The Crystal Structure of the Y66L Variant of Green Fluorescent Protein Supports a Cyclization−Oxidation−Dehydration Mechanism for Chromophore Maturation[†,‡]

Matthew A. Rosenow, Holly A. Huffman, Marlene E. Phail, and Rebekka M. Wachter*

Department of Chemistry and Biochemistry, Arizona State University, Tempe, Arizona 85287-1604

Received November 26, 2003; Revised Manuscript Received February 5, 2004

ABSTRACT: The crystal structure of a colorless variant of green fluorescent protein (GFP) containing the Y66L substitution has been determined to 1.5 Å. Crystallographic evidence is presented for the formation of a trapped intermediate on the pathway of chromophore maturation, where the peptide backbone of residues 65−67 has condensed to form a five-membered heterocyclic ring. The hydroxyl leaving group remains attached to the ring as confirmed by high-resolution electrospray mass spectrometry. The α-carbon of residue 66 exhibits trigonal planar geometry, consistent with ring oxidation by molecular oxygen. Side chain positions of surrounding residues are not perturbed, in contrast to structural results obtained for the GFPsol-S65G/Y66G variant [Barondeau, D. P., Putnam, C. D., Kassmann, C. J., Tainer, J. A., and Getzoff, E. D. (2003) *Proc. Natl. Acad. Sci. U.S.A. 100*, 12111−12116]. The data are in accord with a reaction pathway in which dehydration is the last of three chemical steps in GFP chromophore formation. A novel mechanism for chromophore biosynthesis is proposed: when the protein folds, the backbone condenses to form a cyclopentyl tetrahedral intermediate. In the second step, the ring is oxidized by molecular oxygen. In the third and final step, elimination of the hydroxyl leaving group as water is coupled to a proton transfer reaction that may proceed via hydrogen-bonded solvent molecules. Replacement of the aromatic Tyr66 with an aliphatic residue appears to have a profound effect on the efficiency of ring dehydration. The proposed mechanism has important implications for understanding the factors that limit the maturation rate of GFP.

Green fluorescent protein (GFP)[1] and its homologues constitute a family of small, compact, 11-stranded β-barrel proteins that spontaneously generate their own fluorophore from three internal residues (*1*, *2*). Because of their bright colors ranging from cyan to green to red, these proteins are used extensively as research tools in molecular and cell biology (*3*). GFP was originally cloned from the Pacific Northwest jellyfish *Aequorea victoria*, and the biochemical properties of its chromophore have been characterized in detail over the past decade (*3*). More recently, a number of homologues with broadly varying colors have been cloned from a series of coral species (*4*−*6*). These include red-shifted GFP-like proteins, which contain a chromophore π-system that is further modified from the green intermediate stage (*7*−*10*).

Gaining a detailed understanding of the reaction mechanism of GFP chromophore biosynthesis is of fundamental scientific interest. In addition, knowledge of the reaction pathway and identification of intermediates have significant practical and medical implications. GFP has a profound impact on the versatility of techniques available to cell biologists studying diseases ranging from cancer to developmental disabilities to genetic defects. Though GFP is most often used as a fusion tag to visually detect protein expression and cellular trafficking, an increasing number of variants have been engineered for specific biosensor applications that are useful in cell biology (*3*). A better understanding of the factors involved in forming the chromophore would be helpful in designing variants with novel and improved properties.

Chromophore formation in GFP-like proteins is a highly unusual post-translational modification that spontaneously follows protein folding and requires molecular oxygen. The autocatalytic mechanism is initiated by an intrachain ring closure, which leads to the formation of a five-membered ring from the backbone atoms of the original polypeptide chain (Ser65, Tyr66, and Gly67 in wild-type GFP) (*11*, *12*). This event is generally thought to be followed by a dehydration step according to mechanism A (Scheme 1), generating the five-membered imidazolinone ring (*11*, *12*). Full chromophore maturation, and hence the development of visible fluorescence, requires further processing steps that follow spontaneously. In all GFP-like proteins characterized to date, these steps include at least one oxidation event. To generate the green chromophore present in mature GFP, cyclization and dehydration are thought to be followed by

[†] This work was supported by a grant from the National Science Foundation (MCB-0213091) to R.M.W.

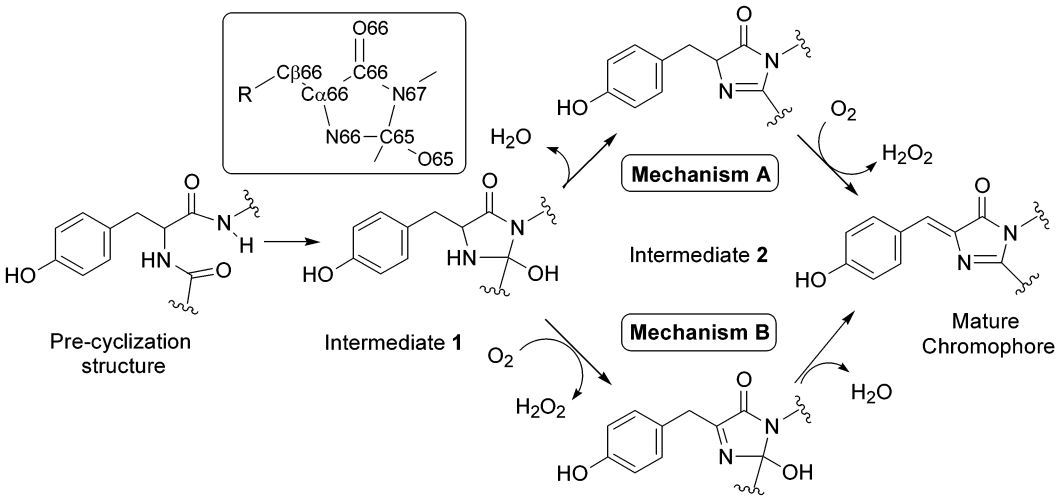
[‡] Coordinates for the X-ray structure of EGFP-Y66L have been deposited in the Protein Data Bank (entry 1S6Z).

* To whom correspondence should be addressed: Department of Chemistry and Biochemistry, Arizona State University, Tempe, AZ 85287. E-mail: RWachter@asu.edu.

[1] Abbreviations: OD, optical density; GFP, green fluorescent protein; HAL, histidine ammonia lyase; MIO, methylidene imidazolone; rms, root-mean-square; MALDI-MS, matrix-assisted laser desorption ionization mass spectrometry; ESI-MS, electrospray ionization mass spectrometry; HPLC, high-pressure liquid chromatography; EGFP, enhanced green fluorescent protein (GFP-F64L/S65T); GFPsol, GFP-F64L/S65T/F99S/M153T/V163A.

