Published in final edited form as:

Biochemistry. 2017 February 07; 56(5): 736–747. doi:10.1021/acs.biochem.6b00800.

Mispacking and the Fitness Landscape of the Green Fluorescent Protein Chromophore Milieu

Shounak Banerjee[†], Christian D. Schenkelberg[†], Thomas B. Jordan[†], Julia M. Reimertz[†], Emily E. Crone[§], Donna E. Crone[†], and Christopher Bystroff^{*,†,‡}

[†]Department of Biological Sciences, Center for Biotechnology and Interdisciplinary Studies, Rensselaer Polytechnic Institute, Troy, New York 12180, United States

[‡]Department of Computer Science, Rensselaer Polytechnic Institute, Troy, New York 12180, United States

[§]Department of Biology, Colgate University, Hamilton, New York 13346, United States

Abstract

The autocatalytic maturation of the chromophore in green fluorescent protein (GFP) was thought to require the precise positioning of the side chains surrounding it in the core of the protein, many of which are strongly conserved among homologous fluorescent proteins. In this study, we screened for green fluorescence in an exhaustive set of point mutations of seven residues that make up the chromophore microenvironment, excluding R96 and E222 because mutations at these positions have been previously characterized. Contrary to expectations, nearly all amino acids were tolerated at all seven positions. Only four point mutations knocked out fluorescence entirely. However, chromophore maturation was found to be slower and/or fluorescence reduced in several cases. Selected combinations of mutations showed nonadditive effects, including cooperativity and rescue. The results provide guidelines for the computational engineering of GFPs.

Graphical Abstract

*Corresponding Author: Department of Biological Sciences, Rensselaer Polytechnic Institute, 110 8th St., Troy, NY 12180. bystrc@rpi.edu.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.biochem.6b00800.

Saturation mutagenesis of E222 (Table S1), spectroscopic properties of selected mutants (Table S2), benchmark for expert side chain modeling of point mutants (Table S3), protein expression (Figure S1), interpolating between the crystallized states (Figure S2), Rosetta predictions of the ΔG of mutation (Figure S3), slow chromophore maturation in dark mutant Q94D (Figure S4), models of mutations that significantly impaired fluorescent function (Figure S5), Supplemental Experimental Procedures, image acquisition parameters for *in vivo* relative fluorescence, and supplemental references (PDF)

Chromophore maturation mechanism (MPG)

The authors declare no competing financial interest.

The applications of the green fluorescent protein (GFP) and its variants are expanding beyond the visualization of gene expression and protein localization in cells.¹ GFP has been engineered to signal successful protein folding² and to detect protein–protein interactions.^{3,4} It has also been implemented as an intracellular pH sensor,⁵ a calcium sensor,⁶ and a biosensor for influenza virus.⁷ Alterations to the chemical structure and environment of its chromophore result in shifted excitation and emission maxima, such as those observed in red fluorescent protein variants.⁸ However, our ability to engineer GFP is dependent on our knowledge of the sequence determinants of its fluorescent function.

The GFP chromophore is autocatalytically generated by a three-step reaction called chromophore maturation (CM) (Video S1).^{9,10} Maturation of the chromophore is activated by localized charged side chains in the protein core.^{11,12} R96, located on strand 4, stabilizes the enol tautomeric form of the Y66–G67 peptide bond, acidifying the G67 nitrogen and priming it for nucleophilic attack on the S65 carbonyl carbon, cyclizing the backbone.¹² R96 is highly conserved among GFP homologues, underscoring its importance in this role. However, debilitating mutation R96A or R96M can be rescued by Q183R, which places a buried positive charge near the Y66 carbonyl oxygen.¹² The rate-limiting step in maturation is oxidative elimination across C α and C β of Y66¹⁰ with a half-life of ~4 h. The rate of oxidation depends on the permeability of the β barrel to molecular oxygen near strand 7.¹³ Chromophore maturation is completed by dehydration of the S65 peptide bond, leaving a conjugated p-hydroxybenzylidene imidazoli-none ring system that fluoresces green in blue light.

The mature GFP chromophore glows when the protein is folded but not when it is unfolded. Several alternative mechanisms have been proposed for how the folded protein blocks nonradiative decay of the photoexcited state of the chromophore and achieves a high quantum yield of fluorescence (QY). Quantum mechanics simulations find that the χ_2 “one-bond-flip” (OBF) rotation of the Y66 side chain in the photoexcited state is the lowest-energy route to nonradiative decay,¹⁴ and this rotation is sterically hindered by side chains in the chromophore microenvironment. The asynchronous “hula-twist” rotation of χ_1 and χ_2 to a conical intersection conformation^{14–16} has been proposed and, unlike the OBF, seems possible within the confines of the folded state, but this rotation has an intrinsic energetic

barrier and therefore probably does not account for a significant fraction of the nonradiative decay of the excited state. Also, the planarity of the chromophore has been used to explain QY differences between GFP homologues,¹⁷ reasoning that the QY increases with the improved alignment of the imidazolinone and hydroxybenzylidene rings, but we see no such differences in chromophore planarity between high-resolution crystal structures of GFP variants that have significant QY differences. *cis-trans* isomerization of the protein backbone was implicated in the lowering of the QY upon the discovery of a nonfluorescent “trapped” state that occurs upon refolding,¹⁸ a state that is eliminated upon mutational redesign of the loop containing a *cis*-peptide at P89.¹⁹ However, QY differences are seen between two variants that both lack the *cis*-peptide (S. Banerjee, unpublished data). Thus, our working model for QY differences focuses on the OBF hypothesis and its dependence on steric interactions in the immediate environment of the chromophore.

This understanding of CM and QY lays the groundwork for our interpretation of the results of the point mutations described in this work.

In this work, we survey seven positions surrounding the GFP chromophore by exhaustive mutation, asking which amino acid side chains permit the retention of GFP function and which knock out that function. The observed relative fluorescence (RF) is the product of the chromophore maturation efficiency, defined as the fraction of proteins that contain the mature chromophore versus any other chemical state of the active tripeptide unit, and the QY, defined as the emission intensity per mole of the mature chromophore for a fixed excitation intensity, relative to wild type (unsplit sOPT-GFP,³ hereafter termed OPT-GFP). For mutations that knocked out, knocked down, or increased the RF, we asked two questions. First, how could that mutation affect the CM process, given our current understanding of it? Second, how could that mutation affect the QY, basing our analysis on the OBF hypothesis? In selected cases, we were able to experimentally determine which of the two processes, CM or QY, was affected by the mutation, and in the remaining cases, we provide plausible explanations.

Note that R96 and T203 were not surveyed for mutational tolerance in this work because they have been extensively studied elsewhere.^{12,20,21} R96 mutations eliminate or slow chromophore maturation. Mutations at T203 have been shown to alter the fluorescence spectrum of the chromophore.²⁰ Nor have we covered all possible requirements of fluorescent function on positions in the GFP sequence, choosing to focus instead on the immediate environment of the chromophore. It is very possible that point mutations elsewhere in the GFP structure exist that will knock out, knock down, or enhance fluorescence. However, the motivation for our work was the hope that a better understanding of the sequence tolerance at positions around the chromophore would inform or allow robust computational design of GFP variants for novel functions.

E222 point mutations have also been extensively studied in previous work. Saturation mutagenesis of E222 was performed as a control, and the results are in qualitative agreement with the findings of Nakano and co-workers.²² Though, like R96, E222 is highly conserved, we found several mutants, including E222K, that displayed wild-type-like fluorescence.

The results described in this paper paint a picture of the chromophore microenvironment that is robust to mutation with certain exceptions. Surprisingly, we found conservative mutations that knocked out fluorescence, either by severely impeding CM or by drastically reducing the QY. We also found nonconservative mutations that were well tolerated. Equally surprising were several mutations that enhanced the RF. Herein, we attempt to explain these results in the context of our current understanding of the structural determinants of GFP's fluorescent function.

