

Kinetic Isotope Effect Studies on the de Novo Rate of Chromophore Formation in Fast- and Slow-Maturing GFP Variants[†]

Lauren J. Pouwels,[‡] Liping Zhang,[‡] Nam H. Chan,[‡] Pieter C. Dorrestein,[§] and Rebekka M. Wachter^{*,‡}

The Department of Chemistry and Biochemistry, Arizona State University, Tempe, Arizona 85287-1604, and The Departments of Pharmacology, Chemistry and Biochemistry, University of California San Diego, La Jolla, California 92093-0636

Received April 23, 2008; Revised Manuscript Received July 11, 2008

ABSTRACT: The maturation process of green fluorescent protein (GFP) entails a protein oxidation reaction triggered by spontaneous backbone condensation. The chromophore is generated by full conjugation of the Tyr66 phenolic group with the heterocycle, a process that requires C–H bond scission at the benzylic carbon. We have prepared isotope-enriched protein bearing tyrosine residues deuterated at the beta carbon, and have determined kinetic isotope effects (KIEs) on the GFP self-processing reaction. Progress curves for the production of H₂O₂ and the mature chromophore were analyzed by global curve fitting to a three-step mechanism describing preoxidation, oxidation and postoxidation events. Although a KIE for protein oxidation could not be discerned ($k_H/k_D = 1.1 \pm 0.2$), a full primary KIE of 5.9 (± 2.8) was extracted for the postoxidation step. Therefore, the exocyclic carbon is not involved in the reduction of molecular oxygen. Rather, C–H bond cleavage proceeds from the oxidized cyclic imine form, and is the rate-limiting event of the final step. Substantial pH-dependence of maturation was observed upon substitution of the catalytic glutamate (E222Q), indicating an apparent pK_a of 9.4 (± 0.1) for the base catalyst. For this variant, a KIE of 5.8 (± 0.4) was determined for the intrinsic time constant that is thought to describe the final step, as supported by ultra-high resolution mass spectrometric results. The data are consistent with general base catalysis of the postoxidation events yielding green color. Structural arguments suggest a mechanism in which the highly conserved Arg96 serves as catalytic base in proton abstraction from the Tyr66-derived beta carbon.

In recent years, the post-translational modifications yielding the colorful chromophores of GFP-like¹ proteins have been studied extensively (1, 2). Members of the family of fluorescent proteins (FPs) are evolutionarily related to its founding member avGFP from the jellyfish *Aequorea victoria* (3), and are generally found in marine organisms such as reef-building corals. FPs have attracted considerable interest, due to their ability to synthesize brightly fluorescing entities from intrinsic amino acid residues, such that the intense coloration of the mature protein appears to be an inherent property of the particular genetic sequence. The interior of the eleven-stranded β -barrel of FPs contains a helical peptide that spans residues 60–71 in avGFP, with position 66 occupied by a conserved tyrosine and 67 by a conserved glycine residue. Autocatalytic backbone cross-linking of residues 65 and 67 is followed by the net transfer

of two redox equivalents to atmospheric oxygen, ultimately yielding a chromophore that remains covalently attached to the polypeptide backbone and is buried in the protein's interior (Scheme 1) (4). Substantial progress has been made in gaining a better understanding of the main chain condensation reaction that triggers the ensuing sensitivity to air oxidation (5). O₂-mediated oxidation is ultimately responsible for the π -orbital extension of the Tyr66 phenolic group over the benzylic carbon center (C _{β} 66), such that desaturation of the C _{α} –C _{β} bond provides conjugation with the heteroaromatic ring constructed from main chain atoms. Aside from H₂O₂, the product of the reaction is a two-ring chromophore embedded in the protein matrix, usually in a planar conformation (Scheme 1) (1).

Previous work has focused on the preorganization of reacting atoms by the protein fold, a feature that allows for nucleophilic addition of the amide nitrogen of Gly67 (N67) to the carbonyl carbon of residue 65 (C65), while eliminating the energetic cost associated with the disruption of hydrogen bonds typically found in buried helical structures (5). This reaction appears to be facilitated by the highly conserved Arg96 with its guanidinium group hydrogen bonded to the main chain oxygen of residue 66. Mutagenesis studies have demonstrated that Arg96 imparts electrostatic catalysis by activating the 66–67 peptide bond (6–8). In addition, Arg96 has been proposed to play a role in stabilizing the superoxide anion radical that may form upon interaction with molecular oxygen (9, 10). In previous work, we have demonstrated that

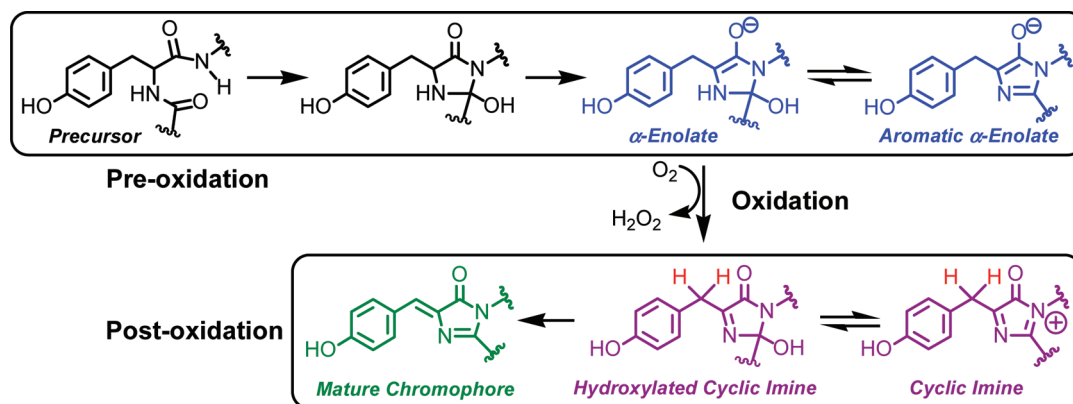
[†] This work was supported by a grant from the National Science Foundation (NSF MCB-0615938) and a grant from the National Institutes of Health (NIH RO3-EB006413) to R.M.W. NSF Grant CHE-0131222 provided funds to purchase the MALDI instrument.

* E-mail: rwachter@asu.edu. Phone: 480-965-8188. Fax: 480-965-2747.

[‡] Arizona State University.

[§] University of California San Diego.

¹ Abbreviations: KIE, kinetic isotope effect; FP, fluorescent protein; GFP, green fluorescent protein; MALDI, matrix-assisted laser desorption ionization mass spectrometry; ESI, electrospray ionization mass spectrometry; LTQ-FTMS, linear ion trap quadrupole Fourier transform mass spectrometry; HPLC, high pressure liquid chromatography; ACN, acetonitrile; TFA, trifluoroacetic acid; amu, atomic mass units.

Scheme 1: Proposed Three-Step Mechanism for GFP Maturation, Consisting of Preoxidation, Oxidation, and Postoxidation Processes^a

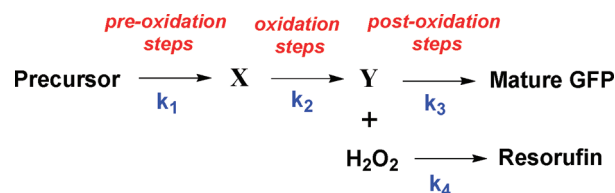
^a Proposed hydration–dehydration equilibria of the α -enolate and cyclic imine forms are also shown.

the main-chain condensation reaction is facile, occurring essentially within the dead time of manual mixing and HPLC injection methods (11). These data have provided support for the existence of an equilibrium between the open and closed forms of the protein, as exposure to the acidic HPLC conditions allows for the separation of pre- and postcyclization species of GFP (11). Likely, the closed five-membered ring is favored in the folded state of the protein. Although the aromatic α -enolate has been proposed to be the dominant preoxidation species (10), a mix of hydrated and dehydrated ring structures has been observed crystallographically in some GFPs (12–15), and the internal hydration–dehydration equilibria are as yet poorly understood (Scheme 1).

Several years ago, we proposed that GFP self-processing entails a protein oxidation reaction that remains centered on the five-membered ring, and does not affect the hybridization state of the benzylic $C_{\beta 66}$ carbon. This idea was based on the X-ray structure of the Y66L variant, where the condensed main-chain was found to be oxidized to the hydroxylated cyclic imine form (Scheme 1) (9). Complete elimination of water from the heterocycle was observed only upon desaturation of the $C_{\alpha 66}$ – $C_{\beta 66}$ bond, which requires C–H bond cleavage at $C_{\beta 66}$. In the Y66L variant, rehybridization of this carbon center is compounded by the low degree of activation of the aliphatic side chain. Desaturation was only observed upon prolonged incubation in high concentrations of a general base such as fluoride, formate or acetate (14). These compounds were able to diffuse into the protein's interior and accept a proton under physiological conditions (16, 17). Redox chemistry did not play a role in rehybridization from sp^3 to sp^2 , rendering a radical mechanism for $C_{\beta 66}$ oxidation unlikely (14). Here, we present further evidence that the chemistry observed in Y66L is analogous to the normal maturation mechanism of intact GFPs.

