J|A|C|S
ARTICLES
Published on Web 03/21/2006

Understanding GFP Posttranslational Chemistry: Structures of Designed Variants that Achieve Backbone Fragmentation, Hydrolysis, and Decarboxylation

David P. Barondeau, Carey J. Kassmann, John A. Tainer, and Elizabeth D. Getzoff*

Contribution from the Department of Molecular Biology, The Skaggs Institute for Chemical Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037

Received September 27, 2005; E-mail: edg@scripps.edu

Abstract: The green fluorescent protein (GFP) creates a fluorophore out of three sequential amino acids by promoting spontaneous posttranslational modifications. Here, we use high-resolution crystallography to characterize GFP variants that not only undergo peptide backbone cyclization but additional denaturation-induced peptide backbone fragmentation, native peptide hydrolysis, and decarboxylation reactions. Our analyses indicate that architectural features that favor GFP peptide cyclization also drive peptide hydrolysis. These results are relevant for the maturation pathways of GFP homologues, such as the kindling fluorescent protein and the Kaede protein, which use backbone cleavage to red-shift the spectral properties of their chromophores. We further propose a photochemical mechanism for the decarboxylation reaction, supporting a role for the GFP protein environment in facilitating radical formation and one-electron chemistry, which may be important in activating oxygen for the oxidation step of chromophore biosynthesis. Together, our results characterize GFP posttranslational modification chemistry with implications for the energetic landscape of backbone cyclization and subsequent reactions, and for the rational design of predetermined spontaneous backbone cyclization and cleavage reactions.

Introduction

Proteins are often driven by evolutionary selection for their biological functions to augment the chemical reactivity or properties of natural amino acids through posttranslational modifications.[1] Although many of these modifications require additional enzymes or chaperones, some are self-catalyzed, including the tripeptide backbone cyclization reactions that create the electrophilic catalysts[2] of the enzymes histidine ammonia lyase (HAL),[3] phenylalanine ammonia lyase (PAL),[4] tyrosine aminomutase (TAM),[5] and the fluorophores of green fluorescent protein (GFP)[6,7] and its homologues, such as red fluorescent protein (RFP).[8,9] These spontaneous backbone cyclization reactions have a significant impact on biology, medicine, and biotechnology. Deficiencies in HAL activity, the nonoxidative elimination of the α-amino group as part of the histidine degradation pathway, lead to the rare hereditary metabolic disorder histidinemia that is often characterized by mental retardation and speech defects;[10] PAL may be used to degrade phenylalanine and combat the metabolic disease phenylketonuria;[11] and TAM is a key enzyme in enediyne antitumor antibiotic C-1027 biosynthesis.[5] Moreover, the spontaneous biosynthesis and tunable fluorescence properties of the chro-

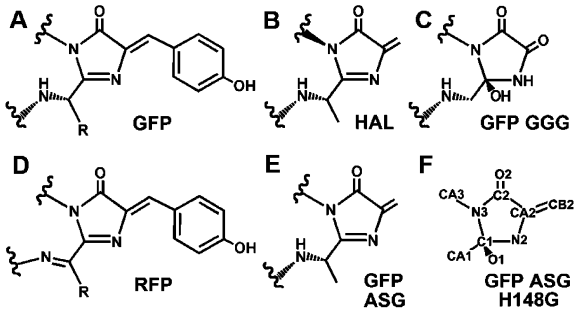


Figure 1. Molecular structures of imidazolone products after backbone cyclization and other posttranslational modifications. (A) GFP chromophore (PDB (Protein Data Bank) 1EMA), (B) HAL MIO catalyst (PDB 1B8F), (C) GFP GGG posttranslational modification (PDB 1QYQ), (D) RFP chromophore (PDB 1GGX), (E) GFP ASG chromophore (PDB 1YJF), and (F) atom labels for GFP ASG H148G (PDB 1YJ2) nonaromatic posttranslational modification.

(1) Okeley, N. M.; van der Donk, W. A. *Chem. Biol.* **2000**, *7*, R159−R171.
(2) Poppe, L. *Curr. Opin. Chem. Biol.* **2001**, *5*, 512−524.
(3) Schwede, T. F.; Retey, J.; Schulz, G. E. *Biochemistry* **1999**, *38*, 5355−5361.
(4) (a) Calabrese, J. C.; Jordan, D. B.; Boodhoo, A.; Sariaslani, S.; Vannelli, T. *Biochemistry* **2004**, *43*, 11403−11416. (b) Ritter, H.; Schulz, G. E. *Plant Cell* **2004**, *16*, 3426−3436.
(5) Christenson, S. D.; Liu, W.; Toney, M. D.; Shen, B. *J. Am. Chem. Soc.* **2003**, *125*, 6062−6063.
(6) Ormö, M.; Cubitt, A. B.; Kallio, K.; Gross, L. A.; Tsien, R. Y.; Remington, S. J. *Science* **1996**, *273*, 1392−1395.
(7) Yang, F.; Moss, L. G.; Phillips, G. N., Jr. *Nat. Biotech.* **1996**, *14*, 1246−1251.
(8) Gross, L. A.; Baird, G. S.; Hoffman, R. C.; Baldridge, K. K.; Tsien, R. Y. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 11990−11995.
(9) (a) Wall, M. A.; Socolich, M.; Ranganathan, R. *Nat. Struct. Biol.* **2000**, *7*, 1133−1138. (b) Yarbrough, D.; Wachter, R. M.; Kallio, K.; Matz, M. V.; Remington, S. J. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 462−467.
(10) Taylor, R. G.; Levy, H. L.; McInnes, R. R. *Mol. Biol. Med.* **1991**, *8*, 101−116.
(11) Sarkissian, C. N.; Shao, Z.; Blain, F.; Peevers, R.; Su, H.; Heft, R.; Chang, T. M.; Scriver, C. R. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 2339−2344.

Table 1. Maturation Properties for GFP and Variants

variant	chromophore tripeptide	mutations[a]	posttranslational outcome	fragmentation/hydrolysis	ref
wt GFP	SYG		mature oxidized chromophore	no	21
GFPsol	TYG		mature oxidized chromophore	no	21
R96A	TYG	R96A	slows maturation to months	no	23
ASG R96A	ASG	S65A Y66S R96A (wt F64)	uncyclized	no	22
ASG G67A	ASA	S65A Y66S G67A	uncyclized	no	22
GSG V68G	GSG	S65G Y66S V68G	uncyclized	no	22
Y66L	TLG	Y66L	cyclized, O1 hydroxyl present	?	24
GGG (Gly-Gly-Gly)	GGG	S65G Y66G	anaerobic: uncyclized aerobic: cyclized, O1 hydroxyl present, oxygen added at Y66 Cα	anaerobic: no aerobic: fragmentation	23
ASG (GFPhal)	ASG	S65A Y66S (wt F64)	cyclized, Y66S side-chain dehydration	fragmentation	22 and Figure 2B
ASG H148G	ASG	S65A Y66S H148G (wt F64)	cyclized, O1 hydroxyl present, Y66S side-chain dehydration	fragmentation	22
ASG F64L	ASG	S65A Y66S	cyclized, O1 hydroxyl present, Y66S side-chain dehydration	fragmentation	Figure 4A
GAG	GAG	S65G Y66A	cleaved 66−67 peptide	hydrolysis	Figure 3A
GSG	GSG	S65G Y66S	cleaved 66−67 peptide, Y66S side-chain dehydration	hydrolysis	Figure 3C
GSG F64	GSG	S65G Y66S (wt F64)	cleaved 66−67 peptide, Y66S side-chain dehydration, S65G decarboxylation	hydrolysis	Figure 4C

[a] Mutants are in a GFPsol (F64L S65T F99S M153T V163A) background unless otherwise stated.

mophores of GFP,[6,7] its homologues,[12] and mutants,[13,14] have revolutionized in vivo molecular tagging and cell labeling. Despite dramatically different protein environments and architectures, these proteins contain posttranslational modifications with similarities in both the resulting five-membered imidazolone moieties (Figure 1) and their biosynthetic reaction mechanisms. Understanding how the protein architectures drive these amino acid transformations may lead to the design of proteins with novel catalytic or reporter properties.

