Understanding the folding of GFP using biophysical techniques

Sophie E Jackson†, Timothy D Craggs and Jie-rong Huang

Green fluorescent protein (GFP) and its many variants are probably the most widely used proteins in medical and biological research, having been extensively engineered to act as markers of gene expression and protein localization, indicators of protein–protein interactions and biosensors. GFP first folds, before it can undergo an autocatalytic cyclization and oxidation reaction to form the chromophore, and in many applications the folding efficiency of GFP is known to limit its use. Here, we review the recent literature on protein engineering studies that have improved the folding properties of GFP. In addition, we discuss in detail the biophysical work on the folding of GFP that is beginning to reveal how this large and complex structure forms.

Expert Rev. Proteomics 3(5), 545–559 (2006)

CONTENTS

†Author for correspondence
Chemistry Department,
Lensfield Road,
Cambridge CB2 1EW, UK
Tel.: +44 1223 762 011
Fax: +44 1223 336 362
sej13@cam.ac.uk

KEYWORDS:
aggregation, denatured state, equilibrium intermediate, kinetic intermediate, misfolding, oligomeric state, protein folding selection

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 [1,2]. 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 encapsulated conjugated π system, the chromophore, which emits green fluorescence (FIGURE 1) [3]. No cofactors are necessary for either the formation or the function of the chromophore [4], which is embedded in the interior of the protein surrounded by an 11-stranded β-barrel (FIGURE 2) [5,6]. GFP is remarkable for its structural stability and high fluorescence quantum yield, a result of the fact that, in the native state, the chromophore is rigid and shielded from bulk solvent. On protein denaturation, the chromophore remains chemically intact but fluorescence is lost. Therefore, the green fluorescence is a sensitive probe of the state of the protein. It is clear that 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. This review focuses on recent engineering studies aimed at improving the folding properties of GFP, as well as recent studies using a range of biophysical techniques to characterize the folding pathway of this complex and important protein.

Improving the folding properties of GFP: engineering studies

Wild-type GFP is very prone to misfolding and aggregation when expressed in *Escherichia coli* [1] and, therefore, there has been considerable effort in engineering variants of GFP that have better folding properties. In addition to the intrinsic tendency of GFP to misfold and aggregate, fusions of GFP with other proteins frequently show reduced folding yields [7]. Furthermore, circular permutants of GFP are often employed as biosensors and these also show a strong tendency to misfold and aggregate [8,9,10]. The need for better folding fluorescent proteins, particularly at higher temperatures, has led to a number of studies that have isolated so-called 'folding' mutants. These are mutants of GFP that are brighter *in vivo*, usually as a result of more efficient folding. However, in some cases they may act by improving

10.1586/14789450.3.5.545

ISSN 1478-9450

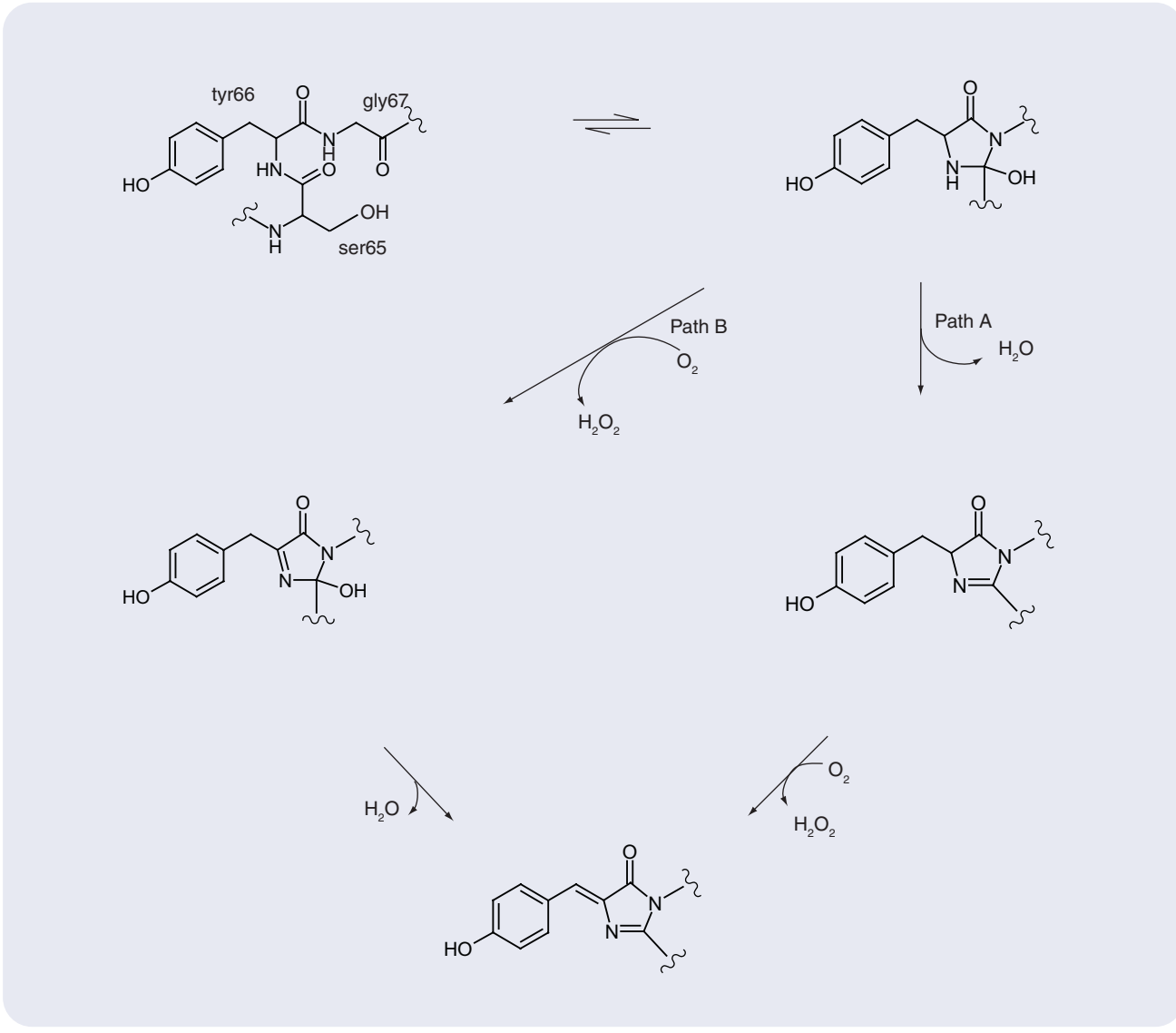

Figure 1. Mechanisms for the formation of and the structure of the *p*-hydroxy-benzylideneimidazolidinone chromophore of GFP.

chromophore formation (see section on chromophore formation). In general, these mutants suppress misfolding and aggregation, rather than accelerating folding. The results from a number of different laboratories are discussed below.

One of the earliest studies in this area was from the Haselof group who combined random mutagenesis and screening of *E. coli* colonies for increased brightness to obtain a mutant of GFP which showed a 35-fold increase in green fluorescence intensity when expressed in *E. coli* and yeast [11]. In this case, two mutations were identified, Val→Ala163 and Ser→Gly175, which were found to suppress the tendency of GFP to misfold and aggregate at 37°C [11]. In a separate study by the Kohno group, another mutation, Ser→Pro147, which enhances green fluorescence at elevated temperatures, was also identified [12].

Perhaps the most important early study on GFP folding mutations is that by the Stemmer group, who isolated the so-called 'cycle 3' mutant also termed GFPuv, which is 42 times more fluorescent than wild-type GFP and is now used extensively in a variety of applications. This variant was obtained using fluorescence screening of a library of GFP mutants created by DNA shuffling techniques [13]. It contains three mutations (Phe→Ser99, Met→Thr153, Val→Ala163) and its structure and folding kinetics have been studied [14,15]. The three mutations lie on the surface of the protein in three different β-strands. While the side chains of Ser99 and Thr153 are exposed, the side chain of Ala163 is buried. It was found to fold with double exponential kinetics with rates very similar to wild type, thus establishing that its enhanced fluorescence *in vivo* is not a result of changes in structure or folding. In a more detailed study of this mutant, Kuwajima and coworkers demonstrated that although its unfolding and refolding kinetics are very similar to the wild type, the cycle-3 mutant is much less prone to aggregation [15]. The mutations, which all lie on the same face of the β-barrel, reduce the overall hydrophobicity of GFP and, thereby, suppress aggregation.

In a recent publication, the Waldo group has used DNA shuffling techniques to create a library of GFP and DsRed mutants fused to a poorly folding bait protein, in this case, bullfrog red cell H-subunit of ferritin, an insoluble protein when expressed at 37°C [16]. Using this 'folding interference method', they went through four rounds of selection and obtained a 'superfolder' GFP that has six additional mutations to the parent GFP – the cycle 3 'folding reporter' GFP. A comprehensive characterization of this superfolder GFP demonstrated that it not only has enhanced folding properties *in vivo*, but shows improved tolerance to circular permutation, chemical denaturants and faster folding kinetics. In addition, this superfolder GFP was mutated to create superfolder CFP, BFP and YFP. Using the same method, a superfolder DsRed was also isolated but not characterized as extensively. Each of the single mutations in superfolder GFP was made in the parent folding reporter GFP to assess the effects on stability and folding. The mutations were found to have different effects, Ser→Arg30 and Tyr→Asn39, folded faster than the parent, and were more stable towards chemical denaturants. A crystal structure showed that a reorganization of side chains around these mutations resulted in increased favorable electrostatic interactions or hydrogen bonding networks. In contrast, Tyr→Phe145 and Ile→Val171 had little effect on folding rates or stability and presumably act by reducing misfolding and aggregation. The final mutations Asn→Thr105 and Ala→Val206 did not show increased folding rates or stability, however, both mutations increase the β-forming propensity of the polypeptide chain.

