Chromogenic Cross-Link Formation in Green Fluorescent Protein

REBEKKA M. WACHTER*
Department of Chemistry and Biochemistry, Arizona State University, Tempe, Arizona 85287

Received May 23, 2006

ABSTRACT
Recent advances in our understanding of the mechanism of chromophore formation in green fluorescent protein (GFP) are presented. GFP is the best-studied member of the family of GFP-like proteins, proteins that exhibit bright coloration spanning most of the visible spectrum. GFPs undergo a post-translational self-modification process that yields an intrinsic fluorophore constructed from an internal main-chain cross-link that is susceptible to air oxidation. A combination of protein X-ray crystallographic and kinetic experiments has led to the development of a mechanistic model that entails conformational pre-organization, electrophilic and base catalysis, and production of hydrogen peroxide upon protein oxidation. The process is concluded by a slow proton abstraction step from a tyrosine-derived carbon acid.

Introduction

The diversity of known protein prosthetic groups generated by post-translational alteration of genetically encoded amino acids has expanded tremendously during the past decade.[1] Several instances are known in which covalent self-processing steps follow protein folding spontaneously. Such autocatalytic modifications may be considered an intrinsic property of the particular amino acid sequence since complete protein maturation requires only a single gene, in contrast to the biosyntheses of protein-derived motifs that necessitate external enzymes or cofactors.

The post-translational oxidation of tyrosine residues plays an essential role in the biogenesis of several redox cofactors. This theme may be illustrated by the formation of quinone-type factors, essentially built-in prosthetic groups that may be covalently cross-linked to nearby residues.[1] Examples are the formation of topaquinone by the oxidative modification of a tyrosine in the copper amine oxidases, oxidative cross-linking of a tyrosine phenol to a cysteine residue in galactose oxidase, and the Met-Tyr-Trp cross-link recently described in catalase-peroxidase (Figure 1).[2] In these systems, protein-bound metal chemistry aids in the spontaneous formation of the cofactor since coordination of O_2 to the metal ion facilitates rapid electron transfer. However, a number of biologically important oxidation reactions utilize O_2 as the final electron acceptor without the aid of redox-active metal ions.[3] In spite of the fundamental importance of oxygen activation by the protein matrix, our mechanistic understanding of this process remains limited.

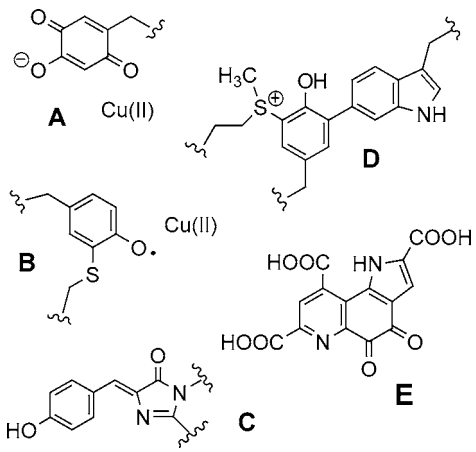


FIGURE 1. Oxidative tyrosine modifications in proteins. (A) Topaquinone redox cofactor in copper amine oxidases. (B) Thioether cross-link in galactose oxidase. (C) Chromogenic cross-link in green fluorescent protein. (D) Met-Tyr-Trp cross-link in catalase-peroxidase. (E) Pyrroloquinoline quinone redox cofactor in bacterial dehydrogenases.[1]

A distinct, metal-independent tyrosine oxidation process is exemplified by a group of colored proteins, whose founding member is green fluorescent protein (GFP).[4,5,6] A tyrosine-bearing tripeptide located in the center of the eleven-stranded β-barrel is spontaneously modified to yield a chromogenic entity that is responsible for the colorful appearance of these proteins (Figure 2). Thus, the brightly fluorescing chromophores that are the trademark of GFP-like proteins are the result of an oxidative tyrosine modification mediated by a peptide cyclization reaction rather than metal chemistry. However, GFPs differ from other tyrosine oxidations in that the phenolic ring atoms are not targeted directly. Instead, all mature chromophores derived from GFP-like proteins are oxidized along the tyrosine $C_\alpha-C_\beta$ bond, although the extent of their π-system may vary (Figure 3).[7]

GFP was discovered in 1961 as part of a bioluminescent system found in photocytes of the jellyfish *Aequorea victoria*, a Pacific Northwest medusa belonging to the class Hydrozoa.[8] The remarkable potential of GFP as a biotechnological tool was not exposed until several decades later, with the demonstration that the jellyfish protein acquires green fluorescence when expressed in unrelated organisms.[9] Today, most bioscience researchers are familiar with GFP as a genetically encodable fluorophore that is well-suited for multicolor reporting and subcellular protein localization. More recently, the precise visualization of dynamic cellular processes by novel GFP-based biosensors has become feasible.[10−12]

Although for many years, GFP remained the only protein of its kind, a series of fluorescent and colored

Rebekka M. Wachter received her B.A. from the University of California, Santa Cruz and carried out her doctoral work at the University of Oregon. After receiving her Ph.D. in Chemistry in 1996, she spent 5 years as a post-doctoral research associate in the Institute of Molecular Biology at the University of Oregon. In 2001, she accepted a position as Assistant Professor of Chemistry and Biochemistry at Arizona State University. Her research interests concern the mechanism of protein self-processing reactions, macromolecular X-ray crystallography, and photoprotein structure and function.

* Corresponding author. E-mail: RWachter@asu.edu.