Like Arg96, Glu222 is completely conserved in all known fluorescent proteins, and may be involved in facilitating multiple proton transfer steps during GFP maturation (3, 18). Substitution with the sterically equivalent glutamine residue results in a variant that exhibits reduced maturation rates at physiological pH values. At pH 8, E222Q matures with a time constant of seven hours, whereas the intact parent molecule EGFP exhibits a time constant of one hour in *in vitro* assays (6). In previous work, we were unable to

Scheme 2: Outline of the Three Major GFP Maturation Steps Modeled in Global Curve Fitting of Kinetic Data



characterize any intermediate states of E222Q by trypsinolysis/MALDI (6), although an intermediate species with a reduced mass (–2 Da) was clearly observed in mGFPsol, an enhanced variant of EGFP (11). In contrast to these intact GFPs, maturation of the E222Q variant provided substantial evidence for base catalysis (6). For this reason, we have previously interpreted the slow maturation rate at pH 8 in terms of inefficient backbone cyclization or inefficient $C_{\alpha 66}$ deprotonation (6), both preoxidation events. However, the data presented here have allowed us to revise our interpretation, adjusting our model for base catalysis to be focused on postoxidation events (Scheme 1).

Recent investigations on chromophore biosynthesis in intact GFPs have provided compelling evidence that hydrogen peroxide is produced prior to the rise of green fluorescence (Scheme 2) (11). We have demonstrated that the third and final step in GFP maturation proceeds with a time constant of about 11 min and is therefore partially rate-determining. This step follows the slowest part of the overall reaction, the net transfer of two hydrogen atoms to molecular oxygen, which proceeds with a time constant of about 34 min (11). Based on trypsinolysis/MALDI data, an intermediate with a mass loss of 2 Da accumulates during the reaction. These results are consistent with dehydrogenation prior to full protein maturation, in support of a mechanism that involves a long-lived oxidized intermediate. However, the previously published data do not address the chemical nature of the final step yielding green color.

An as yet unresolved question in GFP maturation concerns the mechanism of C–H bond scission at the benzylic CH_2 group, a process essential for full π -orbital conjugation of the phenolic with the imidazolinone ring. As the O_2 -mediated chemistry is completed prior to this event (Scheme 2) (11), homolytic C–H bond cleavage appears unlikely. Here, we

demonstrate that a proton is abstracted from C β 66, which is acidified by the electron-withdrawing properties of the oxidized imidazolinone ring (14). Preliminary KIE data for mGFPsol maturation were reported in a recent review article (1). In the present work, we present deuterium isotope effect data for the mGFPsol and E222Q self-processing reactions, and propose that Arg96 is the general base that catalyzes proton abstraction.

EXPERIMENTAL PROCEDURES

Preparation of Immature mGFPsol Protein. 6His-tagged mGFPsol, a GFP variant that has also been termed GFP-trix, bears the mutations F64L/S65T/F99S/M153T/V163A/A206K in wild-type avGFP background (11). Expression in rich media was carried out in strain JM109(DE3), utilizing the pRSET_B expression plasmid (Invitrogen) bearing the N-terminally 6His-tagged mGFPsol gene (12). mGFPsol inclusion bodies were prepared by bacterial expression at 42 °C, solubilized in buffer containing 8 M urea, 50 mM HEPES pH 7.9, 50 mM NaCl and 1 mM DTT, and purified by Ni-affinity chromatography in the denatured state essentially as described (9, 11). Kinetic experiments were carried out immediately after protein purification on the same day.

Preparation of Immature E222Q Protein. Protein folding from inclusion bodies proved to be extremely inefficient for the E222Q variant (avGFP- F64L/S65T/E222Q), termed E222Q in this paper. Therefore, E222Q was purified from soluble bacterial fractions by following a two-hour procedure as reported previously (6, 19). Rapid purification was followed by flash-freezing in 150 μ L aliquots (1.2–1.7 mg/mL protein) and storage at –80 °C. One liter of liquid media and a two-hour induction period (25 °C) yielded a total of 4–5 mg of protein purified to homogeneity. The same yield was obtained when E222Q-expressing *Escherichia coli* were grown in minimal media to incorporate a deuterium label into the expressed protein (see below).

Isotopic Enrichment Procedure. C β -dideuterated tyrosine was purchased from Cambridge Isotope Laboratories, Inc. The pRSET_B plasmid (Invitrogen) bearing DNA coding for 6His-tagged mGFPsol or E222Q was transformed into *E. coli* strain BL21(DE3) for protein expression in minimal media. Minimal media stock solutions were prepared in the following way. The trace elements stock solution was prepared by combining 85 mL of water with 0.60 g of CaCl $_2$ ·2H $_2$ O, 0.60 g of FeSO $_4$ ·7H $_2$ O, 0.115 g of MnCl $_2$ ·4H $_2$ O, 0.08 g of CoCl $_2$ ·6H $_2$ O, 0.07 g of ZnSO $_4$ ·7H $_2$ O, 0.03 g of CuCl $_2$ ·2H $_2$ O, 0.002 g of H $_3$ BO $_3$, 0.025 g of (NH $_4$) $_6$ Mo $_7$ O $_{24}$ ·4H $_2$ O, and 0.50 g of EDTA-Na $_2$, adjusting the volume to 100 mL, and filter-sterilizing the solution. The 5X minimal salts stock solution was prepared by the combination of 150 mL of water, 3 g of KH $_2$ PO $_4$, 11.32 g of Na $_2$ HPO $_4$ ·7H $_2$ O, 0.5 g of NaCl, and 1.0 g of NH $_4$ Cl. The pH was adjusted to 7.2, the volume to 200 mL, and the solution was heat-sterilized. The 5X amino acid stock solution (without tyrosine) was prepared by dissolving 2.5 g of alanine, 2.0 g of arginine, 2.0 g of aspartic acid, 250 mg of cysteine, 2.0 g of glutamine, 3.25 g of glutamic acid, 500 mg of histidine, 1.15 g of isoleucine, 2.1 g of lysine, 1.25 g of methionine, 650 mg of phenylalanine, 500 mg of proline, 10.5 g of serine, 1.15 g of threonine, 1.15 g of valine, 2.75 g

of glycine, and 1.15 g of leucine in a total volume of 1 L of deionized water, followed by heat-sterilization (20). A 10 mg/mL tryptophan stock solution was prepared separately by dissolving 1.0 g of tryptophan in 100 mL of 50 mM HEPES pH 7.9, 300 mM NaCl, followed by filter-sterilization. In addition, 20% glucose, 20 mg/mL thiamine, and 1.0 M MgCl $_2$ stock solutions were prepared in water. A 10 mM uracil stock was prepared in 50 mM HEPES pH 7.9, 300 mM NaCl.

The minimal medium was assembled to a final volume of 1 L by combining 200 mL of minimal salts stock, 20 mL of glucose stock, 2.0 mL of thiamine stock, 1.0 mL of MgCl $_2$ stock, 5.0 mL of trace elements stock, 20.0 mL of uracil stock, 25 mL of tryptophan stock and 200 mL of amino acid stock. 200 mg of ampicillin and 170 mg of C β -dideuterated tyrosine (as dry powder) were added just before inoculation. One liter of minimal medium was inoculated with 8 mL of overnight culture (LB broth supplemented with 90 μ g/mL carbenicillin). Flask cultures were grown in a shaker (300 rpm) at 37 °C until the OD $_{600}$ reached 0.8, then the temperature was adjusted to 42 °C to prepare mGFPsol inclusion bodies and to 25 °C to express soluble E222Q, and shaking was reduced to 250 rpm. mGFPsol was induced for 4.0 h and E222Q for 2 h by addition of 250 mg of IPTG, then harvested by centrifugation (6).

Trypsinolysis and Peptide Purification. To determine the incorporation efficiency of the deuterium label, the molecular mass of proteolytic peptides derived from isotopically labeled and unlabeled material was determined. Tryptic peptides of E222Q were generated exactly as described previously (6). For inclusion body-derived mGFPsol, 55 μ L of water was added to 165 μ L of 8 M urea-denatured protein to yield a final concentration of 6 M urea. Four microliters of 10 mg/mL trypsin (Sigma T-1426, TCPK-treated bovine) was added to 800 mL of digest buffer (50 mM HEPES pH 8.0, 300 mM NaCl, 20 mM CaCl $_2$) immediately before use, and combined with the protein solution. The digest was incubated for 3 min at 30 °C, and dithiothreitol was added to yield a concentration of 10 mM, followed by a two-minute incubation at room temperature. Tryptic peptides were separated by reverse-phase HPLC (Waters 600 HPLC system equipped with a Waters 996 photodiode array detector) on a C18 analytical column (Vydac) utilizing a water/ACN (acetonitrile) gradient containing 0.1% TFA (trifluoroacetic acid). Peptides were eluted after column equilibration at 45% solvent B for 10 min, followed by a linear gradient from 45% to 60% solvent B at 1% per minute. The absorbance was monitored at 220, 280 and 380 nm. Five fractions were collected in eppendorf tubes, split into two equal parts of 0.75 mL each, and lyophilized immediately. One part of each HPLC fraction was subjected to MALDI mass spectrometry, and the other part was saved for ESI ion trap mass spectrometric analysis.

MALDI Mass Spectrometry. Lyophilized peptides were resolubilized in 18 μ L of water, and 1 μ L of peptide solution was mixed with 3 μ L of a saturated solution of α -cyano-4-hydroxycinnamic acid matrix. MALDI data were collected as described previously, utilizing a Voyager DE STR mass spectrometer equipped with a nitrogen laser (6). Masses of all peptides eluted by HPLC were determined by MALDI and matched with calculated masses obtained from theoretical tryptic digests (see Supplementary Table 1, Supporting

Information). Peptides found in the HPLC fraction eluting at 51–52% solvent B (elution time 16 to 17 min) contained the chromophore-forming residues, as well as a C-terminal peptide fragment. This fraction was submitted for ion trap mass spectrometric analysis.