10.1021/bi0361315 CCC: $27.50
Published on Web 03/23/2004

Scheme 1: Three-Step Mechanism of GFP Chromophore Formation[a]



[a] (A) Generally accepted mechanism (*11*) and (B) mechanism proposed in this paper (inset shows atom labels used throughout the text).

dehydrogenation of the $C_\alpha-C_\beta$ bond of Tyr66 in the presence of molecular oxygen, leading directly to the fully conjugated chromophore (Scheme 1, mechanism A).

Several years ago, mass spectrometry results for GFP grown under anaerobic conditions provided support for this mechanistic sequence, since the additional mass loss due to the introduction of oxygen was found to be only 1 ± 4 Da (*11*). These results are consistent with oxidation from a single to a double bond being the last step in chromophore maturation. One of the first GFP X-ray structures solved, that of the bright green variant GFP-S65T (*13*), gave some evidence of a carbinolamine intermediate trapped prior to the elimination of water (*1*). The structure of this intermediate could not be analyzed in detail, since most of the chromophore in the crystal had matured to its final fluorescent state.

Recently, high-resolution X-ray structures of aerobically and anaerobically prepared GFPsol-S65G/Y66G were determined (*14*). In this variant, all three residues involved in chromophore formation (residues 65−67) are glycines. In the aerobic structure, the internal helix was found to be cyclized to a hydroxylated imidazolidine dione, similar to the tetrahedral intermediate generated by ring condensation (intermediate 1 in Scheme 1), though with an oxygen atom bonded to C_α66. Under anaerobic conditions, the same variant was shown to remain in the precyclization state, with the ring-forming atoms positioned in close proximity to each other. Hence, cyclization must be the first step in chromophore formation, as originally proposed by Tsien et al. (*11*), though it has also been argued that initial oxidation of Tyr66 to the $\alpha-\beta$ dehydrotyrosine would aid in subsequent ring closure (*15*). From the structural work of aerobic GFPsol-S65G/Y66G, it is not clear what factors stabilize the cyclopentyl intermediate (*14*). It is unlikely that backbone cross-linking is driven solely by mechanical compression of reacting groups, since helical distortions in the precyclization state are maintained after ring closure (*14*). For this reason, Barondeau et al. (*14*) have proposed that the cyclic intermediate is stabilized by ring conjugation via elimination of water.

To this date, very few examples of autocatalytic peptide backbone cyclizations have been characterized in proteins. Main chain condensations have been described only in the family of GFP-like proteins (*7*), and more recently in histidine ammonia lyase (HAL) (*16*, *17*), though cyclic succinimide structures involving side chains such as aspartic acid are more common (*18*). In HAL, an autocatalytic peptide backbone cyclization initiates the formation of a built-in electrophilic cofactor, the methylidene imidazolone (MIO) (*16*). Dehydroalanine is thought to form after backbone cyclization, and cyclization has been argued to be the result of pressure exerted by the protein fold onto the reacting atoms (*17*). There is no obvious structural homology of any kind between GFP and HAL, and to this date, no sequence or structural motif has been identified that would allow for prediction of backbone condensations in other proteins.

The π-skeleton of the mature GFP chromophore is primarily derived from the phenolic side chain of Tyr66 with the oxidized $\alpha-\beta$ bond. Other aromatic residues (His, Phe, and Trp) have been substituted in this position with retention of fluorescence (*12*); hence, the mechanism of chromophore formation is intact as long as residue 66 is aromatic. In the work presented here, we explored whether the backbone cross-linking reaction is also facilitated in the absence of an aromatic group. We prepared the colorless EGFP-Y66L variant, crystallized the protein, and determined its high-resolution X-ray structure. Based on structural features of the cyclized carbinolamine intermediate and mass spectrometry data, we propose a novel mechanism of GFP chromophore formation (Scheme 1, mechanism B), where the dehydration step occurs as the last of three major chemical steps.

MATERIALS AND METHODS

Site-Directed Mutagenesis and Preparation of Soluble Protein. The Y66L substitution was introduced into the EGFP variant (GFP-F64L/S65T) using the PCR-based QuikChange site-directed mutagenesis kit (Stratagene). EGFP-Y66L, N-terminally tagged with a six-His tag in the $pRSET_B$ plasmid, was overexpressed in *Escherichia coli* strain JM109-DE3. Liquid culture growth and protein purification were carried out essentially as described previously (*1*). To maximize the level of protein expression in soluble form, the growth temperature was adjusted to 25 °C before

induction with IPTG. After Ni−NTA affinity purification (Qiagen), the six-His tag was cleaved with α-chymotrypsin (Sigma) (1:50 chymotrypsin/GFP mixture, w/w) by incubation at 4 °C for 18 h. The digest was then loaded onto a DEAE-Sepharose column (Sigma) and washed with buffer containing 20 mM HEPES (pH 7.9) and 50 mM NaCl. EGFP-Y66L was eluted with buffer containing 20 mM HEPES (pH 7.9) and 300 mM NaCl. Final protein purity was at least 95% as judged by SDS−PAGE. The molecular mass and homogeneity were verified by MALDI mass spectrometry.

Protein Preparation from Inclusion Bodies. To prepare inclusion bodies of EGFP-Y66L, the growth temperature was equilibrated to 42 °C before induction with IPTG (*19*). Otherwise, the same protein expression protocol that was used for the soluble protein was followed. After cell disruption, the pellet was homogenized in inclusion body wash buffer [100 mM Tris (pH 7.9), 500 mM NaCl, 20 mM EDTA, 2% Triton X-100, and 5 mM DTT] and washed a total of four times, with Triton X-100 included in the first wash only. The final homogenate was quick-frozen in liquid nitrogen and stored at −80 °C. Aliquots were thawed on ice, solubilized in 50 mM HEPES (pH 7.9), 50 mM NaCl, 8.0 M urea, and 1 mM DTT, and then loaded onto a Ni−NTA chromatography column equilibrated in 50 mM HEPES, 50 mM NaCl, and 8.0 M urea. The column was washed in wash buffer with 20 mM imidazole, and purified denatured EGFP-Y66L was eluted by increasing the imidazole concentration to 100 mM.

Electrospray Ionization (ESI) Mass Spectrometry. Samples of EGFP-Y66L, isolated both in soluble form and from inclusion bodies, were purified by C18 reverse-phase HPLC (acetonitrile gradient in a water/0.1% TFA mixture). The single protein peak was collected, dried in a speedvac, and then sonicated in 50% acetonitrile and 50% water before analysis by ESI mass spectrometry. Mass spectra were collected on a 4.7 T Fourier transform ion cyclotron resonance (FT-ICR) instrument (Ionspec, Inc.) using ubiquitin as an internal standard. Multiply charged ions were acquired, and the data were deconvoluted by computer analysis.