The results from these four studies are summarized in FIGURE 3, which shows the position of the folding mutations in the GFP structure, and in TABLE 1, which summarizes the effect of the mutations.

Chromophore formation

In order to form the mature chromophore, the polypeptide backbone must undergo four distinct processes: folding, cyclization, oxidation and dehydration [17,18]. Peptide cyclization is initiated by nucleophilic attack of the Gly67 amide nitrogen on the Ser65 carbonyl carbon, forming an imidazolone ring. Dehydration of the Ser65 carbonyl oxygen and dehydrogenation of the Thr66 Cα–Cβ bond produces the fully conjugated, *p*-hydroxybenzylidene-imidozolidinone chromophore (FIGURE 1).

Initial studies by Reid and Flynn established that oxidation is the rate-limiting step in chromophore formation and confirmed that it is an autocatalytic process [4]. Since then, many studies have focused on the detailed mechanism of chromophore formation, and the involvement of residues outside the chromo-tripeptide. In particular, Arg96 and Glu222 have been found to be involved in the autocatalysis of chromophore formation [19,20].

Two mechanisms for chromophore formation have been proposed. In the first, cyclization is followed by dehydration and oxidation. In this mechanism, the heterocycle formed after cyclization is stabilized by dehydration, and then dehydrogenation of the Tyr66 Cα-Cβ occurs to form the fully conjugated chromophore [18,21]. In the second mechanism, it has been proposed that the oxidation step precedes the dehydration step,

Figure 2. Ribbon diagram of the structure of GFP as determined by x-ray crystallography [5,6]. The *p*-hydroxy-benzylideneimidazolidinone chromophore is located in the central α-helix and is inaccessible to solvent.

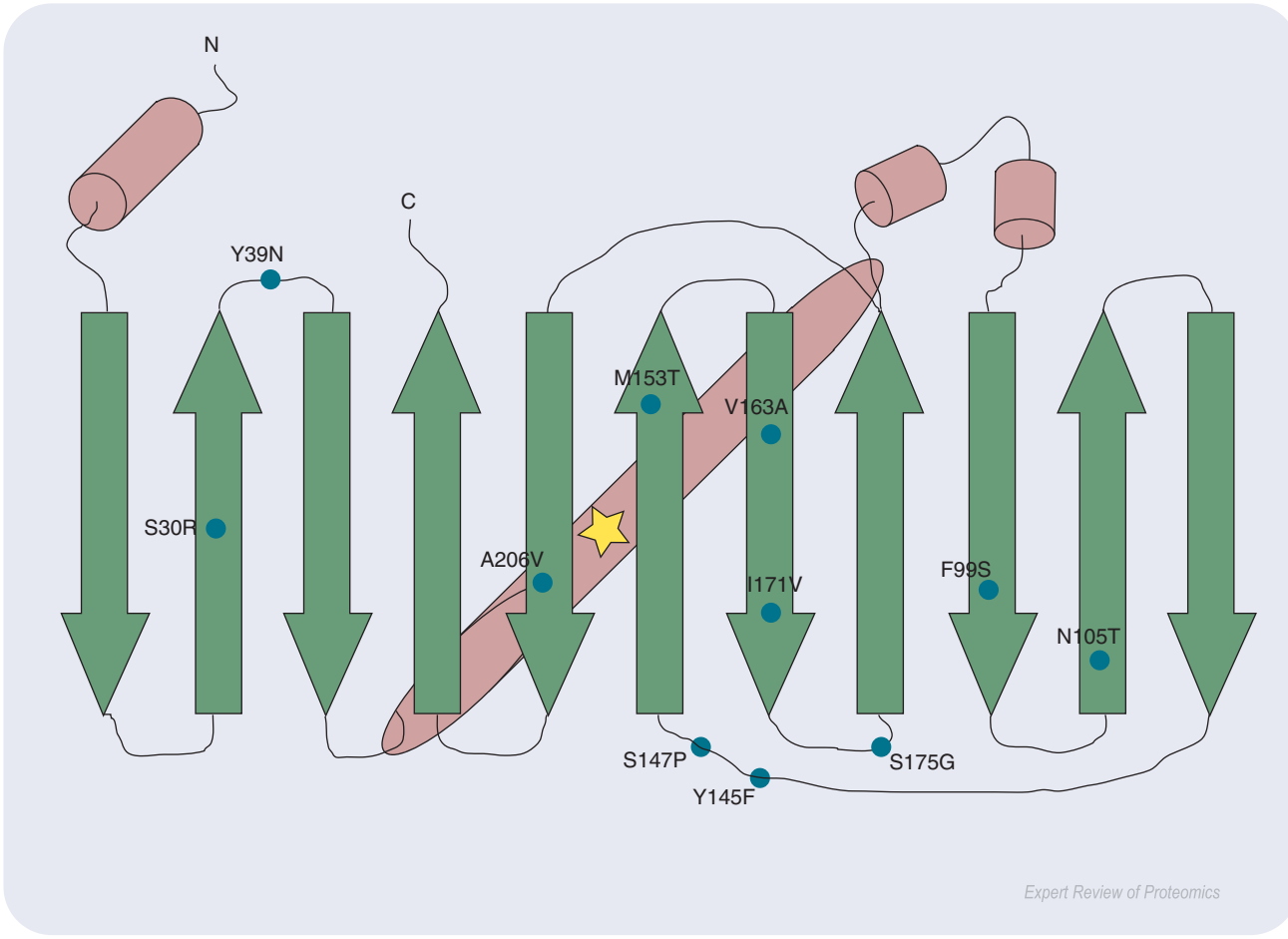

Figure 3. Topological map of green fluorescent protein showing the elements of secondary structure and the positions of the folding mutants.

This is supported by structural studies on the Y66L variant of GFP wherein a trapped intermediate was observed in which cyclization had occurred, and in which the hydroxyl leaving group remained attached to the heterocyclic ring [22]. However, the α-carbon of residue 66 was shown to be trigonal planar, consistent with ring oxidation by molecular oxygen. Further evidence in support of this mechanism was obtained from kinetic studies on chromophore formation and the concomitant production of H_2O_2 [23]. In this study, the Wachter group reported time constants for three kinetic steps. The first step, involving folding and peptide cyclization proceeded with a time constant of 1.5 min. The second step, corresponding to the oxidation, which was found to be rate limiting, proceeded with a time constant of 34 min, whilst the final step proceeded with a time constant of 11 min. Under highly aerobic conditions, it was proposed that the dominant path to chromophore formation follows the cyclization–oxidation–dehydration mechanism. Both mechanisms may occur in parallel, the relative flux being dependent on oxygen concentration and the efficiency of ring dehydration for the particular GFP variant.

In contrast to the above mechanisms, one computational DFT study suggested that oxidation could precede cyclization [24]. Getzoff and coworkers have interpreted these data slightly differently and suggested that although the cyclization reaction appears thermodynamically unfavorable (consistent with the relative thermodynamic stabilities calculated by DFT [24] it still occurs first and is then trapped by the dehydration of the ring [21].

Both computational methods [25] and x-ray crystallography [21] have shown that the central α-helix exhibits a dramatic approximately 80° bend during chromophore formation. The resultant strained structure is proposed to raise the energy of the precyclized state closer to that of the cyclized intermediate, hence reducing the activation energy. It also serves to position the Gly67 nitrogen (the nucleophile) and the Ser65 carbonyl oxygen in close contact, priming the cyclization step.

A recent computational study from the Zimmer group has proposed that the cycle-3 mutations and the Ser→Pro147 mutant exhibit increased fluorescence at room temperature due to the formation of a tighter turn than wild type in the precyclized protein around residues 65–67 [26]. Thus, these mutations may improve the rate of chromophore maturation, in addition to reducing the overall hydrophobicity, and hence aggregation propensity.

Preventing dimerization

Wild-type GFP is known to dimerize at high concentrations [6]. In their fluorescence resonance energy transfer (FRET) study of lipid rafts, Zacharias and coworkers measured the homoaffinity

of YFP by sedimentation equilibrium analytical ultracentrifugation and found a K_d of 0.11 mM [27]. By replacing hydrophobic residues at the crystallographic interface of the dimer with positively charged residues (A206K, L221K or F223R), they went on to engineer mutants in which dimerization was essentially eliminated. The A206K mutant was so extremely monomeric in nature that it was difficult to determine an accurate dissociation constant for a hypothetical dimer. This mutation has now been introduced into the full range of fluorescent proteins [2].

High-energy barriers & slow equilibrium

In their study of the cycle-3 variant, Fukuda and colleagues, established that the unfolding and refolding of GFP was slow and, as a consequence, the unfolding equilibrium is reached over a period of days (rather than an hour or less for smaller proteins) and the protein appears to be very stable with respect to chemical denaturants [15]. These results have now been corroborated by a number of different groups that have also shown that GFP unfolds and refolds very slowly compared to small, monomeric proteins [28]. FIGURES 5 & 6 shows the rate at which the unfolding equilibrium is reached for GFP at different temperatures and pH, as measured in our laboratory [29]. Even at 37°C, a true equilibrium is reached only after several weeks (FIGURE 6). Careful analysis of these data to two- and three-state models reveals that there is a stable intermediate state populated under equilibrium conditions [29]. The thermodynamic parameters obtained from the analyses shows that the intermediate state is compact compared to the denatured state; however, there has still been a significant increase in the solvent accessible surface area on unfolding of the native to the intermediate state. Although the intermediate state has very little green fluorescence (approximately 10% of the native state, consistent with the access of water to the chromophore) it is still remarkably stable with respect to the denatured state, with a free energy of unfolding of over 10 kcal/mol at 25°C, pH 6 The data are consistent with an intermediate state in which β-strands 7–9 have unfolded and exposed the chromophore but in which the rest of the β-barrel structure remains intact [29].