10.1021/ar040086r CCC: $37.00
Published on Web 11/14/2006

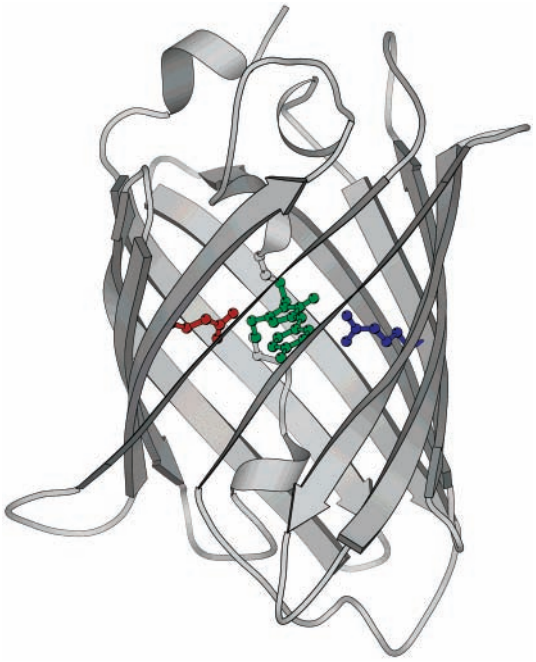


FIGURE 2. Schematic of the GFP fold (pdb ID 1EMA).[5] The chromophore (green), Arg96 (blue), and Glu222 (red) are shown as ball-and-stick models. Two β-strands are shown as thin ribbons to better visualize the protein's interior.

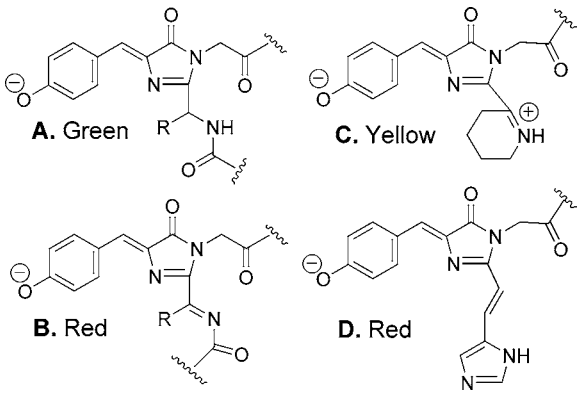


FIGURE 3. Chemical diversity of chromophores generated in GFP-like proteins. (A) GFP, (B) DsRed, (C) Zoanthus yellow, and (D) Kaede after photolysis.[7]

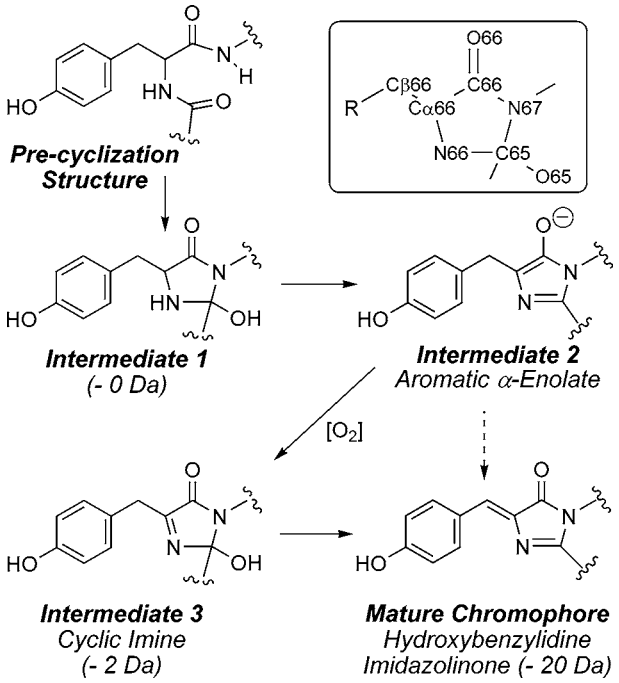
proteins with distant homology to GFP was cloned in 1999 from reef building corals of the class Anthozoa.[13] Because of their broad color variety and amazing photophysical properties, they have opened the door to a diverse set of novel applications in live cell imaging.[14] In spite of this success, the biological function of GFP-like proteins in their native organisms continues to be debated. Mechanical stimulation of the jellyfish is known to trigger emission of a green flash due to a luciferase/GFP bioluminescent system, whereas in the nonbioluminescent Anthozoa species, GFP-like proteins are the origin of the vivid colors observed on the organism's surface.[15] Proposed functions for these proteins in reef-building corals include photoprotection of endosymbiotic dinoflagellates in bright sun light.[15,16]

During the past few years, important advances have been made toward the development of an atomistic model of the GFP self-processing mechanism. Building upon several earlier investigations,[17−19] recent kinetic and X-ray crystallographic studies arising from our and other research groups have provided valuable insight into the biochemistry of fluorophore maturation.[20,21,22] In this Account, some of the structural and functional features will be described that are unique to GFP-like proteins and contribute most significantly toward self-modification. This work is focused on the current state of knowledge of chromophore biogenesis in GFP, the best-studied member and a representative of the cyan-green color class of GFP-like proteins.[14,23,24] Here, the green stage embodies the end point of post-translational modification, whereas in the red and yellow homologues, additional processing steps provide more extensive π-orbital conjugation concomitant with broader color diversity (Figure 3).[7] However, in all color classes, the formation of an intermediate similar to the GFP green form appears to be a prerequisite for the ensuing chemistry associated with red-shifted optical properties. An improved understanding of the process operational in GFP is hoped to provide more general insight into the biochemistry of all members of this fascinating family of colored proteins.

