant concentration. Only residues on β-strand 1 are shown, the rest are shown in supplementary material. V12 (squares), L18 (triangles), D19 (reverse triangles), G20 (diamonds), and D21 (circles).

molecules enter the protein core through transiently formed channels and cavities, which are small and exhibit rapid fluctuations of interior atoms (on the order of tenths to several angstroms). In this case, the exchange process is highly localized and, as such, amide protons in adjacent residues can exchange independently rather than cooperatively. In contrast, chemical exchange in the "local unfolding model" stemming from subglobal or global unfolding events are a cooperative process according to their position within elements of secondary structure. This is characterized by the exchange rates of a group of amide protons having a similar dependence on denaturant concentration, i.e., similar m values and values of ΔG_{HX}, the exchange free energy; this generates the so-called isotherms (Bai *et al.*, 1995) shown in Fig. 2(A). In fact, both models are observed and can operate simultaneously (Miller and Dill, 1995). Data can be fit to a combined model using Eq. (1) (Yan *et al.*, 2002)

$$\Delta G = -RT \ln\left[\exp\left(\frac{-\Delta G_{pt}}{RT}\right) + \exp\left(\frac{m[\mathrm{D}] - \Delta G^0_{HX}}{RT}\right)\right]. \quad (1)$$

The first term in the parenthesis is the contribution from the penetration-induced exchange (where ΔG_{pt} is the energy required for penetration) and the second term the contribution from global/subglobal unfolding events (where ΔG^0_{HX} is the H/D exchange energy of the global/subglobal event in the absence of denaturant, and [D] is the concentration of denaturant). If the first term is small compared to the second, then the equation simplifies to

$$\Delta G = \Delta G^0_{HX} - m[\mathrm{D}]. \quad (2)$$

This produces a straight line with slope of minus m in a plot of ΔG_{HX} versus [D], e.g., line (a) in Fig. 2(A). However, if the penetration term dominates the exchange mechanism, the slope of this straight line decreases, Fig. 2(A).

In order to distinguish between the penetration model and the subglobal/global unfolding model, several groups have investigated the dependence of amide exchange rates on denaturant concentration. This method has also been used to provide information on the cooperativity of unfolding events for proteins which have amide groups which exchange through subglobal unfolding events (Mayo and Baldwin, 1993; Bai *et al.*, 1995; Chamberlain *et al.*, 1996; Yan *et al.*, 2002; Bhutani and Udgaonkar, 2003).

H/D exchange rates for GFP at pH 7.4 in the presence of denaturant

In order to establish the exchange mechanism of the slow-exchanging amide protons in GFP whose exchange rates could be determined at pH 7.4, 37 °C (Huang *et al.*, 2007), a series of experiments were undertaken at different concentrations of chemical denaturant. ^{1}H–^{15}N heteronuclear NMR spectroscopy was used to determine the H/D exchange rates at pH 7.4 in six different concentrations of GdmCl ranging from 0 to 1.0 M. Samples were pre-equilibrated in GdmCl solutions at 37 °C for 8 weeks prior to the measurements in order that the samples had reached equilibrium (Huang *et al.*, 2007). The values of ΔG_{HX} calculated from the measured exchange rate constants as a function of denaturant concentration are shown in Fig. 2(B) and Supplementary Material S1. Figure 2(B) shows typical data for just a few amide protons. However, none of the amide exchange rates showed any dependence on denaturant concentration, Supplementary Materials S1, suggesting that the solvent penetration model dominates under these conditions (Woodward *et al.*, 1982). In order to test this, values of the free energy ΔG_{HX} as a function of denaturant concentration were calculated using a range of penetration energies ΔG_{pt}, from 4 to 16 kcal mol^{-1}, and previously determined values for the global unfolding energy, ΔG^0_{HX}, (16 kcal mol^{-1}) and m value (6.5 kcal mol^{-1} M^{-1}) obtained from fluorescence equilibrium measurements (Huang *et al.*, 2007) [Fig. 2(A)]. Even when ΔG_{pt} is as high as 10 kcal mol^{-1}, the penetration mechanism still dominates the H/D exchange processes for

GFP and the denaturant dependency only becomes evident when the concentration of denaturant is higher than 1.0 M GdmCl. These simulated curves, which have been generated using reasonable estimated parameters, agree with our experimental results which show that, at *p*H 7.4, ΔG_{HX} are independent of denaturant concentration.

These results are somewhat unusual [the exchange rates of most proteins studied to date show some dependence on denaturant concentrations (Takei *et al.*, 2002; Yan *et al.*, 2002; Krishna *et al.*, 2004)], and we attribute this to the fact that GFP is a very stable protein, i.e., it has a large free energy of unfolding (Dietz and Rief, 2004; Huang *et al.*, 2007). It is interesting to note that the values of ΔG_{HX} obtained from the exchange experiments are all lower than those obtained from chemical denaturation unfolding curves monitored by fluorescence and, therefore, exchange is unlikely to be caused by a global unfolding event. As ΔG_{HX} does not depend upon denaturant concentration, it is also unlikely that exchange is caused by a subglobal unfolding event, although a very localized unfolding event where there is only a very small change in solvent accessible surface area cannot be ruled out completely. In other words, for GFP, it is likely that the penetration energy is sufficiently small compared to the subglobal or global unfolding energies that, under these conditions, the penetration model dominates the exchange mechanism.