MATERIALS AND METHODS

All residue numbers in this work are based on the numbering in the structure of superfolder GFP [Protein Data Bank (PDB) entry 2B3P].

Saturation Mutagenesis.

OPT-GFP³ was used as a starting point for mutations. Saturation mutagenesis was performed on seven residues within 3.8 Å of any atom on the *p*-hydroxybenzylidene imidazolinone chromophore. Another independent round of selection was performed, on the basis of the amount of contact area with the chromophore that residues had.

Two methods for saturation mutagenesis were used. (1) Assembly polymerase chain reaction (PCR) was performed with two degenerate codons (NBH and NHB) using the Phusion polymerase (36 codons each with redundancy; encodes all 20 amino acids and stop codons). (2) Inverse PCR-directed mutagenesis was performed using the Phusion polymerase and three degenerate codons (NDT, VHG, and TGG) described in the literature (Kille et al., 2013). This method reduces screening effort by eliminating synonymous codons for all amino acids except valine. Even for valine, two of the four possible codons are encoded. Clones carrying the NDT library, VHG library, and TGG (tryptophan) were transformed separately, to prevent screening bias introduced by errors in mixture ratios. The effects of mutating all positions to tryptophan were captured.

Cloning and Expression.

All constructs were expressed in Acella cells (Edgebio) using the pET28a+ expression vector with an N-terminal six-His tag. The cells were cultured on nitrocellulose membranes atop LB-Agar plates with kanamycin. Gene expression was induced by transferring the membrane to LB-Agar plates with kanamycin and 0.1 mM isopropyl β-D-1-thiogalactopyranoside, at room temperature. Relative expression levels were analyzed by polyacrylamide gel electrophoresis (PAGE) of cell lysates for selected mutants. Gels were imaged by transillumination using a Biodocit Gel documentation system. ImageJ was used to integrate bands in images of PAGE gels. Lanes were densitometrized, and peaks (GFP and a reference peak) were integrated from the peak base as shown in Figure S1c. Relative expression levels were reported as the integrated density of the GFP peak over the reference peak.

Plate Imaging.

Induced colonies were incubated for a total of 21 h. Images of the plates, illuminated by a blue light source (DarkReader, Clare Chemical), were acquired for the first 5 h, in 1 h intervals. See Supplemental Experimental Procedures for details. An ultraviolet (UV) shield was used to ensure that cellular DNA was not damaged during imaging. Another image was recorded at the end of 21 h. The plates were then transferred to a cold room at 4 °C. ImageJ²³ was used to estimate the fluorescence intensity of glowing colonies, and the results were classified using Excel 2013.

Sequencing.

The DNA sequences of genes expressed in colonies of interest were extracted by colony PCR using the OneTaq DNA polymerase following the standard manufacturer's (NEB) procedure. The PCR products were shipped on 96-well plates to MCLAB for Sanger sequencing.

Protein Purification.

Proteins were purified under nondenaturing conditions using immobilized metal affinity chromatography (IMAC) on Ni-NTA agarose bead columns. Standard protocols specified by Thermo Scientific, the manufacturer of the HisPur columns, were used.

Five milliliters of the eluted protein was double-dialyzed into 10 mM sodium phosphate and 5 mM sodium chloride (pH8.0).

Chromophore Maturation.

The chromophore maturation efficiency for selected representative mutants was measured by taking the ratio of absorbances at wavelengths of 280 and 485 nm in the native state in PBS and correcting for the protein's molar extinction coefficient. Maturation efficiency η is defined by the following equation:

$$\eta = A_{485}/(A_{280}/\epsilon) \quad (1)$$

where ϵ is the molar extinction coefficient for the protein and was determined for each mutant using PROTPARAM.^{24–26} Absorbances were measured using a Nanodrop 2000c spectrophotometer (Thermo Scientific) on affinity-purified solutions of protein at approximately 9 mg/mL.

Fluorescence Spectra and Quantum Yield.

Fluorescence spectra were recorded on a Fluorolog Tau 3A spectrofluorometer after diluting proteins to a final concentration of 0.33 μ M. The 3 mL samples were placed in a 4.5 mL rectangular quartz cuvette. The chromophores were excited with 485 nm light from a monochromator with a 3 nm slit width (bandpass). Emission was monitored at a right angle to the excitation beam, using a monochromator set to 508 nm with a 5 nm slit width to allow sufficient photonic flux to the photomultiplier tube. Corrections for lamp intensity fluctuations were applied dynamically, and only the corrected spectra are shown.

The quantum yield was measured after exchanging proteins into PBS [10 mM sodium phosphate and 5 mM salt (pH 8.0)] and diluting all proteins to have an optical density of 0.5 at a wavelength of 485 nm and an optical path length of 10 mm, as measured by a Shimadzu UV-1650PC spectrophotometer (slit width of 3 nm).

Signal averaging was performed over four scans for all spectral acquisitions.

Circular Dichroism.

Far-UV circular dichroism was measured by placing 400 μL of 10 μM protein in a quartz cuvette with a 2 mm path length and tracking the ellipticity introduced, at wavelengths between 190 and 260 nm. A Jasco 815 CD spectrometer was used for all measurements.

For near-UV circular dichroism studies, 3 mL of proteins at 0.5 mg/mL was placed in a quartz cuvette with a 10 mm path length. Ellipticity was measured at wavelengths between 260 and 320 nm.

For all measurements, the sensitivity was set to 5 mdeg and ellipticities were recorded at wavelength intervals of 1 nm. All readings are the automatically calculated signal averages of four scans

Modeling.

Structures of the mutant positions discussed in the experimental setup were generated computationally using different states of the GFP chromophore maturation process and are shown in Figure S2, including the precyclized precursor (PDB entry 1QXT), the cyclized-only intermediate (PDB entry 1QYQ), and the mature state (PDB entry 2B3P). The precursor structure 1QXT had been crystallized by mutating Arg96 to Ala, which disrupts a critical chromophore interaction, which thereby hinders chromophore maturation and causes the maturation process to occur over a couple of months, rather than a couple of hours. As a result, MOE was used to mutate Ala96 back to Arg to model the known wild-type sequence of GFP in the precursor state. The precursor Tyr rotamer present in 1QXT was also rotated to place the Tyr side chain on the opposite side of the newly modeled Arg96. It was theorized that in 1QXT, the R96A mutation created a void pocket allowing the Tyr rotamer to occupy an unnatural rotameric state. Failing to rotate the Tyr side chain to the opposite side of the modeled Arg would prevent the necessary rotamer transition needed to reach the mature rotameric state. In cyclized intermediate state 1QYQ, the chromophore Tyr residue was mutated to a Gly to trap the chromophore in the intermediate state. Hence, it was necessary to model a hydroxybenzoyl group onto the cyclized intermediate structure, taking care to note correct chirality on the insertion ring carbon atom. The tyrosine rotamer chosen was the rotamer most similar to the mature rotamer.

Additionally, three interpolations between the precursor and the cyclized intermediate and one interpolation between the cyclized intermediate and the mature GFP were generated computationally. The interpolations were performed by taking two adjacent GFP states, fixing the chromophore atoms to the midpoint between their coordinates in the two adjacent states, transferring this interpolated chromophore to the nearest, experimentally determined GFP structure, and minimizing the chromophore within this GFP structure in MOE 2013

(CCG, Montreal, QC) using the Amber94 force field and an implicit solvation model (Figure S2). This process produced chromophore states that roughly model a more fine-grained transition between the crystallized states of the chromophore pathway.