Both the GFP fluorophore and the HAL/PAL/TAM 4-methylidene-imidazole-5-one (MIO) posttranslational modifications (Figure 1A,B, respectively) entail three major biosynthetic steps: backbone cyclization via covalent bond formation between glycine nitrogen (Gly67 in GFP, Gly144 in HAL) and carbonyl carbon atoms (Ser65 in GFP, Ala142 in HAL), dehydration of this same carbonyl carbon atom, and either Tyr66 oxidation (GFP)[13,15] or Ser143 dehydration (HAL)[3,16] reactions to generate Cα−Cβ exocyclic double bonds and mature ring systems. In RFP, the electronic conjugation of the GFP-like chromophore is extended with a fourth major synthetic step, a second oxidation reaction that generates a double bond between backbone nitrogen and Cα atoms (Figure 1D).[8,9] Alternately, the RFP homologues Kaede[17] and the kindling fluorescent protein (KFP)[18,19] extend the electronic conjugation of the chromophore by creating double bonds through distinct, spontaneous, backbone cleavage reactions. Interestingly, the cleavage reaction for KFP may be critical for its light-induced isomerization and photoswitching.[20]

In GFP, biosynthesis of the tripeptide chromophore (Ser65-Tyr66-Gly67) appears robust; proteins with mutations that alter Ser65, modify residues adjacent to the chromophore, or substitute Tyr66 with an aromatic amino acid are able to synthesize fluorophores.[21] Recent studies to better understand the driving force and chemistry for GFP posttranslational modifications have focused on further modifications of the Ser-Tyr-Gly tripeptide and adjacent residues (Table 1), structural characterization of the resultant proteins, and extrapolation to the native biosynthetic mechanism.[22−24] We summarize these results in Table 1 and highlight four examples of altered or unusual posttranslational chemistry in GFP relevant to this study: (1) an oxidative cross-link between the Y66L-substituted chromophore and His148;[24,25] (2) chromophore-mediated decarboxylation of Glu222 upon intense light excitation;[26] (3) oxygen incorporation at the Y66G Cα position for the S65G Y66G (renamed GGG for the chromophore residues) variant (Figure 1C);[23] and (4) creation of dehydroalanine moieties through HAL-like dehydration rather than GFP-like oxidation chemistry for the S65A Y66S (renamed ASG) variants (Figure 1E,F).[22] Interestingly, the cyclized GGG and ASG variants under denaturing (but not native) conditions also undergo peptide backbone fragmentation (PBF).[22,23] It is unclear how these GFP cleavage reactions are related to the denaturation-induced cleavage in RFP[8] and the native cleavage reactions that appear critical for chromophore maturation in RFP homologues.[17−20]

(12) (a) Matz, M. V.; Fradkov, A. F.; Labas, Y. A.; Savitsky, A. P.; Zaraisky, A. G.; Markelov, M. L.; Lukyanov, S. A. *Nat. Biotech.* **1999**, *17*, 969−973. (b) Labas, Y. A.; Gurskaya, N. G.; Yanushevich, Y. G.; Fradkov, A. F.; Lukyanov, K. A.; Lukyanov, S. A.; Matz, M. V. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 4256−4261.
(13) Heim, R.; Prasher, D. C.; Tsien, R. Y. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 12501−12504.
(14) Heim, R.; Tsien, R. Y. *Curr. Biol.* **1996**, *6*, 178−182.
(15) (a) Cubitt, A. B.; Heim, R.; Adams, S. R.; Boyd, A. E.; Gross, L. A.; Tsien, R. Y. *Trends Biochem. Sci.* **1995**, *20*, 448−455. (b) Reid, B. G.; Flynn, G. C. *Biochemistry* **1997**, *36*, 6786−6791.
(16) Baedeker, M.; Schulz, G. E. *Structure* **2002**, *10*, 61−67.
(17) Mizuno, H.; Mal, T. K.; Tong, K. I.; Ando, R.; Furuta, T.; Ikura, M.; Miyawaki, A. *Mol. Cell* **2003**, *12*, 1051−1058.
(18) Quillin, M. L.; Anstrom, D. M.; Shu, X.; O'Leary, S.; Kallio, K.; Chudakov, D. M.; Remington, S. J. *Biochemistry* **2005**, *44*, 5774−5787.
(19) Wilmann, P. G.; Petersen, J.; Devenish, R. J.; Prescott, M.; Rossjohn, J. *J. Biol. Chem.* **2005**, *280* (4), 2401−4.
(20) Andresen, M.; Wahl, M. C.; Stiel, A. C.; Grater, F.; Schafer, L. V.; Trowitzsch, S.; Weber, G.; Eggeling, C.; Grubmuller, H.; Hell, S. W.; Jakobs, S. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 13070−13074.
(21) Tsien, R. Y. *Annu. Rev. Biochem.* **1998**, *67*, 509−544.
(22) Barondeau, D. P.; Kassmann, C. J.; Tainer, J. A.; Getzoff, E. D. *Biochemistry* **2005**, *44*, 1960−1970.
(23) Barondeau, D. P.; Putnam, C. D.; Kassmann, C. J.; Tainer, J. A.; Getzoff, E. D. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 12111−12116.
(24) Rosenow, M. A.; Huffman, H. A.; Phail, M. E.; Wachter, R. M. *Biochemistry* **2004**, *43*, 4464−4472.
(25) Rosenow, M. A.; Patel, H. N.; Wachter, R. M. *Biochemistry* **2005**, *44*, 8303−8311.
(26) (a) Bell, A. F.; Stoner-Ma, D.; Wachter, R. M.; Tonge, P. J.; van Thor, J. J. *J. Am. Chem. Soc.* **2003**, *125*, 6919−6926; (b) Gensch, T.; Hellingwerf, K. J.; Johnson, L. N. *Nat. Struct. Biol.* **2002**, *9*, 37−41.

Table 2. Diffraction Data Collection and Refinement Statistics

	ASG heat	GAG	GSG	ASG F64L	GSG F64
resolution (Å)	20.0−2.00	20.0−1.35	20.0−1.20	20.0−1.30	20.0−1.80
last shell (Å)[a]	2.07−2.00	1.40−1.35	1.24−1.20	1.35−1.30	1.86−1.80
observations	53653	194574	472236	572311	214158
unique observations	15582	49224	68670	56899	21833
R_{sym}[b] (%)	7.5 (25.4)	4.3 (29.6)	4.4 (29.2)	5.2 (34.8)	8.6 (34.3)
completeness (%)	97.8 (97.1)	98.2 (98.5)	99.3 (94.7)	99.7 (99.4)	100.0 (99.9)
$I/\sigma I$	16.0 (3.9)	25.3 (3.6)	44.5 (5.1)	47.8 (4.6)	28.8 (6.4)
refinement parameters	7864	19821	19427	20094	8511
R_{work}/R_{free}[c] (%)	20.2/24.1	14.2/21.5	12.9/17.4	13.1/18.7	16.0/23.2
PDB code	2G16	2G3D	2G2S	2G6E	2G5Z

[a] Values in parentheses are the statistics for the highest resolution shell of data. [b] $R_{sym} = \Sigma|I_{hkl} - \langle I\rangle|/\Sigma\langle I\rangle$, where $\langle I\rangle$ is the average individual measurement of I_{hkl}. [c] $R_{work} = (\Sigma||F_{obs}| - |F_{calc}||)/\Sigma|F_{obs}|$, where F_{obs} and F_{calc} are the observed and calculated structure factors, respectively.