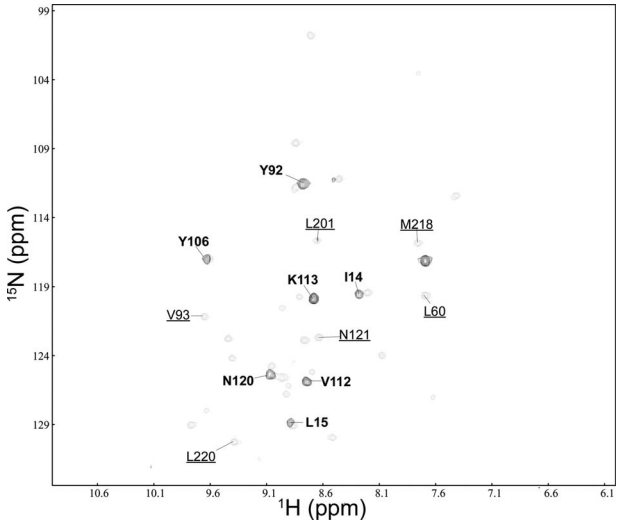

Figure 3. Comparison of the residual peaks in the HSQC spectrum of GFP measured under different conditions. The peaks obtained at *p*H 7.4, 37 °C after 1 month of exchange are shown in grey. The 15 peaks that exchange very slowly are labeled: the six that exchanged after 12 weeks are underlined, while the seven that exchange extremely slowly are shown in bold.

A superstable core of GFP characterized at extreme conditions

To further examine the 40 extremely slow-exchanging amide groups identified in our earlier work, experiments were carried out at a higher *p*H (8.5) in the presence of 0–1.6 M GdmCl in order to accelerate the exchange processes. In all cases, the exchange profiles show no dependence on denaturant concentration. Three groups of amide protons were identified under these new conditions: (i) those that exchange in four weeks (grey peaks in Fig. 3); (ii) those that exchange in 12 weeks (Leu60, Val93, Asn121, Met218, Leu201, and Leu220) (peak labels underlined in Fig. 3); and (iii) those that show no exchange during the period of the experiment, includes residues Ile14, Leu15, Tyr92, Tyr106, Val112, Lys113, and Asn120 [bold labels in Fig. 3; their representative exchange profiles measured in 0.5 M GdmCl are shown in Supplementary Material S2.

Although the seven residues which show most protection from exchange do not exchange sufficiently fast for an accurate exchange rate to be measured, it is possible to estimate an upper limit of $k_{ex} < 10^{-8}$ s^{-1} based on the experimental data (see Supplementary Material S2). It is well established from studies on other systems, that the highest energy for H/D exchange is likely to correspond to the global unfolding energy if isotopic effects and proline isomerization are considered (Bai *et al.* 1994; Huyghues-Despointes *et al.*, 1999). Assuming the seven slowest-exchanging residues are at the EX2 limit, their corresponding ΔG_{HX} would be in the range of 14–18 kcal mol^{-1} (Table I) based on the estimated exchange rates. These high values indicate that the structure is very stable, which is consistent with previous studies on GFP (Huang *et al.*, 2007). A direct comparison of ΔG_{HX} with the free energy of unfolding obtained from equilibrium measurements is not possible as the two experiments were carried out under different conditions. However, if it is assumed that the stability of GFP does not vary much between *p*H 7.4 and 8.5, and there is some experimental evidence to support this (Alkaabi *et al.*, 2005), the free energy of unfolding at 1.6 M GdmCl can be estimated to be approximately 6 kcal mol^{-1} [Fig. 2(A)]. As the values for ΔG_{HX} measured are considerably higher than this, then this suggests that there is considerable residual structure in the denatured state of GFP under native conditions. If, on the other hand, residues exchange at the EX1 limit, then $k_{ex}=k_{op}$, i.e., the opening rate constant is around 10^{-8}–10^{-10} s^{-1}. This is within the range of values obtained for the unfolding rate constant at *p*H 7.4 in 1.6 M GdmCl calculated from extrapolation of the unfolding rate constants measured by stopped-flow spectroscopy at high concentrations of denaturant (Huang *et al.*, 2007). It is unlikely the high levels of amide protection are due to aggregation as no change in chemical shift or line broadening is observed during the course of the experiment. Thus, the seven slowest exchanging residues in GFP (Fig. 4) all show extremely strong protection. These results are consistent with previous studies which show that GFP has a high stability and slow unfolding rates (Huang

Table I. Estimation of the ΔG_{HX} for the extremely slow-exchanging amide proton in GFP using different limiting exchange rates.

		ΔG_{HX} (kcal mol^{-1})		
Residues	k_{int} (s^{-1})	(If $k_{ex}=10^{-8}$ s^{-1})	(If $k_{ex}=10^{-9}$ s^{-1})	(If $k_{ex}=10^{-10}$ s^{-1})
I14	68.4	14.0	15.4	16.8
L15	98.9	14.2	15.6	17.0
Y92	507	15.2	16.6	18.0
Y106	717	15.4	16.8	18.3
V112	90.2	14.1	15.6	17.0
K113	422	15.1	16.5	17.9
N120	78.6	14.1	15.5	16.9

et al., 2007), but also with a model in which there is significant residual structure in this region of the protein under native conditions.