The energy scores of all possible mutants (20 for each single mutant and 400 for each double mutant) used in Figure S3 were calculated using both the fixed backbone single-state (SSD) and multistate (MSD) Rosetta²⁷ protein design protocols and the Talaris2013 scoring function.²⁸ Fixed backbone designs for each point mutation were generated using the mature GFP as a scaffold state. Each point mutation was performed by mutating to the desired amino acid and allowing all other residue side chains to repack around the mutation. Each point mutant design was performed 20 times, and the energy of the point mutant was reported as the lowest energy of these 20 designs.

Exploratory and interpretive modeling of the mutations was performed in MOE and InteractiveRosetta. A benchmark validation of this method was done using point mutants of known structure. The benchmark consists of nine high-resolution structures of T4 lysozyme point mutants,²⁹ each differing from the wild type at a single position, and nine variants of *Escherichia coli* dihydrofolate reductase,³⁰ each with one or two sequence changes. All lysozyme mutations were V to T or A to S, adding hydroxyl groups. MOE's Rotamer Explorer, along with energy minimization and expert knowledge, was used to predict side chain χ_1 rotamers ($t = 180^\circ$; $m = -60^\circ$; $p = 60^\circ$) and sometimes χ_2 and χ_3 rotamers. The results are listed in Table S3. The overall accuracy in χ_1 rotamer prediction was 80%.

Chromophore Maturation Pathway.

High-resolution structures of chromophore maturation intermediates observed for the R96M and Q183E mutants of cycle 3 GFP¹² were used to generate a video depiction of chromophore maturation. Smoothed morphs from one intermediate to the next were generated by linear interpolations between them using all atoms that were present in both crystal structures. The Morph Conformation tool in UCSF Chimera version 1.10.1³¹ was used to generate the interpolated structures. Final rendering of these frames and integration into a movie were performed in PyMol.³² Molecular oxygen was added to illustrate hydride transfer in the oxidation step. The resulting movie is provided as Supporting Information (Video S1).

RESULTS

A total of 134 point variants of GFP were characterized, including the wild type, comprising an exhaustive survey of all 20 amino acids at seven positions surrounding the chromophore (Figure 1). Additionally, two sets of double mutants were created, from which 22 unique sequences were recovered. Colonies carrying variants were classified in vivo as “bright” (>60% RF), “dim” (1%–60% RF), or “dark” (0% RF) green under blue light excitation. Intensities were measured from images with standard exposure conditions (see the Supplemental Experimental Procedures) using ImageJ.²³ Sodium dodecyl sulfate (SDS)–PAGE analysis of whole cell lysates of select point mutants with significantly altered RF showed that protein expression levels varied by at most 2-fold (Figure S1), showing that large differences in colony RF were not due to protein expression levels. Each of the mutants

was found in abundance in the soluble fractions of lysed cells, indicating that the variants are soluble and well-folded. Of the 134 variants, 4 were “dark”, 10 were “dim”, and 120 were “bright” (Table 1).

Sequence tolerance, defined as the diversity of amino acids observed at aligned positions in a multiple-sequence alignment (PFAM ID PF01353),³³ is variable for the positions studied and does not correlate with the distance to the chromophore. For example, F145 and F165 both directly pack against chromophore atoms, yet mutants of F165 are, on average, less fluorescent than those of F145. Surprisingly, all of the positions studied were highly tolerant of mutation.

No change in color was observed for any of the mutants except F145W (Figure 2), which exhibits a distinct yellow color and glows dimly when irradiated with blue light.

Dark Mutants.

Careful examination of all four dark mutants (H181D, Q183P, Q183L, and Q94D) and one very dim mutant (F145W) by far-UV circular dichroism spectroscopy (Figure 3a) showed a smaller positive ellipticity near 200 nm for the Q183P and F145W mutants, indicating a lower degree of hydrogen bonding.^{34,35} A larger negative ellipticity near 217 nm was observed for F145W and three of the dark mutants (H181D, Q183P, and Q183L), suggesting more β -sheet character in these mutants than in the wild-type protein.³⁶ The fourth dark mutant, Q94D, produced a far-UV spectrum that was indistinguishable from that of either the wild type or a bright mutant (F145M).

The near-UV CD ellipticity profile between wavelengths of 260 and 320 nm (Figure 3b) reveals differences in the packing of aromatic side chains.³⁶ All four dark mutants show low profiles in this region, while the bright variants and the very dim variant F145W have similar high profiles. Absorption at 485 nm showed that the wild type, F145W, and all other fluorescent variants have the mature chromophore present but none of the dark mutants have the chromophore. From this information, we can hypothesize that nativelike side chain packing is required for CM.

All four of the dark mutants were found to acquire fluorescence after storage at 4 °C for 3 months. The Q94D mutant was re-expressed and purified, and its chromophore content was measured 1 week, 6 weeks, and 6 months after induction. A >100 day half-life of chromophore maturation was found, assuming single-phase first-order kinetics. An additional >30 day half-life would explain the slow increase in RF that was observed in the wild type and F145W, but not in F145M. This slow increase cannot be due to the previously described half-life of 34 min.³⁷

Quantum Yield of Fluorescence.

The relative QY was measured as the ratio of the integrated fluorescence intensity³⁸ obtained for a mutant to that of the wild type. Glowing mutants differ significantly in their QY. F145W had the lowest value (0.18), while F145M is 17% brighter than the wild type (Figure 4a).

Inspection of the structure suggests energetic reasons for the observed differences. The wild-type phenylalanine is tightly packed into a rare side chain conformation with a χ_1 of -142° , which is 40° from the closest rotamer (trans; $\chi_1 = 180^\circ$) and 80° from the most favorable rotamer (minus; $\chi_1 = -60^\circ$). The F145W mutation exacerbates the steric clashes. In contrast, the methionine adopts a favorable conformation, suggesting that the F145M mutation relieves side chain packing stress.

The colonies of F145W appear to be intensely yellow under normal lab lighting, indicating robust chromophore maturation, but they are weakly fluorescent (RF = 0.18), pointing to a low QY. The F145W mutation creates clashes with the buried residues H169 and V167. In modeling studies, backbone atom shifts that could alleviate the clashes but at the expense of several broken hydrogen bonds were found, implying that poor core packing is inevitable for this mutant. Mispacking and the resultant higher chromophore mobility are the best explanation for the low QY.

Double Mutants.

Two pairs of residues were subjected to double-mutant saturation mutagenesis. Residues 69 and 183 form a “handshake” hydrogen bonding pair. Residues 145 and 165 contact the two *e*-carbons of the chromophore hydroxybenzyl moiety. These side chains can block rotation of that moiety and thereby inhibit quenching. The goal was to find either pairs of high-RF mutations that resulted in a significantly lower RF or pairs of low-RF mutations that resulted in a significantly higher RF. The preliminary results outlined in Table 2 show instances of the first phenotype and a single instance of the second phenotype, discussed further below.

Long-Term Chromophore Maturation and Aggregation.

Variants Q94D, H181D, Q183L, and Q183P showed no CM immediately after purification. However, after storage at 4°C for 3 months, all variants acquired a green color, albeit faint. Insoluble aggregates were also found for these variants. No such aggregation was observed for the wild type, F145M, or F145W, even though the concentration of each of three proteins was higher.

Therefore, a second round of quantitation was performed to estimate the amount of mature chromophore formed in these variants after 3 months. The results are shown and superimposed with spectra acquired immediately after purification (Figure 5). F145M showed no additional CM. The wild type showed a 7% gain in the amount of mature chromophore, while F145W showed a 12% gain (Figure 5).

DISCUSSION

A similar study was performed³⁹ in which 16 residues within $9\text{ \AA}$ of the chromophore were mutated to either cysteine or alanine. However, exhaustive saturation mutagenesis was not performed, leaving open the question of whether their findings can be generalized or dependent on the type of mutation. Our work shows the effect of adding atoms to as well as removing atoms from the protein core, mutations from polar to nonpolar side chains and vice versa, and the effect of introducing charged side chains into the protein core that are not necessarily paired by a countercharge.

