Oxidative Chemistry in the GFP Active Site Leads to Covalent Cross-Linking of a Modified Leucine Side Chain with a Histidine Imidazole: Implications for the Mechanism of Chromophore Formation[†,‡]

Matthew A. Rosenow, Hetal N. Patel, and Rebekka M. Wachter*

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

Received February 28, 2005; Revised Manuscript Received April 25, 2005

ABSTRACT: The mechanism of chromophore biosynthesis in green fluorescent protein (GFP) is triggered by a spontaneous main chain cyclization reaction of residues 65−67. Here, we demonstrate that the initially colorless Y66L variant, designed to trap chromophore precursor states, is oxidatively modified to generate yellow chromophores that absorb at 412 and 374 nm. High- and low-pH crystal structures determined to 2.0 and 1.5 Å resolution, respectively, are consistent with π-orbital conjugation of a planar Leu66-derived adduct with the imidazolinone ring, which is ~90 and 100% dehydrated, respectively. Time-, base-, and oxygen-dependent optical properties suggest that the yellow chromophores are generated from a 338 nm-absorbing intermediate, interpreted to be the Y66L analogue of the wild-type GFP chromophore. Generation of this species is catalyzed by a general base such as formate, and proceeds via a cyclization−oxidation−dehydration mechanism. The data suggest that a hydration−dehydration equilibrium exists in the cyclic form of the peptide, and that dehydration is favored upon extensive conjugation with the modified side chain. We conclude that the mechanism of GFP chromophore biosynthesis is not driven by the aromatic character of residue 66. In the low-pH X-ray structure, a highly unusual cross-link is observed between His148 and the oxidized Leu66 side chain, suggesting a conjugate addition reaction of the imidazole nitrogen to the highly electrophilic diene group of the yellow chromophore. The reactivity described here further expands the chemical diversity observed in the active site of GFP-like proteins, and may allow for covalent attachment of functional groups to the protein scaffold for catalytic purposes.

Green fluorescent protein (GFP)[1] is the founding member of a new class of proteins, the family of GFP-like proteins. Members of this family are used extensively as research tools in molecular and cell biology (*1*, *2*), because they are brightly colored with their appearance ranging from cyan to green to red (*3*−*5*). The coloration originates from a fluorescent entity that forms spontaneously in the protein's interior once the native fold is formed. Extensive color diversity in Anthozoa-derived GFPs appears to be coupled to the chemical diversity of chromophore structures, which depends on the extent of post-translational processing (*6*). These proteins have found exciting new applications as "fluorescent timers" in developmental biology experiments (*7*), and have been developed into novel imaging agents for multicolor and FRET labeling experiments (*1*, *8*). Highly sensitive fluorescence imaging techniques in conjunction with a large variety of genetically encodable fluorescent probes are providing cell biologists with the tools needed for precise resolution of cellular events (*9*−*11*).

The single-domain fold of GFP consists of an 11-stranded β-barrel with a central coaxial helix (*12*, *13*) that bears the three chromophore-forming residues, Ser65, Tyr66, and Gly67 in the wild type and Thr65, Tyr66, and Gly67 in enhanced GFP [EGFP, GFP-F64L/S65T, engineered for improved brightness (*14*)]. The mature chromophore is green-fluorescent and consists of an imidazolinone ring linked to a phenolic group by a methylidene bridge (Scheme 1). In some GFP homologues, further modifications from the green stage result in an extension of the chromophore π-system concomitant with a red shift of optical properties (*15*−*18*). However, the internal backbone cross-linking reaction appears to trigger all ensuing covalent modifications, irrespective of the protein's final color.

Soon after the gene was cloned (*19*) and the chemical nature of the chromophore was characterized (*20*, *21*), an outline for the GFP chromophore biosynthetic process was proposed. According to this mechanism (Scheme 1, mechanism A), nucleophilic attack of the amide nitrogen of Gly67 at the carbonyl carbon of Ser65 (*22*, *23*) leads to peptide ring closure followed by the elimination of water. The last step in this mechanism is the dehydrogenation of the Tyr66 $C_\alpha-C_\beta$ bond by oxidation via molecular oxygen, leading to full π-conjugation of the phenolic group with the imidazolinone ring. In all known GFP crystal structures, the guani-

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

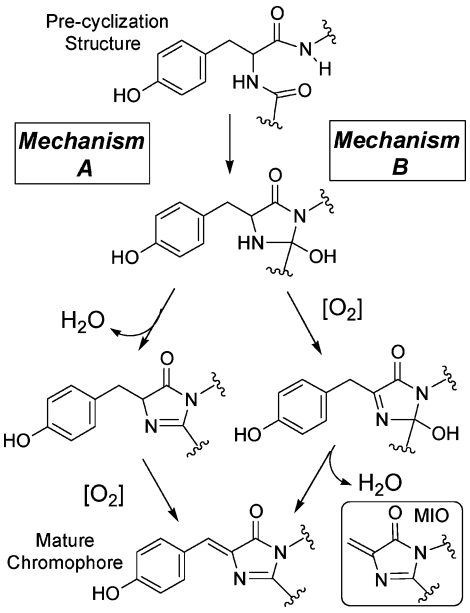
[‡] Coordinates for the X-ray structures of yellow EGFP-Y66L (374 and 412 nm species) have been deposited in the Protein Data Bank (entries 1Z1Q and 1Z1P).

* To whom correspondence should be addressed: Department of Chemistry and Biochemistry, Arizona State University, Tempe, AZ 85287-1604. Phone: (480) 965-8188. Fax: (480) 965-2747. E-mail: RWachter@asu.edu.

[1] Abbreviations: GFP, green fluorescent protein; EGFP, enhanced green fluorescent protein (GFP-F64L/S65T); MIO, methylidene imidazolone; HAL, histidine ammonia lyase.

10.1021/bi0503798 CCC: $30.25
Published on Web 05/21/2005

Scheme 1: Outline of the Mechanism of GFP Chromophore Formation[a]



[a] Mechanism A is cyclization followed by dehydration and oxidation; mechanism B is cyclization followed by oxidation and dehydration. The inset shows the chemical structure of the MIO cofactor of histidine ammonia lyase.

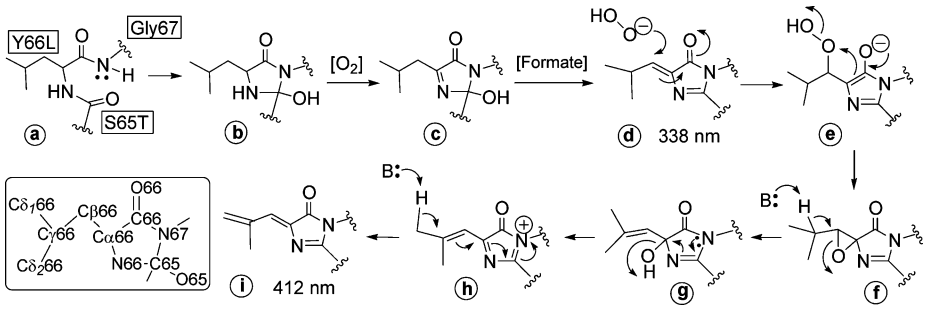
dinium group of Arg96 is hydrogen-bonded to the Tyr66 carbonyl oxygen, and the carboxyl group of Glu222 is positioned near the opposite face of the chromophore (*3*). Both Arg96 and Glu222 are highly conserved (*2*) and appear to facilitate the internal main chain cyclization reaction. Mechanistic proposals include abstraction of a proton from the Gly67 amide nitrogen (*24*) to activate the peptide for nucleophilic cycloaddition.

