

# Computationally Guided Monomerization of Red Fluorescent Proteins of the Class *Anthozoa*

Thesis by  
Timothy M. Wannier

In Partial Fulfillment of the Requirements for the  
Degree of  
Doctor of Philosophy



CALIFORNIA INSTITUTE OF TECHNOLOGY

Pasadena, California

2015

(Defended December 19<sup>th</sup>, 2014)

© 2015

Timothy M. Wannier

All Rights Reserved

## ACKNOWLEDGEMENTS

I took a fairly circuitous route to finishing my Ph.D. here at Caltech. Nine and a half years ago I was accepted to the graduate program, but instead chose to attend UCLA. I bailed out of the graduate program at UCLA after two years to pursue a career in finance, which had been the focus of my undergraduate studies. Looking back on this decision, it is clear that it was made for the wrong reasons, but I would not be where I am today without having traveled that path. Two years later in London, with the small private equity firm that I had joined closing up shop, I yearned for my days in science, and back at the lab. This led me to contacting Steve Mayo for a position as an unpaid volunteer that eventually morphed into a tech job and re-admission to Caltech's Ph.D. program, which I am now proudly graduating from.

Thank you Steve Mayo for accepting me into your lab, and for helping me regain my footing here at Caltech. In your lab I discovered a field of science that bridged biology and engineering, and never failed to present difficult and sometimes impossible-seeming problems. I appreciate the degree to which you let me find my own way and direct my research. Your insight has been invaluable, but your advice never smothering. I especially appreciate your support in helping me to continue to the end of my degree, which at times seemed very far off. I've enjoyed trips to Protein Society in San Diego, conferences at UCLA, and had the personal freedom to explore the field with the benefit of your stewardship. I will always be grateful to you for my time at Caltech.

Thank you Frances Arnold for making me feel at home with your lab, and for encouraging me to explore opsins with some of your bright and talented lab members. I have benefitted from your sharp insight and intuition. The experience with your group has been one of the best parts of the latter stage of my graduate career.

Thank you Niles Pierce for helping me fall in love with science all those years ago as a SURF student in your lab. I enjoyed engineering a control system for a DNA enzyme, and

learned some valuable laboratory techniques and the lay of the land here at Caltech. Thank you also for being there and listening to me as I tried to find my way in the world.

Many thanks also to the rest of my committee. Pamela Bjorkman, you were very fun to work for as a TA in Bi 1, and listened patiently to my crystallography woes. David Prober, you provided some critical support in grant writing and experimental design.

Thank you Arnold Berk, Yi Sun, and the staff at UCLA's Molecular Biology Institute, for even if I did not finish out my degree, I learned a lot from my two years at school. The basics of laboratory biology that I learned spanned a wide range of disciplines from studying mouse brains to stem cell tissue culture to *Drosophila* to basic yeast molecular biology. I will always remember the years at UCLA in the company of Ryan Schmidt, Matthew McBrian, Ian McHardy, David Shanley, and Brian Duistermars fondly.

My work on fluorescent proteins would never have been possible without the prior scientific endeavors of Roberto Chica and Matthew Moore. Matthew especially took me under his wing and taught me more than I could have ever hoped for about biochemistry, protein engineering and fluorescent proteins. I've also enjoyed collaborations with a variety of people over the course of my research, including Kevin Brown at the University of Connecticut, who is a world expert at analyzing protein sequence data, Jens Kaiser and Pavle Nikolovski at the Molecular Observatory, who have been extremely helpful to the cause of many a fluorescent protein crystal, and Scott McIsaac who helped electroshock a fluorescent protein design project that had run into hard times.

My time in the Mayo lab has been made ever enjoyable by interactions with the best sort of scientists and labmates. Jan Kostecki, Alex Nisthal, Matthew Moore, Kurt Mou, Emzo de los Santos, Jackson Cahn, Bernardo Sosa Padilla Araujo, Seth Lieblich, Mohsen Chitsaz, Gene Kym, Samy Hamdouche, Toni Lee and Alexandria Berry; I will miss you all. Also, the whole Arnold lab was fun and welcoming to me, especially team opsin, which comprised Scott McIsaac, Austin Rice, and Lukas Herwig.

Finally, a big thank you to my family and friends in and around LA who have kept me afloat through the years in lab. Kelly Roberts, Nate Lam, Vince Mateus, Phil Kenney, Chris Garnic, Dan Gross, Eric Murtfeldt, mom, dad, grandma, and the whole Wannier/Saffman/Roberts clan. Thanks for all the love.

## ABSTRACT

Red fluorescent proteins (RFPs) have attracted significant engineering focus because of the promise of near infrared fluorescent proteins, whose light penetrates biological tissue, and which would allow imaging inside of vertebrate animals. The RFP landscape, which numbers ~200 members, is mostly populated by engineered variants of four native RFPs, leaving the vast majority of native RFP biodiversity untouched. This is largely due to the fact that native RFPs are obligate tetramers, limiting their usefulness as fusion proteins. Monomerization has imposed critical costs on these evolved tetramers, however, as it has invariably led to loss of brightness, and often to many other adverse effects on the fluorescent properties of the derived monomeric variants. Here we have attempted to understand why monomerization has taken such a large toll on *Anthozoa* class RFPs, and to outline a clear strategy for their monomerization. We begin with a structural study of the far-red fluorescence of AQ143, one of the furthest red emitting RFPs. We then try to separate the problem of stable and bright fluorescence from the design of a soluble monomeric  $\beta$ -barrel surface by engineering a hybrid protein (DsRmCh) with an oligomeric parent that had been previously monomerized, DsRed, and a pre-stabilized monomeric core from mCherry. This allows us to use computational design to successfully design a stable, soluble, fluorescent monomer. Next we took HcRed, which is a previously unmonomerized RFP that has far-red fluorescence ( $\lambda_{\text{emission}} = 633 \text{ nm}$ ) and attempted to monomerize it making use of lessons learned from DsRmCh. We engineered two monomeric proteins by pre-stabilizing HcRed's core, then monomerizing in stages, making use of computational design and directed evolution techniques such as error-prone mutagenesis and DNA shuffling. We call these proteins mGinger0.1 ( $\lambda_{\text{em}} = 637 \text{ nm} / \Phi = 0.02$ ) and mGinger0.2 ( $\lambda_{\text{em}} = 631 \text{ nm} \Phi = 0.04$ ). They are the furthest red first generation monomeric RFPs ever developed, are significantly thermostabilized, and add diversity to a small field of far-red monomeric RFPs. We anticipate that the techniques we describe will facilitate future RFP monomerization, and that further core optimization of the mGingers may allow significant improvements in brightness.

## TABLE OF CONTENTS

Acknowledgements.....	iii
Abstract.....	vi
Table of Contents.....	vii
Nomenclature.....	ix
Chapter I: Introduction.....	1
Background.....	1
Statement of Problem.....	2
Experimental Strategy.....	4
Summary of Results.....	5
References.....	9
Chapter II: The Structure of a Far-Red Fluorescent Protein, AQ143, Shows Evidence in Support of Reported Red-Shifting Chromophore Interactions.....	13
Abstract.....	13
Introduction.....	14
Results and Discussion.....	15
Conclusion.....	20
Materials and Methods.....	20
Acknowledgements.....	21
Tables and Figures.....	22
Supplementary Tables and Figures.....	25
References.....	34
Chapter III: Computational Design of the $\beta$ -sheet Surfaces of Red Fluorescent Proteins Allows Control of Protein Oligomerization.....	37
Abstract.....	37
Introduction.....	38
Results.....	40
Discussion.....	44
Materials and Methods.....	47
Acknowledgements.....	50
Tables and Figures.....	51
Supplementary Tables and Figures.....	58

References .....	61
Chapter IV: mGinger: A Soluble Far-Red Fluorescent Monomer Derived from HcRed.....	64
Abstract .....	64
Introduction .....	65
Results .....	68
Discussion .....	75
Conclusion.....	79
Materials and Methods.....	80
Acknowledgements.....	85
Tables and Figures .....	86
Supplementary Tables and Figures.....	93
References .....	103
Chapter V: Accurate Characterization of Some Common <i>Anthozoa</i> Class Red Fluorescent Proteins.....	107
Abstract .....	107
Introduction.....	108
Results and Discussion .....	109
Conclusion.....	114
Materials and Methods.....	114
Tables and Figures .....	115
References .....	120