The fact that GFP reaches an unfolding equilibrium only very slowly is indicative of high-energy barriers for both the folding and unfolding reactions. The rate constants for folding and unfolding have been measured by a number of different groups and are consistently found to be small compared with those measured for small, monomeric proteins [28].

Regan and coworkers have used the β-barrel structure of GFP to study the effect of pairs of interacting residues across parallel β-strands on stability and folding [30]. Positions 17 (β1) and 122 (β4) were mutated using library cassette mutagenesis methods, and a series of mutants produced and analyzed with different cross-strand pairs. Unfolding and folding rate constants were measured *in vitro* using pH-jump experiments and, under the experimental conditions used, unfolding half-lives in the order of 3 minutes and refolding half lives in the order of 1–8 min were observed. In addition to the *in vitro* measurements, the rate of maturation of wild-type and mutant GFPs were measured *in vivo*. Different rates for the maturation of GFP were observed for the mutants *in vivo* and *in vitro* a ten-fold range in folding rates was observed but differences in the unfolding rates were undetectable. In this case, wild type was found to have the highest folding rate and was the most fluorescent in cells. The results established that there is a correlation between folding rates measured *in vitro* and levels of intracellular fluorescence.

Table 1. Structural and folding parameters for GFP folding mutants.

Mutation	Position	Properties	Ref.
S175G	Loop exposed	Suppresses aggregation	[11]
V163A	β-strand 8 buried	No effect on folding rates, reduces hydrophobicity.	[11,13]
S147P	Loop buried	Increased maturation rate	[12,26]
M153T	β-strand 7 exposed	No effect on folding rates, reduces hydrophobicity.	[13]
F99S	β-strand 4 exposed	No effect on folding rates, reduces hydrophobicity.	[13]
S30R	β-strand 2 exposed	Mutation stabilizes protein, faster folding.	[16]
Y39N	Loop exposed	Mutation stabilizes protein, faster folding.	[16]
N105T	β-strand 5 exposed	No effect on stability or folding, increased β propensity.	[16]
Y145F	Loop buried	No effect on stability or folding, presumably reduces aggregation.	[16]
I171V	β-strand 8 buried	No effect on stability or folding, presumably reduces aggregation.	[16]
A206V	β-strand 10 buried	No effect on stability or folding, increased β– propensity.	[16]

In another study, GFP has been cyclized using intein technology and the effects on unfolding and refolding measured [31]. The cyclic variant is identical in sequence and structure to the linear parent GFP, except that the N- and C-terminals are covalently linked together through a short region of peptide. Unfolding half lives of approximately 0.1–0.6 min were measured directly in high concentrations of the chemical denaturant guanidinium chloride. These data were extrapolated to extract an unfolding half life in water in the order of 3 min consistent with the Regan results. In this case, two phases, fast and slow, were identified in the refolding reactions with half lives of 1–2 and 40 min, respectively. The cyclized GFP was found to be more stable than the linear parent GFP and unfold at approximately half the rate.

Careful detection and analysis of the refolding of GFP as probed by green fluorescence shows that there are more than two kinetic phases. FIGURE 5 shows the results of a double

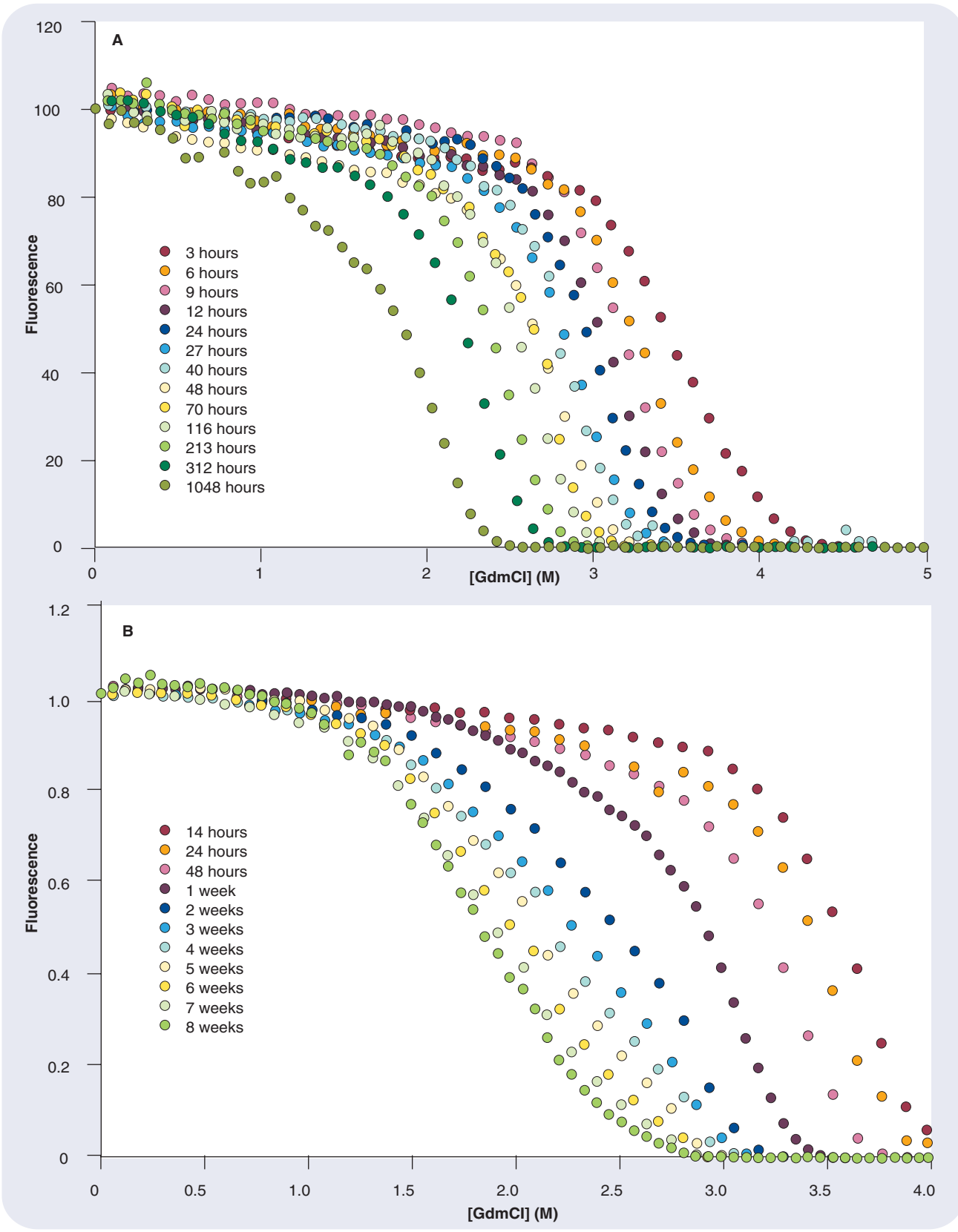

Figure 4. Equilibrium denaturation studies of GFP. Guanidium chloride denaturation of GFP monitored by green fluorescence. (A) at 25°C and pH 6.0 (B) at 37°C and pH 7.5. Craggs, Huang andJackson [UNPUBLISHED RESULTS].
GFP: Green fluorecent protein.