Nano-HPLC Ion Trap Mass Spectrometry. The extent of biosynthetic incorporation of C_{β} -dideuterated tyrosine into expressed protein was determined by the Proteomics Core Facility at the University of Arizona, Tucson, AZ. Lyophilized peptides from the HPLC fraction collected at 51–52% solvent B (see above) were solubilized in 50 μ L of water. A microbore HPLC system (Surveyor; ThermoFinnigan, Jan Jose, CA) was modified to operate at capillary flow rates using a simple T-piece flow-splitter. Columns (6 cm \times 100 μ m i.d.) were prepared by packing 100 Å, 5 μ m Zorbax C18 resin at 500 psi pressure into columns with integrated electrospray tips made from fused silica, pulled to a 5 mm tip using a laser puller (Sutter Instrument, Novato, CA). An electrospray voltage of 1.8 kV was applied using a gold electrode via a liquid junction upstream of the column. For each run, a 10 μ L sample aliquot was injected into the analytical column using a Surveyor autosampler. The HPLC column eluent was eluted directly into the ESI (electrospray ionization) source of a ThermoFinnigan LCQ-Deca XP Plus ion trap mass spectrometer. Peptides were eluted utilizing a linear gradient from 0% to 50% solvent B over a 30 min interval (solvent A, water/0.1% formic acid; solvent B, ACN/0.1% formic acid) at a flow rate of 400 nL/min. In order to resolve the isotope pattern of the doubly charged ($z = 2+$) peptides of interest, three separate MS runs were performed acquiring continuous zoom scans with a 10 amu (atomic mass unit) window for each of the three peptides previously characterized by MALDI (see above): $m/z = 1190$ (1185.5–1195.5), $m/z = 1296$ (1291.5–1301.5), and $m/z = 1565$ (1560.5–1570.5). For each doubly charged ($z = 2+$) peptide examined by the zoom method, the major isotopic fingerprint matched the calculated D2 peptide mass exactly (see Supplementary Table 2, Supporting Information). However, very small peaks were observed with m/z consistent with the H2 peptide. Although these peaks were essentially at noise level, they were utilized to calculate a lower limit for the isotopic incorporation efficiency. For the H2 and D2 versions of each peptide, peak areas were determined for the second isotope peak $M + 1$. For the peptide with experimentally determined monoisotopic $m/z = 2196.6$, minimum incorporation efficiency was calculated to be 97.5%, for $m/z = 1565.8$, 99.2%, and for $m/z = 1190.6$, 93.3%, providing an average of $96.6 \pm 3.1\%$ for the lower limit of isotope incorporation efficiency in our labeled protein preparations.

Kinetic Assay for Hydrogen Peroxide Evolution during mGFPsol Maturation. As described previously (11), immature urea-solubilized mGFPsol protein (approximately 0.2 mg/mL) was rapidly diluted into a stirred quartz fluorimeter cell with the thermostat set to 30 °C. To measure the generation of hydrogen peroxide, a newly purchased Amplex Red Hydrogen Peroxide Assay Kit (Molecular Probes, Inc.) was utilized. A 10 μ L aliquot of denatured protein was added to a 290 μ L aliquot of rapidly stirring solution consisting of 240 μ L folding buffer (50 mM HEPES pH 7.9, 300 mM NaCl, 1 mM EDTA) and 50 μ L Amplex Red (10-acetyl-10H-phenoxazine-3,7-diol) working solution prepared ac-

cording to the manufacturer's instructions (21). The red fluorescence of the resorufin indicator dye produced in response to hydrogen peroxide was monitored via emission at 580 nm ($\lambda_{\text{ex}} = 550$) as described previously (11). The conversion rate of Amplex Red to resorufin was determined by subjecting a 0.3 μ M H_2O_2 standard solution to identical assay conditions. The amount of hydrogen peroxide in the standard was approximately equal to that generated during a typical protein maturation experiment. The pseudo-first order rate constant for this reaction, termed k_4 , was extracted by computer-fitting the rise of fluorescence to the equation $F = F_{\text{max}} - e^{(-k_4t)} \times F_{\text{max}}$, utilizing the program Kaleidagraph. As before, the extracted rate constant k_4 was utilized to account for the reagent response time during protein maturation, and was determined anew for each individual experiment (11). In the current work, the reagent response time was found to be more rapid compared to previously published data (11), likely because newly purchased reagent was used. The reagent is light, air and temperature sensitive, and exhibits a limited shelf life, therefore all assays and control experiments for one kinetic data set were performed on the same day utilizing the same batch of Amplex Red working solution. Due to low levels of reagent autoxidation, all protein and hydrogen peroxide kinetic traces were baseline-corrected by subtracting the trace generated by an appropriate blank.

Kinetic Assay for Acquisition of Green Fluorescence upon mGFPsol Maturation. To monitor the rise of green fluorescence upon chromophore formation, a fresh 10 μ L aliquot of urea-denatured protein was diluted into folding buffer (50 mM HEPES pH 7.9, 300 mM NaCl, 1 mM EDTA) to yield a final volume of 300 μ L, and fluorescence emission was monitored at 510 nm ($\lambda_{\text{ex}} = 480$) in a rapidly stirring quartz cuvette at 30 °C (11). For the maturation study involving unlabeled protein, fluorimeter settings were as described previously (11). For the deuterium-labeled protein, fluorescence emission data were collected in one-minute intervals for the first 120 min, followed by two-minute intervals up to 240 min and five-minute intervals up to 360 min, to avoid photobleaching of the newly formed chromophore. All fluorescence experiments were carried out on a Jobin Ivon Horiba FluoroMax-3 instrument.

Kinetic Assay for E222Q Maturation. Kinetic experiments on immature E222Q were carried out at eight different pH values between pH 7.72 and 10.02. UV-vis absorbance scans were collected on a Shimadzu UV-2401 spectrophotometer as a function of time (6), and kinetic runs were performed on both unlabeled and isotope-labeled protein. For each experiment, aliquots of flash-frozen E222Q protein were removed from the freezer and diluted 10-fold into buffer of appropriate pH (20 mM PIPES, HEPES, CHES or CAPS, 300 mM NaCl, 1 mM EDTA), and an absorbance scan was collected immediately. Subsequently, the samples were capped to avoid evaporation and incubated in a temperature-controlled chamber at 30 °C. The increase in absorbance at 483 nm (chromophore anion) was determined by collection of absorbance scans at various time points until saturation was observed. At pH 10, time points were collected over a total period of 0.73 h for unlabeled protein and 4.0 h for isotope-labeled protein. At pH 7.7, the data collection time ranged from 32 h for unlabeled protein to 76 h for labeled protein. For each kinetic experiment, the change in chromophore absorbance was plotted as a function of time, and

computer-fitted to a rate equation describing a unimolecular reaction [$A = A_{\max} - e^{(-kt)} \times A_{\max}$], using the program Kaleidagraph (Figure 3 inset) (6). The extracted pseudo-first order rate constants k_{obs} were plotted as a function of pH. The pH-rate profiles for unlabeled and isotope-labeled protein pools were each computer-fitted to an expression for a base-catalyzed reaction of the form $k_{\text{obs}} = (k_{\text{int}} \times K_a) / (10^{[-\text{pH}]} + K_a)$. According to this model, the observed rate depends on the protonation equilibrium of an ionizable group with acid dissociation constant K_a , whereas k_{int} represents the intrinsic (pH-independent) rate constant.

Ultrahigh Resolution Mass Determination of Immature E222Q by LTQ-FTMS. 200 μL of 1.8 mg/mL flash-frozen E222Q was removed from the -80°C freezer, a UV/vis absorbance scan was collected, and 160 μL was purified by reverse-phase HPLC. The eluting protein peak was collected in two fractions, frozen immediately in liquid nitrogen and lyophilized overnight. The lyophilized samples were redissolved into 49.5:49.5:1.0 MeOH/water/formic acid. The samples were infused by nanoelectrospray ionization with a Biversa Nanomate (Advion Biosystems, Ithaca, NY), and analyzed on a Finnigan LTQ-FTMS (Thermo-Electron Corporation, San Jose, CA) running the Tune Plus software version 1.0 (Thermo). The data were collected with the autogain set to 5E6 and the maximum accumulation time set to 8000 ms and 200,000 resolution. The final spectrum was obtained by averaging 394 scans with the QualBrowser software version 1.4 SR1 (Thermo). The average spectrum was then subjected to deconvolution using the Extract option within the QualBrowser software. Theoretical spectra superimposed onto the experimental spectra are based on the calculated amino acid composition of the protein, and were generated with the freeware Isopro 3.0 and matched to the experimental spectra.

RESULTS

A Primary Deuterium Isotope Effect for the Postoxidation Step in GFP Maturation. To better understand the maturation process of green fluorescent protein, we have examined the mechanism of chromophore biogenesis upon incorporation of an isotopic label. To investigate whether carbon-hydrogen bond cleavage at $C_{\beta 66}$ is an integral part of O_2 -mediated oxidation, or whether this bond is disrupted at a later stage, we incorporated C_{β} -dideuterated tyrosines into the intact, fast-maturing variant mGFPsol (Figure 1 inset) (11). We argued that the protons attached to the benzylic carbon of Tyr66 are likely the least acidic protons to be removed during GFP maturation, suggesting that this process may entail significant rate retardation. Therefore, we produced the denatured precursor form of mGFPsol from inclusion bodies, and carried out maturation kinetic assays with and without the tyrosine-based deuterium label.