The acid-denatured state of GFP has residual structure

A variety of NMR methods have been used to obtain information on the structure of denatured states of proteins (Dyson and Wright, 2004). Typically, chemical shifts, coupling constants, ^{15}N-relaxation rates, amide-proton temperature coefficients, as well as NOE data have all been used as probes of structure in the denatured ensemble. These experiments have revealed that there can be both native-like and non-native secondary structure present, the exact structure depending upon the protein and experimental conditions. For example, both native and non-native secondary structural preferences were observed in the urea-denatured state of barstar (Bhavesh *et al.*, 2004), in the GdmCl-denatured state of HIV-1 protease (Bhavesh *et al.*, 2001; 2003; Chatterjee *et al.*, 2005), and in the acid-denatured state of hUBF HMG box 1 (Zhang *et al.*, 2005). In contrast, β-type preferences were seen in denatured apomyoglobin which is a highly helical protein (Mohana-Borges *et al.*, 2004). As GFP is a large protein and an assignment of the denatured state is not possible without extensive and expen-

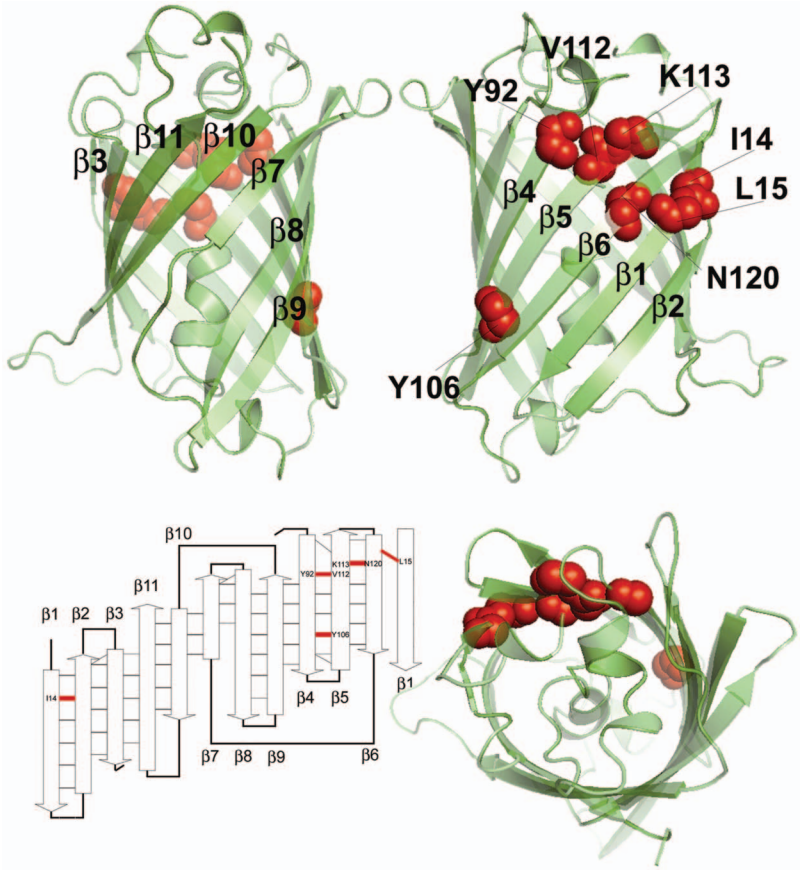

Figure 4. The superstable core of GFP. The three-dimensional structure of GFP showing the positions of all the extremely slow exchanging residues (red balls). Note that they are all located on one side of GFP, and connect β4, β5, β6, β1, and β2. The topology diagram (bottom left) showing the hydrogen bonding by the seven very slow-exchanging amide protons.

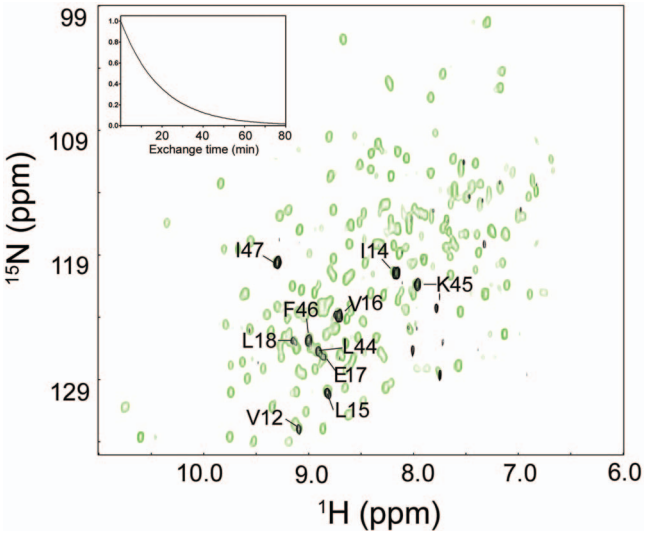

Figure 5. Comparison of acid-denatured-refolded spectrum (black) and native state spectrum (green). The sample was incubated under acid-denaturing conditions for 80 min before refolding. The inset is the estimated exchanging profile of the amide proton with the slowest intrinsic rate constant ($8.67 \times 10^{-4}\,s^{-1}$) under the conditions used.

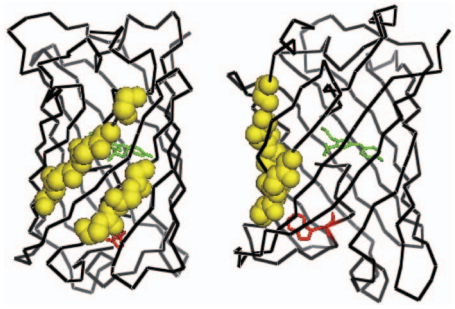

Figure 6. Three-dimensional representation of the residues with protected amide protons under acid denaturing conditions. Yellow balls: residues with protected hydrogens in the acid-denatured state. Trp57 is shown in red and the chromophore in yellow.

sive labeling strategies, an alternative approach to those discussed above was employed here. H/D exchange of amide protons in the denatured state was measured by refolding the protein and probing the native state with conventional NMR experiments.

