Applications of Computational Protein Design to

Red Fluorescent Proteins

Thesis

by

Matthew Michaels Moore

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

California Institute of Technology,

Pasadena, California

2013

(Defended September 13, 2012)

Acknowledgements

hope they understand in some small way how much their wagging tails and excited faces filled me with joy, even in the darkest of moments.

I thank my lab-mates from Caltech for all their advice, suggestions, friendship, and caring. In no particular order, I would like to thank Alexandria Berry, Jennifer Keeffe, Heidi Privett, Benjamin Allen, Kurt Mou, Toni Lee, Alex Nisthal, Tim Wannier, Christina Vizcarra, Bernardo Sosa Padilla, Mohsen Chitsaz, Emzo de los Santos, Corey Wilson, Erin Drez, Marie Ary, and Rhonda DiGiusto. I will likely always think fondly of our lunch-time discussions, our arguments, and our kinships.

I thank my long time friends Cody Cannell and Gene Chuang for being around in the times when I most needed them. I wish them both well and hope they can forgive me that my ever-busier life made it hard to stay close as I continued towards the completion of my PhD.

I thank Pavle Niklovski for his tireless efforts in the laboratory, for his attention to detail, and for his eagerness to see people satisfied. I thank Mandy Casani and Cathy J. Miles for their continuous support and encouragement during my long search to find a job and start a career upon completion of my PhD. Their advice and kind words did not go unappreciated.

I thank my thesis committee members—Linda Hsieh-Wilson, Francis Arnold, and Douglas Rees—for their valuable time and precious intellectual input during our meetings together.

I especially thank Roberto Antonio Chica, without whom I would never have completed my PhD studies. Professor Chica was always enthusiastic, encouraging, and excited to

Part Two

Chapter 4. A method for rapid identification of monomeric fluorescent protein mutants
guided by computational protein design

Abstract

The research comprising this thesis is primarily concerned with applications of computational protein design (CPD) to red fluorescent proteins (FPs); it is presented in three chapters divided into two parts. Part One, Chapters 2 & 3, of the thesis focuses on the application of rational design and CPD to the core residues of FPs. Chapter 2 applies CPD to the well known red fluorescent protein mCherry. Design hypotheses in this work were driven by a desire to red-shift the fluorescence emission of the parent protein. Chapter 3 takes the most successful results from the mCherry system and attempts to applies these results to the far-red FP mPlum. These two proteins, mCherry and mPlum, share a directed evolution parent, mRFP1. Part Two, consisting of Chapter 4, presents the beginnings of a comprehensive study into the applications CPD for designing FP surfaces. The system used is based on DsRed, the oligomeric parent of monomer mCherry.

Introduction

Fluorescent proteins (FPs) are a dream system for the biophysical protein chemist. These molecules marry the studies of quantum mechanisms, kinetics, protein chemistry, and spectroscopy into an intellectually engaging blend of experimental phenomena. Not only is there a wealth of information about these proteins and their experimental systems, but there is also always a new wealth of information to collect in any FP project. The vibrant colors and predictable behaviors of these proteins make them a joy to work worth, as well.

In 2008, the Nobel Prize in Chemistry was awarded "for the discovery and development of the green fluorescent protein, GFP" [1]. This prize signified the importance of FPs in research science. These proteins share a common "β-can" motif with a coaxial α-helix located within an 11-stranded β-barrel [2, 3]. A chromophore located in the middle of this helix is responsible for the fluorescent properties that make these proteins so useful for *in vivo* labeling experiments. This chromophore is formed by autocatalytic cyclization of a tripeptide sequence (Xaa-Tyr-Gly) in the core of the β-barrel fold [3]. Although variations exist, the chromophore comes mainly in green and red forms [3, 4]. The same key steps are believed to underlie the initial chromophore formation process regardless of the final covalent structure [5, 6].

Accordingly, the most salient chemical feature for all colors of FPs is this ability to undergo autocatalytic backbone cyclization resulting in the formation of a fluorescent chromophore. No external cofactors or enzymes are needed to obtain fluorescence during heterologous expression of these proteins [2, 3]. The biosynthesis of this chromophore is therefore completely encoded in the primary structure of these proteins. The mechanism

of chromophore formation has been intensely studied, and the knowledge gained has been successfully applied toward the rational engineering of FP variants with novel properties [7-9].

With regards to this mechanism, the most thoroughly characterized contributor to rapid autocatalytic backbone cyclization in GFP is the presence of a buried lysine or arginine residue [10]. This residue, which is conserved as an arginine in all known FPs [3], contributes a buried positive charge to the area near the chromophore-forming residues of the central helix. In GFP, this R96 residue has been successfully replaced by a lysine (GFP-R96K) with no significant effect on the rate of fluorescence development after overnight expression in *E. coli* [10]. Furthermore, rapid generation of fluorescence was recovered when the Q183R mutation was incorporated into GFP-R96A [10]. These results were confirmed by crystallography whereby the position 96-adjacent R183 residue was shown to fulfill all H-bonds of the wild-type R96 to the chromophore and to occupy the internal pocket created by the R96A mutation [10].

It is still not completely clear how the universally conserved glutamate residue in FPs contributes to the mechanism of chromophore formation. A variety of compensating and tolerated mutations have been investigated for the corresponding E222 in GFP, such as E222Q [11, 12] and E222G [13]. In fact, a colorless and non-fluorescent GFP homolog from *Aequorea coerulescens* was made fluorescent by an E222G mutation discovered through random mutagenesis [14]. The optimized fluorescent mutant, termed aceGFP, has 92% sequence identity with GFP from *A. victoria*, and reversal of the required E222G mutation severely compromises chromophore maturation in this mutant (only ~3% of soluble aceGFP-G222E reaches the mature state) [14].

However, it is clear that E222 in GFP has some effect on the environment inside the GFP core. For instance, the E222G mutation in *A. victoria* GFP lowers the pK$_a$ of the mature chromophore to the point that the deprotonated chromophore is the only species detected at pH 7.0 [2, 13]. Additionally, Wachter *et al.* reported that the maturation rate of GFP-E222Q increases dramatically with increasing pH, whereas the maturation rate of wild-type GFP is constant above pH 8.0 [11]. They interpreted this change in pH dependence to mean that E222 effectively lowers the pK_a of the nucleophilic amide nitrogen involved in backbone cyclization.

The sequence dependence of chromophore formation for core residues in the 11-stranded β-barrel fold of FPs has never been intentionally investigated in an hypothesis-directed experimental fashion [2, 3]. Countless mutagenesis protocols have deeply inquired into what residue substitutions allow chromophore formation while attenuating various properties of the protein [4], such as maturation rate [15], stability [16], emission wavelength [17-19], and fluorescence brightness [20, 21]. However, very few if any reports have addressed what core substitutions maintain protein stability but disallow chromophore formation at positions other than the catalytic residues or those immediately adjacent to the chromogenic tripeptide [10-12, 22, 23].

Red fluorescent proteins (RFPs) derived from organisms in the class *Anthozoa*, in particular, have gained widespread notoriety through their applications to cell biology [3, 24, 25]. For example, these proteins have been used as markers of gene expression, expressed as fusion partners for the tracking of intracellular endogenous protein-RFP chimeras, and complemented with other FPs for use in fluorescence resonance energy transfer (FRET) experiments [3, 25, 26]. The availability of monomeric versions of these

proteins has even further bolstered their worth as fusion tags and vaulted them into routine experimental use [24, 27].

In the past, a variety of techniques for large-scale random mutagenesis and screening have been used to push the emission wavelength of RFPs to its far-red limits [16, 24, 28]. Starting from different wild-type precursors, these procedures have produced some of the best and brightest mutants to date: mCherry [16], mKate [29], mNeptune [30], and mPlum [10]. However, these approaches were based on the exploration of a sequence space that by design differs from the template by single nucleotide mutations. The limits of these methods can be circumvented by a rational approach using computational protein design (CPD). This approach can be used to identify focused combinatorial libraries of mutants, and these libraries correspond to sequences that satisfy a given three-dimensional structure according to a molecular dynamics based force field [31, 32].

Wild-type RFPs are typically obligate oligomers [3, 26]. Generation of monomeric FPs for molecular tagging is usually accomplished by a disrupt-and-recover approach. This approach starts with the sampling of mutations that disrupt known protein-protein interfaces [27]. However, disruption often leads to significant interference with the fluorescent properties of an FP, making it necessary to search for mutations to recover fluorescence [27]. Moreover, this approach can be tedious because full recovery may require multiple rounds of directed evolution and screening, and in some cases, may never be achievable. Here as well, rational design-based approaches guided by the use of CPD can benefit the course of this process.

The research comprising this thesis is presented in three chapters divided into two parts. Part One, Chapters 2 & 3, of the thesis focuses on the application of rational design and

CPD to the core residues of FPs. Chapter 2 applies CPD to the well known red fluorescent protein mCherry. Design hypotheses in this work were driven by a desire to red-shift the fluorescence emission of the parent protein. Chapter 3 takes the most successful results from the mCherry system and attempts to applies these results to the far-red FP mPlum. These two proteins, mCherry and mPlum, share a directed evolution parent, mRFP1. Part Two, consisting of Chapter 4, presents the beginnings of a comprehensive study into the applications CPD for designing FP surfaces. The system used is based on DsRed, the oligomeric parent of monomer mCherry. Professional and personal acknowledgments conclude the thesis.

References

1. Nobelprize.org (7 May 2012) The Nobel Prize in Chemistry 2008. http://www.nobelprize.org/nobel_prizes/chemistry/laureates/2008/.

2. Tsien RY (1998) The green fluorescent protein. *Annual Review of Biochemistry* 67:509-544.

3. Chudakov DM*, et al.* (2010) Fluorescent proteins and their applications in imaging living cells and tissues. *Physiological Reviews* 90(3):1103-1163.

4. Subach FV, Piatkevich KD, & Verkhusha VV (2011) Directed molecular evolution to design advanced red fluorescent proteins. *Nature Methods* 8(12):1019-1026.

5. Zhang LP*, et al.* (2006) Reaction progress of chromophore biogenesis in green fluorescent protein. *Journal of the American Chemical Society* 128:4766-4772.

6. Strack RL*, et al.* (2010) Chromophore formation in DsRed occurs by a branched pathway. *Journal of the American Chemical Society* 132(24):8496-8505.

7. Wachter RM*, et al.* (1998) Structural basis of spectral shifts in the yellow-emission variants of green fluorescent protein. *Structure With Folding & Design* 6:1267-1277.

8. Griesbeck O*, et al.* (2001) Reducing the environmental sensitivity of yellow fluorescent protein: Mechanism and applications. *Journal of Biological Chemistry* 276:29188-29194.

9.   Mishin AS*, et al.* (2008) The first mutant of the Aequorea victoria green fluorescent protein that forms a red chromophore. *Biochemistry* 47:4666-4673.

10.   Wood TI*, et al.* (2005) Defining the role of arginine 96 in green fluorescent protein fluorophore biosynthesis. *Biochemistry* 44:16211-16220.

11.   Sniegowski JA*, et al.* (2005) Base catalysis of chromophore formation in Arg96 and Glu222 variants of green fluorescent protein. *Journal of Biological Chemistry* 280(28):26248-26255.

12.   Sniegowski JA, Phail ME, & Wachter R (2005) Maturation efficiency, trypsin sensitivity, and optical properties of Arg96, Glu222, and Gly67 variants of green fluorescent protein. *Biochemical and Biophysical Research Communications* 332(3):657-663.

13.   Ehrig T, Okane DJ, & Prendergast FG (1995) Green-fluorescent protein mutants with altered fluorescence excitation-spectra. *FEBS Lett* 367:163-166.

14.   Gurskaya NG*, et al.* (2003) Colourless green fluorescent protein homologue from the non-fluorescent hydromedusa Aequorea coerulescens and its fluorescent mutants. *Biochemistry Journal* 373:403-408.

15.   Terskikh AV*, et al.* (2002) Analysis of DsRed mutants: Space around the fluorophore accelerates fluorescence development. *Journal of Biological Chemistry* 277:7633-7636.

16.   Shaner N*, et al.* (2004) Improved monomeric red, orange and yellow fluorescent proteins derived from Discosoma sp. red fluorescent protein. *Nature Biotechnology* 22(12):1567-1572.

17.   Wang L*, et al.* (2004) Evolution of new nonantibody proteins via iterative somatic hypermutation. *Proceedings of the National Academy of Sciences, USA* 101:16745-16749.

18.   Chica RA*, et al.* (2010) Generation of longer emission wavelength red fluorescent proteins using computationally designed libraries. *Proceedings of the National Academy of Sciences, USA* 107(47):20257-20262.

19.   Morozova KS*, et al.* (2010) Far-Red Fluorescent Protein Excitable with Red Lasers for Flow Cytometry and Superresolution STED Nanoscopy. *Biophysical Journal* 99(2):L13-L15.

20.   Ai H*, et al.* (2006) Directed evolution of a monomeric, bright and photostable version of Clavularia cyan fluorescent protein: structural characterization and applications in fluorescence imaging. *Biochemistry Journal* 400(3):531-540.

21.   Goedhart J*, et al.* (2010) Bright cyan fluorescent protein variants identified by fluorescence lifetime screening. *Nature Methods* 7(2):137-139.

22. Barondeau DP, *et al.* (2005) Understanding GFP chromophore biosynthesis: Controlling backbone cyclization and modifying post-translational chemistry. *Biochemistry* 44:1960-1970.

23. Barondeau DP, *et al.* (2006) Understanding GFP posttranslational chemistry: Structures of designed variants that achieve backbone fragmentation, hydrolysis, and decarboxylation. *Journal of the American Chemical Society* 128:4685-4693.

24. Shaner N, Steinbach P, & Tsien R (2005) A guide to choosing fluorescent proteins. *Nature Methods* 2(12):905-909.

25. Giepmans BNG, *et al.* (2006) The fluorescent toolbox for assessing protein location and function. *Science* 312:217-224.

26. Verkhusha V & Lukyanov K (2004) The molecular properties and applications of Anthozoa fluorescent proteins and chromoproteins. *Nature Biotechnology* 22(3):289-296.

27. Campbell RE, *et al.* (2002) A monomeric red fluorescent protein. *Proceedings of the National Academy of Sciences, USA* 99:7877-7882.

28. Merzlyak E, *et al.* (2007) Bright monomeric red fluorescent protein with an extended fluorescence lifetime. *Nature Methods* 4(7):555-557.

29. Shcherbo D, *et al.* (2007) Bright far-red fluorescent protein for whole-body imaging. *Nature Methods* 4(9):741-746.

30. Lin MZ, *et al.* (2009) Autofluorescent proteins with excitation in the optical window for intravital imaging in mammals. *Chemistry & Biology* 16(11):1169-1179.

31. Mena M, *et al.* (2006) Blue fluorescent proteins with enhanced brightness and photostability from a structurally targeted library. *Nature Biotechnology* 24(12):1569-1571.

32. Treynor TP, *et al.* (2007) Computationally designed libraries of fluorescent proteins evaluated by preservation and diversity of function. *Proceedings of the National Academy of Sciences, USA* 104:48-53.

Generation of Longer Emission Wavelength Red Fluorescent Proteins Using Computationally Designed Libraries

Roberto A. Chica[a,b], Matthew M. Moore[c], Benjamin D. Allen[c], Stephen L. Mayo[a,c,1]

[a]Division of Biology, California Institute of Technology,

1200 E. California Blvd., Pasadena, CA 91125 USA.

[b]Present address: Department of Chemistry, University of Ottawa,

10 Marie Curie, Ottawa, ON, Canada K1N 6N5.

[c]Division of Chemistry and Chemical Engineering, California Institute of Technology,

1200 E. California Blvd., Pasadena, CA 91125 USA.

[1]Corresponding author[1]: steve@mayo.caltech.edu

Abbreviations: CPD, computational protein design; RFP, red fluorescent protein; FP, fluorescent protein; λ_{em}, fluorescence emission wavelength; λ_{ex}, fluorescence excitation wavelength; Φ_F, fluorescence quantum yield

Abstract

The longer emission wavelengths of red fluorescent proteins (RFPs) make them attractive for whole-animal imaging because cells are more transparent to red light. Although several useful RFPs have been developed using directed evolution, the quest for further red-shifted and improved RFPs continues. Herein, we report a structure-based rational design approach to red shift the fluorescence emission of RFPs. We applied a combined computational and experimental approach that uses computational protein design as an *in silico* pre-screen to generate focused combinatorial libraries of mCherry mutants. The computational procedure helped us identify residues that could fulfill interactions hypothesized to cause red shifts without destabilizing the protein fold. These interactions include stabilization of the excited state through H-bonding to the acylimine oxygen atom, destabilization of the ground state by hydrophobic packing around the charged phenolate, and stabilization of the excited state by a π-stacking interaction. Our methodology allowed us to identify three mCherry mutants (mRojoA, mRojoB, and mRouge) that display emission wavelengths >630 nm, representing red shifts of 20–26 nm. Moreover, our approach required the experimental screening of a total of ~5000 clones, a number several orders of magnitude smaller than those previously used to achieve comparable red shifts. Additionally, crystal structures of mRojoA and mRouge allowed us to verify fulfillment of the interactions hypothesized to cause red shifts, supporting their contribution to the observed red shifts.

Red fluorescent proteins (RFPs) derived from organisms in the class *Anthozoa* have found widespread application in cell biology. For example, these proteins have been used as markers of gene expression [2], expressed as fusions to track endogenous protein within cells [3], and applied with other fluorescent proteins (FPs) for use in FRET experiments [4]. The availability of monomeric versions of these proteins has bolstered their worth as fusion tags and vaulted them into routine experimental use [5, 6].

The red fluorescence displayed by these proteins arises from the presence of an acylimine group conjugated with the standard *p*-hydroxybenzylideneimidazolinone GFP chromophore [7]. The additional double bond extends the size of the chromophore conjugated system leading to an increase in emission wavelength. The longer emission wavelength of RFPs makes them attractive for whole-animal imaging because cells are more transparent to red light. For imaging applications, higher emission wavelengths (650–900 nm) are desirable because they tend to minimize background absorption and light scattering by tissue components and are less damaging to cells, enabling longer acquisition times.

Naturally-occurring *Anthozoa* RFPs, such as zRFP574 [8], eqFP578 [9], DsRed [10], and eqFP611 [11], are obligate oligomers that display emission wavelengths ranging from 574 nm to 611 nm. Significant effort has been made to monomerize and red-shift the emission wavelength of these RFPs using directed evolution. Starting from various wild-type precursors, these procedures have produced several far-red ($\lambda_{em} > 630$ nm) monomeric RFPs such as mPlum [12], mKate2 [13], and mNeptune [14]. Each of these useful monomeric RFPs was developed using random mutagenesis [6, 14, 15]. Although directed evolution has successfully yielded red-shifted monomeric RFPs, a strictly

rational methodology to red shift *Anthozoa* class FPs has not yet been described. Aside from the T203Y mutation in *Aequorea victoria* GFP that red-shifts emission by 20 nm to yield the yellow variant YFP [16], and mutations leading to a *trans*-to-*cis* isomerization in the chromophore of eqFP611 [3], the rational prediction of mutations causing red shifts has proven difficult.

We were interested in developing a structure-based rational design approach to red shift the fluorescence emission of RFPs. Toward this end, we proposed three structure-based hypotheses supported by results reported in the literature and tested them using the mCherry scaffold. We applied a combined computational and experimental approach that used computational protein design (CPD) as an *in silico* pre-screen to generate focused combinatorial libraries of mCherry mutants. The computational procedure helped us identify residues that could fulfill the interactions hypothesized to cause red shifts without destabilizing the protein fold. This methodology allowed us to identify mutants displaying bathochromic emission wavelength shifts of up to 26 nm. Moreover, our approach required the experimental screening of libraries several orders of magnitude smaller than those previously used to achieve comparable red shifts. Additionally, crystal structures of two of our most red-shifted mutants allowed us to verify fulfillment of our structure-based design hypotheses and gain a better understanding of the causes of red shifts in RFPs.

Results and Discussion

Design hypotheses. To red shift the emission wavelength of a monomeric RFP, three literature-based hypotheses were tested by screening for red shifting in libraries of computationally designed mCherry variants. RFP chromophores contain a *p*-hydroxybenzylideneimidazolinone group prepended with an acylimine substituent at the C1 atom. The acylimine group is the result of an oxidation coupled to *trans*-to-*cis* isomerization of the peptide bond between the first chromophore-forming residue and the preceding residue (F65-M66 bond in mCherry) [7]. The presence of the acylimine electron-withdrawing substituent creates a resonance structure for the deprotonated form of the chromophore in which the partial negative charge is located on the acylimine oxygen atom (Fig. 1*A*, bottom right). The deprotonated chromophore has been demonstrated to be the species responsible for red fluorescence in RFPs [6] and quantum mechanical studies have indicated that resonance forms of this species with the partial negative charge localized on the imidazolinone and acylimine oxygen atoms make higher contributions to the description of the excited state [1, 17, 18]. Design hypothesis I (Fig. 1*A*) involves stabilization of the excited state by H-bonding the acylimine oxygen of the chromophore.

In all known crystal structures of FPs [19], a strictly conserved Arg residue forms an H-bond to the imidazolinone oxygen atom (R95, Fig. 1*A*). However, no conserved residue has been identified as H-bonding to the acylimine oxygen. CPD was used to identify potential H-bond donor residues located within direct or water-mediated H-bonding distance to the acylimine oxygen. The goal was to preferentially stabilize the resonance form of the chromophore in which the negative charge is located on the

acylimine oxygen. We hypothesized that such an H-bond would stabilize the excited state, thereby decreasing the energy difference between the excited and ground states, resulting in an increased emission wavelength. Literature results suggest that this type of interaction would have the predicted effect. The crystal structures of mPlum and its E16Q variant demonstrate an H-bond between the acylimine oxygen and E16 or Q16, respectively [20]. Mutation of E16 to hydrophobic residues leads to hypsochromic emission wavelength shifts [12]. Additionally, in the course of this work, the crystal structure of the far-red RFP Neptune demonstrated a water-mediated H-bond between the acylimine oxygen and neighbouring residue S28 that leads to red-shifted excitation and emission [14].

Design hypothesis II (Fig. 1*B*) involves destabilizing the ground state by hydrophobic packing around the charged chromophore phenolate group. We theorized that destabilization of negative charge localization at the oxygen atom of the phenolate would raise the energy of the ground state relative to the excited state, thereby creating a further red-shifted emission spectrum. To test this hypothesis, CPD was used to identify hydrophobic residues that could be accommodated at positions surrounding the phenolate negative charge (residues 143, 146, 161, and 163). Campbell *et al.* [5] postulated that residues just above the plane of the chromophore of DsRed (residues 161, 163, 175, and 177) influence polarization of the RFP chromophore, thus affecting the emission wavelength. For example, DsRed has a Lys at position 163, which can stabilize the phenolate negative charge through an electrostatic interaction. In mCherry, an uncharged Gln is found at this position, perhaps partially accounting for its red-shifted emission relative to DsRed.

Design hypothesis III (Fig. 1*C*) involves stabilizing polarization of the excited state with a π-stacking interaction, similar to that seen in the crystal structures of yellow variants of *Aequorea victoria* GFP containing the T203Y mutation [21, 22]. In these mutants, π-stacking between Y203 and the chromophore phenolate is thought to stabilize the polarization of the chromophore excited state [16], leading to a 20 nm bathochromic shift in emission wavelength.

Computational design. To test these hypotheses, we applied a novel CPD approach to generate focused combinatorial libraries of mCherry variants containing mutations that could potentially fulfill the interactions hypothesized to cause red shifts [23]. The library design procedure takes as input a list of scored sequences, and two sets of constraints: a list of allowed sets of amino acids, and a range of desired library sizes. The algorithm generates a list of the combinatorial libraries that satisfy these constraints, and then ranks the libraries by the degree to which they reflect the energetic preferences present in the list of scored sequences. Thus, CPD was used to perform an *in silico* pre-screen to eliminate sequences incompatible with the protein fold and generate combinatorial libraries amenable to rapid experimental screening. Thirteen positions with side chains that point towards the chromophore were divided into 3 groups intended to test the 3 design hypotheses (Table 1). Design hypothesis I (library 1) was tested by varying residues with side chains ≤ 6 Å from the acylimine oxygen (positions 14, 16, and 120). Design hypothesis II (library 2) was tested by varying residues with side chains within 5 Å of the chromophore phenolate (positions 143, 146, 161, and 163). Positions 175 and 177 were also varied, since they have been suggested to influence polarization of the chromophore [5]. Design hypothesis III (library 3-1) was tested by introducing a

π-stacking Tyr at position 197 and varying neighbouring residues (73, 146, 195, and 217) to stabilize the Tyr in the correct orientation. In all the designs, all 20 amino acids were allowed at each of the designed positions.

Prior to computational design, these 13 positions were subjected to site-saturation mutagenesis and screening to identify single mutations producing emission red shifts. We found several point mutations that caused small red shifts (3-8 nm): F14S, F14N, V16S, V16T, W143M, S146C, S146T, I161M, Q163M, I197T, I197Y, and A217S. I197Y was already known to cause red shifts in mRFP1.1 [14] and GFP [22], F14S was shown to red-shift DsRed [24], and V16T was critical in red-shifting mGrape2 [14]. I197R and I197H led to hypsochromic shifts (to 593 nm and 605 nm, respectively), a result that correlates well with the fact that basic residues near the chromophore phenolate, such as K163 in DsRed and H199 in amFP486 [25], decrease emission wavelength. This effect can be explained by stabilization of the ground state through a favourable electrostatic interaction with the phenolate negative charge.

For each of the designed positions, the red-shifting point mutations identified above and the wild-type residue were required in the final library composition during computational library design. A217C, which was found to enhance brightness, was also required for some libraries, as was S146G (to allow similar degenerate codon composition for libraries 3-1 and 3-2).

For libraries 1 and 2, we specified a size of ~500 sequences—a compromise between sufficient sequence diversity and ease of screening. Library 3 was divided into two half-size libraries of ~250 sequences. Library 3-1 only allowed Tyr in its π-stacking

conformation (see Materials and Methods) at position 197 (to test design hypothesis III) and 3-2 only allowed the red-shifting Thr at position 197, but in this case its conformation was allowed to vary.

Computational library design results are listed in Table 1. For library 1, 540 sequences predicted to be the most energetically favourable were identified for screening. The final library composition included a large variety of amino acids of different sizes and properties. Each position included multiple H-bond donor residues, allowing comprehensive testing of design hypothesis I. Interestingly, Glu and Gln found in mPlum and its E16Q mutant were also found to be favourable at position 16 even though they were not required during library design. For library 2, the wild-type residue was the most energetically favourable at positions 175 and 177. Hydrophobic residues were predicted at each designed position, allowing design hypothesis II to be adequately tested. In addition to the required wild-type Gln and red-shifting Met at position 163, another hydrophobic residue (Ile) was predicted as well as Lys, the residue found in parent DsRed. The composition of libraries 3-1 and 3-2 was very similar but not identical, as expected, given the similarity of the required amino acids.

Library screening. The four mutant libraries were screened for emission wavelength red shifts using a 96-well plate fluorimetric assay. Screening of library 1 identified 7 mutants with red shifts ≥ 4 nm ($\lambda_{em} \geq 615$ nm) (Table S1). All retained the wild-type Tyr at position 120 and contained an H-bond donor (Ser or Thr) at positions 14 or 16. These findings correlate well with design hypothesis I. Neither Glu nor Gln at position 16 caused a red shift, which is seen with mPlum and its E16Q variant. However, it has been shown that Glu must be combined with F65I to induce a red shift in mPlum [12].

Although the point mutants F14S and F14N showed small red shifts, library members containing these mutations did not exhibit $\lambda_{em} \geq 615$ nm.

Screening of library 2 produced 16 mutants with red shifts ≥ 9 nm. All contained a hydrophobic residue at position 163 (Leu or Met), which correlates well with design hypothesis II. I161M and Q163M found in library 2 are also observed in mRaspberry and mPlum, which could partially account for their far red-shifted emission wavelengths. Neither the wild-type Gln nor Lys (found in DsRed) was found at position 163 in our most red-shifted mutants. All had Cys, Thr, or Ala at position 146, with Cys strongly preferred (14 of 16). Given the low propensity of Cys to form H-bonds, it is possible that no or weak H-bonding to the chromophore phenolate would shift electron density away from it and towards the imidazolinone and acylimine groups, resulting in a red shift. At position 143, nonpolar, polar, and aromatic amino acids were found, indicating no preference. Screening of library 3-1 identified 9 mutants with red shifts ≥ 9 nm, whereas only 1 was found for library 3-2. The brightest from these 2 libraries were the triple mutants A195-Y197-C217 (AYC) and T195-T197-N217 (TTN). Interestingly, only the small amino acids Ala, Val, and Thr were found at position 195 in red-shifted mutants from library 3-1 (Table S1), indicating that a small residue at this position may be required to accommodate the Y197 mutation. The larger amino acids Met and Ile, although included in the library (Table 1), were not found at this position in the most red-shifted mutants.

We next combined the best red-shifting mutations from each library to determine if the red-shifting effects were additive. Starting with either the AYC or TTN triple mutants from libraries 3-1 and 3-2 as templates, two combinatorial libraries were prepared. V16T

from library 1 and S146C/T, I161I/M, and Q163L/M from library 2 were introduced into both templates. Wild-type S146 was also included as it is important in maintaining the chromophore in its deprotonated state [26]. Thus, a total of 24 additional mutants were screened using the same 96-well plate assay. These second-generation mutants displayed red-shifted emission wavelengths of up to 26 nm, well over the 8-10 nm induced by AYC or TTN alone (Table 2). The results demonstrate that red-shifting mutations from the different libraries can have an additive effect and that this is independent of the template used. The observed additivity could arise from the fact that these mutations cause red shifts through independent processes. For example, the V16T mutation, which satisfies design hypothesis I, caused a 4 nm red shift on either the AYC or TTN templates. Moreover, mutants containing additional mutations that satisfy design hypothesis II (I161M, Q163M, Q163L) displayed even greater red shifts, adding 9-18 nm to the red shifts induced by the template. Noticeably, proteins with combinations of mutations from all three libraries had decreased quantum yields, leading to lower brightness.

Spectroscopic characterization. Absorption, excitation and emission spectra were measured for the most red-shifted mutants and are reported in Fig. S1. We identified three second-generation mutants with $\lambda_{em} > 630$ nm and named them mRojoA, mRojoB, and mRouge, after the Spanish and French words for red, respectively. Both mRojoA and mRojoB were derived from the AYC triple mutant, whereas mRouge was derived from TTN. These proteins differ from mCherry by 5-7 mutations, display λ_{em} of 631-637 nm, and have λ_{ex} near 600 nm, 9-12 nm higher than λ_{ex} of mCherry (Table 2, *SI text*, and Fig. S1). All three mutants achieve a longer λ_{em} than mRaspberry with fewer mutations away

from mCherry. This indicates that these mutations mostly result in λ_{em} bathochromic shifts, which is expected since we screened only for red-shifted emission.

Extinction coefficients are reported for each mutant in Table 2. The extinction coefficients determined for mCherry and mRaspberry differ from those reported in the literature (72,000 $M^{-1}cm^{-1}$ for mCherry and 86,000 $M^{-1}cm^{-1}$ for mRaspberry) [6, 12]. This discrepancy is caused by differences in the experimental conditions during the critical chromophore maturation phase. These conditions include the availability of soluble oxygen, the time allowed for maturation, and variations in temperature. Nonetheless, all the extinction coefficient values reported in Table 2 are for proteins purified in the same 96-well plate under identical maturation conditions. mRojoA, mRojoB, and mRouge exhibit much lower quantum yields (<0.10) than mCherry even though their extinction coefficients are not drastically smaller. In all cases, our mutants are not as bright as mCherry. This is not unexpected as we did not screen for improved brightness nor did our computational design target residues towards this goal. Since mRaspberry is also brighter than our mutants, we expect that this is accounted for by the extra mutations in mRaspberry (with respect to mCherry).

To better understand the role of each mutation on the observed spectral properties, 11 point mutants accounting for all the mutations found in mRojoA, mRojoB, and mRouge were prepared and characterized. As shown in Table 2, mutations V16T, S146C, Q163M, I197T, and I197Y are largely responsible for the red shifts, while S146C, Q163L, I197T, and I197Y primarily account for the decreases in quantum yield. Lower quantum yields could result from decreased rigidity of the chromophore. It is unclear how these mutations affect this property. A217N is the only point mutant that causes a

hypsochromic shift. The Q163M and A217N point mutants display higher quantum yields, but A217C is the only single mutant that is as bright as wild-type mCherry.

Causes of red shifts in RFPs. Crystal structures were solved for our two most red-shifted mCherry variants, mRojoA and mRouge (Table S2 and *SI text*). The conformational predictions of the CPD software agree well with the crystal structures obtained (Fig. S2), as indicated by the atomic RMSDs for the 13 designed residues (0.30 ± 0.02 Å for mRojoA and 0.23 Å for mRouge). Analysis of the crystal structures demonstrates that two of the three design hypotheses were clearly achieved. Hypothesis I (Fig. 1*A*) was fulfilled by the T16 mutation found in both variants, which provides a water-mediated H-bond with the acylimine group of the chromophore (Fig. 2). This interaction should stabilize the resonance form of the chromophore in which negative charge is localized on the acylimine oxygen. A similar interaction was demonstrated to red-shift the emission wavelength in Neptune [14]. Hypothesis III (Fig. 1*C*) was fulfilled by the I197Y mutation in mRojoA, which causes a clearly observable π-stacking interaction with the chromophore phenolate group (Fig. 2*A*), similar to that observed in YFP [22]. The I197Y mutation has been shown to contribute to red-shifted emission wavelength in a series of recently reported mFruits termed mGrapes [27], presumably through a similar process. Additionally, a recently reported variant of DsRed-Express2 called E2-Crimson [28] contains the S197Y mutation, which along with Q66F is largely responsible for the observed red shift ($\lambda_{em} = 646$ nm).

It is more difficult to determine whether design hypothesis II (increasing hydrophobicity around chromophore phenolate) was fulfilled. Compared to mCherry, mRojoA and mRouge have more hydrophobic residues at position 163 directly above the chromophore

phenolate ring (Leu and Met, respectively), replacing the more polar Gln found at this position in mCherry. M163 is found in the two far-red mFruits, mRaspberry and mPlum, suggesting that it is involved in the red shifting observed. The measured pKa values for mRojoA, mRojoB, and mRouge are higher than for mCherry (Table 2, *SI text*, and Fig. S3), indicating that the hydroxyl hydrogen of the chromophore phenol group is less acidic. Since a more hydrophobic environment would destabilize the ionized form of the chromophore phenol, a higher pKa would be an expected result of satisfying design hypothesis II.

Far-red fluorescent proteins. The quest for higher emission wavelength monomeric RFPs is ongoing given the enormous potential for applications in whole-body imaging of research model animals. Far-red RFPs (> 630 nm) that are already available include mKate/mKate2 [13, 15], HcRed [29], RFP639 [3], E2-Crimson [28], mPlum [12], Neptune/mNeptune [27], AQ143 [30], and TagRFP657 [31]. With λ_{em}s ranging from 630-657 nm, most of these far-red RFPs have been successfully used in imaging experiments. However, they all have disadvantages, such as pKa's near physiological pH (mKate/mKate2), oligomeric states (RFP639, HcRed, E2-Crimson, and AQ143), monomer-dimer equilibriums (Neptune/mNeptune), slower maturation (TagRFP657), low brightness (HcRed, AQ143), and incomplete maturation (mPlum). Recently, monomeric bacterial phytochrome-derived FPs that emit in the IR region (>700 nm) have been developed [20]. These IR-FPs are very useful but require an exogenous cofactor for fluorescence. Thus, engineering GFP homologues that emit brightly in the far-red or near-IR region of the spectrum is still highly desirable. Until now, using structure-based rational design to increase emission wavelength has proven difficult. For this reason,

most far-red FPs were discovered by screening very large libraries of mutants generated by random mutagenesis in a directed evolution approach.

Here, using CPD as a pre-screen, we combined structure-based rational design with experimental screening to increase the emission wavelength of mCherry, a monomeric RFP. The computational pre-screen discarded mutations that were incompatible with the protein fold, allowing us to drastically decrease the size of libraries required for experimental screening. We were thus able to achieve a 26 nm red shift by screening libraries several orders of magnitude smaller than those previously used to achieve comparable red shifts in other RFP scaffolds. Performing 2 rounds of combinatorial mutagenesis, we screened a total of ~5000 clones to identify several mutants with λ_{em}s > 630 nm. In contrast, starting from mRFP1.2 (whose λ_{em} of 612 nm is similar to that of our starting structure), Wang *et al*. performed 10 rounds of somatic hypermutation random mutagenesis and screened millions of cells to obtain mRaspberry (λ_{em} = 625 nm, Φ_F = 0.15) [12]. Perhaps the best comparison to mRaspberry is our 3-2.A9 mutant, which is red shifted by the same amount (λ_{em} = 623 nm) and has a similar Φ_F (0.11) (Table 2). We were able to isolate 3-2.A9 by combining 3 mutations (found after screening only ~700 clones from library 3-2) with 1 mutation (V16T) that we had previously identified through site-saturation mutagenesis. To obtain mPlum from mRaspberry, Wang *et al*. performed an additional 13 rounds of somatic hypermutation, screening millions of clones [12].

The development of the far-red-shifted RFP mKate was guided by identification of an optimized dimeric mutant of eqFP578 termed TurboRFP. Using 4 rounds of random

mutagenesis and screening >100,000 clones, Shcherbo *et al.* isolated an RFP with an λ_{em} of 635 nm termed Katushka [15]. The red-shifting mutations of Katushka were then inserted into a monomeric variant of TurboRFP, yielding mKate. TagRFP657, a further optimized mKate mutant with an λ_{em} of 657 nm, is currently the most red-shifted monomeric RFP [31]. Recently, Strack *et al.* [28] developed E2-Crimson, a very red-shifted DsRed-derived RFP. E2-Crimson was isolated after screening more than 1,000,000 colonies in 3 rounds of targeted combinatorial mutagenesis and 2 rounds of random mutagenesis. Thus, directed evolution approaches, though generally more time-consuming and costly than CPD-based approaches, have yielded several useful RFPs.