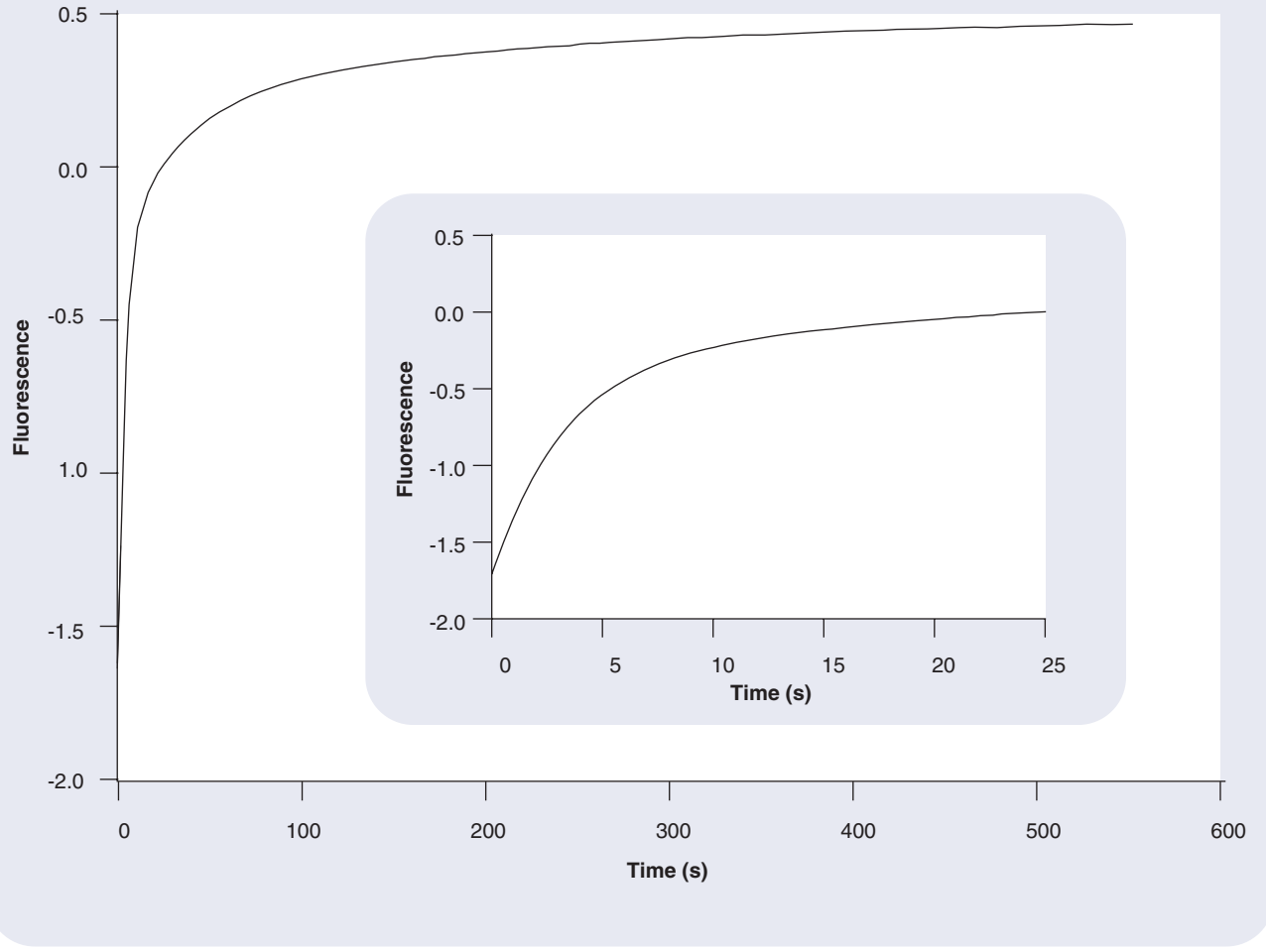

Figure 5. Kinetics of folding of GFP. Refolding kinetics of GFP monitored by green fluorescence in a stopped-flow apparatus. Refolding is initiated by a rapid pH jump from pH 1.5 to 6.3. Craggs & Jackson, [UNPUBLISHED RESULTS]. Inset shows the faster phases observed between 0 – 25 s. Data fit to a triple exponential process.
GFP: Green fluorescent protein.

pH-jump experiment performed in our laboratory that shows at least three distinct kinetic phases. Recently, Kuwajima and coworkers have undertaken the most comprehensive study of the folding of GFP to date [32]. Multiple probes including the green fluorescence, tryptophan fluorescence and far-UV CD were employed to reveal five folding phases including a rapid-burst phase. Half-lives for these phases range from 20 ms to 6 min under the conditions used. A complex kinetic scheme has been proposed by the Kuwajima group based on their results (FIGURE 6) in which there is heterogeneity in the denatured state due to proline isomerisation; there are several intermediate states including the rapidly formed 'burst-phase' intermediate, in which there is a nonspecific collapse of the polypeptide chain, and an on-pathway intermediate with molten-globule-like properties; that there are at least two slow phases which are limited by proline isomerisation. A structure for the second intermediate state is proposed based on their results and those from other groups: the intermediate is known to be compact with significant secondary structure, but it does not show green fluorescence or rigid tertiary structure. This late intermediate identified from kinetic experiments appears to be similar to the equilibrium intermediate observed in our studies [29].

In a paper just published online, the Kuwajima group have reported an equilibrium intermediate that is populated at pH 4 [33]. This intermediate was shown to be the same as one of the intermediates detected during refolding and through a combination of fluorescence and small-angle x-ray scattering (SAXS) experiments shown to have properties similar to a molten-globule state.

Oligomeric fluorescent proteins

The folding behaviour of other GFP-like fluorescent proteins (FPs) has also been investigated. DsRed is a FP from the coral *Discosoma* and is a homotetramer of β-barrels whose fluorescence is red shifted compared to GFP. The acid denaturation of DsRed was measured and partial renaturation achieved on alkalization [34]. Several distinct states of the protein were

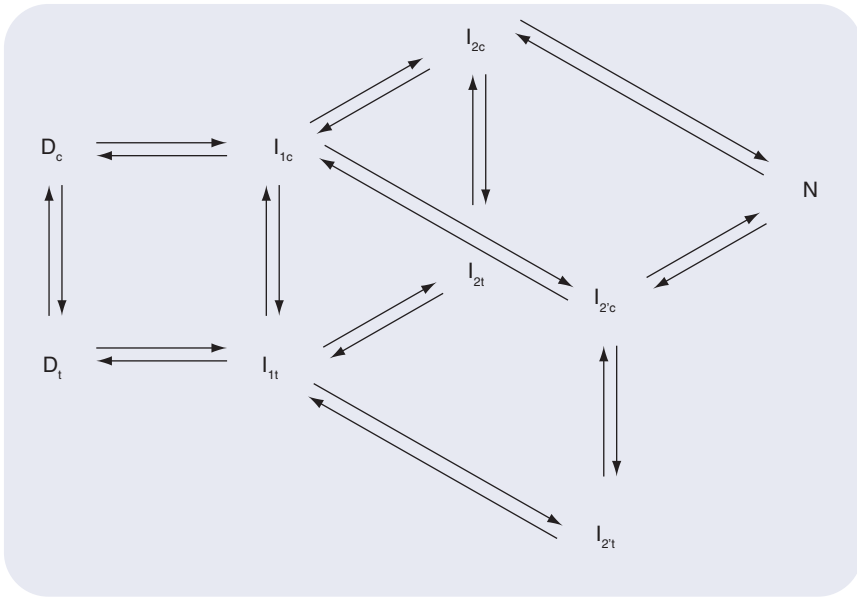

Figure 6. Kinetics of folding of GFP. Kinetic scheme for the folding of green fluorescent protein as proposed by Kuwajima and coworkers [32]. D_c and D_t are the denatured states with *cis* and *trans* proline isomers, N is the native state, and I_{1c}, I_{2c}, $I_{2'c}$, I_{1t}, I_{2t}, $I_{2't}$ are the different intermediate states populated.

found during the unfolding and refolding processes corresponding to different oligomeric states – monomer, dimer, trimer and tetramer.

This work was followed up by a comparative study of the folding and stability of five different FPs with varying oligomeric states and spectral properties [35]. Folding and unfolding kinetics were measured in addition to stability measurements using a wide range of probes. For all five proteins, the kinetics were found to be slow in agreement with other studies and a quasi-equilibrium produced. Despite some limitations, this study clearly showed that there is a wide range in stabilities of FPs, in general the higher-order oligomers being more stable. However, this study also demonstrated that FPs with the same oligomeric state can have very different stabilities.

Partially structured & denatured states of GFP

A pressure-induced unfolding study of red-shifted GFP (rsGFP) has combined different optical probes of the native state of the protein – fluorescence and absorbance measurements probing tertiary structure whilst FT-IR was employed to monitor changes in secondary structure [36]. Although very stable to high pressures at lower temperatures, and unfolding irreversibly at high temperatures and high pressures, conditions were found where rsGFP was reversibly unfolded by pressure. Two transitions were revealed by the different probes, the first at about 4 kbar where changes associated with the tertiary structure were observed and attributed to penetration of water into the β-can structure, particularly in the region of the chromophore. This creates a 'swollen pretransitional' state, which has relatively small changes in tertiary structure and which retains its secondary structure. The second transition, observed at higher pressures (~8 kbar), represents a global unfolding of the protein with a loss of secondary structure of the β-barrel. Interestingly, the helical structure of GFP seems to be maintained in the denatured state under these conditions.

In addition to the studies described above, there is further evidence for residual structure in the denatured state of GFP. In their analysis of the acid denaturation and refolding of GFP, the Kuwajima group found that there is significant secondary structure in the acid-unfolded state as shown by far-UV circular dichroism and this residual structure was shown to be sensitive to salt [32]. We have investigated this residual structure further using ^{19}F-NMR spectroscopy in combination with photochemically induced dynamic nuclear polarisation (CIDNP) techniques. The ^{19}F spectrum of ^{19}F-tyrosine-labeled GFP is shown in FIGURE 8 along with the full assignment. The assignment was made by combining the results of ^{19}F-NMR spectra of single Tyr→Phe mutations, relaxation data and results from photo-CIDNP experiments [37]. The ^{19}F-NMR spectrum of denatured GFP is shown in FIGURE 8 and clearly shows two peaks, the larger corresponding to the ^{19}F-labeled tyrosines which are all in a chemically identical environment in the denatured state, the smaller peak corresponding to the single ^{19}F-labeled tyrosine, which has a chemically distinct environment as a result of its position within the chromophore. Although there is little evidence for residual structure from the ^{19}F spectrum of the denatured state, the results from photo-CIDNP experiments provide support for the far-UV CD results. The photo-CIDNP experiments report on the solvent exposure of the tyrosine side chains and FIGURE 8 shows the results for the native state of GFP. Four peaks are observed corresponding to the four solvent exposed tyrosines. Interestingly, a strong correlation is found between the solvent accessibility of the highest occupied molecular orbital (HOMO) of a given tyrosine, and its photo-CIDNP signal rather than the solvent accessible surface area (SASA) of the residue. The SASA data suggest that five, rather than four fluorotyrosine residues should be polarizable in GFP (i.e., Tyr39, Tyr151, Tyr182, Tyr200 and Tyr143). Although Tyr143 has a very similar overall solvent accessibility to Tyr200, the HOMO accessibilities differ by an order of magnitude. Inspection of the crystal structure shows that the part of Tyr143 that protrudes into the solvent is an unreactive $C\beta H_2$-$C\alpha$H-NH fragment that bears little HOMO electron density, whereas it is the reactive aromatic side chain that is exposed for Tyr200 [5,6].