A large part of the π-skeleton of the mature GFP chromophore is derived from the phenolic side chain of Tyr66. Other aromatic residues (His, Phe, and Trp) have been substituted in this position with retention of fluorescence (*22*). To investigate the role of the aromatic group in fluorophore biosynthesis, GFP variants have been prepared that contain the Y66G (*25*) or Y66L substitution (*26*), mutants unable to generate the wild-type chromophore. Crystal structures of these variants are consistent with backbone condensation being the first chemical step in chromophore formation, as originally proposed by Tsien and co-workers (*23*). In the structure of the aerobic form of Y66G (*25*), the internal helix was reported to be cyclized to a hydroxylated imidazolidine dione, an entity derived from the tetrahedral intermediate generated by ring closure (intermediate 1, Scheme 1) with an additional oxygen atom bonded to C_α66 (for atom labels, see the inset of Scheme 2). Remarkably, under anaerobic conditions, the same variant was found to remain in the precyclization state (*25*), indicating a role for molecular oxygen in stabilizing the cyclic intermediate (*26*). This work demonstrates that GFP backbone cross-linking is not solely driven by mechanical compression of reacting atoms, as has been suggested (*27*), since distortions in the internal helix are observed in both the pre- and postcyclization state (*25*).

Recently, we have considered an alternate mechanism for GFP chromophore biosynthesis that consists of a cyclization−oxidation−dehydration sequence (Scheme 1, mechanism B) (*26*). This mechanistic hypothesis is based on the high-resolution X-ray structure of the colorless Y66L variant, in which the cyclic peptide is apparently trapped in the tetrahedral intermediate state by ring oxidation (Scheme 2, structure **c**). In this structure, the hydroxyl leaving group remains attached to the five-membered ring and the α-carbon of residue 66 exhibits trigonal planar geometry, suggesting that replacement of the aromatic Tyr66 with an aliphatic residue has a profound effect on the efficiency of ring dehydration. The structural data are in accord with a reaction pathway in which oxidation immediately follows ring closure, and dehydration completes the process of chromophore formation.

Aside from those in the family of GFP-like proteins, main chain condensation reactions have been described in unrelated proteins such as the ammonia lyases (*28*−*30*), and most recently in a tyrosine amino mutase (*31*). In these enzymes, an autocatalytic peptide backbone cyclization initiates the formation of a built-in electrophilic cofactor, the methylidene imidazolone group (MIO) (Scheme 1, inset) (*28*). Main chain condensation of an Ala-Ser-Gly sequence in histidine ammonia lyase (HAL) leads to formation of an imidazolinone ring analogous to the five-membered ring of the GFP chromophore. The attached dehydroalanine group is subsequently generated by elimination of water from the serine side chain, which is in a position equivalent to Tyr66 in GFP. Getzoff and co-workers have demonstrated that a transplantation of the HAL tripeptide into GFP (GFP_{hal}) results in a series of internal modifications consistent with generation of MIO in the center of the GFP β-barrel (*24*). The X-ray structures of GFP_{hal} support either formation of the exocyclic methylidene group of MIO or incorporation of an exocyclic carbonyl oxygen in place of the Y66S side chain. Dehydration of the five-membered ring appears to be incomplete, consistent with a chemical equilibrium between the hydration

Scheme 2: Proposed Mechanism of Formation of the Yellow Form of Y66L[a]



[a] The inset shows atom labels used throughout the text.

Table 1: Names of Y66L Spectroscopic Species

UV−vis absorbance bands (λ > 300 nm)	description	proposed chemical structures (Schemes 2 and 3)
none (colorless)	hydrated form of the heterocycle, likely the cyclic imine form (nonaromatic)	**c**
338 nm	C_α66−C_β66-desaturated adduct to the imidazolinone ring (dehydrated, aromatic), analogue of mature GFP chromophore, structurally similar to MIO	**d**
412 nm (yellow)	diene adduct to the imidazolinone ring (dehydrated, aromatic)	**i**
374 nm (yellow hue)	dienolate form of imidazolinone ring (dehydrated, aromatic), covalently cross-linked to His148 imidazole nitrogen	**k**

adduct and the dehydrated product (*26*). Covalent addition of nucleophilic compounds to the MIO-like group was not observed.

Here, we demonstrate that the initially colorless Y66L variant of EGFP is oxidatively modified to generate a UV−vis-absorbing chromophore with electrophilic properties similar to those of MIO. The data suggest that this entity is formed from an intermediate species that is the Y66L analogue of the green-fluorescent chromophore. Thus, the entire process of GFP chromophore formation may be mimicked in Y66L, and the mechanism does not appear to be driven by the aromatic character of residue 66.

EXPERIMENTAL PROCEDURES

Preparation of Yellow Y66L. Expression, purification, and removal of the His_6 tag of Y66L were carried out as described previously (*26*). Colorless Y66L was converted to a UV−vis-absorbing yellow species by incubating the protein with 2.0 M sodium formate, and either 100 mM cacodylic acid (pH 6.5) or 100 mM HEPES (pH 8.0). For kinetic experiments, 0.5 mg/mL protein was incubated at 31 °C for 18 days, and absorbance scans were collected in 2 day intervals. For experiments designed to investigate the effect of formate concentration, 0.5 mg/mL protein samples containing between 0 and 2.0 M sodium formate were incubated at 31 °C for 14 days, at both pH 6.5 and 8.0 (buffers described above). For experiments designed to investigate the effect of different types of anions, 0.5 mg/mL colorless EGFP-Y66L was incubated at pH 8.0 (50 mM HEPES and 20 mM NaCl) and pH 6.5 (50 mM sodium cacodylate and 20 mM NaCl) for a total of 15 days at 31 °C. Each sample contained one of the following additives at 1.0 M: sodium formate, sodium acetate, sodium gluconate, sodium chloride, potassium fluoride, and potassium bromide. Control samples contained protein and buffer with and without 20 mM NaCl. Spectra were collected every 4 days for 16 days. The maximum absorbance ratio (obtained after incubation for 3 weeks in 1.0 M formate) was used to determine the yield of yellow protein at each time point and for each type of additive. Optical absorption spectra were collected on a Shimadzu UV-2401PC spectrophotometer. Fluorescence measurements were collected on a SPEX Fluoromax-3 instrument (Jobin Yvon Horiba, Inc.).