A pulse-labeling strategy was used to carry out a H/D exchange study on the acid-denatured state of GFP. Denatured samples at *p*H 2.9 were left in D_2O for 5, 20, 40, or 80 min and then refolded and ^{15}N, ^{1}H correlation spectra of the native state acquired. A one-dimensional proton spectrum of the native state of a sample of GFP which had never been unfolded was compared with samples which had been labeled in the acid-denatured state and subsequently refolded, in order to ensure that refolding was complete and successful. After labeling with D_2O in the acid-denatured state, a significant number of crosspeaks were observed to disappear from the amide proton region of the native spectrum, however, the aliphatic and aromatic regions of the spectra remained identical (see Supplementary Material S3). This shows that the protein refolded to the correct native state after denaturation. As expected, a large number of amide protons exchange rapidly in the denatured state. However, a number of peaks showed significant remaining intensity after 5 min of denaturation under acid conditions and some of these peaks remained even after prolonged exchange times such as 20, 40, and 80 min (Supplementary Material S4). If there were no protection against exchange in the denatured state, the residue with the slowest intrinsic exchange rate at *p*H 2.9, 25 °C would have less than 5% of the initial intensity after 80 min of exchange and therefore would be undetectable (inset of Fig. 5). The fact that ten crosspeaks remained clearly visible after 80 min of incubation at *p*H 2.9 in D_2O indicates that there is a significant degree of protection of these amide groups as a result of residual structure (Fig. 5). The results clearly show that there is significant residual structure in GFP at low *p*H, although these types of experiments do not directly reveal whether the structure is native or non-native.

If the residual structure in acid-denatured GFP is native-like, then structural mapping of the highly protected residues show that they form a well-defined cluster spanning across β-strands 1, 2, 3, 6, and 11 (see Fig. 6). Although the residues that show protection are located only in β-strands 1 and 3, in the native state they hydrogen bond with groups in β-strands 2, 6, and 11. For example, the amide proton of Leu44 is protected in the denatured state, and its hydrogen bond acceptor is the carbonyl oxygen of Leu220 on β-strand 11. However, the amide group of Leu220 is not protected in the acid-denatured state. Although this may suggest that the residual structure is not native-like, it is possible that in a highly dynamic and flexible state, such as a denatured state, only one of the hydrogen-bonding partners is protected. An alternative explanation is that the protection of amide groups in β-strands 1 and 3 in the acid-denatured state is a result of non-native residual structure. It is interesting to note that eight out of the ten protected amides are associated with residues which have hydrophobic sidechains and that these are all located relatively close in sequence in the region spanning from residue 12 to 47 on β-strands 1 and 3. Thus, residual structure in the acid-denatured state of GFP may be driven by hydrophobic collapse which results in non-native contacts, and non-native hydrogen bonding. This would explain why the native hydrogen-bonding partners of residues 12–18 and 44–47 do not show protection. Such hydrophobic collapse in a denatured state has been observed for other proteins (Bhavesh *et al.*, 2003; Iimura *et al.*, 2007), for a recent review see reference (Bowler, 2007).

Whether the residual structure in the acid-denatured state of GFP is native or non-native, it is clearly present in the N-terminal region of the protein and this has some po-

tential significance with regard to the folding of the large β-barrel protein. Such residual structure is likely to reduce the initial conformational search the polypeptide chain has to undertake during the first steps of folding and this may lead to faster and more efficient folding. For GFP, this is supported by results from folding studies which have shown that the refolding efficiency and rate from the acid denatured state (Fukuda *et al.*, 2000; Enoki *et al.*, 2004), which we and others have shown contains significant residual structure, is considerably higher than from the chemically denatured state produced by high concentrations of GdmCl which appears to have little or no residual structure (Khan *et al.*, 2006).

Kuwajima and co-workers have shown that a burst-phase intermediate is transiently populated during the refolding of GFP from the acid-denatured state. A rapid increase in tryptophan fluorescence was observed and this was attributed to the formation of an early intermediate in which the single tryptophan residue (Trp57) was partially buried (Enoki *et al.*, 2004). This observation can be explained in more detail by our results: in the acid-denatured state, the whole polypeptide is flexible and without fixed tertiary structure including Trp57, except for the region containing the residual structure identified in this study (yellow balls in Fig. 6). Trp57 (red stick in Fig. 6) is very close in sequence and space to the residual structure in the N-terminal region of the protein, and it is therefore likely that it interacts rapidly with the hydrophobic cluster formed by residues 12–18 and 44–47 immediately after refolding is initiated, thereby protecting the aromatic side chain from bulk solvent and increasing the fluorescence yield (Enoki *et al.*, 2004; 2006). The results presented here are consistent with this model and provide a mechanism for Trp57 burial. After this step, the on-pathway intermediate is formed as discussed in our previous studies (Huang *et al.*, 2007) and by Enoki *et al.* (2004).

Implications for the folding mechanism of GFP

Using two different types of H/D exchange measurement by NMR, we have probed the energy landscape for folding of GFP in detail and have gleaned important details on the mechanism of amide exchange, the regions of the protein which are highly protected from exchange, as well as identifying residual structure in the acid-denatured state. Although in the past there has been some debate about the validity of using H/D exchange results measured under equilibrium conditions to inform on folding pathways (Clarke *et al.*, 1997; Clarke and Itzhaki, 1998), recent work has established that the results from H/D exchange experiments measured under such conditions are similar to those obtained in a kinetic pulsed H/D exchange experiment (Englander *et al.*, 2007). Together, our results provide additional and valuable information relevant to the folding mechanism of GFP.