The photo-CIDNP spectrum of the acid denatured state of GFP shows two peaks consistent with the ^{19}F spectrum of the acid denatured state. However, in this case and in contrast to

the results on the native state, the signs of the two peaks are opposite. This provides strong evidence that the denatured state is heterogeneous, containing subensembles with significantly different rotational correlation times [37].

Dynamics

The dynamics of GFP have been measured using a variety of experimental approaches. The dynamics associated with the chromophore have been probed by fluorescence correlation spectroscopy (FCS) [38]. In this technique, time-resolved fluctuations in fluorescence are used to report on the dynamic and thermodynamic processes that affect the fluorescence of the protein (in this case, enhanced GFP, eGFP). Protonation of the hydroxyl group of Tyr66 is shown to induce large changes in absorption and emission spectra with a pK_a of 5.8. The autocorrelation function of fluorescence emission shows contributions from two chemical relaxation processes, one pH dependent and the other pH independent. The FCS data provide information on the dynamics and equilibrium properties of the protonation process.

GFP and its variants, like all known fluorescent proteins exhibit complex photophysical and photochemical behavior. This interesting area falls outside the scope of this review but the interested reader is directed to [39] as a good starting point for further reading.

^{15}N-NMR relaxation measurements have been used to study the dynamics of GFP on a ps–ns timescale and have shown that most of the β-barrel backbone is rigid on these timescales [40]. H/D exchange techniques were also employed to study conformational dynamics on a μs–ms timescale. The rates of exchange were found to vary enormously and were assigned to four classes – fast, intermediate, slow and very slow. The slowest exchanging amide protons did not show significant levels of exchange over the time course of the experiment. These studies identified a region comprising of β-strands 7, 8 and 10 that show increased rates of exchange compared to the rest of the protein and indicate that this region has a higher degree of flexibility in agreement with molecular dynamic simulations. The spectra of a mutant of GFP (His-Gly148 located on β-strand 7 and known to affect the chromophore) showed interesting additional peaks showing that this mutant is in slow exchange between two conformations.

Single-molecule unfolding & folding

Single-molecule force spectroscopy has been used to investigate the mechanical unfolding of GFP [41]. Here, GFP is sandwiched into a multidomain construct with either Ig8 or DdFLN and the unfolding of the chimeric protein studied with pulling Atomic Force Microscopy techniques. The results suggest that GFP mechanically unfolds via two intermediate states, the first is characterized by the detachment of the seven-residue N-terminal α-helix to form a kinetically stable but thermodynamically unstable state that retains the β-barrel structure. The second metastable intermediate state has one

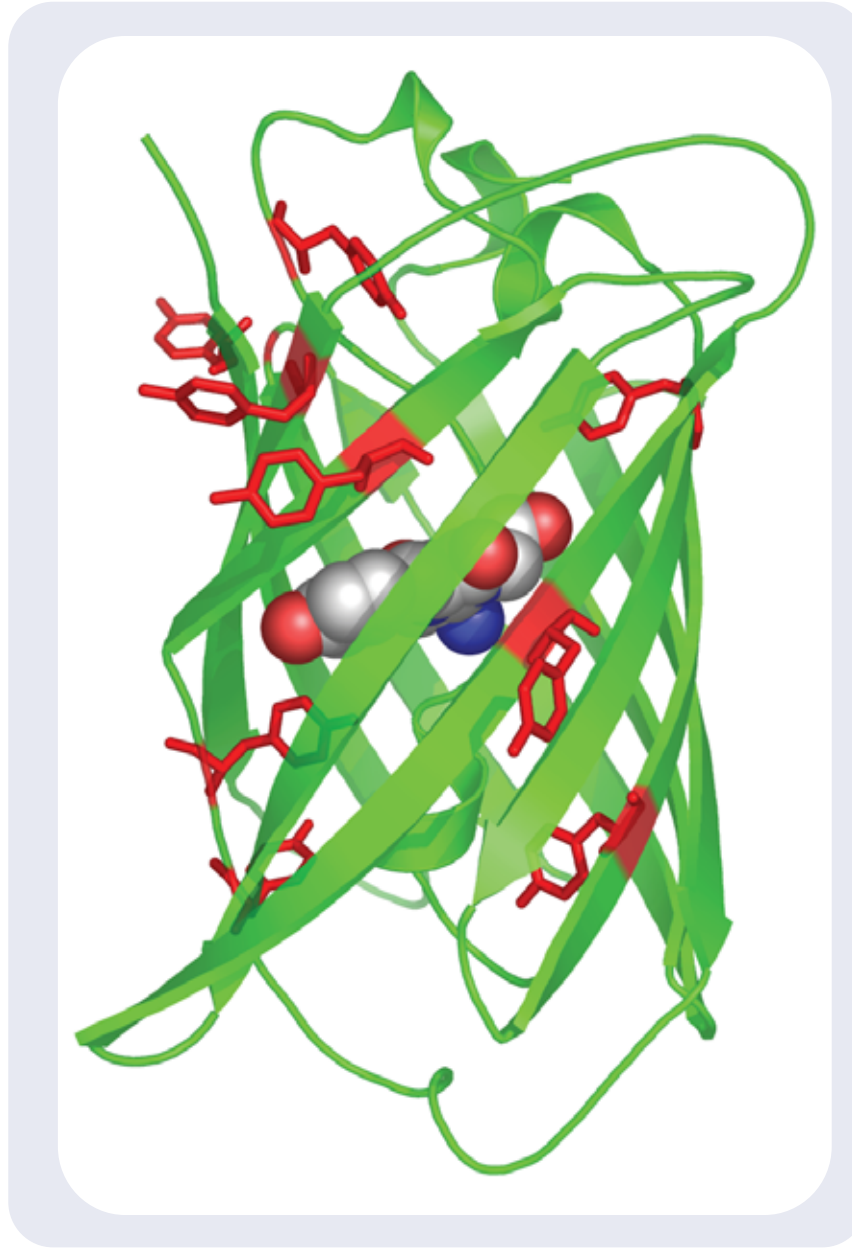

Figure 7. Results from the ^{19}F NMR studies on GFP [37]. The structure of GFP is shown with the side chains of the ten tyrosine residues shown in red. The chromophore is shown in space filling mode in the centre of the β-barrel.

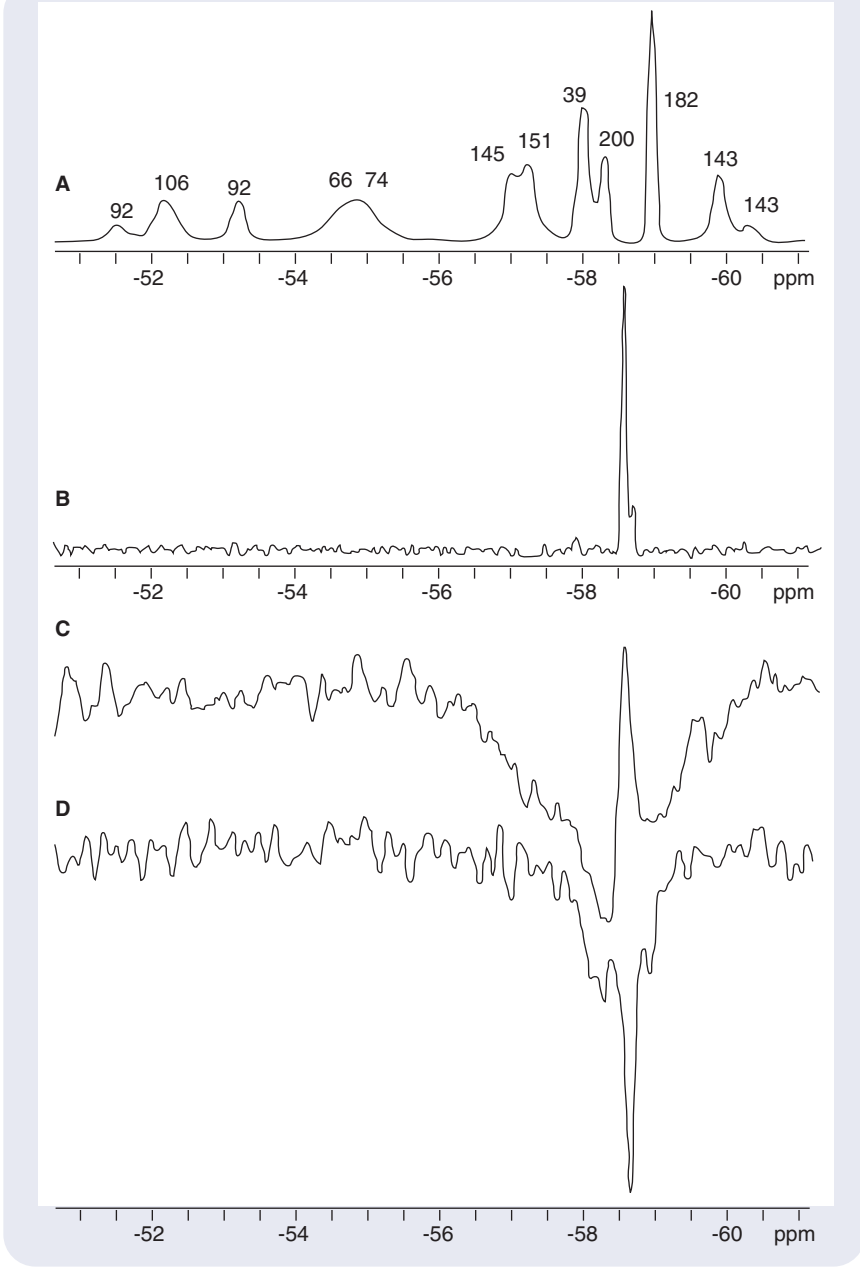

Figure 8. Results from the ^{19}F NMR studies on GFP [37]. (A) ^{19}F-NMR spectrum of labeled wild-type GFP with the assignments shown. (B) Photo-CIDNP spectrum of labeled wild-type GFP denatured in 6 M GdmCl. (C) Photo-CIDNP spectrum of labeled wildtype GFP denatured at pH 2.9 and (D) Photo-CIDNP spectrum of labeled wild-type GFP in native buffer conditions.
CIDNP: Chemically induced dynamic nuclear polarisation; GFP: Green fluorescent protein.

complete β-strand detached from the barrel. A schematic of the free energy surface of GFP and the mechanical unfolding pathway are shown in FIGURE 9.