Crystallization and Diffraction Data Collection. EGFP-Y66L was expressed in soluble form in *E. coli*, purified, concentrated to 15 mg/mL, and crystallized via the hanging drop vapor diffusion method (2 μL of protein and 2 μL of mother liquor). The mother liquor contained 15−20% polyethylene glycol 8000, 200 mM calcium acetate, and 100 mM cacodylic acid (pH 6.5) (Sigma). Rod-shaped crystals, typically 0.1 mm across and 0.5 mm long, grew at 4 °C within 3−5 days. Single crystals were soaked in 30% glycerol for cryoprotection and flash-frozen in a stream of N_2 gas (100 K). X-ray diffraction data were collected using an Raxis IV++ image plate detector mounted on a Rigaku RU200HB rotating anode generator (Cu Kα radiation) and equipped with osmic confocal mirrors.

Structure Determination and Refinement. Diffraction data were collected to 1.5 Å and were processed with MOSFLM (*20*) and SCALA within the CCP4 suite of programs (*21*). The space group was $P2_12_12_1$ with the following unit cell parameters: $a = 50.80$ Å, $b = 62.74$ Å, and $c = 69.96$ Å (Table 1). Its structure was nearly isomorphous to the X-ray structure of GFP-S65T (*1*) (PDB entry 1EMA). This model was used for phasing after deletion of residue 64 (which is mutated in EGFP-Y66L) and the chromophore (equivalent to residues 65−67). After removal of 10% of the data to estimate R_{free}, refinement was carried out against 90% of the data using maximum likelihood least-squares minimization in CNS (*22*). Initial rigid body refinement was followed by positional refinement to 3.0 Å resolution, and then stepwise to the full resolution limit with inclusion of *B*-factor refinement. Electron density maps ($2F_o - F_c$ and $F_o - F_c$) were displayed intermittently, and the model was adjusted manually using O (*23*). Ordered solvent molecules were modeled on the exterior protein surface and within hydrogen bonding distance of appropriate protein atoms. Modeling of interior water molecules was postponed until the very end of refinement.

Table 1: Crystallographic Statistics for EGFP-Y66L

Data Collection and Processing	
space group	$P2_12_12_1$
cell dimensions *a*, *b*, *c* (Å)	50.80, 62.74, 69.97
total no. of observations	205920
no. of unique reflections	35157
completeness[a] (%) (shell)	96.5 (86.5)
$I/\sigma(I)$[a] (shell)	8.0 (2.1)
R_{merge}[a,b] (%) (shell)	6.1 (33.0)
resolution (Å)	1.50
Refinement Statistics	
resolution (Å)	20−1.50
no. of reflections	35088
R_{work} (R_{free})	0.189 (0.223)
no. of protein atoms	1815
no. of solvent atoms	295
bond length deviations (Å)	0.015
bond angle deviations (deg)	1.88
B-factor deviation for main chain atoms	3.6
B-factor deviation for side chain atoms	8.5

[a] Completeness is the ratio of the number of observed reflections for which $I > 0$ divided by the theoretically possible number of reflections. Values in parentheses are for the highest-resolution shell between 1.58 and 1.50 Å. [b] $R_{merge} = \sum|I_{hkl} - \langle I\rangle|/\sum\langle I\rangle$, where $\langle I\rangle$ is the average of individual measurements of I_{hkl}.

Strong positive difference electron density was observed in the protein interior, indicating the presence of a five-membered cyclic structure. The density was clearly consistent with peptide backbone cross-linking between residues 65 and 67, as is the case in GFPs containing the mature chromophore. When the crystallographic *R*-factor was reduced to 24%, the heterocyclic ring and the leucine side chain of residue 66 (Y66L) were manually positioned into the density. Further refinement was carried out while the bond lengths within the five-membered ring were only weakly restrained. No bond angle or torsion angle restraints were used for the ring atoms, since the oxidation state of these bonds was not known. The peptide bond between residues 66 and 67, which is part of the cyclic structure, was refined in the same manner as other peptide bonds. Clear and strong positive difference density was observed adjacent to and within bonding distance of $C_\alpha 65$, indicating that this atom bears a fourth substituent. Modeling of the attached entity as an oxygen atom completely accounted for the density. The final model includes residues 2−230, a surface-bound chloride ion, and 295 ordered water molecules. Data collection and refinement statistics are presented in Table 1.

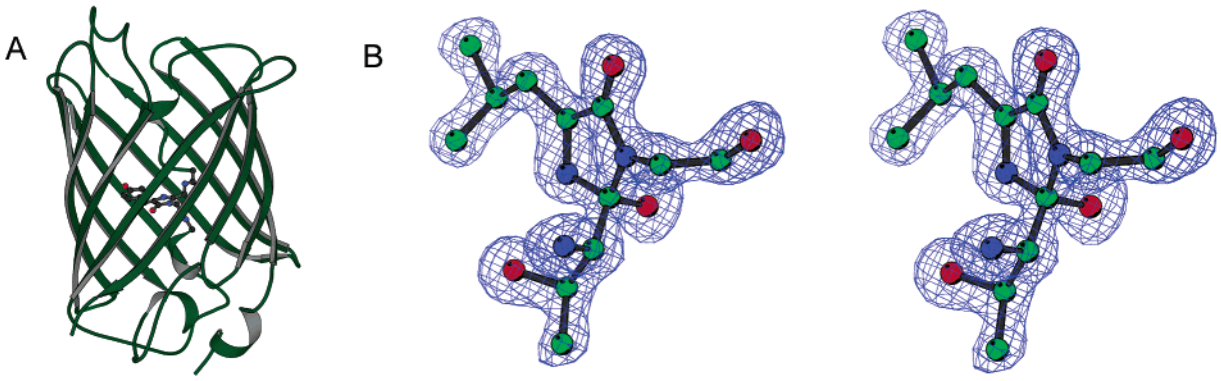
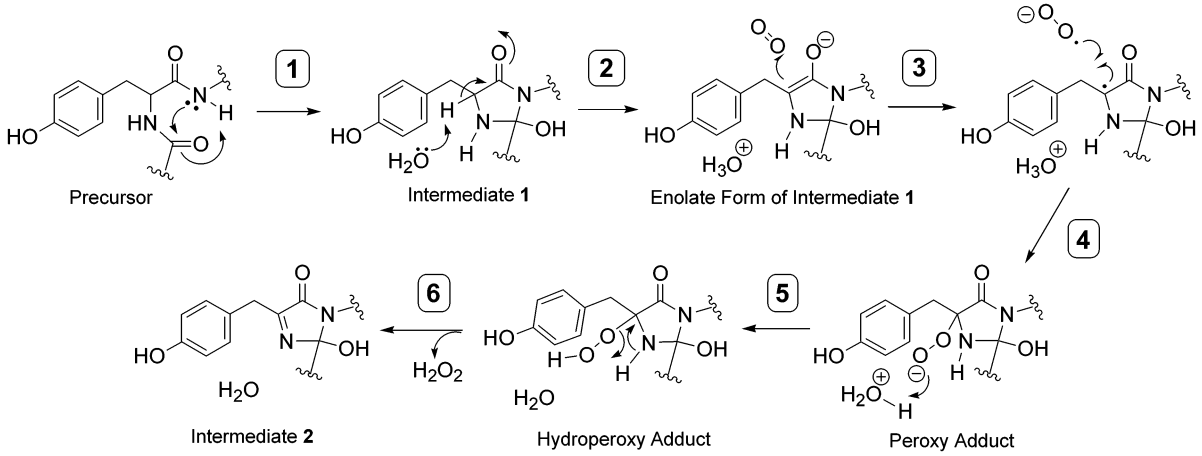