Outline of GFP Maturation

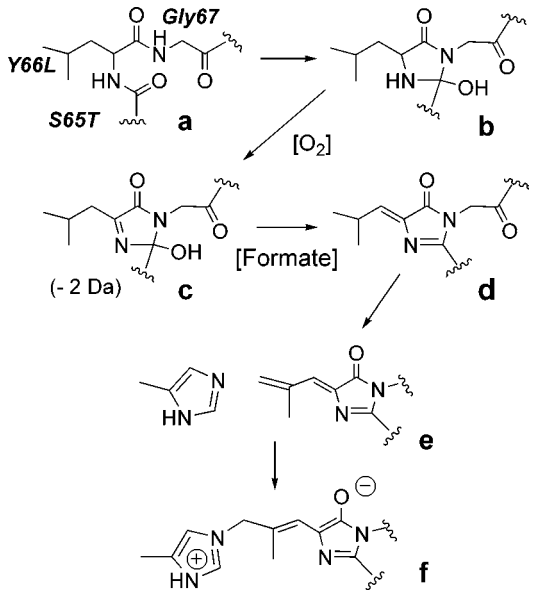
The GFP fluorophore is formed only under conditions permissive of protein folding (i.e., the expressed polypeptide must be able to attain its native three-dimensional structure to partition through the ensuing covalent modification steps). Although fluorophore formation is triggered by a main-chain cyclization reaction, the rate-limiting process in the acquisition of color depends on the availability of molecular oxygen.[17,18,25] On the basis of this observation, a multistep mechanism for chromophore biosynthesis in GFP was proposed some years ago,[17,18] in which nucleophilic attack of the amide nitrogen of Gly67 (N67) onto the carbonyl carbon of Ser65 (C65) leads to a heterocyclic intermediate composed of the backbone atoms of Ser65, Tyr66, and Gly67 (Scheme 1, intermediate **1**; inset: atom labels used throughout the text). Intermediate **1** is thought to undergo enolization at Cα66 to augment its reactivity toward molecular oxygen.[21] The elimination of water from the heterocycle has been suggested to generate a stable pre-oxidation species,[17,18] although hydration−dehydration equilibria have also been described.[26,27] On the basis of recent experimental evidence,[28] the ensuing intermediate **2** may comprise the aromatic imidazolinone ring (Scheme 1), from which chromophore formation is thought to continue slowly by way of air oxidation.[4] Recent evidence from our laboratory suggests the formation of an oxidized cyclic imine form, intermediate **3**, that is generated upon production of H_2O_2 and subsequently matures to the final green stage via the net loss of a water molecule.[21,25] Full π-orbital conjugation of the phenolic group with the imidazolinone ring yields a cis-coplanar system that is highly fluorescent when

Scheme 1. Outline of GFP Maturation Mechanism[a]



[a] Indicated masses are those observed by trypsinolysis/MALDI according to ref 25. Inset: Atom labels used throughout the text.

Scheme 2. Self-Modification Mechanism in Compromised Variant Y66L, Illustrating Formation of Y66L Analogue of Mature GFP Chromophore (d) and Cross-Link with His148 (f)



embedded in the protein's interior (Scheme 1). Early mechanistic data were derived from in vivo studies, in which bacterial cultures expressing wild-type GFP were first grown under anaerobic conditions.[17,18] The introduction of air resulted in a mass loss of roughly 1 Da, as determined by electrospray ionization mass spectrometry. These results were interpreted in terms of dehydrogenation of the Cα66−β66 bond to yield mature protein directly from intermediate **2** (Scheme 1, dashed arrow),[17,18] although the results are also consistent with a mechanism that partitions through intermediate **3** (solid arrows).[25] Time constants for fluorescence acquisition in live cells were reported to be 120 min for wild-type GFP and 27 min for GFP-S65T, a particularly bright and fast-maturing variant that was discovered serendipitously some years ago.[17,29]

A popular technique to study GFP maturation under well-defined conditions is based on triggering protein folding via rapid dilution of urea-solubilized inclusion bodies.[19] In our laboratory, we have made extensive use of this approach to monitor de novo chromophore biosynthesis. We have chosen a highly soluble, monomeric, and rapidly maturing variant that carries the S65T substitution.[25] This variant has been termed GFP-trix or alternatively mGFPsol, following the nomenclature developed by Tsien and co-workers.[30,31] We have measured a time constant of about 50 min for in vitro maturation of mGFPsol at 30 °C.[25] In line with this result, a re-oxidation time constant of 56 min was recently reported after chemical reduction of mature GFPsol.[28] To our knowledge, the most rapid re-oxidation process reported to-date is that for a yellow GFP variant termed Venus, where a time constant of 2.1 min was measured at 37 °C.[32]

Mechanistic Role of Tyr66

Cyclization and Oxidation in the Absence of an Aromatic Residue. Among the three residues from which the chromophore is constructed, Tyr66 and Gly67 are completely conserved in all known GFP-like proteins. Although the bulk of the chromophore π-system is derived from the tyrosine phenolic group, it has long been known that GFP is able to synthesize an intact chromophore upon substitution of Tyr66 with other aromatic residues.[17] To investigate whether the aromatic character of residue 66 is essential for GFP cyclization or oxidation, we replaced Tyr66 with a leucine residue (Scheme 2).[21] The X-ray structure of the colorless Y66L variant gave clear evidence toward a hydroxylated heterocycle, consistent with a trapped tetrahedral intermediate in the center of the β-barrel (Scheme 2, species **c**). The trigonal-planar geometry of Cα66 suggested a cyclic imine form of the protein that bears a double bond between N66 and Cα66 due to air oxidation. Additional support in favor of an oxidized protein species was provided by mass spectrometric data indicating a mass loss of 2 Da in colorless Y66L.[20] More recent experimental evidence suggests that species **c** may be related to an intermediate formed during maturation of intact GFPs (Scheme 1, intermediate **3**).[25] Collectively, the structural and biochemical findings imply that an aromatic group in position 66 is not essential for GFP cyclization or oxidation.

Heterocyclic Hydration−Dehydration Equilibria. In the colorless Y66L variant, the hydration adduct of the oxidized heterocycle appears to be the energetically favored form.[21] Substantial support for a chemical equilibrium between the hydrated and the dehydrated forms of the five-membered ring was subsequently provided by X-ray structures of partial hydration adducts (about 10 and 85%) in GFP variants bearing the Y66S substitution.[27] In these mutants, a covalent modification was described that

is reminiscent of the methylidene imidazolinone cofactor found in the ammonia lyases.[33]

We have further characterized the hydration equilibria of the GFP imidazolinone ring by analyzing X-ray structures of Y66L species that have acquired a yellow tint upon prolonged incubation with a general base.[26] Compounds such as fluoride, formate, and acetate are known to be small enough to partition into the GFP active-site cavity.[34] At high concentrations, they appear to catalyze deprotonation of Cβ66, which is acidified by the electron-withdrawing properties of the heterocycle. On the basis of its optical characteristics, the resulting form of the protein (Scheme 2, species **d**) is thought to embody the Y66L analog of the mature GFP chromophore. This form is susceptible to further oxidation, and eventually, a highly conjugated, flat entity with yellow appearance ($\lambda_{max} = 412$ nm) is generated (Scheme 2, species **e**).[26] The electrophilic properties of this group render it a target for conjugate addition reactions by nearby nucleophiles, such as the imidazole group of His148, which is observed to be cross-linked to the modified Leu66 side chain in one of the X-ray structures (Scheme 2, species **f**).[26] In combination, the crystallographic data are consistent with a hydration−dehydration equilibrium that is modulated by the extent of π-orbital conjugation with the side chain and the associated ring aromaticity. In colorless Y66L, which bears an aliphatic adduct (Scheme 2, species **c**), the heterocycle is completely hydrated, whereas in the yellow forms, extensive delocalization yields a primarily dehydrated aromatic heterocycle (Scheme 2, species **d**−**f**). In intact GFPs, the phenolic group of Tyr66 may play a role in shifting the equilibrium toward the complete elimination of water.