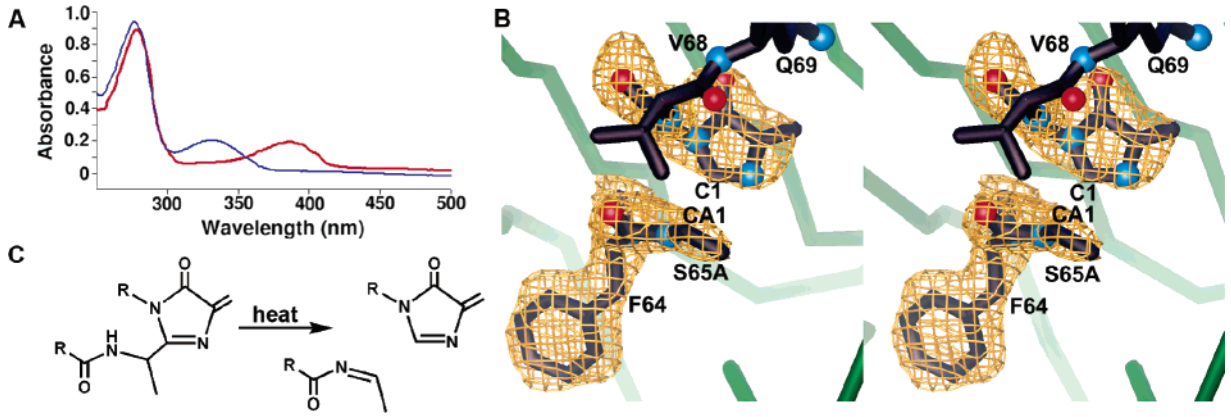


Figure 2. Heat-induced backbone cleavage for the ASG variant. (A) UV−vis absorbance spectra of the ASG variant displayed before (red) and after (blue) heat treatment at 85 °C for 10 min. (B) Stereoimage showing cleaved structure of heat-treated ASG variant with simulated annealing omit electron density map contoured at 3σ (yellow). (C) Reaction for the heat-induced modification of the ASG variant.

Here, we identified conditions that allowed structural characterization of the GFP PBF products and established that this reaction is distinct from the cleavage reactions in RFP homologues. We discovered and structurally characterized GFP variants that surprisingly underwent spontaneous peptide hydrolysis and decarboxylation reactions. We provide peptide fragmentation, peptide hydrolysis, and decarboxylation reaction mechanisms and explain this chemistry in terms of the unique protein environment of GFP. Together, these results have implications for the spontaneous peptide hydrolysis in RFP homologues, provide predictions for how to design mutations to achieve specific backbone and peptide cleavage reactions, and improve our detailed understanding of the driving force and mechanism for GFP chromophore biosynthesis.

Materials and Methods

Mutagenesis and Protein Purification. Using the QuikChange method (Stratagene), we replaced the S65T Tyr66 Gly67 (TYG chromophore) amino acids of GFPsol (F64L S65T F99S M153T V163A) with S65G Y66S (GSG), S65G Y66A (GAG), S65A Y66S (ASG F64L), and F64 S65G Y66S (GSG F64). The resulting plasmids were transformed into BL21-CodonPlus(DE3)-RIL *Escherichia coli* cells (Stratagene), which were grown at 25 °C in 3-L batches. At an optical density of 0.5 at 600 nm, protein expression was induced with 0.2 mM isopropyl-β-D-thiogalactoside. The bacteria cells were pelleted 6−12 h later and frozen in liquid nitrogen until purification. Proteins were purified by modifying a published protocol[27] to incorporate HQ (26 mm × 30 cm) (PerSeptive Biosystems, Inc.) and S-100 (26 mm × 60 cm) (Pharmacia) columns.[28]

Crystal Structure Determination and Refinement. GFP variants were crystallized at 8−12 mg/mL in hanging drops by modifying a published protocol.[6,28] Initial crystal clusters were crushed, serially diluted in a stabilizing mother liquor solution (50 mM Hepes pH 8.0, 50 mM $MgCl_2$, 19% poly(ethylene glycol) 4000), and used as microseeds to grow large single crystals. Diffraction data were collected from crystals that were cryocooled immediately after immersion in the stabilizing solution plus 20% ethylene glycol. Data sets were collected at the Stanford Synchrotron Radiation Laboratory: heat-treated ASG (λ = 0.85 Å), GSG (λ = 0.95369 Å), and GAG (λ = 0.979 Å) on beamline 11-1 and ASG F64L and GSG F64 (λ = 0.97946 Å) on beamline 9-1. Data sets were indexed and reduced in the $P2_12_12_1$ space group with the *hkl* package,[29] and phases were determined by molecular replacement with AMoRe.[30] The search model was a refined 1.0 Å GFPsol structure, determined by molecular replacement from a previous GFP structure.[6] Difference electron density and omit maps were manually fit with the XtalView package[31] and refined in either CNS[32] or Shelx-97[33] using all the diffraction data, except for 5% used for R_{free} calculations.[34] Standard uncertainties were determined by inverting

(27) Deschamps, J. R.; Miller, C. E.; Ward, K. B. *Protein Expr. Purif.* **1995**, *6*, 555−558.
(28) Barondeau, D. P.; Kassmann, C. J.; Tainer, J. A.; Getzoff, E. D. *J. Am. Chem. Soc.* **2002**, *124*, 3522−3524.
(29) Otwinowski, Z.; Minor, W. *Macromol. Cryst. A* **1997**, *276*, 307−326.
(30) Navaza, L. *Acta Crystallogr. A* **1994**, *50*, 157−163.
(31) McRee, D. E. *J. Struct. Biol.* **1999**, *125*, 156−165.
(32) Brunger, A. T.; Adams, P. D.; Clore, G. M.; DeLano, W. L.; Gros, P.; Grosse-Kunstleve, R. W.; Jiang, J.-S.; Kuszewski, J.; Nilges, N.; Pannu, N. S.; Read, R. J.; Rice, L. M.; Simonson, T.; Warren, G. L. *Acta Crystallogr. D* **1998**, *54*, 905−921.
(33) Sheldrick, G. M.; Schneider, T. R. *Methods Enzymol.* **1997**, *277*, 319−343.
(34) Brunger, A. T. *Nature* **1992**, *355*, 472−474.