It is clear from a large number of studies on small proteins that residual structure in denatured states can play an important role in the early states of folding (Oliveberg and Fersht, 1996; Religa *et al.*, 2005). From studies of GFP, it is known that there is considerable residual structure under acid-denaturing conditions as shown by far-UV CD (Enoki *et al.*, 2004), small-angle x-ray scattering (Enoki *et al.*, 2006) and ^{19}F-NMR photo-CINDP (Khan *et al.*, 2006) experiments, however, it was not know which regions of the protein were involved in the residual structure. Here, we have identified the residues associated with this residual structure and shown that they locate to β-strands 1 and 3. As many of the residues involved in the residual structure have hydrophobic sidechains, in addition to the fact that the native-state hydrogen-bonding partners of these residues do not show protection in the acid-denatured state, it is likely that this residual structure is driven by a localized hydrophobic collapse or clustering and that non-native interactions are involved. We propose that the first step in folding either from an extended, random-coil-like state, or from the ribosome, is the rapid formation of this residual structure in the extreme N-terminus of the protein. Formation of this structure may result in the partial burial of Trp57 and correspond to the early intermediate identified by Kuwajima and co-workers (Enoki *et al.*, 2004).

The identification and characterization of a superstable core of GFP, in which amide protons show an extremely high level of protection from exchange, reveals a partially structured state of the protein that is transiently populated under native conditions. The superstable core involves residues in β-strands 1, 4, 5, and 6 which together form a four-stranded mini β-sheet structure. In this case, and in contrast to the results in the acid-denatured state, most of the native hydrogen-bonding partners are protected to the same degree, strongly suggesting that native-like secondary structure is formed in this region. In fact, the protection of these amides is so strong that it is possible that these residues form residual structure in the denatured state which is transiently populated under native conditions. This structure may be formed cooperatively although this cannot be stated with certainty as it was not possible to measure the exchange rates for these amide groups. From our results, however, we propose that this region is either structured in the denatured state populated under native conditions, or constitutes a nucleus for folding formed very early on the folding pathway, and subsequent to its formation the rest of the β-barrel structure forms over several steps.

From the H/D exchange rates described in this study, in conjunction with our previous studies and those of others, we propose a model where, after β-strands 1, 4, 5, and 6 have formed some native-like structure, then this structure is first consolidated by a strengthening of β-strands 4 and 6, with some interactions with residues in the central α-helix and residues in adjacent strands 10 and 11 forming (residues Leu60, Gln94, Asn121, Leu201, Met218, and Leu220 all show levels of protection intermediate between the super-

stable core residues and those identified in an earlier study which we have attributed to a late folding intermediate (Huang *et al.*, 2007). At the final stage, considerable native secondary structure is made between β-strands 1, 4, 5, 6 and to a lesser extent the N-terminal region of the central α helix with additional structure emerging and some protection seen in all the other β strands except β-strand 9. This constitutes the formation of the late folding intermediate which has been shown to be highly structured by fluorescence, H/D exchange and other studies, with partial unfolding in the region of β-strands 7-10 (Helms *et al.*, 1999; Seifert *et al.*, 2003; Enoki *et al.*, 2006; Andrews *et al.*, 2007; Huang *et al.*, 2007). At some point, either in the last transition state or during the energetically downhill process post transition state, these last three to four strands become completely structured and the native state is achieved.

MATERIAL AND METHODS

Protein expression and reagents

Truncated GFPuv expression and purification is described elsewhere (Huang *et al.*, 2007). Ultrapure guanidinium chloride (GdmCl) was purchased from ICN Biomedicals, Inc. D_2O (99.9%) 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-de-ionized water was used throughout.

H/D exchange experiments

Protein samples were pre-equilibrated at 37 °C in various concentrations of GdmCl. Deuterated GdmCl solutions at the required *p*H, buffer, and concentration were prepared in D_2O and were used for initiating H/D exchange by means of a NAP-5 Column (Amersham Bioscience). This buffer exchange step took about 5 min before the first heteronuclear single-quantum coherence (HSQC) spectrum was acquired, longer than in previous studies where lyophilized protein was dissolved directly into D_2O (Huang *et al.*, 2007). However, in these studies, buffer exchange using the NAP-5 column was necessary to prevent the aggregation of GFP during lyophilization in the presence of GdmCl. As the aim of these experiments was to focus on the very slow-exchanging amide protons, the increased dead time was not an issue. The final concentration of GFP was about 600 μM, and all experiments were performed at 37 °C and the samples were stored at 37 °C between measurements.

All NMR spectra were acquired on a Bruker AVANCE 500 or 700 MHz spectrometer equipped with 5 mm ^1H, ^{13}C, and ^{15}N triple-resonance cryogenic probe heads. Phase-sensitive HSQC using Echo/Antiecho-TPPI gradient selection with decoupling during acquisition (Palmer *et al.*, 1991; Grzesiek and Bax, 1993; Schleucher *et al.*, 1994) was used to record the two-dimensional $^1H-^{15}N$ spectra. All HSQC spectra were acquired with 1024 (t_2) and 128 (t_1) complex points. All spectra were acquired at 37 °C. Acquired data were processed by NMRPipe (Delaglio *et al.*, 1995) and analyzed with NMRView (Johnson and Blevins, 1994). Intensities of peaks were exported from NMRView and analyzed by PRISM (GraphPad, Inc.).