To incorporate deuterated tyrosine residues, mGFPsol was expressed in minimal media supplemented with a complete set of canonical amino acids, with the exception of tyrosine, which was replaced by C_{β} -dideuterated tyrosine (20). Inclusion bodies of mGFPsol containing either dihydro- or dideutero-tyrosines were prepared by protein expression at 42°C . The inclusion bodies were isolated and washed extensively, followed by solubilization in 8 M urea and affinity-purification. Maturation kinetic experiments were

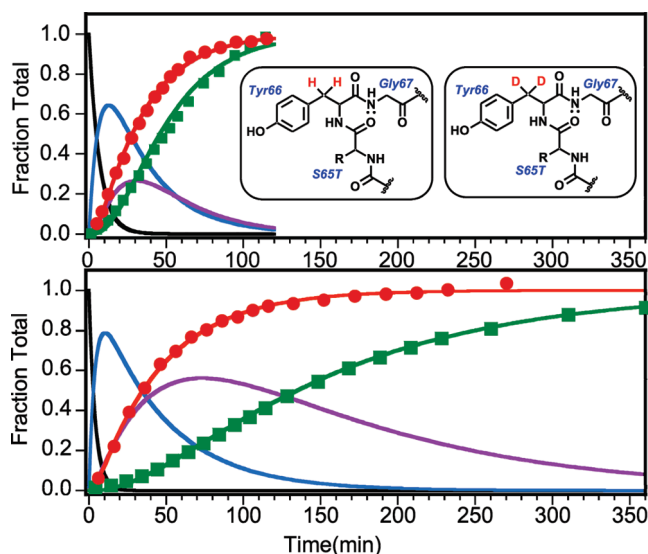


FIGURE 1: Progress curves for mGFPsol maturation of unlabeled protein (top panel) and isotopically enriched protein (bottom panel). Red (—●—): rise in red fluorescence due to resorufin production from hydrogen peroxide. Green (—■—): rise in green fluorescence due to formation of the mature GFP chromophore. The symbols represent experimental data (for clarity, only 10% of the collected data points are shown). The solid lines describe theoretical progress curves calculated from the rate constants extracted by global curve fitting of experimental data. Black: precursor protein. Blue: pre-oxidation intermediate (α -enolate). Purple: postoxidation intermediate (cyclic imine). Red: resorufin (hydrogen peroxide indicator). Green: mature protein. Inset: Schematic of precyclization structures with C_{β} -dihydro (unlabeled) and C_{β} -dideutero (labeled) Tyr66.

carried out as reported previously by triggering protein folding via rapid dilution of the chaotrope (11). For unlabeled mGFPsol, four independent kinetic runs were performed in the present study, such that inclusion of the recently published data allowed for statistical evaluation of a total of five complete data sets. Similarly, five independent kinetic experiments were performed on deuterium-labeled mGFPsol. Each complete data set consisted of a series of baseline-corrected kinetic traces that were collected on the same day utilizing the same aliquot of freshly prepared urea-solubilized protein. First, the rate of hydrogen peroxide evolution upon protein maturation was monitored by employing an enzyme-linked fluorogenic assay that produces the red-fluorescent compound resorufin in the presence of H_2O_2 (21). Second, the rate of resorufin production from H_2O_2 was monitored by use of a hydrogen peroxide standard solution. Third, the progress curve for complete protein maturation was determined by monitoring the green fluorescence emission of the mature chromophore. The protein self-processing reaction was initiated by 30-fold dilution from 8 M urea, and the reaction was followed for either two hours (unlabeled protein) or six hours (labeled protein) (Figure 1).

For each of the kinetic data sets, the progress curves for resorufin production and chromophore formation were fitted to a sequential three-step mechanism (Scheme 2) using the program DynaFit (22). In the global curve-fitting procedures, the rate constant for resorufin production from H_2O_2 (k_4) was held constant, and ranged from 1.1 to 2.3 min^{-1} ($\tau_4 = 0.43$ to 0.93 min), whereas the rate constants for the three major protein maturation steps were modeled by the variable parameters k_1 , k_2 , and k_3 . For each kinetic data set, values for these rate constants were extracted by modeling the time

Table 1: Kinetic Constants for mGFPsol Maturation from Inclusion Bodies

	preoxidation τ_1 (min)	oxidation τ_2 (min)	postoxidation τ_3 (min)	overall τ_{total} (min)
mGFPsol, Unlabeled ($n = 5$)				
av (\pm std dev)	1.93 (\pm 1.53)	37.10 (\pm 3.14)	15.28 (\pm 5.53)	54.31 (\pm 6.54)
mGFPsol, Deuterium-Labeled ($n = 5$)				
av (\pm std dev)	4.99 (\pm 2.45)	42.06 (\pm 6.45)	90.04 (\pm 27.95)	137.09 (\pm 28.79)
Deuterium Isotope Effects Calculated for mGFPsol				
k_H/k_D (\pm std dev)	2.58 (\pm 2.40)	1.13 (\pm 0.20)	5.89 (\pm 2.81)	2.52 (\pm 0.61)

evolution of six species (precursor, intermediate X, intermediate Y, mature GFP, H_2O_2 , and resorufin) as described previously (Scheme 2) (11). Average values and sample standard deviations were calculated from five individual determinations for each rate constant, obtained from curve fitting of five independent data sets each for dihydro and dideutero mGFPsol (Table 1).

We found that the three-step kinetic model provided an excellent fit to the data obtained for both unlabeled and deuterium-labeled protein (Figure 1). For the preoxidation process consisting of protein folding and main-chain cyclization, the extracted time constants $\tau_{1H} = 1.93 \pm 1.53$ min and $\tau_{1D} = 4.99 \pm 2.45$ min exhibit large standard deviations, such that the deuterium isotope effect, $k_H/k_D = 2.58 \pm 2.40$, remains indeterminate (Table 1). Likely, the large error in τ_1 is the result of inhomogeneous protein folding rates (23). The second major process, which produces hydrogen peroxide, proceeds on a much slower time scale, yielding the time constants $\tau_{2H} = 37.10 \pm 3.14$ min and $\tau_{2D} = 42.06 \pm 6.45$ min. From these data, the deuterium isotope effect k_H/k_D for oxidation is calculated to be 1.13 ± 0.20 (Table 1). Therefore, within experimental error, the oxidation rate does not appear to be modified upon deuteration, as the deviation from unity is smaller than the standard deviation determined by error propagation. In contrast, the average time constant for the postoxidation step is extended substantially upon deuteration, with a value of $\tau_{3H} = 15.28 \pm 5.53$ min for the unlabeled protein and a value of $\tau_{3D} = 90.04 \pm 27.95$ min for the dideutero protein. Therefore, τ_3 is largely responsible for the overall change in maturation rate observed upon tyrosine deuteration (Figure 1). The KIE for this step is calculated to be 5.89 ± 2.81 , consistent with a full primary deuterium isotope effect. Therefore, C–H bond scission at $\text{C}_{\beta 66}$ is part of the third step of protein maturation (Scheme 2), and constitutes the major rate-limiting event of the final process leading to green fluorescence. We conclude that desaturation of the bridging carbon $\text{C}_{\beta 66}$ is not an integral part of protein oxidation, but instead follows hydrogen peroxide production (Scheme 1).

Maturation of Soluble E222Q Proceeds from the Oxidized Intermediate Stage. To better define the involvement of catalytic residues in the three-step self-modification process, we purified immature E222Q protein in soluble form as described previously (6). This variant is derived from EGFP as is mGFPsol, and both bear the substitutions F64L/S65T adjacent to the chromophore π -system known to enhance brightness and folding (24–26). However, E222Q does not bear the cycle 3 folding mutations F99S/M153T/V163A distant from the active site, nor does it contain the A206K substitution that eliminates dimerization (27). With respect to the arguments presented in this work, these differences

are judged to be irrelevant, since immature E222Q was expressed in soluble form (this variant could not be folded efficiently from inclusion bodies (7)). We rapidly purified E222Q after a short induction period and flash-froze the partially mature protein. Absorbance measurements were utilized to determine that 33% of the freshly purified material had fully matured to the green-fluorescent stage. In previously performed trypsinolysis/MALDI experiments, we had been unsuccessful in detecting any oxidized intermediate states (6), therefore we proceeded with kinetic assays that monitor hydrogen peroxide evolution. In spite of numerous attempts under a wide variety of experimental conditions and pH values, we were unable to observe hydrogen peroxide production upon E222Q maturation, suggesting that the protein may already be oxidized.