Generating targeted, computationally designed combinatorial libraries has a distinct advantage in that it allows one to identify combinations of mutations that would have been difficult to predict rationally or to obtain through random mutagenesis (which often builds on point mutations discovered in individual rounds). Synergistic effects are more easily obtained by a semi-rational approach involving combinatorial mutagenesis [32]. For instance, under other experimental circumstances the A217N mutation found in the TTN triple mutant motif described earlier would have been discarded during screening as it causes a hypsochromic shift. However, these three mutations together resulted in an 8 nm bathochromic shift. Another example of synergistic effects is seen in mutant 3-1.A3 (Table 2), which contains the red-shifting but quantum yield-reducing I197Y mutation. The presence of V195A and A217C compensate for the decrease in quantum yield, bringing it up from 0.03 to 0.08. Since neither A217C nor V195A increase the quantum yield by themselves (Table 2), this increase is the result of the interplay between these

three residues. Synergistic effects are similarly seen in mutant 3-2.A3, likely due to the presence of the quantum yield-increasing A217N.

Conclusion

Using a structure-based rational approach that combines CPD with experimental screening, we were able to identify three mutants exhibiting emission red shifts of 20–26 nm: mRojoA, mRojoB, and mRouge. Although these mutants are not as bright or as red-shifted as other useful RFPs, these results suggest that this approach may be applicable to red shift the emission wavelength of other RFPs. It could also be used to further red shift the recently engineered phytochrome-based IR-FPs. We expect that other useful properties, such as increased quantum yield and maturation rate, could also be improved using this method. Additionally, red-shifted mutants developed using our design procedures could serve as templates for optimization of other properties such as brightness through random mutagenesis, thereby combining the benefits of both rational design and directed evolution.

Materials and Methods

Computational design. Four independent libraries were computationally designed using the procedure described in *SI text*. Residues 14, 16, and 120 of mCherry were designed in library 1. Residues 143, 146, 161, 163, 175, and 177 were designed in library 2. Residues 73, 146, 195, and 217 were designed in libraries 3-1 and 3-2. These latter two libraries differ by the presence of a Tyr (library 3-1) or Thr (library 3-2) residue at position 197. All 20 proteinogenic amino acids were allowed at designed positions; residues with side

chains pointing towards and within 4 Å of the designed residues were allowed to sample alternative conformations during the design, but their identities were not modified. The crystallographic conformer at each designed position was also allowed. A standard backbone-dependent side chain rotamer library [33] with expansions by one standard deviation about χ_1 and χ_2 was used. Prior to our design procedure, we generated an *in silico* structure of mCherry mutant I197Y in which residue Y197 is stacked next to the chromophore phenolate group. To generate this structure, which would serve as the input structure for computational design in library 3-1, we sampled different conformations of Y197 using a large backbone-independent conformer library [34]. During computational design of library 3-1, the conformation of the π-stacked Y197 residue was not allowed to vary. The energy function used is described in *SI text*.

Screening. The DNA libraries prepared as described in *SI text* were transformed into chemically competent *E. coli* BL21-Gold(DE3) cells (Stratagene). Colonies were picked into individual wells of Nunc V96 MicroWell polypropylene plates containing 200 μL of medium (LB with 100 μg/mL ampicillin supplemented with 10% glycerol). The plates were covered with a sterile Breathe-Easy gas permeable sealing membrane (Sigma) and incubated overnight at 37 °C with shaking. After incubation, these mother plates were used to inoculate sterile Nunc V96 MicroWell polypropylene plates ("daughter" plates) containing 300 μL of Overnight Express Instant TB media (Novagen) supplemented with ampicillin per well. Daughter plates were sealed with breathable membranes and incubated overnight (37 °C, 250 rpm). After incubation, the cells were harvested by centrifugation and the cell pellets were washed twice with PBS (pH 7.4). Washed cell pellets were then incubated at 4 °C for 72 h to allow chromophore maturation. These

pellets were resuspended in PBS and transferred to a Fluotrac 96-well plate (Greiner Bio-One) for screening. Screening was performed using a Tecan Safire2 plate reader equipped with a plate stacker. Emission spectra (λ_{ex} = 565 nm) and excitation spectra (λ_{em} = 650 nm) were measured. Purification and spectroscopic characterization of mutants is described in *SI text*.

References

1. Marques MAL, *et al.* (2003) Time-dependent density-functional approach for biological chromophores: The case of the green fluorescent protein. *Physical Review Letters* 90:258101.

2. Jach G, *et al.* (2001) Use of red fluorescent protein from Discosoma sp. (DsRed) as a reporter for plant gene expression. *Plant Journal* 28(4):483-491.

3. Kredel S, *et al.* (2008) Optimized and far-red-emitting variants of fluorescent protein eqFP611. *Chemistry & Biology* 15:224-233.

4. Mizuno H, *et al.* (2001) Red fluorescent protein from Discosoma as a fusion tag and a partner for fluorescence resonance energy transfer. *Biochemistry* 40(8):2502-2510.

5. Campbell RE, *et al.* (2002) A monomeric red fluorescent protein. *Proceedings of the National Academy of Sciences, USA* 99:7877-7882.

6. Shaner N, *et al.* (2004) Improved monomeric red, orange and yellow fluorescent proteins derived from Discosoma sp. red fluorescent protein. *Nature Biotechnology* 22(12):1567-1572.

7. Wall MA, Socolich M, & Ranganathan R (2000) The structural basis for red fluorescence in the tetrameric GFP homolog DsRed. *Nature Structural Biology* 7(12):1133-1138.

8. Pletneva N, *et al.* (2006) Structure of a red fluorescent protein from Zoanthus, zRFP574, reveals a novel chromophore. *Acta Crystallographica Section D: Biological Crystallography* 62(Pt 5):527-532.

9. Merzlyak E, *et al.* (2007) Bright monomeric red fluorescent protein with an extended fluorescence lifetime. *Nature Methods* 4(7):555-557.

10. Baird GS, Zacharias DA, & Tsien RY (2000) Biochemistry, mutagenesis, and oligomerization of DsRed, a red fluorescent protein from coral. *Proceedings of the National Academy of Sciences, USA* 97:11984-11989.

11. Wiedenmann J, *et al.* (2002) A far-red fluorescent protein with fast maturation and reduced oligomerization tendency from Entacmaea quadricolor (Anthozoa, Actinaria). *Proceedings of the National Academy of Sciences, USA* 99(18):11646-11651.

12. Wang L, *et al.* (2004) Evolution of new nonantibody proteins via iterative somatic hypermutation. *Proceedings of the National Academy of Sciences, USA* 101:16745-16749.

13. Shcherbo D, *et al.* (2009) Far-red fluorescent tags for protein imaging in living tissues. *Biochemistry Journal* 418(3):567-574.

14. Lin MZ, *et al.* (2009) Autofluorescent proteins with excitation in the optical window for intravital imaging in mammals. *Chemistry & Biology* 16(11):1169-1179.

15. Shcherbo D, *et al.* (2007) Bright far-red fluorescent protein for whole-body imaging. *Nature Methods* 4(9):741-746.

16. Ormo M, *et al.* (1996) Crystal structure of the Aequorea victoria green fluorescent protein. *Science* 273:1392-1395.

17. Olsen S & Smith SC (2007) Radiationless decay of red fluorescent protein chromophore models via twisted intramolecular charge-transfer states. *Journal of the American Chemical Society* 129:2054-2065.

18. Tozzini V & Nifosi R (2001) *Ab Initio* molecular dynamics of the green fluorescent protein (GFP) chromophore: An insight into the photoinduced dynamics of green fluorescent proteins. *Journal of Physical Chemistry B* 105(24):5797-5803.

19. Wood TI, *et al.* (2005) Defining the role of arginine 96 in green fluorescent protein fluorophore biosynthesis. *Biochemistry* 44:16211-16220.

20. Shu X, *et al.* (2009) Unique interactions between the chromophore and glutamate 16 lead to far-red emission in a red fluorescent protein. *Protein Science* 18:460-466.

21. Griesbeck O, *et al.* (2001) Reducing the environmental sensitivity of yellow fluorescent protein: Mechanism and applications. *Journal of Biological Chemistry* 276:29188-29194.

22. Wachter RM, *et al.* (1998) Structural basis of spectral shifts in the yellow-emission variants of green fluorescent protein. *Structure With Folding & Design* 6:1267-1277.

23. Allen BD, Nisthal A, & Mayo SL (2010) Computational protein design of structural ensembles: validation by automated screening of combinatorial libraries. *Proceedings of the National Academy of Sciences, USA*.

24. Eli P & Chakrabartty A (2006) Variants of DsRed fluorescent protein: Development of a copper sensor. *Protein Science* 15(10):2442-2447.

25. Henderson JN & Remington SJ (2005) Crystal structures and mutational analysis of amFP486, a cyan fluorescent protein from Anemonia majano. *Proceedings of the National Academy of Sciences of the United States of America* 102:12712-12717.

26. Yarbrough D*, et al.* (2001) Refined crystal structure of DsRed, a red fluorescent protein from coral, at 2.0-A resolution. *Proceedings of the National Academy of Sciences, USA* 98(2):462-467.

27. Lin MZ*, et al.* (2009) Autofluorescent proteins with excitation in the optical window for intravital imaging in mammals. *Chemistry & Biology* 16(11):1169-1179.

28. Strack RL*, et al.* (2009) A rapidly maturing far-red derivative of DsRed-Express2 for whole-cell labeling. *Biochemistry* 48(35):8279-8281.

29. Gurskaya NG*, et al.* (2001) GFP-like chromoproteins as a source of far-red fluorescent proteins. *FEBS Letters* 507:16-20.

30. Shkrob M*, et al.* (2005) Far-red fluorescent proteins evolved from a blue chromoprotein from Actinia equina. *Biochemistry Journal* 392(3):649.

31. Morozova KS*, et al.* (2010) Far-Red Fluorescent Protein Excitable with Red Lasers for Flow Cytometry and Superresolution STED Nanoscopy. *Biophysical Journal* 99(2):L13-L15.

32. Chica R, Doucet N, & Pelletier J (2005) Semi-rational approaches to engineering enzyme activity: Combining the benefits of directed evolution and rational design. *Current Opinion in Biotechnology* 16(4):378-384.

33. Dunbrack RL, Jr. & Cohen FE (1997) Bayesian statistical analysis of protein side-chain rotamer preferences. *Protein Science* 6(8):1661-1681.

34. Lassila JK*, et al.* (2006) Combinatorial methods for small-molecule placement in computational enzyme design. *Proceedings of the National Academy of Sciences, USA* 103:16710-16715.

Figures & Legends.
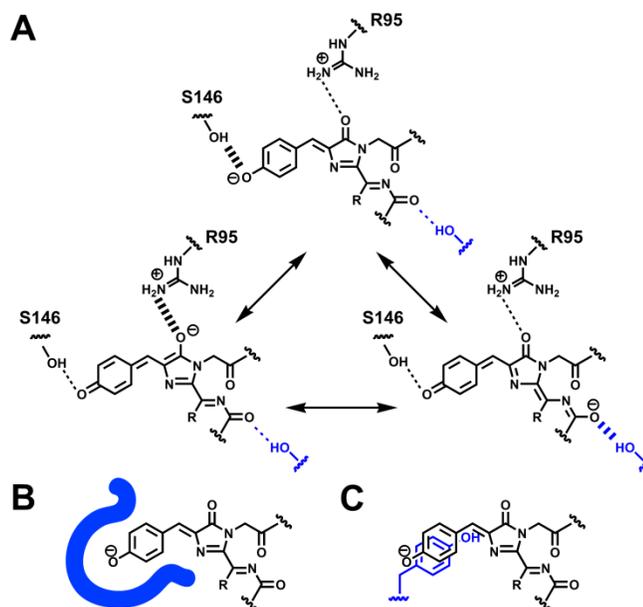
Fig. 1. Structure-based design hypotheses tested on mCherry scaffold. (*A*) Design hypothesis I: stabilization of the excited state form on lower right by H-bonding to acylimine oxygen. H-bonds that stabilize each resonance structure are indicated by wide dashed lines. (*B*) Design hypothesis II: destabilization of ground state by hydrophobic packing around charged phenolate. (*C*) Design hypothesis III: stabilization of polarization of excited state by π-stacking the chromophore phenolate with a tyrosine. (*A-C*) Designed interactions are shown in blue.
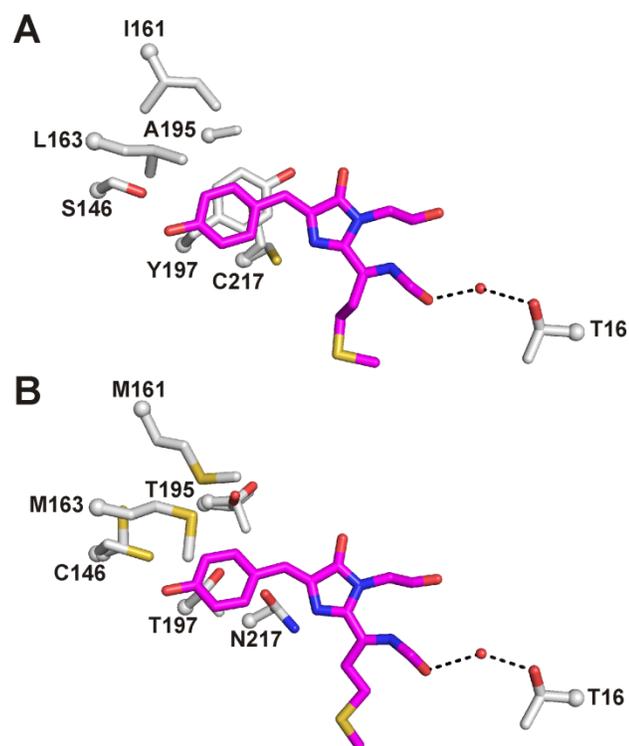
Fig. 2. Crystal structures of designed positions in the most red-shifted mutants, mRojoA (*A*) and mRouge (*B*). The chromophore is shown in magenta. Residue C217 in mRojoA (*A*) and residues C146 and T195 in mRouge (*B*) exhibit two conformations in the crystal structure. H-bonds are indicated by dashed lines.

Table 1. Computational library design results

Library	Design hypothesis tested	Protein sequences sampled	Library size*	Designed position†	Wild-type residue	Required amino acids	Library amino acids
1	I. Stabilization of the excited state by H-bonding to acylimine oxygen	8×10^3	540	14	F	FNS	FINSTY
				16	V	STV	ADEGHIKLNPQRSTV
				120	Y	Y	FHILNY
2	II. Destabilization of ground state by hydrophobic packing around charged phenolate	6.4×10^7	520	143	W	MW	ACFGILMPRSTVW
				146	S	CST	ACGST
				161	I	IM	IM
				163	Q	MQ	KLMQ
				175	A	A	A
				177	V	V	V
3-1	III. Stabilization of polarization of excited state by π-stacking the chromophore phenolate with Tyr197	1.6×10^5	250	73	V	V	IV
				146	S	CST	ACGST
				195	V	V	AIMTV
				217	A	ACS	ACGST
3-2‡		1.6×10^5	240	73	V	V	AV
				146	S	CGS	ACGST
				195	V	V	AITV
				217	A	AS	ADGNST

* Defined by user.
† Numbering based on DsRed.
‡ Red-shifting mutation Thr197 was included and stabilized by designing neighbouring residues.

Table 2. Spectral properties of various red fluorescent proteins

Protein	λ_{ex} (nm)	λ_{em} (nm)	λ_{em} shift from mCherry (nm)	Extinction coefficient ($M^{-1}cm^{-1}$)	Φ_F	Relative brightness* (% mCherry)	pKa	Mutations away from mCherry	Mutations						
									Library 1	Library 2			Libraries 3-1 and 3-2		
									16	146	161	163	195	197	217
mFruits															
mCherry	588	611	-	101,000	0.22	100	4.7	-	V	S	I	Q	V	I	A
mRaspberry	600	625	+14	62,000	0.15	42	-	10[†]	-	-	M	M	T	-	-
AYC template															
3-1.A3	596	621	+10	63,000	0.08	23	-	3	-	-	-	-	A	Y	C
3-1.A6	595	625	+14	61,000	0.06	16	-	4	T	-	-	-	A	Y	C
mRojoA	597	633	+22	48,000	0.02	4	5.4	6[‡]	T	-	-	L	A	Y	C
mRojoB	598	631	+20	61,000	0.06	16	5.0	5	T	-	-	M	A	Y	C
TTN template															
3-2.A3	596	619	+8	56,000	0.13	33	-	3	-	-	-	-	T	T	N
3-2.A9	596	623	+12	39,000	0.11	19	-	4	T	-	-	-	T	T	N
3-2.H11	597	628	+17	28,000	0.03	4	-	5	T	-	-	L	T	T	N
mRouge	600	637	+26	43,000	0.02	4	6.1	7	T	C	M	M	T	T	N
Point mutants															
V16T	590	615	+4	73,000	0.20	66	-	1	T	-	-	-	-	-	-
S146C	593	619	+8	50,000	0.06	14	-	1	-	C	-	-	-	-	-
I161M	590	614	+3	89,000	0.16	64	-	1	-	-	M	-	-	-	-
Q163L	590	614	+3	73,000	0.09	30	-	1	-	-	-	L	-	-	-
Q163M	590	615	+4	65,000	0.25	73	-	1	-	-	-	M	-	-	-
V195A	588	610	-1	48,000	0.21	45	-	1	-	-	-	-	A	-	-
V195T	590	613	+2	71,000	0.23	73	-	1	-	-	-	-	T	-	-
I197Y	598	618	+7	79,000	0.03	11	-	1	-	-	-	-	-	Y	-
I197T	587	616	+5	79,000	0.05	18	-	1	-	-	-	-	-	T	-
A217C	588	613	+2	101,500	0.22	100	-	1	-	-	-	-	-	-	C
A217N	568	597	-14	31,000	0.28	39	-	1	-	-	-	-	-	-	N

* Brightness is calculated as the product of the extinction coefficient and the fluorescence quantum yield.

[†] This number does not include the N- and C-terminal sequences coming from GFP and other spacer sequences.

[‡] Contains an additional undesired mutation (R125H).

Supporting Information

SI Results and Discussion

Absorption spectroscopy. The absorption spectrum of mRojoB is unique among the most red-shifted mutants in that it has an additional peak centered at around 510 nm (Fig. S1*C*). This peak has been suggested to correspond to the deprotonated green form of the chromophore [1]. The green form is equivalent to the *p*-hydroxybenzylidene-imidazolinone group shown in Fig. 1 without the acylimine substitution on the five-membered ring. The deprotonated green form has been postulated to be a dead-end product created during chromophore maturation [2]. The 510 nm peak is absent in the excitation spectrum of mRojoB monitored at 650 nm, indicating that the species associated with this peak does not contribute to far-red fluorescence. This result further validates that the 510 nm peak does not correspond to the acylimine substituted red form of the chromophore. The 510 nm peak is also found in the absorption spectrum of mRFP1 [3] from which mCherry is derived, as well as in the spectrum of mCherry point mutant Q163M (Fig. S1*E*). The 510 nm species could be caused by the combination of residues M163 and I161, which is found in mRojoB, mRFP1, and mCherry mutant Q163M, but is absent in mCherry, mRojoA, and mRouge.

mRouge has an additional absorbance peak at ~390 nm (Fig. S1*D*) which could correspond to the protonated green form of the chromophore. This peak is also absent in the excitation spectrum monitored at 650 nm. The protonated green chromophore could be favoured in this protein due to the S146C mutation, since the substitution of a Cys at

this position would significantly increase the pKa of the phenol group of the chromophore, given the low H-bonding propensity of thiols. In line with this supposition, it is significant to note that the measured pKa of the red chromophore in mRouge is one pH unit higher than that of mRojoB (Table 2, Fig. S3). Since the microenvironments of red and green chromophore species will be identical in these cases, it is safe to assume that the pKa of the green chromophore in mRouge will also be higher than that of mRojoB. This implies that the assignment of the ~390 nm peak for mRouge to the protonated green chromophore is correct.

Crystal Structure of mRojoA. The crystal structure of mRojoA (PDB accession code 3NEZ) was solved at 1.70 Å resolution by molecular replacement using the structure of mCherry (PDB code: 2H5Q [4]) (Table S2). mRojoA crystallized in space group $P12_11$ with unit cell dimensions of a = 61.2 Å, b = 97.4 Å, c = 84.2 Å. The asymmetric unit consisted of four molecules organized into two pairs of dimers that slant toward each other to form an A-frame-like tent structure. The tetrameric association of protomers in mRojoA is significantly different than the tight tetramer seen in DsRed and other class Anthozoa FPs [5, 6]. Superimposition of the backbone atoms of individual molecules of mRojoA with mCherry yielded an average RMSD of 0.21 ± 0.01 Å.

Treatment of the mRojoA chromophore was left until the end of refinement. Weak electron density for the phenolate group of the chromophore necessitated building the chromophore one atom at a time between refinement cycles. Building of the chromophore revealed a mixture of mature and immature species. The mature species is the acylimine-containing red chromophore. It is characterized by a *cis* peptide bond between F65 and M66. The immature species corresponds to the green chromophore,

which is characterized by a *trans* peptide bond between F65 and M66. Occupancy refinement of the red and green forms was performed as described for DsRed [7]. Each associated dimer in the structure was found to have one chromophore with ~70%/30% red/green split and one chromophore with ~50%/50% red/green split. The presence of the green chromophore indicates incomplete maturation. No evidence for the *trans* configuration of the chromophore phenolate ring was observed during refinement.

The H-bonding network around the chromophore of mRojoA is illustrated in Fig. S4*A*. The conserved H-bond between the chromophore imidazolinone O2 atom and catalytic R95 is present. The other catalytic residue, E215, makes a close H-bond with the chromophore imidazolinone N2 atom through $O_{\epsilon 1}$ with a similar configuration to what is seen in mCherry [4]. The $O_{\epsilon 2}$ atom of E215 shows an H-bond to a conserved crystallographic water above the imidazolinone ring of the chromophore. The H-bond between S146 and the chromophore phenolate oxygen atom present in mCherry is also present in mRojoA. This H-bond has been suggested to stabilize the deprotonated phenolate form of the chromophore [8]. Additionally, a water-mediated H-bond between residue T16 and the acylimine oxygen atom is observed (Fig. 2*A*). The interoxygen distances between the acylimine and water, and between water and T16 in monomer A were 2.5 Å and 2.9 Å, respectively. This water-mediated H-bond differs from the one observed in the crystal structure of Neptune [9]. The H-bond in Neptune involves residue S28, analogous to G31 in mRojoA, and is located on the opposite side of the acylimine oxygen.

The structure of mRojoA shows the same type of chromophore π-stacking interaction observed in GFP-derived YFPs [10-12]. In mRojoA, π-stacking occurs with Y197. The

centroid to centroid distance from the chromophore phenolate to the Y197 side chain in mRojoA is 3.9 Å, and the angle between the normals of these planes is ~8°. These values are comparable to those found in the crystal structure of the yellow fluorescent protein citrine (PDB code: 3DQO [10]) with a centroid to centroid distance of 3.6 Å and an angle of 6°.

Crystal Structure of mRouge. The crystal structure of mRouge (PDB accession code 3NED) was solved at 0.95 Å resolution using direct methods (Table S2). mRouge crystallized in space group $P12_11$ with unit cell dimensions of a = 48.9 Å, b = 42.9 Å, c = 61.3 Å. Unlike mRojoA, the structure of mRouge contains only one protein molecule per asymmetric unit. Superimposition of the protein with mCherry gives a backbone-atom RMSD of 0.10 Å.

For mRouge, in-depth treatment of the chromophore was again left until the end of the refinement procedure. Throughout most of the refinement only the imidazolinone heterocycle was fit due to very weak electron density corresponding to the Tyr-derived phenol ring of the chromophore (Fig. S4*B*). After refinement on the rest of the protein was completed, significant electron density in the difference map was observed within covalent bonding distance of the chromophore C1 atom. Prior studies on GFP mutants suggested that this density corresponds to hydroxylation at the C1 position [13-15]. Moreover, a recent investigation of the maturation pathway for DsRed (the wild-type parent of mRouge) indicated that red chromophore maturation likely proceeds through an intermediate in which the acylimine bond preceding the chromophore is already oxidized and there is hydroxylation at the chromophore C1 position [16]. Evidence for the presence of this putative intermediate in the mRouge crystal structure was strengthened

by trypsin-digest mass spectrometry data, which showed a peptide fragment containing the chromophore tripeptide at the appropriate mass (−4 Da from the uncyclized chromophore tripeptide) (Fig. S5C). Mass spectrometry analysis also demonstrated the presence of the green chromophore in mRouge (Fig. S5B), which had already been observed spectroscopically (Fig. S1D). Given the spectroscopic evidence for both red and green chromophores in mRouge along with the mass spectrometry evidence and crystallographic indication for a hydroxylated species, all three of these molecular entities were modeled into the crystal structure of mRouge. Occupancy refinement yielded 22% for the red chromophore, 45% for the green chromophore, and 33% for the hydroxylated species. No evidence for the *trans* configuration of the chromophore phenolate ring was observed during refinement in any of the difference maps.

For mRouge, many of the residues that form key interactions with the chromophore occur in multiple conformations (Fig. S4B). For example, C146 has two conformations occupied at 56% and 44%. In many RFPs (including mCherry and mRojoA), a Ser residue at this position is H-bonded to the chromophore phenolate OH atom [4, 5, 17]. This interaction stabilizes the anionic chromophore, which is recognized as the dominant fluorescent species in FPs [5, 17-20]. The major conformer of C146 points away from the chromophore. However, in the minor conformation, the sulfur atom points toward the chromophore in a similar conformation to what is observed for S146 in mCherry. Considering the low H-bonding strength of thiols, it is unlikely that this interaction represents a significant H bond.

K70 has two conformations in mRouge (populated at 82% and 18%), with the major conformer pointing away from the bridging CB2 carbon of the chromophore and making

an H-bond to T195. The minor conformer points toward the CB2 atom and forms a salt bridge with E148, as seen for K70 in mCherry [4].

Another position of interest with multiple conformations in mRouge is the catalytic residue E215 (populated at 75% and 25%). E215 is speculated to be deprotonated in DsRed [21], but to be protonated in mCherry because its $O_{\epsilon 1}$ atom comes within H-bonding proximity to the chromophore N2 atom in mCherry [4]. The major conformation of E215 in mRouge is most similar to the conformation of that same residue in mCherry (Fig. S4*B*), with its $O_{\epsilon 1}$ oxygen forming an H-bond with the N2 atom of the chromophore. This implies that the major conformer of E215 in mRouge is protonated, analogous to what is seen in mCherry.

It is worth noting that the minor conformer of E215 in mRouge (Fig. S4*B*) is not in the same conformation as that seen for E215 in DsRed. In DsRed, E215 is oriented in such a way as to create a salt bridge with K70 across one face of the chromophore ring system [5, 7]. The minor conformer of the E215 residue in mRouge attains a previously unobserved conformation for DsRed variants. In this conformation, the $O_{\epsilon 1}$ atom is within H-bonding distance of both the chromophore N2 atom and the hydroxyl group attached to the chromophore C1 atom (2.4 Å and 2.6 Å, respectively). This conformer of E215 mainly differs from the protonated, major conformer in mRouge by a ~90° twist about the side chain χ_3 angle.

A water-mediated H-bond between residue T16 and the acylimine oxygen atom is also present in the structure of mRouge (Fig. 2*B*). It is nearly identical to the one observed in

the structure of mRojoA, with interoxygen distances between the acylimine and water, and between water and T16 of 3.0 Å and 2.7 Å, respectively.

A sequence and structural motif consisting of T195-T197-N217 occurs in mRouge. This set of mutations leads to an 8 nm red shift in the emission spectra of mCherry mutants (Table 2). In mRouge, the O$_{\epsilon 2}$ oxygen of the major conformer of E215 is H-bonded to the terminal nitrogen on the side chain amide of N217 (Fig. S6). Moreover, the side chain amide oxygen from N217 forms an H-bond with the hydroxyl group of T197. Finally, the side chain of T197 forms a water mediated H-bond with the side chain of the major conformer of T195. The means by which this set of interactions causes red-shifted emission is unclear. However, the network of H-bonds resulting from this set of mutations may explain the synergistic effects on emission wavelength encountered by including them in the mCherry scaffold.

SI Materials and Methods

Materials. All reagents used were of the highest available purity. Restriction enzymes and DNA-modifying enzymes were from New England Biolabs. Synthetic oligonucleotides were obtained from Integrated DNA Technologies, and Ni-NTA agarose resin was obtained from Qiagen. CelLytic B buffer and lysozyme were purchased from Sigma-Aldrich. All aqueous solutions were prepared using water purified with a Millipore BioCell system.

Computational design. Hydrogens were added to the crystal structure of mCherry (PDB code: 2H5Q) using Molprobity [22]. Following removal of all water molecules and ions,

any strain or steric clashes in the structure were removed by performing 50 steps of conjugate gradient energy minimization [23]. Partial atomic charges for the chromophore were parameterized using the charges described by Sitkoff *et al.* [24]. Computational design was performed using the PHOENIX protein design software. The energy function used was based on the DREIDING force field [23] and included a scaled van der Waals term [25], hydrogen bonding and electrostatic terms [25], and terms for implicit solvation and phi-psi propensities. Implicit solvation energies were evaluated using a model based on occluded volume [26] described below. Amino acid phi-psi propensities were derived and applied following the method of Shortle [27]. Sequence optimization was carried out with FASTER; a Monte Carlo-based algorithm was then used to sample sequences around the minimum energy configurations identified by FASTER [28] and generate a list of high-scoring sequences. This was followed by a computational library design step in which combinatorial sequence libraries were defined as described by Allen *et al.* [29]. Briefly, based on the list of scored sequences generated by CPD and the list of required amino acids specified by the user, the algorithm determines the library composition that represents the best set of top-scoring sequences that can be encoded by a single degenerate codon at each position for a desired range of library sizes.

Implicit solvation energies. To account for the contributions of solvent to the free energies of folding, we applied an implicit solvation potential inspired by the methods of Dahiyat and Mayo [25] and Lazaridis and Karplus [26]. The potential is intended to reward the burial of nonpolar groups, penalize the burial of polar groups, and penalize the exposure of nonpolar groups. In this scheme, atomic groups are scored based on how

their volumes are occluded by other groups in the protein structure. We used a Gaussian function to compute the occlusion O_i of atom i by other atoms j:

$$O_i = \sum_j V_j e^{-\frac{R_{ij}^2}{2\lambda^2}}$$

where V_j is the volume of atom j, R_{ij} is the distance between i and j, and λ is the constant correlation length. We define the fractional exposure Θ_i for atom i as:

$$\Theta_i = \frac{O_{i,\max} - O_i}{O_{i,\max} - O_{i,\min}}$$

where the O_i minima and maxima are found from a database survey of protein crystal structures. When $O_i = O_{i,\max}$, atom i is fully buried, and $\Theta_i = 0$; when $O_i = O_{i,\min}$, atom i is fully exposed, and $\Theta_i = 1$.

The nonpolar solvation energy of atom i is given by:

$$E_{np,i} = s_i \sigma_{np} (\kappa_{np} + 1)\Theta_i - s_i \sigma_{np}$$

where s_i is a scaling parameter specific to the amino acid type and atom type for i, σ_{np} is the nonpolar desolvation energy benefit, and κ_{np} is the nonpolar exposure scale factor. This formula allows the balance between favourable nonpolar desolvation and unfavourable nonpolar exposure to be tuned between $s_i \kappa_{np} \sigma_{np}$ energy units when i is fully exposed, and $-s_i \sigma_{np}$ energy units when fully buried. For each nonpolar atom i, s_i is the mode of exposed surface areas observed for the residue and atom type for i in truncated tripeptides with coordinates taken from proteins in a structural database previously used to generate conformer libraries [30]. In this way, nonpolar atoms with greater exposed surface area in the truncated tripeptide unfolded state model receive larger magnitude

energies than those with smaller unfolded state surface areas with the same fractional exposure value.

The formula for polar solvation is analogous to that for nonpolar solvation, but with opposite signs, indicating that the preference for exposure versus burial is reversed:

$$E_{p,i} = -s_i\sigma_p(\kappa_p + 1)\Theta_i + s_i\sigma_p$$

The s_i parameters for polar groups were derived via a linear fit between the Θ_i values and Poisson-Boltzmann reaction field energies calculated for side chains in a computationally tractable subset of the structural database.

Mutagenesis. The mCherry and mRaspberry genes were PCR-amplified from plasmids mCherry-pBAD and mRaspberry-pBAD (provided by R.Y. Tsien, UCSD) and subcloned into pET11-a (Novagen) via *Nde*I/*Bam*HI. The plasmids were then transformed into *E. coli* XL-1 Blue and BL21(DE3) cells. The entire *Nde*I/*Bam*HI fragments, including the whole coding region, were verified by DNA sequencing. All mutations were introduced into the mCherry gene by overlap extension mutagenesis [31] using VentR DNA polymerase. Briefly, external primers were used in combination with sets of complementary pairs of degenerate oligonucleotides containing the desired mutations in individual PCR reactions. The resulting overlapping fragments were gel-purified (Qiagen) and recombined by overlap extension PCR. The resulting amplicons were digested with *Nde*I/*Bam*HI, gel-purified, and ligated into pET11-a expression vector with T4 ligase. Library composition was verified by sequencing 96 clones per library (Agencourt Biosciences).

Protein expression and purification for spectral characterization. Mother plates containing the mutant libraries were used to inoculate 24-well culture plates (Whatman) containing 5 mL Overnight Express Instant TB media (Novagen) supplemented with ampicillin in each well. The 24-well plates were sealed with sterile Bugstopper venting capmats (Whatman) and incubated at 37 °C overnight with shaking. After expression, cells were harvested by centrifugation and washed twice with PBS. After maturation at 4 °C for one week, the cell pellets were resuspended in 400 µL lysis buffer (50 mM sodium phosphate buffer, pH 8.0, 300 mM NaCl, 2.5 mM imidazole, 1X CelLytic B, 1 mg/mL lysozyme, and 25 U/mL benzonase nuclease (Novagen)) and incubated at 30 °C for 30 min with shaking. After centrifugation, clarified lysates were recovered and proteins were purified by affinity chromatography using His-Select plates (Sigma) according to the manufacturer's protocol.

Protein expression and purification for crystallization. Protein was expressed in 1.0 L cultures by transformation of a pET11-a vector containing the gene of interest into *E. coli* BL21-Gold(DE3) and purified by Ni-NTA affinity chromatography according to the manufacturer's protocol. Column elutions were desalted by gel filtration using a Superdex 75 10/300 GL Tricorn resin column (GE Healthcare) into a final buffer solution of 50 mM phosphate buffer, pH 7.5, and 150 mM NaCl.

Spectroscopic characterization. Proteins purified as described above were quantified using the alkali denaturation method [32]. Briefly, RFPs were alkali-denatured with an equal volume of 2 M NaOH. It is known that the alkali-denatured RFP chromophore converts to a GFP-like one, with extinction coefficient 44,000 M^{-1} cm^{-1} at 452 nm under these conditions. Absorbance, emission, and excitation spectra were recorded in PBS

with a Tecan Safire2 plate reader. Path lengths for each well were calculated ratiometrically using the difference in absorbance of PBS at 900 nm and 998 nm. Based on the absorbance spectra of native proteins and the concentration determination of alkali-denatured proteins, molar extinction coefficients were calculated. For determination of quantum yields, the integrated fluorescence intensity of mutants of interest was compared with that of equally absorbing samples of mCherry and mRaspberry (quantum yields 0.22 and 0.15, respectively) with excitation at 550 nm.

pKa measurements. pH titrations were performed using a range of buffers from pH 2 to 9. Proteins were diluted into these buffers to a concentration of 5-10 µM. Fluorescence scans were taken at each pH value using a Tecan Safire2 plate reader. The Henderson-Hasselbach equation was used to calculate the pKa for each protein (Fig. S3).

Growth of Crystals and Screening. Light blue crystals of mRouge were grown in 1 µL × 1 µL hanging drops with a precipitant solution of 200 mM ammonium acetate, 100 mM Bis-Tris, pH 6.5, and 25% (w/v) polyethylene glycol 3350. Large multi-crystalline chunks (0.5 mm × 2.0 mm × 0.25 mm) were prodded with a nylon loop to break off smaller shards for isolation and collection of diffraction data. All the pieces screened were crystallographically identical in terms of space group and unit cell dimensions. Dark purple crystals of mRojoA were grown in hanging drops with 1 µL protein solution and 1 µL of the same precipitant used to crystallize mRouge. These long stick-like rhomboidal crystals of mRojoA had approximate dimensions of 1.0 mm × 0.02 mm × 0.02 mm.

Data Collection and Processing. For mRouge, a dataset was collected locally at 2.0 Å resolution, and another dataset resolved at 0.95 Å was collected at the Stanford Synchrotron Radiation Lightsource (SSRL) beamline 12-2. The low-resolution dataset

was integrated, merged, and scaled with IMOSFLM [33] and SCALA [34]. For the high-resolution dataset, IPMOSFLM [33] was used for integration and ACORN was used for merging and scaling. For mRojoA, crystals were sent to the SSRL where the dataset was collected at 1.70 Å. This dataset was integrated, merged, and scaled with IMOSFLM and SCALA.

Solution and Refinement of Crystal Structures. The 2.0 Å dataset of mRouge, and the 1.70 Å dataset of mRojoA were solved by molecular replacement using PHASERMR [35]. For mRouge and mRojoA, the search model used consisted of the PDB coordinates from mCherry (2H5Q [4]) with the chromophore removed. After the higher resolution dataset of mRouge was solved by direct methods, the initial coordinates from the 2.0 Å structure were used as a starting point for further refinement of the sub-atomic resolution structure.