Anaerobic Incubation of Colorless EGFP-Y66L. Colorless EGFP-Y66L at 1.8 mg/mL was incubated in 100 mM cacodylate (pH 6.5) or 100 mM HEPES (pH 8.0), with and without 2.0 M formate and under either atmospheric or anaerobic conditions, all at room temperature. To remove molecular oxygen, samples were degassed under vacuum and purged with argon three times for 15 min. Vials were sealed, transferred to an anaerobic chamber flooded with a mixture of N_2 and He gas, and incubated for 3 weeks at 23 °C. Absorption spectra were collected immediately after the vials were removed from the glovebox. Samples switched from anaerobic to aerobic conditions were allowed to incubate aerobically for an additional 3 weeks at 23 °C.

Crystallization and Diffraction Data Collection. EGFP-Y66L used for crystallization was incubated at 31 °C for 6 weeks in 2.0 M sodium formate, with 100 mM cacodylate (pH 6.5) or 100 mM HEPES (pH 8.0). The protein was concentrated to 15 mg/mL and crystallized via the hanging-drop vapor-diffusion method using 1 μL of protein and 1 μL of mother liquor. For crystals of the 412 nm-absorbing protein, the mother liquor contained 18−20% polyethylene glycol 8000, 200 mM calcium acetate, and 100 mM HEPES (pH 8.0). Rod-shaped crystals, typically 0.1 mm across and 0.5 mm long, grew at 4 °C within 7−10 days. For crystals of the 374 nm-absorbing species, the mother liquor contained 15−20% polyethylene glycol 8000, 200 mM calcium acetate, and 100 mM cacodylic acid (pH 6.5). Small rod-shaped crystals approximately 0.02 mm across and 0.1 mm long grew at room temperature in 10−15 days. Larger crystals, 0.3 mm across and 1.5 mm long, were grown in 3−5 days by macroseeding techniques, in which protein and mother liquor were mixed as described above and allowed to equilibrate for 3 days. Small crystals were washed in mother liquor, transferred to the equilibrated protein solution, and allowed to grow undisturbed for several weeks. To examine the optical properties of the crystal-derived protein, some crystals were washed in mother liquor and then redissolved in buffer to collect an absorbance scan. X-ray diffraction data were collected from single crystals soaked in a 30% glycerol/mother liquor solution for cryoprotection and flash-frozen in a stream of liquid N_2 gas (100 K). Data were collected using a 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 of 412 and 374 nm-absorbing EGFP-Y66L were collected (Table 1), and data sets were processed using MOSFLM (*32*) and SCALA within the CCP4 suite of programs (*33*). The colorless Y66L structure (PDB entry 1S6Z) (*26*), with residues 65−67 omitted, was used as a molecular replacement model. Refinement was carried out against 90% of the data using maximum likelihood least-squares minimization in CNS (*34*), where 10% of the data were used to estimate R_{free}. Initial rigid body refinement was followed by positional refinement in a stepwise fashion 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 (*35*). Ordered solvent molecules were modeled near the protein surface if within hydrogen bonding distance of appropriate protein atoms. Modeling of interior water molecules was postponed until the very end of the refinement. Strong positive difference electron density was observed in the protein interior, indicating the presence of a five-membered ring generated from the protein backbone atoms of residues 65−67, as previously observed in the colorless Y66L (*26*). When the crystallographic *R*-factor was reduced to 22% (374 nm species) and 23% (412 nm species), the heterocyclic ring and the leucine side chain of residue 66 (Y66L) were manually positioned into the density. A series of refinement tests in which the geometry restraints on the ring atoms were varied ruled out a puckered ring conformation. The use of weak bond length restraints and no bond angle restraints lead to a planar ring geometry and the elimination of difference density features above and below the ring.

Inspection of the 1.50 Å $2F_o - F_c$ map for the 374 nm species gave evidence of significant difference density features when the Leu66 side chain was modeled with tetrahedral carbons. Most of these features disappeared when angle restraints were eliminated, and upon convergence, all side chain atoms were found to be in-plane with the five-membered ring. Remaining positive difference density at the 4.5σ levels was observed near the terminal methyl group of Leu66 and adjacent to the imidazole ring of His148, which was not well-positioned in its density. Therefore, the van der Waals repulsion term for nonbonded contacts was reduced for C_γ, $C_{\delta1}$, and $C_{\delta2}$ of Leu66. Upon further refinement, the imidazole group was repositioned within covalent bonding distance of the terminal methyl group $C_{\delta1}$ of Leu66, completely accounting for the difference density. On the basis of the refined model for the 374 nm species, the electron density of the 412 nm species was re-examined. Though diffraction data quality for the 412 nm species was inferior, as indicated by a higher R_{merge} and a lower resolution (Table 1), the model refined to a planar Leu66 side chain when side chain geometry restraints were relaxed. The density for the 412 nm species was inconsistent with a covalent cross-link to His148 for the majority of the protein population.

RESULTS

The initially colorless Y66L variant slowly converts to a yellow form upon prolonged incubation in the presence of certain small anions, either at room temperature or at 30 °C. Incubation at pH 6.5 leads to a strong absorbance band centered around 374 nm (Figure 1AC), whereas incubation at pH 8.0 leads to a broad band centered at 412 nm, with significant contributions at 400 and 430 nm (Figure 1B,C). Both species are essentially nonfluorescent. To examine whether these modifications are relevant to the mechanistic chemistry of GFP chromophore biosynthesis, the optical properties and the chemistry of formation of the yellow chromophores were investigated and their structures determined.

Formation of Yellow Y66L Depends on an Aerobic Environment. We find that colorless Y66L is first converted to a 338 nm-absorbing form, a protein species that is observed only at high pH or in the presence of certain anions

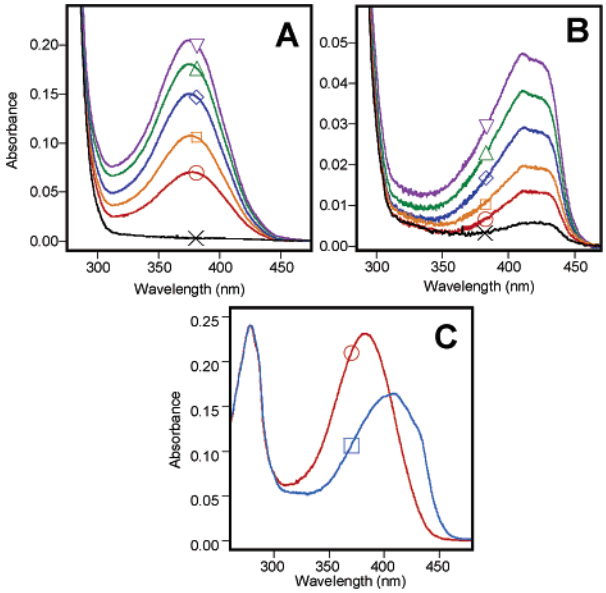


FIGURE 1: Absorbance spectra of yellow Y66L. Colorless Y66L was incubated (A) at pH 6.5 and (B) at pH 8.0 for 14 days at 31 °C in the presence of various concentrations of sodium formate: black (×) for 0 M, red (○) for 0.25 M, orange (□) for 0.50 M, blue (◇) for 1.0 M, green (△) for 1.5 M, and purple (▽) for 2.0 M. (C) Absorbance spectra of the yellow forms of Y66L used for crystal growth: red (○) for the protein incubated at pH 6.5 in 2 M formate for 3 weeks and blue (□) for the protein incubated at pH 8.0 in 2 M formate for 3 weeks.