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

The mechanism of H/D exchange is well studied and reviewed (Linderstrom-Lang, 1958; Hvidt, 1966; Huyghues-Despointes *et al.*, 1999; Englander, 2000; Dempsey, 2001; Ferraro *et al.*, 2004; Krishna *et al.*, 2004), and can be described by the following two-step model:

$$NH_{closed} \underset{k_{cl}}{\overset{k_{op}}{\rightleftharpoons}} NH_{open} \overset{k_{int}}{\rightharpoonup} ND_{open}, \tag{3}$$

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 (*p*H, temperature, neighboring amino acids, and isotope effects) and can be estimated on the basis of model compound data (Bai *et al.*, 1993). The exchange rate constant k_{ex} determined from the model is

$$k_{ex} = \frac{k_{op} \cdot k_{int}}{k_{cl} + k_{int}}, \tag{4}$$

which has two limits, so-called EX1 and EX2 (Dempsey, 2001; Krishna *et al.*, 2004). At the EX2 limit, $k_{cl} \gg k_{int}$, ΔG_{HX} can be calculated from the measured exchange rate constants using the following equation:

$$\Delta G_{HX} = -RT \ln K_{op} = RT \ln \frac{k_{int}}{k_{ex}}, \tag{5}$$

where K_{op} is the equilibrium opening constant and is equal to k_{op}/k_{cl}. At the EX1 limit, $k_{int} \gg k_{cl}$, the exchange rate constant is simply equal to the opening rate constant

$$k_{ex} = k_{op}. \tag{6}$$

The web-based program SPHERE (Zhang and Roder; Bai *et al.*, 1993; Connelly *et al.*, 1993) was used for calculating k_{int}, and pDs of 7.8 and 8.9 (0.4 *p*H units above the measured *p*H in D_2O) were used.

The acid-denatured sample preparation and NMR spectra acquisition

1 mL of 600 μM GFP in PBS, *p*H 7.4 was diluted ten times into predeuterated PBS, *p*H 2.9. The denatured sample was left at 25 °C (incubated in water bath) for varying lengths of time. During this time, unprotected amide protons exchanged with solvent deuterons, but amide hydrogens which are involved in residual structure remain protected from exchange. The denatured solution was then diluted ten times into deuterated PBS buffer, *p*H 7.4. In this step, protein refolded and the hydrogen-bonded protons in the native structure were protected against exchange. The sample was left at 25 °C for 5 min for complete refolding to take place,

then concentrated at 4 °C to minimize further H/D exchange. The concentration step from a diluted sample of 100 mL to 600 μL took around 4 h using Vivaspin centricons. The NMR spectra of concentrated samples were acquired at 37 °C with Bruker AVANCE 500 or 700 MHz spectrometers. The band-selective optimized flip-angle short-transient HMQC pulse sequence (Schanda and Brutscher, 2005) was applied in order to minimize the experimental time.

ACKNOWLEDGMENTS

STDH is a recipient of a Netherlands Ramsay and a Human Frontier Long-term Fellowship (LT0798/2005). JC acknowledges funding from the Wellcome and Leverhulme Trusts. JC is a recipient of an HFSP Young Investigator's Award (RGY67/2007). We thank the staff and the use of the Biomolecular NMR Facility, Department of Chemistry, University of Cambridge.

REFERENCES

Alkaabi, KM, Yafea, A, and Ashraf, SS (2005). "Effect of *p*H on thermal- and chemical-induced denaturation of GFP." *Appl. Biochem. Biotechnol.* **126**, 149–156.

Andrews, BT, Schoenfish, AR, Roy, M, Waldo, G, and Jennings, PA (2007). "The rough energy landscape of superfolder GFP is linked to the chromophore." *J. Mol. Biol.* **373**, 476–490.

Bai, Y, Milne, JS, Mayne, L, and Englander, SW (1993). "Primary structure effects on peptide group hydrogen exchange." *Proteins* **17**, 75–86.

Bai, Y, Milne, JS, Mayne, L, and Englander, SW (1994). "Protein stability parameters measured by hydrogen exchange." *Proteins* **20**, 4–14.

Bai, Y, Sosnick, TR, Mayne, L, and Englander, SW (1995). "Protein folding intermediates: native-state hydrogen exchange." *Science* **269**, 192–197.

Bhavesh, NS, Juneja, J, Udgaonkar, JB, and Hosur, RV (2004). "Native and nonnative conformational preferences in the urea-unfolded state of barstar." *Protein Sci.* **13**, 3085–3091.

Bhavesh, NS, Panchal, SC, Mittal, R, and Hosur, RV (2001). "NMR identification of local structural preferences in HIV-1 protease tethered heterodimer in 6 M guanidine hydrochloride." *FEBS Lett.* **509**, 218–224.

Bhavesh, NS, Sinha, R, Mohan, PM, and Hosur, RV (2003). "NMR elucidation of early folding hierarchy in HIV-1 protease." *J. Biol. Chem.* **278**, 19980–19985.

Bhutani, N, and Udgaonkar, JB (2003). "Folding subdomains of thioredoxin characterized by native-state hydrogen exchange." *Protein Sci.* **12**, 1719–1731.

Bowler, BE (2007). "Thermodynamics of protein denatured states." *Mol. Biosyst.* **3**, 88–99.

Chamberlain, AK, Handel, TM, and Marqusee, S (1996). "Detection of rare partially folded molecules in equilibrium with the native conformation of RNaseH." *Nat. Struct. Biol.* **3**, 782–787.