Conformational Pre-organization and Hydrogen-Bonding Patterns

The helix threaded through the center of the GFP β-barrel comprises a total of 11 residues (residues 60−71) that include the chromophore forming tripeptide. However, crystal structures of mature GFPs suggest that only three main-chain hydrogen bonds are formed within this helix: 60−64 (α-helical), 61−65 (α-helical), and 68−71 (3_{10}-helical) (Figure 4A).[5,6] Protein folding has long been thought to lock residues 65−67 into a tight-turn conformation that brings the reacting groups into close proximity.[4,35,36] This idea has recently been confirmed by the comparison of pre- and post-cyclization X-ray structures. In a seminal piece of work, Getzoff and co-workers have demonstrated that the anaerobic crystal structure of a variant carrying the S65G/Y66G substitutions is that of the pre-cyclization state.[31] Remarkably, residues 65−67 adopt a tight-turn conformation, and the buried helix exhibits the same main-chain hydrogen-bonding pattern as in mature GFP (Figure 4B). The lack of hydrogen bonds to the backbone atoms partaking in the nucleophilic addition reaction, O65 and N67, and the structural pre-organization of the tripeptide are likely essential features of the machinery facilitating main-chain condensation. In the

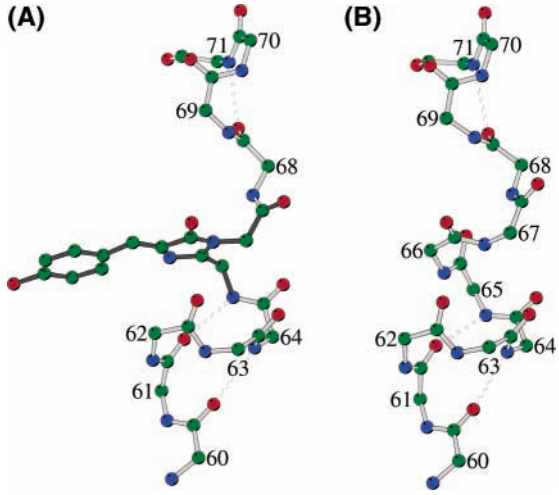


FIGURE 4. (A) Backbone conformation of the central helix comprising residues 60−71, bearing the mature chromophore (dark bonds), generated from pdb ID 1EMA.[5] (B) Backbone conformation of residues 60−71 in the pre-cyclization X-ray structure of a compromised GFP variant (pdb ID 1QYO).[31] Hydrogen bonds are shown as dashed lines. Carbon atoms are shown in green, oxygen in red, and nitrogen in blue.

precyclization structure, unfavorable steric interactions could not be identified,[31] arguing against the mechanical compression hypothesis put forth some years ago.[33] In line with these results, we have demonstrated that the proteolytic susceptibility of compromised GFPs is not correlated with the extent of chromophore formation.[37]

In all GFP X-ray structures, the helical axis is bent by about 80° near Tyr66, a distortion that has been linked to the pre-organization of atoms involved in ring closure.[31] This angle may be important in maintaining a highly conserved hydrogen bond between O66 and the guanidinium group of Arg96 (Figure 5), an interaction that plays an essential role in chromophore biogenesis. A disruption of the internal hydrogen-bonding network around Arg96, in conjunction with a perturbation of interior packing interactions, has been linked to protein destabilization and inefficient chromophore biosynthesis.[37] On the basis of limited proteolysis data, we have observed a loss of conformational stability when the highly conserved Gly67 was replaced with an alanine residue.[37] Steric interference by the introduction of an extra methyl group likely disrupts interior side-chain packing, thus breaking the hydrogen bond between O66 and Arg96. This interpretation is supported by a substantial elevation of chromophore pK_a in the G67A variant.[37] Interestingly, the phi and psi angles of Gly67 in the pre-cyclization structure, −90 and −16°, place this residue into an allowed region of the Ramachandran plot, suggesting that the driving force for the conservation of Gly67 does not lie in the inherent conformational flexibility of glycine residues but rather in the necessity to maintain a hydrogen bond between O66 and Arg96.

Catalytic Roles of Active-Site Residues

Electrophilic and Base Catalysis. Aside from proper conformational alignment, several chemical features play