Owing to its unique spectroscopic properties and high quantum yield, GFP and its variants have been the subject of many single-molecule experiments using optical fluorescence techniques. In general, Yellow Fluorescent Proteins (YFPs), created by the substitution of Thr203 to an aromatic amino acid, have been used due to their optical properties. They have been shown to exhibit interesting photophysical properties, including on/off blinking [42] and flickering [43]. The folding of one GFP variant, the GFPmut2 construct, has been investigated at the single-molecule level by encapsulation in wet nanoporous silica gels [44].

In collaboration with the group of David Klenerman, our own group has recently undertaken single-molecule studies of variants of GFP under non-equilibrium conditions where we can monitor both unfolding and folding reactions. A confocal microscope is used in conjunction with novel nanopipette technology [45] to observe both equilibrium behaviour and unfolding kinetics of the YFP, Citrine [46] labeled with an acceptor dye, Alexa 647, by single-pair FRET (sp-FRET) and dual-colour single molecule fluorescence coincidence spectroscopy (sm-FCS). The citrine mutant was chosen for its increased photostability compared with other YFPs [2].

Initial single-molecule kinetic unfolding studies were conducted by injection of native, labeled citrine into varying concentrations of guanidinium chloride (GdmCl) contained in a one millilitre sample chamber. Using diffusion sp-FRET, histograms of FRET efficiencies were generated for each 14 min interval over the time course of the reaction (FIGURE 10). Two populations were observed at each time point, with FRET efficiencies of approximately 0.65 corresponding to folded, labeled Citrine, and 0.00, made up of unlabeled citrine and GdmCl impurities. The unfolding rate constants were obtained by plotting the number of acceptor events against time and fitting to a single exponential decay (FIGURE 11). Alternatively, unfolding rate constants were obtained by plotting the change in the Gaussian fit to the labeled and unlabeled FRET peaks. The labeled and unlabeled peaks decreased with the same rate suggesting that the attachment of the dye has not significantly affected the unfolding rate or the stability of the protein. The unfolding rate was also monitored by sm-FCS, observing the decrease in the number of coincident events (fluorescence above the background count in both the Alexa and the Citrine channel

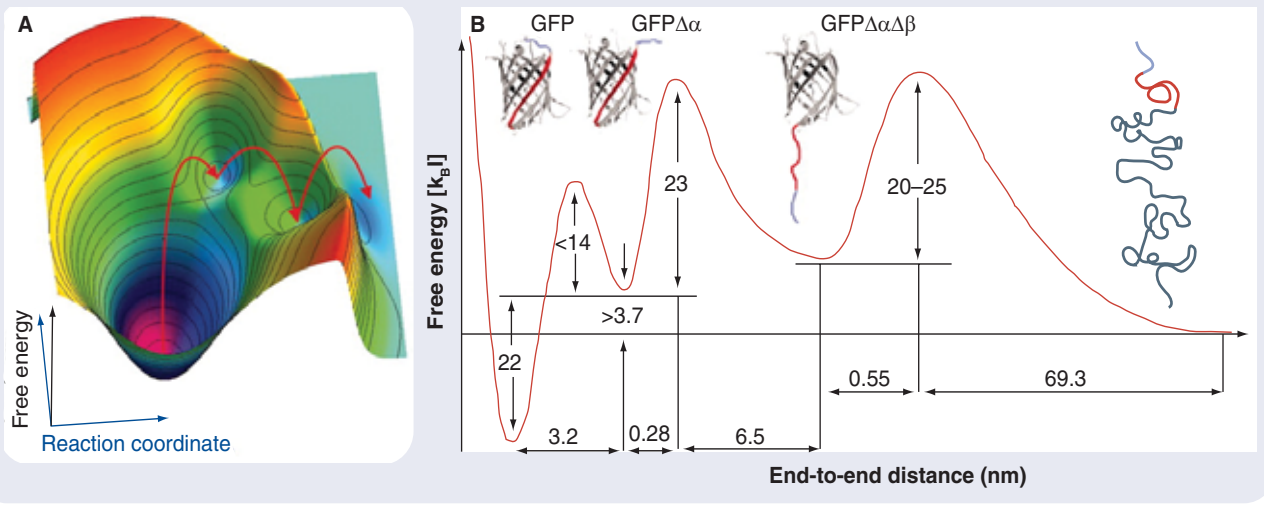

Figure 9. Free-energy surface for the mechanical unfolding of GFP by atomic force microscopy. Taken from [23].

in the same 1 ms bin). The unfolding rate constants from the single-molecule experiments agreed well with results from bulk solution study.

Expert commentary & five-year view

There is no doubt about the importance of GFP in current biological and medical research: a quick literature search on publications using GFP comes up with more than 9000 hits from the year 2000 onwards. Its widespread use results from the inherent, unique spectroscopic properties of GFP, in addition to the comprehensive engineering that has been performed on the protein to modify and optimize optical, chemical and physical properties. An exhaustive review on the protein is just not possible and, in this article, we have focused on the folding properties of the protein. Folding is an essential step in the maturation and use of GFP, but one that is sometimes limited by the competing reactions of misfolding and aggregation. Described in the review are the best results achieved so far to improve the folding properties of the protein. Remarkably, it has been 10 years since the publication of the cycle 3 mutant in 1996 [13], and only recently has a new GFP with improved folding been reported [16]. It is unclear whether this is because it is inherently difficult to generate better folding variants of GFP, or whether selection procedures are limiting in the process.

The recent interest and work done in characterizing the folding pathway of GFP *in vitro*, leads us into an exciting new area – that of rationally designing mutants to aid in the folding of GFP. Once a stable core or folding nucleus has been identified, then mutations can be designed which specifically stabilize this nucleus and increase folding rates. As well as understanding more about the folding pathway, in particular, characterizing the intermediate states in addition to the rate-limiting transition state barriers, we also require a better understanding of how and why this protein misfolds and aggregates. Mutations can then be designed which not only aid folding but also suppress aggregation. Combinations of both types of mutation should be particularly effective in creating a new ultrafolding GFP.

Information resources

Recent review articles

For anyone confused as to which of the many available fluorescent proteins to use in their experiments, a recent review by Shaner, Steinbach and Tsien, is an excellent guide [2]. The review discusses the different factors to be considered when choosing an FP such as spectral properties, brightness, expression, toxicity, photostability, oligomerisation and sensitivity to environmental effects, in addition to summarizing this data for a large number of FPs. A slightly earlier review from the Tsien laboratory also contains a section on the design and construction of FP-based fluorescent reporting systems [47]. In addition, this review describes applications of FPs as 'passive' markers of protein expression and localisation, and as 'active' indicators of small-molecule messenger dynamics, enzyme activation and protein–protein interactions. For a more specific review on using mutants of GFP to monitor protein conformations and interactions by fluorescence resonance energy transfer see review by Miyawaki and Tsien [48]. It must be stressed that there are now numerous reviews on the applications of GFP in biology. The three reviews cited are just a starting point for interested readers.

In addition to review articles, there is a great deal of information also available on GFP on both academic and commercial websites. Highly recommended is Roger Tsien's website which contains a complete list of publications from the lab, as well as images, movies, discussion documents and links.