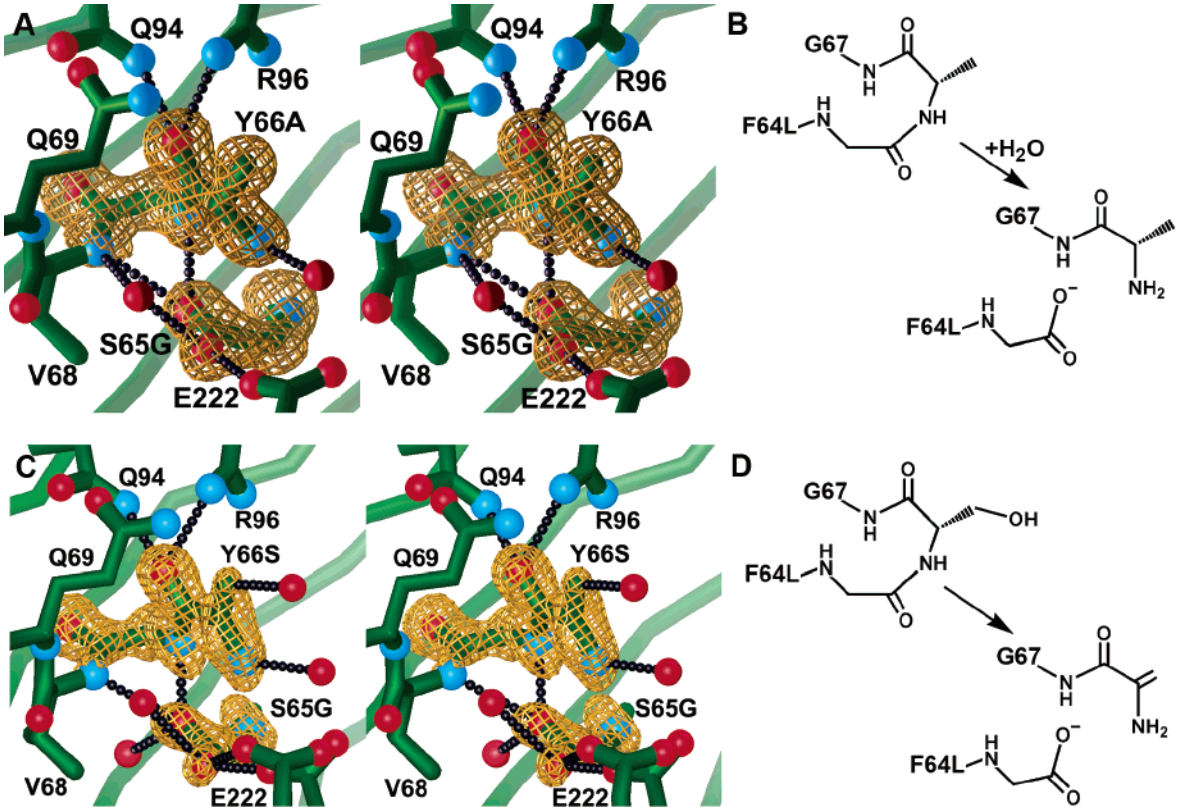


Figure 3. Crystallographic structures of the GAG and GSG variants. (A) Stereopair for the GAG variant (green) displayed with simulated annealing omit electron density map contoured at 3σ (yellow) at 1.35 Å resolution, emphasizing peptide hydrolysis of the S65G Y66A peptide bond, tetrahedral CA2 atom, and hydrogen-bonding interactions (black spheres) for the new resultant N- and C-termini. Water molecules are shown as red spheres. (B) Reaction and posttranslational products for the GAG variant. (C) Stereopair for the GSG variant (green) displayed with simulated annealing omit electron density map contoured at 3σ (yellow) at 1.20 Å resolution, emphasizing peptide hydrolysis, trigonal CA2 atom, and hydrogen bonding interactions (black spheres) for the resultant N- and C-termini. Glu222 exhibits dual conformations. (D) Reaction and posttranslational products for the GSG variant.

the full least squares covariance matrix in Shelx-97.[33] GFP polyclonal antibodies were obtained from Novus.

Results

Peptide Backbone Fragmentation for the ASG GFP Variant. Cyclized GFP variants can exhibit peptide backbone fragmentation (PBF), whereas uncyclized GFP variants do not (Table 1). For cyclized variants, PBF can occur whether the five-membered imidazolone ring is aromatic (Figure 1E) or nonaromatic (Figure 1C,F), but PBF was not found for variants, like wild-type GFP, that contain an aromatic side-chain for central chromophore residue 66 (Table 1). Here, we determined that cleavage produced two peptide fragments (~20 and ~7 kDa), the larger of which bound GFP polyclonal antibodies (a SDS−PAGE (sodium dodecyl sulfate polyacrylamide gel electrophoresis) gel displaying the cleavage products is provided as Supporting Information Figure 1). Backbone fragmentation products are observed for ASG variants[22] under either heating or 4 M urea denaturing conditions.

To understand the products and basis for the PBF reaction, we heat-treated, crystallized, and determined the 2.0 Å structure of the ASG variant (Table 2). We discovered that incubation of the purified ASG protein at 85 °C for 10 min irreversibly shifted the absorbance maxima from 385 to 340 nm (Figure 2A), correlated this spectral shift with PBF, and heat-treated a similar sample prior to crystallization. The retention of visible light absorbance for heat-treated samples suggested that the imidazolone ring is likely intact and still aromatic. Omit and difference electron density maps for the heat-treated ASG structure revealed bond cleavage between the S65A CA1 and the C1 atoms (Figure 2B), which resulted in fragments (calculated as 19 710 and 6943 Da) consistent with its SDS-gel cleavage pattern. The S65A CB1 atom is coplanar with the F64−S65A peptide bond, which suggested formation of a double bond between the S65A N1 and CA1 atoms (Figure 2C). The electron density further revealed an intact and planar five-membered imidazolone ring, as predicted by its visible absorbance signature. Thus, PBF occurs between the CA1 and the C1 atoms, which results in an N1−CA1 double bond and intact imidazolone ring.

Peptide Hydrolysis for the GAG and GSG GFP Variants. We discovered GFP variants that undergo spontaneous peptide hydrolysis under native conditions. The colorless S65G Y66A (renamed GAG) and S65G Y66S (renamed GSG) variants exhibited backbone cleavage under nondenaturing conditions to generate ~20 and ~7 kDa fragments (Supporting Information Figure 1). The omit electron density (Figure 3A) for the 1.35 Å resolution structure of GAG (Table 2) revealed that the backbone is cleaved at the peptide bond between the S65G carbonyl carbon and the Y66A nitrogen atoms (C−N distance 3.44 ± 0.06 Å). The S65G carbonyl carbon moved >2 Å from its position in an intact imidazolone ring and is bonded to two terminal oxygen atoms that form hydrogen bonds to the side-chain of Glu222 and to the Gly67 and Val68 backbone nitrogen atoms. Thus, either the new S65G carboxy terminus or Glu222 is protonated. Interestingly, the Y66A Cα atom exhibited sp^3-

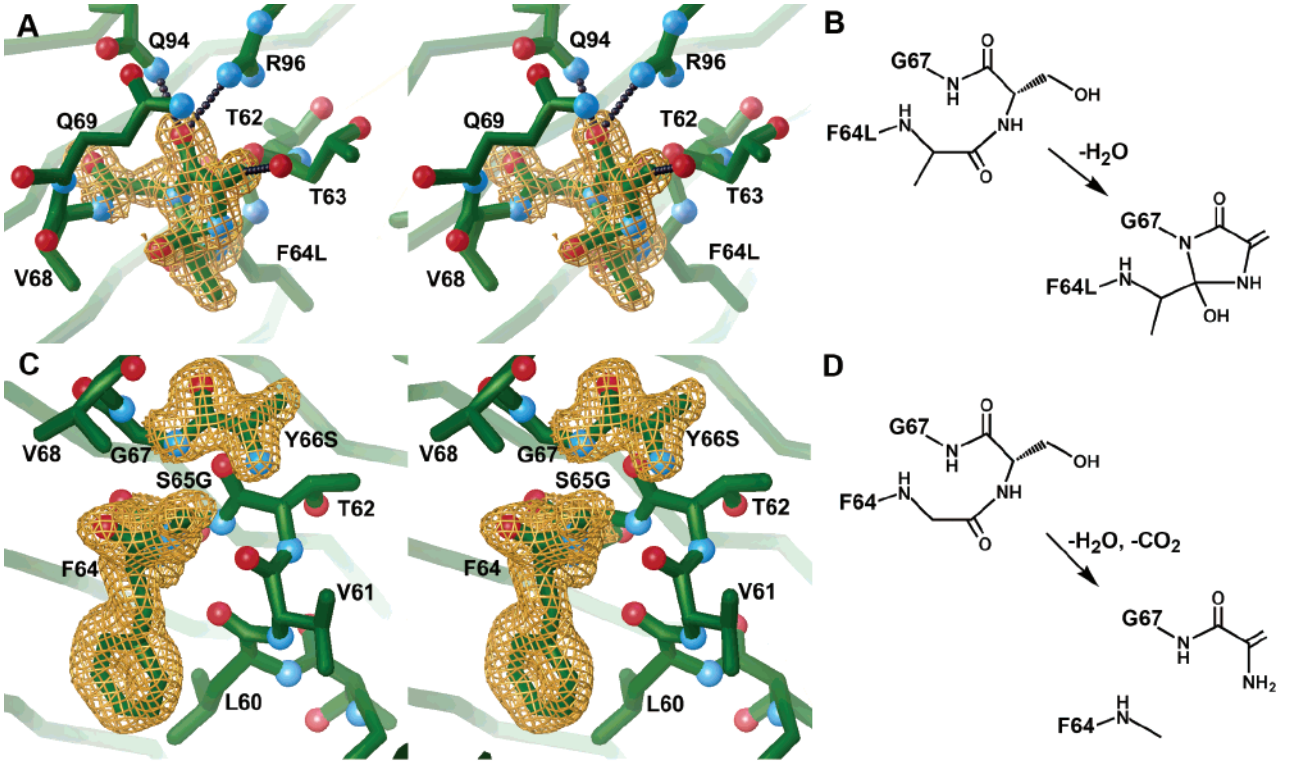


Figure 4. Crystallographic structures of the GFP variants ASG F64L at 1.30 Å resolution and GSG F64 at 1.80 Å resolution. (A) Stereopair for the ASG F64L variant (green) displayed with simulated annealing omit map contoured at 3σ (yellow) and emphasizing the intact five-membered ring with attached hydroxyl group. (B) Reaction and posttranslational products for the ASG F64L variant. (C) Stereopair for the GSG F64 variant (green) displayed with simulated annealing omit map contoured at 3σ (yellow), emphasizing the peptide hydrolysis and decarboxylation. (D) Reaction and posttranslational products for the GSG F64 variant.