FIGURE 1: (A) Ribbon diagram of the GFP protein fold generated from the coordinates of GFP-S65T (*1*), with the chromophore shown in ball-and-stick representation. (B) Stereoview of the $F_o - F_c$ electron density omit map for EGFP-Y66L, contoured at 4.0σ. Residues 65−67 were omitted from the final atomic model, and the map was calculated after 200 steps of crystallographic refinement. The atoms of residues 65−67 from the refined model were then superimposed onto the map (oxygen in red, nitrogen in blue, and carbon in green).

Scheme 2 Process of Backbone Cyclization and Oxidation, Part of the Proposed Mechanism of GFP Chromophore Maturation[a]



[a] These events are shown in six microscopic steps and lead to formation of intermediate 2.

RESULTS

Description of the X-ray Structure of EGFP-Y66L. To investigate structures of precursor GFPs and intermediates with immature chromophores, the EGFP-Y66L variant was generated via site-directed mutagenesis. Replacement of Tyr66, the central of the three chromophore-forming residues, with the nonaromatic leucine resulted in a colorless mutant as expected. The protein was expressed in *E. coli*, purified, and crystallized in space group $P2_12_12_1$ with the following unit cell dimensions: $a = 50.80$ Å, $b = 62.74$ Å, and $c = 69.96$ Å. A complete diffraction data set was collected; the data were processed to 1.5 Å resolution, and the structure was determined via molecular replacement using the X-ray structure of GFP-S65T (PDB entry 1EMA, Figure 1A) for phasing (*1*). The structure contains one monomer per asymmetric unit. Refinement gave high-quality statistics with a final *R*-factor of 19% and good stereochemistry (Table 1).

Electron density near the center of the 11-stranded β-barrel clearly indicated the formation of a five-membered ring, even before residues 65−67 were modeled (Figure 1B). This cyclic structure was modeled as an imidazolidinone (fully reduced imidazole with an attached carbonyl oxygen), analogous to the imidazolinone ring of the mature chromophore (partially reduced imidazole with an attached carbonyl oxygen). During further refinement, the ring bond lengths were only weakly restrained, and ring bond angles were not restrained at all, since the oxidation state and hybridization of ring atoms were not known. On the basis of clear electron density, O65 was modeled as a hydroxyl group bonded to the Thr65−Leu66 peptide bond C65 (Figure 1B), which exhibits tetrahedral geometry and four substituents. EGFP-Y66L is trapped in an intermediate chromophore maturation stage that precedes the elimination of the hydroxyl leaving group as a water molecule. There are no significant features in the final $F_o - F_c$ difference electron density map near residues 65−67. In particular, no electron density consistent with the precyclization or the postdehydration state of the protein is observed.

Geometry of the EGFP-Y66L Cyclic Intermediate. Since the peptide backbone of the internal helix is cross-linked, this variant has progressed through the first chemical step toward GFP chromophore formation (Schemes 1 and 2). A five-membered cyclic structure buried in the protein interior has formed due to nucleophilic attack of the amide nitrogen of Gly67 on the carbonyl carbon of Thr65. Hence, this variant constitutes a trapped intermediate on the maturation pathway to a fully intact, fluorescent GFP. The ring geometry is essentially planar (Figure 1B).

The electron density is striking in that the Leu66 $C_\alpha - C_\beta$ bond lies in-plane with the five-membered ring (Figure 1B). The trigonal planar geometry of $C_\alpha 66$ is also observed in mature GFPs, which is part of the mature chromophore

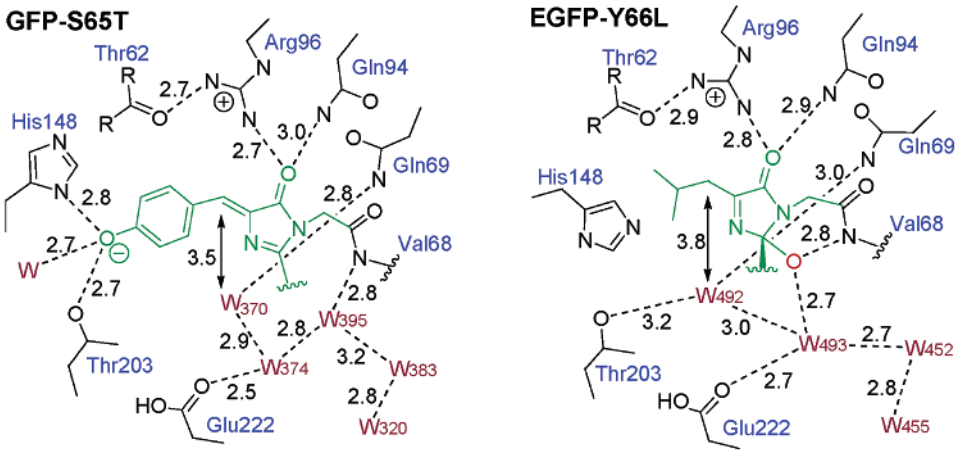
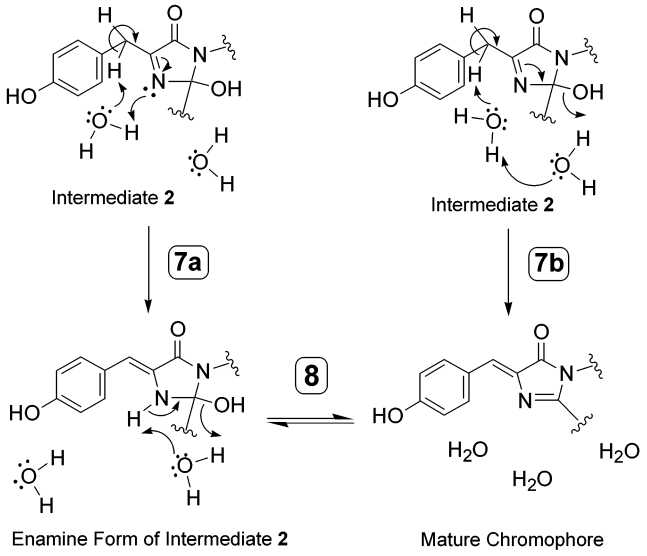


FIGURE 2: Schematic of protein groups positioned around the chromophore in GFP-S65T (*1*) and around the cyclized intermediate in EGFP-Y66L. Distances are in angstroms and are taken from the respective crystal structures. Ordered waters are denoted with W. Dashed lines represent hydrogen bonds.