(e.g., formate; see below). The absorbance data suggest that the 338 nm form contains a dehydrated aromatic heterocycle with a α-β desaturated leucine side chain, and is therefore related to the ammonia lyase cofactor MIO (*31*) (Table 1 and Scheme 2, structure **d**). The yellow Y66L forms absorbing at 374 and 412 nm appear to be further modifications of the 338 nm species, and are generated under aerobic conditions only (Figure 2A,B). Their yield is greatly improved in the presence of high concentrations of formate or a similar type of anion. No reversion of this process is detected over a period of several weeks, even if formate is removed from the sample by ultrafiltration. In the absence of formate at pH 6.5, no significant spectral features due to the yellow form are observed over a 19 day period, regardless of whether the incubation is carried out in the presence of absence of oxygen (black spectrum in Figure 1A and light blue and red spectra in Figure 2A). However, in the absence of formate at pH 8.0, weak absorbance bands do arise, with peaks centered at 412 nm under aerobic conditions (black spectrum in Figure 1B and light blue spectrum in Figure 2B) and at 330−340 nm under anaerobic conditions (red spectrum in Figure 2B). When samples are incubated anaerobically in the presence of 2 M formate, weak absorbance in the range of 330−340 nm appears at both pH 6.5 and 8.0 (green spectra in Figure 2A,B). If the anaerobic formate-containing samples are then transferred to aerobic conditions for an additional 19 days, the characteristic strong absorbance bands at 374 or 412 nm develop (orange spectra in Figure 2A,B), as in control reactions of entirely aerobic incubations (dark blue and purple spectra). The 338 nm band is also observed aerobically, but only early in the conversion to the yellow form. After 2 days, the 338 nm species rises (blue spectra in Figure 2C,D) and then falls at later time points as the 374 and 412 nm species become more

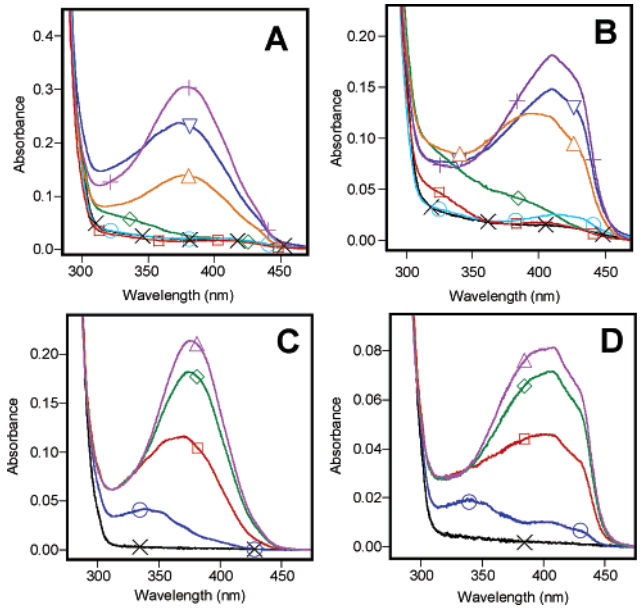


FIGURE 2: (A and B) Incubation of colorless Y66L at room temperature at (A) pH 6.5 and (B) pH 8.0: black (×) for starting material, light blue (○) for 19 days aerobic (no additives), red (□) for 19 days anaerobic (no additives), green (◇) for 19 days anaerobic (2 M formate), orange (△) for 19 days anaerobic followed by 19 days aerobic (2 M formate), dark blue (▽) for 19 days aerobic (2 M formate), and purple (+) for 38 days aerobic (2 M formate). (C and D) Incubation of colorless Y66L in the presence of 2.0 M formate at 31 °C under atmospheric conditions at (C) pH 6.5 and (D) pH 8.0: black (×) for 0 days, blue (○) for 2 days, red (□) for 6 days, green (◇) for 12 days, and purple (△) for 18 days.

prominent, consistent with a 338 nm precursor that is oxidized under atmospheric conditions to yield yellow protein. We conclude that dioxygen acts synergistically with formate to oxidize the 338 nm species, leading to the broad bands that lend the proteins their yellow hue. Though the 412 nm form of Y66L undergoes small spectral shifts as a function of pH, this species cannot be converted to the 374 nm form by simple pH adjustment, not even after prolonged incubation. However, incubation of the 374 nm form at pH 8.5 leads to a gradual red shift to 412 nm over a period of 2 months, indicating slow conversion of the 374 nm form to the 412 nm form.

Formation of Yellow Y66L Is Catalyzed by a General Base. The apparent oxidation process for generating absorbance bands at 374 or 412 nm is accelerated by large concentrations of anions with basic properties. For example, absorbance peak intensities triple when formate concentrations are increased from 0.25 to 2.0 M (Figure 1A,B). To characterize this process in more detail, colorless Y66L was incubated for 16 days at 31 °C in the presence of a series of additives, each at 1 M. At pH 6.5, 33.0% of the protein was converted to the yellow form with fluoride as the additive ($pK_a = 3.45$), 25.1% with formate ($pK_a = 3.75$), 16.1% with acetate ($pK_a = 4.74$), 1.0% with gluconate ($pK_a = 3.60$), 1.0% with chloride, 0.8% with bromide, and 1.0% in a negative control without additives. At pH 8.5, the yield was 16.5% with fluoride, 12.0% with formate, 8.2% with acetate, 4.7% with gluconate, 3.6% with chloride, 2.7% with bromide, and 3.6% for the negative control. Clearly, fluoride exhibits the strongest effect, though other anions with similar activity include formate and acetate. On the other hand, gluconate, chloride, and bromide do not facilitate the conversion to the yellow form. Overall, the trend in activity follows the anion size and the pK_a of the corresponding acid, but not its redox potential. A particular anion is active if it displays basic properties over the physiological pH range of 6−8 (pK_a = 3−5). However, the data rule out a role for redox chemistry mediated by the additive. Though formate is a strong reducing agent (standard reduction potential of −0.2 V vs the normal hydrogen electrode) and its presence could lead to the formation of superoxide anion radicals, neither fluoride (2.87 V) nor acetate is redox-active despite the effectiveness of each in catalyzing the modification of Y66L to a yellow form.

The simplest model for the generation of yellow Y66L is one in which, at high concentrations, the anion partitions into an interior cavity, and acts as a general base in a proton abstraction reaction that involves the imidazolinone ring or the remnant of the Leu66 side chain (Scheme 2). In GFP X-ray structures, a large interior cavity filled with several water molecules is located adjacent to the GFP chromophore (*12*, *36*). Similarly, Y66L contains a cavity filled with ordered solvent molecules that may play a role in proton shuttling events to facilitate chromophore maturation (*2*, *26*). Fluoride, formate, and acetate are small enough to partition into this cavity, whereas gluconate either is too large to be accommodated or cannot adopt the orientations necessary for proton abstraction. In the yellow-fluorescent GFP variants, binding of small monovalent anions has previously been characterized (*37*, *38*), and formate was shown to interact directly with the chromophore.