## NOMENCLATURE

**RFP** – Red fluorescent protein

**FP** – Fluorescent protein

$\Phi$  – Quantum yield of fluorescence

$\epsilon$  – Extinction coefficient

$\lambda_{\text{abs}}$  – Maximum intensity absorbance wavelength

$\lambda_{\text{ex}}$  – Maximum intensity excitation wavelength

$\lambda_{\text{em}}$  – Maximum intensity emission wavelength

$\text{\AA}$  – Angstrom

nm – Nanometer

AUC – Analytical ultracentrifugation

SEC – Size exclusion chromatography

HPLC – High pressure liquid chromatography

T<sub>m</sub> – Melting temperature

FRET – Förster Resonance Energy Transfer

Homo-FRET – FRET between two molecules of a homodimer

mP – Milli-polarization units



# CHAPTER 1

## Introduction

### 1.1 Background

Fluorescent protein (FP) engineering emerged as a field in the late 1990's, shortly after green fluorescent protein (GFP) from *Aequorea victoria* was first isolated and characterized (1-3). The field blossomed quickly, with FPs quickly becoming indispensable imaging agents across large swaths of biological research (4-6). GFP was engineered to fluoresce at a broad spectrum of wavelengths and brightened via directed evolution (7-9). The first major expansion of FP diversity came with the discovery of new naturally occurring fluorescent proteins with altered chromophore environments that absorbed and emitted light over a broad spectrum of colors, ranging from cyan to red (10-12). This encouraged molecular biologists and protein engineers to attempt to modify FP spectra artificially with mutations to the chromophore and residues in the immediate chromophore environment, which further expanded the color palette to include deep blues and far-reds (13-18). Shortly following the discovery of the diversity of FP spectra, a broad range of attributes were discovered and engineered including halide detection and photoactivatable and photoswitchable fluorescence (19-23). The field of FP engineering blossomed quickly and branched out into a number of smaller more specialized fields such as calcium imaging, long stokes-shift fluorescent proteins, and two-photon microscopy (24-27).

Red fluorescent protein (RFP) engineering has garnered considerable sustained interest because of the possibility of engineering RFPs that are bright, photostable, and excite and emit at near-infrared (NIR) wavelengths. NIR light penetrates biological tissue with minimal absorption from the biological molecules that are the primary absorbers of light in visible to infrared wavelengths: melanin, hemoglobin, and water (28, 29). Two other classes of FPs have been reported that fluoresce at NIR wavelengths, but they are inferior

as universal biological markers. The first, a protein family derived from bacterial phytochrome called IFPs exhibit peak fluorescent emission anywhere from 670 nm to 720 nm, but require a heme-derived cofactor, biliverdin, that is not ubiquitously present in mammalian tissue (30-33). Indeed the group that engineered the IFPs needed to engineer an artificial heme oxygenase 1 into a mouse to achieve bright fluorescence from the proteins. The other family of near-infrared FPs recently reported is the transmembrane bacterial opsin family (34, 35). These proteins' peak fluorescence can range up to 731 nm, but they are very dim, located in the cell membrane, and like the bacterial phytochromes, require a cofactor, which in this case is retinal. The *Aequorea victoria* class of FPs remains the most desirable class so far identified.

## 1.2 Statement of Problem

The primary drawback of the *Aequorea victoria* class is that all native RFPs that have been characterized to date from this class are obligate tetramers, which is a hindrance to the standard usages of FPs as markers in a biological context. Most of these RFPs have been isolated from *Anthozoa* class corals and anemones (36-38). One of the primary usages of FPs is as a part of a fusion protein to track the cellular movements of a fused protein target (5). Fusing an oligomeric FP to a target protein necessarily means, though, that the target protein's cellular localization and diffusion will be impacted by the oligomeric tendency of the FP it is fused to. This could mean aggregation of the protein target, causing it to fall into inclusion bodies, signaling or other downstream effects of improper oligomerization of the target protein, or false clustering in the case of structural or membrane-bound proteins whose cellular locations are being tracked (39, 40).

Roger Tsien and colleagues engineered the first monomeric derivative of an *Anthozoa* class RFP, which they called mRFP1 (41). There are two symmetrical oligomeric interfaces that are part of each RFP tetramer, which are called the AB and the AC interface, and were so-named because of the original chain names given to the individual monomers in the first structure of DsRed (42). The engineering process of mRFP1 began with the break of the

AB interface, by sticking charged residues into the primarily hydrophobic interface. This succeeded in breaking the interface, but severely compromised fluorescence. Many rounds of directed evolution were used to recover this fluorescence, whereupon the second of the two interfaces, the more stable AC interface, was broken via the same strategy. The break of the AC interface, however, completely extinguished fluorescence, necessitating much more extensive directed evolution to recover it. The final monomeric variant, mRFP1 necessitated 33-point mutations from the native DsRed. mRFP1 was also much dimmer than DsRed, and exhibited a bathochromic shift to its fluorescent emission. mRFP1 has since been further optimized into the mFruits (mCherry, mRaspberry, tdTomato, mPlum, etc.), and other far-red FPs, but none of these variants has come anywhere close to recovering the brightness of the parent protein, DsRed (14, 43, 44).

Since the first RFP monomerization of mRFP1 in 2002, scientists have identified approximately 50 RFPs and 20 chromoproteins that absorb in far-red wavelengths, but there have only been four other instances of RFP monomerization. One of these instances was a repeated monomerization of DsRed by Benjamin Glick and colleagues (45), while the other three were monomerizations of eqFP578 to FusionRed (46), eqFP611 to mRuby (47), and COCP to mKeima (48). COCP is a native chromoprotein; meaning that fluorescence was first induced into it with a well known cysteine to serine mutation that provides a hydrogen bond to the phenolate oxygen of the chromophore, serving to stabilize a fluorescent *cis* conformation of the chromophore. In all of these cases of RFP monomerization, the process used to engineer the final monomeric variant looked very similar to the path taken to engineer mRFP1, that is to say it was long, labor intensive, and involved significant mutation to the native protein's chromophore environment and hence afforded little control over the final spectroscopic character of the resultant monomer. The lack of ability to readily monomerize RFPs has left the vast majority of native RFP biodiversity untapped in efforts to engineer improved RFP variants, as oligomerization is such a drawback to the potential use of any such marker.

In addition to the difficulty of monomerizing native RFP tetramers, there is a considerable lack of interest in targeting novel RFPs for monomerization because there has been such

significant spectroscopic change to the proteins every time they have been monomerized. These changes to fluorescence have been difficult to predict and are often harmful to the usefulness of the monomeric variants. Some common negative repercussions of RFP monomerization include diminished brightness, decreased photostability, disrupted chromophore maturation, and a hypsochromic shift to fluorescent emission (21, 49). It is not well understood why monomerization is so harmful to fluorescence as it has been hard to disentangle the effects of all of the numerous mutations made to monomers over successive rounds of evolution. More direct approaches to understanding the fluorescent changes in RFPs have not had much success. Quantum mechanical studies of chromophores have difficulty even predicting the wavelength of fluorescent emission very accurately, and struggle to capture the effects of any nearby amino-acid residues that interact with the chromophore (50, 51).

Devising a strategy to facilitate *Anthozoa* class RFP monomerization would allow greater diversity to be sampled in a monomeric context, possibly allowing for a brighter, more red-shifted variant to be monomerized. Furthermore, by better understanding RFP monomerization, the forces behind the negative spectroscopic consequences that accompany the process might be diagnosed and addressed.

### **1.3 Experimental Strategy**

It is a good guess that the fluorescent impacts of monomerization are due to changes to an RFP's immediate chromophore environment, as the actual chemical makeup of its chromophore remains unchanged. Data demonstrating this have never been thoroughly presented, and doing so constitutes the first part of our efforts to rationalize and explore the landscape of RFP monomerization and far-red FP engineering. We attempt to disentangle the problems of (1) engineering a soluble monomeric protein and (2) ensuring that the new monomeric protein remains fluorescent and retains the desired spectroscopic attributes. Separately addressing these two problems involves innovative approaches to protein design, as distinct structural areas that overlap in primary sequence space (opposite sides of

a  $\beta$ -sheet alternate residues) are difficult to design libraries around. We sought to rationally target small groups of structural regions (5-20 residues), and use random or computationally designed libraries to query these spaces.