Fluorescent function could be affected by a change in the QY, a change in the amount of CM, or both. Alterations in the chromophore's excitation and emission maxima cannot be ruled out. An analysis of the crystal structures of GFP homologues that originated from the jelly fish *Aequorea victoria* (avGFP) reveals nearly optimal packing of all the side chains studied. The two exceptions are F145 and F165, which adopt uncommon rotamers. Because both of these side chains interact with the chromophore's phenolate ring, the interactions are likely to impose significant constraints on chromophore rotational isomerization. Therefore, mutations at these positions were expected to produce significant changes in fluorescent behavior.

All other positions studied have well-conserved polar side chains in all avGFP derivatives. They form hydrogen bonds with chromophore atoms directly or via water molecules, and they act as cornerstones of a buried, solvent-mediated hydrogen bonding network (Figure 1). This is very unusual in the cores of globular proteins. Q69, Q94, Q183, H181, and T62 create a polar environment for the buried charge of the catalytic R96 side chain. Nonpolar interactions (Y92, V112, V150, and F165) surround and seem to hold this polar core together. The positioning of a positive charge near the imidazolinone (Y66) oxygen is crucial for CM, and mutation of Q94 to an aspartic acid or glutamic acid might neutralize the R96 positive charge, severely reducing CM efficiency.¹¹ Mutations to positively charged side chains at position 69, 94, or 183 could have either deleterious effects on folding due to charge repulsion or benefits due to increased rates of backbone cyclization for CM and increased rates of delocalization of the imidazolinone oxygen's electron cloud in the excited state. The latter is predicted to produce the species of the chromophore that has the so-called "B" (477 nm) excitation peak and 508 nm emission peak.⁴⁰

Our understanding of protein backbone dynamics further contributes to our expectations. The plasticity of side chain packing in the core of GFP may be limited by the rigidity of the surrounding β barrel, which is more constrained in the range of its curvature than is an open β sheet. Without the possibility of backbone accommodation, side chain packing is subject to rotameric preferences, but because most side chains share rotameric preferences, the energetic effects will be mostly due to shape and charge complementarity. Note that GFP's extreme kinetic stability is expected to compensate for some degree of mispacking, underpacking, or overpacking in the core.

One of the aims of this work was to test the ability of energy calculations to predict mutations that can impede chromophore maturation and function. Unfortunately, the calculated ΔG of mutation using Rosetta 3.5 was found to be uncorrelated with fluorescence as shown in Figure S3.

Requirements for Chromophore Maturation.

Only four of the 134 mutations were able to severely impede chromophore maturation. Here we discuss the remarkable tolerance of these positions to mutation and speculate about the special character of the four "dark" mutations. This discussion of each of the seven targeted positions is summarized in Table 3.

Single-Site Mutations.

Q69.—The central helix has been shown to fold early on GFP's folding pathway,⁴¹ suggesting that Q69 would be sensitive to mutations. Models of several mutants suggested that these would result in a distortion of the central helix and/or prevention of barrel closure. An example of this is the Q69Y mutation (Figure S4b). However, no knockout mutations were found at this position, nor were there any side chains that produced an increased fluorescence. Q69 sits in the polar core of GFP, making a pair of handshake hydrogen bonds with Q183. It has nonpolar and polar neighbors and has plenty of space to accommodate any side chain, explaining the tolerance of mutation at this position.

Q94.—Residue 94 lies on strand 4, which is adjacent to strands 5 and 9. The contacts between strands 5 and 4 are believed to form early in folding,^{42–44} whereas strands 9 and 4 are thought to come together in the last stages of folding based on the observed kinetic perturbations of a disulfide bridge that crosses that gap.⁴⁵ If so, the protein is mostly native at the time point in folding where the Q94 side chain must find its lowest-energy conformation. Q94 is in contact with R96 and may influence the rotameric state of that arginine, which would in turn affect the ability of GFP to catalyze the backbone cyclization event that creates the chromophore. Surprisingly, Q94 tolerates all amino acids except aspartic acid.

The one dark mutation, Q94D, was explored by modeling. It was found that a D at position 94 has only one possible rotamer. In the presence of the buried charge of D94, the preferred rotameric state of R96 is predicted to be different than the wild-type rotamer and shifts >4 Å from the backbone oxygen of Y66. It is therefore likely that R96 is no longer close enough to the Y66 backbone to stabilize the transition state of the cyclization reaction, explaining the observation that CM is stalled in this mutant.

To follow up on this hypothesis, we looked at chromophore absorbance for purified Q94D mutants 7 days, 6 weeks, and 6 months after expression in bacteria. The 7-day-old protein had barely detectable levels of green fluorescence; however, fluorescence was observable after storage for 6 weeks at 4 °C, and the amount tripled after 6 months at 4 °C. The rate of chromophore maturation for this mutant can be estimated to be at most 0.007 per day ($t_{1/2}$ 100 days). The wild-type GFP has been shown to lead to chromophore maturation with a half-life of 34 min,³⁷ where the slow step is the oxidative elimination at the bridging methylene. However, both OPT-GFP and F145W, but not F145M, displayed a very slow phase of maturation ($t_{1/2}$ 30 days) that cannot be explained by a 34 min half-life.

We attempted to thermally unfold Q94D in 1 M GndHCl along with the wild-type control, but neither showed signs of unfolding after minutes at 95 °C using circular dichroism, giving us no reason to suspect a thermodynamically impaired Q94D structure. A significantly slower maturation rate along with no large changes in thermodynamic stability lends support for the non-native R96 rotamer model described above.

The excitation maximum for this mutant was at 385 nm (Figure 5), corresponding to the protonated chromophore,⁴⁶ whereas all of the other mutants and the wild-type GFP had the more typical 485 nm excitation maximum of the anionic state of the chromophore. The

Q94D mutation adds a negative charge near the imidazolinone ring of the chromophore, stabilizing the protonated state.

F145.—Position 145 is on strand 7. Strand 7 has the fewest hydrogen bonds to its neighboring strands in the barrel, six H-bonds to strand 10, and seven H-bonds to strand 8, compared to 10–12 H-bonds for the other nine barrel seams. In a disulfide engineering experiment, both of the fast phases of folding kinetics were perturbed by placing a covalent linkage from the 6–7 loop on the 8–9 loop, implicating the 7–8 seam as one of the last strand pairings to form.⁴⁵ Also, H/D exchange NMR studies found the greatest amount of backbone amide labeling in and around strand 7.⁴⁴ Thus, we conclude that strand 7 folds last. F145W alone knocks out fluorescence, although not completely. The rationalization of this result is that packing in the GFP core is tight. When the F145W mutation was modeled, the indole side chain adopted the same χ_1 rotamer as the wild-type F but is flipped by 180° in χ_2 between the cyclized-only stage of chromophore maturation and the fully mature stage (Figure S5d). By flipping, it can make a favorable hydrogen bond with the hydroxyl group of the mature chromophore. F145W is sterically constrained by contacts with S205 (T205 in precursor structures) and H169, which hinder the movement of the indole ring, forcing clashes with chromophore atoms. No other rotamer is energetically possible. The steric clashes may explain the low quantum yield of F145W. A S205G mutation and/or a truncation of H169 to a shorter side chain might possibly rescue the quantum yield in the context of this mutation.

Indeed, in GFP homologues that have W145, H169 is changed to L and there is a single-residue deletion at Y143. In these homologues, the backbone is shifted by 3.3 Å to create a large space between strands 7 and 10 in this region. F145W is then easily accommodated, and the protein glows normally.