Refinement was accomplished using REFMAC5 [36, 37] and PHENIX (Python-based Hierarchical ENvironment for Integrated Xtallography) [38]. PHENIX was used specifically for refinement of atomic occupancies. Model building was done with COOT [39], wherein water molecules were added manually when they were within H-bonding distance of other heteroatoms (2.3–3.5 Å) and had peaks in the $F_o - F_c$ map of greater than 3.5 σ. In addition, water molecules were removed when they had equivalent isotropic B-factors greater than 60–80 Å^2. During generation of R-factors, 5% of data was excluded for cross-validation with an R_{free} value. Crystallographic R-factors were calculated in the standard fashion ($R = \sum |F_{obs} - F_{calc}| / \sum F_{obs}$).

In the crystal structure of mRouge, refinement was done with anisotropic ADPs after the initial few cycles of refinement. In the crystal structure of mRojoA, the final refinement

steps were carried out with 20 translation-libration-screw (TLS) groups per protein molecule [40]. TLS groups were identified automatically by using the TLS-Motion Determination web server (TLS MD) [40]. Riding hydrogens were included in the refinement of all structures for non-water molecules, but were only retained in the final structure of mRouge due to its high resolution.

The library file for the chromophore was built based on the CH6 chromophore deposited in the Hetero-compound Information Centre - Uppsala (HIC-Up) online database. Appropriate constraints and atom types were added and/or edited to account for the alternative covalencies observed for different conformations of the chromophore.

Mass Spectrometry Analyses. Following separation by SDS-PAGE, the ~25 kDa band from a freshly purified sample of mRouge was excised and destained. Destaining of the Coomassie dye was accomplished by a 100 μL wash of 50 mM ammonium bicarbonate followed by a 50 μL wash of a 1:1 mixture of 50 mM ammonium bicarbonate and acetonitrile; this process was repeated for a total of three times. After destaining, the gel band sample was reduced with 25 μL of 50 mM ammonium bicarbonate plus 50 μL of freshly prepared 10 mM DTT in 100 mM ammonium bicarbonate for 30 minutes at 50°C. The sample was then alkylated in the absence of light with 25 μL of 50 mM ammonium bicarbonate plus 50 μL of freshly prepared 55 mM iodoacetate in 100 mM ammonium bicarbonate for 20 minutes at room temperature. Following additional washes with 100 μL of 50 mM ammonium bicarbonate and 100 μL of acetonitrile, the gel band sample was digested overnight at 37°C with 75 μL of 50 mM ammonium bicarbonate plus 25 μL of 6 ng/μL sequencing grade porcine trypsin (Promega).

After digestion, the supernatant from the gel band sample was collected. The gel band was washed three times: once with 100 μL of 1% formic acid/2% acetonitrile in water, once with 100 μL of a 1:1 acetonitrile and water mixture, and once with 100 μL of 1% formic acid in acetonitrile. The pooled supernatant and wash solutions were then vacuum-dried overnight and resuspended in 0.1% formic acid in preparation for collection of mass spectrometry data. Samples of this nature were prepared in triplicate from the same freshly expressed and purified sample of mRouge.

These tryptic-digest samples were desalted on a 150 μm × 3 cm C18AQ pre-column (Magic 5 μm, Michrom). After desalting, separation of peptides was performed with a CapLC XE HPLC system (Waters) using a 5 to 35% acetonitrile gradient in 0.2% formic acid on a 100 μm × 15 cm column packed with the same resin as the pre-column. The flow rate during separation was 0.35 μL/min and the HPLC column was connected directly to the mass spectrometer used for MS/MS analysis. Tandem mass spectra were acquired in data-dependent acquisition mode on a hybrid LTQ FT-ICR Ultra mass spectrometer (Thermo Fisher Scientific) with a nanoelectrospray ion source. Full scan mass spectra (400-1600 m/z) were acquired after accumulating 500,000 ions (with a resolution of 50,000 at 400 m/z). The seven most intense ions from the full scans were trapped in the linear ion trap and fragmented by CID after accumulating 5,000 ions (collisional energy: 35%, isolation width: 3 Da). Ion charge state screening was employed for singly and multiply charged ions. A dynamic exclusion list was set (maximum retention time: 60 s, relative mass window: 10 ppm) and early expiration was permitted.

References

1.   Tsien RY (1998) The green fluorescent protein. *Annual Review of Biochemistry* 67:509-544.

2.   Verkhusha VV, *et al.* (2001) An enhanced mutant of red fluorescent protein DsRed for double labeling and developmental timer of neural fiber bundle formation. *Journal of Biological Chemistry* 276:29621-29624.

3.   Campbell RE, *et al.* (2002) A monomeric red fluorescent protein. *Proceedings of the National Academy of Sciences, USA* 99:7877-7882.

4.   Shu X, *et al.* (2006) Novel chromophores and buried charges control color in mFruits. *Biochemistry* 45:9639-9647.

5.   Wall MA, Socolich M, & Ranganathan R (2000) The structural basis for red fluorescence in the tetrameric GFP homolog DsRed. *Nature Structural Biology* 7(12):1133-1138.

6.   Petersen J, *et al.* (2003) The 2.0 Å crystal structure of eqFP611, a far red fluorescent protein from the sea anemone Entacmaea quadricolor. *Journal of Biological Chemistry* 278(45):44626-44631.

7.   Tubbs JL, Tainer JA, & Getzoff ED (2005) Crystallographic structures of Discosoma red fluorescent protein with immature and mature chromophores: Linking peptide bond trans-cis isomerization and acylimine formation in chromophore maturation. *Biochemistry* 44:9833-9840.

8.   Gurskaya NG, *et al.* (2001) GFP-like chromoproteins as a source of far-red fluorescent proteins. *FEBS Letters* 507:16-20.

9.   Lin MZ, *et al.* (2009) Autofluorescent proteins with excitation in the optical window for intravital imaging in mammals. *Chemistry & Biology* 16(11):1169-1179.

10.   Barstow B, *et al.* (2008) Alteration of citrine structure by hydrostatic pressure explains the accompanying spectral shift. *Proceedings of the National Academy of Sciences of the United States of America* 105:13362-13366.

11.   Wachter RM, *et al.* (1998) Structural basis of spectral shifts in the yellow-emission variants of green fluorescent protein. *Structure With Folding & Design* 6:1267-1277.

12.   Griesbeck O, *et al.* (2001) Reducing the environmental sensitivity of yellow fluorescent protein: Mechanism and applications. *Journal of Biological Chemistry* 276:29188-29194.

13. Barondeau DP, *et al.* (2003) Mechanism and energetics of green fluorescent protein chromophore synthesis revealed by trapped intermediate structures. *Proceedings of the National Academy of Sciences, USA* 100:12111-12116.

14. Barondeau DP, *et al.* (2005) Understanding GFP chromophore biosynthesis: Controlling backbone cyclization and modifying post-translational chemistry. *Biochemistry* 44:1960-1970.

15. Rosenow MA, *et al.* (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.

16. Strack RL, *et al.* (2010) Chromophore formation in DsRed occurs by a branched pathway. *Journal of the American Chemical Society* 132(24):8496-8505.

17. Wilmann PG, *et al.* (2005) The 2.1 Å crystal structure of the far-red fluorescent protein HcRed: Inherent conformational flexibility of the chromophore. *Journal of Molecular Biology* 349:223-237.

18. Pletnev S, *et al.* (2008) A crystallographic study of bright far-red fluorescent protein mKate reveals pH-induced cis-trans isomerization of the chromophore. *Journal of Biological Chemistry* 283(43):28980-28987.

19. Quillin ML, *et al.* (2005) Kindling fluorescent protein from Anemonia sulcata: Dark-state structure at 1.38 Å resolution. *Biochemistry* 44:5774-5787.

20. Battad J, *et al.* (2007) A structural basis for the pH-dependent increase in fluorescence efficiency of chromoproteins. *Journal of Molecular Biology* 368(4):998-1010.

21. Taguchi N, *et al.* (2009) Fragment molecular orbital calculations on red fluorescent proteins (DsRed and mFruits). *Journal of Physical Chemistry B* 113:1153-1161.

22. Davis IW, *et al.* (2007) MolProbity: All-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Research* 35:W375-W383.

23. Mayo S, Olafson B, & Goddard W (1990) Dreiding - A generic force-field for molecular simulations. *Journal of Physical Chemistry-US* 94(26):8897-8909.

24. Sitkoff D, Sharp KA, & Honig B (1994) Accurate calculation of hydration free energies using macroscopic solvent models. *Journal of Physical Chemistry* 98(7):1978-1988.

25. Dahiyat BI & Mayo SL (1997) De novo protein design: Fully automated sequence selection. *Science* 278:82-87.

26. Lazaridis T & Karplus M (1999) Effective energy function for proteins in solution. *Proteins* 35:133-152.

27. Shortle D (2003) Propensities, probabilities, and the Boltzmann hypothesis. *Protein Science* 12(6):1298-1302.

28. Allen BD & Mayo SL (2006) Dramatic performance enhancements for the FASTER optimization algorithm. *Journal of Computational Chemistry* 27:1071-1075.

29. Allen BD, Nisthal A, & Mayo SL (2010) Computational protein design of structural ensembles: validation by automated screening of combinatorial libraries. *Proceedings of the National Academy of Sciences, USA*.

30. Lassila JK*, et al.* (2006) Combinatorial methods for small-molecule placement in computational enzyme design. *Proceedings of the National Academy of Sciences, USA* 103:16710-16715.

31. Ho SN*, et al.* (1989) Site-directed mutagenesis by overlap extension using the polymerase chain reaction. *Gene* 77(1):51-59.

32. Gross LA*, et al.* (2000) The structure of the chromophore within DsRed, a red fluorescent protein from coral. *Proceedings of the National Academy of Sciences, USA* 97(22):11990-11995.

33. Leslie AGW (1992) Recent changes to the MOSFLM package for processing film and image plate data. *Joint CCP4 and ESF-EAMCB Newsletter on Protein Crystallography* 26.

34. Evans P (2006) Scaling and assessment of data quality. *Acta Crystallographica Section D: Biological Crystallography* 62:72-82.

35. Mccoy A*, et al.* (2007) Phaser crystallographic software. *Journal of Applied Crystallography* 40:658-674.

36. Murshudov GN, Vagin AA, & Dodson EJ (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallographica Section D: Biological Crystallography* 53:240-255.

37. Vagin AA*, et al.* (2004) REFMAC5 dictionary: Organization of prior chemical knowledge and guidelines for its use. *Acta Crystallographica Section D: Biological Crystallography* 60:2184-2195.

38. Afonine PV, Grosse-Kunstleve RW, & Adams PD (2005) The Phenix refinement framework. *CCP4 Newsletter* 42.

39.   Emsley P & Cowtan K (2004) Coot: Model-building tools for molecular graphics. *Acta Crystallographica Section D: Biological Crystallography* 60:2126-2132.

40.   Painter J & Merritt EA (2006) TLSMD web server for the generation of multi-group TLS models. *Journal of Applied Crystallography* 39:109-111.

SI Figures & Legends
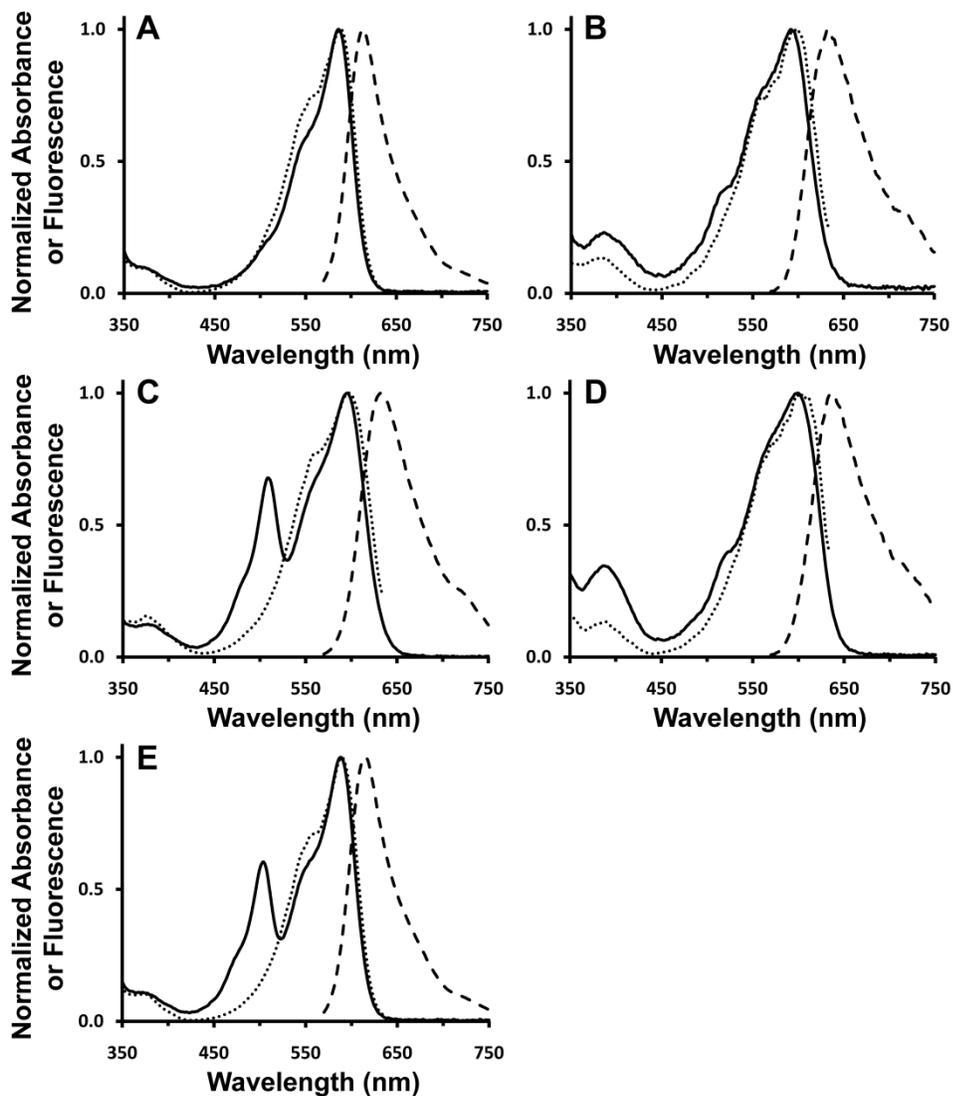
Fig. S1. Fluorescence and absorption spectra of mCherry (*A*), mRojoA (*B*), mRojoB (*C*), mRouge (*D*) and mCherry Q163M (*E*). The absorption spectra are shown with a solid line, the excitation spectra are shown with a dotted line and the emission spectra are shown with a dashed line.
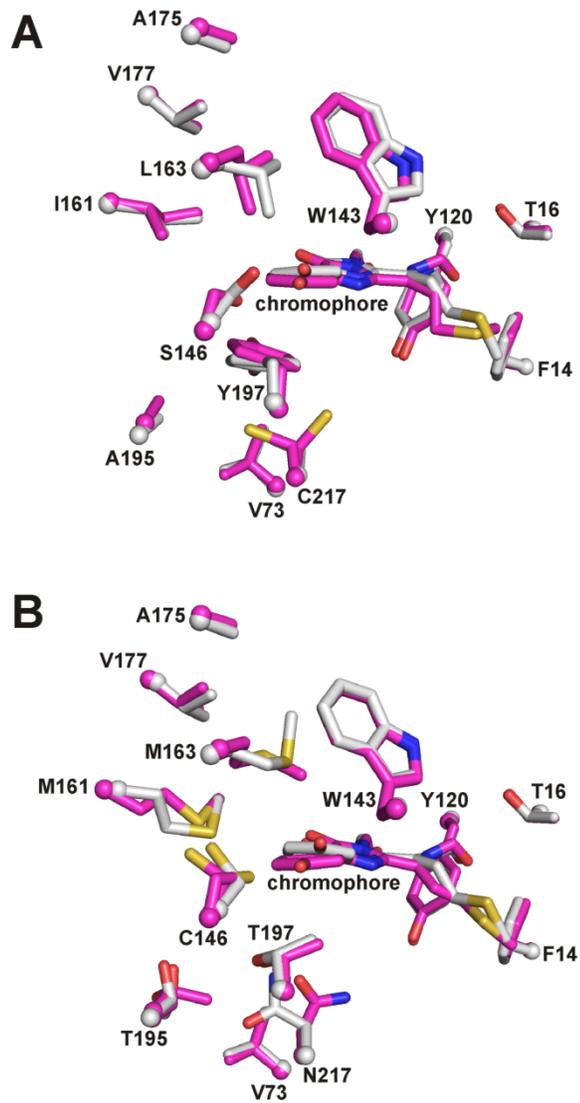
Fig. S2. Comparison of structure predicted by CPD (white) and crystal structure (magenta) for red-shifted mutants mRojoA (*A*) and mRouge (*B*). Residue C217 in mRojoA (*A*) and residues C146 and T195 in mRouge (*B*) exhibit two conformations in the crystal structure. The chromophore was not designed.
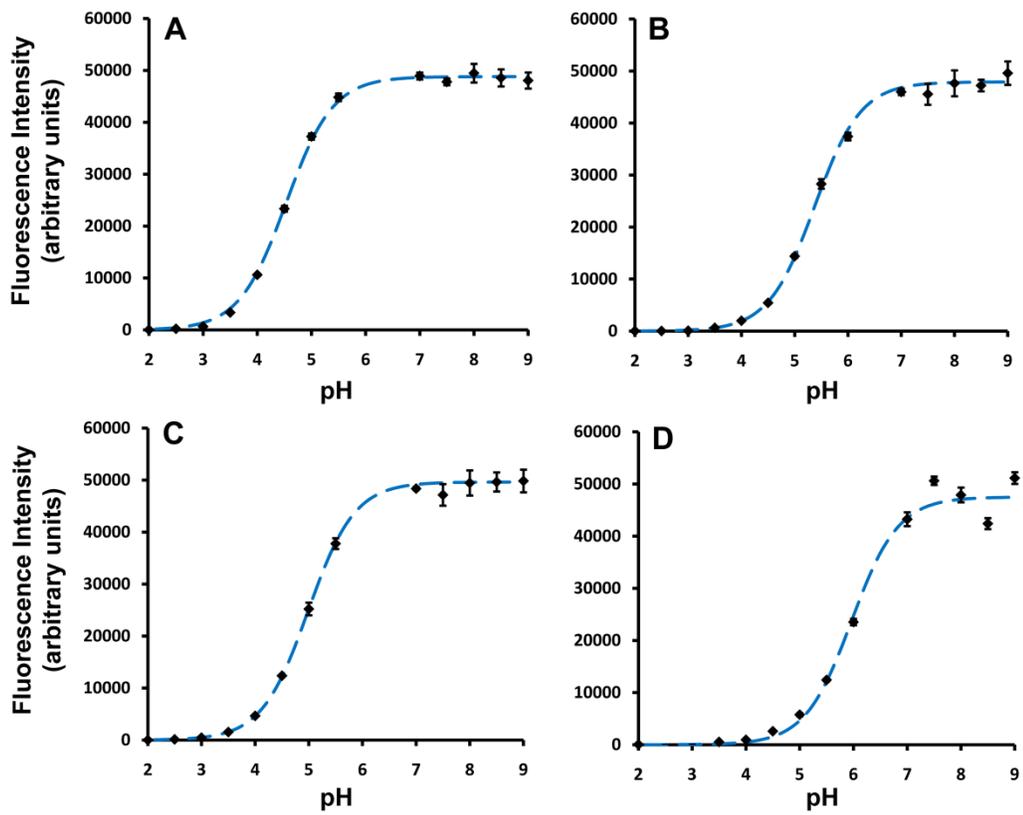
Fig. S3. pH profiles for mCherry (*A*), mRojoA (*B*), mRojoB (*C*) and mRouge (*D*).
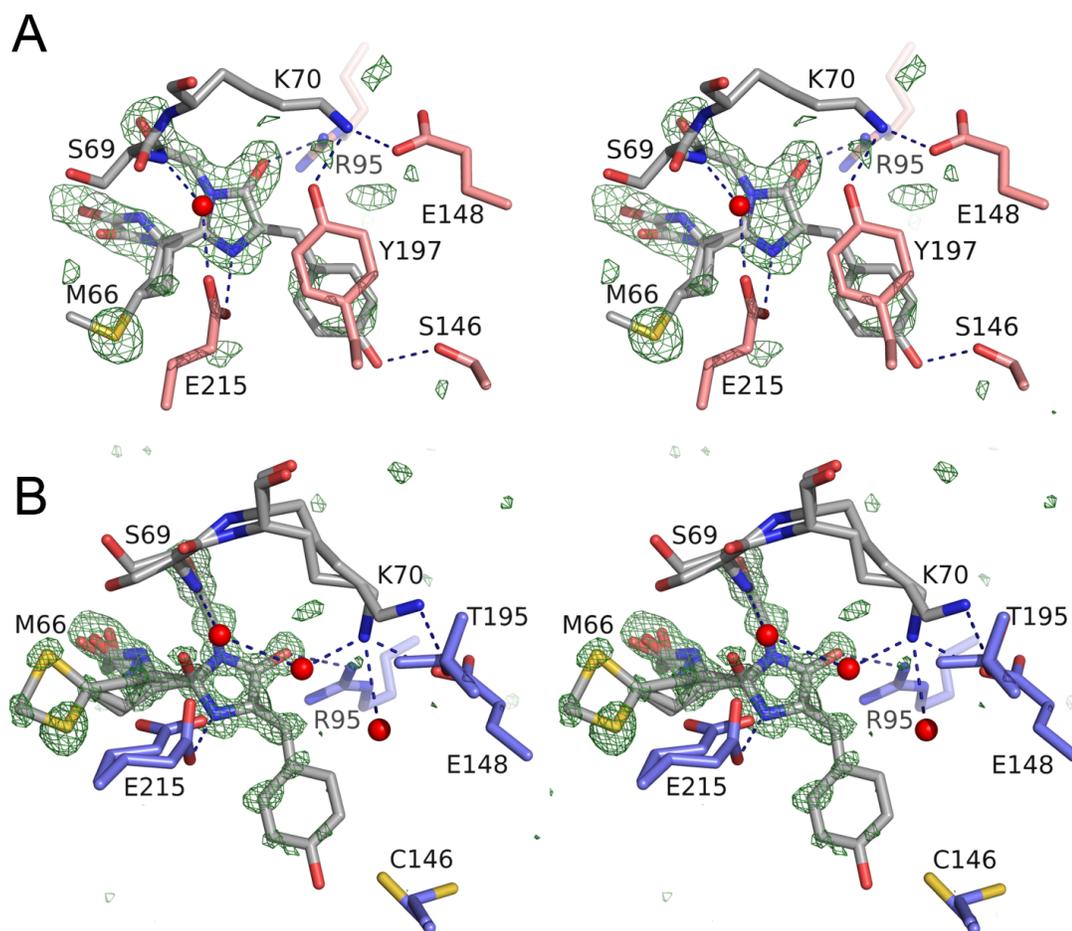
Fig. S4. Chromophore environment of mRojoA (*A*) and mRouge (*B*). H-bonds are indicated by dashed lines and crystallographic waters are indicated by red spheres. Omit maps (green) for the chromophores were contoured at 3.0 s for mRojoA and at 3.5 s for mRouge. Non-contiguous residues are colored pink (mRojoA) and purple (mRouge). The chromophore and its contiguous residues are colored gray. Note that the O_g atom in both conformations of T195 for mRouge (*B*) is hidden behind the C_b atoms of that same residue. The refinement model for the hydroxylated conformation of mRouge did not include atoms from the phenolate ring of the chromophore due to a lack of electron density.
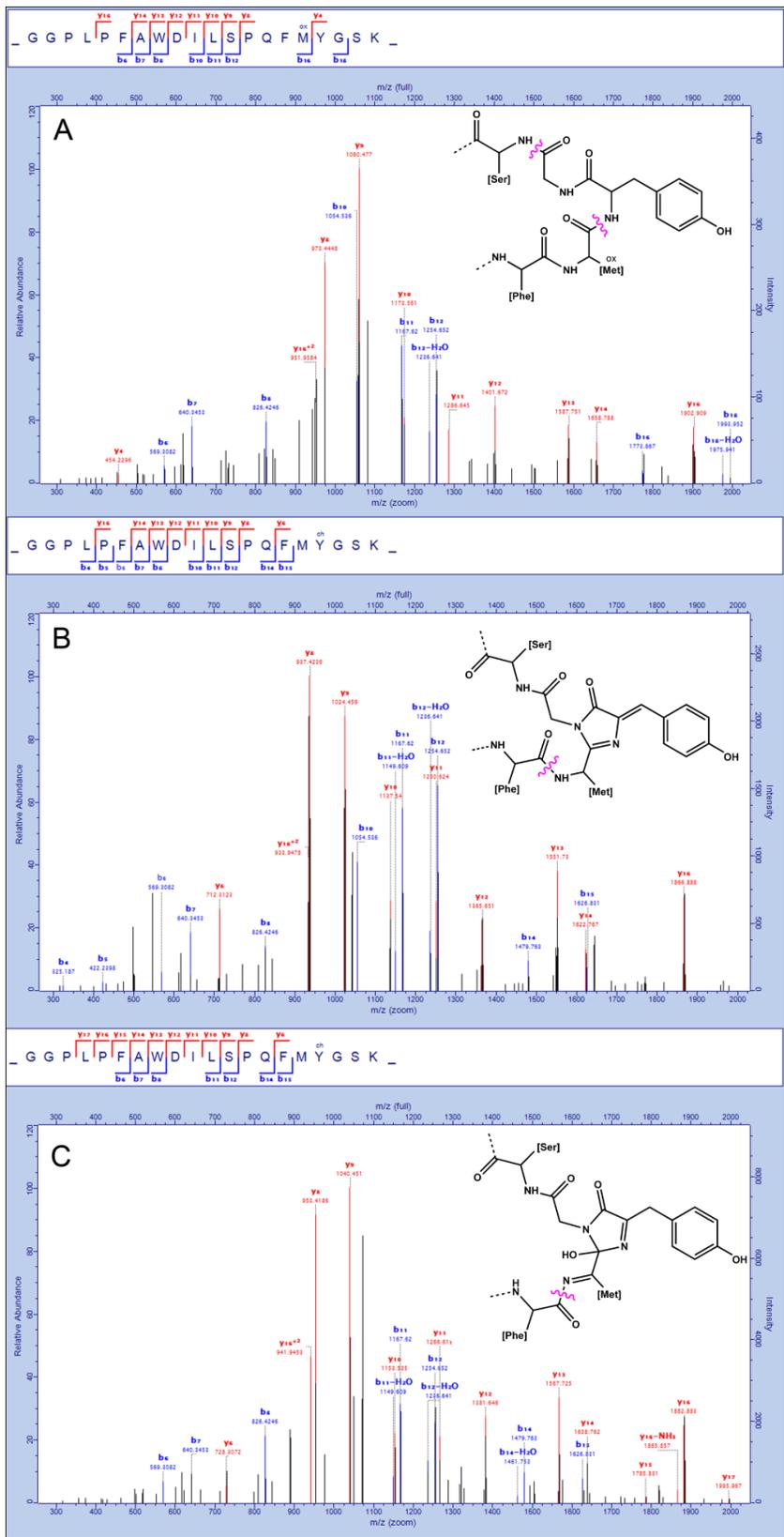
Fig. S5. Trypsin-digest mass spectrometry confirms the presence of a hydroxylated chromophore species in mRouge. Peptide fragments isolated by tandem mass spectrometry from an in gel trypsin-digest of mRouge indicate the presence of the uncyclized chromophore (*A*), the green chromophore species (*B*), and a hydroxylated species consistent with current literature on the maturation of DsRed (*C*). Note that in the uncyclized chromophore (*A*), the side chain sulfur atom of M66 was spontaneously oxidized to a sulfoxide during the process of data acquisition.
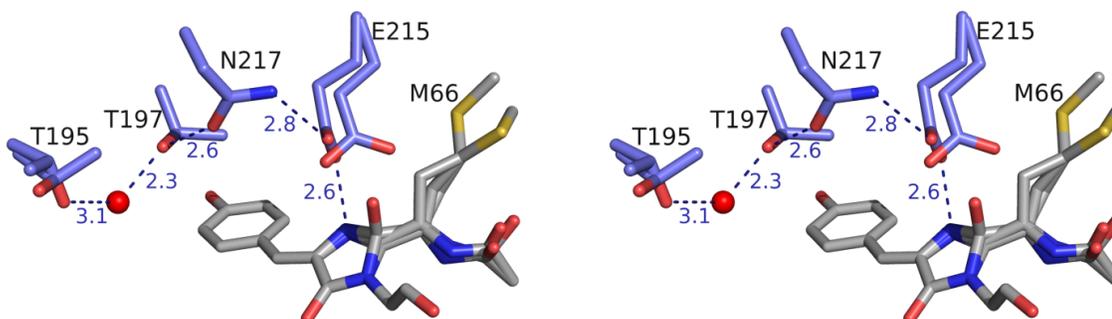
Fig. S6. H-bond network in the red-shifting TTN (T195-T197-N217) motif of mRouge. H-bonds are indicated by dashed lines. Numbers in purple indicate distances in Å.

Table S1. Sequences and emission wavelengths of the most red-shifted mutants found in each library

Protein	Design hypothesis I			Design hypothesis II				Design hypothesis III				λ_{em}
	14	16	120	143	146	161	163	73	195	197	217	(nm)
mCherry	F	V	Y	W	S	I	Q	V	V	I	A	611
Library 1												
1.13B9	-	T	-	-	-	-	-	-	-	-	-	615
1.10B1	Y	T	-	-	-	-	-	-	-	-	-	616
1.8D2	T	I	-	-	-	-	-	-	-	-	-	615
1.9H8	T	T	-	-	-	-	-	-	-	-	-	616
1.1B10	T	S	-	-	-	-	-	-	-	-	-	617
1.9F2	I	T	-	-	-	-	-	-	-	-	-	617
1.5G10	I	S	-	-	-	-	-	-	-	-	-	617
Library 2												
2.14H12	-	-	-	L	C	-	L	-	-	-	-	621
2.14C2	-	-	-	L	C	-	M	-	-	-	-	620
2.10E2	-	-	-	L	C	M	L	-	-	-	-	620
2.1H7	-	-	-	-	C	-	M	-	-	-	-	622
2.7A11	-	-	-	V	C	-	M	-	-	-	-	620
2.15E2	-	-	-	V	C	M	L	-	-	-	-	621
2.13B11	-	-	-	S	C	M	L	-	-	-	-	621
2.18G3	-	-	-	C	C	M	M	-	-	-	-	621
2.14D10	-	-	-	A	C	M	L	-	-	-	-	621
2.2E11	-	-	-	T	C	M	M	-	-	-	-	621

2.5H4	-	-	-	-	C	M	M	-	-	-	-	622
2.8F1	-	-	-	-	T	-	M	-	-	-	-	624
2.18C7	-	-	-	V	C	M	M	-	-	-	-	621
2.15H12	-	-	-	I	C	M	L	-	-	-	-	621
2.15G9	-	-	-	I	C	M	M	-	-	-	-	622
2.16H8	-	-	-	A	A	M	M	-	-	-	-	620
Library 3-1												
3-1.5H10	-	-	-	-	-	-	-	-	-	Y	C	621
3-1.7F9	-	-	-	-	-	-	-	-	-	Y	S	620
3-1.2G4	-	-	-	-	-	-	-	-	T	Y	C	621
3-1.3G3	-	-	-	-	-	-	-	I	-	Y	C	620
3-1.2D8	-	-	-	-	-	-	-	-	T	Y	S	620
3-1.4D8	-	-	-	-	-	-	-	-	A	Y	C	621
3-1.7H9	-	-	-	-	-	-	-	-	-	Y	T	620
3-1.7H1	-	-	-	-	-	-	-	I	A	Y	C	621
3-1.8D4	-	-	-	-	-	-	-	I	-	Y	T	620
Library 3-2												
3-2.8C9	-	-	-	-	-	-	-	-	T	T	N	619

Table S2. Crystallographic Data

	mRouge	mRojoA
Resolution, Å	56.7–0.95	38.0–1.70
Space group	$P12_11$	$P12_11$
Observations	568,101	234,983
Unique observations	135,356	102,737
R_{merge}, %	12.1 (49.6)	4.2 (39.3)
Completeness, %	91.9 (62.9)	95.4 (97.0)
I/σI	9.7 (1.7)	11.9 (2.2)
R_{work}/R_{free}, %	13.3/15.6	18.2/22.0

Values in parenthesis are statistics for the highest resolution shell of data.

Recovery of Red Fluorescent Protein Chromophore Maturation Deficiency through Rational Design

Matthew M. Moore[a], Samuel Oteng-Pabi[b], Antonia T. Pandelieva[b], Stephen L. Mayo[a,c], Roberto A. Chica[b, d, 1]

[a]Division of Biology, California Institute of Technology,

1200 E. California Blvd., Pasadena, CA 91125 USA.

[b]Department of Chemistry, University of Ottawa, Ottawa, ON, Canada K1N 6N5

[c]Division of Chemistry and Chemical Engineering, California Institute of Technology,

1200 E. California Blvd., Pasadena, CA 91125 USA.

[d]Centre for Catalysis Research and Innovation, University of Ottawa,

Ottawa, ON, Canada K1N 6N5

[1]Corresponding author: Email: rchica@uottawa.ca

Abbreviations: RFP, red fluorescent protein; FP, fluorescent protein; λ_{em}, fluorescence emission wavelength; λ_{ex}, fluorescence excitation wavelength; Φ_F, fluorescence quantum yield

Abstract

Red fluorescent proteins (RFPs) derived from organisms in the class *Anthozoa* have found widespread application as imaging tools in biological research. For most imaging experiments, RFPs that mature quickly to the red chromophore and produce little or none green chromophore are most useful. In this study, we converted a yellow fluorescent mPlum mutant to a red-emitting RFP by replacing Glu16 with small non-polar amino acids. We also created an optimized mPlum mutant (mPlum-E16P) that matures almost exclusively to the red chromophore. Analysis revealed two structural characteristics that may be critical for efficient red chromophore maturation in DsRed-derived RFPs. The first is the presence of a lysine residue at position 70 that is able to interact directly with the chromophore. The second is an absence of non-bonding interactions limiting the conformational flexibility at the peptide backbone that is oxidized during red chromophore formation. Our data, as well as structural features for known acylimine-forming fluorescent proteins, supports this proposition.

Red fluorescent proteins (RFPs) derived from organisms in the class *Anthozoa* are widely employed as imaging tools in biological research. For example, these proteins have been used as markers of gene expression, expressed as fusion partners for the tracking of intracellular endogenous protein-RFP chimeras, and complemented with other fluorescent proteins (FPs) for use in fluorescence resonance energy transfer experiments [1, 2]. The longer emission wavelengths of RFPs result from a chromophore whose p-hydroxybenzylidene-imidazolinone group is prepended with an acylimine substituent at the C1 atom as a result of backbone oxidation [3, 4]. The extended conjugation afforded by unsaturation from the acylimine lowers the energy of both excitation and emission from the analogous fluorophore found in green fluorescent proteins (GFPs), resulting in a bathochromic wavelength shift [5].

Chromophore maturation in RFPs is a long-studied phenomenon. In the past, green chromophore maturation was seen as an intermediate step on the pathway to complete red chromophore maturation [6]; this interpretation turned out to be false [7]. Recent work has revealed much about how the red and green chromophore pathways branch off from one another during the maturation process. A study using deuterium labeling uncovered evidence that a branch upstream to green chromophore formation splits the population of proteins in solution into two: a population that will mature to the GFP-like green chromophore and one that will pass through a blue intermediate species to form the RFP red chromophore [7]. Previous to this study, characterization of the RFP HcRed revealed a peak at 410 nm during maturation that corresponds to this blue intermediate species [8]. This peak is significant in that it shows the expected behavior of an intermediate species in the red chromophore maturation process. In addition, X-ray crystallography and

computational modeling support the existence of a species with the structure that has been proposed for this blue intermediate [9, 10].

A schematic illustrating the proposed mechanism for chromophore formation in RFPs is shown in Fig. 1. Briefly, the uncyclized chromogenic tripeptide (Xaa-Tyr-Gly) *1* undergoes a cyclization reaction that is trapped by oxidation to a colorless intermediate species *3*. This colorless intermediate reversibly eliminates a hydroxide moiety yielding *6*, thus enabling a branch in the chromophore maturation pathway. Colorless intermediates that eliminate hydroxide are poised to undergo further oxidation, resulting in the crystallographically observable blue intermediate with an absorption peak at 410 nm (*7*) that was mentioned above. The exact electronic structure of this blue intermediate is unknown—both cationic [7] and anionic [10] structures have been proposed. Completion of this second oxidation reaction seals the fate of FPs and directs their maturation to the red chromophore (*9* and *10*). A base-induced elimination of water follows this second oxidization as evidenced by a kinetic isotope effect on the C-beta carbon of the conserved tyrosine residue central to the chromogenic tripeptide [7]. Colorless intermediates that retain hydroxide (*3*) may also undergo elimination of water by proton abstraction at the C-beta carbon of the conserved tyrosine residue, leading to production of FPs with green chromophores (*4* and *5*) in the other branch of this maturation pathway [11].

Red and green chromophores cannot interconvert—both are dead-end products of chromophore maturation [6, 7]. Consequently, some RFPs express as mixtures of proteins with either green or red chromophores. DsRed and its mutant mPlum are notable examples of RFPs that express in this fashion [12, 13]. Studies of Lys70 in various

mutants of DsRed indicate that this residue is crucial for the formation of the acylimine group and thus red fluorescence [3, 14, 15]. Additionally, the proximity of the Lys70 terminal amino group to the chromophore ring system has been shown to correlate with increasing quantum yield (Φ_F) among DsRed mutants [4, 14].

Fast and complete maturation to the red chromophore is a desired property for RFPs used in imaging experiments. To improve maturation efficiency, investigators have used directed evolution and have obtained RFPs that display fast and efficient red chromophore maturation such as mCherry [16], DsRed.T4, [17] and mKate2 [18]. Here, we present a rational approach to enhance RFP chromophore maturation in the DsRed mutant mPlum. We hypothesized that the identity of the amino acid at position 16 may be crucial to promoting or inhibiting red chromophore maturation. Because residues at position 16 interact directly with the peptide bond that oxidizes during red chromophore formation, we postulated that this interaction could interfere with the oxidation reaction at this position. Using rational design, we converted a yellow-emitting mPlum mutant to a red-emitting RFP by replacing Glu16 with small non-polar amino acids. We also created an optimized mPlum mutant (mPlum-E16P) that matures almost exclusively to the red chromophore.