To determine the types of intermediates trapped in flash-frozen E222Q, we collected reverse-phase HPLC fractions of full-length protein, and analyzed them by ultrahigh resolution mass spectrometry entailing nanoelectrospray ionization and LTQ Fourier transform detection. The spectra provided isotopic resolution of the intact protein chain, which has a calculated mass of 31,078 Da in the absence of any covalent modifications. We found that the early fraction (“up-slope”) of the HPLC peak consisted almost entirely of mature E222Q. This HPLC fraction provided mass spectral peaks with isotopic envelopes that fit well to species with masses 31,058 amu (-20 Da) and 31,074 amu ($-20 + 16 = -4$ Da) (Figure 2A). However, the mass spectrum did not provide clear evidence of either the precursor form (31,078 amu) or the dehydrogenated (oxidized) form (31,076 amu). The 31,074 species is consistent with the addition of one oxygen atom ($+16$ Da) to the mature protein, likely due to secondary oxidative events such as methionine sulfoxide formation. In support of nonspecific oxidative events, the addition of oxygen atoms is further demonstrated by LTQ-FTMS spectra collected on mature E222Q, where the observed bands are consistent with masses of 31,058 amu (mature), 31,074 amu (mature + 16 Da), and 31,090 amu (mature + 32 Da) (see Supplementary Figure 1, Supporting Information).

The mass spectrum collected for the bulk of the HPLC peak (“down-slope”) entailed three main bands, each with isotopic resolution (Figure 2B). The envelope of the most prominent band was reasonably well modeled by a species of mass 31,076 amu, i.e. a 2 Da loss from the mass of the precursor form, in line with protein oxidation via the net loss of two hydrogen atoms. A minor species with slightly larger mass appeared as a small shoulder in the isotopic envelope, consistent with the unmodified precursor form (31,078 amu). In addition, a peak consistent with 31,092 Da ($-2 + 16$ Da) was apparent, suggesting a secondary, non-

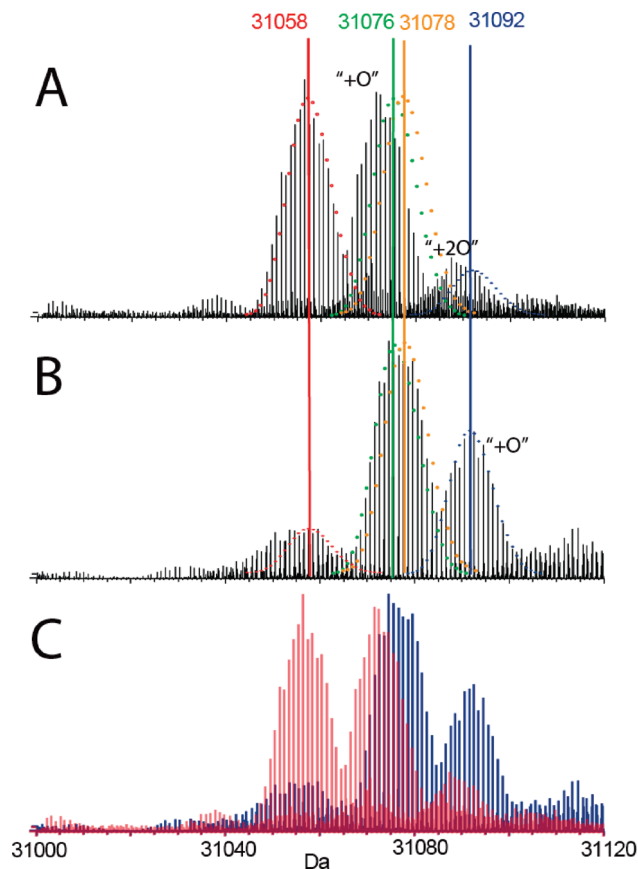


FIGURE 2: Mass spectra of full-length immature E222Q protein isolated in soluble form and purified by HPLC. The isotope-resolved experimental spectra were collected on a Finnigan LTQ-Fourier transform mass spectrometer. The isotopic envelopes represented by colored dots are the theoretical spectra calculated for the following masses: red, 31,058 amu (mature protein, -20 Da); green, 31,076 amu (protein oxidized via dehydrogenation, -2 Da); orange, 31,078 amu (protein precursor, mass unmodified); blue, 31,092 amu (dehydrogenated protein with addition of one oxygen atom, -2 Da + 16 Da). A. Spectrum of the early eluting HPLC fraction. B. Spectrum of the late-eluting HPLC fraction. C. Superimposition of spectra shown in A and B (red = A, blue = B).

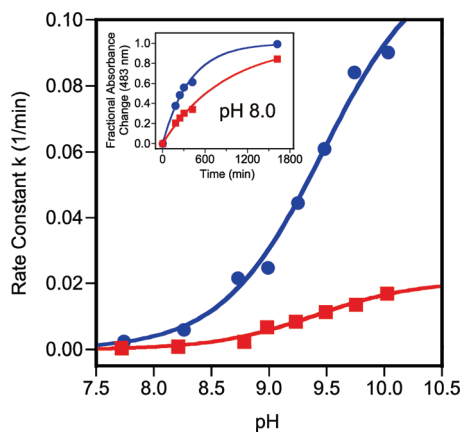


FIGURE 3: pH-rate profile of E222Q maturation, demonstrating the dependence of the observed rate constant on pH. Blue (\bullet): unlabeled protein. Red (\blacksquare): C_β -dideuterotyrosine-labeled protein. The symbols represent experimental data, and the solid lines represent curve fits to kinetic equations (see Experimental Procedures). Inset: Increase in chromophore absorbance (483 nm) over time, measured at pH 8.

specific oxygen atom addition to the dehydrogenated intermediate of the protein. Although this mass would also be

consistent with a dehydrated peroxy intermediate ($-18 + 32$ Da), peroxy adducts tend to be chemically unstable and would hydrolyze rapidly under the acidic HPLC conditions, rendering this interpretation unlikely. Another species is observed slightly above noise level, providing some evidence for the mature form at 31,058 Da, although the isotopic envelope is not reliable at such low peak intensities. A species consistent with dehydration (-18 Da) that would provide a mass of 31,060 amu could not be identified clearly, as its contribution (if present) would fall within baseline noise of the spectrum (Figure 2B).

The LTQ-FTMS data demonstrate that the early eluting protein population consists primarily of mature protein (-20 Da), whereas the late-eluting population consists primarily of the dehydrogenated (oxidized) intermediate (-2 Da). Trace amounts of the precursor are also observed, as indicated by the small amount of experimental peak area lying outside of the theoretical band modeled at 31,076 amu (Figure 2B, green). Based on the mass spectral analysis, the dehydrogenated protein species (with or without oxidative side reactions) dominate the immature protein population, as the peaks with monoisotopic masses of 31,076 and 31,092 amu are most pronounced (see spectral overlay shown in Figure 2C). Therefore, we estimate that the preoxidation species presents less than 10% of the starting material, consistent with the lack of H_2O_2 evolution (see above). Although the presence of small amounts of precursor could interfere with data interpretation, the time evolution of chromophore formation is well modeled by a single exponential over the entire pH range examined (Figure 3 inset). Therefore, the extracted rate constants are interpreted in terms of postoxidation chemistry only, as the third step appears to be the dominant process under the experimental conditions employed.

Maturation of Soluble E222Q Involves Removal of the Tyr66 C_β -Proton. Kinetic experiments on immature E222Q were carried out between pH 7.7 and pH 10.0 as described previously (6). Biosynthetic labeling of E222Q protein was performed as described for mGFPsol, and kinetic assays were carried out with both C_β -dihydro- and C_β -dideutero-tyrosines incorporated into the protein. The chromophore maturation rate was determined at 30°C by collecting UV-vis absorbance scans as a function of time, and determining the fractional increase in chromophore absorbance (483 nm) over time (Figure 3 inset). The time requirement for completion of the reaction varied substantially depending on pH and deuteration, and ranged from less than one hour to three days. Below pH 7.5, the maturation rate was too slow to allow for accurate data collection. For each kinetic experiment, the data were fit to a rate equation describing a pseudo-first order reaction (Figure 3 inset) (6), and a complete pH-rate profile was generated for the dihydro and the dideutero E222Q proteins (Figure 3). In both cases, the profile fit well to an expression describing a base-catalyzed reaction, in which the observed rate constant k_{obs} depends on the acid dissociation constant K_a of a titratable group, and an intrinsic, pH-independent rate constant k_{int} . The corresponding time constant τ_{int} was extracted to be 8.38 ± 0.55 min for unlabeled E222Q, and that for the isotopically enriched material was determined to be 48.21 ± 3.92 min (Table 2). From these values, the deuterium isotope effect k_H/k_D was determined to be 5.75 ± 0.35 , consistent with a full primary

Table 2: Kinetic Constants for E222Q Maturation Extracted from pH–Rate Profiles

pH–rate profile	C β -tyrosine labeling of protein pool	intrinsic time constant τ_{int} (min) (\pm std dev)	pK _a (\pm std dev)
present study	dihydrotyrosines	8.38 (\pm 0.55)	9.46 (\pm 0.07)
	dideuteroxyrosines	48.21 (\pm 3.92)	9.41 (\pm 0.09)
	$k_{\text{H}}/k_{\text{D}}$ (\pm std dev)	5.75 (\pm 0.35)	—
previous study (ref6)	dihydrotyrosines	14.3 (\pm 3.0)	9.2 (−0.2; +0.4)

KIE. Therefore, C–H bond cleavage at the β -carbon of Tyr66 constitutes the major rate-limiting event of the process under consideration. As proton abstraction is fully rate-determining, and the immature protein population consists primarily of oxidized species (−2 Da), we conclude that the rate constant k_{int} determined for E222Q describes the third of the three-step maturation process. This rate constant is therefore equivalent to k_3 in mGFPsol maturation (Scheme 2).