Interestingly, F145W, in addition to being very dim, is blue-shifted by 2 nm in its emission spectrum (Figure 4b, right inset). Similar spectral shifts have been seen in designed red fluorescent proteins,⁴⁷ and in those cases, the phenomenon was explained by the supposed perturbation of the electron density in the excited state. The modeled F145W has a hydrogen bond to the chromophore hydroxyl group, which may favor an increased negative charge on that oxygen, which could in turn increase the energy gap between the ground state and the excited state, explaining the observed blue-shift.

The F145M mutation had the opposite effect, a 2 nm red-shift in the excitation spectrum. A red-shift indicates a decreased energy difference between the ground state and the excited state. Methionine occupies less space in the partially solvent filled cavity next to the chromophore, leaving room for more water molecules. The increased polarity of the surroundings may have decreased the energy of the excited state relative to the ground state, especially if the excited state electronic structure is more polarized.

The molar absorption at 485 nm (Figure 5) shows that both F145M and F145W have the same CM efficiency as the wild type, within experimental error. However, after incubation for 3 months at 4 °C, both the wild type and F145W variants showed significant increases in

the amount of mature chromophore, whereas no additional chromophore maturation was seen in the F145M variant. The QY did not change with time.

H148.—Histidine 148 also occurs in strand 7, the late-folding β strand. The histidine side chain rotates upon CM to accommodate the chromophore hydroxyl oxygen. A mutation to an aspartate has been previously shown to restore 21% of the excitability of the chromophore with near-UV light (390 nm), the so-called “A” peak in which the hydroxybenzyl oxygen is protonated.⁴⁸ Position 148 tolerates all 20 amino acid side chains in our study, perhaps because it is both a surface residue and mobile.

F165.—Residue 165 lies on strand 8 bordering the chromophore and the essential residue R96. The side chain pocket contains both polar and nonpolar surfaces. All 20 amino acids were found to permit CM and fluorescence, although to differing degrees.

A mutation here is also likely to make contacts with strand 7 more labile as F165 packs in an unusual $p\chi_1$ rotamer (60°) against V150 on strand 7. This is due to a lack of space for any other more energetically favorable rotamer. Thus, residues with weaker preferences for this rotamer are more likely to adopt non-native packing. Previous work has shown that such non-native packing near the chromophore can make it more mobile, thereby reducing the QY.¹⁸

H181.—Position 181 lies on strand 9 near the essential residue R96. An H181D mutation knocked out RF, and H181W reduced it by 64%. All other side chains at this position conserved at least 72% of the RF. The strong resemblance of the far-UV CD spectra of this mutant and that of the wild-type protein assures us that the backbone atoms have not shifted far from their native positions. Formation of a salt bridge between the H181D mutation and R96 would bend the arginine side chain away from the imidazolinone (Y66) carbonyl oxygen, and this would preclude CM, similar to the situation with the Q94D “dark” variant described above. Both of these aspartate placements created salt bridges to the catalytic arginine in our modeling studies. H181 is tightly packed, too tight to accommodate H181W, which produces steric clashes with A179. The clashes may be removed by energy minimization but at the cost of backbone atom shifts that are probably too great to be energetically plausible. The loss of RF in H181W may be attributed to the loss of stability because loosening the core would cause the chromophore to be more flexible and therefore more likely to decay nonradiatively from the excited state.¹⁴ For the same reasons, this mutation may have caused inefficient chromophore maturation. In other studies, the CM reaction depends strongly on the stability of the folded protein (K. Fraser and C. Bystroff, unpublished data).

Q183.—Q183 is conserved as an uncharged polar side chain throughout GFP homologues, and that side chain is a part of the internal hydrogen bonding network. Therefore, it is surprising that mutating 183 to nonpolar residues did not drastically reduce fluorescence, excepting for leucine and proline. Both Q183L and Q183P greatly slow CM while preserving the protein’s folding ability, as shown by their far-UV CD spectra (Figure 3). Q183P shows a drop in positive ellipticity near 200 nm, indicating a loss of secondary structure,^{34,36,49} while Q183L resembles the wild type in this respect. A loss of secondary

structure of Q183P could be due to local overpacking, the loss of a backbone H-bond, and/or a poor backbone ϕ angle ($\phi = -110^\circ$ is favorable for a glutamine but not for a proline). Q183P may be locally distorted to preclude the closure of the 9/4 seam, one of the final steps in the proposed folding pathway.⁴⁵ This would leave the structure in an open-barrel conformation. Open-barrel structures created by leaving out one strand do not form the chromophore, or do so very slowly.^{3,50} Nonetheless, the Q183P mutant is soluble, suggesting that it must be nearly natively folded. The dark phenotype of Q183P is rescued by Q69I, a side chain with a different shape, hinting at packing defects as the reason for the loss of function. The loss of chromophore maturation in Q183L is probably due to packing against R96, blocking the catalytic rotamer. Both dark mutants Q183L and Q183P acquired dim green fluorescence over a period of months. Wild-type GFP refolds to a kinetically trapped, nonfluorescent state,¹⁸ suggesting that these dark mutants may be adopting a similar but much more nearly native trapped state.

E222.—E222 is highly conserved across natural fluorescent protein homologues. Mutations of the latter, to residues that do not possess positively charged side chains, can be rescued.^{12,51–54} E222 has been postulated to have a catalytic role in CM.¹¹ It is also the terminal proton acceptor for the proposed excited state proton transfer (ESPT), a mechanism to explain the 397 nm excitation peak in avGFP, which is still the subject of debate.^{21,55} Note that our template, OPT-GFP, contains the S65T mutation and lacks the 397 nm excitation peak, making ESPT irrelevant as an argument for the conservation of E222. In our study (Table S1), several E222 mutants were found that are brightly fluorescent, including charge reversal mutations E222K and E222R, though E222R exhibits a noticeable drop in fluorescence. Only three mutations, E222A, E222P, and E222W, completely knock out RF. This surprisingly high tolerance to point mutations, with 17 of 20 side chains at position 222 yielding fluorescent proteins, is in rough agreement with the observations of Nakano,²² who found that 11 of 20 amino acids were tolerated in the context of wild-type avGFP. It shows that this residue is not a key structural element and not essential for the fluorescent function, as was previously thought on the basis of its conservation in fluorescent proteins. In the context of protein design efforts such as the recently published work on computationally designed GFP-based biosensors,⁷ this finding adds 222 to the design palette.⁵⁶

Double Mutants.

The single-mutant libraries show that the microenvironment of the GFP chromophore is highly tolerant to point mutations. Of the 133 mutants, 120 were bright and only four were dark. High sequence tolerance has previously been associated with local structural flexibility. It implies that the wild-type residues in the microenvironment are optimally packed and can tolerate the small perturbation represented by a point mutation. Double mutants, on the other hand, are more likely to be dim ($RF < 0.60$) than one would expect by chance; 94% of the single mutants were bright, suggesting that 88% (94% squared) of double mutants would be bright if the perturbing effect were additive and uncorrelated. However, only five of 22 double mutants were bright. We may hypothesize that there is a threshold in energy perturbation beyond which RF is quickly lost, and that this threshold is crossed between single and double mutations. If this is true, then random triple mutants in the GFP core would be predicted to be very rarely bright.

Our double mutations were focused on the pairs of positions 69 and 183 and positions 145 and 165. Q69 and Q183 are in direct contact. F145 and F165 are not in contact. Strong covariance was observed in both the positive sense (super-additive) and the negative sense (rescue). For example, a point mutation that knocked out fluorescence, Q183P, was bright in the context of Q69I, showing negative covariance in the energetic effect of mutation. Superadditivity was observed in the combination of two bright point mutants, Q69K and Q183S, producing a very dim double mutant. Furthermore, all three dim double mutants of the pair of positions 145 and 165 consist of bright point mutants. Nakano²² found superadditivity between contacting positions 65 and 222, with 61% of point mutations tolerated but only 13% of double mutations.