Results and Discussion

Red chromophore maturation deficient mPlum mutant. In a previous study, we developed a structure-based, rational design approach that combined computational protein design and experimental screening of combinatorial mutant libraries to red shift the emission wavelength of RFPs [19]. One of the hypotheses that we formulated to

achieve this goal was that the introduction of a π-stacking interaction with the chromophore phenol ring (similar to the one observed in yellow fluorescent proteins [20]) would red shift the emission wavelength of mCherry by stabilizing the excited state. Using our combined computational and experimental approach, we successfully introduced a π-stacking interaction with the mCherry chromophore by making a I197Y mutation (numbering based on DsRed). Two additional mutations, T195A and A217C, improved the quantum yield of the single mutant I197Y. This mutational motif, termed "AYC" for A195/Y197/C217, resulted in a 10 nm bathochromic shift in emission wavelength relative to mCherry [19].

To create a further red-shifted monomeric RFP, we investigated whether the AYC motif could red shift the emission wavelength of the far-red emitting RFP mPlum [21]. mPlum is a member of the mFruit family of monomeric RFPs derived from DsRed [16] and differs from mCherry by 13 mutations (excluding N- and C-terminal tags, see Table 1). Its fluorescence emission wavelength (λ_{em}) of 649 nm is the longest of the mFruits and is red-shifted 38 nm compared to that of mCherry ($\lambda_{em} = 611$ nm). Ile65 and Glu16 appear to be major contributors to mPlum's long emission wavelength as replacement of either residue results in an up to 40 nm hypsochromic shift [21]. These residues interact through an H-bond between the side chain of Glu16 and the acylimine carbonyl oxygen atom of Ile65 [22], and this interaction gives rise to a dynamic Stokes shift that is responsible for the far-red emission wavelength of 649 nm [13]. Given the very high sequence identity between mCherry and mPlum (> 90%), we hypothesized that the AYC motif would red shift the emission wavelength of mPlum as was observed in mCherry, yielding a monomeric RFP with $\lambda_{em} > 650$ nm.

Transplanting the AYC motif into mPlum to generate mPlumAYC resulted in a mutant protein that was completely deficient in red chromophore maturation; i.e., red chromophore formation was essentially undetectable by either absorption or fluorescence spectroscopy (Fig. 2). The absorption peak corresponding to the anionic red chromophore (Fig. 1, *10*), which is at 588 nm in the spectrum of mPlum, is absent in the spectrum of mPlumAYC. This variant instead displays peaks centered at 396 nm and 508 nm; the 396 nm peak corresponds to the neutral green chromophore (Fig. 1, *4*), and the 508 nm peak corresponds to the anionic green chromophore (Fig. 1, *5*). As seen in Fig. S1, when the pH is increased from 7.0 to 9.5, the intensity of the 396 nm peak decreases and a concomitant increase in intensity of the 508 nm peak is observed, indicating ionization of the green chromophore. This same behavior is observed in mPlum (Fig. S1). Given that the mPlumAYC green chromophore pKa is 7.3 (Table 1, Fig. S2), peaks corresponding to both ionization states of the green chromophore should be observable at the pH value of the measurements (pH 7.0). When excited at 508 nm, mPlumAYC displayed dim fluorescence at 527 nm with a quantum yield of 0.02, and no red fluorescence was detected with excitation at 590 nm. Thus, the AYC motif transformed mPlum into a dim yellow fluorescent protein.

To examine the effect of the AYC motif on chromophore maturation, we performed maturation experiments. RFPs were expressed anaerobically in airtight culture tubes and purified rapidly at 4 °C in deoxygenated solutions to obtain protein samples that contained few or no fully mature chromophores. Absorption scans were then done to follow the maturation process at 28 °C. Chromophore maturation of mPlum at pH 7.5 (Figs. 3 and S5A) demonstrates an increase in absorbance as a function of time for the

anionic green (506 nm) and red (588 nm) chromophore peaks. Additionally, a peak at 410 nm appears to shift to 391 nm during the maturation process at pH 7.5. The peak at 410 nm corresponds to the blue intermediate species [7, 10] (Fig. 1, *7*), and the peak at 391 nm corresponds to the neutral green chromophore (Fig. 1, *4)*. To remove spectral overlap of these two peaks, we performed the same maturation experiments at pH 9.5 (Fig. 3). At this pH, the neutral green chromophore (pKa 7.3–7.5) is deprotonated, thereby decreasing the peak intensity at 391 nm and leaving the 410 nm peak unaffected. mPlum maturation at pH 9.5 (Figs. 3 and S5B) shows an increase in absorbance as a function of time for the anionic green (508 nm) and red (588 nm) chromophore peaks. Additionally, a peak at 410 nm increases in intensity and then disappears. This result confirms that the blue intermediate is formed transiently during maturation, as expected [6, 7, 23].

For mPlumAYC, increases in absorbance are observed during maturation at pH 7.5 for the neutral (390 nm) and anionic (508 nm) green chromophore peaks. At pH 9.5, the 390 nm peak disappears due to ionization of the neutral green chromophore, which would allow a 410 nm peak to be observed, if present. For mPlumAYC, however, no such peak is observed (Fig. 3), indicating that the intermediate blue species does not form and that the second oxidation reaction (Fig. 1, *6* to *7)* does not occur. No evidence for the presence of the red chromophore was observed in mPlumAYC as indicated by the absence of any absorption peak at > 550 nm (Fig. 2). Thus, introduction of the AYC motif into mPlum prevents red chromophore maturation by inhibiting the second oxidation reaction, which results in a dim yellow fluorescent protein containing only the green chromophore.

Recovery of red chromophore maturation through rational design. Chromophore maturation in mPlum results in a mixture of protein molecules containing either the red chromophore ($\lambda = 588$ nm) or the green chromophore (peaks at 391 nm and 506 nm) [13] (Fig. 2). This is not the case for mCherry, a fast and efficient maturing RFP that matures almost exclusively to the red chromophore with a single peak centered at 586 nm (Fig. S1) [19]. We hypothesized that Glu16 plays an important role in mPlum's inefficient red chromophore maturation compared to mCherry. Our hypothesis is based on the fact that Glu16 H-bonds to the chromophore acylimine oxygen atom [22]. We postulated that this H-bonding interaction could restrict the conformational freedom of the peptide backbone, which may be required for the second oxidation reaction generating the acylimine group to take place (Fig. 1, **6** to **7**). Previously, researchers suggested that a *cis-trans* isomerization of the peptide backbone between residues 65 and 66 was necessary for this second oxidation reaction to occur [5, 24]. However, such an isomerization would result in a red chromophore with an acylimine oxygen atom pointing up away from catalytic residue Glu215; this has never been observed [25]. During the second oxidation reaction, the C-alpha carbon of residue 66 is transformed from sp^3-hybrization to sp^2-hybridization, changing its geometry from tetrahedral to trigonal planar. This process clearly requires a conformational change in the peptide bond between residues 65 and 66. An H-bonding interaction between Glu16 and the carbonyl oxygen of residue 65 would very likely decrease conformational flexibility around this bond.

An additional observation that lends support to our hypothesis comes from the absorption spectrum of mRojoA [19], an mCherry mutant that contains the AYC motif as well as a water-mediated H-bond between the side chain of Thr16 and the chromophore acylimine

oxygen. Absorption spectra for mRojoA demonstrate that as the pH is increased from 7.5 to 9.5, a peak at 517 nm is revealed, indicating that mRojoA expresses as a mixture of green and red chromophore-containing proteins (Fig. S1). mCherry, which contains a non-H-bonding valine at position 16, does not exhibit this behavior (Fig. S1), indicating that purified preparations of this protein are essentially homogeneous in their maturation to the red chromophore.

Based on these observations, we hypothesized that replacement of Glu16 with residues that cannot donate H-bonds to the acylimine oxygen would improve maturation to the red chromophore in maturation-deficient mPlumAYC. To test this hypothesis, we performed mutagenesis at position 16 in mPlumAYC followed by fluorescence detection of mutants that displayed emission at wavelengths > 550 nm. A total of 11 different amino acids including glutamine were tested at position 16; these included mutations to polar, hydrophobic, aromatic, acidic, and basic residues (Fig. S3). The E16D and E16T mutations did not recover red fluorescence in mPlumAYC, suggesting that side chains that can participate in H-bonding with the acylimine oxygen are detrimental to the second oxidation reaction leading to the red chromophore. The polar residues tyrosine and arginine also did not recover red fluorescence.

Mutation to either an alanine or a proline resulted in the highest recovery of red fluorescence intensity. Some recovery was seen for other non-polar mutations (E16G, E16V, E16I and E16F), but the average integrated red fluorescence signal for these mutants was less than that seen for E16P and E16A (Fig. S3). Decreased fluorescence intensity may be due to lower protein expression levels or impaired fluorescence. However, since cellular densities were controlled, discrepancies in integrated

fluorescence intensity are unlikely to result from well-to-well differences in cellular growth levels. Thus, our mutagenesis study revealed that small non-polar residues improve red chromophore maturation efficiency, possibly by allowing for conformational flexibility during oxidation of the peptide backbone that generates the acylimine. Additionally, the presence of side chains smaller than glutamine at position 16 may increase molecular oxygen accessibility to the peptide backbone, thereby improving the initiation efficiency of this oxidation reaction.

The two mutants that displayed the highest red fluorescence intensity, mPlumAYC-E16A and mPlumAYC-E16P, were further characterized. Their absorption spectra displayed three peaks centered at ~390 nm (neutral green chromophore), ~510 nm (anionic green chromophore), and ~590 nm (red chromophore), similar to what is observed for mPlum (Fig. 2). However, unlike mPlum, the highest intensity peaks for both of these mutants in the pH 7.0–9.5 range are the ~390 nm and ~510 nm peaks (Figs. 2 and S1), suggesting that the majority of RFP molecules contain a green chromophore instead of a red chromophore. The red/green chromophore ratios are reported in Table 2 as ratios of absorbance at 590 nm and 510 nm (A_{590}/A_{510}). These ratios were computed using absorbance values measured at pH 9.5 in order to deprotonate all green chromophores present in solution, converting the entirety of green chromophore-related absorbance to a single peak around 510 nm. The red/green chromophore ratio for mPlum is close to unity, whereas it is approximately 0.15 for mPlumAYC-E16A and mPlumAYC-E16P. Since the extinction coefficients of the red and green chromophores in FPs do not typically differ by orders of magnitude, this change in red/green ratio is significant. When excited at 510 nm, mPlumAYC-E16A and mPlumAYC-E16P display yellow fluorescence at 523 nm

and 525 nm, respectively, with $\Phi_F = 0.02$. These proteins thus contain a significant proportion of molecules that contain a yellow-emitting green chromophore, similar to mPlumAYC.

mPlumAYC-E16A and mPlumAYC-E16P also mature to molecules containing red chromophores that emit at 639 nm and 637 nm, respectively (Table 1). Interestingly, these mutants emit at wavelengths 10–12 nm shorter than mPlum ($\lambda_{em} = 649$ nm). This is presumably because they have lost the dynamic Stokes shift resulting from the H-bonding interaction between the chromophore acylimine and Glu16. Other mPlum mutants with non-polar amino acids at position 16 have been shown to emit at 615–626 nm [21, 22]. Since our mPlumAYC-E16A and mPlumAYC-E16P mutants emit at 637–639 nm, we propose that their 11–24 nm longer emission wavelengths result from the red-shifting effect of the π-stacking interaction incorporated through the AYC motif. Although the E16P and E16A mutations in mPlumAYC partially restore red chromophore maturation, the red fluorescence quantum yields for these mutants ($\Phi_F = 0.04$ and 0.05, respectively) are about half that of mPlum (Table 1).

Maturation kinetics experiments on mPlumAYC-E16A and mPlumAYC-E16P revealed that both variants display a time-dependent increase in the peaks corresponding to the green (~390 and ~510 nm) and red (~590 nm) chromophores at pH 7.5 (Fig. S4). As observed for mPlum, the ~390 nm peak overlaps with the 410 nm peak of the blue intermediate species. To remove this spectral overlap, maturation experiments were also performed at pH 9.5 (Fig. S4). At this higher pH, we observe a peak at 410 nm that increases in intensity and then disappears, suggesting that the blue intermediate species has been consumed to form the red chromophore, as seen for mPlum (Fig. 3). Thus, loss

of the H-bonding interaction between residue 16 and the acylimine oxygen appears to partially restore oxidation of the backbone with subsequent formation of the intermediate blue species (Fig. 1, *7*).

A fast and efficient red chromophore maturing mPlum mutant. A disadvantage of mPlum for imaging experiments is that preparations of this protein result in a mixture of molecules that contain either red or green chromophores [13], as observed in Fig. 2. This is an undesirable property for multicolor imaging applications. To assess whether removal of the H-bonding interaction between Glu16 and the chromophore would result in improved maturation to the red chromophore in mPlum, we prepared the point mutant mPlum-E16P and measured its absorption spectrum. This single amino acid change resulted in an RFP that matures almost exclusively to the red chromophore (Fig. 2). The 506 nm anionic green chromophore peak at pH 7.0 was ablated, and only one major peak located at 590 nm is observed. When the pH is increased to 9.5 (Fig. S1), a small but observable peak appears at 509 nm, corresponding to the anionic green chromophore. These results indicate that maturation to the red chromophore in mPlum-E16P, although dramatically improved, is not complete. Nevertheless, the absorbance ratio A_{590}/A_{510}, representing the ratio of red chromophore to green chromophore, is four times that of mPlum (Table 1), indicating that red chromophore-containing molecules constitute the majority of the RFP population in solution. Maturation kinetics experiments of mPlum-E16P at both pH 7.5 and 9.5 show a clearly observable 410 nm peak corresponding to the intermediate blue species (Fig. S5*D*). The blue species is formed rapidly and is consumed to produce the red chromophore. No peak for the neutral green chromophore is present to overlap with the 410 nm peak, again suggesting that little

green chromophore is present. Therefore, the E16P mutation improved red chromophore maturation efficiency in mPlum, similar to its effect in mPlumAYC.

mPlum-E16P emits red fluorescence at 630 nm, which is almost 20 nm blue-shifted relative to mPlum. This is not unexpected as the dynamic Stokes shift responsible for the large red shift in mPlum is lost when Glu16 is replaced by a proline. Interestingly, this 630 nm emission wavelength is identical to that of mPlum-E16Q [22], which can still form an H-bond with the chromophore acylimine group. Moreover, the emission wavelength of mPlum-E16P is 12 nm longer than that of the E16L mutant, which cannot form such an interaction [22]. Notably, the quantum yield of mPlum-E16P is 40% higher than the quantum yield of mPlum (Table 1) and its extinction coefficient of $29,350 \pm 2000 \ M^{-1}cm^{-1}$ is similar to that of mPlum ($22,000 \ M^{-1}cm^{-1}$ [18]), resulting in a brighter RFP. Another characteristic of mPlum-E16P is that its excitation at 509 nm does not result in detectable yellow fluorescence. Hence, removal of the Glu16-acylimine H-bond largely improved red chromophore maturation in mPlum, resulting in a brighter RFP that matures almost exclusively to the red chromophore.

Chromophore maturation mechanism. To better understand the roles of the AYC motif and position 16 residue identity during chromophore maturation, we performed maturation kinetics experiments for all the mPlum-derived RFPs described above (mPlum, mPlum-E16P, mPlumAYC, mPlumAYC-E16P, and mPlumAYC-E16A). These experiments were performed at pH 7.5 to approximate physiological pH and at pH 9.5 to avoid spectral overlap of the ~390 nm and 410 nm absorption peaks. Maturation half-times for the anionic green and red chromophores (Fig. 1, *5* and *10*) were measured, as well as the time to reach maximum 410 nm absorbance. Maturation half-times for the

formation of the neutral green chromophore (Fig. 1, *4*) were not measured for two reasons. First, the ~390 nm peak corresponding to this species overlaps with that of the 410 nm absorbing blue intermediate, introducing measurement errors. Second, maturation data from mPlumAYC, which matures exclusively to the green chromophore, revealed that maturation half-times for the neutral and anionic green chromophores were identical at the pH of measurement (Fig. S5*C*). This result indicates that ionization of the green chromophore occurs on a much faster timescale than chromophore formation. Thus, the maturation half-time of the anionic green chromophore is sufficient to accurately reflect the maturation half-time of the dehydration step leading to the neutral green chromophore (Fig. 1, *3* to *4*).

For the RFPs studied here, maturation half-times of both green and red chromophores were faster at pH 9.5 than at pH 7.5, except for mPlumAYC-E16P (Table 2). In this protein, red chromophore formation half-times were not significantly different at these two pHs. A shortening of maturation half-times in response to increased pH suggests that a base may be involved in the rate-limiting step on both the green and red chromophore maturation pathways. This assessment is supported by previous studies in which dehydration along both maturation pathways was shown to be rate limiting [7, 11]. Since dehydration reactions in the chromophore maturation mechanism (Fig. 1) require proton abstraction to eliminate hydroxide, it is expected that a higher pH would accelerate these reactions. Interestingly, the greatest acceleration in maturation as a function of pH occurs for mPlum (Table 2 and Fig. S5*A*, *B*). For green chromophore maturation, the half-time decreases from 2.8 h at pH 7.5 to 0.47 h at pH 9.5, whereas for the red chromophore, maturation half-time decreases from 7.3 to 3.9 h. It is unclear why only mPlum would

display such a large acceleration of chromophore maturation, but ionization of the Glu16 residue at higher pH could be a factor.

For all the RFPs studied here, the first peak that appears during chromophore maturation is the 410 nm blue species peak, followed by the green chromophore ~510 nm peak, and finally the red chromophore ~590 nm peak (Figs. 3, S4, and S5A, B, D). Formation of the blue intermediate before the red chromophore is expected. Formation of this blue species by the second oxidation reaction has been demonstrated to shunt intermediates down the red chromophore-forming pathway [7]. Studies in other RFPs have shown similar behaviors [6, 26]. The times at which the 410 nm peak reaches maximum absorbance are reported in Table 2 for all RFPs studied here. After this time point, the production rate of the blue species falls behind the consumption rate, and this species is continually transformed into red chromophore. Except in mPlum-E16P, the 410 nm peak maximum is reached after the green chromophore maturation half-time. This observation indicates that the blue intermediate species is still forming when half of the total green chromophores in solution have been produced. Importantly, the relative rates for the dehydration step leading to green chromophore formation (Fig. 1, *3* to *4*) and the oxidation step leading to the acylimine-containing blue species (Fig. 1, *6* to *7*) should determine the final ratio of red to green chromophore-containing molecules. Thus, a kinetic effect at the branch point of the mechanism (Fig. 1, boxed) should determine the efficiency of RFP chromophore maturation. Consequently, if the half-time of green chromophore maturation is shorter than the time of peak maximum for the blue species, a significant proportion of molecules should contain green chromophores. This is indeed what is observed for mPlum, mPlumAYC-E16A, and mPlumAYC-E16P, which have

red/green chromophore ratios less than 1.0 (Table 2). On the other hand, if formation of the blue species is faster than the dehydration step, the final ratio of red/green chromophores should increase. This is what we observe for our mPlum variant with the most efficient red chromophore maturation, mPlum-E16P, whose 410 nm peak maximum is reached 1.8 h before its green chromophore maturation half-time (Table 2). Thus, in mPlum-E16P, before half of the total green chromophore is formed, the majority of flux through the blue species has already occurred via consumption along the pathway to yield red chromophore. As a result, at the completion of the maturation process, this protein displays the highest red/green ratio of all mPlum mutants described here ($A_{590}/A_{510} = 3.7$).

In summary, our data support the notion that to obtain RFPs displaying more efficient red chromophore maturation, the blue species formation rate must exceed the green chromophore formation rate. The overall red chromophore formation rate need not be increased outright, since the half-time of red chromophore maturation is always longer than the half-time of green chromophore maturation (Table 2) [6, 7]. Presumably, the E16A and E16P mutations speed up the second oxidation step (Fig. 1, *6* to *7*) and possibly slow down the dehydration step leading to green chromophore formation (Fig. 1, *3* to *4*). Alternatively, the competition for intermediate substrates (*3* and *6*) becomes more pronounced when the second oxidation step is made more efficient, which then impedes dehydration. For example, Table 2 shows that the maximum absorbance of the 410 nm peak is reached ~0.2 h faster during mPlum-E16P maturation than during maturation of mPlum. Simultaneously, the maturation half-time of the green chromophore is lengthened in mPlum-E16P relative to mPlum at both pH 7.5 and pH 9.5. Longer green chromophore

- 76 -

maturation half-times are also observable for the mPlumAYC-E16P and mPlumAYC-E16A mutants relative to mPlumAYC (Table 2).

Lastly, the AYC motif affects maturation by slowing down formation of the blue species, and in the case of mPlumAYC, by completely inhibiting it. As seen in Table 2, the time of 410 nm peak maximum is much longer for the AYC motif-containing mPlum mutants than for mPlum and mPlum-E16P, which do not contain this motif. At the same time, the AYC mutations appear to speed up dehydration leading to the green chromophore. For example, at pH 7.5, green chromophore maturation half-time is ~2 h shorter in mPlumAYC than in mPlum and is ~1.7 h shorter in mPlumAYC-E16P than in mPlum-E16P. At pH 9.5, green chromophore formation remains faster for mPlumAYC-E16P compared to mPlum-E16P. Notably, the green chromophore formation rate for mPlum exceeds that for mPlumAYC at pH 9.5, but this may be related to interfering effects from ionization of Glu16 at higher pH. On balance, the AYC motif appears to favor the dehydration reaction leading to formation of the green chromophore (Fig. 1, *3* to *4*) over the second oxidation reaction (Fig. 1, *6* to *7*) that eventually leads to red chromophore formation.

Crystallographic support of maturation studies. The various mPlum mutants described here constitute a useful set of proteins for studying the structure-function relationships involved in RFP chromophore maturation. Thus, to help elucidate the mechanisms underlying the observed effects of our mutations on chromophore maturation, we solved the crystal structures of mPlum-E16P, mPlumAYC, and mPlumAYC-E16A. Data collection and refinement statistics are reported in Table S1. The structure of mPlumAYC demonstrates that the intended π-stacking interaction

between Tyr197 and the chromophore was successfully introduced (Fig. 4*A*). The centroid-to-centroid distance between the Tyr197 phenol ring and the chromophore phenol ring was 3.0 Å and the interplanar angle between these rings was 12.3°. In the mPlumAYC-E16A structure (Fig. 4*B*), the centroid-to-centroid distances were 4.1 Å and 4.0 Å and the interplanar angles were 11.6° and 11.1° for chains A and B, respectively. For comparison, the π-stacking interaction of the yellow fluorescent protein citrine has a centroid-to-centroid distance of 3.6 Å and an interplanar angle of 6.1° [27]. In mRojoA, an mCherry mutant equipped with the AYC motif, centroid-to-centroid distances are tightly clustered in a range of 3.8–3.9 Å for all subunits in the crystal, and the interplanar angles span 4.2°–12.8° [19]. Thus, our crystallographic data demonstrate that introduction of the AYC motif in mPlum resulted in successful creation of the desired π-stacking interaction with the chromophore (Fig. 4*A*, *B*).

The crystal structures of mPlumAYC and mPlumAYC-E16A did not contain appreciable omit map density (contoured at 3σ) corresponding to the presence of an oxidized peptide backbone with sp^2-hybridized C-alpha carbon at residue 66 (Fig. 4*C*), which is indicative of red chromophore formation. However, these structures did show well-resolved tetrahedral sp^3-hybridized C-alpha atoms at position 66 (Fig. 4*D*), indicating green chromophores. This observation is consistent with the weaker intensity observed for the red chromophore absorption peak (~590 nm) of mPlumAYC-E16A and for the absence of such a peak for mPlumAYC (Fig. 2). This weaker or absent peak corresponds to a concentration of red chromophore-containing molecules that is substantially lower or absent. In turn, this smaller population of red chromophore molecules in solution contributes to a smaller population of these molecules in the crystal and consequently an

absence of appreciable electron density. Conversely, the structure of mPlum-E16P presents no appreciable omit map electron density for the green chromophore when contoured at 3σ. Only density corresponding to a planar sp^2-hybridized C-alpha carbon at residue 66 (Fig. 4C), indicating the red chromophore, is apparent. Notably, this correlates with the near complete red chromophore maturation observed for mPlum-E16P. During analysis of the electron density in all three structures, occupancy refinement starting with 50% of both red and green chromophores terminated by driving the appropriate green or red population to 0% occupancy.

For chains A and B, the dihedral angles of the acylimine-containing peptide bonds in mPlum-E16P are 87.5° and 88.3°, which are both close to the theoretically expected value of 90° for RFP red chromophores [24]. For mCherry, this angle is about 30° larger at 119.7°, and for the red chromophore of mPlum, this angle is 168.8°. On the acylimine side of the chromophore, the Pro16 residue in mPlum-E16P appears to slightly distort the secondary structure of the first β-strand in the protein, but does not cause any major structural changes. Specifically, the φ and ψ angles for Pro16 in chain A of mPlum-E16P are −97.9° and 141.1°, respectively; in chain B, they are −91.0° and 139.2°. For mPlum, the corresponding angles at the red chromophore-proximal Glu16 are −126.2° (φ) and 145.7° (ψ). The canonical angles for a protein β-sheet are φ = −135° and ψ = 135°. The allowed φ angles of a proline residue are restricted due to the five-membered ring in this amino acid involving the backbone amine nitrogen. Thus, a deviation of the expected φ angle for a β-sheet at Pro16 would be anticipated.

Except for the expected φ angle change at position 16 due to the presence of a proline residue, there are no substantial φ and ψ angle-changing effects propagated down the rest of the first β-strand in mPlum-E16P in either direction. In mPlumAYC-E16A, the first β-strand of the protein, which contains position 16, is indistinguishable in structure from that of the Glu16-containing proteins mPlum and mPlumAYC. For Ala16 in mPlumAYC-E16A, φ = −129.4° for chain A and −129.1° for chain B, and the ψ angles for this residue are 138.9° in chain A and 139.0° in chain B. In mPlumAYC, the Glu16 residue has φ = −128.6° and ψ = 140.0°.

When comparing the crystal structures of mPlum, mPlum-E16P, mPlumAYC, and mPlumAYC-E16A, we observed strong overall similarity for residues surrounding the chromophore, including the catalytic residues Arg96 and Glu215. The only appreciable difference we could find among these three structures was in the crystallographic side chain conformation of Lys70. In mPlum-E16P, the Lys70 side chain adopts a conformation similar to that found in the fast and efficient maturing RFP mCherry [4]. However, in mPlum, the red chromophore-adjacent Lys70 residue adopts a slightly different conformation, causing the NZ atom to retract away from the chromophore by 1.4 Å with respect to its position in mPlum-E16P (Fig. 4C). In mPlumAYC and mPlumAYC-E16A, the presence of Tyr197 blocks the space where the NZ atom of the Lys70 side chain in mPlum-E16P and mCherry would sit (Fig. 4D). In these AYC motif-containing proteins, the side chain of Lys70 forms three H-bonding interactions: one with the side chain of Glu148, another with the side chain of Tyr197, and a third with a water molecule (Fig. 4A, B). In contrast, the side chain of Lys70 in mPlum and mPlum-E16P only forms two H-bonding interactions: one with the side chain of Glu148 and another

with a water molecule. The AYC motif therefore appears to lock the terminal amine of Lys70 into a position roughly 6 Å away from the chromophore (Table 2). In contrast, the terminal amino groups of Lys70 in mPlum-E16P and mPlum are able to sit about 4 Å and 5 Å away (Table 2), respectively, from the nearest red chromophore heavy atom (CB2, the bridging methylene carbon between the phenol and imidazolinone moieties of the chromophore). Thus, we propose that the AYC motif slows down (mPlumAYC-E16A and mPlumAYC-E16P) or completely hinders (mPlumAYC) the second oxidation reaction during chromophore maturation (Fig. 1, **6** to **7**). It does so by pushing Lys70 away from the chromophore and locking it into a more distal conformation through an additional H-bond. Thus, sequestration of the Lys70 side chain away from the chromophore results in the observed maturation deficiency.

In support of the above statement, we see a general trend toward less efficient red chromophore maturation for longer NZ-to-chromophore distances in our mPlum mutants. Moreover, we observed a correlation between a longer crystallographic distance from the Lys70 terminal amino group to the chromophore and the final proportion of red versus green chromophore-containing molecules in the RFP population. Specifically, we observe a negative correlation between the Lys70 NZ-to-chromophore distance and the red/green ratio (Table 2). Most DsRed-derived RFPs have a conserved lysine at position 70, and it was shown that the K70M mutation in DsRed suppresses red chromophore formation completely [3, 12]. Recently, Lys70 was proposed to be important for the formation of the intermediate blue species through an electrostatic interaction with the chromophore's imidazolinone oxygen atom [10]. Our observation that proximity to the

chromophore of the Lys70 side chain correlates with RFPs that mature more efficiently agrees with this proposal.

Maturation promoting mutations. Because FPs need to fold before any of the chemical transformations required to produce the chromophore can occur [28, 29], chromophore maturation is the rate-limiting step for FPs to become fluorescent. For most imaging experiments, RFPs that mature quickly to the red chromophore are preferred. Commonly-used FPs mature with half-times of 40 min to 2 h at 37°C [30]. To eliminate background signal, complete maturation to the red chromophore is also desirable. Furthermore, RFPs that express as mixtures of molecules containing either the green or red chromophore are inconvenient for multicolor labeling applications. As a result, much effort has been devoted to the optimization of red chromophore maturation in RFPs [16, 17, 31].

Mutations that improve red chromophore maturation in RFPs have been discovered primarily through screening of large mutant libraries obtained through random mutagenesis. Much work has focused on improving DsRed maturation, resulting in a number of mutants displaying accelerated maturation and decreased green chromophore formation [16, 17, 31, 32]. Positions in the sequence of DsRed that are known to affect chromophore maturation include residues 42, 66, 70, 71, 83, 105, 163, and 217. DsRed mutations N42Q [17], V105A [32], Q66M [3, 16], and M163Q [16] have been shown to improve maturation by accelerating red chromophore formation or decreasing green chromophore formation. Conversely, mutations V71M [32], K70M [12] and K83R [33] have been shown to inhibit red chromophore formation and result in green fluorescent proteins. Of the residues involved in the AYC motif, position 217 has been shown to accelerate red chromophore maturation when mutated to alanine [17].

In this study, mutations that hinder or improve red chromophore maturation in DsRed-derived RFPs were rationally designed. We converted the red fluorescent protein mPlum into a strictly yellow-emitting FP by introducing three mutations around the chromophore (T195A, I197Y, and A217C). Then, we used rational design to successfully reintroduce red chromophore formation into mPlumAYC and to drastically improve red chromophore maturation efficiency in mPlum. The rationally designed mutations that resulted in the greatest improvement of chromophore maturation efficiency were E16A and E16P.

One other example of the conversion of green chromophore-containing FPs into RFPs via mutagenesis has been reported. In 2008, Mishin et al. used random mutagenesis and high-throughput screening to convert *Aequorea victoria* GFP into a mutant that emits at 585 nm [34]. The authors proposed that the red fluorescence-emitting GFP mutant undergoes a dehydration reaction at Ser65 of the chromogenic tripeptide (Ser65-Tyr66-Gly67), which results in a dehydroalanine residue that tautomerizes to an acylimine group. The presence of this acylimine extends the conjugation of the chromophore, resulting in red fluorescence at 585 nm. Unlike this prior study, our results represent the first instance in which a green chromophore-containing FP was converted to a red chromophore-containing FP by rational design. Moreover, this conversion was accomplished without reverting any of the mutations causing the maturation deficiency and without altering the chromophore's covalent structure.

Conclusion

In this study, we converted a yellow fluorescent mutant of mPlum into a red-emitting RFP using principles of rational design. Based on our observations, we propose two structural features that are important for efficient red chromophore formation in DsRed-

derived RFPs. The first is the presence of a lysine residue at position 70 that is able to interact directly with the chromophore. Such an interaction has already been shown to be important for red chromophore formation [3, 10]. To interact efficiently with the chromophore, this lysine must be within a certain distance range and must not be locked in a conformation that sequesters the side chain NZ atom away from the chromophore. The second structural feature is an absence of non-bonding interactions limiting the conformational flexibility at the peptide bond that is oxidized to form an acylimine during red chromophore formation. Since the geometry of the C-alpha atom in the first amino acid of the chromogenic tripeptide (residue 66 for mPlum) must change from tetrahedral to trigonal planar during oxidation, flexibility is needed for this process to occur efficiently. Our data, as well as structural features for known acylimine-forming FPs, support this proposition. Satisfying or improving these structural features in other maturation-deficient RFPs may result in RFPs with faster and more complete maturation to the red chromophore.

Materials and Methods

Materials. All reagents used were of the highest available purity. Restriction enzymes and DNA-modifying enzymes were from New England Biolabs. Synthetic oligonucleotides were obtained from Integrated DNA Technologies, and Ni-NTA agarose resin was obtained from Qiagen. CelLytic B buffer and lysozyme were purchased from Sigma-Aldrich. All aqueous solutions were prepared using water purified with a Millipore BioCell system.

Mutagenesis. The mPlum gene was PCR-amplified from plasmid mPlum-pBAD (provided by R.Y. Tsien, UCSD) and subcloned into pET11-a (Novagen) via

NdeI/BamHI. The plasmid was then transformed into *Escherichia coli* XL-1 Blue. The entire NdeI/BamHI fragments, including the whole coding region, were verified by DNA sequencing. All mutations were introduced into the mPlum gene by overlap extension mutagenesis [35] using VentR DNA polymerase (NEB). Briefly, external primers were used in combination with sets of complementary pairs of degenerate oligonucleotides containing the desired mutations in individual PCR reactions. The resulting overlapping fragments were gel-purified (Qiagen) and recombined by overlap extension PCR. The resulting amplicons were digested with NdeI/BamHI, gel-purified, and ligated into pET11a expression vector with T4 ligase.

Characterization of mPlum mutants. The plasmids prepared as described above were transformed into chemically competent *E. coli* BL21-Gold(DE3) cells (Stratagene). Colonies were picked into individual wells of Nunc V96 MicroWell polypropylene plates containing 200 μL of medium (LB with 100 μg/mL ampicillin supplemented with 10% glycerol). The plates were covered with a sterile Breathe-Easy gas permeable sealing membrane (Sigma) and incubated overnight at 37°C with shaking. After incubation, these mother plates were used to inoculate sterile Nunc V96 MicroWell polypropylene plates (daughter plates) containing 300 μL of Overnight Express Instant Terrific Broth media (Novagen) supplemented with 100 μg/μL ampicillin per well. Daughter plates were sealed with breathable membranes and incubated overnight (37°C, 250 rpm shaking). After incubation, the cells were harvested by centrifugation and the cell pellets were washed twice with PBS (pH 7.4). Washed cell pellets were then incubated at 4°C for 72 h to allow chromophore maturation. These pellets were resuspended in PBS and transferred to a Fluotrac 96-well plate (Greiner Bio-One) for screening. Fluorescence was measured

with a Tecan Safire2 plate reader. Emission spectra (λ_{ex} = 570 nm) were measured from 590 nm to 700 nm.

Protein expression and purification for spectral characterization and crystallization. Protein was expressed in 1.0 L cultures by transformation of a pET11-a vector containing the gene of interest into *E. coli* BL21-Gold(DE3) and purified by Ni-NTA affinity chromatography according to the manufacturer's protocol. Column elutions were desalted by gel filtration using a Superdex 75 10/300 GL Tricorn resin column (GE Healthcare) into a final buffer solution of 50 mM phosphate buffer, pH 7.5, and 150 mM NaCl.

Spectroscopic characterization. All absorbance and emission spectra were recorded with a Tecan Safire2 plate reader in Greiner UV-Star 96-well plates. Proteins purified as described above were quantified using the alkali denaturation method [33]. Briefly, RFPs were alkali-denatured with an equal volume of 2 M NaOH. It is known that the alkali-denatured RFP chromophore converts to a GFP-like one, with extinction coefficient 44,000 M^{-1} cm^{-1} at 452 nm under these conditions. Absorbance, emission, and excitation spectra were recorded in PBS. Path lengths for each well were calculated ratiometrically using the difference in absorbance of PBS at 900 nm and 998 nm. Based on the absorbance spectra of native proteins and the concentration determination of alkali-denatured proteins, molar extinction coefficients were calculated. For determination of quantum yields, the integrated fluorescence intensity of mutants of interest was compared with that of equally absorbing samples of mCherry and DsRed (quantum yields 0.22 and 0.80, respectively) with excitation at 550 nm.

pH Studies and pKa measurements. pH titrations were performed using a range of buffers from pH 2.0 to 9.5. Proteins were diluted into these buffers to a concentration of

5–10 µM. Fluorescence and absorbance scans were taken at each pH value using a Tecan Safire2 plate reader. The Henderson-Hasselbach equation was used to calculate the pKa for each protein.

Maturation studies. To start, 30 mL cultures were inoculated with frozen cell stocks of *E. coli* BL21-Gold(DE3) containing the gene of interest in a pET11-a vector. After growing for 2–3 h at 37°C with shaking, cultures were induced with 5 mM IPTG, then capped and sealed to create an anaerobic environment. Proteins were expressed under these anaerobic conditions for 3–4 h at room temperature. After expression, cells were harvested by centrifugation and resuspended in 1.8 mL deoxygenated lysis buffer (50 mM Tris-HCl, pH 8.0, 300 mM NaCl, 5 mM imidazole, 1X CelLytic B, 1 mg/mL lysozyme, and 25 U/mL benzonase nuclease (Novagen)). Resuspended cell lysates were sealed from the air, then incubated at room temperature for 30–40 min without shaking to allow for complete cell lysis by lysozyme. After centrifugation, clarified lysates were recovered and proteins were quickly purified by Ni-NTA affinity chromatography at 4°C. Absorbance and emission spectra were recorded with a Tecan Safire2 plate reader. Spectroscopic data collection for maturation studies was performed at 28°C for all proteins studied here. All experiments were performed in triplicate.