Scheme 3: Proposed Alternate Mechanisms for the Dehydration Event that Is Proposed To Complete the Process of GFP Chromophore Maturation[a]



[a] In step 7a, the enamine tautomer of the imine (intermediate 2) is formed prior to dehydration, whereas in step 7b, dehydration is initiated by proton abstraction at C_β66.

π-system (Figure 2). Since the geometry of C_α66 is indicative of sp^2 hybridization, one of the three connecting bonds must be desaturated (C_α66−C66, C_α66−N65, or C_α66−C_β66). Hence, the observed X-ray data are consistent with either the enolate form of intermediate 1 (Scheme 2), intermediate 2 with an oxidized heterocycle (Scheme 2), or the enamine tautomer of intermediate 2 (Scheme 3).

Conjugation State of the Cyclized Intermediate. Intermediate 1, the fully saturated hydroxylated imidazolidinone ring, is formed immediately after ring closure according to step 1 (Scheme 2). This intermediate may be in equilibrium with its enolate form according to step 2. In the X-ray structure of EGFP-Y66L, a water molecule, W493, is positioned 3.5 Å from C_α66 (Figure 2). This water is localized at the appropriate face of the trigonal center, i.e., "behind" the plane of the page, to be responsible for proton abstraction from the L-amino acid precursor (Scheme 2). The negative charge of the enolate, localized on O66, would be electrostatically stabilized via a salt bridge interaction with the guanidinium group of Arg96 (Figure 2). In addition, W493 is directly hydrogen bonded to the side chain of Glu222, and polarization by the carboxylate may increase its basicity. The hydronium ion generated by proton abstraction may transfer the proton to the Glu222 carboxylate. Since the enolate form of intermediate 1 has a trigonal planar geometry at C_α66, the X-ray data of EGFP-Y66L are consistent with this intermediate, though it is unlikely that the high-energy enolate would be the predominant species present in the crystal.

All experiments were carried out under aerobic conditions. Hence, a second possibility is that oxidation of the five-membered ring occurs spontaneously in EGFP-Y66L. A possible oxidation mechanism for intermediate 1 is proposed in Scheme 2 (steps 3−6). According to this mechanism, transfer of an electron from the enolate to molecular oxygen (step 3) is followed by a radical−radical combination step (step 4) that leads to the peroxy adduct of the ring at C_α66. Upon protonation of the peroxy group (step 5) from the hydronium ion of W493 generated in step 2, the hydroperoxy adduct abstracts a proton from N66 and dissociates as the hydrogen peroxide product (step 6). These events are proposed to lead to the formation of intermediate 2, in which the ring is partially oxidized to the hydroxylated imidazolinone with a desaturated C_α66−N65 bond (Scheme 2). There are several lines of evidence that argue in favor of spontaneous oxidation of the initial cyclization product, intermediate 1. These include facilitated transfer of an electron to dioxygen via the enolate (step 3), and oxidative derivatization of a similar cyclized intermediate at C_α66 (*14*). These points are discussed in more detail in Discussion.

In addition to the intermediate 1 enolate and the intermediate 2 imine, the X-ray data (Figure 1B) are also consistent with the enamine form of intermediate 2 (Scheme 3). Imine-to-enamine tautomerizations are relatively facile interconversions since the only requirement is a carbon-to-nitrogen proton transfer (step 7a, Scheme 3). In addition, the enamine would contain a C_α66−C_β66 double bond, as found in the green fluorescent mature chromophore. Once the enamine is generated, transfer of a proton from N66 to O65 via an ordered solvent molecule (W493, Figure 2) would facilitate the final dehydration event (step 8, Scheme 3), and the

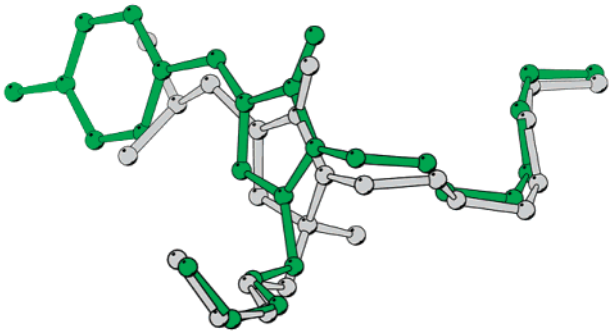


FIGURE 3: Overlay of structures derived from residues 65−67 in GFP-S65T (green) and EGFP-Y66L (gray), generated by superposition of C_α atoms of residues 5−64 and 68−229 (GFP-S65T, PDB entry 1EMA).

process would be complete. An alternate route for generating the mature chromophore directly from intermediate 2 would be deprotonation at C_β66 with concomitant protonation of O65. According to this scheme, the proton would be transferred via two ordered water molecules positioned near the reactive groups in the crystal structure, W492 and W493 (step 7b, Scheme 3). Alternate mechanisms 7a and 7b are discussed in more detail in Discussion.

Structural Comparison of EGFP-Y66L with GFP-S65T. The refined EGFP-Y66L structure exhibits a root-mean-square deviation of 0.30 Å with respect to the GFP-S65T structure for all C_α atoms of residues 5−229, excluding residues 65−67. The Leu66 C_α−C_β bond occupies the same region of space as the Tyr66 side chain in GFP-S65T (Figure 3). The hydroxylated imidazolinone ring in EGFP-Y66L is shifted by ∼0.67 Å when compared to the equivalent ring in GFP-S65T, in which it is part of the mature planar chromophore (Figure 3). Because of the tetrahedral carbon center, C_α65 is shifted by 0.84 Å in EGFP-Y66L, tilting the ring out of plane by ∼23° in comparison to GFP-S65T. This movement is not translated along the main chain atoms of the protein. The C_α atom of Leu64, immediately preceding the heterocycle, overlays closely on the equivalent atom in GFP-S65T (0.26 Å difference), with no structural perturbation. The C_α atom of the succeeding Val68 is shifted by 0.37 Å, whereas the following main chain atoms are shifted by less than 0.3 Å (Figure 3). The displacement of the Y66L hydroxylated imidazolinone ring appears to be primarily a consequence of the tetrahedral geometry around the carbinolamine, and hence strictly a local effect.