hybridization geometry and a Cα−Cβ bond length (1.52 ± 0.05 Å) that are consistent with an unmodified Ala side-chain, indicating that this variant had not undergone a GFP-like oxidation reaction. The peptide hydrolysis reaction for GAG (Figure 3B) and PBF for ASG (Figure 2C) are therefore dissimilar; GAG fragments at a distinct site, undergoes a hydrolysis reaction, and cleaves without prior denaturation or exocyclic double bond formation.

The 1.20 Å resolution structure of the GSG variant (Table 2) revealed peptide hydrolysis between the S65G carbonyl carbon and the Y66S nitrogen atoms (Figure 3C,D), similar to that of GAG. For the resulting GSG N-terminal fragment, the S65G carbonyl carbon atom again was bonded to two terminal oxygen atoms that formed hydrogen bonds to the side-chain of Glu222 and to the Gly67 and Val68 backbone nitrogen atoms. For the C-terminal fragment, the structure revealed that the GSG variant underwent Y66S dehydration: there is no electron density for the Y66S oxygen atom, the bond length for the Y66S Cα−Cβ bond is consistent with a double bond (1.38 ± 0.05 Å), and the Cα atom appears planar and sp^2-hybridized. Furthermore, the GSG variant exhibited a rare carbon−oxygen hydrogen bond (2.64 ± 0.06 Å) between the Y66S Cβ atom and a water molecule, previously postulated to be the product of the Y66S dehydration reaction in ASG variants.[22] The GSG and ASG variants therefore share the ability to dehydrate the Y66S side-chain hydroxyl group. Despite this commonality, the GSG variant underwent ring decomposition and peptide hydrolysis (Figure 3D), whereas the ASG variant retained a stable ring moiety.

Structural Analysis of ASG F64L and GSG F64 Variants. Here, we created and characterized two additional GFP mutants to understand the determinants for ring decomposition and peptide hydrolysis. Two sequence differences between the GSG and the ASG constructs could be responsible for the distinct posttranslational outcomes; GSG contains the S65A to S65G substitution plus the additional solubility mutation F64L. To test if peptide hydrolysis depends on the F64L and/or S65G substitution, we constructed and characterized both the ASG F64L and the GSG F64 variants. (Note that we now specify the identity of residue 64 to distinguish the ASG F64L and GSG F64 variants from ASG and GSG.) The 1.30 Å omit electron density for the ASG F64L structure (Table 2) revealed a cyclized HAL-like moiety (Figure 4A). The Y66S side-chain was dehydrated, the Y66S Cβ atom exhibited an unusual carbon−oxygen hydrogen bond to a water molecule (like the ASG and GSG variants), and a hydroxyl group was bound at the S65A C1 (formerly the carbonyl carbon) atom (Figure 4B). The structure of the ASG F64L and ASG[22] variants overlay well with a minor 0.4 Å backbone shift due to the F64L substitution (Supporting Information Figure 2). Therefore, the F64L substitution does not control peptide hydrolysis.

In contrast, the 1.80 Å structure of the GSG F64 variant (Table 2) revealed S65G−Y66S peptide hydrolysis (Figure 4C), like that of GSG and GAG. The omit electron density indicated that the Y66S hydroxyl group had undergone dehydration to form a dehydroalanine moiety. Unexpectedly, the electron density for GSG F64 also revealed that the S65G carbonyl carbon and terminal oxygen atom(s) are absent, suggesting a

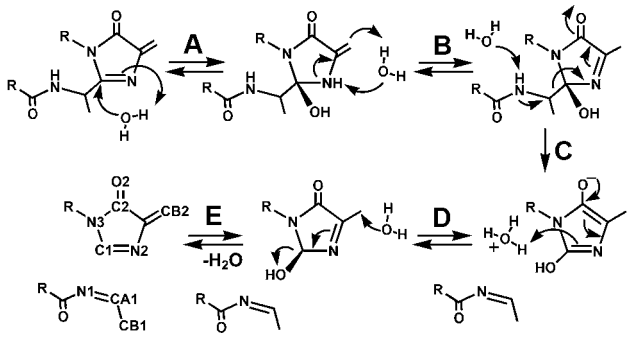
subsequent decarboxylation reaction (Figure 4D). This GFP variant therefore appears to have undergone spontaneous peptide cyclization, HAL-like dehydration, peptide hydrolysis, and decarboxylation posttranslational modifications (see Scheme 2). The GSG F64 and GSG variant structures differ by a backbone shift for residues 64 and 65 of 0.7−1.4 Å (Supplemental Figure 2), which we suggest is primarily a consequence of GSG decarboxylation and the resulting relaxed steric constraints, rather than the F64L substitution. Together, these data indicated that the S65G substitution in the chromophore tripeptide favored peptide hydrolysis, whereas reverting the preceding F64L substitution to wild-type Phe64 favored a subsequent decarboxylation reaction.

Discussion

The biosynthesis of the GFP/RFP fluorophore occurs spontaneously after protein folding without cofactors or accessory proteins, making fluorescent protein fusions tractable in a variety of organisms and fundamentally altering in vivo molecular tagging and cell labeling.[21,35] Importantly, GFP/RFP mutants and homologues have now been identified that exhibit fluorescent emission maxima ranging from blue to far red,[12−14,36] allowing concurrent surveillance of multiple targets. To achieve its red-shifted spectral properties, RFP extends the electronic conjugation of a GFP-like chromophore through a second oxidation reaction, which creates a double bond between the backbone nitrogen and the Cα atoms of the first chromophore residue (Figure 1D).[8,9] The resulting acylimine moiety for RFP is reactive; the backbone is cleaved upon prolonged hydrolysis under denaturing conditions[8] (Figure 5A). Similarly, Quillin et al. report that the kindling fluorescent protein (KFP) undergoes this same acylimine hydrolysis[18] as a chromophore maturation step under nondenaturing conditions (Figure 5A). In a contrasting report, Wilmann et al. report that KFP undergoes native peptide bond cleavage[19] through an unknown mechanism to create an imino substituent that extends the conjugation of the chromophore (Figure 5B). Analogously, the Kaede protein tunes the spectral properties of its chromophore through the N−Cα backbone cleavage and Cα−Cβ double bond formation via a photoinduced β-elimination mechanism[17] (Figure 5C). A common theme for these red-shifted fluorescent proteins is the protein-driven extension of the chromophore conjugation through oxidation and cleavage reactions. An in-depth understanding of the driving force and protein chemistry for these reactions is therefore critical for controlling and further engineering the posttranslational modifications and the design of new fluorescent proteins.

Here, we characterize two additional backbone cleavage reactions for GFP: PBF reactions diagnostic of modified chromophore variants with an exocyclic double bond on the cyclized ring (Figure 5D) and peptide hydrolysis reactions that occur under native conditions (Figure 5E). Denaturation conditions are required for the fragmentation reactions of both RFP[8] (Figure 5A) and GFP variants (Figure 5D), but the mechanisms for these reactions appear unrelated (Scheme 1). The GFP PBF reaction is reminiscent of one proposal for the natural chromophore maturation step in KFP[19] (Figure 5B); both involve cleavage reactions to create N−Cα double bonds. GFP PBF occurs on the C-terminal or chromophore side of the N−Cα bond (Figure 5D) and blue shifts the spectral properties (Figure 2A), whereas KFP extends the electronic conjugation of the chromophore and red shifts the spectral properties by cleaving on the N-terminal or nonchromophore side of the N−Cα bond (Figure 5B). In summary, the GFP/RFP protein scaffold appears to activate peptide cleavage reactions at a variety of positions at or near the chromophore (Figure 5F).