Chatterjee, A, Mridula, P, Mishra, RK, Mittal, R, and Hosur, RV (2005). "Folding regulates autoprocessing of HIV-1 protease precursor." *J. Biol. Chem.* **280**, 11369–11378.

Clarke, J, and Itzhaki, LS (1998). "Hydrogen exchange and protein folding." *Curr. Opin. Struct. Biol.* **8**, 112–118.

Clarke, J, Itzhaki, LS, and Fersht, AR (1997). "Hydrogen exchange at equilibrium: a short cut for analysing protein-folding pathways?" *Trends Biochem. Sci.* **22**, 284–287.

Connelly, GP, Bai, Y, Jeng, MF, and Englander, SW (1993). "Isotope effects in peptide group hydrogen exchange." *Proteins* **17**, 87–92.

Delaglio, F, Grzesiek, S, Vuister, GW, Zhu, G, Pfeifer, J, and Bax, A (1995). "NMRPipe: a multidimensional spectral processing system based on UNIX pipes." *J. Biomol. NMR* **6**, 277–293.

Dempsey, CE (2001). "Hydrogen exchange in peptides and proteins using NMR-spectroscopy." *Prog. Nucl. Magn. Reson. Spectrosc.* **39**, 135–170.

Dietz, H, and Rief, M (2004). "Exploring the energy landscape of GFP by single-molecule mechanical experiments." *Proc. Natl. Acad. Sci. U.S.A.* **101**, 16192–16197.

Dyson, HJ, and Wright, PE (2004). "Unfolded proteins and protein folding studied by NMR." *Chem. Rev. (Washington, D.C.)* **104**, 3607–3622.

Englander, SW (2000). "Protein folding intermediates and pathways studied by hydrogen exchange." *Annu. Rev. Biophys. Biomol. Struct.* **29**, 213–238.

Englander, SW, and Kallenbach, NR (1983). "Hydrogen-exchange and structural dynamics of proteins and nucleic-acids." *Q. Rev. Biophys.* **16**, 521–655.

Englander, SW, Mayne, L, and Krishna, MMG (2007). "Protein folding and misfolding: mechanism and principles." *Q. Rev. Biophys.* **40**, 287–326.

Enoki, S, Maki, K, Inobe, T, Takahashi, K, Kamagata, K, Oroguchi, T, Nakatani, H, Tomoyori, K, and Kuwajima, K (2006). "The equilibrium unfolding intermediate observed at *p*H 4 and its relationship with the kinetic folding intermediates in green fluorescent protein." *J. Mol. Biol.* **361**, 969–982.

Enoki, S, Saeki, K, Maki, K, and Kuwajima, K (2004). "Acid denaturation and refolding of green fluorescent protein." *Biophys. J.* **43**, 14238–14248.

See EPAPS Document No. E-HJFOA5-2-001807 for supplemental material. This document can be reached through a direct link in the online article's HTML reference Section or via the EPAPS homepage (http://www.aip.org/pubservs/epaps.html).

Ferraro, DM, Lazo, ND, and Robertson, AD (2004). "EX1 hydrogen exchange and protein folding." *Biophys. J.* **43**, 587–594.

Fukuda, H, Arai, M, and Kuwajima, K (2000). "Folding of green fluorescent protein and the cycle3 mutant." *Biophys. J.* **39**, 12025–12032.

Grzesiek, S, and Bax, A (1993). "Amino acid type determination in the sequential assignment procedure of uniformly 13C/15N-enriched proteins." *J. Biomol. NMR* **3**, 185–204.

Helms, V, Straatsma, TP, and McCammon, JA (1999). "Internal dynamics of green fluorescent protein." *J. Phys. Chem. B* **103**, 3263–3269.

Huang, JR, Craggs, TD, Christodoulou, J, and Jackson, SE (2007). "Stable intermediate states and high energy barriers in the unfolding of GFP." *J. Mol. Biol.* **370**, 356–371.

Huyghues-Despointes, BM, Scholtz, JM, and Pace, CN (1999). "Protein conformational stabilities can be determined from hydrogen exchange rates." *Nat. Struct. Biol.* **6**, 910–912.

Hvidt, A, and Nielsen, SO (1966). "Hydrogen exchange in proteins." *Adv. Protein Chem.* **21**, 287–386.

Iimura, S, *et al.* (2007). "Characterization of the denatured structure of pyrrolidone carboxyl peptidase from a hyperthermophile under nondenaturing conditions: role of the C-terminal alpha-helix of the protein in folding and stability." *Biophys. J.* **46**, 3664–3672.

Jackson, SE, Craggs, TD, and Huang, JR (2006). "Understanding the folding of GFP using biophysical techniques." *Expert Rev. Proteomics* **3**, 545–559.

Johnson, BA, and Blevins, RA (1994). "NMR view—a computer-program for the visualization and analysis of NMR data." *J. Biomol. NMR* **4**, 603–614.

Karplus, M, and McCammon, JA (1981). "The internal dynamics of globular proteins." *Crit. Rev. Biochem.* **9**, 293–349.

Khan, F, Kuprov, I, Craggs, TD, Hore, PJ, and Jackson, SE (2006). "(19)F NMR studies of the native and denatured states of green fluorescent protein." *J. Am. Chem. Soc.* **128**, 10729–10737.

Klein-Seetharaman, J, *et al.* (2002). "Long-range interactions within a nonnative protein." *Science* **295**, 1719–1722.

Krishna, MM, Hoang, L, Lin, Y, and Englander, SW (2004). "Hydrogen exchange methods to study protein folding." *Methods* **34**, 51–64.