Environment of the Cyclized Chromophore Precursor. The hydrogen bonding interactions of protein groups with the hydroxylated imidazolinone ring are identical to those observed in GFP-S65T and other GFP structures. Arg96 and Glu222 are strictly conserved among all known GFP homologues (*4*) and are found in equivalent positions in X-ray structures of GFP variants and GFP-like proteins (*8*, *24*−*27*). Arg96 is always found to be hydrogen-bonded to the carbonyl oxygen of the heterocycle, as it is in EGFP-Y66L. In EGFP-Y66L, the side chain of Glu222 is hydrogen bonded to the side chain hydroxyl of Thr65. This is also the case in GFP-S65T (*1*), though the exact position of the Glu222 carboxylate tends to vary in different GFP mutants (see Discussion). Both Arg96 and Glu222 are stabilized by a network of hydrogen bonding interactions with nearby residues and ordered solvent molecules (Figure 2), and the pattern is identical in EGFP-Y66L and GFP-S65T (*1*). In addition, the side chains of Gln94, Gln69, and Gln183 and the backbone of Thr62 are positioned in nearly identical spatial relationships with respect to the imidazolinone ring as in GFP structures containing a mature chromophore (Figure 2).

A series of ordered water molecules form a hydrogen bonding network in the vicinity of the condensed backbone, and are highly conserved in GFP variants (Figure 2) (*1*). In the X-ray structure of EGFP-Y66L, the solvent molecule equivalent to W395 (GFP-S65T) is not observed, since O65 remains bonded to C65 in the hydration adduct.

In the vicinity of the Leu66 side chain, structural rearrangements are observed to accommodate the shorter and more hydrophobic leucine. The imidazole ring of His148 is flipped toward the protein interior by a 110° rotation, and occupies part of the cavity created by elimination of the chromophore phenolic group. A similar movement of His148 was observed in the low-pH structure of the blue fluorescent GFP variant containing the Y66H substitution (*28*), where the chromophore has been shortened significantly. In EGFP-Y66L, the Tyr145 backbone atoms have moved toward the cavity by 0.7 Å, and the side chain is rotated by ∼24°. The terminal methyl groups of Leu66 are near van der Waals contact with the side chains of Val150, His148, Tyr145, and Thr203 (separation of <4.0 Å). The side chain of Ser205 has rotated by approximately 90° and is not involved in any obvious hydrogen bonding interactions.

Mass Spectrometry. To determine whether the chemical rearrangements observed in EGFP-Y66L lead to a change in mass, high-resolution electrospray ion cyclotron resonance mass spectrometry (ESI-ICR) was carried out. Since backbone cyclization in GFP is initiated by proper protein folding to its native state (*3*), the precursor molecular mass can only be determined by isolating protein that has remained unfolded or misfolded during its entire lifetime. To compare the molecular mass of the EGFP-Y66L precursor with that of the mature soluble protein used for crystallization (see above), EGFP-Y66L was expressed separately in two different forms, as inclusion bodies and in soluble form. Partitioning of the protein into inclusion bodies was achieved by adjusting the growth temperature to 42 °C (*19*), whereas the soluble protein yield was maximized by expression at 25 °C (*3*). The inclusion bodies were solubilized and purified in 8 M urea, and directly injected into reverse-phase HPLC under denaturing conditions. This procedure ensured that the unfolded form of the protein was not allowed to reach its native three-dimensional structure at any time during purification. Hence, inclusion body-derived protein cannot contain a cyclized backbone.

The mass of the soluble form of EGFP-Y66L was found to be 31 025.0 ± 3 Da, and that of the inclusion body-derived protein was 31 026.0 ± 3 Da. No additional species were present in either of the two samples. The theoretical mass for unmodified EGFP-Y66L is 31 029 Da, which includes the 37-residue six-His tag. Mass loss due to folding and maturation is therefore only 1 ± 3 Da, consistent with soluble EGFP-Y66L in the intermediate 1 or intermediate 2 state (or respective tautomers). Elimination of water (−18 Da) has clearly not occurred. Hence, the mass spectral results are in accord with the X-ray data.

DISCUSSION

In EGFP-Y66L, the tetrahedral intermediate on the pathway to chromophore formation is trapped due to substitution of an aromatic residue in position 66 with an aliphatic amino acid. On the basis of the X-ray structure and mass spectrometry results presented here, a novel mechanism for GFP chromophore formation is proposed. This mechanism is a modification of the generally accepted mechanism originally put forth by Tsien and co-workers (*11*), in which the three major chemical steps leading to chromophore maturation were proposed to be cyclization, dehydration, and oxidation, in that order. Here, we propose that the sequence of events is cyclization followed by oxidation, and that dehydration occurs last and completes the process (mechanism B, Scheme 1). ESI-ICR mass spectrometry data for EGFP-Y66L support the cyclization−oxidation−dehydration mechanism, since mass loss as a function of protein folding and covalent modification is consistent with oxidation but not dehydration.

The details of the modified mechanism can be described in seven or eight microscopic steps (Schemes 2 and 3). According to this mechanism, nucleophilic attack of the amide N67 onto the carbonyl C65 leads to backbone ring closure, as originally proposed by Tsien and co-workers (*11*). This cyclization reaction may be facilitated by a proton transfer from N67 to O65, which would lead to charge neutralization of intermediate 1 (step 1, Scheme 2).

Oxidation of the Cyclized Intermediate by Molecular Oxygen. Crystallographic data presented here argue in favor of backbone condensation coupled to ring oxidation during the early steps of chromophore formation. The X-ray structure of EGFP-Y66L exhibits clear planar geometry around C_α66. This carbon must therefore be sp^2-hybridized (Figure 1), with a desaturated bond to one of its three substituents. For this reason, C_α66 is proposed to be the site of oxidation. A trigonal carbon center is consistent with three different forms of the cyclization product: the enolate form of intermediate 1 (C_α66−C66 double bond, Scheme 2), intermediate 2 (C_α66−N66 double bond, Schemes 2 and 3), and the enamine tautomer of intermediate 2 (C_α66−C_β66 double bond, Scheme 3). The enolate of intermediate 1 is likely too high in energy to be the main species found in the crystal. More likely candidates for the entity trapped in EGFP-Y66L are the two forms of intermediate 2 generated by oxidation of the five-membered ring of intermediate 1 (Scheme 3). The initial condensation product, intermediate 1 (Scheme 2), is structurally related to a doubly reduced imidazole, an imidazolidine. The reduction of the singly reduced imidazoline to the fully reduced imidazolidine is known to be difficult (*29*), suggesting that oxidation of the imidazolidine may be an energetically favorable process occurring spontaneously under aerobic conditions.