X-ray Structures of Yellow Y66L Are Consistent with Extensive Side Chain Desaturation and Ring Dehydration. The three-dimensional structures of the two yellow protein species absorbing at 474 and 412 nm were determined by X-ray crystallography. Protein utilized for crystallization was incubated for several weeks at pH 6.5 or 8.0 in the presence of 2 M formate (Figure 1C), and crystals were grown by the hanging drop vapor diffusion method. To better correlate absorption spectra with crystallographic results, a subset of crystals was washed in mother liquor and resolubilized in buffer. The crystal-derived protein gave solution spectra identical to those of the original protein pool used for crystal growth (Figure 1C), with an A_{280}/A_{380} absorbance ratio of 1.0 at pH 6.5, and an A_{280}/A_{410} absorbance ratio of 1.5 at pH 8.0, each judged to be the end point of the conversion.

In-house diffraction data were collected, and the structures were determined by molecular replacement using the colorless Y66L structure (*26*) as a model (PDB entry 1S6Z). Both protein species crystallized in the same space group as the colorless protein ($P2_12_12_1$) with nearly identical unit cell dimensions (Table 2). The structure of the 374 nm-absorbing protein was refined to 1.5 Å with an *R*-factor of 17.5%, and the structure of the 412 nm-absorbing protein was refined to 2.0 Å with an *R*-factor of 17.3%, each with good geometry. In both structural models, the hydroxyl leaving group was found to be ejected from C65, consistent with dehydration of the five-membered ring (Figure 3A,B, omit electron density map of the 374 nm species; Figure 3C, omit map of the 412 nm species). The electron densities are consistent with a localized water molecule that is 3.1 Å (374 nm form) and 2.9 Å (412 nm form) from C65. Dehydration is 100 and ~90% complete, respectively, as estimated from occupancy tests. As in the colorless Y66L structure, all ring atoms are

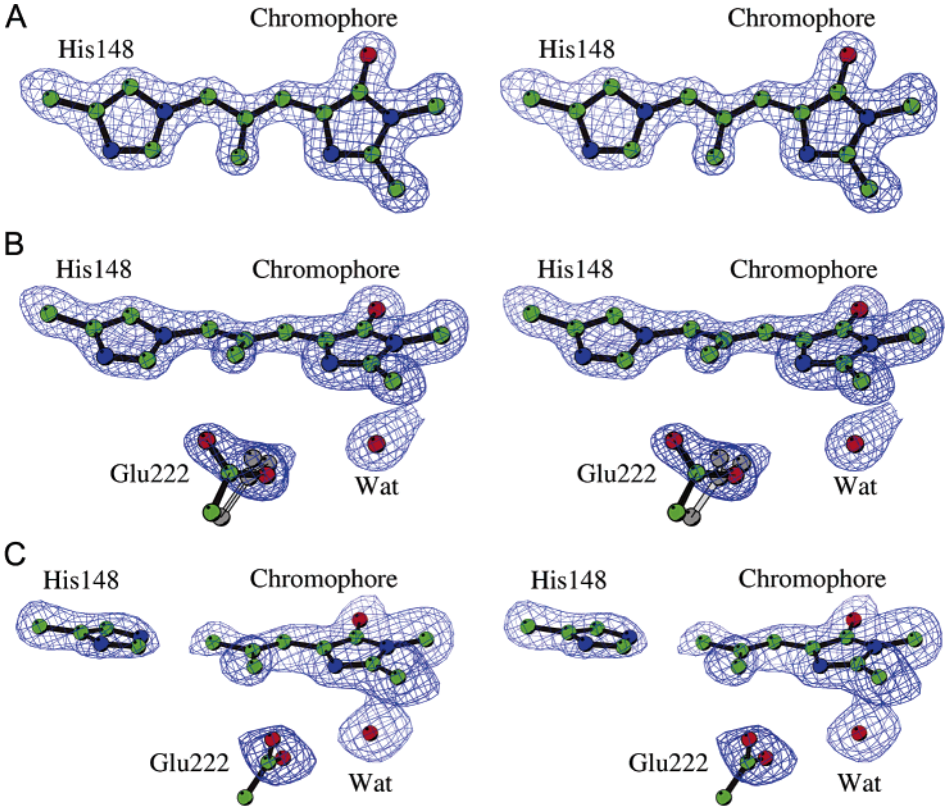


FIGURE 3: Stereoview of the $F_o - F_c$ electron density omit maps for the yellow forms of Y66L. Residues 65−67, the side chain carboxylate of Glu222, Wat449, and the imidazole ring of His148 were omitted from the final atomic model, and the map was calculated after 200 steps of crystallographic refinement. The atoms originally omitted from the refined model were then superimposed onto the map (oxygen in red, nitrogen in blue, and carbon in green). (A and B) Two different orientations of the 374 nm-absorbing form of Y66L, with the omit map contoured at 4σ. (C) The 412 nm-absorbing form of Y66L, with the omit map contoured at 3σ.

Table 2: Crystallographic Statistics of UV−Vis-Absorbing Forms of Y66L

	412 nm species	374 nm species
data collection and processing		
space group	$P2_12_12_1$	$P2_12_12_1$
cell dimensions *a*, *b*, *c* (Å)	50.91, 62.59, 69.62	50.93, 62.46, 69.39
total no. of observations	65188	173627
no. of unique reflections	14782	33847
completeness[a] (%) (shell)	95.3 (95.3)	94.1 (94.1)
$I/\sigma(I)$[a] (shell)	7.1 (2.0)	10.0 (2.6)
R_{merge}[a,b] (%) (shell)	10.3 (39.3)	5.0 (28.8)
resolution (Å)	2.00	1.50
refinement statistics		
resolution (Å)	30−2.00	30−1.50
no. of reflections	14757	33818
R_{work} (R_{free})	17.3 (23.1)	17.5 (20.7)
no. of protein atoms	1814	1814
no. of solvent atoms	162	296
bond length deviations (Å)	0.013	0.019
bond angle deviations (deg)	1.92	2.02
B-factor deviation for main chain atoms	2.87	1.82
B-factor deviation for side chain atoms	4.40	3.08

[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 of 2.11−2.00 Å for pH 8.0 and 1.58−1.50 Å for pH 6.5. [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}.

in-plane and C_α66 is sp^2-hybridized (trigonal planar). Visual inspection of difference density maps during refinement indicated that an aliphatic Leu66-derived side chain gave a poor fit to the electron density. Careful refinement with

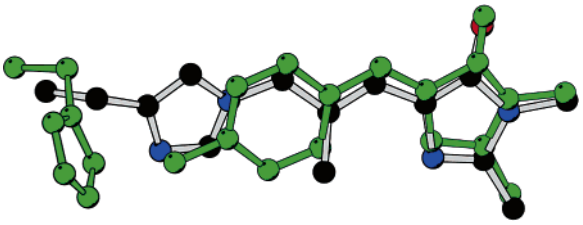
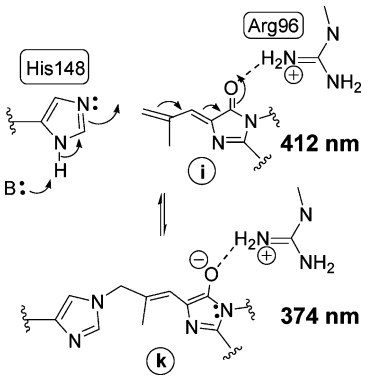


FIGURE 4: Overlay of crystallographic models derived from residues 65−67 and 148 in GFP-S65T (green) and the 374 nm form of Y66L (gray bonds, oxygen in red, nitrogen in blue, and carbon in black), generated by superposition of C_α atoms of residues 5−64 and 68−228 [GFP-S65T, PDB entry 1EMA (*13*)].

relaxed geometry restraints resulted in a model containing a flat side chain, with all atoms in-plane with the five-membered ring (Figure 3). The best crystallographic model for both yellow forms is consistent with extensive desaturation of the altered leucine and π-orbital overlap with the heterocycle (Scheme 2, structure **i**).