Crystallography. A very large purple crystal of mPlum-E16P was grown in 0.3 µL × 0.3 µL sitting drops with a precipitant solution of 200 mM MgCl, 100 mM sodium cacodylate, pH 6.5, and 50% (v/v) polyethylene glycol 200. This large rhomboidal crystal had approximate dimensions 0.4 mm × 0.5 mm × 0.1 mm. Bluish crystalline chunks of mPlumAYC-E16A were grown in sitting drops with 0.3 µL protein solution and 0.3 µL of the same precipitant used to crystallize mPlum-E16P; approximate dimensions were

0.05 mm × 0.10 mm × 0.08 mm. Bright yellow cubes of mPlumAYC were grown in sitting drops with 0.3 µL protein solution and 0.3 µL of a precipitant solution consisting of 200 nM sodium cacodylate, 100 mM Tris-HCl, pH 8.5, and 30% (w/v) polyethylene glycol 4000. These cubes had approximate dimensions of 0.05 mm × 0.05 mm × 0.05 mm. All datasets were collected at the Stanford Synchrotron Radiation Lightsource (SSRL) beamline 12-2. IPMOSFLM [36] was used for integration and SCALA [37] was used for merging and scaling. All datasets collected were solved by molecular replacement using PHASERMR [38]. The search model used consisted of the PDB coordinates from mPlum (2QLG [39]) with the chromophore removed.

Refinement was accomplished using REFMAC5 [40, 41] and PHENIX [42]. PHENIX was used specifically for refinement of atomic occupancies. Model building was done with COOT [43], wherein water molecules were added manually when they were within H-bonding distance of other heteroatoms (2.3–3.5 Å) and had peaks in the $F_o - F_c$ map of greater than 3.5 σ. In addition, water molecules were removed when they had equivalent isotropic B-factors greater than 60–80 $Å^2$. During generation of R-factors, 5% of data was excluded for cross-validation with an R_{free} value. Crystallographic R-factors were calculated in the standard fashion ($R = \sum |F_{obs} - F_{calc}| / \sum F_{obs}$). For all crystal structures, the final refinement steps were carried out with 20 translation-libration-screw (TLS) groups per protein molecule [44]. TLS groups were identified automatically by using the TLS Motion Determination web server (TLSMD) [45]. Riding hydrogens were included in the refinement of all structures for non-water molecules. The library file for the chromophore was built based on the CH6 chromophore deposited in the Hetero-compound Information Centre - Uppsala (HIC-Up) online database.

References

1.   Verkhusha VV & Lukyanov KA (2004) The molecular properties and applications of *Anthozoa* fluorescent proteins and chromoproteins. *Nature Biotechnology* 22(3):289-296.

2.   Giepmans BN*, et al.* (2006) The fluorescent toolbox for assessing protein location and function. *Science* 312(5771):217-224.

3.   Tubbs JL, Tainer JA, & Getzoff ED (2005) Crystallographic structures of Discosoma red fluorescent protein with immature and mature chromophores: Linking peptide bond trans-cis isomerization and acylimine formation in chromophore maturation. *Biochemistry* 44:9833-9840.

4.   Shu X*, et al.* (2006) Novel chromophores and buried charges control color in mFruits. *Biochemistry* 45:9639-9647.

5.   Wall MA, Socolich M, & Ranganathan R (2000) The structural basis for red fluorescence in the tetrameric GFP homolog DsRed. *Nature Structural Biology* 7(12):1133-1138.

6.   Verkhusha VV*, et al.* (2004) Common pathway for the red chromophore formation in fluorescent proteins and chromoproteins. *Chemistry & Biology* 11(6):845-854.

7.   Strack RL*, et al.* (2010) Chromophore formation in DsRed occurs by a branched pathway. *Journal of the American Chemical Society* 132(24):8496-8505.

8.   Wilmann PG*, et al.* (2005) The 2.1 Å crystal structure of the far-red fluorescent protein HcRed: inherent conformational flexibility of the chromophore. *Journal of Molecular Biology* 349(1):223-237.

9.   Subach OM*, et al.* (2010) Structural characterization of acylimine-containing blue and red chromophores in mTagBFP and TagRFP fluorescent proteins. *Chemistry & Biology* 17(4):333-341.

10.   Bravaya KB*, et al.* (2012) Insight into the common mechanism of the chromophore formation in the red fluorescent proteins: The elusive blue intermediate revealed. *Journal of the American Chemical Society* 134(5):2807-2814.

11.   Pouwels LJ*, et al.* (2008) Kinetic isotope effect studies on the de novo rate of chromophore formation in fast- and slow-maturing GFP variants. *Biochemistry* 47(38):10111-10122.

12. Baird GS, Zacharias DA, & Tsien RY (2000) Biochemistry, mutagenesis, and oligomerization of DsRed, a red fluorescent protein from coral. *Proceedings of the National Academy of Sciences, USA* 97:11984-11989.

13. Abbyad P, *et al.* (2007) Dynamic Stokes shift in green fluorescent protein variants. *Proceedings of the National Academy of Sciences, USA* 104(51):20189-20194.

14. Strongin DE, *et al.* (2007) Structural rearrangements near the chromophore influence the maturation speed and brightness of DsRed variants. *Protein Engineering Design and Selection* 20(11):525-534.

15. Baird GS, Zacharias DA, & Tsien RY (2000) Biochemistry, mutagenesis, and oligomerization of DsRed, a red fluorescent protein from coral. *Proceedings of the National Academy of Sciences, USA* 97:11984-11989.

16. Shaner NC, *et al.* (2004) Improved monomeric red, orange, and yellow fluorescent proteins derived from Discosoma sp. red fluorescent protein. *Nature Biotechnology* 22(12):1567-1572.

17. Bevis BJ & Glick BS (2002) Rapidly maturing variants of the Discosoma red fluorescent protein (DsRed). *Nature Biotechnology* 20(1):83-87.

18. Shcherbo D, *et al.* (2009) Far-red fluorescent tags for protein imaging in living tissues. *Biochemistry Journal* 418(3):567-574.

19. Chica RA, *et al.* (2010) Generation of longer emission wavelength red fluorescent proteins using computationally designed residues. *Proceedings of the National Academy of Sciences, USA* 107(47):20257-20262.

20. Wachter RM, *et al.* (1998) Structural basis of spectral shifts in the yellow-emission variants of green fluorescent protein. *Structure* 6(10):1267-1277.

21. Wang L, *et al.* (2004) Evolution of new nonantibody proteins via iterative somatic hypermutation. *Proceedings of the National Academy of Sciences, USA* 101(48):16745-16749.

22. Shu X, *et al.* (2009) Unique interactions between the chromophore and glutamate 16 lead to far-red emission in a red fluorescent protein. *Protein Science* 18(2):460-466.

23. Wiehler J, von Hummel J, & Steipe B (2001) Mutants of Discosoma red fluorescent protein with a GFP-like chromophore. *FEBS Letters* 487(3):384-389.

24. Yarbrough D, *et al.* (2001) Refined crystal structure of DsRed, a red fluorescent protein from coral, at 2.0 Å resolution. *Proceedings of the National Academy of Sciences, USA* 98(2):462-467.

25. Wachter RM, Watkins JL, & Kim H (2010) Mechanistic diversity of red fluorescence acquisition by GFP-like proteins. *Biochemistry* 49(35):7417-7427.

26. Wilmann PG, *et al.* (2005) The 2.1 Å crystal structure of the far-red fluorescent protein HcRed: Inherent conformational flexibility of the chromophore. *Journal of Molecular Biology* 349:223-237.

27. Griesbeck O, *et al.* (2001) Reducing the environmental sensitivity of yellow fluorescent protein. Mechanism and applications. *Journal of Biological Chemistry* 276(31):29188-29194.

28. Sniegowski JA, *et al.* (2005) Base catalysis of chromophore formation in Arg96 and Glu222 variants of green fluorescent protein. *Journal of Biological Chemistry* 280(28):26248-26255.

29. Rosenow MA, *et al.* (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.

30. Chudakov DM, *et al.* (2010) Fluorescent proteins and their applications in imaging living cells and tissues. *Physiological Reviews* 90(3):1103-1163.

31. Strack RL, *et al.* (2008) A noncytotoxic DsRed variant for whole-cell labeling. *Nature Methods* 5(11):955-957.

32. Terskikh AV, *et al.* (2002) Analysis of DsRed Mutants. Space around the fluorophore accelerates fluorescence development. *Journal of Biological Chemistry* 277(10):7633-7636.

33. Gross LA, *et al.* (2000) The structure of the chromophore within DsRed, a red fluorescent protein from coral. *Proceedings of the National Academy of Sciences, USA* 97(22):11990-11995.

34. Mishin AS, *et al.* (2008) The first mutant of the *Aequorea victoria* green fluorescent protein that forms a red chromophore. *Biochemistry* 47(16):4666-4673.

35. Ho SN, *et al.* (1989) Site-directed mutagenesis by overlap extension using the polymerase chain reaction. *Gene* 77(1):51-59.

36. Leslie AGW (1992) Recent changes to the MOSFLM package for processing film and image plate data. *Joint CCP4 and ESF-EAMCB Newsletter on Protein Crystallography* 26.

37. Evans P (2006) Scaling and assessment of data quality. *Acta Crystallographica Section D: Biological Crystallography* 62:72-82.

38. Mccoy A*, et al.* (2007) Phaser crystallographic software. *Journal of Applied Crystallography* 40:658-674.

39. Shu X*, et al.* (2006) Novel chromophores and buried charges control color in mFruits. *Biochemistry* 45(32):9639-9647.

40. Murshudov GN, Vagin AA, & Dodson EJ (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallographica Section D: Biological Crystallography* 53:240-255.

41. Vagin AA*, et al.* (2004) REFMAC5 dictionary: Organization of prior chemical knowledge and guidelines for its use. *Acta Crystallographica Section D: Biological Crystallography* 60:2184-2195.

42. Afonine PV, Grosse-Kunstleve RW, & Adams PD (2005) The PHENIX refinement framework. *CCP4 Newsletter* 42.

43. Emsley P & Cowtan K (2004) Coot: Model-building tools for molecular graphics. *Acta Crystallographica Section D: Biological Crystallography* 60:2126-2132.

44. Painter J & Merritt EA (2006) Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. *Acta Crystallographica Section D: Biological Crystallography* 62:439-450.

45. Painter J & Merritt EA (2006) TLSMD web server for the generation of multi-group TLS models. *Journal of Applied Crystallography* 39:109-111.

Figures & Legends.
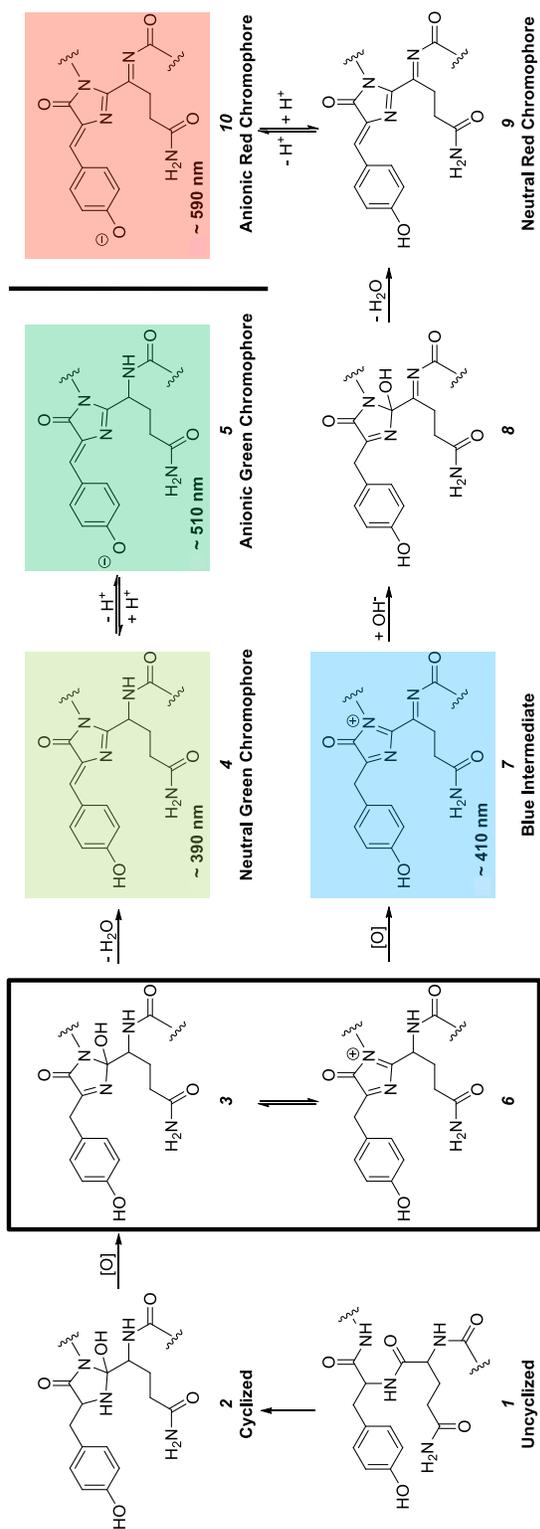
Fig. 1. Chromophore maturation mechanism. The chromophore maturation mechanism proposed by Strack *et al.* is a branched pathway ending with either red or green chromophore formation. Intermediates and final products on this pathway that are observable by visible spectroscopy are color-coded and labeled with the approximate wavelength of their peak absorption.
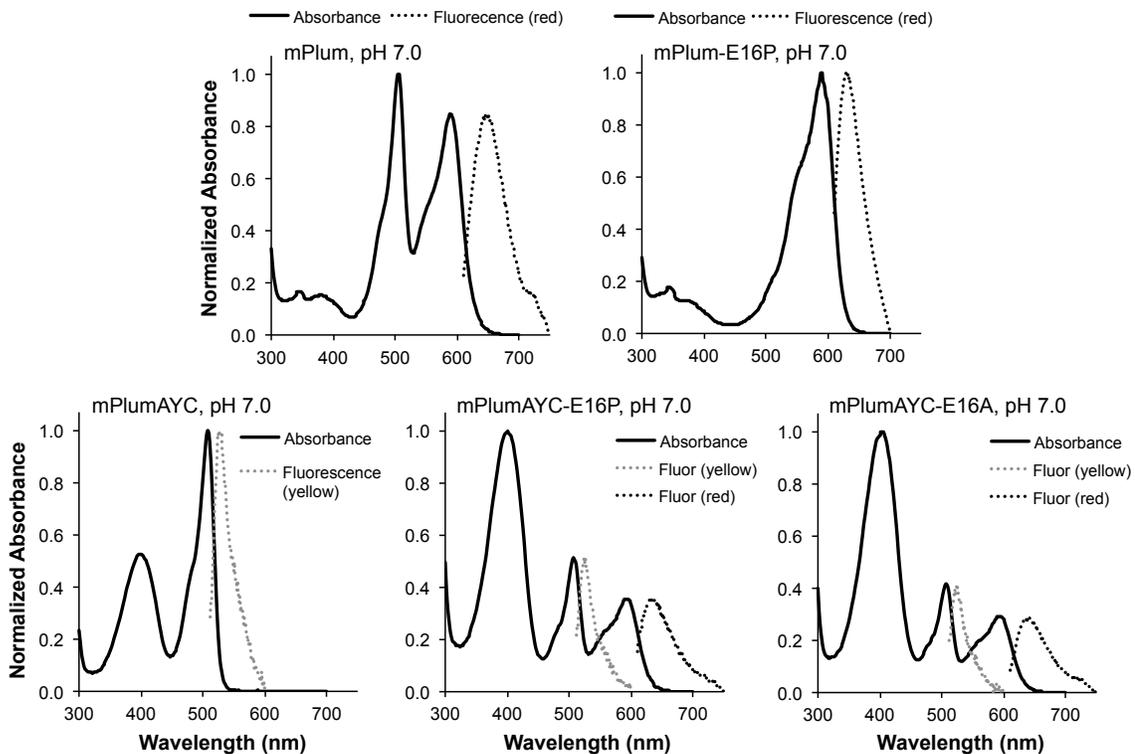
Fig. 2. Absorption and fluorescence spectra of various RFPs. Absorption spectra (heavy black lines) are normalized to the largest intensity absorbance peak present in each spectrum. Fluorescence emission spectra (dotted lines) are normalized to the absorbance peak in each spectrum corresponding to the excitation wavelength used to induce fluorescence. All spectra were measured at pH 7.0.
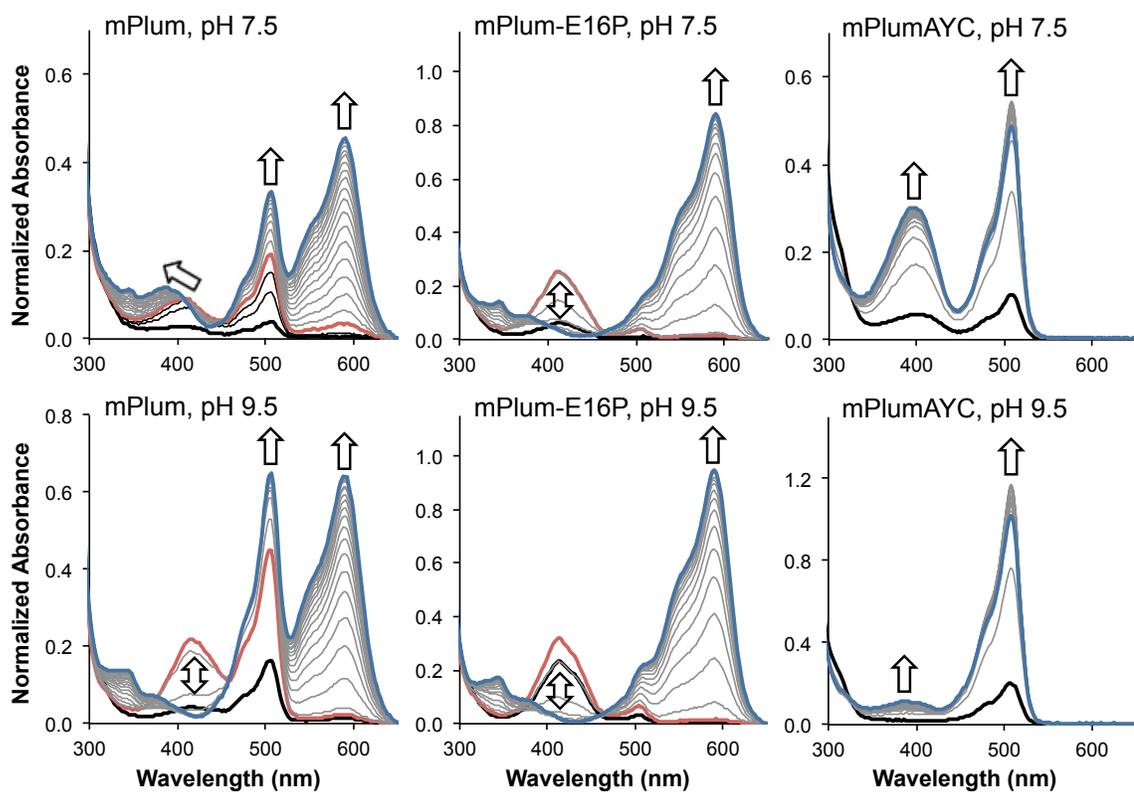
Fig. 3. Maturation experiments. All spectra are normalized to the 280 nm absorbance peak. Heavy black and blue traces represent the beginning (t = 0 h) and end (t = 20 h) of the maturation experiment, respectively. The distance in time between each gray or black trace is 1.0 h. Arrows indicate the primary direction of peak movement during maturation. Each heavy red trace indicates the point in time when the 410 nm absorbance peak reached its maximum during the course of maturation. Black traces occur before the 410 nm peak reaches its maximum level; gray traces occur after the maximum.
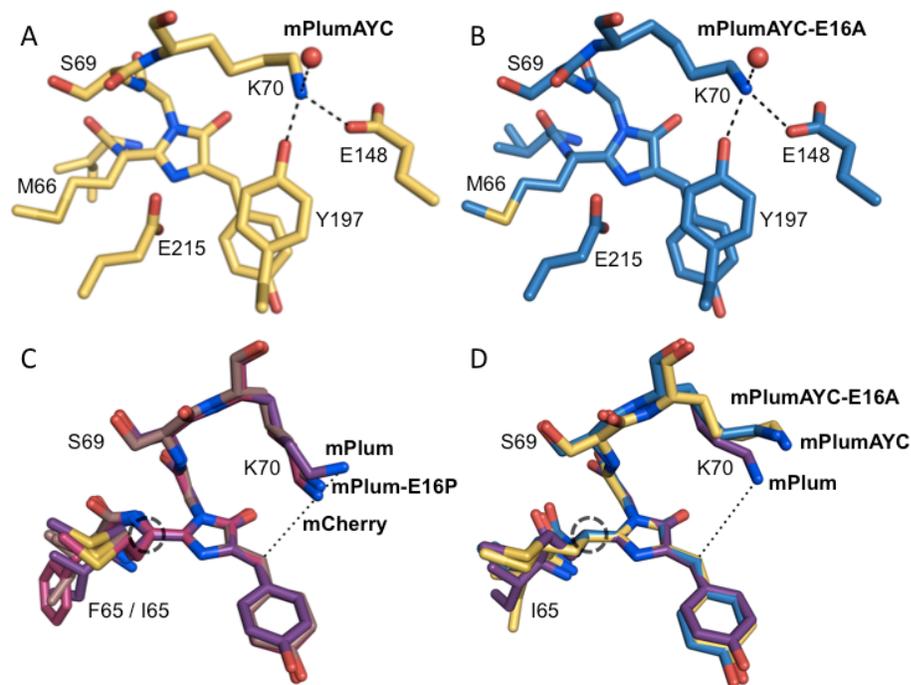
Fig. 4. Crystal structures. (*A* and *B*) Introduction of the AYC motif results in π-stacking interactions between the chromophore and Tyr197 in both mPlumAYC (*A*) and mPlumAYC-E16A (*B*). H-bonding interactions with Lys70 are illustrated with dashed lines. These interactions combine to sequester the terminal amino group of Lys70 away from the chromophore. (*C* and *D*) Comparisons of Lys70-to-chromophore distance are illustrated between mCherry (pink), mPlum (purple), and mPlum-E16P (mauve) (*C*), as well as between mPlum (purple), mPlumAYC (yellow), and mPlumAYC-E16A (blue) (*D*). A dotted line connects the NZ atom of Lys70 in mPlum to the CB2 atom of the mPlum red (*C*) or green (*D*) chromophore. Note that Lys70 in mPlum adopts a slightly different conformation in the red versus green chromophore contexts. All proteins were aligned by the atoms of their five-membered heterocyclic chromophore ring. A dashed circle highlights the C-alpha carbon of residue 66, which is either sp^2-hybridized in the red chromophore (*C*) or sp^3-hybridized in the green chromophore (*D*).

Table 1. Properties of the RFP mPlum and its Mutants

Protein	Mutations				Post-Maturation Absorbance Peaks, nm		Fluorescence Emission Properties: Peak Wavelength, nm & Quantum Yield				pKa	
	16	195	197	217	I(green)	I(red)	I(green)	Φ_F(green)†	I(red)	Φ_F(red)†	green	red
mPlum*	E	T	I	A	391	588	---	---	649	0.10	N.D.	4.2
mPlum-E16P	P	T	I	A	---	590	---	---	630	0.14	---	4.7
mPlumAYC	E	A	Y	C	396	---	527	0.02	---	---	7.3	---
mPlumAYC-E16P	P	A	Y	C	395	594	525	0.02	637	0.04	7.4	5.4
mPlumAYC-E16A	A	A	Y	C	393	592	523	0.02	639	0.05	7.5	5.4

* mPlum also contains the following mutations with respect to mCherry: I7V, K45R, L124V, I161M, Q163M, K166R, K182M, N194K, and N196D (numbering based on DsRed)

† Quantum yields for all proteins were measured at pH 7.5. mPlum exhibits measureable green-yellow fluorescence emission at pH > 7.5, but at pH 7.5, it was not possible to measure a Φ_F(green) value

N.D. = Not Determined

Table 2. Maturation Data

Protein	Maturation Half-Time (green)*, h		Maturation Half-Time (red), h		Time of 410 nm peak maximum, h	A_{590}/A_{510} in mature protein	NZ-CB2 distance
	pH 7.5	pH 9.5	pH 7.5	pH 9.5	pH 9.5	pH 9.5	Å†
mPlum	2.8 ± 0.2	0.47 ± 0.04	7.3 ± 0.1	3.90 ± 0.09	0.62 ± 0.03	0.9	5.6
mPlum-E16P	3.12 ± 0.06	2.18 ± 0.06	3.77 ± 0.03	3.12 ± 0.08	0.38 ± 0.03	3.7	4.2
mPlumAYC	0.89 ± 0.04	0.79 ± 0.01	--¶	--¶	--¶	--¶	6.4
mPlumAYC-E16P	1.43 ± 0.03	1.24 ± 0.04	5.4 ± 0.3	5.6 ± 0.4	2.8 ± 0.1	0.14	---
mPlumAYC-E16A	1.83 ± 0.07	1.20 ± 0.05	5.9 ± 0.2	4.6 ± 0.1	3.8 ± 0.4	0.15	6.5

* Maturation half-times are for the anionic green chromophore that absorbs at ~510 nm.

† Crystallographic NZ-CB2 distances are presented for the major conformer of Lys70 in each structure.

¶ Does not mature to the 410 nm absorbing blue intermediate or to the anionic red chromophore.

Supporting Information

SI Figures & Legends
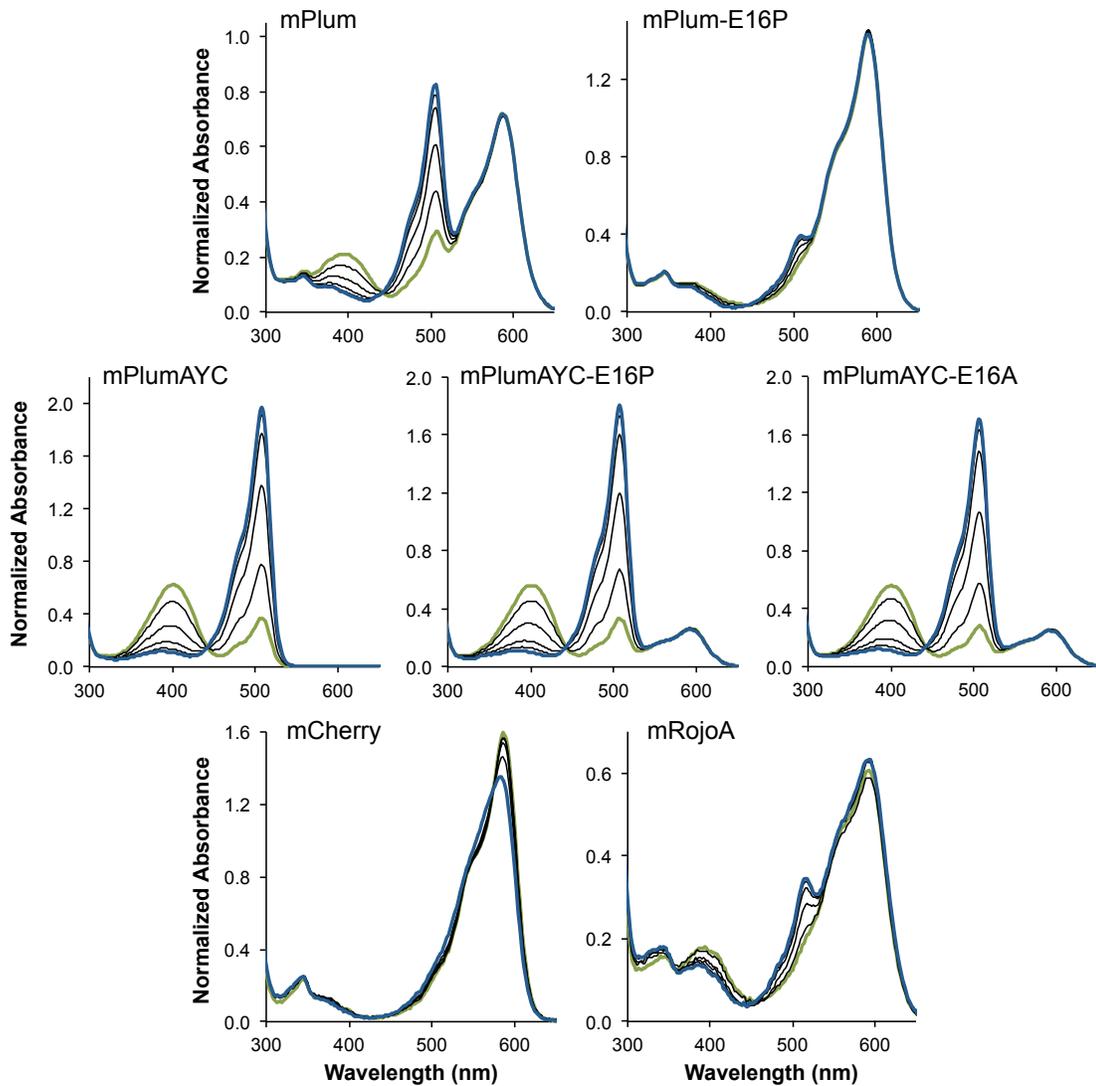
Fig. S1. pH affects the absorption spectra of various RFPs. All spectra are normalized to the 280 nm absorbance peak. Heavy green and blue traces represent spectra taken at pH 7.0 and 9.5, respectively. All remaining black traces are separated from each other by 0.5 pH units (specifically: pH 7.5, 8.0, 8.5, and 9.0).
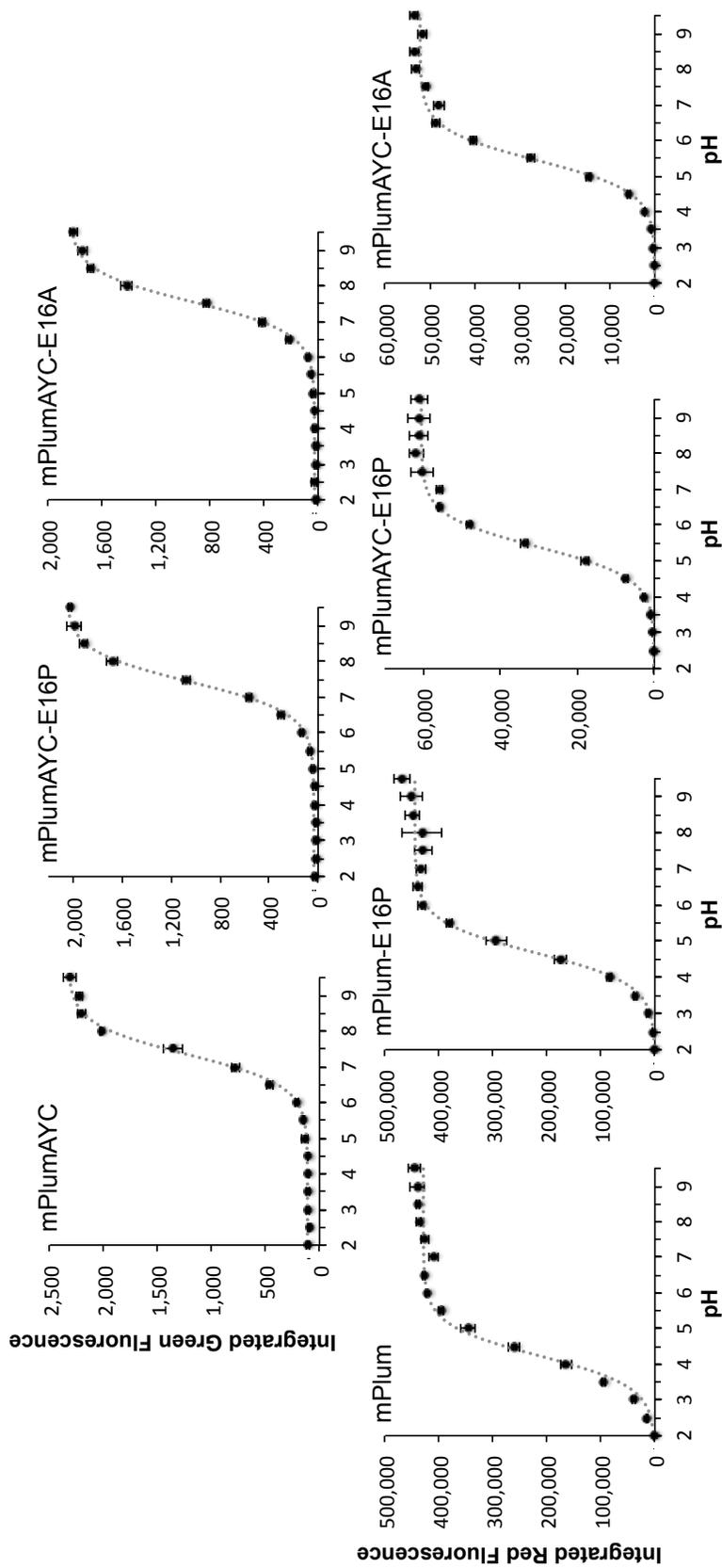
Fig. S2. pH profiles of mPlum and its mutants. Top row: integrated green fluorescence intensity between 510 nm and 600 nm; bottom row: integrated red fluorescence intensity between 590 nm and 700 nm. Data were fit with a Henderson-Hasselbalch model, which was used to determine chromophore pKa's.
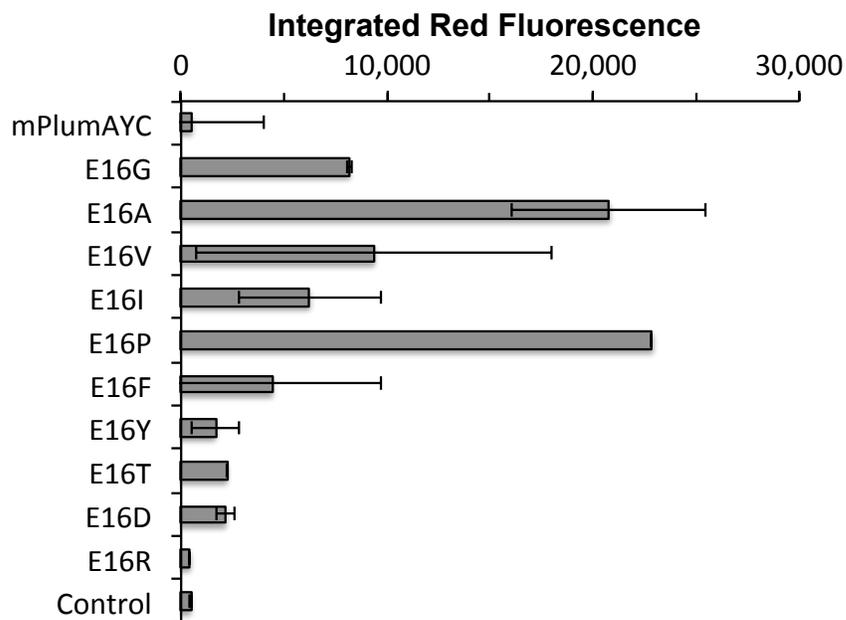
Fig. S3. Red fluorescence level in mPlumAYC E16 mutants. Integrated red fluorescence intensity between 590 nm and 700 nm with excitation at 550 nm was measured for one to six samples of each mPlumAYC E16 point mutant. Shown are the average integrated fluorescence intensities with error bars representing ± 1 σ. The control sample contained an empty protein expression plasmid.
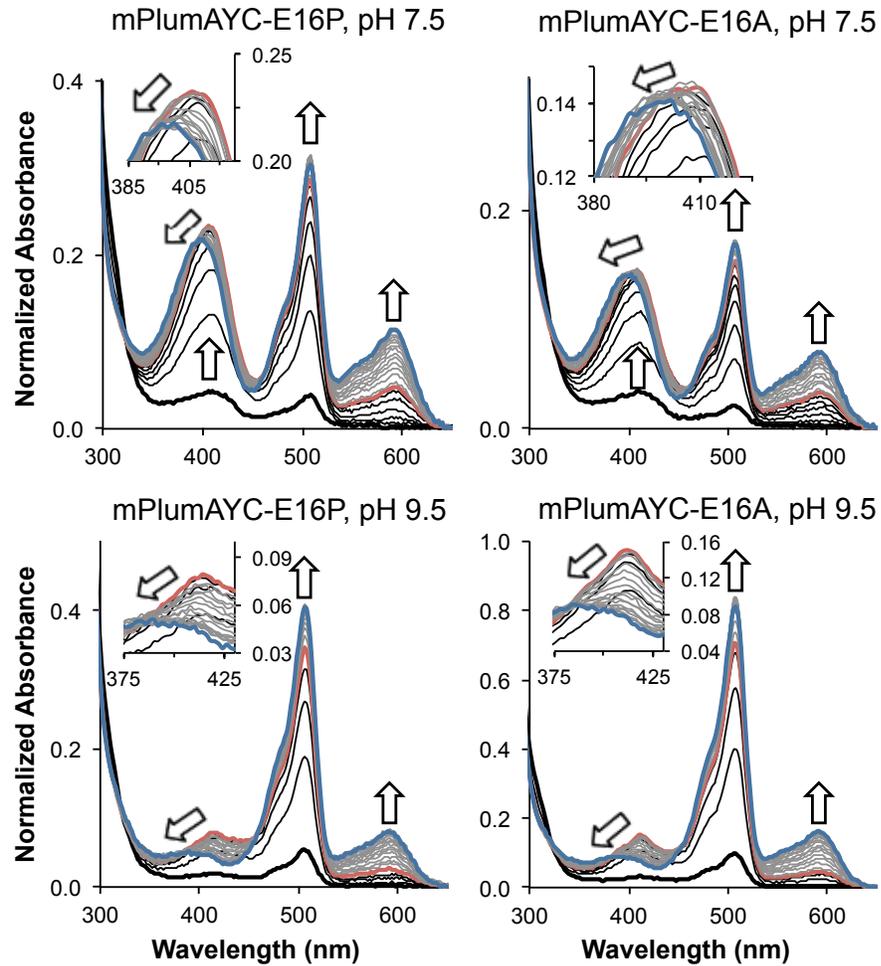
Fig. S4. Maturation experiments with recovery mutants. All spectra are normalized to the 280 nm absorbance peak. Heavy black and blue traces represent the beginning (t = 0 h) and end (t = 20 h) of the maturation experiment, respectively. The time between each gray or black trace is 1.0 h. Arrows indicate the primary direction of peak movement during maturation. Each heavy red trace indicates the point in time when the 410 nm absorbance peak reached its maximum level during the course of maturation. Black traces occur before the 410 nm peak reaches its maximum level; gray traces occur after the maximum. Selected regions of each panel (*insets*) are magnified to illustrate the revelation of a more pronounced peak at 410 nm during maturation at pH 9.5.
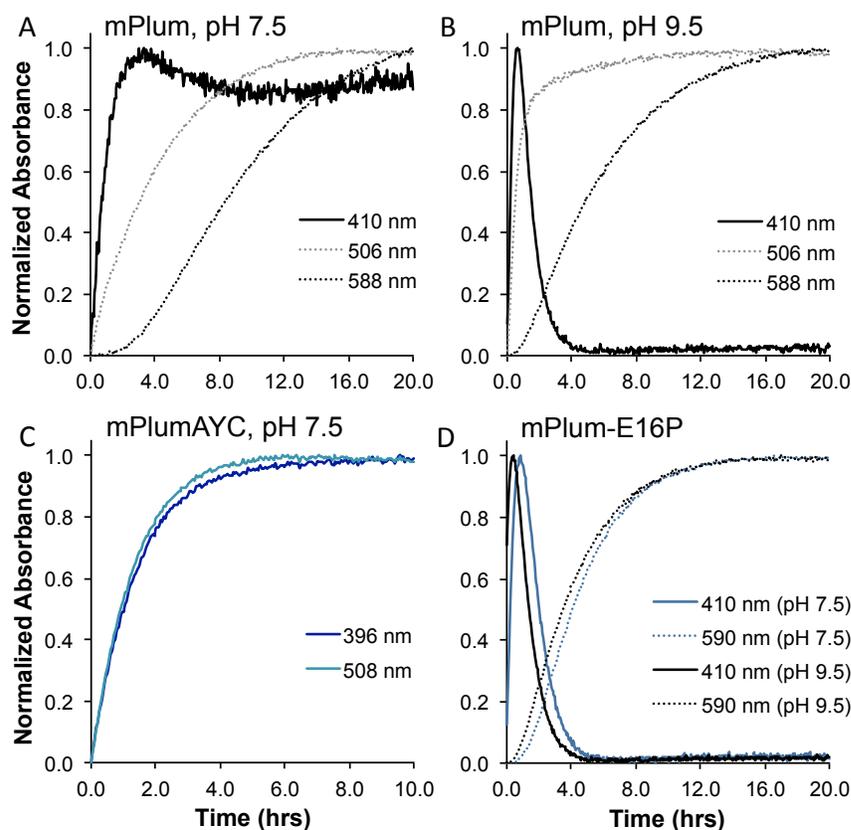
Fig. S5. Maturation kinetics plots. All spectral data is normalized to the maximum peak intensity observed over the course of maturation for each wavelength depicted. Suppression of spectral interference involving the 410 nm absorbance peak is illustrated for mPlum when maturation is tracked at higher pH (*A* and B). In mPlumAYC (*C*), green chromophore maturation half-time is equivalent when tracking both the neutral green chromophore (396 nm) and the ionized green chromophore (508 nm). This result indicates that green chromophore ionization and maturation occur on much different timescales. Lastly, a shift to faster red chromophore maturation half-time and faster arrival at the 410 nm peak maximum occurs when tracking maturation at pH 9.5 in mPlum-E16P (*D*). This shift to shorter half-times can be seen in mPlum as well for pH 7.5 (*A*) versus pH 9.5 (*B*).