We can envision a general principle that explains the data. It makes sense to us that the highly cooperative nature of side chain packing in the protein core would produce a very rough fitness landscape, where each low-energy state is the solution of a multibody packing problem. The random sampling of sequences during the course of evolution would settle on sequences that tolerate point mutations by re-enforcing the natively packed core over mispacked states, thereby increasing the chances of the survival of function.

Spectroscopic Properties of Selected Mutants.

Perturbations to the milieu, most notably at position 203, have been previously shown to alter spectral properties in fluorescent proteins.^{47,48} These alterations are primarily the result of differential stabilization of the ground state and excited states of the chromophore. Electrostatic interactions are thought to play an important role in maintaining the planarity of the chromophore.⁵⁷ Therefore, mutations of polar residues or residues that interact with these affect properties like the direction and strength of their dipole moments. Such changes are reflected in their spectral properties. For example, a preferential stabilization of the excited state over the ground state causes a red-shift in the emission wavelength. Changes in the vibrational mobility of the chromophore in its ground state are manifested in changes in shape to the excitation profile and fluorescence lifetime⁵⁸ (Table S2).

The far-UV CD spectra of the bright and dark mutants are very similar to that of the wild type. For dark mutants Q183P and F145W, a decreased positive ellipticity in the 200 nm region is observed, suggesting loss of order in core aromatic side chains, most likely by adopting an open-barrel structure. However, they are more folded than peptide-bound LOOGFPs in the presence of their missing strands, yet CRO maturation and/or function is precluded, unlike the case for less folded LOO-GFPs.⁴⁸

The absence of chromophore maturation leaves tyrosine 66 unmodified in dark mutants Q94D, H181D, and Q183P. This extra tyrosine does not appear to contribute to the ellipticity in the 270–280 nm region for the dark mutants Q94D, H181D, and Q183P. Q183L shows a very low chromophore maturation efficiency and a higher positive ellipticity in the region of 260–290 nm. This suggests that Q183L assumes a structure that is closer to native and one that is likely to constrain the extra tyrosine.

H181D and Q183L have almost identical near-UV CD spectra. Along with modeling, this suggests that H181D folds to the native structure and precludes CRO maturation by ion pairing with R96.

Chromophore Maturation and Trapped States.

It is known that GFP samples numerous off-pathway intermediates during folding.^{18,59} Some of these fold to near-native states separated from the native state by high energy barriers. We hypothesize that chromophore maturation is possible only in the native state, because like any enzyme it requires precise positioning of the catalytic moieties. If so, then the wild type, F145M, and F145W sample nearly native mispacked states to different extents. These mispacked states undergo a transition to the native state very slowly, by backtracking through an unfolded state, for instance. The F145M variant, exhibiting early chromophore maturation, logically must sample the native state the most, followed by the wild type and F145W variant, which sample increasing amounts of mispacked state(s). The tryptophan is predicted to pack well into the nascent protein before chromophore maturation, as shown in Figure S5d. However, upon chromophore maturation, the bulky indole side chain must move to avoid collisions. This explains the experimental finding of slow but robust chromophore maturation in this mutant.

CONCLUSIONS

Systematic exploration of the fitness landscape of the chromophore milieu of GFP reveals that point mutations are well tolerated but that double mutations are less so. Circular dichroism analysis shows a correlation between fluorescence and side chain order. Molecular modeling analysis, summarized in Table 3, suggests that side chain mispacking would explain most of the observations, either by creating a trapped, mispacked state that fails to catalyze chromophore maturation or by favoring a less well packed, more flexible chromophore environment that would allow rotational isomerization in the photoexcited state and its concomitant fluorescence quenching. We postulate that chromophore maturation is possible only in the native state, because like any enzyme it requires precise positioning of the catalytic moieties. If this is true, then the wild type and the bright mutants such as F145M favor the natively folded and packed state, while dim and dark mutants such as F145W favor off-pathway states with the side chains mispacked in the core. Because we observed very slow chromophore maturation in some of the dark mutants, and at least one of these mutants (Q94D) is kinetically stable, these nearly native states must undergo a transition to the native state very slowly, possibly only by backtracking through an unfolded state. Superadditivity in the effect of multiple mutations is not surprising when the energetic effect of mutation is to perturb the equilibrium between the unique, perfectly packed native core and any number of possible mispacked states.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Funding

This work was supported by National Institutes of Health Grant R01GM099827 to C.B.

REFERENCES

- (1). Chalfie M, Tu Y, Euskirchen G, Ward WW, and Prasher DC (1994) Green fluorescent protein as a marker for gene expression. *Science* (Washington, DC, U. S.) 263, 802–805.
- (2). Waldo GS, Standish BM, Berendzen J, and Terwilliger TC (1999) Rapid protein-folding assay using green fluorescent protein. *Nat. Biotechnol* 17, 691–695. [PubMed: 10404163]
- (3). Cabantous S, Terwilliger TC, and Waldo GS (2005) Protein tagging and detection with engineered self-assembling fragments of green fluorescent protein. *Nat. Biotechnol* 23, 102–107. [PubMed: 15580262]
- (4). Cabantous S, Nguyen HB, Pedelacq J-D, Koraichi F, Chaudhary A, Ganguly K, Lockard MA, Favre G, Terwilliger TC, and Waldo GS (2013) A New Protein-Protein Interaction Sensor Based on Tripartite Split-GFP Association. *Sci. Rep* 3, 2854. [PubMed: 24092409]
- (5). Bizzarri R, Arcangeli C, Arosio D, Ricci F, Faraci P, Cardarelli F, and Beltram F (2006) Development of a Novel GFP-based Ratiometric Excitation and Emission pH Indicator for Intracellular Studies. *Biophys. J* 90, 3300–3314. [PubMed: 16603505]
- (6). Nakai J, Ohkura M, and Imoto K (2001) A high signal-to-noise Ca²⁺ probe composed of a single green fluorescent protein. *Nat. Biotechnol* 19, 137–141. [PubMed: 11175727]
- (7). Huang YM, Banerjee S, Crone DE, Schenkelberg CD, Pitman DJ, Buck PM, and Bystroff C (2015) Toward Computationally Designed Self-Reporting Biosensors Using Leave-One-Out Green Fluorescent Protein. *Biochemistry* 54, 6263–6273. [PubMed: 26397806]
- (8). Shaner NC, Campbell RE, Steinbach PA, Giepmans BNG, Palmer AE, and Tsien RY (2004) Improved monomeric red, orange and yellow fluorescent proteins derived from *Discosoma* sp. red fluorescent protein. *Nat. Biotechnol* 22, 1567–1572. [PubMed: 15558047]
- (9). Reid BG, and Flynn GC (1997) Chromophore Formation in Green Fluorescent Protein. *Biochemistry* 36, 6786–6791. [PubMed: 9184161]
- (10). Rosenow MA, Huffman HA, Phail ME, and Wachter RM (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. [PubMed: 15078092]
- (11). Barondeau DP, Putnam CD, Kassmann CJ, Tainer JA, and Getzoff ED (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. [PubMed: 14523232]
- (12). Wood TI, Barondeau DP, Hitomi C, Kassmann CJ, Tainer JA, and Getzoff ED (2005) Defining the Role of Arginine 96 in Green Fluorescent Protein Fluorophore Biosynthesis. *Biochemistry* 44, 16211–16220. [PubMed: 16331981]
- (13). Evdokimov AG, Pokross ME, Egorov NS, Zaraisky AG, Yampolsky IV, Merzlyak EM, Shkoporov AN, Sander I, Lukyanov KA, and Chudakov DM (2006) Structural basis for the fast maturation of Arthropoda green fluorescent protein. *EMBO Rep* 7, 1006–1012. [PubMed: 16936637]
- (14). Zhang Q, Chen X, Cui G, Fang W-H, and Thiel W (2014) Concerted Asynchronous Hula-Twist Photoisomerization in the S65T/H148D Mutant of Green Fluorescent Protein. *Angew. Chem., Int. Ed* 53, 8649–8653.
- (15). Megley CM, Dickson LA, Maddalo SL, Chandler GJ, and Zimmer M (2009) Photophysics and Dihedral Freedom of the Chromophore in Yellow, Blue, and Green Fluorescent Protein. *J. Phys. Chem. B* 113, 302–308. [PubMed: 19067572]
- (16). Stoner-Ma D, Jaye AA, Matousek P, Towrie M, Meech SR, and Tonge PJ (2005) Observation of Excited-State Proton Transfer in Green Fluorescent Protein using Ultrafast Vibrational Spectroscopy. *J. Am. Chem. Soc* 127, 2864–2865. [PubMed: 15740117]