This approach is novel, as past RFP engineering, and protein engineering in general, have for the most part focused on design and evolution at two very opposite ends of a continuum. On the one hand, site-saturation mutagenesis and rational design consider small numbers of residues and attempt to exhaustively search sequence space for the best combination of mutations at these positions (52-54). The other extreme is the classic directed evolution by random mutagenesis, primarily with error-prone PCR, that searches slowly through sequence space by making mostly random changes to a few residues at a time in a protein sequence, as large numbers of random mutations cause a protein to lose its functionality (55). These strategies have their benefits and their limitations. Site-saturation mutagenesis and rationally guided mutation can answer very specific mechanistic questions and exhaustively query the mutational landscape around small groups of residues and structural regions. By contrast, they are limited in their scope as in mutating only a few residues, it is easy to miss important effects at distal positions. Error-prone mutagenesis, on the other hand, can access vast areas of sequence space that might not be targeted by any rational approach, and is very easily implemented, but lacks direction, and necessitates the screening of very large libraries of variants to isolate synergistic mutations that are individually deleterious or neutral with regard to the screened attribute. In this work we try to access an area between rational design and random mutagenesis on the protein design spectrum.

## **1.4 Summary of Results**

As an initial study of far-red FP engineering, we conducted an in depth analysis of the mechanisms behind the far-red fluorescence of AQ143, which is a tetrameric RFP that was designed from the chromoprotein aeCP597 (56, 57). We solved a crystal structure of AQ143, which demonstrates novel red-shifting chromophore interactions and confirms

some other hypothesized interactions. The far-red chromophore environment helps to expand our view of RFP core design and from this work we began to think about RFP cores as unique environments that are separate from the surface of the protein.

In beginning the design, the protein surface seemed an easier place to start than the chromophore environment, and so we created a core-stabilized RFP variant that provided us a more facile test of our design methodologies. We successfully designed an RFP surface, succeeding in repeating the monomerization of DsRed for the third time. In this case, however, we made use of a pre-optimized fluorescent core found in the monomeric derivative, mCherry. These core positions allow us to use computational protein design to monomerize the protein, moving directly from a tetramer to a monomer without any directed evolution or the screening of large libraries of variants. In fact our computationally designed library contains 95 members, 97% of which are fluorescent and monomeric. However, this was only accomplished in the context of an RFP that had previously been monomerized, which afforded us a fluorescent core that was optimized in the context of a monomeric scaffold. The next step was to apply this technique to a previously unmonomerized RFP.

We chose to work with HcRed (11) because it was reported to have a far-red fluorescent emission peak of 645 nm, and a broken AB oligomeric interface. As we will discuss later, there is a systemic problem in the FP engineering field of inaccurately reported fluorescent properties, so as it turns out neither of these attributes were exactly as advertised. HcRed's fluorescent emission turned out to be closer to 633 nm, and it was dimeric, but with significant tetrameric tendency visible by SEC and in the fact that it crystallizes as a tetramer. As HcRed still exhibited relatively far-red fluorescent emission and had not previously been monomerized, we continued and took some of the lessons learned in our repeated monomerization of DsRed and tried to apply them to the monomerization of HcRed. We wanted to divide the engineering process into distinct steps, each of which could be targeted by small to medium-sized libraries of  $10^2$ - $10^5$  variants, which we felt would allow the entire diversity of the library, or at least a significant proportion of it, to be accessed with medium-throughput screening techniques (expression in 96-well plates of

individual clones and assay by plate-reader). For each of these distinct steps, we determined a specific attribute to target, and then identified groups of residues that we felt rationally played a role in that attribute of the designed protein, whether it was oligomericity, stability, fluorescence, or solubility. We hoped to access large but targeted areas of sequence space that would be enriched with desired variants.

We divided the task of engineering HcRed into distinct steps. The first step was core design, in which we attempted to build diversity into the HcRed core by targeting key structural regions with a library that was shaped by the amino acid diversity of engineered far-red FP variants (32, 44, 57, 58). We then took a group of fluorescent diversified core variants and perturbed them with a partial disruption of their AC dimeric interface. We isolated a variant that remained fluorescent after a partial tail deletion, and was both brighter and bathochromically shifted from wild-type HcRed. The mechanism of the brighter and far-red fluorescence in the core-optimized variant (HcRed7) was that an introduced tyrosine formed a  $\pi$ -stacking interaction with the chromophore phenolate. This optimization was not sufficient, however, as HcRed7 error prone mutagenesis was needed to recover dim fluorescence both after a full C-terminal tail deletion and again after completing the monomerization. We eventually arrived at monomeric variants mGinger0.1 and mGinger0.2, which are the furthest red-shifted first generation monomeric RFPs reported to date. The design process was improved in a number of ways over previous RFP monomerizations. First, we managed to break the process into segments, each of which focused on one particular attribute of the design, and made use of a mutational strategy that fit the particular goal. Second, we maintained a baseline level of fluorescence throughout the process, not allowing the fluorescent emission to shift hypsochromically, and obviating the need for mutations to the chromophore environment to restore fluorescence – no core residue within 6.0 Å of the chromophore was mutated. Finally, the design procedure only involved two rounds of error prone mutagenesis, far less than other RFP monomerizations, and these rounds were supplemented with smaller, targeted libraries. mGinger0.1 and mGinger0.2 occupy a place among the brightest far-red monomers, and there is opportunity to further optimize their fluorescence.

Finally, prior studies of RFPs have inconsistently measured a variety of their spectroscopic properties. There are often large discrepancies between values reported in the literature, and we have encountered many incorrectly reported values in this work. Oligomerization, which is a principal focus of this work, is one of the most incompletely characterized properties of RFPs, which is troubling, as it is also one of the most important to determine prior to using an RFP in a biological context. We express and characterize a group of important RFPs, finding that many are improperly claimed to be monomeric, and propose standardized measurement techniques that we use to assay their brightness and determine maximum intensity fluorescent excitation and emission wavelengths. In characterizing this array of RFPs, we present a dataset that can be queried for relationships between various spectroscopic properties, and their structural basis. We remark upon a stark correlation between quantum yield and thermal stability among RFPs.

## 1.5 References

1. Cody CW, Prasher DC, Westler WM, Prendergast FG, & Ward WW (1993) Chemical structure of the hexapeptide chromophore of the *Aequorea* green-fluorescent protein. *Biochemistry* 32(5):1212-1218.
2. Ormö M, *et al.* (1996) Crystal Structure of the *Aequorea victoria* Green Fluorescent Protein. *Science* 273(5280):1392-1395.
3. Prasher DC, Eckenrode VK, Ward WW, Prendergast FG, & Cormier MJ (1992) Primary structure of the *Aequorea victoria* green-fluorescent protein. *Gene* 111(2):229-233.
4. Baulcombe DC, Chapman S, & Cruz S (1995) Jellyfish green fluorescent protein as a reporter for virus infections. *The Plant Journal*.
5. Chalfie M, Tu Y, Euskirchen G, Ward W, & Prasher D (1994) Green fluorescent protein as a marker for gene expression. *Science* 263(5148):802-805.
6. Marshall J, Molloy R, Moss GWJ, Howe JR, & Hughes TE (1995) The jellyfish green fluorescent protein: a new tool for studying ion channel expression and function. *Neuron*.
7. Cormack BP, Valdivia RH, & Falkow S (1996) FACS-optimized mutants of the green fluorescent protein (GFP). *Gene*.
8. Heim R, Prasher DC, & Tsien RY (1994) Wavelength mutations and posttranslational autoxidation of green fluorescent protein. *Proceedings of the National Academy of Sciences of the United States of America* 91(26):12501-12504.
9. Heim R & Tsien RY (1996) Engineering green fluorescent protein for improved brightness, longer wavelengths and fluorescence resonance energy transfer. *Current biology : CB* 6(2):178-182.
10. Fradkov AF, *et al.* (2000) Novel fluorescent protein from *Discosoma* coral and its mutants possesses a unique far-red fluorescence. *FEBS letters* 479(3):127-130.
11. Gurskaya NG, *et al.* (2001) GFP-like chromoproteins as a source of far-red fluorescent proteins1. *FEBS Letters* 507(1):16-20.
12. Matz MV, *et al.* (1999) Fluorescent proteins from nonbioluminescent Anthozoa species. *Nature biotechnology* 17(10):969-973.
13. Mena MA, Treynor TP, Mayo SL, & Daugherty PS (2006) Blue fluorescent proteins with enhanced brightness and photostability from a structurally targeted library. *Nature biotechnology* 24(12):1569-1571.
14. 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.