The X-ray Structure of the 374 nm Form of Y66L Implies a Covalent Cross-Link between the Modified Leu66 Side Chain and the His148 Imidazole. In both yellow Y66L structural models, refinement resulted in positioning of the His148 imidazole ring close to the Leu66 side chain, essentially filling the interior cavity generated by the Y66L substitution (Figure 3). In addition, refinement of the 374 nm species gave evidence of a remarkably close approach (2.1 Å) of the imidazole ring to the leucine-derived $C_{\delta1}$66, with significant positive difference density (4.5σ) remaining between these groups. Therefore, additional refinement was carried out in the absence of van der Waals repulsion terms

Scheme 3: Covalent Adduct Formation between the His148 Imidazole Ring and the Oxidized Leu66 Side Chain



for $C_{\delta 1}66$. The ensuing crystallographic model displayed a distance of 1.51 Å between His148 N_{ϵ} and $C_{\delta 1}66$, providing clear evidence that these atoms are covalently linked, probably by a single bond (Figure 3A,B). Though the X-ray data cannot distinguish between His148 rotamers that would place either N_{ϵ} or C_{δ} close to $C_{\delta 1}66$, chemical arguments favor bonding of the nucleophilic nitrogen to the modified Leu66 side chain (Scheme 3). The His148−Leu66 cross-linking reaction appears to be facilitated by a shift of His148 main chain atoms by 1.0 Å toward the protein interior. In the 374 nm form of the protein, Glu222 is best modeled in two conformations with roughly equal occupancy (Figure 3B), possibly due to different charge states of the carboxylic acid. One conformation is reminiscent of that present in GFP-S65T (*12*) and colorless Y66L (*26*), whereas the other is more similar to that found in wild-type GFP (*39*).

As opposed to the 374 nm form, the best crystallographic model for the 412 nm species indicates that most of the protein population does not contain the His148−Leu66 cross-link (Figure 3C). In the absence of van der Waals repulsion terms for $C_{\delta 1}66$, refinement results in a separation of 2.64 Å between His148 N_{ϵ} and $C_{\delta 1}66$. However, weak positive difference density features remain between these atoms. Evidently, the 412 nm form of the protein contains a mixture of states that, in addition to the major species modeled, include at least two minor populations: an electrophilic adduct to His148 (Scheme 3) and the hydrated form of the imidazolinone ring.

DISCUSSION

Air Oxidation Appears To Involve the Y66L Analogue of the Wild-Type GFP Chromophore. In this work, we describe the spontaneous formation of an UV−vis-absorbing chromophore in the Y66L variant of EGFP (for a description of different Y66L species, see Table 1). We have previously demonstrated that the colorless form of this protein contains an oxidized five-membered ring between residues 65 and 67 (Scheme 2, structure **c**) (*26*). Evidently, additional oxidative chemistry is responsible for the generation of a fully π-conjugated system that includes a diene adduct to the imidazolinone ring (Scheme 2, structure **i**). On the basis of optical properties in conjunction with X-ray structural data, we propose the following mechanism for generation of this yellow chromophore (Scheme 2). Dehydration of the colorless cyclic intermediate **c** (*26*) is base-catalyzed, and leads to a 338 nm-absorbing species (structure **d)**, proposed to be the Y66L analogue of the mature GFP chromophore. Generation of the 338 nm species requires the removal of a proton from $C_{\beta}66$, a process that may be relevant to the normal mechanism of GFP chromophore formation (*26*). In Y66L, proton abstraction appears to be facilitated by an elevated pH or high concentrations of formate as a general base. Under aerobic conditions, the electrophilic nature of the 338 nm form leads to additional oxidation events that may be mediated by activated oxygen species such as hydrogen peroxide. After elimination of water, the proposed epoxide intermediate (Scheme 2, structure **f**) may yield the fully conjugated species **i**, in accord with X-ray crystallographic observations of a planar chromophore (Figure 3). Species **i** exhibits a broad absorbance band centered at 412 nm (Figure 1B,C), and is proposed to be the diene adduct to the imidazolinone ring.

We propose that the 338 nm form is a conjugated π-system analogous to the wild-type GFP chromophore, and chemically related to the built-in MIO cofactor of the ammonia lyases (*29*) and tyrosine amino mutase (*31*) (Scheme 1, inset). Optical evidence for MIO was first provided by UV difference spectroscopy in the histidine and phenylalanine ammonia lyases (*40*), where an absorbance band at 305−310 nm was detected. The MIO group in these enzymes is apparently not aromatic, since the ring nitrogen equivalent to N67 (GFP) is sp^3-hybridized in the histidine ammonia lyase crystal structure (*28*). Mutants of this enzyme that do contain an aromatic heterocycle with a sp^2-hybridized ring nitrogen display a red-shifted band with absorbance at 335−340 nm (*29*, *41*). Reaction of the highly electrophilic MIO group with cysteine and dioxygen has been reported to lead to a nitrogen adduct that also absorbs at 335−340 nm, in accord with optical properties of a small-molecule model compound containing the aromatic imidazolinone ring (*42*). Therefore, the Y66L absorbance band centered at 338 nm (Scheme 2, structure **d)** is explained well by an aromatic imidazolinone ring, entirely analogous to the wild-type GFP chromophore (Scheme 1).

A Highly Unusual Side Chain Cross-Link between His148 and Modified Leu66. The covalent cross-link observed by X-ray crystallography (Figure 3) is thought to be a result of the highly electrophilic nature of the 412 nm species, the diene adduct to the imidazolinone ring (Scheme 2, structure **i**). We propose that at low pH, His148 moves toward the oxidized Leu66 side chain, and the nucleophilic imidazole nitrogen adds to the fully oxidized leucine side chain in a Michael-type conjugate addition reaction (Scheme 3). The attachment of the yellow chromophore to His148 via a single bond effects a blue shift of the absorbance maximum from 412 to 374 nm, consistent with less π-orbital delocalization. The 374 nm species is ascribed to the dienolate form, with the negative charge electrostatically stabilized by a salt bridge to Arg96 (Scheme 3, structure **k**). Dienolates have been identified as intermediates in enzymatic reactions such as in ketosteroid isomerase (*43*) and 2,4-dienoyl-CoA reductase (*44*). In yellow Y66L, the covalent adduct between Leu66 and His148 appears to slowly reverse at elevated pH. An appropriately cross-linked peptide could not be identified by proteolysis and mass spectrometry, lending additional support to a chemically labile adduct.