Scheme 1. Proposed Mechanism for Peptide Backbone Fragmentation (PBF) of GFP Variants[a]



[a] Atom labels are shown for the final products. (A) Reversible hydration; (B) N2−CB2 tautomerization; (C) denaturation-induced CA1-C1 backbone cleavage; (D) enol-keto tautomerization and C1 protonation; and (E) water ejection to form final products.

Peptide Backbone Fragmentation (PBF). From the conditions and products that we identified, we propose the following PBF mechanism (Scheme 1): (A) attack on the aromatic 4-methylidene-imidazole-5-one (MIO) by a water nucleophile hydroxylates the C1 atom; (B) tautomerization between the exocyclic CB2 and the five-membered ring N2 nitrogen atoms; (C) proton abstraction from the backbone nitrogen N1 atom resulting in N1−CA1 double bond formation, CA1-C1 bond cleavage, and generation of an enolate moiety at C2; (D) enol-keto tautomerization leading to protonation of the S65A C1 atom; and (E) CB2 deprotonation and water ejection thereby creating the aromatic species observed in our crystallographic structure and implied by our spectroscopic data.

This mechanism for PBF is consistent with the results we present here and those published previously.[22−24] The reversible dehydration at C1 in Scheme 1A matches that proposed to account for mixed populations of nonaromatic and aromatic (dehydrated) imidazolone ring species for the ASG GFP variants (Figure 1E,F).[22] The resultant hydroxyl moiety at C1 was first observed in the refined crystal structure of the cyclized GGG variant, which also undergoes PBF.[23] The water-based N1 deprotonation (Scheme 1C) likely requires prior protein denaturation, as cleavage products are only observed under denaturation conditions (Supporting Information Figure 1) and the ASG structures[22] contain no appropriately placed base under native conditions. The double bond arrangement in the intermediate preceding dehydration (Scheme 1E) resembles that proposed by Roesenow et al.[24] for a cyclized and oxidized, but not dehydrated, intermediate in GFP chromophore biosynthesis. Our PBF mechanism requires an exocyclic Cα−Cβ double bond at residue 66 (resulting from GFP-like oxidation or HAL-like dehydration reactions) for the N2−CB2 tautomerization step (Scheme 1B) and subsequent cleavage. In wild-type GFP, we predict that the energetic cost of this tautomerization, which

(35) (a) Ehrhardt, D. *Curr. Opin. Plant Biol.* **2003**, *6*, 622−628. (b) Miyawaki, A. *Curr. Opin. Neurobiol.* **2003**, *13*, 591−596.
(36) Shaner, N. C.; Campbell, R. E.; Steinbach, P. A.; Giepmans, B. N.; Palmer, A. E.; Tsien, R. Y. *Nat. Biotechnol.* **2004**, *22*, 1567−1572.

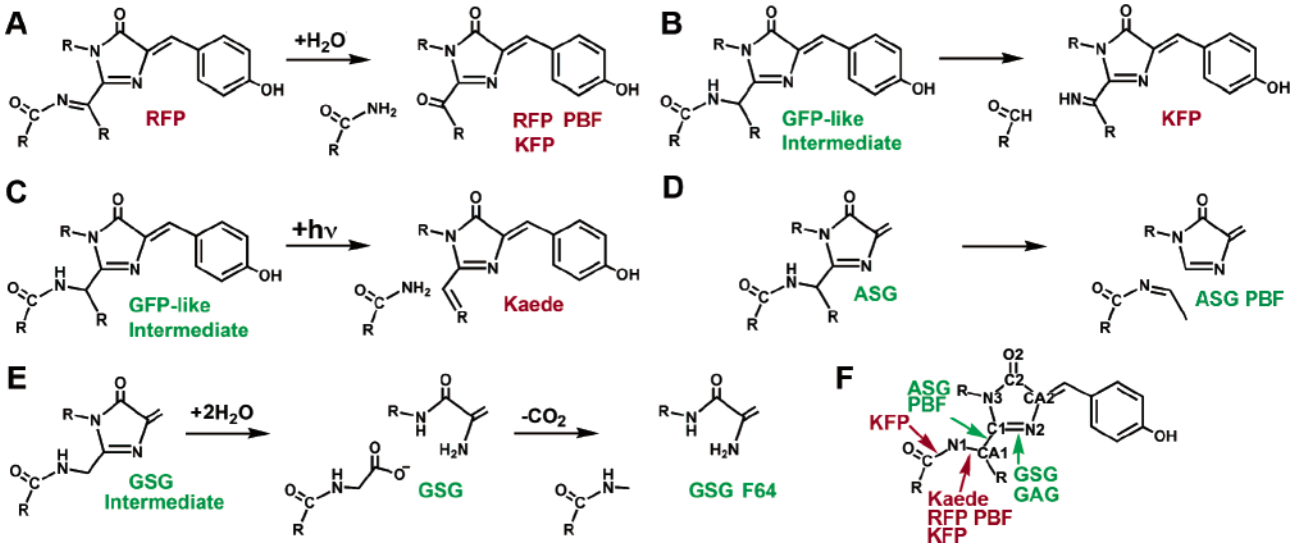
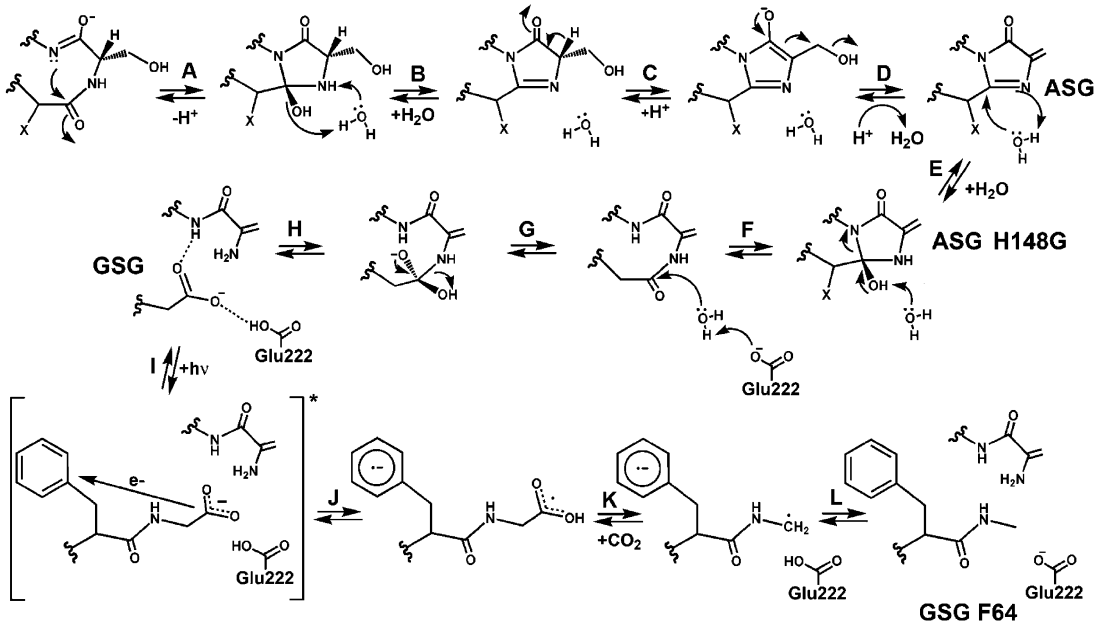


Figure 5. Peptide backbone cleavage reactions for GFP and RFP homologues. (A) PBF of RFP[8] and proposal[18] for spontaneous KFP maturation; (B) alternative proposal[19] for KFP maturation; (C) light-induced backbone cleavage for Kaede protein;[17] (D) PBF for ASG GFP; (E) peptide hydrolysis for GSG GFP and decarboxylation for GSG F64; and (F) summary of cleavage positions for GFP and its RFP homologues. Labels in green are species that are conjugated similarly to GFP, whereas those in red are further conjugated, as in RFP.