15. Shcherbo D, *et al.* (2007) Bright far-red fluorescent protein for whole-body imaging. *Nature methods* 4(9):741-746.
16. Subach OM, *et al.* (2008) Conversion of red fluorescent protein into a bright blue probe. *Chemistry & biology* 15(10):1116-1124.
17. Tomosugi W, *et al.* (2009) An ultramarine fluorescent protein with increased photostability and pH insensitivity. *Nature methods* 6(5):351-353.
18. Wang L, Jackson WC, Steinbach PA, & Tsien RY (2004) Evolution of new nonantibody proteins via iterative somatic hypermutation. *Proceedings of the National Academy of Sciences of the United States of America* 101(48):16745-16749.
19. Arosio D, *et al.* (2007) Spectroscopic and structural study of proton and halide ion cooperative binding to gfp. *Biophysical journal* 93(1):232-244.
20. Gurskaya NG, *et al.* (2006) Engineering of a monomeric green-to-red photoactivatable fluorescent protein induced by blue light. *Nature biotechnology* 24(4):461-465.
21. Patterson GH & Lippincott-Schwartz J (2002) A photoactivatable GFP for selective photolabeling of proteins and cells. *Science (New York, N.Y.)* 297(5588):1873-1877.
22. Wachter RM & Remington SJ (1999) Sensitivity of the yellow variant of green fluorescent protein to halides and nitrate. *Current biology : CB* 9(17):9.
23. Chen TW, Wardill TJ, Sun Y, Pulver SR, & Renninger SL (2013) Ultrasensitive fluorescent proteins for imaging neuronal activity. *Nature*.
24. Heim N & Griesbeck O (2004) Genetically encoded indicators of cellular calcium dynamics based on troponin C and green fluorescent protein. *Journal of biological chemistry*.
25. Piatkevich KD, *et al.* (2010) Monomeric red fluorescent proteins with a large Stokes shift. *Proceedings of the National Academy of Sciences of the United States of America* 107(12):5369-5374.
26. Subach OM, *et al.* (2011) A photoswitchable orange-to-far-red fluorescent protein, PSmOrange. *Nature methods* 8(9):771-777.
27. Yang J, *et al.* (2012) mBeRFP, an Improved Large Stokes Shift Red Fluorescent Protein. *PloS one* 8(6).
28. Tromberg BJ, *et al.* (1999) Non-invasive in vivo characterization of breast tumors using photon migration spectroscopy. *Neoplasia (New York, N.Y.)* 2(1-2):26-40.
29. Weissleder R (2001) A clearer vision for in vivo imaging. *Nature biotechnology* 19(4):316-317.
30. Yu D, *et al.* (2013) An improved monomeric infrared fluorescent protein for neuronal and tumour brain imaging. *Nature communications* 5:3626.

31. Filonov GS, *et al.* (2011) Bright and stable near-infrared fluorescent protein for in vivo imaging. *Nature biotechnology* 29(8):757-761.
32. Kiryl DP, Fedor VS, & Vladislav VV (2013) Far-red light photoactivatable near-infrared fluorescent proteins engineered from a bacterial phytochrome. *Nature Communications*.
33. Shcherbo D, *et al.* (2010) Near-infrared fluorescent proteins. *Nature methods* 7(10):827-829.
34. Flytzanis NC, Bedbrook CN, & Chiu H (2014) Archaeorhodopsin variants with enhanced voltage-sensitive fluorescence in mammalian and *Caenorhabditis elegans* neurons. *Nature* ....
35. McIsaac SR, *et al.* (2014) Directed evolution of a far-red fluorescent rhodopsin. *Proceedings of the National Academy of Sciences* 111(36):13034-13039.
36. Alieva NO, *et al.* (2007) Diversity and evolution of coral fluorescent proteins. *PloS one* 3(7).
37. Labas YA, *et al.* (2002) Diversity and evolution of the green fluorescent protein family. *Proceedings of the National Academy of Sciences of the United States of America* 99(7):4256-4261.
38. Meyers M, Porter JW, & Wares JP (2012) Genetic diversity of fluorescent proteins in Caribbean agariciid corals. *The Journal of heredity* 104(4):572-577.
39. Wang W, Li G-W, Chen C, Xie SX, & Zhuang X (2011) Chromosome organization by a nucleoid-associated protein in live bacteria. *Science* 333(6048):1445-1449.
40. Siyuan W, Jeffrey RM, Graham TD, Sunney XX, & Xiaowei Z (2014) Characterization and development of photoactivatable fluorescent proteins for single-molecule-based superresolution imaging. *Proceedings of the National Academy of Sciences*.
41. Campbell RE, *et al.* (2002) A monomeric red fluorescent protein. *Proceedings of the National Academy of Sciences*.
42. Ma W, M S, & R R (2000) The structural basis for red fluorescence in the tetrameric GFP homolog DsRed. *Nature structural biology* 7(12):1133-1138.
43. Michael ZL, *et al.* (2009) Autofluorescent Proteins with Excitation in the Optical Window for Intravital Imaging in Mammals. *Chemistry & Biology*.
44. Chica RA, Moore MM, Allen BD, & Mayo SL (2010) Generation of longer emission wavelength red fluorescent proteins using computationally designed libraries. *Proceedings of the National Academy of Sciences of the United States of America* 107(47):20257-20262.
45. Strongin DE, *et al.* (2007) Structural rearrangements near the chromophore influence the maturation speed and brightness of DsRed variants. *Protein engineering, design & selection : PEDS* 20(11):525-534.

46. Shemiakina II, *et al.* (2012) A monomeric red fluorescent protein with low cytotoxicity. *Nature communications*.
47. Kredel S, *et al.* (2008) mRuby, a bright monomeric red fluorescent protein for labeling of subcellular structures. *PloS one* 4(2).
48. 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.
49. Moore MM, Oteng-Pabi SK, Pandelieva AT, Mayo SL, & Chica RA (2011) Recovery of Red Fluorescent Protein Chromophore Maturation Deficiency through Rational Design. *PloS one* 7(12).
50. Hasegawa J-yY, *et al.* (2010) Excited states of fluorescent proteins, mKO and DsRed: chromophore-protein electrostatic interaction behind the color variations. *The journal of physical chemistry. B* 114(8):2971-2979.
51. Sun Q, Doerr M, Li Z, Smith SC, & Thiel W (2010) QM/MM studies of structural and energetic properties of the far-red fluorescent protein HcRed. *Physical chemistry chemical physics : PCCP* 12(10):2450-2458.
52. Dalby PA (2011) Strategy and success for the directed evolution of enzymes. *Current opinion in structural biology*.
53. Privett HK, *et al.* (2012) Iterative approach to computational enzyme design. *Proceedings of the National Academy of Sciences of the United States of America* 109(10):3790-3795.
54. Parra LP, Agudo R, & Reetz MT (2013) Directed Evolution by Using Iterative Saturation Mutagenesis Based on Multiresidue Sites. *ChemBioChem*.
55. Farinas ET, Bulter T, & Arnold FH (2001) Directed enzyme evolution. *Current opinion in biotechnology*.
56. Shkrob MA, *et al.* (2005) Far-red fluorescent proteins evolved from a blue chromoprotein from *Actinia equina*. *Biochemical Journal*.
57. Wannier T & Mayo SL (2014) The structure of a far-red fluorescent protein, AQ143, shows evidence in support of reported red-shifting chromophore interactions. *Protein Science* 23(8):1148-1153.
58. Pletnev S, *et al.* (2012) Structural basis for bathochromic shift of fluorescence in far-red fluorescent proteins eqFP650 and eqFP670. *Acta crystallographica. Section D, Biological crystallography* 68(Pt 9):1088-1097.