- (17). Maddalo SL, and Zimmer M (2006) The role of the protein matrix in green fluorescent protein fluorescence. *Photochem. Photobiol* 82, 367–372. [PubMed: 16613487]
- (18). Andrews BT, Roy M, and Jennings PA (2009) Chromophore Packing Leads to Hysteresis in GFP. *J. Mol. Biol* 392, 218–227. [PubMed: 19577576]
- (19). Rosenman DJ, Huang Y, Xia K, Fraser K, Jones VE, Lamberson CM, Van Roey P, Colon W, and Bystroff C (2014) Green-lighting green fluorescent protein: Faster and more efficient folding by eliminating a cis–trans peptide isomerization event. *Protein Sci* 23, 400–410. [PubMed: 24408076]
- (20). Tsien RY (1998) THE GREEN FLUORESCENT PROTEIN. *Annu. Rev. Biochem* 67, 509–544. [PubMed: 9759496]
- (21). Jung G, Wiehler J, and Zumbusch A (2005) The photophysics of green fluorescent protein: influence of the key amino acids at positions 65, 203, and 222. *Biophys. J* 88, 1932–1947. [PubMed: 15613627]
- (22). Nakano H, Okumura R, Goto C, and Yamane T (2002) In vitro combinatorial mutagenesis of the 65th and 222nd positions of the green fluorescent protein of *Aequorea victoria*. *Biotechnol. Bioprocess Eng* 7, 311–315.
- (23). Abramoff MD, Magalhaes PJ, and Ram SJ (2004) Image processing with Image. *Journal of Biophotonics International* 11, 36–42.
- (24). Edelhoch H (1967) Spectroscopic determination of tryptophan and tyrosine in proteins*. *Biochemistry* 6, 1948–1954. [PubMed: 6049437]
- (25). Gill SC, and von Hippel PH (1989) Calculation of protein extinction coefficients from amino acid sequence data. *Anal. Biochem* 182, 319–326. [PubMed: 2610349]
- (26). Pace CN, Vajdos F, Fee L, Grimsley G, and Gray T (1995) How to measure and predict the molar absorption coefficient of a protein. *Protein Sci* 4, 2411–2423. [PubMed: 8563639]
- (27). Leaver-Fay A, Tyka M, Lewis SM, Lange OF, Thompson J, Jacak R, Kaufman K, Renfrew PD, Smith CA, Sheffler W, Davis IW, Cooper S, Treuille A, Mandell DJ, Richter F, Ban YE, Fleishman SJ, Corn JE, Kim DE, Lyskov S, Berrondo M, Mentzer S, Popovic Z, Havranek JJ, Karanicolas J, Das R, Meiler J, Kortemme T, Gray JJ, Kuhlman B, Baker D, and Bradley P (2011) ROSETTA3: an object-oriented software suite for the simulation and design of macromolecules. *Methods Enzymol* 487, 545–574. [PubMed: 21187238]
- (28). O'Meara MJ, Leaver-Fay A, Tyka MD, Stein A, Houlihan K, DiMaio F, Bradley P, Kortemme T, Baker D, Snoeyink J, and Kuhlman B (2015) Combined Covalent-Electrostatic Model of Hydrogen Bonding Improves Structure Prediction with Rosetta. *J. Chem. Theory Comput* 11, 609–622. [PubMed: 25866491]
- (29). Blaber M, Lindstrom JD, Gassner N, Xu J, Heinz DW, and Matthews BW (1993) Energetic cost and structural consequences of burying a hydroxyl group within the core of a protein determined from Ala.fwdarw. Ser and Val.fwdarw. Thr substitutions in T4 lysozyme. *Biochemistry* 32, 11363–11373. [PubMed: 8218201]
- (30). Oyen D, Fenwick RB, Stanfield RL, Dyson HJ, and Wright PE (2015) Cofactor-Mediated Conformational Dynamics Promote Product Release From *Escherichia coli* Dihydrofolate Reductase via an Allosteric Pathway. *J. Am. Chem. Soc* 137, 9459–9468. [PubMed: 26147643]
- (31). Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, and Ferrin TE (2004) UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem* 25, 1605–1612. [PubMed: 15264254]
- (32). The PyMOL Molecular Graphics System, version 1.8 (2015) Schrödinger, LLC, Portland, OR.
- (33). Finn RD, Coghill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, Potter SC, Punta M, Qureshi M, Sangrador-Vegas A, Salazar GA, Tate J, and Bateman A (2016) The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res* 44, D279–285. [PubMed: 26673716]
- (34). Kent KP, Childs W, and Boxer SG (2008) Deconstructing Green Fluorescent Protein. *J. Am. Chem. Soc* 130, 9664–9665. [PubMed: 18597452]
- (35). Do K, and Boxer SG (2011) Thermodynamics, Kinetics, and Photochemistry of β -Strand Association and Dissociation in a Split-GFP System. *J. Am. Chem. Soc* 133, 18078–18081. [PubMed: 21981121]