Consistent with this notion, C_α66 has been found to be oxidized to a carbonyl in the aerobic X-ray structure of GFPsol-S65G/Y66G (*14*). The additional atom observed in this structure, the carbonyl oxygen, is presumably derived from atmospheric oxygen, lending support to the proposed susceptibility of C_α66 to spontaneous oxidative modification. Under anaerobic conditions, the same variant was found to remain in the precyclization state (*14*). Hence, ring oxidation is thought to play a critical role in trapping the initial cyclization product, intermediate 1 (Schemes 1 and 2). C_α66 may be the site of covalent attachment of a hydroperoxide, subsequently leading to the formation of intermediate 2 on the pathway of chromophore biosynthesis (Scheme 2, steps 3−6). There is precedence for the spontaneous formation of a hydroperoxide adduct in an enzyme's active site (*30*). For example, in the crystal structures of the photoproteins obelin and aequorin, a hydroperoxide was observed to be covalently attached to the enzyme-bound ceolenterazine ligand (*31*).

Cyclization Is Not Trapped by Dehydration. The EGFP-Y66L structure presented here clearly indicates that dehydration does not play a role in trapping the backbone condensation product. Hydroxyl O65 is found to be attached to the tetrahedral carbon, with full occupancy and a low *B*-factor. In addition, this oxygen is hydrogen-bonded to the backbone nitrogen of Val68 and a water molecule, W493 (Figure 2). W493, in turn, is part of a network of hydrogen bonding interactions that include the Glu222 side chain and several ordered solvent molecules, and extends to the carbonyl oxygen of Phe71. Hydrogen bonding of protein groups and ordered waters with O65 may aid in favoring the hydration adduct at C65 over the elimination of water. Structural rearrangements around the five-membered ring are minor and appear to be primarily due to the tetrahedral geometry of C65 (Figure 3). For these reasons, dehydration in EGFP-Y66L is not likely stalled because of structural perturbations in the vicinity of the ring, as has been suggested for the GFPsol-S65G/Y66G variant (*14*). Instead, the aliphatic substitution in position 66 must be responsible for arresting chromophore maturation, and ejection of water is not required for stabilization of the cyclized intermediate.

Ejection of the Leaving Group from C65 via Enamine Formation. To complete the process of maturation of the chromophore to its final fluorescent state, the hydroxyl group (O65) must be eliminated as a water molecule (Scheme 3). To this end, we propose two alternate mechanisms, both supported by the X-ray structure of EGFP-Y66L. In the first mechanism, the imine of intermediate 2 is converted to its enamine tautomer by the transfer of a proton from C_β66 to N66 via a water molecule such as W492 (step 7a, Scheme 3). This step is followed by a second transfer of a proton from N66 to O65, again mediated by a water molecule such as W493. Protonation of hydroxyl O65 will facilitate its elimination as water. W493 is 3.1 Å from N66 in the X-ray structure (Figure 2) and localized below the plane of the ring. If one considers small-scale diffusion of waters buried in a protein's interior, W493 or its equivalent may act as a proton shuttle in the enamine dehydration step, which completes maturation of the chromophore. Though in GFP the green fluorescent dehydration product is clearly the most stable species, the hydrated adduct may be favored in EGFP-Y66L because of the less conjugated system.

Ejection of the Leaving Group from C65 via Deprotonation of C_β66. On the basis of the X-ray structure of EGFP-Y66L, another mechanism must be considered for the final event in chromophore biosynthesis. In GFPs containing an aromatic residue 66, dehydration may be facilitated by deprotonation of C_β66 (step 7b, Scheme 3), since proton loss at this site would be coupled to ejection of O65 via the N66−C_α66 π-bond. Hence, step 7b would directly lead to formation of the extended π-system of the mature chromophore. The C_β66 hydrogens are acidified because of the electron withdrawing properties of the adjacent desaturated

heterocycle (intermediate 2, Scheme 3). In wild-type GFP, additional acidification of C_β66 is achieved due to π-conjugation of the negative charge with the Tyr66 phenolic group. This additional acidification may allow for completion of the chromophore maturation process, whereas in EGFP-Y66L, proton abstraction would be less favorable so that intermediate 2 is trapped (Scheme 2).

X-ray crystal structures of GFPs (*1*) (Figure 2) indicate that no protein atoms are appropriately positioned to be candidates for playing a role as a base in step 7b (Scheme 3). Deprotonation of C_β66 could be achieved by a nearby water molecule such as W492, which is 3.8 Å from C_β66 in the EGFP-Y66L structure (Figure 2). W492 is hydrogen bonded to W493, which in turn forms a hydrogen bond with hydroxyl O65 of the tetrahedral intermediate. Once a proton is removed from C_β66, a proton relay mechanism could shuttle the proton to O65 via W492 and W493 (Figure 2 and Scheme 3). Elimination of O65 as a water molecule is facilitated by protonation of the hydroxyl group, as it is in the enamine mechanism. In the proton relay mechanism, charges are neutralized via conserved solvent molecules that connect the proton donor C_β66 to the proton acceptor O65. The mature chromophore, more stable than the hydration adduct in wild-type GFP, is generated by proton transfer and elimination of H_2O.

Both dehydration mechanisms (steps 7a and 7b, Scheme 3) propose a proton shuttle via hydrogen-bonded water bridges, and both mechanisms may be operational in GFP. A role for a proton relay system in chromophore biosynthesis is not unreasonable, since most high-resolution X-ray structures of GFPs (*1*, *25*, *28*) contain an array of positionally conserved, well-ordered solvent molecules. These waters are localized in a large interior cavity (*1*) that is found directly adjacent to the chromophore (Figure 2). In addition, there is extensive evidence for the existence of proton relay systems in enzyme active sites. One of the best-studied examples is the enzyme carbonic anhydrase, in which intramolecular proton transfer proceeds through a series of hydrogen-bonded water bridges, apparently over a distance of 9 Å between the proton donor and acceptor (*32*).

Role of Glu222 and Thr62. It has been known for some time that Glu222 is not essential in generating the chromophore (*33*, *34*), though this residue is completely conserved among all known GFP-like proteins (*4*). The X-ray structure of EGFP-Y66L supports the idea that Glu222 is not directly involved in catalyzing chemical steps of chromophore formation. In this structure, the carboxylate of Glu222 is 3.4 Å from N66, essentially in the same position as in GFP-S65T (Figure 2). In the aerobic structure of GFPsol-S65G/Y66G, Glu222 is within 3.1 Å of N66, albeit in a different rotamer conformation (*14*). If Glu222 were the base responsible for deprotonation of N66 immediately after cyclization (mechanism A, Scheme 1) as has been suggested (*15*), proton abstraction and dehydration would have occurred in both EGFP-Y66L and GFPsol-S65G/Y66G. Yet, this is not observed in either case. In addition, the charge state of Glu222 appears to vary in mature GFPs (*1*, *35*) and is inferred to be neutral in EGFP-Y66L. The exact position of the Glu222 carboxylate is not well conserved in available crystal structures of GFP mutants (*1*, *25*, *35*). Such changes in position and charge make it difficult to assign a particular catalytic or steric role in facilitating ring closure or deprotonation reactions. More likely, this residue may have a function in protein folding and stability.