## CHAPTER 2

### **The Structure of a Far-Red Fluorescent Protein, AQ143, Shows Evidence in Support of Reported Red-Shifting Chromophore Interactions.**

Wannier, Timothy M.<sup>a</sup> and Mayo, Stephen L.<sup>a</sup>

<sup>a</sup> *Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, CA 91125*

*(This work appeared as a structure report in 2014 in the journal Protein Science)*

#### **2.1 Abstract**

Engineering fluorescent proteins (FPs) to emit light at longer wavelengths is a significant focus in the development of the next generation of fluorescent biomarkers, as far-red light penetrates tissue with minimal absorption, allowing better imaging inside of biological hosts. Structure-guided design and directed evolution have led to the discovery of red FPs with significant bathochromic shifts to their emission. Here, we present the crystal structure of one of the most bathochromically shifted FPs reported to date, AQ143, a nine-point mutant of aeCP597, a chromoprotein from *Actinia equina*. The 2.19 Å resolution structure reveals several important chromophore interactions that contribute to the protein's far-red emission and shows dual occupancy of the green and red chromophores.

## 2.2 Introduction

Fluorescent proteins (FPs) that emit light in the near-infrared (NIR) window (~650-900 nm) are in demand as biological imaging agents. The NIR window is a local minimum at which light penetrates tissue with minimal absorption from biological molecules such as melanin, hemoglobin, and water.<sup>1</sup> FPs natively do not emit light in the NIR; the longest maximum intensity emission wavelength ( $\lambda_{em}$ ) reported to date for a native red FP (RFP) is 613 nm, found in NvFP-7R from *Nematostella vectensis*.<sup>2</sup> FPs with significant bathochromic shifts to  $\lambda_{em}$  have been produced with both rational design and directed evolution but these molecules tend to have low quantum yields, poor brightness and other characteristics that compromise their utility.<sup>3-8</sup> Many FP engineering strategies, including those that have induced bathochromic shifts in the  $\lambda_{em}$ , have relied on atomic-resolution structural data to guide intuition-based design, motivating continued efforts to obtain additional structural information for far-red FPs. AQ143, which was engineered from aeCP597, a chromoprotein from *Actinia equine*,<sup>9</sup> is one of only seven known FPs of the *Aequorea victoria* FP-like superfamily that exhibit a peak emission wavelength of at least 650 nm. The other five proteins (Neptune,<sup>10</sup> eqFP650,<sup>7</sup> TagRFP657,<sup>11</sup> mCardinal<sup>12</sup>, eqFP670,<sup>7</sup> and TagRFP675<sup>8</sup>) are all variants of eqFP578, a native RFP from *Entacmaea quadricolor*.<sup>13</sup> There are known structures for five of these proteins (Neptune: 3IP2, eqFP650: 4EDO, mCardinal: 4OQW, eqFP670: 4EDS, TagRFP675: 4KGF), but as they are all derived from the same ancestral protein, there is limited sequence diversity among these structures. Here we report the 2.19 Å crystal structure of AQ143, which is derived from a more distantly related protein, aeCP597 (~60% sequence identity to eqFP578 and its variants). AQ143 has a novel chromophore environment (defined as all internal-facing residues within 5 Å of the chromophore), which shares no more than 70% (16 of 23 positions) sequence identity with any other red fluorescent protein. Glu41 plays an important role in red-shifting AQ143's emission spectrum and is not seen in any other fluorescent protein. The reported structure also provides evidence in support of recently reported red-shifting chromophore interactions.<sup>6,8,14</sup>

## 2.3 Results and Discussion

The asymmetric unit contains eight protein molecules, which align with an all-atom *r.m.s.d.* of 0.27–0.74 Å with differences between molecules concentrated mostly in the loop regions and in the C- and N-terminal tails. The chromophore region also varies somewhat between molecules and shows weak electron density around the phenolate side chain, which could be attributed to mobility in the phenolate side chain and co-occupancy of two different chromophores -- green and red.

### *Oligomerization*

AQ143 is a native tetramer, which is clear in the crystal packing. The asymmetric unit, however, contains eight monomers, or two such tetrameric assemblies with the C-terminal tail of each monomer involved in making inter-tetramer contacts. To verify the oligomerization state of AQ143, we ran both size exclusion chromatography (SEC) and analytical ultracentrifugation (AUC). SEC analysis indicates that AQ143 behaves as a tetramer, but that it has slight octomeric properties, while AUC confirms that the protein is predominantly tetrameric (Figs. S1 and S2). As oligomerization is an important consideration in the engineering of RFPs, all of which are natively tetrameric,<sup>15</sup> we compared the AB and AC interfaces of AQ143 with those of four other native RFPs (Table S1) using the PISA server of the European Bioinformatics Institute<sup>16</sup> and report average buried surface area and average  $\Delta^iG$  (the solvation free energy gain upon formation of the interface). The AC interface is known to be the tighter of the two interfaces,<sup>17</sup> which is consistent with the AC interface having more negative  $\Delta^iGs$ . Interestingly, although AQ143 showed a similar amount of buried AB interface surface area, the  $\Delta^iG$  for AQ143's AB interface is very high, indicating a large amount of hydrophobic residues at this interface. The AC interface is more difficult to compare as the amount of buried surface area varies widely, although this is in part due to the lack of crystallographic density at the C-terminal tails (which participate intimately in this interface) in many of these structures. DsRed, the first successfully monomerized RFP<sup>17</sup> shows the lowest  $\Delta^iG$  for its AC interface, possibly indicating that future monomerization efforts of AQ143 may be more difficult.

### ***Green and red chromophores***

Many engineered far-red FPs exhibit slow or incomplete maturation to the red chromophore,<sup>4,11,18</sup> and it has recently been shown that maturation to the green and red chromophores in DsRed-type FPs occurs via a branched pathway (i.e., the two forms of the chromophore are separate endpoints in chromophore maturation; the green is not an intermediate in the maturation to the red chromophore as had been previously proposed).<sup>19</sup> AQ143 is a DsRed-type FP with a chromophore composed of a methionine/tyrosine/glycine triad (MYG) that matures to both a green and a red chromophore (Fig. 1), as evidenced by its absorbance, excitation, and emission spectra (Figs. S3, S4, and S5). To calculate the percentage of chromophores that mature to the green and to the red, we determined the extinction coefficients of the two species by the dynamic difference method. In this procedure, AQ143 was pH-adjusted to alkaline conditions, in which the green and red chromophores denature at different rates and their respective contributions to the 450 nm alkali-denatured absorbance peak can be determined (Figs. S6, S7, S8 and Supplementary Methods).<sup>5</sup> We calculated the extinction coefficient to be  $58,000 \pm 11,000 \text{ M}^{-1}\text{cm}^{-1}$  for the red chromophore and  $47,000 \pm 5,000 \text{ M}^{-1}\text{cm}^{-1}$  for the green chromophore. From these data, we estimated the percentages in the fully mature protein to be  $33 \pm 6\%$  for the red and  $67 \pm 6\%$  for the green chromophore. Measurements of the protein in the crystal condition suggested that this fraction did not change upon crystallization. Corroborating the spectroscopic evidence, we observed that the refined electron density map of AQ143 shows a mixture of chromophores containing both the oxidized N-acylimine (red) and the unoxidized N-acylamine (green) at the N-terminal residue of the chromophoric triad. The estimated occupancy of the red and green chromophores averaged across all eight monomers in the asymmetric unit is  $24 \pm 9\%$  and  $76\%$ , respectively. Thus the spectroscopic calculations of chromophore occupancy in the crystal condition are consistent with the crystallographic refinement.