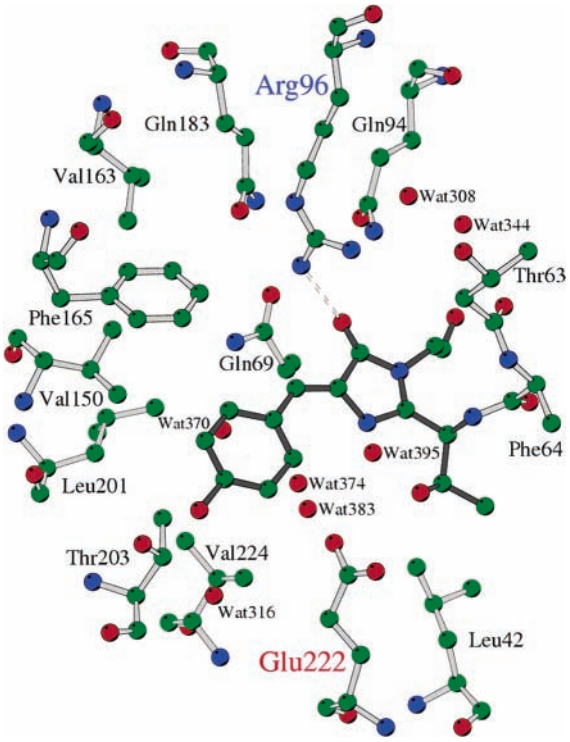


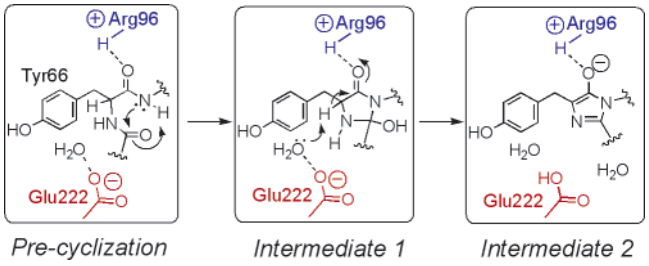
FIGURE 5. Model of the chromophore (dark bonds) and its protein environment (light bonds), generated from pdb ID 1EMA.[5] Carbon atoms are shown in green, oxygen in red, and nitrogen in blue.

a role in GFP color acquisition. The chromophore-bearing cavity in the interior of the β-barrel is surrounded by hydrophilic residues and filled with several crystallographically ordered water molecules (Figure 5). The highly conserved residues Arg96 and Glu222 originate from strands on opposing sides of the β-barrel (Figure 2), and their side-chain functional groups are positioned adjacent to the chromophore, slightly above and below the plane of the heterocycle (Figure 5). Charges carried by these groups are stabilized by hydrogen bonding, but not salt bridge interactions, and the spatial separation between the Glu222 carboxylate and the Arg96 guanidinium group is about 9 Å.[5]

Both Arg96 and Glu222 have been implicated in fulfilling catalytic roles in chromophore biogenesis.[4,5,13,37] Not surprisingly, the substitution of Arg96 with an alanine or methionine was reported to lead to extremely slow chromophore maturation extending over several months.[20,22,31] Evidently, the positive charge of Arg96 and its highly conserved hydrogen bond to O66 lower the activation barrier for peptide modification substantially.[4,22,31,36] Although the function of Arg96 can be partially compensated for by a lysine residue,[37] the recently published crystal structure of R96K indicates that the ϵ-amino group is not hydrogen bonded to O66.[22]

We have constructed a complete pH-rate profile for de novo chromophore biogenesis in the compromised variants R96M and E222Q[20] and compared the results to the intact, fast-maturing variant EGFP (enhanced green fluorescent protein).[29,38] We found that in vitro maturation of EGFP proceeds with a time constant of 1 h and is independent of pH, whereas the single substitution variants

Scheme 3. Proposed Mechanism for Pre-oxidation Steps in GFP Chromophore Biosynthesis, as Facilitated by Catalytic Residues Arg96 and Glu222[a]



[a] Sequence of proton-transfer reactions is as yet unknown.

R96M and E222Q display strong pH dependence.[20] The lack of a pH effect in EGFP as compared to E222Q suggests a model for chromophore synthesis in which the carboxylate of Glu222 functions as a general base, facilitating deprotonation of protein groups via transition state stabilization of proton-transfer reactions (Scheme 3).[20] This model requires that the buried Glu222 side chain be entirely deprotonated over a broad pH range.

The kinetic data suggest that in the compromised variants E222Q and R96M, the mechanism of peptide activation has changed from general to specific base catalysis.[20] The pH-rate profiles obtained for these variants fit well to a base-catalyzed reaction with a rapid preliminary equilibrium. According to this model, a protein functional group equilibrates with solution pH prior to the rate-determining step, yielding the catalytically active conjugate base in proportion to the hydroxide ion concentration. In the R96M variant, we have extracted an apparent pK_a value of 6.5 that may arise from titration of Glu222. Likely, this group displays a somewhat elevated pK_a in R96M since the positive charge of Arg96 has been removed.

pK_a Depression of Peptide Main-Chain Groups. Kinetic experiments on the E222Q variant indicate that at physiological pH, chromophore biogenesis extends over many hours (time constant $\tau = 32$ h at pH 7 and 7.4 h at pH 8).[20] Mass spectrometry data on immature proteins suggest the accumulation of a pre-oxidation species, given that no mass loss is observed by trypsinolysis/MALDI. Remarkably, at a high pH such as pH 9.5, the time constant for E222Q chromophore formation is reduced to 14 min, indicating a 4-fold rate enhancement over the parent molecule EGFP. The kinetic model supports rate modulation by protein-derived basic groups with an apparent pK_a value of about 9.2.[20] We have interpreted this proton dissociation constant as arising from either the Gly67 amide nitrogen (N67), the Tyr66 α-carbon (Cα66), or a combination thereof. For both groups, pK_a depression by many orders of magnitude is thought to have its foundation in electrophilic catalysis by Arg96. Although a normal amide nitrogen is estimated to have a pK_a value of about 15, the hydrogen bond of O66 to Arg96 in the protein's core is expected to lower this value substantially. In support of this notion, the small molecule acetamide is known to undergo a nitrogen pK_a shift from 15 to about 7 upon O-protonation.[39] In addition to the

GFP amide nitrogen N67, Cα66 is expected to be acidified as well, in analogy to enzymes such as enolases and racemases, where acidification of a carbon acid is often aided by coordination of a magnesium ion to the attached carboxylate.[40]

Although the nucleophilic reactivity of N67 in GFP is augmented by the positive charge of Arg96, it is not essential that the nitrogen lose its proton prior to nucleophilic attack. One of several plausible cyclization mechanisms may comprise proton transfer from N67 to O65, yielding a charge-neutral intermediate (Scheme 3).[21] Perhaps the apparent dissociation constant of about 9.2 (in E222Q) primarily reflects titration of Cα66 rather than N67. Energetic arguments suggest that the carbon acid would be more difficult to titrate than the amide nitrogen, implying that titration of Cα66 would contribute more strongly toward rate modulation. However, the sequence of proton-transfer steps is as yet unknown, and further experiments are necessary to elucidate these details.