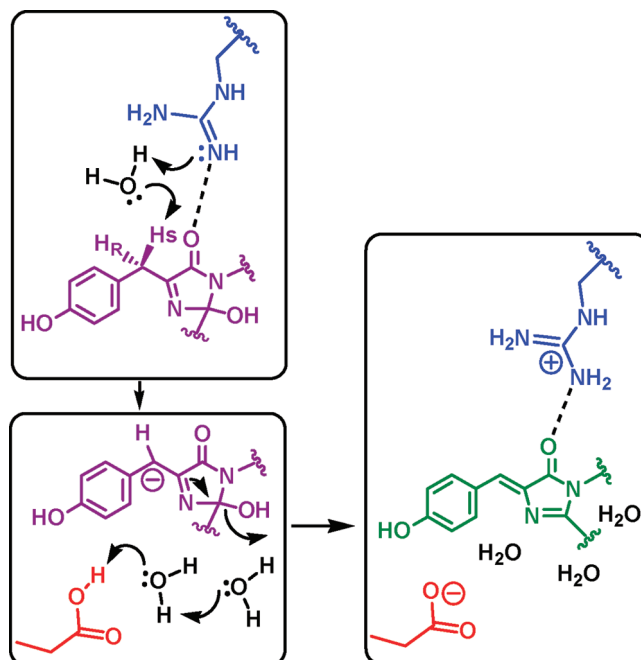
The substantial pH-dependence of k_{obs} (Figure 3) leads to extraction of an apparent pK_a value of 9.46 (\pm 0.07) for dihydro protein, and 9.41 (\pm 0.09) for dideutero protein (Table 2). These data suggest that a protein group with high pK_a value functions as a general base in proton abstraction, and that the acid dissociation constant of the titrating catalyst is not affected by the deuterium label. The large deuterium isotope effect supports a model in which proton transfer occurs in the transition state of the rate-limiting event, suggesting that Arg96 may be the general base (see Discussion).

Incorporation Efficiency of C β -Dideuterated Tyrosine. To quantize the level of biosynthetic incorporation of the deuterium label, we first determined the masses of proteolytic peptides by MALDI. Deuterated protein derived from inclusion bodies was digested with trypsin, and peptides were separated by reverse-phase HPLC. Masses of eluting peptides were matched with calculated masses obtained from theoretical tryptic digests. The HPLC fraction eluting at 51–52% ACN contained three peptides, each bearing one tyrosine residue. Two of these peptides included the chromophore-forming residues 65–67, and one was derived from the C-terminus (peptides 53–73, 46–73, and 216–238). The monoisotopic masses of principal peaks obtained from the MALDI spectra indicated a mass increase of 2 ± 0.4 Da compared to the calculated masses in the absence of isotopic enrichment (see Supplementary Table 1). To better quantize the percent enrichment, the three peptides were subjected to nano-HPLC ion trap mass spectrometry (see Supplementary Table 2). For each peptide, the mass was shifted by 2.0 Da relative to the theoretical mass, and the observed isotope distribution agreed exactly with the expected distribution calculated for the D2 (dideutero) version. Peak area determination of the second (major) isotope peak for the D2 and H2 (dihydro) peptides was utilized to estimate an incorporation efficiency of $96.6 \pm 3.1\%$. This value constitutes a lower limit, as the H2 peptide peak area was found to be near baseline noise. Therefore, the data are consistent with essentially complete incorporation of the deuterium label, and 100% incorporation efficiency of C β -dideuterated tyrosine residues was assumed for all kinetic calculations.

DISCUSSION

Proton Transfer Steps in GFP Maturation, and Models for Exocyclic C–H Bond Scission. The overall GFP maturation reaction requires several proton transfer steps, such as

Scheme 3: Proposed Mechanism for Stereospecific Deprotonation of C β 66 by General Base Catalysis Involving N η_2 of Arg96^a



^a Carbanion formation is thought to trigger dehydration of the heterocycle, a process that may be facilitated by the Glu222 carboxylic acid as a proton donor.

deprotonation of the Gly67 amide nitrogen during ring closure, and deprotonation of the Tyr66 α -carbon to promote enolization of the heterocycle (Scheme 1) (6). Neither the cyclization reaction nor the associated proton transfer steps limit the overall rate of chromophore biosynthesis, as they appear to proceed within the first few minutes of the reaction (Table 1) (11). However, to produce the mature chromophore, a carbon–hydrogen bond must be broken at the β -carbon of Tyr66 (C β 66), as this carbon center changes its hybridization state from sp^3 to sp^2 upon maturation. A reasonable mechanism may involve homolytic bond cleavage, which would place an unpaired electron onto the exocyclic carbon center upon H-atom transfer to an activated oxygen species (28). Nonetheless, data presented here support heterolytic bond scission with proton abstraction from the exocyclic carbon center by a nearby base. This event could be mediated by a peroxy anion species generated upon reduction of dioxygen (9, 14). However, the kinetic data demonstrate that deprotonation of C β 66 occurs after the release of H₂O₂ from the active site (Scheme 3). Evidently, the nascent carbanion triggers a series of bond rearrangements that result in the expulsion of the hydroxyl leaving group from the heterocycle. C β 66 may be activated by the electron withdrawing properties of the imidazolinone ring, in combination with π -orbital delocalization over the phe-

nolic group. Glu222 may assist in this process by donating a proton to the hydroxyl group attached to C65, making use of a proton relay system that involves two ordered water molecules in highly conserved positions (9, 10) (Figure 3).

Chromophore Biosynthesis Proceeds via Oxidation of the Five-Membered Ring to the Cyclic Imine. Results presented here provide strong support for a mechanism in which the hybridization state of the bridging C β 66 is not affected by the reduction of molecular oxygen. Specifically, H₂O₂ evolution is not accompanied by a detectable deuterium isotope effect for C–H bond cleavage at C β 66 (Table 1). Based on these data, protein oxidation appears to solely involve the heterocyclic functionality (9, 14). The absence of a KIE on oxidation confirms a mechanistic scheme in which the protein is first oxidized to yield the cyclic imine form (Scheme 1), as originally proposed based on the X-ray structure of the Y66L variant (9). In support of this idea, we have previously identified a –2 Da intermediate generated during mGFPsol maturation (11). Here, we present ultrahigh resolution mass spectral evidence that a –2 Da intermediate of the E222Q variant can be trapped by flash freezing rapidly purified protein. The isotopically resolved mass spectrum of full-length E222Q indicates that the majority of the immature protein population has lost the mass equivalent of two hydrogen atoms (Figure 2).

The Final Maturation Process Consists of Rate-Limiting Proton Abstraction from the Bridging Carbon. In this work, we provide direct evidence that bond rearrangement at C β 66 occurs postoxidation, as a 6-fold primary deuterium isotope effect is observed for this step upon deuteration of the exocyclic carbon (Table 1). The kinetic data provide nearly identical KIE values for τ_3 and τ_{int} , determined for mGFPsol and E222Q to be 5.8 and 5.9 respectively (Tables 1 and 2). As these values refer specifically to the third and final step in GFP maturation (Scheme 1), a proton must be removed from the bridging carbon in the transition state of the rate-determining event of this process. Evidently, proton transfer initiates bond rearrangements that lead to full π -orbital conjugation of the phenolic group with the heterocycle, while the hydroxyl adduct is ejected from the five-membered ring (Scheme 3) (1, 9). The pK_a of C β 66 likely resides above physiological pH values, as carbon acids tend to bear pK_a values of 20 or higher (29). For this reason, proton abstraction in the third and final step contributes substantially to the overall rate-retardation observed in chromophore biogenesis.

In the present study, the intrinsic time constant τ_{int} for full E222Q maturation from the oxidized intermediate state is extracted to be 8.4 (± 0.6) min, as determined by curve fitting of the pH–rate profile (Figure 3 and Table 2). The kinetic model entails a base catalyzed reaction, in which the observed rate constant is a function of an intrinsic time constant and an acid dissociation constant. In a previously published study, the time constant was extracted to be 14.3 (± 3.0) min, although fewer data points were collected over the relevant pH range (Table 2) (6). A precise determination of the intrinsic time constant depends critically on the collection of high pH data, however, the pH–rate profile levels off above pH 11, where protein denaturation interferes with kinetic measurements. Taken together, the E222Q data support an intrinsic time constant in the range of 8–14 min, a value comparable in magnitude to that determined for τ_3

in mGFPsol, 15.3 (± 5.5) min (Tables 1 and 2). The similarity in magnitudes suggests that the same protein group is responsible for catalyzing C–H bond cleavage at C β 66 in both variants.

Catalysis Involves a Protein Group with High pK_a Value, Suggesting Arg96 as General Base. The pH–rate profile for E222Q maturation from the oxidized intermediate to the green-fluorescent form supports general base catalysis with a conjugate acid pK_a value of 9.4 (Figure 3). The introduction of C β -dideuterated tyrosine residues does not modify the extracted dissociation constant (Table 2). As a primary isotope effect is observed, proton abstraction must occur in the transition state of this step, requiring that the titrating group is positioned in close vicinity to the carbon acid. Inspection of high-resolution crystal structures such as that of GFP-S65T (12) allows for the identification of only one protein-based group fitting these requirements, the guanidine/guanidinium group of Arg96. Therefore, we propose here that proton abstraction from the exocyclic carbon is facilitated by Arg96. We note that the kinetically equivalent mechanism, whereby catalysis is carried out by a hydroxide ion, may also be consistent with the data (30), because the slope of the curve fitted to a log–log plot is close to unity between pH 7.5 and 8.5. However, the curve levels off above pH 9.0, and the extracted pK_a of 9.4 is judged to be too low to support hydroxide ion catalysis (pK_a of water ~ 16).