### *Cis vs. trans phenolate*

The phenolate side chain of the chromophore (the phenolate group) in DsRed-type FPs and related chromoproteins can occupy either a *cis* or a *trans* conformation, indicating its proximity to the N1 nitrogen of the imidazolinone ring of the chromophore. For many RFPs, a *trans* to *cis* isomerization of this phenolate moiety, which is sometimes pH-inducible,<sup>20,21</sup> has been implicated in fluorescence. In non-fluorescent chromoproteins, for instance, the chromophore is found in the *trans* conformation, and mutations to these chromoproteins that stabilize the *cis* conformation have created FPs such as HcRed and AQ143. In engineering AQ143 from the chromoprotein aeCP597, Cys143Ser was reported to be responsible for inducing weak fluorescence,<sup>9</sup> as the mutation to serine stabilizes the *cis* chromophore by providing a hydrogen bond to the hydroxyl oxygen of the phenolate side chain. In the referenced work, fluorescence was improved by removing a serine hydrogen bond to the hydroxyl of the *trans* phenolate with a Ser158Ala mutation, further stabilizing the *cis* over the *trans* chromophore. By inducing fluorescence in an otherwise non-fluorescent chromoprotein, these mutations seem to imply that the *cis* chromophore represents the fluorescent moiety in AQ143.

Indeed, the refined structure shows good electron density for all parts of the chromophore with the exception of the phenolate side chain, which we modeled in the *trans* configuration. However, the difference map shows that the modeled phenolate is not a perfect fit, as the electron density is not sufficient to describe a chromophore that is solely found in the modeled *trans* configuration, while residual density appears in the position we expect that the *cis* phenolate would occupy. The refined electron density is such that we expect there is a co-occupancy in the crystal of two or more chromophore orientations and also possibly that the phenolate is mobile in one or both of these chromophore species. This would be consistent with a *cis-trans* isomerization of the chromophore upon fluorescence excitation, as has been seen in other FPs,<sup>5,22,23</sup>. The lack of clear electron density for the phenolate moiety implies that the fluorescence-inducing mutations in AQ143 may have had their predicted effects, namely in destabilizing the native *trans* chromophore, and allowing for the phenolate to occupy the *cis* conformation. Given the ambiguity associated with the

chromophore orientation and the lack of clear density for the *cis* conformation, we elected to model-build the *cis* phenolate post-refinement (Fig. 1). The modeled position of the *cis* phenolate accommodates a hydrogen bond between the hydroxyl of the fluorescence-inducing Cys143Ser mutation and the phenolate oxygen, supporting the hypothesis that this interaction is linked to the induction of fluorescence in AQ143 (Fig. 2). A second water-mediated hydrogen bond to the phenolate oxygen appears to further stabilize the *cis* conformation.

Interestingly, neither the *trans* nor the modeled *cis* conformations of the chromophore are coplanar with the imidazoline ring. This non-coplanarity is relatively uncommon in FPs and has been proposed to be responsible for low quantum yields.<sup>10</sup> AQ143 indeed has a very low quantum yield (0.04)<sup>9</sup>, and improving the coplanarity of the two chromophore rings may represent an opportunity to further improve its fluorescence.

### ***Mechanisms of bathochromic shift***

AQ143 exhibits a number of red-shifting chromophore interactions that have been well documented in the literature.<sup>8</sup> A network of direct and water-mediated hydrogen bonds has been proposed to lower the energy of the photoexcited state of the chromophore's conjugated  $\pi$ -electron system, resulting in bathochromic shifts to  $\lambda_{em}$ .<sup>10,24</sup> In AQ143, three hydrogen bonds to the chromophore are good candidates to provide such stabilization including two hydrogen bonds to the acylimine oxygen, as well as one to the phenolate oxygen (Fig. 2).

Glu41 and a water molecule coordinated by Gln106 and the chromophore's C-terminal acyl oxygen both form hydrogen bonds to the chromophore N-acylimine (Fig. 2). To our knowledge, the only other FPs known to have two hydrogen bonds to the acylimine oxygen are CjBlue, the furthest red-shifted chromoprotein, and TagRFP675, the furthest red-shifted FP, although in TagRFP675, the hydrogen bond donor at the position equivalent to Glu41 in AQ143 is a glutamine.<sup>8,14</sup> mPlum, the furthest red-shifted monomeric FP,<sup>3</sup> has a similar hydrogen bonding interaction between Glu16 and the chromophore N-acylimine, but is lacking a coordinated water molecule to provide the second hydrogen bond. The

importance of hydrogen bonds to the N-acylimine was shown in mPlum variants, in which Glu16 is mutated to other residues including proline and glutamine, causing significant hypsochromic shifts to  $\lambda_{em}$ .<sup>18,25,26</sup>

Additionally, flexibility in the hydrogen bonding network to the phenolate oxygen of the chromophore, particularly via water-mediated hydrogen bonds, has been proposed to be responsible for extended Stokes' shifts and significant bathochromic shifts to fluorescence emission.<sup>8</sup> The modeled *cis* chromophore, which we believe to be the fluorescent moiety, can accommodate two hydrogen bonds to the phenolate oxygen from the hydroxyl of Ser143 and a structural water molecule (Fig. 2). The *trans* chromophore, despite the mutation away from Ser158, makes a hydrogen bond contact with a structural water molecule stabilized by Glu145 and Thr176, although the effect of this interaction is less clear as the *trans* chromophore is not thought to be fluorescent.

Finally, many red-shifted fluorescent proteins have been described that exhibit  $\pi$ -stacking interactions with the phenolate group of the chromophore.<sup>6,27,28</sup> Histidine and tyrosine have both been reported at positions analogous to His197 in AQ143 with histidine present in eqFP578, RFP639, and mRuby<sup>13,29,30</sup>, and tyrosine present in mRojoA, TagRFP657, and mGrape3<sup>6,10,11</sup>. In engineering mRojoA, a tyrosine  $\pi$ -stacking interaction with the *cis* phenolate was explicitly designed into the protein which resulted in a 7nm red-shift<sup>6</sup>. In AQ143, His197 appears to form a  $\pi$ -stacking interaction with the *trans* phenolate (Fig. 2), which we presume to be the non-fluorescent entity. Interestingly, in mRuby and eqFP578, the histidine also  $\pi$ -stacks with the *trans* phenolate, whereas in the further red-shifted RFP639, the  $\pi$ -stack occurs with the *cis* phenolate. This implies that there may be room to further stabilize the photo-excited state of the *cis* phenolate of AQ143 and red-shift its emission by optimizing the  $\pi$ -stacking interaction with the *cis* chromophore.

## 2.4 Conclusion

AQ143 is one of the furthest red-emitting FPs of the GFP family, and the structure reported in this study helps elucidate some of the features underlying its far-red emission. A recently reported FP, TagRFP675, shares many of the same chromophore interactions responsible for AQ143's bathochromic shift.<sup>8</sup>

## 2.5 Materials and Methods

### *Protein expression and purification*

A synthetic gene construct encoding an N-terminal poly-histidine tagged AQ143 (GenBank KF479351) was assembled *in vitro*, expressed in *Escherichia coli* BL21(DE3) cells, purified, and crystallized. Cultures were grown at 37°C to an optical density of ~0.6 in LB, induced, then allowed to express protein at 20°C for 24 hours. Protein was purified via His-tag affinity chromatography, run over a size exclusion column to remove trace contaminants and move the protein into storage buffer (1 × PBS pH 7.4), and finally concentrated to 18 mg/ml.

### *Crystallization, data collection, and structure determination*

Rectangular plate crystals grew in 7 days by the sitting-drop vapor diffusion method in 100 mM Tris pH 7.0 with 50 mM lithium sulfate and 20% w/v PEG 3350. Crystals were flash frozen in 2-Methyl-2,4-pentanediol (MPD) and shipped to beamline 12-2 at the Stanford Synchrotron Radiation Lightsource, where a 2.19 Å data set was collected. Phases were obtained through molecular replacement using the crystal structure of the FP asFP595 (PDB ID 1A50).

Following molecular replacement, model building and refinement were run with COOT and PHENIX.<sup>31,32</sup> NCS restraints were applied to early refinement steps and removed at the final stages of refinement. TLS parameters were used throughout. The chromophore was initially left out of the refinement and added at a later stage when clear density became

evident for it. First the chromophore was added without the phenolate side chain, as little density appeared for this group. Subsequently, as density became clearer, a *trans* chromophore was added. The final modeled chromophore has a *trans* phenolate ring, an imidazoline heterocyclic ring, and dual occupancy of a green N-acylamine and a red N-acylimine. Coordinates were deposited in the Protein Data Bank with the code 4OHS. Data collection and refinement statistics are listed in Table 1.