The X-ray structure of EGFP-Y66L does not support a basic role for the carbonyl oxygen of Thr62. Elimination of water from the five-membered ring has been proposed to be initiated by abstraction of a proton from N67 via Thr62 (*14*) (mechanism A, Scheme 1). The separation between N67 and the carbonyl oxygen of Thr62 (O62) is 3.4 Å in EGFP-Y66L, essentially identical to that in GFPsol-S65G/Y66G (3.3 Å) (*14*). Yet, dehydration does not occur in either protein. In accord with this observation, the pK_a of an amide carbonyl oxygen is generally estimated to lie between 0 and −4 (*36*), highly unfavorable for a role as a proton acceptor.

In summary, we have shown that the replacement of an aromatic with an aliphatic residue (Y66L) in EGFP has a profound effect on the efficiency of dehydration of the cyclic intermediate. We propose that ejection of water from the five-membered ring is the final step in GFP chromophore maturation, and that this step is favorable only in the presence of an aromatic amino acid in position 66.

REFERENCES

1. Ormo, M., Cubitt, A. B., Kallio, K., Gross, L. A., Tsien, R. Y., and Remington, S. J. (1996) *Science 273*, 1392−1395.
2. Yang, F., Moss, L. G., and Phillips, G. N. (1996) *Nat. Biotechnol. 14*, 1246−1251.
3. Tsien, R. Y. (1998) *Annu. Rev. Biochem. 67*, 509−544.
4. Matz, M. V., Fradkov, A. F., Labas, Y. A., Savitsky, A. P., Zaraisky, A. G., Markelov, M. L., and Lukyanov, S. A. (1999) *Nat. Biotechnol. 17*, 969−973.
5. Lukyanov, K. A., Fradkov, A. F., Gurskaya, N. G., Matz, M. V., Labas, Y. A., Savitsky, A. P., Markelov, M. L., Zaraisky, A. G., Zhao, X., Fang, Y., Tan, W., and Lukyanov, S. A. (2000) *J. Biol. Chem. 275*, 25879−25882.
6. Matz, M. V., Lukyanov, K. A., and Lukyanov, S. A. (2002) *BioEssays 24*, 953−959.
7. Gross, L. A., Baird, G. S., Hoffman, R. C., Baldridge, K. K., and Tsien, R. Y. (2000) *Proc. Natl. Acad. Sci. U.S.A. 97*, 11990−11995.
8. Yarbrough, D., Wachter, R. M., Kallio, K., Matz, M. V., and Remington, S. J. (2001) *Proc. Natl. Acad. Sci. U.S.A. 98*, 462−467.
9. Wall, M. A., Socolich, M., and Ranganathan, R (2000) *Nat. Struct. Biol. 7*, 1133−1138.
10. Mizuno, H., Mal, T. K., Tong, K. I., Ando, R., Furuta, T., Ikura, M., and Miyawaki, A. (2003) *Mol. Cell 12*, 1051−1058.
11. Cubitt, A. B., Heim, R., Adams, S. R., Boyd, A. E., Gross, L. A., and Tsien, R. Y. (1995) *Trends Biochem. Sci. 20*, 448−455.
12. Heim, R., Prasher, D. C., and Tsien, R. Y. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 12501−12504.
13. Heim, R., Cubitt, A. B., and Tsien, R. Y. (1995) *Nature 373*, 663−664.
14. Barondeau, D. P., Putnam, C. D., Kassmann, C. J., Tainer, J. A., and Getzoff, E. D. (2003) *Proc. Natl. Acad. Sci. U.S.A. 100*, 12111−12116.
15. Siegbahn, P. E. M., Wirstam, M., and Zimmer, M. (2001) *Int. J. Quantum Chem. 81*, 169−186.
16. Schwede, T. F., Retey, J., and Schulz, G. E. (1999) *Biochemistry 38*, 5355−5361.
17. Baedeker, M., and Schulz, G. E. (2002) *Structure 10*, 61−67.
18. Erskine, P. T., Coates, L., Mall, S., Gill, R. S., Wood, S. P., Myles, D. A. A., and Cooper, J. B. (2003) *Protein Sci. 12*, 1741−1749.
19. Reid, B. G., and Flynn, G. C. (1997) *Biochemistry 36*, 6786−6791.
20. Leslie, A. G. (1999) *Acta Crystallogr. D55*, 1696−1702.
21. Collaborative Computational Project Number 4 (1994) *Acta Crystallogr. D50*, 760−763.
22. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) *Acta Crystallogr. D54*, 905−921.

23. Jones, T. A., Zou, J.-Y., Cowan, S. W., and Kjelgaard, M. (1991) *Acta Crystallogr. A47*, 110−119.
24. Elsliger, M.-A., Wachter, R. M., Hanson, G. T., Kallio, K., and Remington, S. J. (1999) *Biochemistry 38*, 5296−5301.
25. Palm, G. J., Zdanov, A., Gaitanaris, G. A., Stauber, R., Pavlakis, G. N., and Wlodawer, A. (1997) *Nat. Struct. Biol. 4*, 361−365.
26. Prescott, M., Ling, M., Beddoe, T., Oakley, A. J., Dove, S., Hoegh-Guldberg, O., Devenish, R. J., and Rossjohn, J. (2003) *Structure 11*, 275−284.
27. Petersen, J., Wilmann, P. G., Beddoe, T., Oakley, A. J., Devenish, R. J., Prescott, M., and Rossjohn, J. (2003) *J. Biol. Chem. 278*, 44626−44631.
28. Wachter, R. M., Brett, A. K., Heim, R., Kallio, K., Tsien, R. Y., Boxer, S. G., and Remington, S. J. (1997) *Biochemistry 36*, 9759−9765.
29. Badger, G. M. (1961) *The chemistry of heterocyclic compounds*, Academic Press, London.
30. Prendergast, F. G. (2000) *Nature 405*, 291−293.
31. Liu, Z.-J., Vysotski, E. S., Chen, C.-J., Rose, J. P., Lee, J., and Wang, B.-C. (2000) *Protein Sci. 9*, 2085−2093.
32. Silverman, D. N. (2000) *Biochim. Biophys. Acta 1458*, 88−103.
33. Ehrig, T., O'Kane, D. J., and Prendergast, F. G. (1995) *FEBS Lett. 367*, 163−166.
34. Wiehler, J., Jung, G., Seebacher, C., Zumbusch, A., and Steipe, B. (2003) *ChemBioChem 4*, 1164−1171.
35. Brejc, K., Sixma, T. K., Kitts, P. A., Kain, S. R., Tsien, R. Y., Ormö, M., and Remington, S. J. (1997) *Proc. Natl. Acad. Sci. U.S.A. 94*, 2306−2311.
36. Cox, R. A., and Yates, K. (1978) *J. Am. Chem. Soc. 100*, 3861−3867.

BI0361315