Arg96 appears well positioned to activate a water molecule for proton abstraction from C β 66 (Scheme 3). One face of the mature chromophore typically entails a series of hydrophobic contacts, such as Phe165 in GFP or Pro63 in Anthozoa fluorescent proteins (31). Positioned near this face is the guanidinium group of Arg96, with each of its terminal nitrogen atoms N η_1 and N η_2 involved in two hydrogen bonding interactions. These include the highly conserved interaction between N η_2 and the carbonyl oxygen of the chromophore's imidazolinone ring, as well as a hydrogen bond between N η_1 and the carbonyl oxygen of Thr62. Geometric considerations render a solvent-mediated interaction of N η_1 with C β 66 unlikely (distance 4.9 Å in S65T), as solvent access is blocked by the carbonyl oxygen of Thr62. In contrast, N η_2 is found within van der Waals distance to C β 66, with a separation of 3.9 Å in S65T (12). Strikingly, in the postoxidation intermediate trapped in Y66L, this separation is further reduced to 3.7 Å (9), and the hydrogen bonding geometry observed for N η_2 is more in line with a tetrahedral amino group, a geometry that would stabilize the neutral guanidine form (see below). Therefore, we propose that proton transfer may be mediated by a water molecule serving as a proton shuttle to N η_2 . The structural arrangement predicts that the *pro-S* proton may be abstracted from the cyclic imine intermediate in a stereospecific manner (Scheme 3). Although an appropriately positioned solvent molecule is not observed crystallographically, several ordered waters are found in the vicinity of the β -methylene bridge, suggesting that diffusion of trapped solvent may aid in full maturation. In support of this notion, the crystal structure of the chemically reduced α -enolate form of Y66H demonstrates that access of solvent to C β 66 is aided by the nonplanar geometry of the nascent chromophore, as three ordered water molecules are observed within 3.3, 3.7, and 3.9 Å of C β 66 (10) (Figure 4). In addition, previous work has provided evidence that small molecules such as iodide

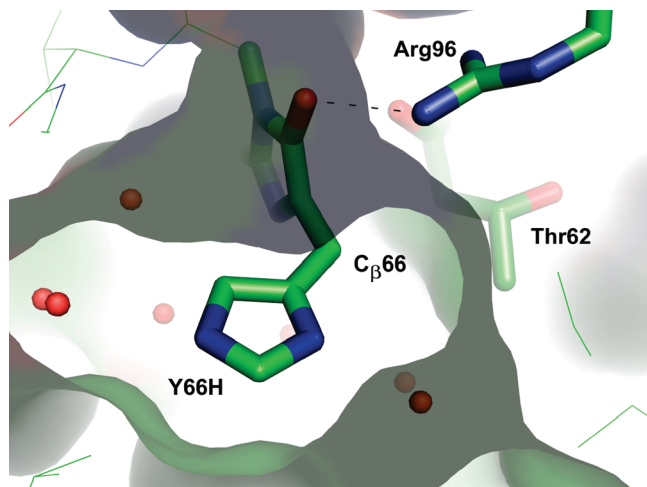


FIGURE 4: Structural arrangement of Arg96 $N_{\eta 2}$ in relation to $C_{\beta 66}$, as exemplified by the X-ray structure of the chemically reduced Y66H variant (10). The image was generated from pdb ID code 2fwq (10) using the program PyMol (36). The molecular surface of the chromophore binding pocket is also shown.

are able to access $C_{\beta 66}$ and Arg96 (16), in line with the observation that formate catalyzes deprotonation at the exocyclic carbon in the Y66L variant (14).

pK_a Depression of Arg96 during GFP Maturation? Typically, arginine residues titrate with a pK_a value of about 12.5 in aqueous solution, suggesting that the acid dissociation constant of Arg96 may be depressed by 3 pK_a units in E222Q. Substantial pK_a perturbations may occur upon protein folding due to the enclosure of the guanidinium group by the protein matrix, and sequestration away from bulk solvent. Based on theoretical calculations, the guanidine/guanidinium group is thought to adopt a series of low-energy conformations that may provide a variety of hydrogen bonding geometries, as energetically favored structures include both pyramidal and planar $N_{\eta 1}$ and $N_{\eta 2}$ centers (29). Rapid conformational interconversions could facilitate transient interactions that differ from those observed in mature GFPs, where Arg96 is almost certainly cationic. Distortion toward a pyramidal amino geometry, which is favored by the neutral form, could be enforced by appropriately positioned hydrogen bonding partners (29). Most importantly, the protonation equilibria in play while maturation is in progress may not reflect the final pK_a value of Arg96 in the green-fluorescent state. The polarizability of the mature chromophore's π -system may lead to electrostatic stabilization of the cationic form by allowing for increased negative charge density at the imidazolinone oxygen.

Previously, we have demonstrated that, in intact GFPs, the total in vitro processing time is essentially invariant over the pH range of 8–10, where overall maturation proceeds with a time constant of 1.0 (± 0.1) h (6). These results suggest that the pK_a of Arg96 may drop below 8 while the biosynthetic reaction is in progress. However, it is difficult to assess the specific pH-dependence of step 3 in relation to all other steps, as the overall process includes protein folding in addition to the chemical steps. Remarkably, at pH 7.0, the time constant for EGFP maturation was observed to be 1.3 (± 0.6) h, suggesting slower chromophore biosynthesis at lower pH values (6). We speculate that these data reflect a rate reduction in step 3 upon protonation of Arg96, however, the measurement error prohibits a detailed mecha-

nistic interpretation. Compared to intact GFPs, the pK_a value of Arg96 may be modulated by the E222Q protein environment such that its value is elevated to 9.4.

Inefficient Maturation in Arg96 Mutants of GFP. Arg96 has been demonstrated to fulfill catalytic functions in chromophore biogenesis, as methionine and alanine substitutions are known to slow down maturation substantially. In particular, the R96M variant exhibits a 5000-fold rate reduction in acquisition of green fluorescence (6). Evidently, this variant remains trapped in the precyclization state, complicating an analysis of the specific involvement of arginine in the postoxidation step. Similar to R96M, the side chain truncation variant R96A exhibits an extremely slow main-chain cyclization reaction as well (5). Although it may be feasible to test the catalytic role of Arg96 by subjecting R96A to chemical rescue experiments involving guanidine (32), this approach may be complicated by the non-native packing arrangements observed for Tyr66 in the truncation variant (5).

Importantly, the activity of Arg96 can be compensated for by a lysine residue (7). The R96K variant is able to mature to the green fluorescent state, albeit with reduced efficiency, as the ϵ -amino group is not directly hydrogen bonded to the chromophore's carbonyl oxygen (8). Although the R96K variant generates visible fluorescence upon overnight expression, only 30 to 50% of the protein population bears a mature chromophore. In addition, the protein fold is destabilized, as indicated by increased trypsin sensitivity (7).

Arginine as a General Base in Enzymatic Reactions. Several enzymatic reactions have been described in which arginine residues are thought to play the role of acid–base catalyst in proton transfer reactions. Likely the best-studied example is the hydrolysis reaction of inosine monophosphate dehydrogenase (IMPDH) (30), where it has been demonstrated that guanidine derivatives can rescue the activity of the active-site R418A mutant in a pH-dependent manner (32). In this system, a lysine residue is able to replace the function of arginine with only a 10-fold rate reduction (33). The IMPDH hydrolysis reaction presents an important analogy to the mechanism proposed here for GFP, as a water molecule appears to be activated by the catalytic Arg418 in the transition state of proton transfer. The appropriately placed solvent molecule was only observable in an inhibitor complex (34), indicative of intermediate states that exhibit altered solvent positions. Surprisingly, kinetic data have suggested that the general base titrates with a pK_a of about 8, in line with a substantially perturbed arginine pK_a in the enzyme's active site (33). Deprotonation of a carbon acid by an arginine residue is further exemplified by the pectate/pectin lyases and by fumarate reductase (29). In pectate lyase, a catalytic arginine residue has been implicated in direct proton abstraction, and theoretical calculations have provided an estimate of 9.5 for its pK_a value (35). However, to date, no direct experimental measurements of pK_a values are available for any of the proposed active site arginines in these enzyme systems.

A Proposed Role for Glu222 in Dehydration. Our data do not support a direct role for Glu222 in proton abstraction from the exocyclic $C_{\beta 66}$. In the E222Q variant, the rate of the postoxidation step is reduced by a factor of 7 near physiological pH values (6), an impairment that seems

insufficient to suggest a basic function for Glu222 in the fully rate-determining process (29). Also, the position of the Glu222 carboxyl group in relation to the nascent β -methylene bridge provides an inappropriate geometry for proton removal. A more reasonable role for Glu222 would be that of proton donor to the hydroxyl leaving group, a function that would facilitate the ejection of water from the heterocycle by making use of a proton relay involving two water molecules (Scheme 3).