- www.tsienlab.ucsd.edu/

For those interested in single-molecule studies of GFP, the website of William Moerner is recommended:

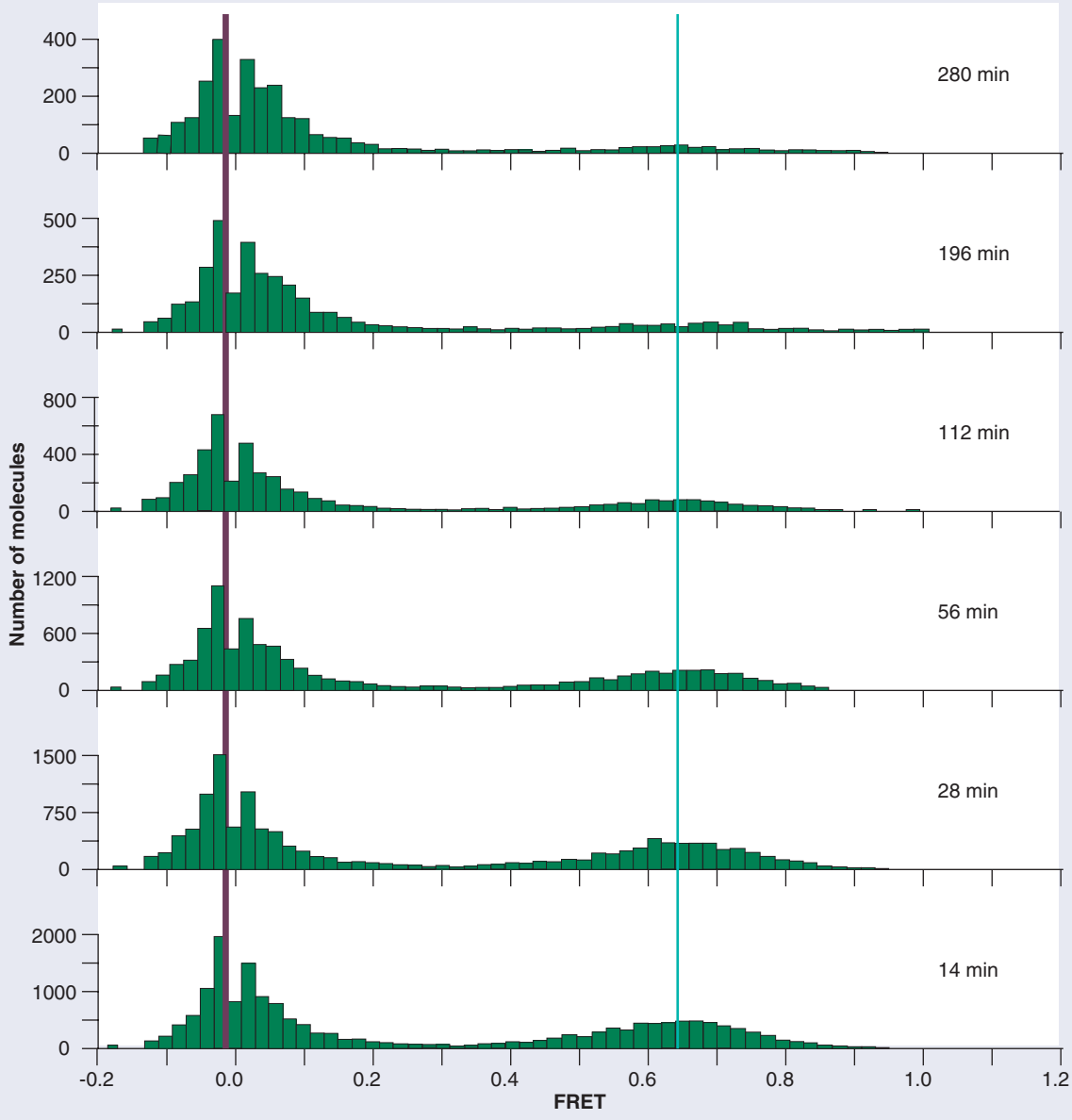

Figure 10. Typical single molecule histograms of YFP unfolding. 50 pM YFP in 4 M GdmCl. Data were acquired at 14 min intervals over 280 min at 25°C. The left- and right-hand lines are lines through the centre of the two Gaussian peaks.

- www.stanford.edu/group/moerner/
- Information on commercially available FPs can be obtained from the Clontech website, in the Living Colours Fluorescent Proteins section.

  www.clontech.com/clontech/

  Recommended books on GFP include:

- GFP, Properties, Applications and Protocols 2nd Ed. Chalfie and Kain, John Wiley & Sons Inc
- Glowing Genes: A Revolution in Biotechnology (2005) Marc Zimmer, Prometheus Books
- Aglow in the Dark (2005) Vincent Pieribone & David Gruber, Belknapp Press of Harvard University Press, Cambridge, Massachusetts, USA and London, England

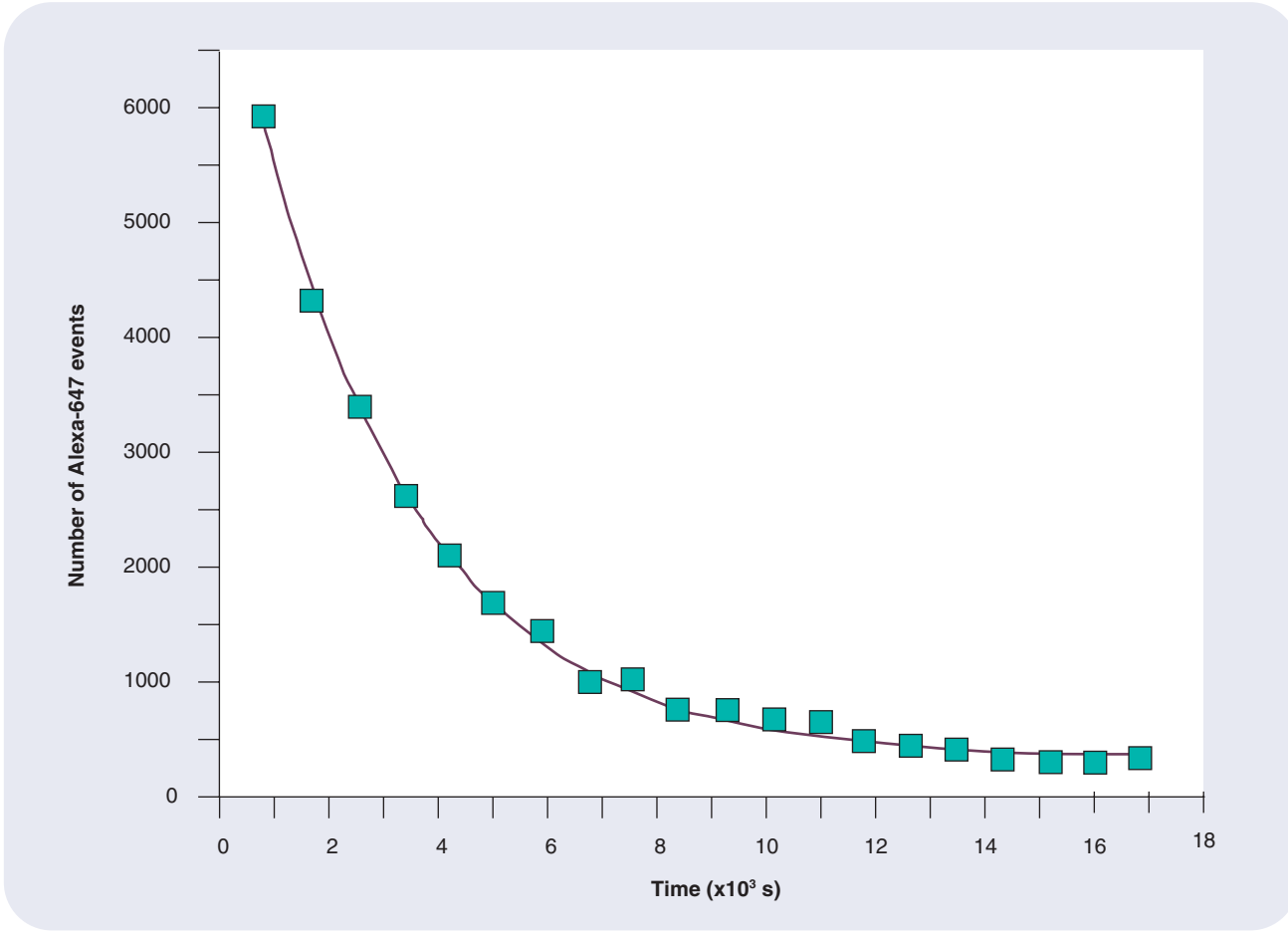

Figure 11. A single molecule unfolding time course by monitoring the number of acceptor events (>25 counts) of 50 pM YFP in 4M GdmCl. Data were fitted to a first-order equation (line of best fit).

Key issues

- Protein engineering techniques and selection methods used to generate variants of GFP with improved folding properties.
- Mechanism of chromophore formation.
- Mutations that reduce the tendency of GFP to dimerize.
- Complex kinetic mechanism for folding involving multiple intermediate states and parallel pathways.
- Partially structured states of GFP and residual structure in the denatured state.
- Chromophore and backbone dynamics
- Single-molecule folding and unfolding studies.

References

Papers of special note have been highlighted as:
• of interest
•• of considerable interest

1 Tsien RY. The green fluorescent protein. *Ann. Rev. Biochem.* 67, 509–544 (1998).

2 Shaner NC, Steinbach PA, Tsien RY. A guide to choosing fluorescent proteins. *Nat. Methods* 2, 905–916 (2005).
• Very useful practical guide to selecting fluorescent proteins for use in a wide range of applications.

3 Zimmer M. Green fluorescent protein (GFP): applications structure, and related photophysical behavior. *Chem. Rev.* 102, 759–781 (2002).

4 Reid BG, Flynn GC. Chromophore formation in green fluorescent protein. *Biochemistry* 36, 6786–6791 (1997).

5 Ormo M, Cubitt AB, Kallio K, Gross LA, Tsien RY, Remington SJ. Crystal structure of the Aequorea victoria green fluorescent protein. *Science* 273, 1392–1395 (1996).

6 Yang F, Moss L, Phillips G. The molecular structure of green fluorescent protein. *Nat. Biotech.* 14, 1246–1251 (1996).

7 Waldo GS, Standish BM, Berendzen J, Terwillger TC. Rapid protein folding assay using green fluorescent protein. *Nat. Biotechnol.* 17, 691–695 (1999).

8 Baird GS, Zacharias DA, Tsien RY. Circular permutation and receptor insertion within green fluorescent proteins. *Proc. Natl Acad. Sci. USA* 96, 11241–11246 (1999).