Table S1. Crystallography Data

	mPlum-E16P	mPlumAYC	mPlumAYC-E16A
Resolution, Å	33.9–1.65	37.9–1.75	40.3–2.0
Space Group	$P2_12_12_1$	$P22_12_1$	$P2_12_12_1$
Observations	178,646	74,017	61,173
Unique Observations	53,773	22,710	23,844
R_{merge}, %	8.2 (51.1)	22.9 (38.7)	10.3 (15.0)
Completeness, %	97.9 (97.5)	95.5 (99.9)	92.2 (95.1)
I / σI	7.2 (1.7)	3.5 (2.1)	6.0 (4.4)
R_{work} / R_{free}, %	19.0 / 22.7	22.1 / 28.1	21.5 / 27.4
Molec. / Asymm. Unit	2	1	2
Unit Cell Dimensions, Å	a = 61.2	a = 38.7	a = 61.2
	b = 76.9	b = 61.6	b = 64.7
	c = 95.8	c = 96.3	c = 94.8
	$\alpha = \beta = \gamma = 90°$	$\alpha = \beta = \gamma = 90°$	$\alpha = \beta = \gamma = 90°$

Values in parentheses are statistics for the highest resolution shell of data.

A Method for Rapid Identification of Monomeric Fluorescent Protein Mutants Guided by Computational Protein Design

Timothy Wannier[a], Matthew M. Moore[a], Stephen L. Mayo[a,b]

[a]Division of Biology, California Institute of Technology,

1200 E. California Blvd., Pasadena, CA 91125 USA

[b]Division of Chemistry and Chemical Engineering, California Institute of Technology,

1200 E. California Blvd., Pasadena, CA 91125 USA

Abbreviations: RFP, red fluorescent protein; FP, fluorescent protein; CPD, computational protein design; λ_{max}^{abs}, fluorescence emission wavelength maximum; λ_{max}^{ems}, absorbance wavelength maximum; Φ_F, fluorescence quantum yield; DF, dilution factor; AUC, analytical ultracentrifugation

Abstract

Often the usefulness of a fluorescent protein tag is limited by its oligomerization state. To address this issue, a computational protein design (CPD) method was explored for efficiently identifying monomeric mutants of fluorescent proteins derived from class *Anthozoa* organisms. Focused combinatorial libraries of surface mutations were made in a variant of the tetrameric red fluorescent protein DsRed; this variant had a truncated C-terminal tail and thirteen point mutations corresponding to the core of mCherry, a monomeric derivative of DsRed. One library (M-Lib) was composed of sequences designed with an occlusion-based implicit solvation model using CPD. Another, smaller library (R-Lib) was produced through random sampling of amino acids from a distribution of residues found on the surface of proteins in mesophilic bacteria. Protein expression and characterization revealed that 97% of the M-Lib members were more than minimally fluorescent (quantum yield, Φ_F, > 0.1) compared to only 27% of the R-Lib members. A monomeric state was confirmed for fluorescent members of both M-Lib and R-Lib using fluorescence anisotropy, size exclusion chromatography, and analytical ultracentrifugation. We also discovered that only certain sets of core residues enable chromophore formation in an oligomerization state-independent manner. Specifically, we found that the set of core residues in mCherry supports chromophore formation in both oligomeric and monomeric contexts, but the core residues of DsRed only support chromophore formation in an oligomeric context.

Fluorescent proteins (FPs) have revolutionized the world of cell biology [1]. They have proven to be particularly valuable when employed as molecular tags for the *in vivo* monitoring of protein-protein interactions [2-5]. Most of the naturally occurring FPs discovered so far are high-order oligomers [1] and although their oligomeric state does not significantly affect their ability to function as reporters of gene expression [2, 3, 6], it has a dramatic effect on their success as molecular tags [3, 5]. When a protein of interest is fused with an FP tag, any protein-protein interactions that are observed can be called into question if the tag is not monomeric. The ambiguity stems from the fact that the fusion constructs can oligomerize through their non-monomeric FP tags, resulting in non-native protein-protein interactions that are indistinguishable from any native protein-protein interactions that may also occur. This artificial aggregation can also affect diffusion rates and cause problems in protein transport and trafficking.

To avoid these problems, researchers have sought to generate monomeric FPs from oligomeric ones. This is usually accomplished by a disrupt-and-recover approach [7], which starts by identifying mutations that disrupt known protein-protein interfaces. Unfortunately, the disruption often alters the fluorescent properties of the protein significantly, making it necessary to search for mutations to recover fluorescence [7]. This approach can be tedious because full recovery may require multiple rounds of directed evolution and screening, and in some cases, may never be achievable.

There are ten notable literature examples wherein oligomeric FPs have been intentionally converted to monomeric ones [7-16]. Some variation of the disrupt-and-recover approach was used in all these cases, and in the majority of them, some type of fluorescence recovery process was required. There are only two instances (mAG1 [8] and mEosFP

[10]) in which the authors were fortunate enough to find that their interface-disrupting mutations did not cause significant interference with wild-type fluorescent properties. In addition, a recent study using dynamic light scattering indicates that the solution state behavior of some well-known "monomeric" proteins is actually oligomeric [17]. This result illustrates another salient problem with the disrupt-and-recover approach; namely, limited forethought given to a plan for restructuring oligomer-forming interfaces can lead to proteins that are "monomeric enough" to behave appropriately in test systems but that exhibit unreliable oligomerization state fidelity when more scrutiny is applied.

An alternate strategy that may prove more successful is to use a rational design-based approach facilitated by computational protein design (CPD). CPD can be combined with empirical data to produce a medium- to high-throughput process for FP monomerization. CPD-generated focused combinatorial libraries can be screened for monomeric FP mutants using a high-throughput assay based on fluorescence anisotropy measurements [18]. The monomeric nature of these mutants can then be confirmed by size exclusion chromatography [8, 14] and analytical ultracentrifugation [17]. The utility of this approach was explored here. Our work focused on monomerization of the tetrameric FP DsRed, but this method could also be applied to other class *Anthozoa* FPs for which crystal structures already exist.

Results and Discussion

Analysis of C-terminal deletions in DsRed. We initially hypothesized that C-terminal residues after position 220 could be cleaved from any class *Anthozoa* FP of interest as they are not required for protein folding or fluorophore formation [1, 2, 7]. Removal of these residues is desirable because it decreases the computational complexity for CPD

calculations and simplifies gene assembly during library construction. Furthermore, because these C-terminal residues are intimately involved in one of the conserved oligomerization interfaces found in class *Anthozoa* FPs [19-26], we wanted to clarify the extent to which they affect the oligomerization state of DsRed.

We created several DsRed mutants by deleting one to five C-terminal residues and found that deletions of any length significantly altered the solubility and fluorescent properties of this FP. Notably, DsRedΔ225 was the only C-terminal deletion mutant that expressed soluble protein at levels sufficient to purify samples for further characterization. DsRedΔ225 ranks lower than wild-type DsRed in terms of quantum yield (Φ_F) (Table 1). However, λ_{max}^{abs} and λ_{max}^{ems} for these two proteins are the same, giving them very similar absorbance and emission spectra (Fig. S1*A*). Importantly, DsRedΔ225 nearly co-elutes with wild-type DsRed by size exclusion chromatography (Fig. S1*B*), indicating that they have the same oligomerzation state in solution. DsRed is known to primarily form homotetrameric oligomers in solution [2, 7, 16, 24, 27], which we confirmed using analytical ultracentrifugation (Fig. S3).

Characterization of DsRed and mCherry core mutants. Since C-terminal deletions were poorly tolerated in DsRed, we sought a means for buffering this protein against the disruptive effects of this modification. We postulated that mutating the core of DsRed to the core of a well known monomeric derivative of this protein, mCherry [28, 29], would provide the intended buffering effect. In addition, we aimed to isolate any effects oligomerization state would have on the fluorescent properties of our mutants. To this end, we constructed genes for two proteins (Table S1). The first protein, DsRed with 13 core mutations corresponding to the core residues found in mCherry, was dubbed

DsRmCh (DsRed with an mCherry core). The second protein was the inverse mutant and was dubbed mChDsR (mCherry with a DsRed core).

We found the fluorescence and spectral properties of DsRmCh to be nearly identical to those of mCherry (Table 1). The most notable difference is that DsRmCh has 36% higher quantum yield. As expected from its DsRed surface, DsRmCh co-elutes with DsRed by size exclusion chromatography (Fig. S2*A*) and gives a peak corresponding to a solution state homotetramer by analytical ultracentrifugation (Fig. S3).

Removing the C-terminal residues of DsRmCh (DsRmChΔ221–225) had no effect on the protein's fluorescence and spectral properties (Fig. S1*B*). The quantum efficiency of this deletion mutant is statistically equivalent to that of mCherry (Fig. 1*A*). Size exclusion chromatography shows it elutes close to DsRed and DsRedΔ225 (Fig. S2*A*), and analytical ultracentrifugation reveals it is homotetrameric in solution (Fig. S3). Taken together, these results indicate that the C-terminal tail does not determine DsRed's oligomerization state in solution.

In contrast, the inverse mutant of DsRmCh, mChDsR, was completely non-fluorescent. Fortunately, its soluble expression was sufficient to characterize the protein by size exclusion chromatography. The elution profile indicates that its oligomerization state in solution is more similar to mCherry than it is to any of the DsRed surface-bearing proteins discussed above (Fig. S2*A*). Trypsin digest mass spectrometry of mChDsR revealed no evidence of peptide modifications in the string of protein sequence where modifications due to chromophore formation would produce a mass loss. For mCherry, equivalent trypsin digest mass spectrometry indicated mass losses of –20 Da and –2 Da,

which are consistent with chromophore formation and intermediate oxidation on the path to chromophore formation, respectively [1, 2, 30, 31].

To further investigate the distinct non-fluorescent behaviour of mChDsR, we constructed a series of single mutants in which we modified the core of mChDsR or the surface of DsRed. First, we explored single mutations reverting the core of mChDsR back to mCherry (Fig. S4*A*). We discovered that the chromophore-proximal F177V and S197I mutations led to significant red fluorescence recovery in mChDsR. The significance of these mutations is unclear, as a survey of the ten FP monomerization studies mentioned previously indicates that position 177 is typically a phenylalanine [7-16]. Position 197 is known to be involved in π-stacking interactions with the mature FP chromophore [32-34], but its role in promoting or inhibiting red chromophore maturation has not been studied. Similarly, we examined the effect of point mutations on the surface of DsRed leading to the construction of mChDsR (Fig. S4*B*). These experiments revealed the following surface mutations to cause the least disruption of red fluorescence in DsRed: V127T, A164R, Y192A, and Y194N. The Y192A mutation was the best in terms of red fluorescence retention. All four of these mutations correspond to positions central to DsRed's dimeric interfaces [24].

Monomerization methodology. The two dimeric interfaces of tetrameric FPs from class *Anthozoa* organisms have been well characterized in a number of crystal structures [19-26]. These two interfaces (AB and AC) were thoroughly described in the original crystallographic analysis of DsRed [24]. The AB interface is the weaker of the two and its center occurs around residue 125. This interface is reminiscent of most protein-protein interfaces in that the core interfacial residues are hydrophobic and are surrounded by a

shell of hydrophilic residues [35]. The AC interface is the stronger of the two and includes symmetrical interactions between surface patches on each protein as well as their C-terminal tails.

The general layout of these interfaces can be captured schematically using a grid-like diagram, and the number of mutations made at a given interface position can be tabulated from FP monomerization studies reported in the literature [7-16] (Figs. S5 and S6). For oligomeric wild-type FPs from class *Anthozoa*, the geometry for the association of β-barrel subunits is highly conserved [2, 3]. Thus, one can infer the importance of a position toward disruption of an interface from its mutational frequency. This analysis led to 7 positions being identified as most important for disruption of the AB interface (positions 21, 96, 106, 125, 127, 180, and 182). Disruption of the AC interface was associated with 10 positions (positions 145, 149, 153, 162, 164, 174, 176, 192, 194, and 216). Consequently, our CPD monomerization procedure aimed to include diversity at those 17 positions.

We hypothesized that computational design of the surface residues of any class *Anthozoa* oligomeric FP would lead to a set of surface-compatible mutations incapable of supporting the wild-type oligomeric interfaces. In the past, explicit design for homodimeric interactions by CPD has been a challenge, even when abundant design constraints aimed at ensuring molecular complementarity were employed [36, 37]. Thus, we reasoned that the stochastic nature of CPD algorithms coupled with a lack of explicit design toward retention of any molecular complementarity found in the wild-type interfaces would result in a very low probability of new oligomer formation and a reasonable probability of monomerization.

Library design with CPD. We applied a CPD software package developed in-house [32] to design the AB and AC interfaces of a well-studied DsRed crystal structure (PDB code: 1G7K) [38]. Interfaces were designed for one molecule of the DsRed homotetramer using an occlusion-based implicit solvation model [32, 39, 40]. The 17 positions described above were varied in amino acid identity during the design calculation. In addition to the wild-type residue, only non-hydrophobic, non-cysteine, and non-proline residue identities were allowed. The design results were ranked by a molecular dynamics force field-based scoring function [41, 42], and the top 89 sequences were chosen for gene assembly. These sequences made up a putative monomer library and were numbered sequentially as M-Lib #1 through M-Lib #89, with M-Lib #1 being the top-scoring sequence (with the lowest energy) (Table S2).

The results from our characterization of DsRmCh and mChDsR indicated that certain sets of core mutations are incompatible with proper chromophore formation outside the homotetrameric context found in DsRed. In contrast, the constellation of core residues in mCherry sustains proper chromophore formation in both monomeric and oligomeric contexts. These findings led us to conclude that wild-type DsRed would be an inappropriate starting scaffold for designing monomeric FPs with CPD. Accordingly, DsRmChΔ221–225 was chosen to be the parent protein for our putative monomer library of top-ranked CPD sequences. During gene assembly of this library, 6 other derivative sequences were produced (M-Lib #90–95), which did not rank among the top 100 sequences produced by our CPD calculations. Importantly, these derivative sequences did not differ substantially from the other 89 library members of M-Lib, and the consensus

sequence of the inclusive 95 member library is equivalent to the sequence of M-Lib #1 (Fig. S7*A*).

Generation of random library. We also produced a smaller library of sequences using the same interface positions and allowing the same set of residue identities as in the CPD calculation, minus glycine (Table S3). However, in this case, we randomly sampled residue identities from a published amino acid distribution for surface residues in mesophilic bacteria [43]. This distribution was adjusted to only include our chosen design residues plus the wild-type residue at each position—mirroring the residue identities allowed for CPD. This library thus provided a set of sequences with randomly chosen interface residues (Fig. S7*B*) consistent with observed amino acid distributions on the surface of proteins in bacteria. A random library of this nature serves as a check against the effectiveness of our CPD monomerization methodology in producing libraries enriched with soluble protein mutants [44]. This 22-member random library (R-Lib) was also constructed via gene assembly using DsRmChΔ221–225 as the parent protein.

Monomer library and random library results. Expression of the 95-member M-Lib revealed that nearly all (92) were more than minimally fluorescent ($\Phi_F > 0.1$); only 3 mutants showed $\Phi_F < 0.1$: M-Lib #53, 66, and 79. There were no prominent sequence-based differences between these 3 proteins and the rest of the sequences in M-Lib (Table S2). Also, the 6 derivative sequences (M-Lib #90–95), which were not ranked among the top 100 CPD design sequences, had spectral and fluorescent properties similar to the parent protein, DsRmChΔ221–225 (Table S4). In contrast, 9 of the 22-member R-Lib (41%) were minimally fluorescent, and 6 other R-Lib sequences showed no evidence of chromophore maturation or exhibited zero quantum yield (Table S5).

Therefore, in total, 97% of the M-Lib sequences were more than minimally fluorescent compared to only 27% for R-Lib. This result demonstrates that sequences predicted by CPD are more likely to retain fluorescence in our DsRed-based experimental system.

M-Lib and R-Lib sequences were fairly invariant with regard to λ_{max}^{abs} and λ_{max}^{ems} (Table 2). Distinctly leptokurtic (excess kurtosis > 0.0) distributions were observed for both of these parameters. Leptokurtic distributions are more sharply peaked than a corresponding normal distribution. On balance, distributions measured for M-Lib values were more leptokurtic than for R-Lib values (Table 2). Strong negative skewness (< -1.0) was observed for the distribution of λ_{max}^{abs} in both M-Lib and R-Lib. In most of the parameters measured for M-Lib, distributions were observed to be negatively skewed (Table 2). Negative skewness indicates that the mean of a distribution is less than the median and mode for that distribution. Thus, the majority of measurements for the sequences in M-Lib fall above the mean value for this library.

M-Lib and R-Lib members were also evaluated by the ratio of their λ_{max}^{abs} to λ_{280nm}^{abs} peak heights (Fig. 1B). This ratio gives an indication of the extent of red chromophore maturation, the size of the chromophore extinction coefficient, and the protein expression level in our system. For reference, mCherry consistently produces a $\lambda_{max}^{abs}/\lambda_{280nm}^{abs}$ of 1.3–1.4 after affinity purification in our medium throughput expression system. The average $\lambda_{max}^{abs}/\lambda_{280nm}^{abs}$ for all sequences measurable in M-Lib was 1.00, with a standard deviation of 0.18 (Table 2). Notably, seven of the ten mutants with the highest quantum yields measured for M-Lib showed λ_{max}^{abs} to λ_{280nm}^{abs} ratios lower than one standard deviation below this average (M-Lib #1, 2, 7, 16, 17, 55, & 57) (Table S4). For R-Lib,

the average $\lambda_{max}^{abs}/\lambda_{280nm}^{abs}$ was 0.34, and the standard deviation was 0.45 (Table 2). Larger standard deviations for R-Lib were apparent in nearly all measurement metrics used to evaluate M-Lib and R-Lib (Fig. 1 and Table 2).

With regard to the scoring function used to rank library members, four of the top ten ranked CPD sequences were among the top ten highest measured Φ_F mutants in M-Lib (Table S2 and Table S4). Furthermore, M-Lib #1 and M-Lib #2, the two top-ranking sequences by our CPD scoring function, also had the highest measured Φ_F. A consensus sequence produced for the top ten highest Φ_F mutants in M-Lib did not substantially differ from the consensus sequence for all of M-Lib (Fig. S7*C*). However, N182 was more prevalent among high Φ_F mutants than among any other consensus sequence produced using the top ten ranked mutants along different library metrics (Fig. S7).

Expanding the Φ_F selection, we observed that eight of the top ten mutants with the highest measured Φ_F were within the top twenty CPD ranked sequences. A strong linear correlation between Φ_F and our scoring function was not observed ($R^2 = 0.31$). However, the absence of any correlation between Φ_F and the scoring function value could not be rejected in a two-tailed Spearman rank correlation test at the 99% confidence level. The Spearman test, as opposed to a standard linear correlation coefficient (R^2), is not restricted by the assumption that both dependent and independent variables are normally distributed.

By multiplying Φ_F and $\lambda_{max}^{abs}/\lambda_{280nm}^{abs}$, then scaling by 100, we developed a brightness proxy metric that could be used to rank the effective fluorescence of our mutants (Fig. 1). The average brightness proxy metric for measurable sequences in M-Lib (26.5) was > 3.5

times higher than the R-Lib average (7.2) (Table 2). For comparison, in our expression system, mCherry scores 29.5 ± 1.1 on this metric, whereas DsRed scores 62.5 ± 5.0; DsRmChΔ221–225, the parent protein to M-Lib and R-Lib, scores 24.8 ± 0.6 (Fig. 1*C*). Fourteen members of M-Lib rank higher than mCherry on this metric, the best of which are M-Lib #13 (34.8) and M-Lib #49 (33.0). None of R-Lib rank above mCherry on this metric (Table S5), and only two of R-Lib rank near their parent protein (R-Lib #2 and R-Lib #14).

The ultimate test of these libraries, however, is in determining the oligomerization state of their members. When a fluorophore is excited with polarized light, the extent to which that polarization is maintained upon emission depends on the amount of time it takes the molecule to tumble in solution (the rotational correlation time). This time is longer for larger molecular complexes up to a certain point [18]. For FPs, the lifetime of the excited state (the fluorescence lifetime) is much shorter than the rotational correlation time. Because of this large difference in lifetime versus correlation time, incoming polarized light used for excitation should remain relatively unchanged after emission. However, in homomolecular complexes of FPs, homomolecular fluorescence resonance energy transfer (homoFRET) can occur [18].

When homoFRET occurs, the polarization of incoming light is changed according to the orientation of the dipoles in the excited states in both molecules. The more members that a homomolecular complex has, the higher the probability that homoFRET can occur. Thus, the anisotropy of incoming polarized light is better retained by monomeric FPs than by oligomeric FPs. This means emitted light from oligomeric FPs has lower anisotropy—it is more depolarized—than that emitted from monomeric FPs [18]. The

degree to which polarization is maintained (the fluorescence anisotropy, r) was measured for all members of M-Lib, R-Lib, mCherry, DsRed, DsRmCh, and our C-terminal truncations of these proteins. For this parameter to be meaningful, the rotational correlation time must be longer than the lifetime of the excited state for an FP; this is true in general for FPs [18]. As shown in Fig. 2, monomeric and oligomeric proteins were clearly distinguishable by fluorescence anisotropy measurements. As expected, r for monomeric mCherry is higher than for our oligomeric proteins, which maintain the same surface residues as DsRed. M-Lib and R-Lib measurements were also consistent with the r value for mCherry, a solution state monomer (Table 2).

Since oligomerization is concentration dependent, the fluorescence anisotropy effect is also concentration dependent [45]. This means that the degree to which fluorescence anisotropy varies with dilution factor (DF) belies information about the strength of homooligomer formation [45], or K_d. In our expression system, we observe the magnitude of the mCherry r-versus-DF slope to be about six times larger than for DsRed (Table 1). This result affirms the idea that mCherry (Table 1), a monomer, forms much weaker homooligomeric interactions in solution, and thus greater dilutions of mCherry (smaller DF) result in higher measurements for fluorescence anisotropy. The average r-versus-DF slope for M-Lib is not as large in magnitude as for mCherry, but it is still nearly five times larger than the slope for DsRed. For R-Lib, only two r-versus-DF slope measurements could be made, and both are larger in magnitude than for DsRed (Table 2 and Table S5).

To further confirm the oligomerization state of M-Lib and R-Lib members, we performed size exclusion chromatography on eight members of M-Lib and four members of R-Lib

(Fig. S2). Library members from M-Lib were chosen because of high Φ_F (M-Lib #1, 2, and 16), high $\lambda_{max}^{abs}/\lambda_{280nm}^{abs}$ (M-Lib #23), high brightness proxy score (M-Lib #13 and 49), or lack of fluorescence (M-Lib #53 and 79). The three most fluorescent members of R-Lib were chosen (R-Lib #2, 14, and 17) along with one other randomly chosen minimally fluorescent member (R-Lib #7). All twelve of these proteins showed elution profiles by size exclusion chromatography consistent with a monomeric oligomerization state in solution (Fig. S2). To bolster evidence for monomerization even further, we performed analytical ultracentrifugation on M-Lib #23, which produced a peak overlapping with mCherry in molecular weight distribution (Fig. S3). These results show that both CPD and random sampling approaches produce monomeric FP sequences. However, the proportion of solubly expressed fluorescent proteins produced is significantly higher in M-Lib versus R-Lib.

Conclusion

A CPD-guided method for generation of monomeric FPs was demonstrated for a DsRed-based experimental system. A higher proportion of sequences resulted in soluble protein expression for the CPD-generated library of FPs than for a library generated by random sampling from a distribution consistent with the known surface residue distribution in mesophilic bacteria [43]. Both CPD and random sampling methods were successful in the production of monomeric FPs. Our current methodology is limited by an incomplete understanding of which constellations of core residues in FPs support chromophore formation in both an oligomeric and a monomeric context. We discovered that the set of core residues in mCherry supports chromophore formation in both contexts, but the core residues of DsRed only support chromophore formation in an oligomeric context. Further

investigation of the factors that enable context-independent chromophore formation will make our methodology extensible to class *Anthozoa* FPs in general.

Materials and Methods

Library Design. Surface mutations for M-Lib were computationally designed using the PHOENIX protein design software [32, 42]. Chain A of the PDB structure 1G7K for DsRed was used for surface designs [38]. The C-terminal residues 221–225 were deleted from this structure and hydrogen atoms were added using Molprobity [46]. Strain or steric clashes were removed by performing 50 steps of conjugate gradient energy minimization prior to computational design calculations. The energy function used was based on the DREIDING force field [41] and included a scaled van der Waals term, hydrogen bonding and electrostatic terms, and terms for implicit solvation and phi-psi propensities. Implicit solvation energies were evaluated using a model based on occluded volume [39, 40], as described previously [32]. Sequence optimization was carried out with FASTER [47].

The following design positions were included in the CPD calculation: AB interface positions 21, 96, 106, 125, 127, 180, and 182; and AC interface positions 145, 149, 153, 162, 164, 174, 176, 192, 194, and 216. In addition to the wild-type residue identity, the following 12 amino acid identities were allowed at each of these 13 design positions: Ala, Arg, Asn, Asp, Glu, Gln, Gly, His, Lys, Ser, Thr, and Tyr. The size of this sequences space is approximately 10^{14}.

For R-Lib, the following 11 amino acids were allowed for sampling: Ala, Arg, Asn, Asp, Glu, Gln, His, Lys, Ser, Thr, and Tyr. A published distribution of surface amino acid

residues for mesophilic bacteria was used as the basis for sampling [43]. This distribution was renormalized to 100% for each surface position in DsRed following removal of residues not listed above and not equivalent to the wild-type residue at each position. Amino acid identities for each of the 13 AB and AC interface positions designed above were chosen for R-Lib using a random number table with sampling based on the modified distributions just described.

Construction of Library Mutants. All protein sequences except C-terminal deletions and point mutations were constructed by gene assembly using a Tecan Freedom Evo robotic liquid handling system equipped with a thermocycler. During design of the gene assembly primers, codons were optimized for *E. coli* expression. Genes were cloned into pET-DEST53 or pET-11a after assembly using the same robotics system. The C-terminal deletions in DsRed and DsRmCh were produced using QuikChange inverse PCR mutagenesis on supercoiled plasmid DNA isolated from *E. coli* BL21-Gold(DE3) cells (Stratagene). All point mutations for mChDsR and DsRed were produced by QuikChange as well. Purification and characterization of mutants are described in *SI text*.

References

1.   Tsien RY (1998) The green fluorescent protein. *Annual Review of Biochemistry* 67:509-544.

2.   Chudakov DM*, et al.* (2010) Fluorescent proteins and their applications in imaging living cells and tissues. *Physiological Reviews* 90(3):1103-1163.

3.   Verkhusha V & Lukyanov K (2004) The molecular properties and applications of Anthozoa fluorescent proteins and chromoproteins. *Nature Biotechnology* 22(3):289-296.

4.   Giepmans BNG*, et al.* (2006) The fluorescent toolbox for assessing protein location and function. *Science* 312:217-224.

5.   Shaner N, Steinbach P, & Tsien R (2005) A guide to choosing fluorescent proteins. *Nature Methods* 2(12):905-909.

6.   Kredel S*, et al.* (2008) Optimized and far-red-emitting variants of fluorescent protein eqFP611. *Chemistry & Biology* 15:224-233.

7.   Campbell RE*, et al.* (2002) A monomeric red fluorescent protein. *Proceedings of the National Academy of Sciences, USA* 99:7877-7882.

8.   Karasawa S*, et al.* (2003) A green-emitting fluorescent protein from Galaxeidae coral and its monomeric version for use in fluorescent labeling. *Journal of Biological Chemistry* 278(36):34167-34171.

9.   Karasawa S*, et al.* (2004) Cyan-emitting and orange-emitting fluorescent proteins as a donor/acceptor pair for fluorescence resonance energy transfer. *Biochemistry Journal* 381(1):307.

10.   Wiedenmann J*, et al.* (2004) EosFP, a fluorescent marker protein with UV-inducible green-to-red fluorescence conversion. *Proceedings of the National Academy of Sciences, USA* 101:15905-15910.

11.   Ando R, Mizuno H, & Miyawaki A (2004) Regulated fast nucleocytoplasmic shuttling observed by reversible protein highlighting. *Science* 306:1370-1373.

12.   Kogure T*, et al.* (2006) A fluorescent variant of a protein from the stony coral Montipora facilitates dual-color single-laser fluorescence cross-correlation spectroscopy. *Nature Biotechnology* 24(5):577-581.

13.   Gurskaya NG*, et al.* (2006) Engineering of a monomeric green-to-red photoactivatable fluorescent protein induced by blue light. *Nature Biotechnology* 24:461-465.

14. Ai H, *et al.* (2006) Directed evolution of a monomeric, bright and photostable version of Clavularia cyan fluorescent protein: structural characterization and applications in fluorescence imaging. *Biochemistry Journal* 400(3):531-540.

15. Merzlyak E, *et al.* (2007) Bright monomeric red fluorescent protein with an extended fluorescence lifetime. *Nature Methods* 4(7):555-557.

16. Strongin DE, *et al.* (2007) Structural rearrangements near the chromophore influence the maturation speed and brightness of DsRed variants. *Protein Engineering Design and Selection* 20(11):525-534.

17. Lin MZ, *et al.* (2009) Autofluorescent proteins with excitation in the optical window for intravital imaging in mammals. *Chemistry & Biology* 16(11):1169-1179.

18. Yan YL & Marriott G (2003) Analysis of protein interactions using fluorescence technologies. *Current Opinion in Chemical Biology* 7:635-640.

19. Nienhaus K, *et al.* (2008) Trans-cis isomerization is responsible for the red-shifted fluorescence in variants of the red fluorescent protein eqFP611. *Journal of the American Chemical Society* 130(38):12578-12579.

20. Nienhaus K, *et al.* (2006) Chromophore-protein interactions in the anthozoan green fluorescent protein asFP499. *Biophysical Journal* 91:4210-4220.

21. Nienhaus K, *et al.* (2006) Exploring chromophore-protein interactions in fluorescent protein cmFP512 from Cerianthus membranaceus: X-ray structure analysis and optical spectroscopy. *Biochemistry* 45:12942-12953.

22. Wilmann PG, *et al.* (2005) The 2.1 Å crystal structure of the far-red fluorescent protein HcRed: Inherent conformational flexibility of the chromophore. *Journal of Molecular Biology* 349:223-237.

23. Chan MC, *et al.* (2006) Structural characterization of a blue chromoprotein and its yellow mutant from the sea anemone Cnidopus japonicus. *Journal of Biological Chemistry* 281(49):37813-37819.

24. Wall MA, Socolich M, & Ranganathan R (2000) The structural basis for red fluorescence in the tetrameric GFP homolog DsRed. *Nature Structural Biology* 7(12):1133-1138.

25. Maio GD, *et al.* (2008) Crystal structure and Raman studies of dsFP483, a cyan fluorescent protein from Discosoma striata. *Journal of Molecular Biology* 378:871-886.

26. Mizuno H, *et al.* (2008) Light-dependent regulation of structural flexibility in a photochromic fluorescent protein. *Proceedings of the National Academy of Sciences, USA* 105:9227-9232.

27. Baird GS, Zacharias DA, & Tsien RY (2000) Biochemistry, mutagenesis, and oligomerization of DsRed, a red fluorescent protein from coral. *Proceedings of the National Academy of Sciences, USA* 97:11984-11989.

28. Shaner N, *et al.* (2004) Improved monomeric red, orange and yellow fluorescent proteins derived from Discosoma sp. red fluorescent protein. *Nature Biotechnology* 22(12):1567-1572.

29. Shu X, *et al.* (2006) Novel chromophores and buried charges control color in mFruits. *Biochemistry* 45:9639-9647.

30. Strack RL, *et al.* (2010) Chromophore formation in DsRed occurs by a branched pathway. *Journal of the American Chemical Society* 132(24):8496-8505.

31. Zhang LP, *et al.* (2006) Reaction progress of chromophore biogenesis in green fluorescent protein. *Journal of the American Chemical Society* 128:4766-4772.

32. Chica RA, *et al.* (2010) Generation of longer emission wavelength red fluorescent proteins using computationally designed libraries. *Proceedings of the National Academy of Sciences, USA* 107(47):20257-20262.

33. Subach FV, Piatkevich KD, & Verkhusha VV (2011) Directed molecular evolution to design advanced red fluorescent proteins. *Nature Methods* 8(12):1019-1026.

34. Wachter RM, *et al.* (1998) Structural basis of spectral shifts in the yellow-emission variants of green fluorescent protein. *Structure With Folding & Design* 6:1267-1277.

35. Jones S & Thornton JM (1995) Protein-protein interactions: A review of protein dimer structures. *Progress in Biophysics and Molecular Biology* 63:31-65.

36. Huang P, Love J, & Mayo S (2005) Adaptation of a fast Fourier transform-based docking algorithm for protein design. *Journal of Computational Chemistry* 26(12):1222-1232.

37. Huang PS, Love JJ, & Mayo SL (2007) A de novo designed protein protein interface. *Protein Science* 16(12):2770-2774.

38. Yarbrough D, *et al.* (2001) Refined crystal structure of DsRed, a red fluorescent protein from coral, at 2.0 Å resolution. *Proceedings of the National Academy of Sciences, USA* 98:462-467.

39. Lazaridis T & Karplus M (1999) Effective energy function for proteins in solution. *Proteins* 35:133-152.

40. Lazaridis T & Karplus M (1999) Discrimination of the native from misfolded protein models with an energy function including implicit solvation. *Journal of Molecular Biology* 288:477-487.

41. Mayo S, Olafson B, & Goddard W (1990) Dreiding - A generic force-field for molecular simulations. *Journal of Physical Chemistry-US* 94(26):8897-8909.

42. Privett HK, *et al.* (2012) Iterative approach to computational enzyme design. *Proceedings of the National Academy of Sciences, USA* 109(10):3790-3795.

43. Fukuchi S & Nishikawa K (2001) Protein surface amino acid compositions distinctively differ between thermophilic and mesophilic bacteria. *Journal of Molecular Biology* 309(4):835-843.

44. Treynor TP, *et al.* (2007) Computationally designed libraries of fluorescent proteins evaluated by preservation and diversity of function. *Proceedings of the National Academy of Sciences, USA* 104:48-53.

45. Letilly V & Royer C (1993) Fluorescence anisotropy assays implicate protein-protein interactions in regulating Trp repressor DNA-binding. *Biochemistry* 32(30):7753-7758.

46. Davis IW, *et al.* (2007) MolProbity: All-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Research* 35:W375-W383.