Linderstrom-Lang, K (1958). "Deuterium exchange and protein structure." In *Symposium of Protein Structure*. Neuberger, A (ed.), pp 23–24, Methuen, London.

Mayo, SL, and Baldwin, RL (1993). "Guanidinium chloride induction of partial unfolding in amide proton exchange in RNase A." *Science*

262, 873–876.
Miller, DW, and Dill, KA (1995). "A statistical mechanical model for hydrogen exchange in globular proteins." *Methods* **4**, 1860–1873.
Mohana-Borges, R, Goto, NK, Kroon, GJ, Dyson, HJ, and Wright, PE (2004). "Structural characterization of unfolded states of apomyoglobin using residual dipolar couplings." *J. Mol. Biol.* **340**, 1131–1142.
Nakanishi, M, Tsuboi, M, and Ikegami, A (1973). "Fluctuation of the lysozyme structure. II. Effects of temperature and binding of inhibitors." *J. Mol. Biol.* **75**, 673–682.
Oliveberg, M, and Fersht, AR (1996). "Thermodynamics of transient conformations in the folding pathway of barnase: reorganization of the folding intermediate at low *p*H." *Biophys. J.* **35**, 2738–2749.
Ormo, M, Cubitt, AB, Kallio, K, Gross, LA, Tsien, RY, and Remington, SJ (1996). "Crystal structure of the *Aequorea victoria* green fluorescent protein." *Science* **273**, 1392–1395.
Otting, G, Liepinsh, E, and Wuthrich, K (1991). "Protein hydration in aqueous solution." *Science* **254**, 974–980.
Pace, CN, Laurents, DV, and Erickson, RE (1992). "Urea denaturation of barnase: *p*H dependence and characterization of the unfolded state." *Biophys. J.* **31**, 2728–2734.
Palmer, AG, Cavanagh, J, Wright, PE, and Rance, M (1991). "Sensitivity improvement in proton-detected 2-dimensional heteronuclear correlation NMR-spectroscopy." *J. Mater. Res.* **93**, 151–170.
Reid, BG, and Flynn, GC (1997). "Chromophore formation in green fluorescent protein." *Biophys. J.* **36**, 6786–6791.
Religa, TL, Markson, JS, Mayor, U, Freund, SM, and Fersht, AR (2005). "Solution structure of a protein denatured state and folding intermediate." *Nature (London)* **437**, 1053–1056.
Robic, S, Guzman-Casado, M, Sanchez-Ruiz, JM, and Marqusee, S (2003). "Role of residual structure in the unfolded state of a thermophilic protein." *Proc. Natl. Acad. Sci. U.S.A.* **100**, 11345–11349.
Schanda, P, and Brutscher, B (2005). "Very fast two-dimensional NMR spectroscopy for real-time investigation of dynamic events in proteins on the time scale of seconds." *J. Am. Chem. Soc.* **127**, 8014–8015.
Schleucher, J, Schwendinger, M, Sattler, M, Schmidt, P, Schedletzky, O, Glaser, SJ, Sorensen, OW, and Griesinger, C (1994). "A general enhancement scheme in heteronuclear multidimensional NMR employing pulsed field gradients." *J. Biomol. NMR* **4**, 301–306.
Seifert, MH, Georgescu, J, Ksiazek, D, Smialowski, P, Rehm, T, Steipe, B, and Holak, TA (2003). "Backbone dynamics of green fluorescent protein and the effect of histidine 148 substitution." *Biophys. J.* **42**, 2500–2512.
Shortle, D, and Ackerman, MS (2001). "Persistence of native-like topology in a denatured protein in 8 M urea." *Science* **293**, 487–489.
Shortle, D, Chan, HS, and Dill, KA (1992). "Modeling the effects of mutations on the denatured states of proteins." *Protein Sci.* **1**, 201–215.
Takei, J, Pei, W, Vu, D, and Bai, Y (2002). "Populating partially unfolded forms by hydrogen exchange-directed protein engineering." *Biophys. J.* **41**, 12308–12312.
Tsien, RY (1998). "The green fluorescent protein." *Annu. Rev. Biochem.* **67**, 509–544.
Wagner, G, and Wuthrich, K (1982). "Amide protein exchange and surface conformation of the basic pancreatic trypsin inhibitor in solution." *J. Mol. Biol.* **160**, 343–361.
Wirmer, J, Schlorb, C, Klein-Seetharaman, J, Hirano, R, Ueda, T, Imoto, T, and Schwalbe, H (2004). "Modulation of compactness and long-range interactions of unfolded lysozyme by single point mutations." *Angew. Chem., Int. Ed.* **43**, 5780–5785.
Woodward, C, Simon, I, and Tuchsen, E (1982). "Hydrogen exchange and the dynamic structure of proteins." *Mol. Cell. Biochem.* **48**, 135–160.
Yan, S, Kennedy, SD, and 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.
Yang, F, Moss, LG, and Phillips, GN, Jr. (1996). "The molecular structure of green fluorescent protein." *Nat. Biotechnol.* **14**, 1246–1251.
Zhang, X, Xu, Y, Zhang, J, Wu, J, and Shi, Y (2005). "Structural and dynamic characterization of the acid-unfolded state of hUBF HMG box 1 provides clues for the early events in protein folding." *Biophys. J.* **44**, 8117–8125.
Zhang, Y-Z, and Roder, H. "SPHERE: a server program for hydrogen exchange rate estimation."
Zimmer, M (2002). "Green fluorescent protein (GFP): applications, structure, and related photophysical behavior." *Chem. Rev. (Washington, D.C.)* **102**, 759–781.