9 Topell S, Hennecke J, Glockshuber R. Circularly permuted variants of the green fluorescent protein. *FEBS Lett.* 457, 283–289 (1999).

10 Topell S, Glockshuber R. Circular permutation of the green fluorescent protein. *Methods Mol. Biol.* 183, 31–48 (2002).

11 Siemering K. R, Golbik R, Sever R, Haseloff J. Mutations that suppress the thermosensitivity of green fluorescent protein. *Current Biol.* 6, 1653–1663 (1996).

12 Kimata Y, Iwaki M, Lim CR, Kohno K. A novel mutation which enhances the fluorescence of GFP at high temperatures. *Biochem. Biophys. Res. Comm.* 232, 69–73 (1997).

13 Crameri A, Whitehorn EA, Tate E, Stemmer WP. C. Improved green fluorescent protein by molecular evolution using DNA shuffling. *Nat. Biotech.* 14, 315–319 (1996).

14 Battistutta R, Negro A, Zanotti G. Crystal structure and refolding properties of the mutant F99S/M153T/V163A of the Green Fluorescent Protein. *Proteins: Struct. Func. Gen.* 41, 429–437 (2000).

• The structure and folding properties of the widely used cycle 3 variant.

15 Fukuda H, Arai M, Kuwajima K. Folding of green fluorescent protein and the cycle 3 mutant. *Biochemistry* 39, 12025–12032 (2000).

16 Pedelacq JD, Cabantous S, Tran T, Terwilliger TC, Waldo GS. Engineering and characterization of a superfolder green fluorescent protein. *Nat. Biotech.* 24, 79–88 (2006).

•• Describes very recent studies on the selection of a superfolding variant of GFP using a folding interference method for selection.

17 Heim R, Prasher DC, Tsien RY. Wavelength mutations and posttranslational autoxidation of green fluorescent protein. *Proc. Natl Acad. Sci. USA* 91, 12501–12504 (1994).

18 Cubitt AB, Heim R, Adams SR, Boyd AE, Gross LA, Tsien RY. Understanding, improving and using green fluorescent proteins. *Trends Biochem. Sci.* 20, 448–455 (1995).

19 Wood TI, Barondeau DP, Hitomi C, Kassmann CJ, Tainer JA, Getzoff ED. Defining the role of arginine 96 in Green fluorescent protein fluorophore biosynthesis. *Biochemistry* 44, 16211–16220 (2005).

20 Sniegowski JA, Phail ME, Wachter RM. Maturation efficiency, trypsin sensitivity, and optical properties of Arg96, Glu222, and Gly67 variants of green fluorescent protein. *Biochem. Biophys. Res. Comm.* 332, 657–663 (2005).

21 Barondeau DP, Putnam CD, Kassmann CJ, Tainer JA, Getzoff ED. Mechanism and energetics of green fluorescent protein chromophore synthesis revealed by trapped intermediate structures. *Proc. Natl Acad. Sci. USA* 100, 12111–12116 (2003).

•• Here, mutations which substantially slow the rate but not the yield of post-translationally modified protein are used and the structure of the trapped precyclisation intermediate and oxidised post-translational state reported.

22 Rosenow MA, Huffman HA, Phail ME, Wachter RM. The crystal structure of the Y66L variant of green fluorescent protein Supports a cyclization-oxidation-dehydration mechanism for chromophore maturation. *Biochemistry* 43, 4464–4472 (2004).

23 Zhang L, Patel HN, Lappe JW, Wachter RM. Reaction Progress of Chromophore Biogenesis in Green Fluorescent Protein. *J. Am. Chem. Soc.* 128, 4766–4772 (2006).

•• This study uses the kinetics of hydrogen peroxide evolution during the maturation of GFP *in vitro* to inform on the mechanism of formation. Under highly aerobic conditions, the dominant pathway is found to involve a cyclisation then oxidation followed by dehydration.

24 Siegbahn PEM, Wirstam M, Zimmer M. Theoretical study of the mechanism of peptide ring formation in green fluorescent protein. *Int. J. Quantum Chem.* 81, 169–186 (2001).

25 Branchini BR, Nemser AR, Zimmer M. A computational analysis of the unique protein-induced tight turn that results in posttranslational chromophore formation in green fluorescent protein. *J. Am. Chem. Soc.* 120, 1–6 (1998).

26 Baffour-Awuah NY, Fedeles F, Zimmer M. Structural features responsible for GFPuv and S147P-GFP's improved fluorescence. *Chem. Phys.* 310, 25–31 (2005).

27 Zacharias DA, Violin JD, Newton AC, Tsien RY. Partitioning of lipid-modified monomeric gfps into membrane microdomains of live cells. *Science* 296, 913–916 (2002).

28 Jackson SE. How do small single-domain proteins fold? *Fold. Des.* 3, 81–91 (1998).

29 Huang J-R, Craggs TD, Christodoulou J, Jackson SE. Stable intermediate states are formed in the equilibrium unfolding of GFP. *Protein Science* submitted (2006).

30 Merkel JS, Regan L. Modulating Protein Folding rates *in vivo* and *in vitro* by side-chain interactions between the parallel strands of green fluorescent protein. *J. Biol. Chem.* 275, 29200–29206 (2000).

31 Iwai H, Lingel A, Pluckthun A. Cyclic green fluorescent protein produced *in vivo* using an artificially split PI-PfuI intein from Pyrococcus furiosus. *J. Biol. Chem.* 276, 16548–16554 (2001).

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

•• This study was the first to report detailed kinetic experiments on the folding of GFP and detect multiple intermediate states and a parallel pathway resulting from proline isomerisation.

33 Enoki S, Maki K, inobe T, Takahashi K *et al*. The equilibrium unfolding intermediate observed at pH 4 and its relationship with the kinetic folding intermediates in green fluorescent protein. *J. Mol. Biol.* 361(5), 969–982 (2006).

34 Vrzheshch PV, Akovbian NA, Varfolomeyev SD, Verkhusha VV. Denaturation and partial renaturation of a tightly tetramized DsRed protein under mildly acidic conditions. *FEBS Lett.* 487, 203–208 (2000).

35 Stepanenko OV, Verkhusha VV, Kazakov VI *et al*. Comparative studies on the structure and stability of fluorescent proteins EGFP, zFP506, mRFP1, "dimer2", and DsRed1. *Biochemistry* 43, 14913–14923 (2004).

36 Herberhold H, Marchal S, Lange R, Scheyhing CH, Vogel RF, Winter R. Characterization of the pressure -induced intermediate and unfolded state of red-shifted green fluorescent protein - a static and kinetic FTIR, UV/VIS and fluorescence spectroscopy study. *J. Mol. Biol.* 330, 1153–1164 (2003).

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

38 Haupt U, Maiti S, Schwille P, Webb WW. Dynamics of fluorescence fluctuations in green fluorescent protein observed by fluorescence correlation spectroscopy. *Proc. Natl Acad. Sci. USA* 95, 13573–13578 (1998).

39 Bonsma S, Purchase R, Jezowski S, Gallus J, Konz F, Volke RS. Green and red fluorescent proteins: photo- and thermally induced dynamics probed by site-selective spectroscopy and hole burning. *Chem. Phys. Chem.* 6, 838–849 (2005).

40 Seifert MH. J, Georgescu J, Ksiazek D *et al*. Backbone dynamics of green fluorescent protein and the effect of histidine 148 substitution. *Biochemistry* 42, 2500–2512 (2003).

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

42 Dickson RM, Cubitt AB, Tsien RY, Moerner WE. On/off blinking and switching behaviour of single molecules of green fluorescent protein. *Nature* 388, 355–258 (1997).

43 Schwille P, Kummer S, Heikal AA, Moerner WE, Webb WW. Dynamics of fluorescence fluctuations in green fluorescent protein observed by fluorescence correlation spectroscopy. *Proc. Natl Acad. Sci. USA* 97, 151–156 (2000).

44 Cannone F, Bologna S, Campanini B, Diaspro A, Bettati S, Mozzarelli A, Chirico G. Tracking unfolding and refolding of single GFPmut2 molecules. *Biophys. J.* 89, 2033–2045 (2005).

45 White SS, Balasubramanian S, Klenerman D, Ying LM. A Simple nanomixer for single molecule kinetics. *Angewante Chem.* (2006) (In Press).

46 Heikal AA, Hess ST, Baird GS, Tsien RY, Webb WW. Molecular spectroscopy and dynamics of intrinsically fluorescent proteins: Coral red (dsRed) and yellow (Citrine). *Proc. Natl Acad. Sci. USA* 97, 11996–12001 (2000).

47 Zhang J, Campbell RE, Ting A, Tsien RY. Creating new fluorescent probes for cell biology. *Nat. Rev. Mol. Biol.* 3, 906–918 (2002).

48 Miyawaki A, Tsien RY. Monitoring protein conformations and interactions by fluorescence resonance energy transfer between mutants of green fluorescent protein. *Methods Enzymol.* 327 472–500 (2000).

Affiliations

- *Sophie E Jackson*
  Chemistry Department, Lensfield Road, Cambridge CB2 1EW, UK
  Tel.: +44 1223 762 011
  Fax: +44 1223 336 362
  sej13@cam.ac.uk
- *Timothy D Craggs*
  Chemistry Department, Lensfield Road, Cambridge CB2 1EW, UK
  Tel.: +44 1223 767 042
  Fax: +44 1223 336 362
- *Jie-rong Huang*
  Chemistry Department, Lensfield Road, Cambridge CB2 1EW, UK
  Tel.: +44 1223 336 357
  Fax: +44 1223 336 362