### ***Modeling the cis chromophore post refinement***

We modeled the *cis* chromophore after refining the structure because there was poor density for this conformation. There was, however, residual density in the region we expected the *cis* chromophore to be. We introduced the alternate conformation in COOT, fit it to the residual density, and ran the model through several rounds of PHENIX refinement, which resulted in the modeled positions shown in figures 1 and 2 in turquoise.

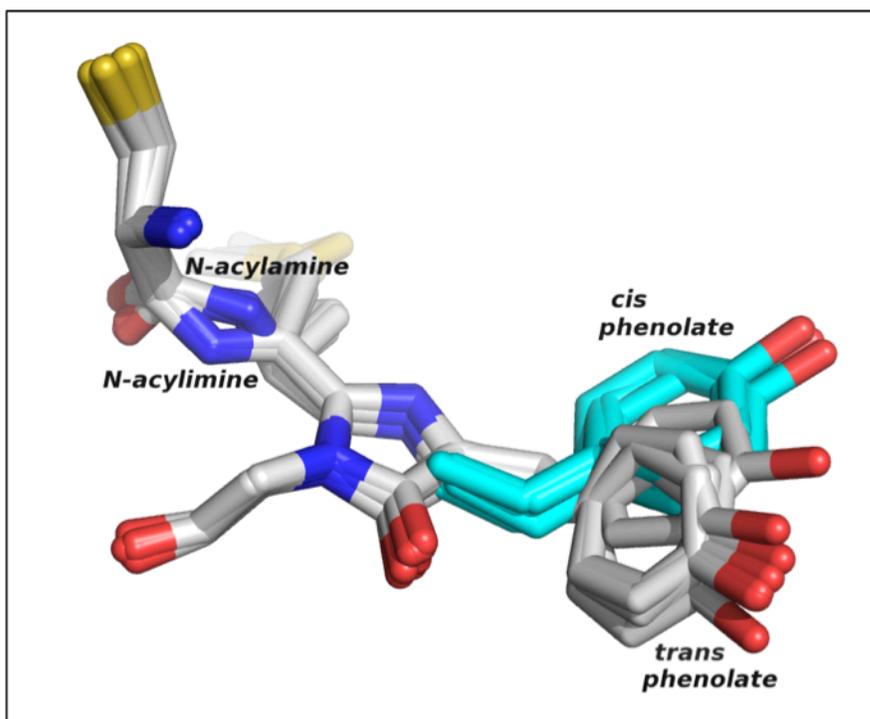
## **2.6 Acknowledgments**

The authors are grateful for the use of beamline 12-2 at the Stanford Synchrotron Radiation Lightsource (SSRL) in Menlo Park, CA, operated by Stanford University and supported by the Department of Energy and the National Institutes of Health. They additionally are thankful to Jens Kaiser and Pavle Nikolovski at the California Institute of Technology for helpful discussions. Finally they thank the Gordon and Betty Moore Foundation, the Beckman Institute, and the Sanofi-Aventis Bioengineering Research Program for support of the Molecular Observatory at the California Institute of Technology.

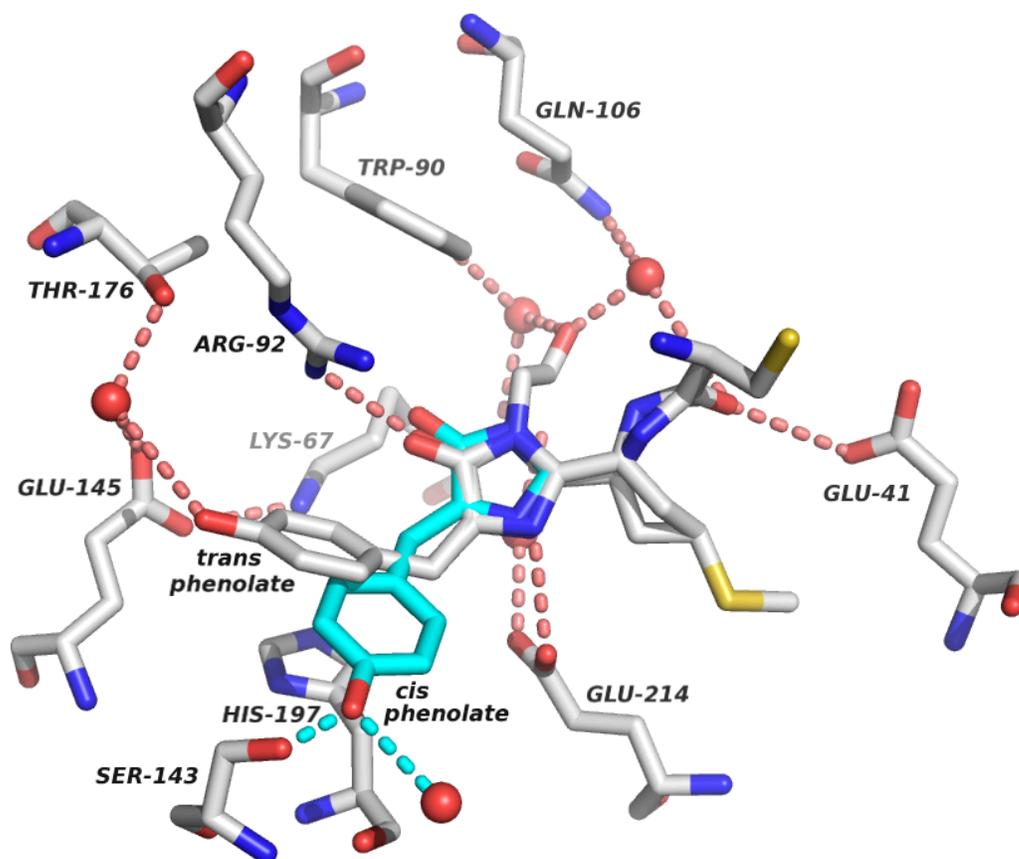
## 2.7 Tables and Figures:

**Table I.** X-ray data reduction and crystallographic refinement statistics

(A) X-ray data reduction statistics	
Space group	P1
Unit cell dimensions ( <i>a</i> , <i>b</i> , <i>c</i> )	51.0 Å, 68.1 Å, 132.8 Å
Resolution	39.1 Å – 2.19 Å
(last shell)	2.31 Å – 2.19 Å
Total measurements (last shell)	281,018 (30,290)
Number of unique reflections (last shell)	72,946 (8,028)
Wavelength	
<i>R</i> -merge (last shell)	0.072 (0.749)
<i>I</i> / $\sigma$ ( <i>I</i> ) (last shell)	11.9 (1.7)
Completeness (last shell)	0.861 (0.648)
Multiplicity (last shell)	3.9 (3.8)
(B) Crystallographic refinement statistics	
Resolution	131.1 Å - 2.19 Å
(last shell)	2.22 Å – 2.19 Å
No. of reflections (working set)	69,234
No. of reflections (test set)	3,647
<i>R</i> -factor (last shell)	0.190 (0.315)
<i>R</i> -free (last shell)	0.221 (0.338)
No. of amino acid residues	1,770
No. of atoms	14,508
No. of solvent molecules	355
Average <i>B</i> -factor	
Protein	62.5 Å <sup>2</sup>
Solvent	49.6 Å <sup>2</sup>
<i>R.m.s.d.</i> from ideal geometry	
Bond lengths	0.006 Å
Bond angles	0.987°



**Figure 1.** Alignment of the chromophores and C-terminal cysteine from each of the eight monomers in the asymmetric unit. The modeled *cis* phenolate is shown in turquoise. The N-acylamine and N-acylimine are present in the green and red chromophores respectively.



**Figure 2.** Chromophore contacts in AQ143. Residues that directly interact with the chromophore or help to coordinate structural waters (red spheres) are shown along with the immediate hydrogen-bonding network. A representative chromophore was chosen (chain E) to illustrate the contacts. Hydrogen bonds (dotted lines) are shown for interactions with the chromophore. The modeled *cis* conformation is shown in turquoise, along with two putative hydrogen bonds to its hydroxyl group. Two hydrogen bonds to the acylimine oxygen from Glu41 and a coordinated water can be seen in the right of the figure.