Scheme 2. Reaction Pathway for GFP Y66S Posttranslational Modifications[a]



[a] Reactions are (A) peptide cyclization; (B) dehydration of main-chain hydroxyl group; (C) enolate formation; (D) dehydration of side-chain hydroxyl group to form the ASG (PDB code 1YJF) product; (E) hydration to form the ASG H148G (PDB code 1YJ2) product; (F) ring decomposition; (G) water attack to make a tetrahedral intermediate; (H) peptide hydrolysis to form the GSG F64L product; (I) light excitation of Phe64 side-chain ring; (J) diradical formation via electron transfer; (K) main-chain decarboxylation; and (L) generation of final GSG F64 products. X is a methyl group for ASG variants and a hydrogen atom for GSG variants.

would remove the electronic conjugation between the imidazolone and the Y66 hydroxyphenyl rings, precludes PBF. Thus, our proposed mechanism (Scheme 1) successfully predicts why only specific GFP variants exhibit PBF (Table 1).

Backbone Cyclization. Our results indicate that the chromophore residues rapidly interconvert between energetically similar precursor and cyclized states. Structural data for the GSG variants (Figures 3C and 4C) revealed posttranslational products with dehydroalanine moieties that were created through Y66S dehydration. We suggest that these GSG variants cyclize (Scheme 2A) and undergo main-chain dehydration (Scheme 2B) leading to an enolate intermediate (Scheme 2C) before side-chain dehydration (Scheme 2D). This scheme is consistent with the observed backbone cyclization requirement for Y66S dehydration[22] and recent structural evidence for an enolate intermediate in GFP chromophore biosynthesis.[37] The GSG variant can then undergo a hydration reaction to create a species similar to the posttranslational product of the ASG H148G[22] variant (Scheme 2E). Deprotonation of the C1 hydroxyl group leads to ring decomposition (Scheme 2F). Next, we propose

(37) Barondeau, D. P.; Tainer, J. A.; Getzoff, E. D. *J. Am. Chem. Soc.* **2006**, *128*, 3166−3168.

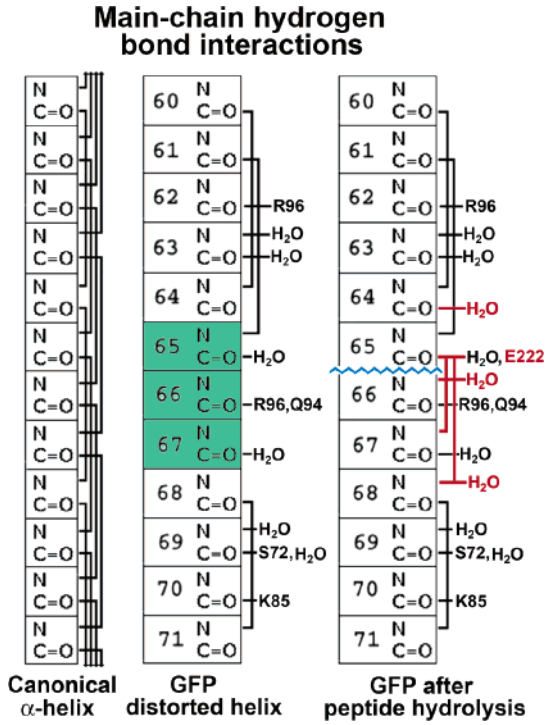


Figure 6. Role of helical main-chain hydrogen bonds in GFP chromophore cyclization. Main chain hydrogen bond interactions for canonical α-helix (left), GFP central helix in precyclized, prehydrolyzed, and mature chromophore states (center), and GFP central helix after peptide hydrolysis for the GSG and GAG variants (right). The residues that form the chromophore are displayed in green, the peptide hydrolysis position is blue, and the new main-chain hydrogen bond interactions are in red (two main-chain, three water, and Glu222 interactions).

that the uncyclized GSG variant either reforms a cyclized moiety (Scheme 2F) or undergoes peptide hydrolysis (Scheme 2G,H). In other words, the GSG variants likely continue to sample both precursor and cyclized conformations until they are irreversibly trapped via peptide hydrolysis. Previous results on the GGG variant, which is cyclized under aerobic but not anaerobic conditions, also support energetically similar precursor and cyclized species.[23] In the GGG variant, oxygen incorporation at the Y66G Cα position (Figure 1C) is required to stabilize the cyclized over the precursor state. Together, these results imply that the GFP/RFP protein scaffold lowers barriers to peptide cyclization and allows population of a cyclized intermediate that can undergo further essentially irreversible chemistry, such as oxidation or backbone cleavage reactions that generate mature fluorophores.

We suggest that the GFP/RFP scaffold lowers these cyclization barriers to promote ring formation by (1) favoring nucleophilic attack by close proximity alignment of the Gly67 amide lone pair with the π^*-orbital of the residue 65 carbonyl and (2) eliminating inhibitory main-chain hydrogen bonds in the precursor state.[22,23] Specifically, the GFP protein scaffold enforces a bend in the central helix that is focused at Gly67 and results in a lack of main-chain hydrogen bonds for the chromophore-forming residues in both precursor and mature chromophore states (Figure 6).[22] The presence of such main-chain hydrogen bonds would stabilize the precursor state and inhibit peptide cyclization. Moreover, mutation-induced structural rearrangements that satisfy these hydrogen bonds, normally desirable for main-chain polar atoms, inhibit backbone cyclization.[22] The protein architecture therefore lowers enthalpic barriers and utilizes a negative design principle (disfavoring alternate stable conformations) to achieve its peptide cyclization function. We propose that some of these same features that favor peptide backbone cyclization also lead to peptide bond hydrolysis in S65G variants.

Peptide Bond Hydrolysis. In GSG and GAG variants, we propose that the absence of main-chain hydrogen bonds for the 65−66 peptide bond (Figure 6) promotes hydrolysis by facilitating the conversion of the S65G carbonyl carbon from an sp^2- to sp^3-hybridized moiety. In the precyclized GGG structure,[23] which is a reasonable model for the prehydrolyzed GSG and GAG variants, a water molecule forms hydrogen bonds to both the S65G carbonyl oxygen and the Glu222 carbonyl oxygen atoms. We therefore propose that Glu222 deprotonates a water molecule, which then attacks the S65G carbonyl carbon atom (Scheme 2G). In the resulting tetrahedral intermediate, one of the oxygen atoms could be stabilized through hydrogen bonds with the backbone amides of Gly67 and Val68, similar to those observed for a terminal oxygen atom in hydrolyzed GAG and GSG variants (Figure 3A,C). Such a scenario is reminiscent of features thought to stabilize the transition state and accelerate peptide hydrolysis in enzymatic systems.[38] Ironically, the same lack of hydrogen bond interactions for residue 65−67 and the close nitrogen−carbon distance that favor peptide cyclization may also favor peptide hydrolysis. In peptide cyclization (Scheme 2A), the Gly67 amide nitrogen attacks the re face of the S65G carbonyl carbon to generate the cyclized, tetrahedral intermediate. Water attack (Scheme 2G) on the opposite face of this carbonyl also generates a tetrahedral intermediate that can be stabilized via Gly67 hydrogen bond interactions for peptide cleavage (Scheme 2H).