Formation of an α-Enolate Intermediate. We have proposed that formation of the α-enolate of Tyr66 is essential in priming the active site for the transfer of redox equivalents to molecular oxygen,[21] in analogy to carbanion forming enzymes that carry out oxygenase side reactions.[41] Recently, the anaerobic X-ray structure of the chemically reduced Y66H variant of GFP was solved, providing evidence of a planar dehydrated heterocycle bound to a tetrahedral Cβ66.[28] Compelling geometric arguments lead the authors to assign this structure to the aromatic α-enolate form of the five-membered ring (Scheme 1, intermediate **2**), likely a pre-oxidation intermediate on the reaction path of chromophore synthesis in which the negative charge is stabilized by Arg96 (Scheme 3).

A striking feature in a large number of GFP X-ray structures is the position of a highly ordered solvent molecule that is located within hydrogen-bonding distance to Glu222 and in van der Waals contact to Cα66 (Figure 5, Wat 395). Geometric arguments provide strong support in favor of proton abstraction from the protein-based carbon acid Cα66 by this water molecule, which is polarized by the carboxylate Glu222 (Scheme 3). Remarkably, in the S65G/Y66G pre-cyclization structure, the distance between solvent molecule and Cα66 is only 3.3 Å.[31] Although Cα66 is substantially acidified by Arg96, additional acidification may be achieved by the conformational preference for a trigonal-planar carbon center in a cyclic main-chain conformation. In support of this notion, dehydro-phenylalanine is known to be a strong inducer of β-bends in short peptides and of 3_{10}-helical conformations in longer peptides,[42] conformations that place the carbonyl carbon of the residue (i) in close proximity to the amide nitrogen of residue ($i + 2$).[43] Although some years ago, a trigonal-planar α-carbon was proposed to be a prerequisite for main-chain condensation,[36] the same argument may be used to propose an additional drop in pK_a driven by an increase in ring planarity upon sp^2-hybridization.

Kinetics of Protein Oxidation

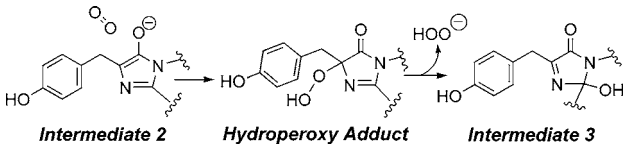
The rate-limiting process in GFP maturation has long been known to depend on an aerobic environment.[18] However, until recently, it has been unclear whether the oxidation process is slow due to weak O_2 binding, slow electron transfer, or slow proton-transfer reactions necessary to complete the reduction of O_2 to H_2O_2. To delineate these processes more precisely, we have used a hydrogen peroxide indicator dye to measure the progress of protein oxidation independently from the acquisition of green fluorescence. The colorimetric assay has provided clear evidence that the maturation of mGFPsol yields the coproduct H_2O_2, which is produced at a 1:1 molar ratio (H_2O_2/chromophore).[25]

Pre-oxidation, Oxidation, and Post-oxidation Kinetics. The kinetic trace for hydrogen peroxide evolution, as triggered by de novo protein folding, demonstrates that the reduced oxygen species is released from the protein prior to the acquisition of green fluorescence.[25] These data have allowed us to resolve the GFP maturation kinetics into pre-oxidation, oxidation, and post-oxidation events and to extract rate constants for each of these processes by the use of global curve fitting procedures. The pre-oxidation steps likely comprise protein folding, main-chain condensation, and associated proton-transfer reactions (Scheme 3). In combination, these events are estimated to proceed with a time constant of roughly 3 min and are thought to yield intermediate **2** (Scheme 1). The ensuing protein oxidation steps lead to the formation of intermediate **3** and the evolution of hydrogen peroxide. These steps proceed with a time constant of 37 (±3) min under ambient aerobic conditions at an O_2 concentration of about 200 μM. A subsequent event completes the process of chromophore biogenesis and exhibits a time constant of 15 (±5) min. Although the post-oxidation step contributes significantly to rate retardation, the production of hydrogen peroxide is the major rate-limiting event.[25]

The separation of pre- and post-cyclization species by reversed-phase HPLC supports the notion that peptide condensation is a rapid process in relation to the ensuing steps. Trypsinolysis/MALDI of intermediates isolated by HPLC gave convincing evidence of an accumulating protein species that has undergone a mass loss of 2 Da, consistent with the formation of intermediate **3** (Scheme 1).[25] Taken together, the kinetic and mass spectrometric data suggest a self-modification process in which the transfer of redox equivalents to molecular oxygen yields the hydroxylated cyclic imine, a species that was found to be trapped in the colorless Y66L variant (Scheme 2 **c**).[21]

Kinetic Isotope Effects. The availability of kinetic tools to monitor GFP oxidation directly has set the stage for a more detailed analysis of the rate-retarding steps. To this end, we have employed biosynthetic labeling methodologies to incorporate Cβ-di-deuterated tyrosine into the intact variant mGFPsol. For deuterated protein, the time constant for the post-oxidation step was found to be increased from 15 (±5) to 90 (±28) min, yielding a full

Scheme 4. Hypothetical Formation of Hydroperoxy Intermediate upon Air Oxidation of Intermediate 2



primary deuterium isotope effect k_H/k_D of 5.9 (±2.8) for the process following H_2O_2 evolution (L. Zhang and R. M. Wachter, unpublished observations). The data are in accord with a mechanism in which the ejection of water from intermediate **3** is prompted by slow C−H bond cleavage at Cβ66, likely involving proton abstraction from the acidified carbon center. In support of this interpretation, no isotope effect could be discerned for peroxide formation ($k_H/k_D = 1.1 \pm 0.2$), providing additional evidence that mGFPsol maturation cannot bypass the cyclic imine stage.