Collectively, the cross-linking results suggest that engineered reactivities in the GFP chromophore pocket could

be used as specific covalent attachment sites in attempts to chemically modify the protein scaffold for catalytic and other purposes (*45*).

Ring Dehydration May Be a Function of Side Chain Conjugation. The crystallographic data obtained for different forms of Y66L indicate that the extent of desaturation of the leucine side chain is an important factor in generating dehydrated populations of the heterocycle. The dehydrated imidazolinone ring appears to be favored in structures that allow for extensive π-conjugation with the modified side chain. In colorless Y66L, a 100% hydration adduct (tetrahedral intermediate) is observed, consistent with the nonaromatic imine form of the heterocycle (Scheme 2, structure **c**) (*26*). However, the 338 nm species is likely dehydrated, suggesting that the ejection of water may be favored upon desaturation of the C_α66−C_β66 bond. In the crystallographic model of the 412 nm species, dehydration is ~90% complete, and most of the side chain is oxidized to the diene form (Figure 3C). In the crystal structure of the 374 nm species, dehydration is 100% complete, in step with 100% side chain oxidation (Figure 3A,B). On the basis of these observations, we propose that the energy difference between the hydrated and dehydrated heterocycle depends on the degree of π-overlap with the side chain, consistent with a hydration−dehydration equilibrium previously suggested for green-fluorescent GFPs (*26*). In support of this hypothesis, X-ray structures of GFP_{hal} have revealed a mixture of populations, with either less than 10% (GFP_{hal}) or more than 85% (GFP_{hal}-H148G) of the cyclopentyl group in the hydration adduct form (*24*).

Implications for the Mechanism of GFP Chromophore Formation. Remarkably, the entire mechanism of chromophore formation appears to be intact in Y66L. Generation of the Y66L analogue of the mature GFP chromophore evidently proceeds via a cyclization−oxidation−dehydration mechanism (Scheme 1, mechanism B). Data presented here support a hydration−dehydration equilibrium in the cyclic form of the peptide, in which the hydration adduct is favored before, and the dehydrated form is favored after oxidation of the C_α66−C_β66 bond to a double bond. Though the modification pathway in Y66L clearly partitions through oxidation prior to the ejection of water, the productive pathway leading to the wild-type GFP chromophore is more difficult to disentangle. This pathway may partition through an unstable dehydrated intermediate that is productive for generation of the green-fluorescent chromophore, but unproductive for generation of the Y66L analogue of the chromophore. Perhaps the most important mechanistic aspect of GFP chromophore biosynthesis is the strategy employed by the protein to stabilize the peptide condensation product. A trapping mechanism implies that the cyclized state of the peptide is modified in such a way that its stability is increased relative to that of the precyclization state. Barondeau et al. (*24*) favor trapping via ejection of water from the enolic form of the ring to generate an aromatic heterocycle. The Y66L data argue against this idea, since this variant generates a stable cyclic structure in the hydrated form (*26*). For this reason, we suggest that stabilization of the heterocycle involves ring oxidation via dehydrogenation by molecular oxygen.

REFERENCES

1. Verkhusha, V., and Lukyanov, K. A. (2004) The molecular properties and applications of Anthozoa fluorescent proteins and chromoproteins, *Nat. Biotechnol. 22*, 289−296.
2. Tsien, R. Y. (1998) The green fluorescent protein, *Annu. Rev. Biochem. 67*, 509−544.
3. Matz, M. V., Fradkov, A. F., Labas, Y. A., Savitsky, A. P., Zaraisky, A. G., Markelov, M. L., and Lukyanov, S. A. (1999) Fluorescent proteins from nonbioluminescent Anthozoa species, *Nat. Biotechnol. 17*, 969−973.
4. 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) Natural animal coloration can be determined by a nonfluorescent green fluorescent protein homolog, *J. Biol. Chem. 275*, 25879−25882.
5. Matz, M. V., Lukyanov, K. A., and Lukyanov, S. A. (2002) Family of the green fluorescent protein: Journey to the end of the rainbow, *BioEssays 24*, 953−959.
6. Shagin, D. A., Barsova, E. V., Yanushevich, Y. G., Fradkov, A. F., Lukyanov, K. A., Labas, Y. A., Semenova, T. N., Ugalde, J. A., Meyers, A., Nunez, J. M., Widder, E. A., Lukyanov, S. A., and Matz, M. V. (2004) GFP-like proteins as ubiquitous metazoan superfamily: Evolution of functional features and structural complexity, *Mol. Biol. Evol. 21*, 841−850.
7. Terskikh, A., Fradkov, A. F., Ermakova, G., Zaraisky, A., Tan, P., Kajava, A. V., Zhao, X., Lukyanov, S., Matz, M., Kim, S., Weissman, I., and Siebert, P. (2000) "Fluorescent timer": Protein that changes color with time, *Science 290*, 1585−1588.
8. Shaner, N. C., Campbell, R. E., Steinbach, P. A., Giepmans, B. N. G., Palmer, A. E., and Tsien, R. Y. (2004) Improved monomeric red, orange and yellow fluorescent proteins derived from *Discosoma* sp. red fluorescent protein, *Nat. Biotechnol. 22*, 1567−1572.
9. Lippincott-Schwartz, J., and Patterson, G. H. (2003) Development and use of fluorescent protein markers in living cells, *Science 300*, 87−91.
10. Hanson, G. T., Aggeler, R., Oglesbee, D., Cannon, M., Capaldi, R. A., Tsien, R. Y., and Remington, S. J. (2004) Investigating mitochondrial redox potential with redox-sensitive green fluorescent protein indicators, *J. Biol. Chem. 279*, 13044−13053.
11. Dooley, C. M., Dore, T. M., Hanson, G. T., Jackson, W. C., Remington, S. J., and Tsien, R. Y. (2004) Imaging dynamic redox changes in mammalian cells with green fluorescent protein indicators, *J. Biol. Chem. 279*, 22284−22293.
12. Ormo, M., Cubitt, A. B., Kallio, K., Gross, L. A., Tsien, R. Y., and Remington, S. J. (1996) Crystal structure of the *Aequorea victoria* green fluorescent protein, *Science 273*, 1392−1395.
13. Yang, F., Moss, L. G., and Phillips, G. N. (1996) The molecular structure of green fluorescent protein, *Nat. Biotechnol. 14*, 1246−1251.
14. Cormack, B. P., Valdivia, R. H., and Falkow, S. (1996) FACS-optimized mutants of the green fluorescent protein (GFP), *Gene 173*, 33−38.
15. Gross, L. A., Baird, G. S., Hoffman, R. C., Baldridge, K. K., and Tsien, R. Y. (2000) The structure of the chromophore within DsRed, a red fluorescent protein from coral, *Proc. Natl. Acad. Sci. U.S.A. 97*, 11990−11995.
16. Yarbrough, D., Wachter, R. M., Kallio, K., Matz, M. V., and Remington, S. J. (2001) Refined crystal structure of DsRed, a red fluorescent protein from coral, at 2.0-Å resolution, *Proc. Natl. Acad. Sci. U.S.A. 98*, 462−467.
17. Wall, M. A., Socolich, M., and Ranganathan, R. (2000) The structural basis for red fluorescence in the tetrameric GFP homolog DsRed, *Nat. Struct. Biol. 7*, 1133−1138.
18. Remington, S. J., Wachter, R. M., Yarbrough, D. K., Branchaud, B. P., Anderson, D. C., Kallio, K., and Lukyanov, K. A. (2005) Zfp538, a yellow fluorescent protein from *Zoanthus*, contains a novel three-ring chromophore, *Biochemistry 44*, 202−212.
19. Prasher, D. C., Eckenrode, V. K., Ward, W. W., Prendergast, F. G., and Cormier, M. J. (1992) Primary structure of the *Aequorea victoria* green-fluorescent protein, *Gene 111*, 229−233.
20. Cody, C. W., Prasher, D. C., Westler, W. M., Prendergast, F. G., and Ward, W. W. (1993) Chemical structure of the hexapeptide chromophore of the *Aequorea* green-fluorescent protein, *Biochemistry 32*, 1212−1218.
21. Niwa, H., Inouye, S., Hirano, T., Matsuno, T., Kojima, S., Kubota, M., Ohashi, M., and Tsuji, F. I. (1996) Chemical nature of the