We suggest that side-chain steric interactions for residue 65 determine which GFP variants undergo peptide hydrolysis (Table 1); side-chains larger than S65G inhibit either the conformations necessary for hydrolysis or the approach of water molecules to the carbonyl carbon atom. The N-terminal fragments for both the GAG and the GSG variants exhibit a S65G conformation more similar to precyclized GGG than to cyclized ASG (Supporting Information Figure 3). Modeling a S65A-like methyl group onto the cleaved GSG structure indicated that ASG variants likely cannot adopt such a conformation due to steric interaction with Leu220. This conformational restriction is essentially a negative design feature as peptide hydrolysis for the GAG and GSG variants creates a remarkable six additional main-chain hydrogen bonds (Figure 6): three hydrogen bonds with water molecules and interactions between the new carboxy terminus with Glu222 and the backbone nitrogen atoms of Gly67 and Val68 (Figure 3A,C). Thus, the GFP scaffold controls hydrogen bonding in the central helix to favor cyclization leading to fluorophore formation, while simultaneously disfavoring additional reactions such as peptide hydrolysis.

Decarboxylation. The decarboxylation of the GSG F64 variant suggests that the GFP scaffold can favor radical formation and one-electron chemistry. Comparison of the GSG F64 and GSG (which carries the F64L substitution) variants

(38) (a) Valina, A. L.; Mazumder-Shivakumar, D.; Bruice, T. C. *Biochem.* **2004**, *43*, 15657−15672. (b) Pelmenschikov, V.; Blomberg, M. R.; Siegbahn, P. E. *J. Biol. Inorg. Chem.* **2002**, *7*, 284−298. (c) Stamper, C.; Bennett, B.; Edwards, T.; Holz, R. C.; Ringe, D.; Petsko, G. *Biochemistry* **2001**, *40*, 7035−7046.

Table 3. Comparison of GFP Decarboxylation Reactions

	GSG F64	wild-type GFP
light requirement	ambient?	high intensity
excitation species	Phe64	chromophore
decarboxylation species	S65G	Glu222
electron transfer	through bond	through space
reference	this paper	26

indicated that the decarboxylation reaction correlates with Phe64. This Phe64 dependence could be due to steric interactions (the F64L substitution in ASG variants result in minor structural rearrangements, Supporting Information Figure 2) or, more likely, the ability of the aromatic ring of Phe64 to mediate photochemistry. On the basis of similar chromophore-mediated decarboxylation of Glu222,[26] plus the precedent for decarboxylation of glycine via radical mechanisms[39] and of Phe-Gly dipeptides by photoinduced radical processes,[40] we propose that the GSG F64 decarboxylation reaction is a light-driven, radical process.

In our mechanism, the excited state of Phe64 (Scheme 2I) generates a radical anion by oxidizing the carboxy terminus of S65G (Scheme 2J). Homolytic cleavage of the S65G carbon−carbon bond releases the unstable carboxy radical as carbon dioxide and generates a new carbon based radical (Scheme 2K). Electron transfer from the Phe64 radical anion and protonation generates the final product revealed by our GSG F64 structural data (Scheme 2L). Differences in electron transfer requirements may explain the distinct excitation requirements for Glu222 and S65G decarboxylation (Table 3). In Glu222 decarboxylation,[26] intense light excitation is required to induce the ∼4 Å electron transfer from Glu222 through space to the chromophore. In contrast, ambient light is evidently sufficient to induce the ∼7 Å electron transfer through the peptide backbone from S65G to the Phe64 side-chain. Although the S65G electron transfer is over a longer distance, it occurs through covalent bonds, which has been shown to increase electron-transfer rates by 20−50-fold.[41] This proposed S65G radical-based chemistry occurs in the absence of the GFP chromophore, implicating the protein environment in facilitating radical-based chemistry. We suggest that nearby Arg96 provides electrostatic stabilization for these proposed chromophore[26] or Phe64-based radical anions (∼5 and ∼9 Å distance between Arg96 and the ring systems, respectively). Similarly, Goodin et al. demonstrated that the buried negative charge of Asp235 plays a critical role in stabilizing the formation of the Trp191 radical cation in cytochrome *c* peroxidase.[42] These results suggest a common theme in which the GFP architecture facilitates one-electron chemistry that is likely important for electron transfer and activation of molecular oxygen.

Posttranslational Modification Chemistry. Remarkably, the GFP/RFP protein architecture promotes not only ring formation and fluorophore biosynthesis but also covalent bond cleavage at four consecutive positions along the polypeptide backbone (Figure 5F). The GFP/RFP scaffold focuses a bend in the central chromophore-containing helix at the location of these cyclization and cleavage reactions that removes inhibitory main-chain hydrogen bonds (Figure 6).[22,23] In addition to this helical bend, this family of fluorescent proteins conserves two buried charged residues, Arg96 and Glu222. Arg96 helps stabilize the helical bend conformation and provides electrostatic interactions that are important for ring formation.[22,23,43,44] Glu222 abstracts the Tyr66 Cα proton for enolate intermediate formation.[43] The unconstrained main-chain atoms and the buried charges of Arg96 and Glu222 provide the primary driving force for fluorophore biosynthesis but also can result in surprising posttranslational reactivity, such as spontaneous peptide hydrolysis and decarboxylation reactions in GFP variants (Scheme 2). We propose that buried Glu222 and the lack of main-chain hydrogen bonds drive the peptide hydrolysis reaction, whereas Arg96 positive electrostatic interactions favor radical anion formation and subsequent decarboxylation. Moreover, the decarboxylation reaction occurs in the absence of a GFP chromophore, thereby implicating the protein environment in initiating radical-based one-electron chemistry, which may also be important for native GFP/RFP oxidation reactions and chromophore biosynthesis. The ability of the GFP/RFP scaffold to activate oxygen would be consistent with variants that undergo additional oxidative chemistry, such as oxygen incorporation[23] and oxidative cross-link reactions.[25] Together, these results provide the groundwork for the design of proteins with novel catalytic or reporter properties and reveal details for how the GFP/RFP protein environment both favors chromophore biosynthesis and utilizes negative design principles to inhibit alternate posttranslational modification chemistry.

Acknowledgment. We thank C. D. Putnam, V. A. Roberts, T. I. Wood, and J. L. Huffman for scientific discussions and O. Sundheim for collecting the GSG F64 and ASG F64L datasets. This work was supported by the La Jolla Interfaces in Sciences (D.P.B.), NIH GM19290 Postdoctoral Fellowships (D.P.B.), and NIH Grant RO1 GM37684 (E.D.G.). Portions of this research were carried out at the Stanford Synchrotron Radiation Laboratory, a national user facility operated by Stanford University on behalf of the U.S. Department of Energy, Office of Basic Energy Sciences. The SSRL Structural Molecular Biology Program is supported by the Department of Energy, Office of Biological and Environmental Research, and by the National Institutes of Health, National Center for Research Resources, Biomedical Technology Program, and the National Institute of General Medical Sciences.

Supporting Information Available: SDS-gel that shows the denaturation and native backbone cleavage products, a structural overlay of precursor and mature chromophore states with the GAG F64L and GSG F64L variants, and a structural overlay of ASG and GSG variants with and without the F64L substitution. This material is available free of charge via the Internet at http://pubs.acs.org.

JA056635L

(39) Bonifacic, M.; Stefanic, I.; Hug, G. L.; Armstrong, D. A.; Klaus-Dieter, A. *J. Am. Chem. Soc.* **1998**, *120*, 9930−9940.
(40) Hill, R. R.; Coyle, J. D.; Birch, D.; Dawe, E.; Jeffs, G. E.; Randall, D.; Stec, I.; Stevenson, T. M. *J. Am. Chem. Soc.* **1991**, *113*, 1805−1817.
(41) Wenger, O. S.; Leigh, B. S.; Villahermosa, R. M.; Gray, H. B.; Winkler, J. R. *Science* **2005**, *307*, 99−102.
(42) Fitzgerald, M. M.; Trester, M. L.; Jensen, G. M.; McRee, D. E.; Goodin, D. B. *Protein Sci.* **1995**, *4*, 1844−1850.
(43) Sniegowski, J. A.; Lappe, J. W.; Patel, H. N.; Huffman, H. A.; Wachter, R. M. *J. Biol. Chem.* **2005**, *280*, 26248−26255.
(44) Wood, T. I.; Barondeau, D. P.; Hitomi, C.; Kassmann, C. J.; Tainer, J. A.; Getzoff, E. D. *Biochemistry* **2005**, *44*, 16211−16220.