Several years ago, we proposed the formation of a peroxy or hydroperoxy adduct to Cα66 as an intermediate form generated during oxidation (Scheme 4).[21] Although the involvement of a hydroperoxy form in GFP is currently speculative, these kinds of intermediates have been proposed in several enzyme systems that utilize molecular oxygen in conjunction with organic cofactors such as flavins and pterins.[3] Perhaps the extent to which a particular GFP variant partitions through intermediate **3** is a function of the heterocylce's hydration equilibrium, which may shift during different stages of the reaction and may be modulated by amino acid substitutions. For example, the reorientation of the Glu222 side chain in variants containing the S65T substitution (such as EGFP and mGFPsol)[5] may result in channeling of the reaction through intermediate **3**, thus accelerating the process in relation to wild-type GFP.

Conclusion

Future studies will allow us to better understand why some GFPs mature more rapidly than others and whether a correlation exists between the maturation rate, the dominant reaction path, and the precise positioning of catalytic groups. We hope that this kind of knowledge will aid in the design of novel amino acid substitutions to develop fast-maturing fluorescent proteins with different optical properties. Temporal resolution is a key factor in the success of live cell imaging of biological processes such as promoter activity. The availability of a palette of GFPs exhibiting rapid color acquisition would be of tremendous advantage to researchers studying stem cell differentiation and organismal development.[44]

Aside from the family of GFP-like proteins,[14] main-chain condensation reactions involving nucleophilic addition−elimination have been described in several ammonia lyases and an amino mutase.[45,46] A detailed understanding of the GFP self-modification process will have broader relevance in the prediction of similar chemistries in unrelated proteins since the type of protein cross-link first identified in GFP may be present in other, less well-characterized proteins.

This work was supported by the National Science Foundation (NSF), Grant MCB-0213091.

References

(1) Okeley, N. M.; van der Donk, W. A. Novel cofactors via post-translational modifications of enzyme active sites. *Chem. Biol.* **2000**, *7*, 159−171.
(2) Ghiladi, R. A.; Medzihradszky, K. F.; de Montellano, P. R. O. Role of the Met-Tyr-Trp cross-link in *Mycobacterium tuberculosis* catalase-peroxidase (KatG) as revealed by KatG(M255I). *Biochemistry* **2005**, *44*, 15093−15105.
(3) Klinman, J. P. Life as aerobes: are there simple rules for activation of dioxygen by enzymes? *J. Biol. Inorg. Chem.* **2001**, *6*, 1−13.
(4) Tsien, R. Y. The Green fluorescent protein. *Ann. Rev. Biochem.* **1998**, *67*, 509−544.
(5) Ormo, M.; Cubitt, A. B.; Kallio, K.; Gross, L. A.; Tsien, R. Y.; Remington, S. J. Crystal structure of the *Aequorea victoria* green fluorescent protein. *Science* **1996**, *273*, 1392−1395.
(6) Yang, F.; Moss, L. G.; Phillips, G. N. The molecular structure of green fluorescent protein. *Nat. Biotech.* **1996**, *14*, 1246−1251.
(7) Wachter, R. M. The family of GFP-like proteins: Structure, function, photophysics, and biosensor applications. *Photochem. Photobiol.* **2006**, *82*, 339−344.
(8) Shimomura, O. Discovery of green fluorescent protein. *Methods Biochem. Anal.* **2006**, *47*, 1−13.
(9) Chalfie, M.; Tu, Y.; Euskirchen, G.; Ward, W. W.; Prasher, D. C. Green fluorescent protein as a marker for gene expression. *Science* **1994**, *263*, 802−805.
(10) Zhang, J.; Campbell, R. E.; Ting, A. Y.; Tsien, R. Y. Creating new fluorescent probes for cell biology. *Nat. Rev. Mol. Cell Biol.* **2002**, *3*, 906−918.
(11) Lippincott-Schwartz, J.; Patterson, G. H. Development and use of fluorescent protein markers in living cells. *Science* **2003**, *300*, 87−91.
(12) Tsien, R. Y. Building and breeding molecules that spy on cells and tumors. *FEBS Lett.* **2005**, *579*, 927−932.
(13) Matz, M. V.; Fradkov, A. F.; Labas, Y. A.; Savitsky, A. P.; Zaraisky, A. G.; Markelov, M. L.; Lukyanov, S. A. Fluorescent proteins from nonbioluminescent Anthozoa species. *Nat. Biotechnol.* **1999**, *17*, 969−973.
(14) Verkhusha, V.; Lukyanov, K. A. The molecular properties and applications of Anthozoa fluorescent proteins and chromoproteins. *Nat. Biotechnol.* **2004**, *22*, 289−296.
(15) Dove, S. G.; Hoegh-Guldberg, O.; Ranganathan, S. Major color patterns of reef-building corals are due to a family of GFP-like proteins. *Coral Reefs* **2001**, *19*, 197−204.
(16) Salih, A.; Larkum, A.; Cox, G.; Kuhl, M. Fluorescent pigments in corals are photoprotective. *Nature* **2000**, *408*, 850−853.
(17) Heim, R.; Prasher, D. C.; Tsien, R. Y. Wavelength mutations and post-translational autoxidation of green fluorescent protein. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 12501−12504.
(18) Cubitt, A. B.; Heim, R.; Adams, S. R.; Boyd, A. E.; Gross, L. A.; Tsien, R. Y. Understanding, improving, and using green fluorescent proteins. *Trends Biochem. Sci.* **1995**, *20*, 448−455.
(19) Reid, B. G.; Flynn, G. C. Chromophore formation in green fluorescent protein. *Biochemistry* **1997**, *36*, 6786−6791.
(20) Sniegowski, J. A.; Lappe, J. W.; Patel, H. N.; Huffman, H. A.; Wachter, R. M. Base catalysis of chromophore formation in Arg96 and Glu222 variants of green fluorescent protein. *J. Biol. Chem.* **2005**, *280*, 26248−26255.
(21) Rosenow, M. A.; Huffman, H. A.; Phail, M. E.; Wachter, R. M. The crystal structure of the Y66L variant of green fluorescent protein supports a cyclization-oxidation-dehydration mechanism for chromophore maturation. *Biochemistry* **2004**, *43*, 4464−4472.
(22) Wood, T. I.; Barondeau, D. P.; Hitomi, C.; Kassmann, C. J.; Tainer, J. A.; Getzoff, E. D. Defining the role of arginine 96 in green fluorescent protein fluorophore biosynthesis. *Biochemistry* **2005**, *44*, 16211−16220.
(23) Matz, M. V.; Lukyanov, K. A.; Lukyanov, S. A. Family of the green fluorescent protein: journey to the end of the rainbow. *BioEssays* **2002**, *24*, 953−959.
(24) Remington, S. J. Negotiating the speed bumps to fluorescence. *Nat. Biotechnol.* **2002**, *20*, 28−29.
(25) Zhang, L.; Patel, H. N.; Lappe, J. W.; Wachter, R. M. Reaction progress of chromophore biogenesis in green fluorescent protein. *J. Am. Chem. Soc.* **2006**, *128*, 4766−4772.