light emitter of the *Aequorea* green fluorescent protein, *Proc. Natl. Acad. Sci. U.S.A. 93*, 13617−13622.
22. Heim, R., Prasher, D. C., and Tsien, R. Y. (1994) Wavelength mutations and posttranslational autoxidation of green fluorescent protein, *Proc. Natl. Acad. Sci. U.S.A. 91*, 12501−12504.
23. Cubitt, A. B., Heim, R., Adams, S. R., Boyd, A. E., Gross, L. A., and Tsien, R. Y. (1995) Understanding, improving, and using green fluorescent proteins, *Trends Biochem. Sci. 20*, 448−455.
24. Barondeau, D. P., Kassmann, C. J., Tainer, J. A., and Getzoff, E. D. (2005) Understanding GFP chromophore biosynthesis: Controlling backbone cyclization and modifying post-translational chemistry, *Biochemistry 44*, 1960−1970.
25. Barondeau, D. P., Putnam, C. D., Kassmann, C. J., Tainer, J. A., and Getzoff, E. D. (2003) Mechanism and energetics of green fluorescent protein chromophore synthesis revealed by trapped intermediate structures, *Proc. Natl. Acad. Sci. U.S.A. 100*, 12111−12116.
26. Rosenow, M. A., Huffman, H. A., Phail, M. E., and Wachter, R. M. (2004) The crystal structure of the Y66L variant of green fluorescent protein supports a cyclization-oxidation-dehydration mechanism for chromophore maturation, *Biochemistry 43*, 4464−4472.
27. Branchini, B. R., Nemser, A. R., and Zimmer, M. (1998) 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.
28. Schwede, T. F., Retey, J., and Schulz, G. E. (1999) Crystal structure of histidine ammonia-lyase revealing a novel polypeptide modification as the catalytic electrophile, *Biochemistry 38*, 5355−5361.
29. Baedeker, M., and Schulz, G. E. (2002) Autocatalytic peptide cyclization during chain folding of histidine ammonia-lyase, *Structure 10*, 61−67.
30. Calabrese, J. C., Jordan, D. B., Boodhoo, A., Sariaslani, S., and Vennelli, T. (2004) Crystal structure of phenylalanine ammonia lyase: Multiple helix dipoles implicated in catalysis, *Biochemistry 43*, 11403−11416.
31. Christenson, S. D., Liu, W., Toney, M. D., and Shen, B. (2003) A novel 4-mehylideneimidazole-5-one-containing tyrosine aminomutase in enediyne antitumor antibiotic C-1027 biosynthesis, *J. Am. Chem. Soc. 125*, 6062−6063.
32. Leslie, A. G. (1999) Integration of macromolecular diffraction data, *Acta Crystallogr. D55*, 1696−1702.
33. Collaborative Computational Project Number 4 (1994) The CCP4 suite: Programs for protein crystallography, *Acta Crystallogr. D50*, 760−763.
34. 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) Crystallography and NMR systems: A new software suite for macromolecular structure determination, *Acta Crystallogr. D54*, 905−921.
35. Jones, T. A., Zou, J.-Y., Cowan, S. W., and Kjelgaard, M. (1991) Improved methods for building protein models in electron density maps and the location of errors in these models, *Acta Crystallogr. A47*, 110−119.
36. Wachter, R. M., Elsliger, M.-A., Kallio, K., Hanson, G. T., and Remington, S. J. (1998) Structural basis of spectral shifts in the yellow-emission variants of green fluorescent protein, *Structure 6*, 1267−1277.
37. Wachter, R. M., and Remington, S. J. (1999) Sensitivity of the yellow variant of green fluorescent protein to halides and nitrate, *Curr. Biol. 9*, R628−R629.
38. Wachter, R. M., Yarbrough, D., Kallio, K., and Remington, S. J. (2000) Crystallographic and energetic analysis of binding of selected anions to the yellow variants of green fluorescent protein, *J. Mol. Biol. 301*, 159−173.
39. Brejc, K., Sixma, T. K., Kitts, P. A., Kain, S. R., Tsien, R. Y., Ormö, M., and Remington, S. J. (1997) Structural basis for dual excitation and photoisomerization of the *Aequorea victoria* green fluorescent protein, *Proc. Natl. Acad. Sci. U.S.A. 94*, 2306−2311.
40. Rother, D., Merkel, D., and Retey, J. (2000) Spectroscopic evidence for a 4-methylidene imidazol-5-one in histidine and phenylalanine ammonia-lyases, *Angew. Chem., Int. Ed. 39*, 2462−2464.
41. Baedeker, M., and Schulz, G. E. (2002) Structures of two histidine ammonia-lyase modifications and implications for the catalytic mechanism, *Eur. J. Biochem. 269*, 1790−1797.
42. Galpin, J. D., Ellis, B. E., and Tanner, M. E. (1999) The inactivation of histidine ammonia-lyase by L-cysteine and oxygen: Modification of the electrophilic center, *J. Am. Chem. Soc. 121*, 10840−10841.
43. Pollack, R. M. (2004) Enzymatic mechanisms for catalysis of enolization: Ketosteroid isomerase, *Bioorg. Chem. 32*, 341−353.
44. Fillgrove, K. L., and Anderson, V. E. (2001) The mechanism of dienoyl-CoA reduction by 2,4-dienoyl-CoA reductase is stepwise: Observation of a dienolate intermediate, *Biochemistry 40*, 12412−12421.
45. Tann, C.-M., Qi, D., and Distefano, M. D. (2001) Enzyme design by chemical modification of protein scaffolds, *Curr. Opin. Chem. Biol. 5*, 696−704.

BI0503798