- (36). Kelly SM, Jess TJ, and Price NC (2005) How to study proteins by circular dichroism. *Biochim. Biophys. Acta, Proteins Proteomics* 1751, 119–139.
- (37). Zhang L, Patel HN, Lappe JW, and Wachter RM (2006) Reaction Progress of Chromophore Biogenesis in Green Fluorescent Protein. *J. Am. Chem. Soc* 128, 4766–4772. [PubMed: 16594713]
- (38). Pedelacq J-D, Cabantous S, Tran T, Terwilliger TC, and Waldo GS (2006) Engineering and characterization of a superfolder green fluorescent protein. *Nat. Biotechnol* 24, 79–88. [PubMed: 16369541]
- (39). Jain RK, and Ranganathan R (2004) Local complexity of amino acid interactions in a protein core. *Proc. Natl. Acad. Sci. U. S. A* 101, 111–116. [PubMed: 14684834]
- (40). Almasian M, Grzetic J, Berden G, Bakker B, Buma WJ, and Oomens J (2012) Gas-phase infrared spectrum of the anionic GFP-chromophore. *Int. J. Mass Spectrom* 330–332, 118–123.
- (41). Huang Y-M, Nayak S, and Bystroff C (2011) Quantitative in vivo solubility and reconstitution of truncated circular permutants of green fluorescent protein. *Protein Sci* 20, 1775–1780. [PubMed: 21910151]
- (42). Fukuda H, Arai M, and Kuwajima K (2000) Folding of Green Fluorescent Protein and the Cycle3Mutant. *Biochemistry* 39, 12025–12032. [PubMed: 11009617]
- (43). Enoki S, Saeki K, Maki K, and Kuwajima K (2004) Acid Denaturation and Refolding of Green Fluorescent Protein†. *Biochemistry* 43, 14238–14248. [PubMed: 15518574]
- (44). Huang J, Craggs TD, Christodoulou J, and Jackson SE (2007) Stable Intermediate States and High Energy Barriers in the Unfolding of GFP. *J. Mol. Biol* 370, 356–371. [PubMed: 17512539]
- (45). Pitman DJ, Banerjee S, Macari SJ, Castaldi C. a., Crone DE, and Bystroff C (2015) Exploring the folding pathway of green fluorescent protein through disulfide engineering. *Protein Sci* 24, 341–353. [PubMed: 25516354]
- (46). Heim R, Prasher DC, and Tsien RY (1994) Wavelength mutations and posttranslational autooxidation of green fluorescent protein. *Proc. Natl. Acad. Sci. U. S. A* 91, 12501–12504. [PubMed: 7809066]
- (47). Chica RA, Moore MM, Allen BD, and Mayo SL (2010) Generation of longer emission wavelength red fluorescent proteins using computationally designed libraries. *Proc. Natl. Acad. Sci. U. S. A* 107, 20257–20262. [PubMed: 21059931]
- (48). Ormö M, Cubitt AB, Kallio K, Gross LA, Tsien RY, and Remington SJ (1996) Crystal Structure of the *Aequorea victoria* Green Fluorescent Protein. *Science* 273, 1392–1395. [PubMed: 8703075]
- (49). Kent KP, and Boxer SG (2011) Light-Activated Reassembly of Split Green Fluorescent Protein. *J. Am. Chem. Soc* 133, 4046–4052. [PubMed: 21351768]
- (50). Huang Y, and Bystroff C (2009) Complementation and Reconstitution of Fluorescence from Circularly Permuted and Truncated Green Fluorescent Protein†. *Biochemistry* 48, 929–940. [PubMed: 19140681]
- (51). Close DW, Paul CD, Langan PS, Wilce MCJ, Traore DAK, Halfmann R, Rocha RC, Waldo GS, Payne RJ, Rucker JB, Prescott M, and Bradbury ARM (2015) Thermal green protein, an extremely stable, nonaggregating fluorescent protein created by structure-guided surface engineering. *Proteins: Struct., Funct., Genet* 83, 1225–1237. [PubMed: 25287913]
- (52). Kiss C, Temirov J, Chasteen L, Waldo GS, and Bradbury ARM (2009) Directed evolution of an extremely stable fluorescent protein. *Protein Eng., Des. Sel* 22, 313–323. [PubMed: 19364809]
- (53). Dai M, Fisher HE, Temirov J, Kiss C, Phipps ME, Pavlik P, Werner JH, and Bradbury ARM (2007) The creation of a novel fluorescent protein by guided consensus engineering. *Protein Eng., Des. Sel* 20, 69–79. [PubMed: 17277006]
- (54). Karasawa S, Araki T, Yamamoto-Hino M, and Miyawaki A (2003) A green-emitting fluorescent protein from *Galaxeidae* coral and its monomeric version for use in fluorescent labeling. *J. Biol. Chem* 278, 34167–34171. [PubMed: 12819206]
- (55). Stoner-Ma D, Jaye AA, Ronayne KL, Nappa J, Meech SR, and Tonge PJ (2008) An Alternate Proton Acceptor for Excited-State Proton Transfer in Green Fluorescent Protein: Rewiring GFP. *J. Am. Chem. Soc* 130, 1227–1235. [PubMed: 18179211]

- (56). Johnson LB, Huber TR, and Snow CD (2014) Methods for library-scale computational protein design. *Methods Mol. Biol* 1216, 129–159. [PubMed: 25213414]
- (57). Park JW, and Rhee YM (2016) Electric Field Keeps Chromophore Planar and Produces High Yield Fluorescence in Green Fluorescent Protein. *J. Am. Chem. Soc* 138, 13619–13629.
- (58). Sauer M, Hofkens J, and Enderlein J (2011) Basic Principles of Fluorescence Spectroscopy. In *Handbook of Fluorescence Spectroscopy and Imaging*, pp 1–30, Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, Germany.
- (59). Andrews BT, Gosavi S, Finke JM, Onuchic JN, and Jennings PA (2008) The dual-basin landscape in GFP folding. *Proc. Natl. Acad. Sci. U. S. A* 105, 12283–12288. [PubMed: 18713871]

Figure 1.

Microenvironment of the chromophore in superfolder GFP (PDB entry 2B3P). Hydrogen bonds (blue dashes) and nonpolar contacts (cyan dashes) are shown. The backbone, represented in translucent cartoon form, is rainbow-colored from the N terminus (violet) to the C terminus (red). Residues that were subject to saturation mutagenesis are shown with thick bonds. Residues R96, T203, and E222 are colored magenta. A buried water-mediated hydrogen bonding network, involving Q69, Q94, R96, and Q183, is colored magenta.

Figure 2. Colonies of mutants expressing (a) F145M, (b) F145W, and (c) Q94[AEKLMPQTV]. Panel c is representative of all other plates. F145W colonies exhibit a distinct yellow color, while all other colonies are different shades of fluorescent green. The variability in fluorescence observed for panel c is shown in panel d.

Figure 3.
(a) Far-UV and (b) near-UV circular dichroism spectra of equimolar amounts of the wild-type and representative mutant proteins.

Figure 4.

(a) Fluorescence spectra of OPT-GFP and F145M with equal concentrations of protein. F145M shows a higher RF because of better CM as well as a higher QY. (b) Fluorescence spectra of OPT-GFP, F145M, and F145W normalized by chromophore concentration, and by (insets) peak emission, showing (left inset) a minor “A” peak in the F145W emission wavelength, (right inset) a 2 nm blue-shift in the F145W emission wavelength, and a 2 nm red-shift in the F145M emission wavelength.

Figure 5. Absorption spectra for the representative mutants, normalized for protein concentration $\eta = (A_{280}/\epsilon_{280})$ at pH 8.0. Dashed lines show data for dark mutants. Thick lines show data for selected mutants after 3 or 6 months at 4 °C. Bars indicate significant differences ($p < 0.01$) in A/η .

Table 1.

In Vivo Relative Fluorescence (RF) Intensities for the Wild Type and All 140 Point Mutants of sfGFP at Seven Positions Surrounding the Chromophore^a

^aNumbers and colors are the molar fluorescence intensity at an excitation wavelength of 485 nm and an emission wavelength of 508 nm after chromophore maturation, relative to that of the wild-type protein (red highlight). Mutations observed in GFP homologues are highlighted in blue. The average, over all mutants, of the root-mean-square deviation across multiple measurements of the same mutant is 0.07.

Table 2.

Preliminary Results of Screening Double Mutants of Residues 69 and 183, Which Usually Form a “Handshake” Hydrogen Bonding Pattern, and of Residues 145 and 165, Which in Unison Can Block Rotation of the Chromophore’s Hydroxybenzylidene Moiety

positions	bright	dim	very dim
Q69, Q183	IP	EK, NS, VT, FH	KS, YT, HL, VN, DT
F145, F165	YL, CL, IT, LI	VI, IQ, YA, AT, LP	HG, IE, CN

Table 3.

Summary of the Discussion of Seven Chromophore Milieu Positions

position	summary
69	forms H-bonds with Q183; loosely packed; tolerant of all mutations; Q69I rescues dark mutant Q183P
Q94	conformation changes during maturation; packs against essential residue R96; single dark mutant; Q94D promotes non-native R96 rotamer; matures very slowly
F145	forms major contacts with the chromophore; F145W matures normally; glows very weakly; F145M is brighter than the wild type
H148	conformation changes during maturation; makes a H-bond to CRO OH; tolerates all side chains
F165	forms major contacts with the chromophore; rare rotameric state; folds late; tolerates all mutations; packing may affect RF
Q181	tightly packed; packs against essential residue R96; dark mutant H181D promotes non-native R96 rotamer
Q183	H-bonds with Q69; packs against R96; dark mutant Q183L interferes with side chain packing; Q183P affects backbone conformation