(26) Rosenow, M. A.; Patel, H. N.; Wachter, R. M. 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. *Biochemistry* **2005**, *44*, 8303−8311.
(27) Barondeau, D. P.; Kassmann, C. J.; Tainer, J. A.; Getzoff, E. D. Understanding GFP chromophore biosynthesis: Controlling backbone cyclization and modifying post-translational chemistry. *Biochemistry* **2005**, *44*, 1960−1970.
(28) Barondeau, D. P.; Tainer, J. A.; Getzoff, E. D. Structural evidence for an enolate intermediate in GFP fluorophore biosynthesis. *J. Am. Chem. Soc.* **2006**, *128*, 3166−3168.
(29) Heim, R.; Cubitt, A. B.; Tsien, R. Y. Improved green fluorescence. *Nature* **1995**, *373*, 663−664.
(30) Campbell, R. E.; Tour, O.; Palmer, A. E.; Steinbach, P. A.; Baird, G. S.; Zacharias, D. A.; Tsien, R. Y. A monomeric red fluorescent protein. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 7877−7882.
(31) Barondeau, D. P.; Putnam, C. D.; Kassmann, C. J.; Tainer, J. A.; Getzoff, E. D. Mechanism and energetics of green fluorescent protein chromophore synthesis revealed by trapped intermediate structures. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 12111−12116.
(32) Nagai, T.; Ibata, K.; Park, E. S.; Kubota, M.; Mikoshiba, K.; Miyawaki, A. A variant of yellow fluorescent protein with fast and efficient maturation for cell-biological applications. *Nat. Biotechnol.* **2002**, *20*, 87−90.
(33) Baedeker, M.; Schulz, G. E. Autocatalytic peptide cyclization during chain folding of histidine ammonia-lyase. *Structure* **2002**, *10*, 61−67.
(34) Wachter, R. M.; Yarbrough, D.; Kallio, K.; Remington, S. J. Crystallographic and energetic analysis of binding of selected anions to the yellow variants of green fluorescent protein. *J. Mol. Biol.* **2000**, *301*, 159−173.
(35) Branchini, B. R.; Nemser, A. R.; Zimmer, M. A computational analysis of the unique protein-induced tight turn that results in post-translational chromophore formation in green fluorescent protein. *J. Am. Chem. Soc.* **1998**, *120*, 1−6.
(36) Siegbahn, P. E. M.; Wirstam, M.; Zimmer, M. Theoretical study of the mechanism of peptide ring formation in green fluorescent protein. *Int. J. Quantum Chem.* **2001**, *81*, 169−186.
(37) Sniegowski, J. A.; Phail, M. E.; Wachter, R. M. Maturation efficiency, trypsin sensitivity, and optical properties of Arg96, Glu222, and Gly67 variants of green fluorescent protein. *Biochem. Biophys. Res. Commun.* **2005**, *332*, 657−663.
(38) Cormack, B. P.; Valdivia, R. H.; Falkow, S. FACS-optimized mutants of the green fluorescent protein (GFP). *Gene* **1996**, *173*, 33−38.
(39) Martin, R. B.; Hutton, W. C. Predominant N-bound hydrogen exchange via O-protonated amide. *J. Am. Chem. Soc.* **1973**, *95*, 4752−4754.
(40) Remington, S. J. Mechanism of citrate synthase and related enzymes (triose phosphate isomerase and mandelate racemase). *Curr. Opin. Struct. Biol.* **1992**, *2*, 730−735.
(41) Abell, L. M.; Schloss, J. V. Oxygenase side reactions of acetolactate synthase and other carbanion forming enzymes. *Biochemistry* **1991**, *30*, 7883−7887.
(42) Jain, R. M.; Rajashankar, K. R.; Ramakumar, S.; Chauhan, V. S. First observation of left-handed helical conformation in a dehydro peptide containing two L-Val residues. Crystal and solution structure of Boc-L-Val-ΔPhe-ΔPhe-ΔPhe-L-Val-OMe. *J. Am. Chem. Soc.* **1997**, *1997*, 3205−3211.
(43) Donnelly, M.; Fedeles, F.; Wirstam, M.; Siegbahn, P. E. M.; Zimmer, M. Computational analysis of the autocatalytic post-translational cyclization observed in histidine ammonia lyase. A comparison with green fluorescent protein. *J. Am. Chem. Soc.* **2001**, *123*, 4679−4686.
(44) Hong, J.-H.; Hwang, E. S.; McManus, M. T.; Amsterdam, A.; Tian, Y.; Kalmukova, R.; Mueller, E.; Benjamin, T.; Spiegelman, B. M.; Sharp, P. A.; Hopkins, N.; Yaffe, M. B. TAZ, a transcriptional modulator of mesenchymal stem cell differentiation. *Science* **2005**, *309*, 1074−1078.
(45) Schwede, T. F.; Retey, J.; Schulz, G. E. Crystal structure of histidine ammonia-lyase revealing a novel polypeptide modification as the catalytic electrophile. *Biochemistry* **1999**, *38*, 5355−5361.
(46) Christenson, S. D.; Liu, W.; Toney, M. D.; Shen, B. A novel 4-methylideneimidazole-5-one-containing tyrosine aminomutase in enediyne antitumor antibiotic C-1027 biosynthesis. *J. Am. Chem. Soc.* **2003**, *125*, 6062−6063.

AR040086R