CONCLUSIONS

In this work, we provide several lines of evidence that GFP maturation involves base-catalyzed proton abstraction at C β 66, and that this process succeeds heterocyclic oxidation. Ultrahigh resolution mass spectrometric data on E222Q demonstrate that the observed acid dissociation constant of 9.4 cannot be ascribed to the Gly67 amide nitrogen as originally proposed (6), as the majority of the protein population is already oxidized to the cyclic imine form and therefore postcyclization. Instead, we propose that this pK_a value may be ascribed to an active site arginine residue, Arg96. To date, Arg96 has been suggested to play a variety of roles in chromophore biosynthesis, ranging from conformational backbone preorganization to electrostatic catalysis, with effects such as pK_a depression of main chain atoms and stabilization of activated oxygen species. The interpretation presented here adds the novel function of general base catalyst for this versatile and indispensable residue found in all fluorescent proteins.

ACKNOWLEDGMENT

We thank Linda Brei at the Proteomics Core Facility of the University of Arizona for the collection of mass spectrometric data to estimate the efficiency of deuterium incorporation.

SUPPORTING INFORMATION AVAILABLE

MALDI and ESI data on tryptic peptides with and without isotope enrichment; LTQ-FTMS spectrum of full-length mature E222Q protein. This material is available free of charge via the Internet at <http://pubs.acs.org>.

REFERENCES

- Wachter, R. M. (2007) Chromogenic cross-link formation in green fluorescent protein. *Acc. Chem. Res.* 40, 120–127.
- Wachter, R. M. (2006) The family of GFP-like proteins: Structure, function, photophysics and biosensor applications. *Photochem. Photobiol.* 82, 339–344.
- Tsien, R. Y. (1998) The Green Fluorescent Protein. *Annu. Rev. Biochem.* 67, 509–544.
- Remington, S. J. (2006) Fluorescent proteins: maturation, photochemistry and photophysics. *Curr. Opin. Struct. Biol.* 16, 1–8.
- Barondeau, D. P., Putnam, C. D., Kassmann, C. J., Tainer, J. A., and Getzoff, E. D. (2003) Mechanism and energetics of green fluorescent protein chromophore synthesis revealed by trapped intermediate structures. *Proc. Natl. Acad. Sci. U.S.A.* 100, 12111–12116.
- Sniegowski, J. A., Lappe, J. W., Patel, H. N., Huffman, H. A., and Wachter, R. M. (2005) Base catalysis of chromophore formation in Arg96 and Glu222 variants of green fluorescent protein. *J. Biol. Chem.* 280, 26248–26255.
- Sniegowski, J. A., Phail, M. E., and Wachter, R. M. (2005) Maturation efficiency, trypsin sensitivity, and optical properties of Arg96, Glu222, and Gly67 variants of green fluorescent protein. *Biochem. Biophys. Res. Commun.* 332, 657–663.
- Wood, T. I., Barondeau, D. P., Hitomi, C., Kassmann, C. J., Tainer, J. A., and Getzoff, E. D. (2005) Defining the role of arginine 96 in green fluorescent protein fluorophore biosynthesis. *Biochemistry* 44, 16211–16220.
- Rosenow, M. A., Huffman, H. A., Phail, M. E., and Wachter, R. M. (2004) The crystal structure of the Y66L variant of green fluorescent protein supports a cyclization-oxidation-dehydration mechanism for chromophore maturation. *Biochemistry* 43, 4464–4472.
- Barondeau, D. P., Tainer, J. A., and Getzoff, E. D. (2006) Structural evidence for an enolate intermediate in GFP fluorophore biosynthesis. *J. Am. Chem. Soc.* 128, 3166–3168.
- Zhang, L., Patel, H. N., Lappe, J. W., and Wachter, R. M. (2006) Reaction progress of chromophore biogenesis in green fluorescent protein. *J. Am. Chem. Soc.* 128, 4766–4772.
- Ormo, M., Cubitt, A. B., Kallio, K., Gross, L. A., Tsien, R. Y., and Remington, S. J. (1996) Crystal structure of the *Aequorea victoria* Green Fluorescent Protein. *Science* 273, 1392–1395.
- Barondeau, D. P., Kassmann, C. J., Tainer, J. A., and Getzoff, E. D. (2005) Understanding GFP chromophore biosynthesis: Controlling Backbone cyclization and modifying post-translational chemistry. *Biochemistry* 44, 1960–1970.
- Rosenow, M. A., Patel, H. N., and Wachter, R. M. (2005) 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* 44, 8303–8311.
- Barondeau, D. P., Kassmann, C. J., Tainer, J. A., and Getzoff, E. D. (2006) Understanding GFP posttranslational chemistry: Structures of designed variants that achieve backbone fragmentation, hydrolysis, and decarboxylation. *J. Am. Chem. Soc.* 128, 4685–4693.
- Wachter, R. M., Yarbrough, D., Kallio, K., and Remington, S. J. (2000) Crystallographic and energetic analysis of binding of selected anions to the yellow variants of green fluorescent protein. *J. Mol. Biol.* 301, 159–173.
- Wachter, R. M., and Remington, S. J. (1999) Sensitivity of the yellow variant of green fluorescent protein to halides and nitrate. *Curr. Biol.* 9, R628–R629.
- Matz, M. V., Fradkov, A. F., Labas, Y. A., Savitsky, A. P., Zaraisky, A. G., Markelov, M. L., and Lukyanov, S. A. (1999) Fluorescent proteins from nonbioluminescent Anthozoa species. *Nat. Biotechnol.* 17, 969–973.
- Remington, S. J., Wachter, R. M., Yarbrough, D. K., Branchaud, B. P., Anderson, D. C., Kallio, K., and Lukyanov, K. A. (2005) zFP538, a yellow fluorescent protein from *Zoanthus*, contains a novel three-ring chromophore. *Biochemistry* 44, 202–212.
- McIntosh, L. P., and Dahlquist, F. W. (1990) Biosynthetic incorporation of ¹⁵N and ¹³C for assignment and interpretation of nuclear magnetic resonance spectra of proteins. *Q. Rev. Biophys.* 23, 1–38.
- Zhou, M., Diwu, Z., Panchuk-Voloshina, N., and Haugland, R. P. (1997) A stable nonfluorescent derivative of resorufin for the fluorometric determination of trace hydrogen peroxide: Applications in detecting the activity of phagocyte NADPH oxidase and other oxidases. *Anal. Biochem.* 253, 162–168.
- Kuzmic, P. (1996) Program DYNAFIT for the analysis of enzyme kinetic data: Application to HIV protease. *Anal. Biochem.* 237, 260–273.
- Steiner, T., Hess, P., Bae, J. H., Wilsch, B., Moroder, L., and Budisa, N. (2008) Synthetic biology of proteins: Tuning GFPs folding and stability with fluoroproline. *PLoS ONE* 3, e1680.
- Heim, R., Cubitt, A. B., and Tsien, R. Y. (1995) Improved green fluorescence. *Nature* 373, 663–664.
- Cormack, B. P., Valdivia, R. H., and Falkow, S. (1996) FACS-optimized mutants of the Green Fluorescent Protein (GFP). *Gene* 173, 33–38.
- Cramer, A., Whitehorn, E. A., Tate, E., and Stemmer, W. P. C. (1996) Improved green fluorescent protein by molecular evolution using DNA shuffling. *Nat. Biotechnol.* 14, 315–319.
- Zacharias, D. A., Violin, J. D., Newton, A. C., and Tsien, R. Y. (2002) Partitioning of lipid-modified monomeric GFPs into membrane microdomains of live cells. *Science* 296, 913–916.
- Walsh, C. T. (2006) *Posttranslational modification of proteins: Expanding nature's inventory*, Roberts and Company Publishers, Greenwood Village, CO.
- Schlippe, Y., V. G., and Hedstrom, L. (2005) A twisted base? The role of arginine in enzyme-catalyzed proton abstractions. *Arch. Biochem. Biophys.* 433, 266–278.

30. Hedstrom, L., and Gan, L. (2006) IMP dehydrogenase: structural schizophrenia and an unusual base. *Curr. Opin. Chem. Biol.* 10, 520–525.
31. Malo, G. D., Wang, M., Wu, D., Stelling, A. L., Tonge, P. J., and Wachter, R. M. (2008) Crystal structure and Raman studies of dsFP483, a cyan fluorescent protein from *Discosoma striata*. *J. Mol. Biol.* 378, 869–884.
32. Schlippe, Y., V. G., and Hedstrom, L. (2005) Guanidine derivatives rescue the Arg418Ala mutation of *Tritrichomonas foetus* IMP Dehydrogenase. *Biochemistry* 44, 16695–16700.
33. Schlippe, Y., V. G., and Hedstrom, L. (2005) Is Arg418 the catalytic base required for the hydrolysis step in the IMP dehydrogenase reaction? *Biochemistry* 44, 11700–11707.
34. Gan, L., Seyedsayamdost, M. R., Shuto, S., Matsuda, A., Petsko, G. A., and Hedstrom, L. (2003) The immunosuppressive agent mizoribine monophosphate forms a transition state analogue complex with inosine monophosphate dehydrogenase. *Biochemistry* 42, 857–863.
35. Scavetta, R. D., Herron, S. R., Hotchkiss, A. T., Kita, N., Keen, N. T., Benen, J. A. E., Kester, H. C. M., Visser, J., and Jurnak, F. (1999) Structure of a plant cell wall fragment complexed to pectate lyase C. *Plant Cell* 11, 1081–1092.
36. DeLano, W. L. The PyMOL molecular graphics system. (2002)World Wide Web<http://www.pymol.org>.

BI8007164