47. Allen BD & Mayo SL (2006) Dramatic performance enhancements for the FASTER optimization algorithm. *Journal of Computational Chemistry* 27:1071-1075.

Figure & Legends
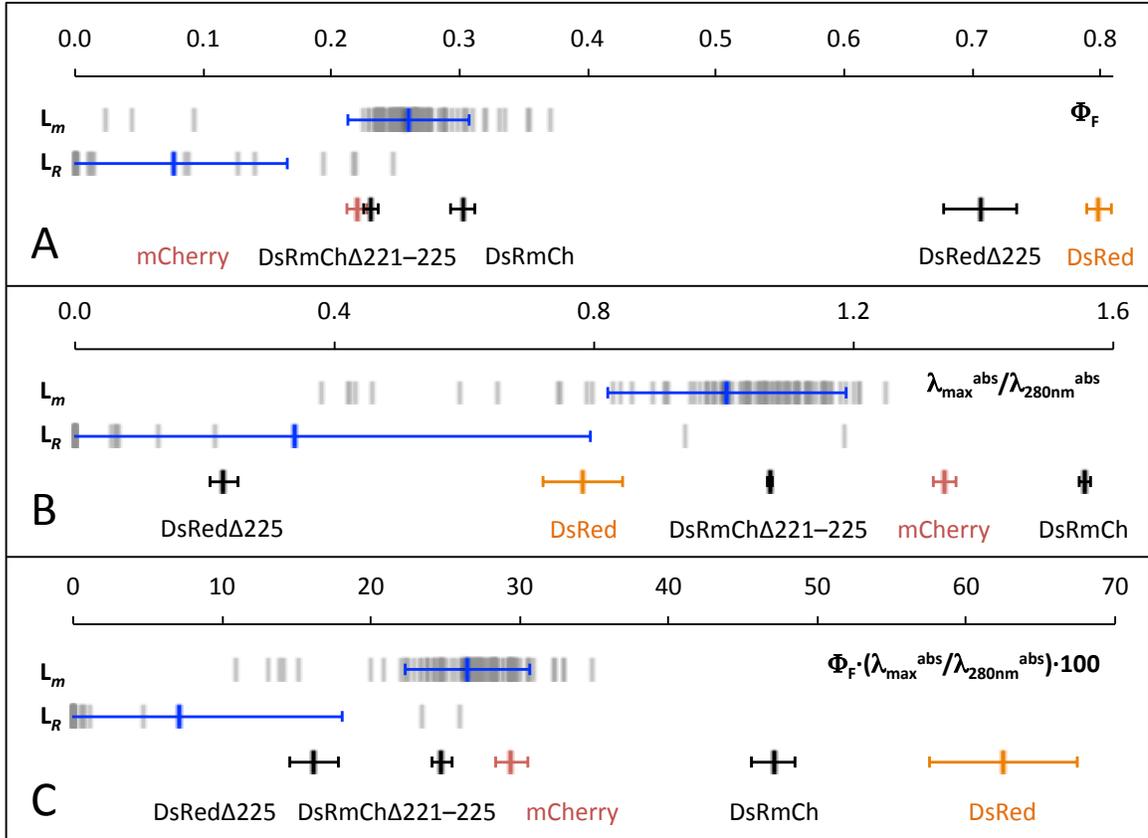
Fig. 1. M-Lib (L_m) and R-Lib (L_R) measurements and scoring metric distributions. Error bars represent $\pm 1\sigma$ for all panels. Each vertical gray tick represents the measurement of one mutant in either M-Lib or R-Lib; overlapping gray ticks appear darker. Mutant protein labels are ordered according to measurement value; mCherry (*red*) and DsRed (*orange*) are color-coded. A blue vertical tick marks the average value for all measurements made along a metric in M-Lib or R-Lib. The metrics shown by panel correspond to: (*A*) quantum yield (Φ_F), (*B*) λ_{max}^{abs} to λ_{280nm}^{abs} peak-height ratio ($\lambda_{max}^{abs}/\lambda_{280nm}^{abs}$), and (*C*) brightness proxy score ($\Phi_F \cdot [\lambda_{max}^{abs}/\lambda_{280nm}^{abs}] \cdot 100$).
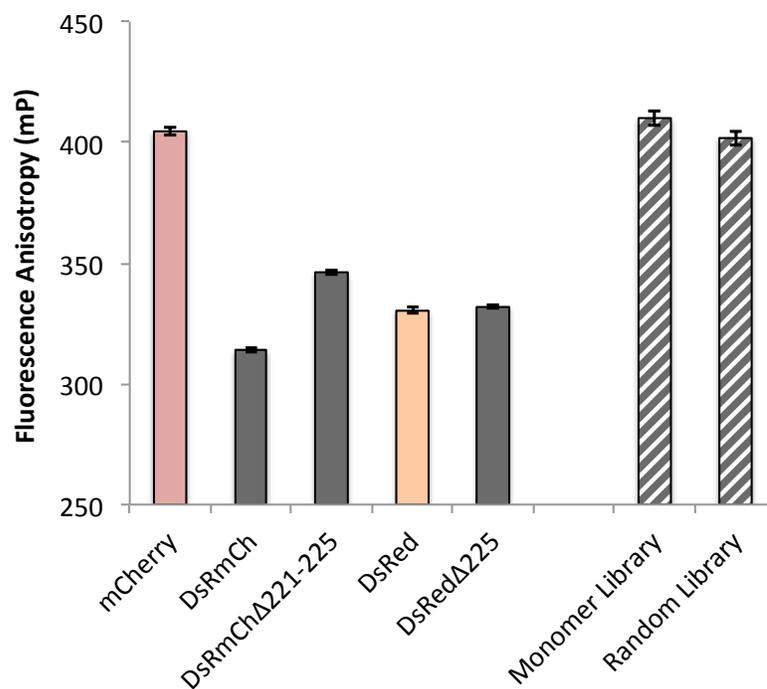
Fig. 2. Fluorescence anisotropy measurements clearly differentiate monomeric FPs (mCherry, M-Lib, and R-Lib) from oligomeric FPs (DsRmCh, DsRed, and their truncations). Fluorescence anisotropy is lower for oligomers, which undergo more homo-FRET. M-Lib and R-Lib values are the average of measurements made on all library members. Other FP values are the average of six measurements. Error bars represent ±1σ from these averages.

Table 1. Spectral Properties and Oligomeric State Characterization

Protein	λ_{max}^{abs}	λ_{max}^{ems}	Φ_F	$\lambda_{max}^{abs}/\lambda_{280nm}^{abs}$	Brightness Proxy*	r, mP†	r vs. DF (slope)†	Elution Volume, CV‡	Mol. Wgt, kDa¶ (calc.)	Mol. Wgt, kDa¶ (via AUC)	ε_{280nm}§ mM^{-1}·cm^{-1}
mCherry	586	613	0.22	1.34	29.5	405	−11.2	0.508	27.7	28.0	31.1
mChDsR	---	---	---	---	---	---	---	0.488	27.8	---	31.1
DsRmCh	586	613	0.30	1.56	47.1	314	---	0.419	108.0	105.6	32.3
DsRmChΔ221-225	586	613	0.23	1.07	24.8	346	---	0.429	105.4	103.2	32.3
DsRed	560	585	0.80	0.78	62.5	330	−1.8	0.418	108.6	101.4	32.3
DsRedΔ225	560	585	0.71	0.23	16.2	332	---	0.423	108.1	---	32.3
Monomer Library◊	586	610	0.26	1.01	26.1	410	−8.8	0.538	26.5	29.4	29.9
Random Library◊	582	610	0.08	0.34	7.2	402	---	0.533	26.3	---	30.9

* The "Brightness Proxy" score is defined as follows: $100 \cdot (\lambda_{max}^{abs}/\lambda_{280nm}^{abs}) \cdot \Phi_F$

† Fluorescence anisotropy (r) was measured in milli-polarization units (mP). DF = Dilution Factor. The slope of r versus DF is a proxy for the dissociation constant (K_d) of an homo-oligomeric FP; a more negative slope indicates a higher K_d.

‡ CV = Column Volume, 1.0 CV = 23.562 mL for the Superdex 75 10/300 GL size exclusion column used.

¶ AUC = Analytical UltraCentrifugation. Molecular weights are presented without chromophore modification, which corresponds to an additional mass loss of 22 Da. DsRmCh, DsRed, and their variants contain the 6xHis tag: MGHHHHHHGSG (1229.9 Da); mCherry and mChDsR use the tag: MGHHHHHHGVSKG (1457.2 Da). Molecular weights of proteins with a DsRed surface are presented for the corresponding homo-tetramer.

§ Extinction coefficients at 280 nm were calculated based on the peptide sequence for one monomer of each protein; structural considerations were ignored.

◊ Molecular weight and ε_{280nm} for the Monomer Library were determined from the consensus sequence for this library, which is also the sequence for M-Lib #1. For the Random Library, these values were calculated as an average of all the sequences in the library ($\sigma_{MolWt} \pm 0.1$ kDa, $\sigma_{280nm} \pm 1.3$ mM^{-1}·cm^{-1}). AUC data for the Monomer Library was only collected on one mutant, M-Lib #23.

The elution volume presented for the Monomer Library is an average of data collected on eight mutants: M-Lib #1, #2, #13, #16, #23, #49, #53, & #79. Likewise, for the Random Library, elution volume is an average of data collected on four mutants: R-Lib #2, #7, #14, & #17. All other values presented for the Monomer Library and the Random Library are averages of the available data collected on all library mutants. See *Table S4* and *Table S5* for measurement values corresponding to each individual mutant in both libraries.

Table 2. Statistics for Monomer Library and Random Library

Monomer Library

	λ_{max}^{abs}	λ_{max}^{ems}	Φ_F	$\lambda_{max}^{abs}/\lambda_{280nm}^{abs}$	Brightness Proxy*	r, mP[†]	r vs. DF (slope)[†]
Avg. ± 1s	586 ± 1	610 ± 1	0.26 ± 0.05	1.00 ± 0.18	26.5 ± 4.2	410 ± 3	−8.8 ± 1.5
Max	586	611	0.37	1.25	34.8	419	−5.5
Min	584	606	0.02	0.38	11.0	404	−12.7
Skewness	−2.6	−1.6	−2.4	−1.9	−1.6	0.4	−0.03
Kurtosis[‡]	4.6	2.9	11.5	3.6	3.6	0.4	−0.4
N	92	94	95	92	92	92	79

Random Library

	λ_{max}^{abs}	λ_{max}^{ems}	Φ_F	$\lambda_{max}^{abs}/\lambda_{280nm}^{abs}$	Brightness Proxy*	r, mP[†]	r vs. DF (slope)[†]
Avg. ± 1s	582 ± 12	610 ± 2	0.08 ± 0.09	0.34 ± 0.45	7.2 ± 10.9	402 ± 3	---
Max	590	613	0.25	1.19	25.9	404	−3.7
Min	552	605	0.00	0.06	0.2	399	−6.7
Skewness	−2.5	−0.6	0.8	1.5	1.4	−1.1	---
Kurtosis[‡]	6.8	1.1	−0.8	0.5	−0.003	---	---
N	8	17	19	8	8	3	2

* The "Brightness Proxy" score is defined as follows: $100 \cdot (\lambda_{max}^{abs}/\lambda_{280nm}^{abs}) \cdot \Phi_F$

[†] Fluorescence anisotropy (r) was measured in milli-polarization units (mP).
DF = Dilution Factor. The slope of r versus DF is a proxy for the dissociation constant (K_d) of an homo-oligomeric FP; a more negative slope indicates a higher K_d.

[‡] Kurtosis as presented here is "excess kurtosis", meaning that a normal distribution produces a value of zero.

Supporting Information

SI Materials and Methods

Materials. All reagents used were of the highest available purity. Synthetic oligonucleotides were obtained from Integrated DNA Technologies, and Ni-NTA agarose resin was obtained from Qiagen. CelLytic B buffer and lysozyme were purchased from Sigma-Aldrich. All aqueous solutions were prepared using water purified with a Millipore BioCell system.

Protein expression and purification. The DNA libraries prepared as above were transformed into chemically competent *E. coli* BL21-Gold(DE3) cells (Stratagene). Colonies were picked into individual wells of Nunc V96 MicroWell polypropylene plates containing 200 µL of medium (LB with 100 µg/mL ampicillin supplemented with 10% glycerol). The plates were covered with a sterile Breathe-Easy gas permeable sealing membrane (Sigma) and incubated overnight at 37 °C with shaking. These mother plates containing the mutant libraries were used to inoculate 24-well culture plates (Whatman) containing 5 mL LB supplemented with ampicillin in each well.

The 24-well plates were sealed with breathable membranes and incubated for 2–3 hrs at 37 °C. After incubation, wells were induced with 1 mM IPTG each, then grown overnight with shaking (25 °C, 250 rpm). After overnight protein expression, cells were harvested by centrifugation. Cell pellets were then incubated at 4 °C for 72 h to allow chromophore maturation. After maturation, cell pellets were resuspended in 400 µL lysis buffer

(50 mM Tris-HCl buffer, pH 8.0, 300 mM NaCl, 2.5 mM imidazole, 1X CelLytic B, 1 mg/mL lysozyme, and 25 U/mL benzonase nuclease (Novagen)) and incubated at room temperature for 1.0 hr. After centrifugation, clarified lysates were recovered and proteins were purified by affinity chromatography using His-Select plates (Sigma) according to the manufacturer's protocol.

Spectroscopic characterization. All absorbance and emission spectra were recorded with a Tecan Safire2 plate reader in Greiner UV-Star clear cyclic olefin copolymer 96-well flat bottom plates. For determination of quantum yields, the integrated fluorescence intensity of mutants of interest was compared with that of equally absorbing samples of mCherry and DsRed (quantum yields 0.22 [1, 2] and 0.80 [3], respectively) with excitation at 550 nm. DsRed was used as an internal standard to check the effectiveness of this measurement procedure.

Size exclusion chromatography. Purified protein samples were eluted on an ÄKTA Purifier FPLC system by gel filtration using a Superdex 75 10/300 GL Tricorn resin column (GE Healthcare) into a final buffer solution of 50 mM phosphate buffer, pH 7.0, and 150 mM NaCl.

Analytical ultracentrifugation. A Beckman XL-1 ultracentrifuge with 4-hole rotor was used for sedimentation velocity measurements and molecular weight determinations. Absorbance scans were taken at 575 nm while spinning at 48,000 rpm overnight. Data was analyzed with SedFit, using C(S) and C(M) distributions.

Fluorescence anisotropy. Rose Bengal (Sigma Aldrich) was used as a standard for measuring an instrument-specific g-factor for subsequent polarization measurements. A

g-factor of 1.115 was determined by calibrating a Tecan Safire2 plate reader with 100μM Rose Bengal, which has a standard fluorescence anisotropy of 349 mP at 610 nm. For calibration, excitation was performed with a 530 nm laser line. Polarization measurements were made for all proteins using excitation at 530 nm and detection at 610 nm.

Mass spectrometry analyses. Following separation by SDS-PAGE, the ~25 kDa band from a freshly purified protein sample was excised and destained. Destaining of the Coomassie dye was accomplished by a 100 μL wash of 50 mM ammonium bicarbonate followed by a 50 μL wash of a 1:1 mixture of 50 mM ammonium bicarbonate and acetonitrile; this process was repeated for a total of three times. After destaining, the gel band sample was reduced with 25 μL of 50 mM ammonium bicarbonate plus 50 μL of freshly prepared 10 mM DTT in 100 mM ammonium bicarbonate for 30 minutes at 50°C. The sample was then alkylated in the absence of light with 25 μL of 50 mM ammonium bicarbonate plus 50 μL of freshly prepared 55 mM iodoacetate in 100 mM ammonium bicarbonate for 20 minutes at room temperature. Following additional washes with 100 μL of 50 mM ammonium bicarbonate and 100 μL of acetonitrile, the gel band sample was digested overnight at 37°C with 75 μL of 50 mM ammonium bicarbonate plus 25 μL of 6 ng/μL sequencing grade porcine trypsin (Promega).

After digestion, the supernatant from the gel band sample was collected. The gel band was washed three times: once with 100 μL of 1% formic acid/2% acetonitrile in water, once with 100 μL of a 1:1 acetonitrile and water mixture, and once with 100 μL of 1% formic acid in acetonitrile. The pooled supernatant and wash solutions were then

vacuum-dried overnight and resuspended in 0.1% formic acid in preparation for collection of mass spectrometry data. Samples of this nature were prepared in triplicate from the same freshly expressed and purified sample of protein.

These tryptic-digest samples were desalted on a 150 μm × 3 cm C18AQ pre-column (Magic 5 μm, Michrom). After desalting, separation of peptides was performed with a CapLC XE HPLC system (Waters) using a 5 to 35% acetonitrile gradient in 0.2% formic acid on a 100 μm × 15 cm column packed with the same resin as the pre-column. The flow rate during separation was 0.35 μL/min and the HPLC column was connected directly to the mass spectrometer used for MS/MS analysis. Tandem mass spectra were acquired in data-dependent acquisition mode on a hybrid LTQ FT-ICR Ultra mass spectrometer (Thermo Fisher Scientific) with a nanoelectrospray ion source. Full scan mass spectra (400-1600 m/z) were acquired after accumulating 500,000 ions (with a resolution of 50,000 at 400 m/z). The seven most intense ions from the full scans were trapped in the linear ion trap and fragmented by CID after accumulating 5,000 ions (collisional energy: 35%, isolation width: 3 Da). Ion charge state screening was employed for singly and multiply charged ions. A dynamic exclusion list was set (maximum retention time: 60 s, relative mass window: 10 ppm) and early expiration was permitted.

Raw files were converted to MGF files using ReAdW4Mascot2 (ver. 20090305a). The MGF files were searched using Mascot (v. 2.2.06) against the target proteins and a set of common contaminant proteins (297 sequences). Precursor error tolerance was set at 10 ppm and fragment ion tolerances was set at 0.5 Da. Trypsin was specified as the digestion enzyme and up to two missed cleavages were allowed. The following variable modifications were specified: oxidation of methionine (+15.99491), protein N-terminal

acetylation (+42.01056), and tyrosine modifications (–22.041865, –20.026215, –2.01565, –4.031300). Carboxamidomethylation of cysteine (+57.02146) was specified as a fixed modification. Scaffold (version Scaffold_3_00_06, Proteome Software Inc., Portland, OR) was used to validate MS/MS based peptide and protein identifications. Peptide identifications were accepted if they could be established at greater than 90.0% probability as specified by the Peptide Prophet algorithm [4]. Protein identifications were accepted if they could be established at greater than 90.0% probability and contained at least one identified peptide. Protein probabilities were assigned by the Protein Prophet algorithm [5]. Proteins that contained similar peptides and could not be differentiated based on MS/MS analysis alone were grouped to satisfy the principles of parsimony.

References

1.   Kredel S*, et al.* (2008) Optimized and far-red-emitting variants of fluorescent protein eqFP611. *Chemistry & Biology* 15:224-233.

2.   Shaner N*, et al.* (2004) Improved monomeric red, orange and yellow fluorescent proteins derived from Discosoma sp. red fluorescent protein. *Nature Biotechnology* 22(12):1567-1572.

3.   Campbell RE*, et al.* (2002) A monomeric red fluorescent protein. *Proceedings of the National Academy of Sciences, USA* 99:7877-7882.

4.   Keller A*, et al.* (2002) Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Analytical Chemistry* 74(20):5383-5392.

5.   Nesvizhskii A*, et al.* (2003) A statistical model for identifying proteins by tandem mass spectrometry. *Analytical Chemistry* 75(17):4646-4658.

SI Figures & Legends
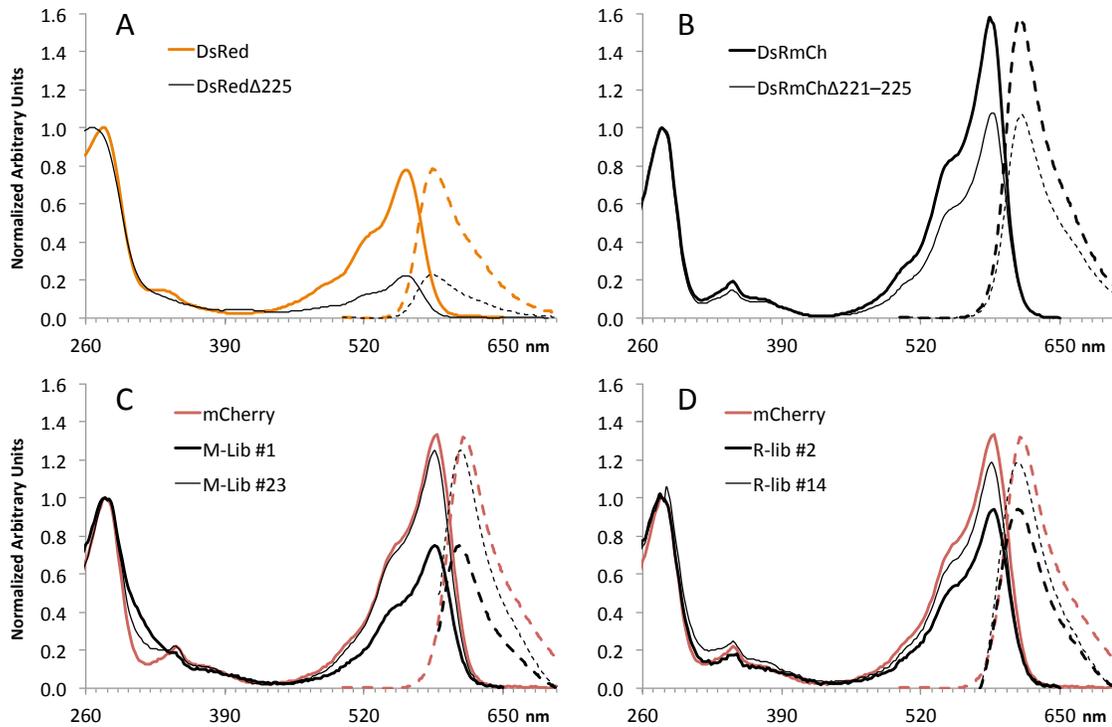
Fig. S1. Absorbance (*solid*) and emission (*dashed*) traces for select proteins. All absorbance spectra are normalized to the 280 nm protein peak. Proteins were excited at 370 nm to collect emission spectra. Emission spectra are normalized to the chromophore absorbance peak. mCherry (*red*) and DsRed (*orange*) are color-coded.
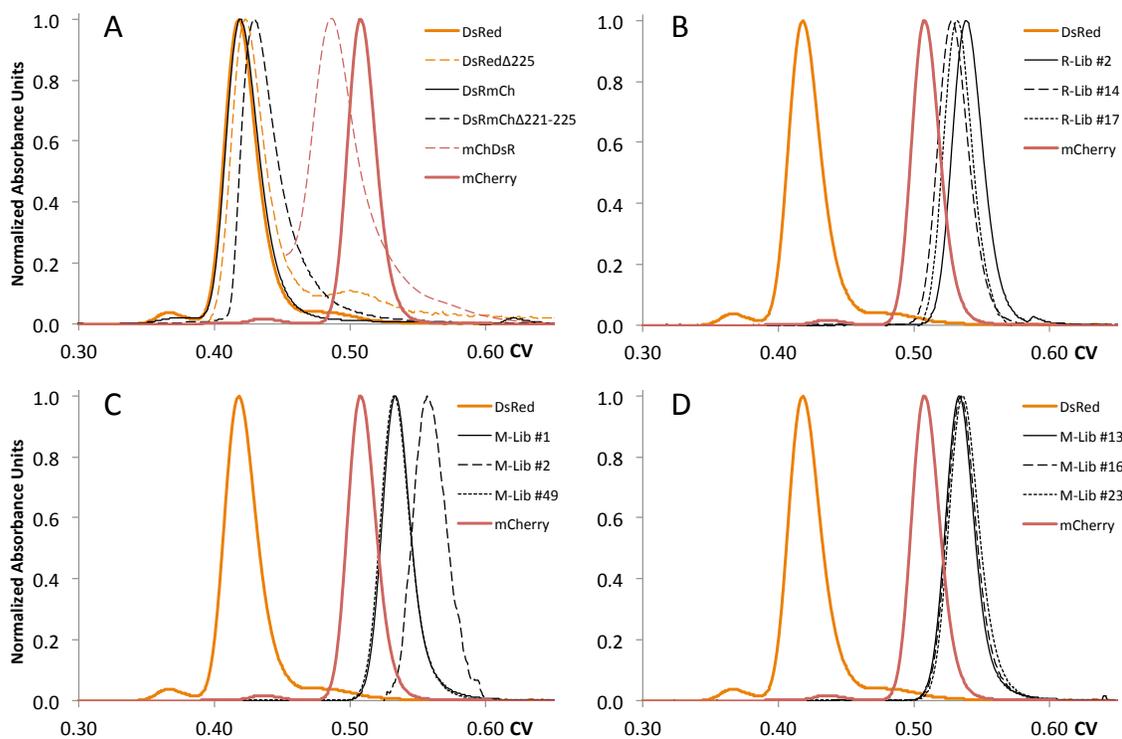
Fig. S2. Size exclusion chromatography traces for select proteins. In all panels, mCherry (*red*) and DsRed (*orange*) are color-coded. (*A*) Data from select proteins in Table 1 suggesting that C-terminal truncation minimally affects oligomerization state. (*B*) Samples from R-Lib elute after mCherry, indicating they are monomers in solution. (*C* & *D*) Samples from M-Lib also elute after mCherry.
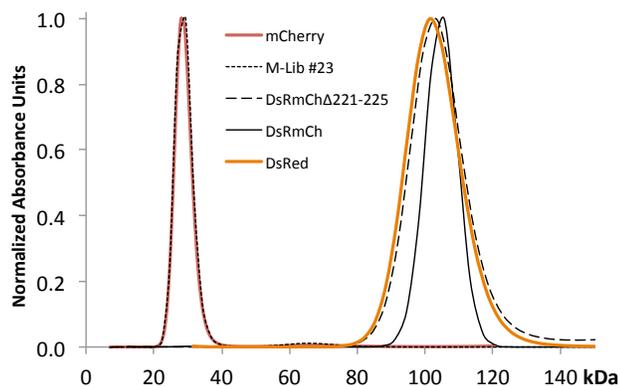
Fig. S3. Analytical ultracentrifuation data for select proteins. Molecular weight distributions for mCherry and M-Lib #23 indicate these proteins are solution state monomers; distributions for DsRed, DsRmCh, and DsRmChΔ221–225 indicate oliogmerization.
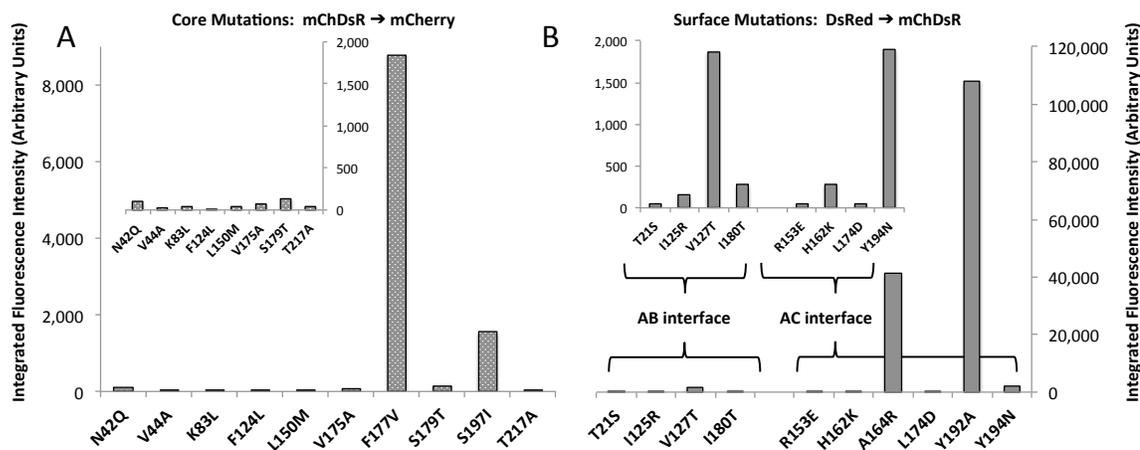
Fig. S4. Point mutations to and from the mChDsR sequence. Integrated red fluorescence intensity is plotted for select mChDsR-to-mCherry core mutations (*A*) and select DsRed-to-mChDsR surface mutations (*B*). Data for mutants with the least intensity is re-plotted on a smaller-scaled axis (*insets*).

A

AB interface

Residue	21	<u>94</u>	96	104	106	108	123	125	127	<u>128</u>	131	<u>154</u>	158	180	182	Mutations per protein
mRFP1	T→S	E	V	V	T	T	K	I→R	V→T	N	S	D	K	I→T	M	4
DsRed.M1	T→S	E	V→S	V	T→E	T→Q	K	I→K	V	N	S→A	D	K	I→V	M→K	8
mTFP1	N	E	T	I	K	K	R→H	D→K	M→E	N	P	D	V→K	I	K→R	5
mAG1	T	E	S	I	T	T	R	D	V→T	N	P	D	K	T	K	1
mKO1	S	E	S	F→S	A→S	S	K	V→T	V	N	A	D	K	T	K	3
mEosFP	N	E	S	I	I	R	R	H	V→T	N	A	D	T	T	K	1
mKeima	T	E	I→S	V	T	S	K	S	V→E	N	P	D	I	T	K	2
TagRFP	T	E	I→V	T	T	T	K	N→R	V	N	S	D	R→E	T	R	3
Dendra	N	E	T	I	T	R	R	N→K	M→T	N	P	D	E	T	K	2
Dronpa	A	E	S	I	I→N	T	R	R	D	N	A	D	K	T	K	1
Mutations per site	2	0	3	1	3	1	1	6	6	0	1	0	1	2	2	

Color-coding for mutational frequency per site

(red)	6
	5
	4
(green)	3
(teal)	2
(blue)	1
(purple)	0

B

C

chain A

D

chain B

E

AB interface

- 138 -
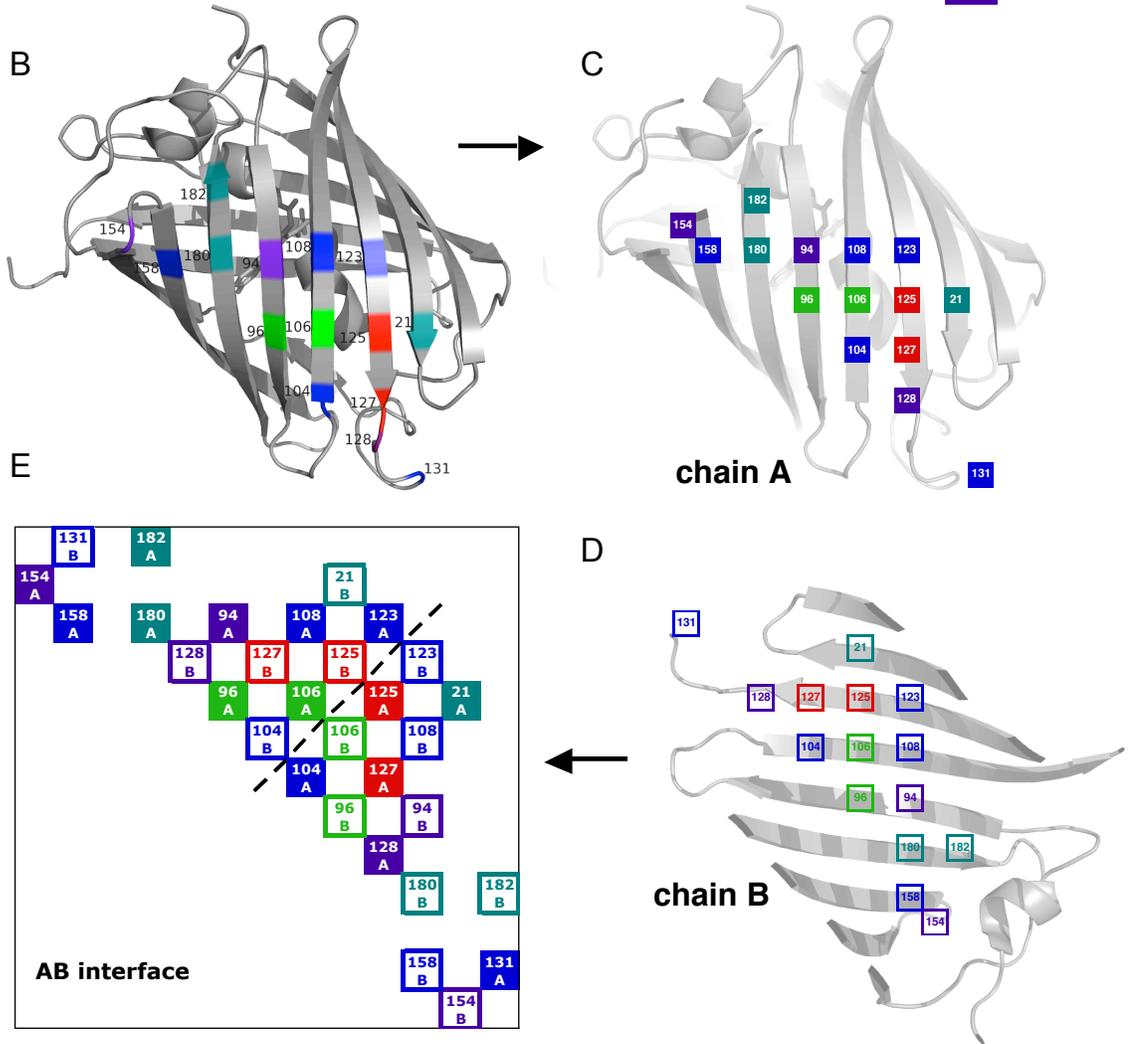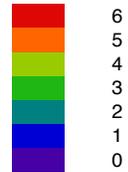
Fig. S5. Generating schematic representations of the dimeric interfaces. (*A*) Literature examples of monomerized proteins; the wild-type residue is indicated on the left wherever mutations have occurred. Underlining indicates conserved positions. (*B*) Color-coded representation of mutational frequency in the AB interface. (*C & D*) Translation of the interface into grid form for (*C*) chain A and (*D*) chain B. (*E*) Schematic representation of the AB interface; a model of the interface as viewed top-down through chain B is also shown (*inset*). Note that this schematic is symmetrical about the diagonal axis (dashed line).

A

AC interface

Residue	100	145	147	149	151	153	160	162	164	172	174	176	192	194	196	198	216	218	220	222	223	224	225	Mutations per protein
mRFP1	E	A	T	R	Y	R→E	E	H→K	A→R	H	L→D	E	Y→A	Y→A	D	K	R	E	R	H→S	L→T	F→G	L→A	10
DsRed.M1	E	A→P	T	R→K	Y	R→Q	E	H→S	A	H	L→T	E→D	Y→N	Y	F	D	K	R→H	E	H→S	L→G	F→S	L→Q	12
mTFP1	E	P→A	T	I→R	Y	R	D	S→K	S→K	H	R	D	Y	F	D	R	N→S	V	R	S	L→T	L→D	P→G	8
mAG1	E	P	T	K	Y	R	D	N	A	H	R	D	Y→A	F→K	D	R	N	V	R	S	M	L	P	2
mKO1	E	P	T	K	T	Y→S	D	T	F→Y	N	K	Q	S	F→Y	G→S	R	D	V	H	–	–	–	–	4
mEosFP	E	P	T	K	Y	R	D	T→H	A	H	R	D	Y	F	D	C	H	V	H	S	G	L	P	1
mKeima	E	P	T	R	F	R	N	F→Y	A	H	L	E	Y→R	Y→E	D	K	I	I	R	S	L	L	G	3
TagRFP	E	A	T	M	Y	A	H→R	Q→D	A	Y→H	H→I	S→N	F→V	F→Y	D	R	M→V	V	K→R	C	D	L	P	9
Dendra	E	P	T	K	H	R	N	N	A	H	L	D	Y→A	F	D	R	H	V	R	S	P	L	P	1
Dronpa	E	P	T	K	Y	R	D	N	A	H	R	D	Y	F	D	R→H	H	E	H	S	G→E	L	P	2
Mutations per site	0	2	0	2	0	3	1	6	3	1	3	2	6	5	1	1	3	0	1	2	4	3	3	

B

Color-coding for mutational frequency per site

6
5
4
3
2
1
0

C
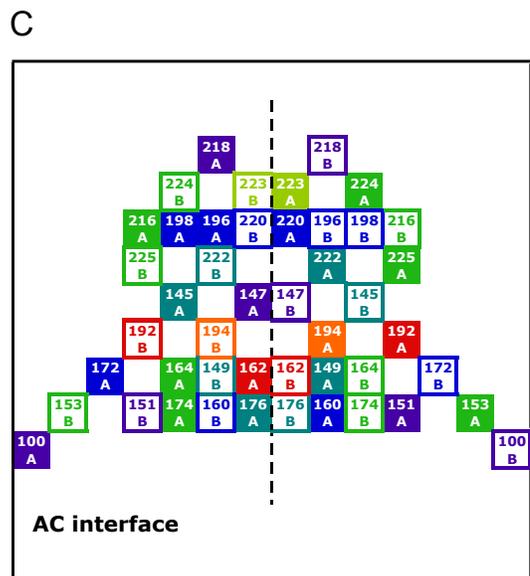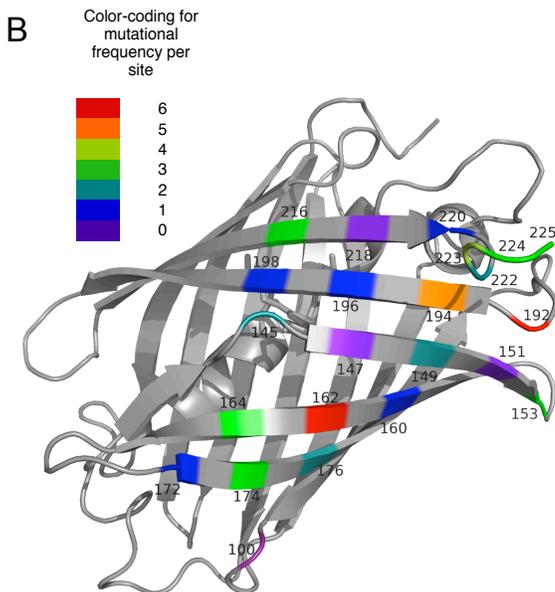
AC interface

Fig. S6. Schematic representation of the AC interface. (*A*) Literature examples of monomerized proteins; the wild-type residue is indicated on the left wherever mutations have occurred. (*B*) Color-coded representation of mutational frequency in the AC interface. (*C*) The resulting schematic representation of the AC interface from application of the procedure shown in Figure 3. Note that b-sheets from opposing subunits of the interface are stacked in an anti-parallel fashion instead of offset by ~90° (as in the AB interface, Fig. S5). Thus, this schematic diagram is symmetrical about the vertical, not the diagonal.
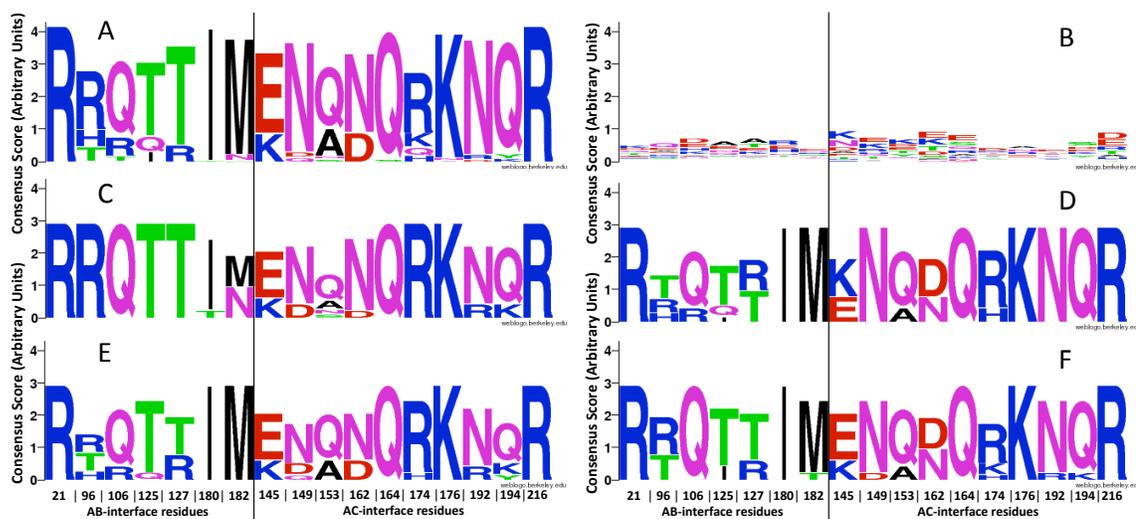
Fig. S7. Protein sequence consensus diagrams for: (*A*) the entire monomer design library (M-Lib), (*B*) the entire random mutation library (R-Lib), (*C*) the top ten ranking sequences in terms of quantum yield (Φ_F), (*D*) the top ten ranking sequences in terms of λ_{max}^{abs} to λ_{280nm}^{abs} peak-height ratio, (*E*) the top ten ranking sequences as scored by the brightness proxy metric ($\Phi_F \cdot [\lambda_{max}^{abs}/\lambda_{280nm}^{abs}] \cdot 100$), and (*F*) the top ten ranked sequences as scored by dilution-dependent fluorescence anisotropy slope (*r* vs. DF slope).

Table S1. Sequences of Selected DsRed, mCherry, and CPD-Generated Mutants

Protein	AB Interface Residues							AC Interface Residues									
	21	96	106	125	127	180	182	145	149	153	162	164	174	176	192	194	216
mCherry	S	V	T	R	T	T	K	A	R	E	K	R	D	E	A	N	R
mChDsR	S	V	T	R	T	T	K	A	R	E	K	R	D	E	A	N	R
DsRmCh	T	V	T	I	V	I	M	A	R	R	H	A	L	E	Y	Y	R
DsRmChΔ221-225	T	V	T	I	V	I	M	A	R	R	H	A	L	E	Y	Y	R
DsRed	T	V	T	I	V	I	M	A	R	R	H	A	L	E	Y	Y	R
DsRedΔ225	T	V	T	I	V	I	M	A	R	R	H	A	L	E	Y	Y	R
Monomer Library*	R	R	Q	T	T	I	M	E	N	Q	N	Q	R	K	N	Q	R

Protein	Core Residues													C-Terminal Tail Residues				
	42	44	66	71	83	124	150	163	175	177	179	197	217	221	222	223	224	225
mCherry	Q	A	M	A	L	L	M	Q	A	V	T	I	A	H	S	T	G	G
mChDsR	N	V	Q	V	K	F	L	K	V	F	S	S	T	H	S	T	G	G
DsRmCh	Q	A	M	A	L	L	M	Q	A	V	T	I	A	H	H	L	F	L
DsRmChΔ221-225	Q	A	M	A	L	L	M	Q	A	V	T	I	A	–	–	–	–	–
DsRed	N	V	Q	V	K	F	L	K	V	F	S	S	T	H	H	L	F	L
DsRedΔ225	N	V	Q	V	K	F	L	K	V	F	S	S	T	H	H	L	F	–
Monomer Library*	Q	A	M	A	L	L	M	Q	A	V	T	I	A	–	–	–	–	–

* Residues shown for the monomer library are a consensus sequence for the CPD design results, which is also the sequence for M-Lib #1.
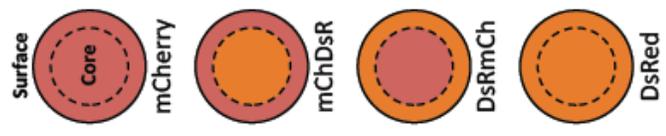
Table S2. Interface Residues Predicted from CPD Interface Designs of DsRed (Monomer Library) (1 of 4)

Protein	AB Interface Positions							AC Interface Positions										Score
	21	96	106	125	127	180	182	145	149	153	162	164	174	176	192	194	216	
M-Lib #1	R	R	Q	T	T	I	M	E	N	Q	N	Q	R	K	N	Q	R	−182.264
M-Lib #2	R	R	Q	T	T	I	N	E	N	Q	N	Q	R	K	N	Q	R	−181.267
M-Lib #3	R	R	Q	T	T	I	M	E	N	Q	N	Q	K	K	N	Q	R	−181.239
M-Lib #4	R	R	Q	T	T	I	M	E	N	A	N	Q	R	K	N	Q	R	−180.756
M-Lib #5	R	H	R	T	T	I	M	E	N	Q	N	Q	R	K	N	Q	R	−180.742
M-Lib #6	R	R	Q	T	T	I	M	E	N	Q	D	Q	R	K	N	Q	R	−180.598
M-Lib #7	R	R	Q	T	T	I	M	E	N	Q	N	Q	R	K	N	Q	R	−180.590
M-Lib #8	R	R	Q	T	T	I	M	K	N	Q	N	Q	R	K	N	Q	R	−180.584
M-Lib #9	R	R	Q	Q	T	I	M	E	N	Q	N	Q	R	K	N	Q	R	−180.540
M-Lib #10	R	T	Q	T	R	I	M	E	N	Q	N	Q	R	K	N	Q	R	−180.429
M-Lib #11	R	R	Q	T	T	I	M	E	N	Q	N	Q	R	K	D	Q	R	−180.102
M-Lib #12	R	R	Q	T	T	I	M	E	N	Q	N	Q	Q	K	N	Q	R	−180.047
M-Lib #13	R	R	Q	T	T	I	M	E	D	Q	N	Q	R	K	R	K	R	−179.780
M-Lib #14	R	R	Q	T	T	I	M	E	N	A	N	Q	K	K	N	Q	R	−179.730
M-Lib #15	R	R	Q	I	T	I	M	E	N	Q	N	Q	R	K	N	Q	R	−179.683
M-Lib #16	R	R	Q	T	T	I	M	E	N	A	N	Q	R	K	N	Q	R	−179.618
M-Lib #17	R	R	Q	T	T	I	N	E	N	Q	N	Q	R	K	N	Q	R	−179.602
M-Lib #18	R	R	Q	T	T	I	M	E	N	Q	N	Q	R	K	N	Q	R	−179.431
M-Lib #19	R	R	Q	T	T	I	M	E	N	Q	N	Q	R	H	N	Q	R	−179.262
M-Lib #20	R	R	Q	T	T	I	M	E	Q	A	N	Q	R	K	N	Y	R	−179.234
M-Lib #21	R	H	R	T	T	I	M	E	N	A	N	Q	R	K	N	Q	R	−179.231
M-Lib #22	R	R	Q	T	T	I	M	E	N	Q	D	Q	H	K	N	Q	R	−179.189
M-Lib #23	R	R	Q	T	T	I	M	K	N	Q	D	Q	R	K	N	Q	R	−179.139
M-Lib #24	R	H	R	Q	T	I	M	E	N	Q	N	Q	R	K	N	Q	R	−179.105

Table S2. Sequences for Monomer Library (2 of 4)

Protein	AB-interface							AC-interface										Scoring Function
	21	96	10 / 6	12 / 5	12 / 7	18 / 0	18 / 2	14 / 5	14 / 9	15 / 3	16 / 2	16 / 4	17 / 4	17 / 6	19 / 2	19 / 4	21 / 6	Value
M-Lib #25	R	R	Q	T	T	I	M	E	N	A	D	Q	R	K	N	Q	R	−179.079
M-Lib #26	R	H	R	T	T	I	M	E	N	Q	D	Q	R	K	N	Q	R	−179.068
M-Lib #27	R	H	R	T	T	I	M	K	N	Q	N	Q	R	K	N	Q	R	−179.064
M-Lib #28	R	R	Q	T	T	I	M	K	Q	Q	N	Q	R	K	N	Y	R	−179.059
M-Lib #29	R	R	Q	T	T	I	M	K	N	A	N	Q	R	K	N	Q	R	−179.047
M-Lib #30	R	R	Q	Q	T	I	M	E	N	A	N	Q	R	K	N	Q	R	−179.032
M-Lib #31	R	T	Q	T	R	I	M	E	N	A	N	Q	R	K	N	Q	R	−178.907
M-Lib #32	R	R	Q	Q	T	I	M	E	N	Q	N	Q	K	K	N	Q	R	−178.888
M-Lib #33	R	R	Q	Q	T	I	M	E	N	Q	D	Q	R	K	N	Q	R	−178.866
M-Lib #34	R	T	Q	T	R	I	M	K	N	Q	N	Q	R	K	N	Q	R	−178.757
M-Lib #35	R	R	Q	I	T	I	M	E	N	Q	N	Q	K	K	N	Q	R	−178.684
M-Lib #36	R	T	Q	T	R	I	M	E	N	Q	N	Q	R	K	N	Q	R	−178.610
M-Lib #37	R	H	R	T	T	I	M	E	N	Q	D	Q	K	K	N	Q	R	−178.581
M-Lib #38	R	R	T	T	R	I	M	E	N	Q	N	Q	R	K	N	Q	R	−178.532
M-Lib #39	R	R	Q	T	T	I	M	K	N	Q	N	Q	Q	K	N	Q	R	−178.531
M-Lib #40	R	R	Q	T	T	I	M	E	N	A	N	Q	Q	K	N	Q	R	−178.496
M-Lib #41	R	H	R	T	T	I	M	E	N	Q	N	Q	Q	K	N	Q	R	−178.491
M-Lib #42	R	R	Q	Q	T	I	M	E	N	Q	N	Q	Q	K	N	Q	R	−178.289
M-Lib #43	R	R	Q	T	T	I	M	E	D	A	N	Q	R	K	R	K	R	−178.270
M-Lib #44	R	R	Q	T	T	I	M	E	N	A	N	Q	R	K	N	Q	R	−178.242
M-Lib #45	R	R	Q	T	T	I	M	E	N	Q	D	Q	K	K	N	Q	R	−178.197
M-Lib #46	R	T	Q	T	T	I	M	E	N	Q	N	Q	Q	K	N	Q	R	−178.184
M-Lib #47	R	R	T	Q	R	I	M	E	N	Q	N	Q	R	K	N	Q	R	−178.146
M-Lib #48	R	K	T	Q	T	I	M	E	N	Q	N	Q	R	K	N	Q	R	−178.123

Table S2. Sequences for Monomer Library (3 of 4)

Protein	AB-interface							AC-interface										Scoring Function
	21	96	10/6	12/5	12/7	18/0	18/2	14/5	14/9	15/3	16/2	16/4	17/4	17/6	19/2	19/4	21/6	Value
M-Lib #49	R	R	Q	T	T	I	M	E	D	Q	D	Q	R	K	R	K	R	−178.122
M-Lib #50	R	T	N	T	T	I	M	E	N	Q	N	Q	R	K	N	Q	R	−178.119
M-Lib #51	R	R	Q	I	T	I	M	E	N	Q	D	Q	R	K	N	Q	R	−178.074
M-Lib #52	R	R	Q	I	T	I	M	E	N	N	N	Q	R	K	N	Q	R	−178.074
M-Lib #53	R	R	Q	Q	T	I	M	E	N	A	N	Q	K	K	N	Q	R	−178.004
M-Lib #54	R	R	Q	T	T	I	M	E	N	A	N	Q	H	K	N	Q	R	−177.918
M-Lib #55	R	R	Q	T	T	I	N	K	N	A	N	Q	R	K	N	Q	R	−177.903
M-Lib #56	R	R	Q	I	T	I	M	K	N	Q	N	Q	R	K	N	Q	R	−177.883
M-Lib #57	R	R	Q	T	T	T	N	E	N	S	N	Q	R	K	N	Q	R	−177.720
M-Lib #58	R	H	R	T	T	I	M	E	N	Q	N	Q	H	K	N	Q	R	−177.702
M-Lib #59	R	R	Q	T	T	I	M	E	N	A	D	Q	H	K	N	Q	R	−177.661
M-Lib #60	R	R	Q	T	T	I	M	K	N	A	D	Q	R	K	N	Q	R	−177.624
M-Lib #61	R	T	Q	Q	R	I	M	E	N	Q	N	Q	H	K	N	Q	R	−177.599
M-Lib #62	R	H	R	T	T	I	M	E	N	A	N	Q	R	K	N	Q	R	−177.594
M-Lib #63	R	H	R	T	T	I	M	E	N	A	N	Q	R	K	N	Q	R	−177.559
M-Lib #64	R	H	R	T	T	I	M	K	N	A	D	Q	R	K	N	Q	R	−177.546
M-Lib #65	R	H	R	I	T	I	M	E	Q	Q	N	Q	R	K	N	Y	R	−177.538
M-Lib #66	R	R	Q	Q	T	I	M	E	N	Q	N	Q	Q	K	N	Q	R	−177.497
M-Lib #67	R	H	R	Q	T	I	M	E	N	Q	D	Q	R	K	N	Q	R	−177.423
M-Lib #68	R	R	Q	Q	T	I	M	K	N	Q	D	Q	R	K	N	Q	R	−177.413
M-Lib #69	R	R	Q	Q	T	I	M	K	N	A	N	Q	R	K	N	Q	R	−177.363
M-Lib #70	R	R	Q	Q	T	I	M	E	N	A	D	Q	R	K	N	Q	R	−177.346
M-Lib #71	R	H	R	T	T	I	M	K	N	S	N	Q	R	K	N	Q	R	−177.317
M-Lib #72	R	R	Q	Q	T	I	M	E	N	Q	D	Q	H	K	N	Q	R	−177.314

Table S2. Sequences for Monomer Library (4 of 4)

Protein	AB-interface							AC-interface										Scoring Function
	21	96	10/6	12/5	12/7	18/0	18/2	14/5	14/9	15/3	16/2	16/4	17/4	17/6	19/2	19/4	21/6	Value
M-Lib #73	R	T	Q	T	R	I	M	K	Q	Q	N	Q	R	K	N	Y	R	−177.218
M-Lib #74	R	T	Q	T	R	I	M	K	N	A	N	Q	R	K	N	Q	R	−177.204
M-Lib #75	R	H	R	T	T	I	M	K	N	Q	N	Q	K	K	N	Q	R	−177.142
M-Lib #76	R	T	Q	T	R	I	M	E	N	A	D	Q	R	K	N	Q	R	−177.071
M-Lib #77	R	T	Q	T	R	I	M	K	N	Q	D	Q	R	K	N	Q	R	−177.068
M-Lib #78	R	R	T	T	T	I	M	E	N	Q	N	Q	K	K	N	Q	R	−177.018
M-Lib #79	R	H	R	T	T	I	M	K	N	Q	N	Q	Q	K	N	Q	R	−177.003
M-Lib #80	R	H	R	T	T	I	M	E	N	A	N	Q	Q	K	N	Q	R	−176.963
M-Lib #81	R	T	Q	I	R	I	M	K	N	Q	N	Q	R	K	N	Q	R	−176.925
M-Lib #82	R	T	Q	I	R	I	M	E	N	Q	D	Q	R	K	N	Q	R	−176.912
M-Lib #83	R	H	R	T	T	I	M	K	N	Q	N	Q	K	K	N	Q	R	−176.906
M-Lib #84	R	T	Q	T	R	I	M	E	N	S	D	Q	R	K	N	Q	R	−176.870
M-Lib #85	R	T	Q	T	T	I	M	E	N	Q	D	Q	R	K	N	Q	R	−176.836
M-Lib #86	R	R	Q	T	T	I	M	K	N	Q	D	Q	K	K	N	Q	R	−176.749
M-Lib #87	R	R	Q	T	T	I	M	K	N	N	D	Q	R	K	N	Q	R	−176.736
M-Lib #88	R	R	T	T	T	I	M	E	N	Q	D	Q	R	K	N	Q	R	−176.733
M-Lib #89	R	R	Q	T	T	I	M	E	N	Q	N	S	K	N	N	Q	R	−176.680
M-Lib #90*	R	R	Q	T	T	I	T	E	N	Q	N	Q	R	K	N	Q	R	---
M-Lib #91*	R	R	Q	I	R	I	M	E	N	Q	D	Q	R	K	N	Q	R	---
M-Lib #92*	R	R	Q	T	T	I	M	E	N	Q	N	S	K	K	N	Q	R	---
M-Lib #93*	R	R	Q	T	T	I	M	E	N	Q	N	Q	K	N	N	Q	R	---
M-Lib #94*	R	R	Q	T	T	I	M	E	N	A	N	Q	R	K	R	Q	R	---
M-Lib #95*	R	R	Q	T	T	I	M	E	D	A	N	Q	R	K	N	Q	R	---

* M-Lib #90 through M-Lib #95 were obtained randomly through the gene assembly method used; these mutants were not ranked among the top 100 sequences by the CPD scoring function.

Table S3. Interface Residues Randomly Sampled from Distribution of Surface Residues in Bacteria (Random Library)

Protein	AB-interface							AC-interface									
	21	96	106	125	127	180	182	145	149	153	162	164	174	176	192	194	216
R-Lib #1	N	A	E	A	K	Q	A	N	R	N	K	K	K	N	T	D	T
R-Lib #2	T	R	D	D	Y	D	R	R	S	A	E	Y	T	S	N	Y	H
R-Lib #3	E	Q	R	Q	E	D	D	E	H	D	K	D	E	R	K	N	R
R-Lib #4	A	Y	D	N	A	N	K	E	K	N	K	E	T	E	Y	Y	E
R-Lib #5	E	E	E	E	K	A	N	R	Y	D	T	A	A	D	A	S	E
R-Lib #6	T	A	K	E	T	H	R	H	S	D	R	Y	E	K	E	Q	D
R-Lib #7	A	K	D	N	E	K	D	S	Q	K	H	K	E	Q	Y	E	E
R-Lib #9	Y	Q	H	–	Y	D	Q	R	A	R	Q	Y	T	R	E	D	E
R-Lib #10	S	D	H	R	Q	D	A	E	T	K	K	D	T	E	R	A	A
R-Lib #12	Q	Q	N	S	T	E	T	Y	D	A	E	E	K	E	Q	K	R
R-Lib #13	K	R	K	Q	H	H	N	H	E	K	E	N	S	K	R	R	D
R-Lib #14	Q	E	H	E	T	S	H	K	R	N	R	K	E	Q	E	E	Y
R-Lib #15	R	T	S	K	Q	E	Y	R	R	E	R	E	K	R	S	D	K
R-Lib #16	K	V	A	A	A	R	A	Q	Q	T	T	S	K	A	K	S	E
R-Lib #17	D	Q	K	N	A	T	R	R	E	H	H	D	N	Q	H	Y	T
R-Lib #18	D	Q	D	Q	V	D	N	T	N	N	A	T	E	S	D	K	R
R-Lib #19	K	Q	D	A	A	K	E	D	N	K	Q	K	D	S	D	T	A
R-Lib #20	T	E	A	E	A	A	E	E	A	T	N	Q	S	N	N	S	D
R-Lib #21	K	D	E	A	T	Y	A	N	N	K	Y	Q	D	S	E	S	D
R-Lib #22	R	S	Q	S	E	R	S	H	S	E	E	Q	D	E	A	K	D

\

Table S4. Data for Individual Mutants in Monomer Library (1 of 3)

Protein	λ_{max}^{abs}	λ_{max}^{ems}	Φ_F	$\lambda_{max}^{abs}/\lambda_{280nm}^{abs}$	Brightness Proxy*	r, mP[†]	r vs. DF (slope)[†]
M-Lib #1	586	609	0.355 ± 0.005	0.75	26.5 ± 3.3	415.0 ± 2.3	−7.1 ± 2.2
M-Lib #2	584	607	0.371 ± 0.073	0.59	22.0 ± 9.8	416.7 ± 0.6	---
M-Lib #3	586	610	0.276 ± 0.005	1.07	29.5 ± 4.0	411.3 ± 5.3	---
M-Lib #4	586	610	0.278 ± 0.007	1.05	29.3 ± 4.9	411.7 ± 3.2	−6.2 ± 2.2
M-Lib #5	584	610	0.266 ± 0.003	1.07	28.3 ± 3.4	411.3 ± 4.9	−8.3 ± 0.3
M-Lib #6	586	609	0.299 ± 0.005	0.83	24.9 ± 3.4	413.7 ± 7.6	---
M-Lib #7	586	608	0.306 ± 0.005	0.65	20.0 ± 2.7	414.7 ± 8.1	---
M-Lib #8	584	609	0.320 ± 0.016	0.74	23.8 ± 5.3	413.7 ± 3.8	---
M-Lib #9	586	610	0.289 ± 0.017	1.06	30.6 ± 7.4	410.3 ± 3.8	−6.0 ± 1.4
M-Lib #10	586	611	0.277 ± 0.011	1.07	29.8 ± 6.1	408.0 ± 6.4	−8.6 ± 2.5
M-Lib #11	586	610	0.265 ± 0.006	1.01	26.9 ± 4.2	408.7 ± 6.1	−8.1 ± 3.1
M-Lib #12	586	610	0.259 ± 0.002	1.06	27.4 ± 2.4	406.0 ± 5.2	−8.3 ± 1.4
M-Lib #13	586	609	0.353 ± 0.004	0.99	34.8 ± 4.1	412.3 ± 0.6	−9.6 ± 1.4
M-Lib #14	586	610	0.275 ± 0.003	1.06	29.3 ± 3.4	412.3 ± 0.6	−5.9 ± 0.8
M-Lib #15	586	611	0.268 ± 0.009	1.10	29.5 ± 5.4	411.7 ± 1.5	−7.4 ± 2.7
M-Lib #16	584	606	0.336 ± 0.012	0.42	14.1 ± 2.7	419.0 ± 2.0	---
M-Lib #17	584	606	0.331 ± 0.011	0.46	15.1 ± 2.8	417.7 ± 0.6	---
M-Lib #18	586	610	0.271 ± 0.002	1.01	27.3 ± 2.8	411.7 ± 1.5	−8.0 ± 2.8
M-Lib #19	586	610	0.249 ± 0.008	1.10	27.3 ± 4.9	409.7 ± 1.5	−10.1 ± 4.3
M-Lib #20	586	610	0.295 ± 0.014	0.91	26.9 ± 5.9	411.7 ± 0.6	−5.5 ± 1.7
M-Lib #21	586	611	0.264 ± 0.011	1.13	30.0 ± 6.1	410.0 ± 0.0	−6.9 ± 2.4
M-Lib #22	586	610	0.247 ± 0.007	1.19	29.4 ± 5.2	407.3 ± 0.6	−9.7 ± 0.6
M-Lib #23	586	610	0.234 ± 0.005	1.25	29.3 ± 4.6	404.0 ± 1.0	−12.2 ± 2.8
M-Lib #24	584	609	0.284 ± 0.003	0.79	22.4 ± 2.6	412.0 ± 1.0	---
M-Lib #25	586	610	0.302 ± 0.001	1.09	32.9 ± 2.2	411.0 ± 0.0	−10.7 ± 0.6
M-Lib #26	586	610	0.275 ± 0.034	1.12	30.7 ± 10.9	412.0 ± 1.0	−9.5 ± 1.7
M-Lib #27	586	610	0.264 ± 0.005	1.13	29.8 ± 4.3	412.0 ± 0.0	−7.0 ± 2.1
M-Lib #28	586	609	0.265 ± 0.005	1.08	28.5 ± 4.1	411.0 ± 1.0	−9.4 ± 1.8
M-Lib #29	584	609	0.270 ± 0.005	1.04	28.1 ± 4.0	412.3 ± 1.2	−8.5 ± 2.6
M-Lib #30	586	609	0.267 ± 0.005	1.05	28.2 ± 4.0	411.0 ± 1.0	−8.8 ± 1.7
M-Lib #31	586	610	0.252 ± 0.008	1.11	27.9 ± 5.1	410.0 ± 1.0	−8.9 ± 3.9
M-Lib #32	586	610	0.265 ± 0.010	1.11	29.4 ± 5.9	409.3 ± 1.5	−10.0 ± 2.2
M-Lib #33	586	611	0.267 ± 0.014	1.08	28.9 ± 6.7	410.0 ± 1.0	−7.0 ± 3.4
M-Lib #34	586	610	0.254 ± 0.004	1.20	30.5 ± 3.9	406.7 ± 0.6	−11.3 ± 2.7
M-Lib #35	586	610	0.246 ± 0.006	1.14	28.1 ± 4.6	407.3 ± 0.6	−10.7 ± 0.6
M-Lib #36	586	611	0.236 ± 0.005	1.16	27.3 ± 4.0	405.7 ± 0.6	−9.3 ± 2.0
M-Lib #37	586	610	0.289 ± 0.001	1.03	29.6 ± 2.0	411.3 ± 0.6	−7.9 ± 1.2
M-Lib #38	586	610	0.274 ± 0.031	0.97	26.6 ± 9.0	413.0 ± 1.0	−6.4 ± 1.9
M-Lib #39	586	610	0.257 ± 0.005	1.06	27.3 ± 4.0	410.3 ± 0.6	−10.1 ± 1.5
M-Lib #40	586	609	0.270 ± 0.003	0.91	24.7 ± 2.9	412.3 ± 1.2	−8.0 ± 1.5

Table S4. Data for Individual Mutants in Monomer Library (2 of 3)

Protein	λ_{max}^{abs}	λ_{max}^{ems}	Φ_F	$\lambda_{max}^{abs}/\lambda_{280nm}^{abs}$	Brightness Proxy*	r, mP[†]	r vs. DF (slope)[†]
M-Lib #41	586	610	0.253 ± 0.007	0.98	24.9 ± 4.3	412.7 ± 1.5	−10.1 ± 1.7
M-Lib #42	586	610	0.254 ± 0.002	0.96	24.3 ± 2.4	411.0 ± 1.0	−9.8 ± 2.3
M-Lib #43	586	609	0.272 ± 0.005	0.86	23.3 ± 3.3	411.3 ± 1.2	−8.2 ± 2.1
M-Lib #44	586	610	0.252 ± 0.004	1.00	25.1 ± 3.5	409.3 ± 1.2	−11.4 ± 3.0
M-Lib #45	586	611	0.241 ± 0.010	1.10	26.6 ± 5.5	408.3 ± 0.6	−9.8 ± 1.1
M-Lib #46	586	610	0.253 ± 0.008	1.03	26.0 ± 4.7	408.3 ± 1.2	−10.1 ± 0.7
M-Lib #47	586	611	0.251 ± 0.007	0.99	24.7 ± 4.2	408.7 ± 0.6	−8.6 ± 1.7
M-Lib #48	586	610	0.229 ± 0.006	1.04	23.8 ± 4.0	404.3 ± 0.6	−8.6 ± 1.3
M-Lib #49	586	611	0.304 ± 0.010	1.08	33.0 ± 6.0	410.0 ± 0.0	−10.6 ± 2.2
M-Lib #50	586	610	0.276 ± 0.028	1.06	29.4 ± 9.4	411.7 ± 0.6	−9.5 ± 1.0
M-Lib #51	586	610	0.264 ± 0.004	1.04	27.4 ± 3.7	412.3 ± 1.2	−6.4 ± 1.2
M-Lib #52	586	610	0.264 ± 0.005	1.00	26.4 ± 3.7	412.7 ± 0.6	-8.1 ± 1.0
M-Lib #53	---	---	0.024	---	---	---	---
M-Lib #54	586	609	0.273 ± 0.003	0.84	23.0 ± 2.6	413.7 ± 0.6	−7.3 ± 1.1
M-Lib #55	586	606	0.311 ± 0.011	0.42	13.1 ± 2.5	417.7 ± 0.6	---
M-Lib #56	586	610	0.254 ± 0.001	1.11	28.4 ± 2.3	409.3 ± 1.2	−10.0 ± 1.4
M-Lib #57	586	607	0.320 ± 0.016	0.43	13.9 ± 3.2	416.0 ± 1.0	---
M-Lib #58	586	610	0.257 ± 0.009	1.05	27.0 ± 5.3	410.0 ± 1.0	−9.2 ± 1.1
M-Lib #59	586	610	0.237 ± 0.011	1.08	25.7 ± 5.5	408.0 ± 1.0	−7.7 ± 0.7
M-Lib #60	584	610	0.229 ± 0.004	1.15	26.4 ± 3.6	404.0 ± 1.0	−7.7 ± 1.8
M-Lib #61	586	610	0.285 ± 0.002	1.13	32.4 ± 3.0	409.7 ± 0.6	−10.7 ± 1.1
M-Lib #62	586	610	0.288 ± 0.033	0.96	27.7 ± 9.5	414.0 ± 1.0	−7.4 ± 0.5
M-Lib #63	586	610	0.244 ± 0.002	1.17	28.4 ± 2.8	409.7 ± 0.6	−9.1 ± 1.7
M-Lib #64	586	610	0.241 ± 0.003	1.13	27.1 ± 3.4	410.7 ± 1.5	−8.6 ± 3.8
M-Lib #65	584	608	0.255 ± 0.004	1.04	26.5 ± 3.4	412.0 ± 1.0	−9.4 ± 2.9
M-Lib #66	---	606	0.044	---	---	---	---
M-Lib #67	586	610	0.233 ± 0.003	1.16	27.1 ± 3.0	409.0 ± 1.0	−10.5 ± 2.5
M-Lib #68	586	610	0.247 ± 0.003	1.15	28.4 ± 3.2	407.7 ± 1.5	−8.0 ± 3.0
M-Lib #69	586	609	0.267 ± 0.018	1.08	28.9 ± 7.5	408.3 ± 1.2	−8.6 ± 2.0
M-Lib #70	586	610	0.276 ± 0.018	0.98	26.9 ± 6.9	409.7 ± 0.6	−7.1 ± 1.3
M-Lib #71	586	609	0.248 ± 0.008	1.11	27.6 ± 5.0	408.3 ± 0.6	−7.3 ± 1.2
M-Lib #72	586	611	0.225 ± 0.001	1.16	26.1 ± 1.8	404.3 ± 0.6	−10.0 ± 1.8
M-Lib #73	586	608	0.290 ± 0.009	1.11	32.3 ± 5.9	410.0 ± 0.0	---
M-Lib #74	586	610	0.256 ± 0.022	1.21	31.0 ± 9.3	408.3 ± 1.2	−9.9 ± 3.9
M-Lib #75	586	609	0.249 ± 0.001	1.13	28.2 ± 2.4	410.3 ± 1.5	−6.7 ± 2.8
M-Lib #76	586	611	0.246 ± 0.001	1.10	26.9 ± 2.1	410.0 ± 1.0	−9.0 ± 2.4
M-Lib #77	586	610	0.235 ± 0.003	1.21	28.3 ± 3.5	407.3 ± 0.6	−11.4 ± 2.4
M-Lib #78	586	610	0.251 ± 0.005	1.13	28.2 ± 4.0	409.3 ± 1.5	−10.1 ± 2.3
M-Lib #79	---	608	0.093	---	---	---	---
M-Lib #80	586	610	0.266 ± 0.018	1.00	26.7 ± 7.0	411.0 ± 1.0	−6.5 ± 0.3

Table S4. Data for Individual Mutants in Monomer Library (3 of 3)

Protein	λ_{max}^{abs}	λ_{max}^{ems}	Φ_F	$\lambda_{max}^{abs}/\lambda_{280nm}^{abs}$	Brightness Proxy*	r, mP[†]	r vs. DF (slope)[†]
M-Lib #81	586	609	0.251 ± 0.018	1.18	29.7 ± 8.1	406.0 ± 1.0	−9.0 ± 1.4
M-Lib #82	586	611	0.238 ± 0.009	1.15	27.5 ± 5.4	406.0 ± 0.0	−10.2 ± 1.7
M-Lib #83	586	610	0.240 ± 0.007	1.03	24.7 ± 4.3	406.7 ± 0.6	−8.9 ± 1.1
M-Lib #84	586	610	0.239 ± 0.008	1.06	25.4 ± 4.8	405.0 ± 0.0	−9.3 ± 1.1
M-Lib #85	586	611	0.267 ± 0.003	1.10	29.3 ± 3.5	407.7 ± 0.6	−12.7 ± 0.7
M-Lib #86	586	610	0.240 ± 0.010	1.16	27.7 ± 5.8	407.0 ± 1.7	−10.0 ± 2.5
M-Lib #87	586	610	0.255 ± 0.005	0.95	24.2 ± 3.6	409.3 ± 1.2	−7.1 ± 3.0
M-Lib #88	586	611	0.236 ± 0.003	1.01	23.8 ± 2.7	408.7 ± 0.6	−7.8 ± 1.4
M-Lib #89	584	608	0.289 ± 0.007	0.38	11.0 ± 1.8	415.7 ± 0.6	---
M-Lib #90	586	610	0.238 ± 0.012	1.04	24.6 ± 5.7	407.0 ± 1.0	−10.6 ± 0.3
M-Lib #91	586	610	0.240 ± 0.002	0.98	23.4 ± 2.2	408.7 ± 0.6	−8.7 ± 1.5
M-Lib #92	586	611	0.236 ± 0.003	1.03	24.3 ± 3.1	407.0 ± 1.0	−8.5 ± 2.1
M-Lib #93	586	611	0.249 ± 0.019	0.91	22.6 ± 6.4	409.3 ± 0.6	---
M-Lib #94	586	609	0.250 ± 0.017	0.89	22.3 ± 5.8	408.3 ± 0.6	−10.2 ± 3.3
M-Lib #95	586	609	0.263 ± 0.014	0.80	20.9 ± 4.8	408.0 ± 0.0	−6.6 ± 0.9

* The "Brightness Proxy" score is defined as follows: $100 \cdot (\lambda_{max}^{abs}/\lambda_{280nm}^{abs}) \cdot \Phi_F$

[†] Fluorescence anisotropy (r) was measured in milli-polarization units (mP). DF = Dilution Factor. The slope of r versus DF is a proxy for the dissociation constant (K_d) of an homo-oligomeric FP; a more negative slope indicates a higher K_d.

Table S5. Data for Individual Mutants in Random Library

Protein	λ_{max}^{abs}	λ_{max}^{ems}	Φ_F	$\lambda_{max}^{abs}/\lambda_{280nm}^{abs}$	Brightness Proxy*	r, mP[†]	r vs. DF (slope)[†]
R-Lib #1	590	609	0.141 ± 0.021	0.06	0.8 ± 0.3	---	---
R-Lib #2	588	611	0.249 ± 0.020	0.94	23.4 ± 6.7	402.7 ± 1.5	−6.7 ± 2.8
R-Lib #3	---	---	0.000 ± 0.001	---	---	---	---
R-Lib #4	---	---	0.001 ± 0.000	---	---	---	---
R-Lib #5	552	610	0.015 ± 0.023	0.13	0.2 ± 0.2	---	---
R-Lib #6	---	609	0.010 ± 0.058	---	---	---	---
R-Lib #7	---	607	0.013 ± 0.005	---	---	---	---
R-Lib #9	---	613	0.077 ± 0.062	---	---	---	---
R-Lib #10	---	610	---	---	---	---	---
R-Lib #12	582	609	0.089 ± 0.051	0.07	0.6 ± 0.5	---	---
R-Lib #13	---	605	0.000 ± 0.001	---	---	---	---
R-Lib #14	586	609	0.219 ± 0.007	1.19	25.9 ± 4.6	398.7 ± 0.6	−3.7 ± 2.4
R-Lib #15	---	610	0.128	---	---	---	---
R-Lib #16	582	609	0.193 ± 0.141	0.06	1.2 ± 1.0	---	---
R-Lib #17	586	610	0.218 ± 0.017	0.21	4.7 ± 1.3	404.3 ± 2.5	---
R-Lib #18	---	---	0.001 ± 0.000	---	---	---	---
R-Lib #19	---	---	---	---	---	---	---
R-Lib #20	---	612	---	---	---	---	---
R-Lib #21	---	---	0.000	---	---	---	---
R-Lib #22	---	612	0.012 ± 0.015	---	---	---	---
R-Lib #23	588	609	0.086 ± 0.038	0.06	0.5 ± 0.4	---	---
R-Lib #24	---	612	0.002	---	---	---	---

* The "Brightness Proxy" score is defined as follows: $100 \cdot (\lambda_{max}^{abs}/\lambda_{280nm}^{abs}) \cdot \Phi_F$

[†] Fluorescence anisotropy (r) was measured in milli-polarization units (mP). DF = Dilution Factor. The slope of r versus DF is a proxy for the dissociation constant (K_d) of an homo-oligomeric FP; a more negative slope indicates a higher K_d.