## 2.8 Supplementary Tables and Figures:

**Table S1.** Surface area analysis of oligomeric interfaces

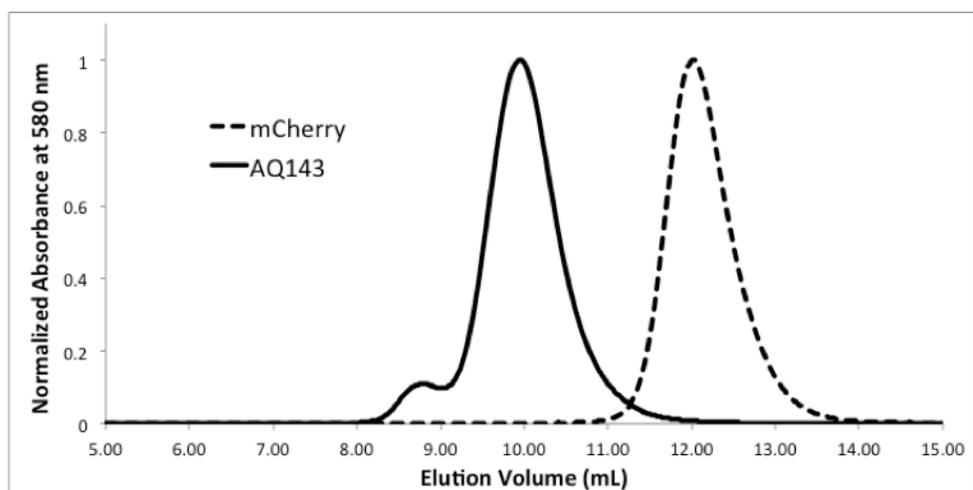
Protein	PDB ID	AB Interface		AC Interface	
		Average buried surface area (Å <sup>2</sup> )	$\Delta^iG$ (kcal/mol)*	Average buried surface area (Å <sup>2</sup> )	$\Delta^iG$ (kcal/mol)*
AQ143	4OHS	979.6	-11.2	1203.9	-13.8
KFP <sup>†</sup>	2A50	977.7	-6.4	1691.4	-19.2
DsRed	1ZGO	993.9	-8.6	1326.8	-7.2
eqFP578	3PIB	1043.1	-2.2	1564.6	-20.8
eqFP611 <sup>†</sup>	1UIS	988.1	-0.5	1185.6	-17.3

\*  $\Delta^iG$  indicates the solvation free energy gain upon formation of the interface.

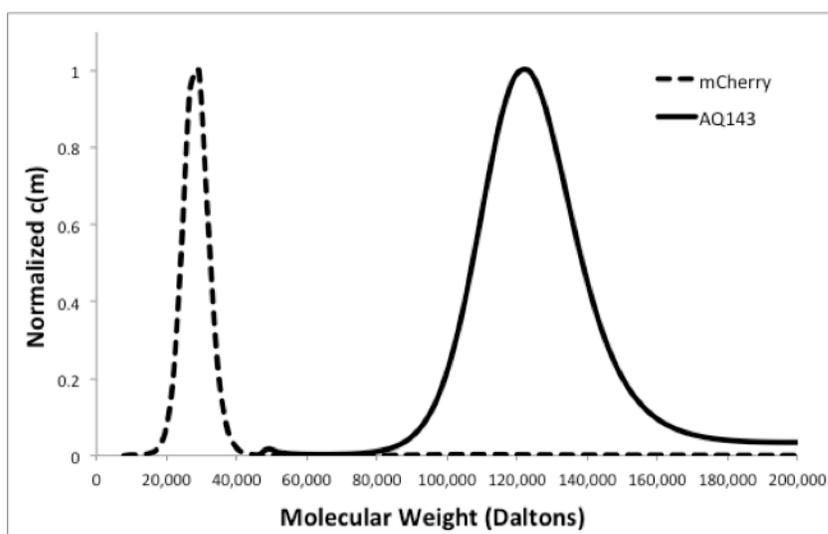
<sup>†</sup> Only two monomers were present in the asymmetric unit, so symmetry mates were generated to visualize the tetramer.

**Table S1.** AQ143 is compared to four other native RFPs (all tetramers). The surface area and solvation free energy of the AB and AC interfaces were evaluated with the PISA server from the European Bioinformatics Institute.

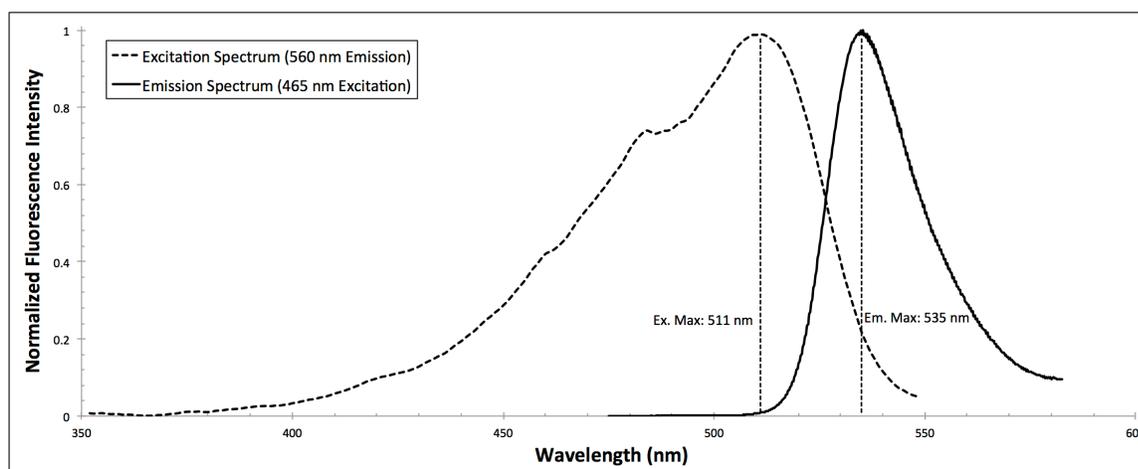
[http://www.ebi.ac.uk/pdbe/prot\\_int/pistart.html](http://www.ebi.ac.uk/pdbe/prot_int/pistart.html)



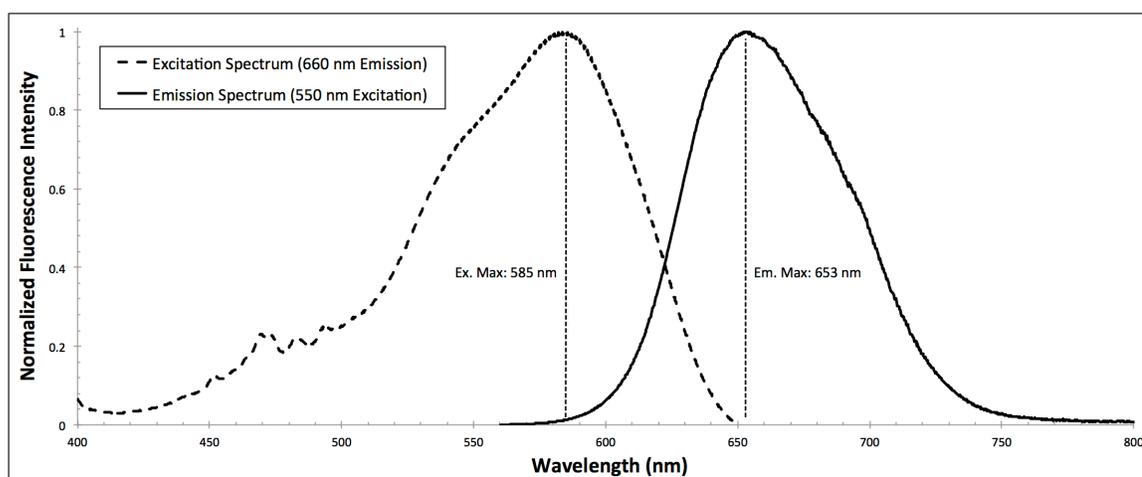
**Figure S1.** SEC traces of mCherry and AQ143. 100  $\mu\text{L}$  of a 60  $\mu\text{M}$  aliquot of each protein was run through a Superdex 75 column. mCherry is a monomer, whereas AQ143 appears to be a tetramer by AUC (Fig. S2), but is an octomeric assembly in crystal packing and demonstrates a clear octomeric peak by SEC.



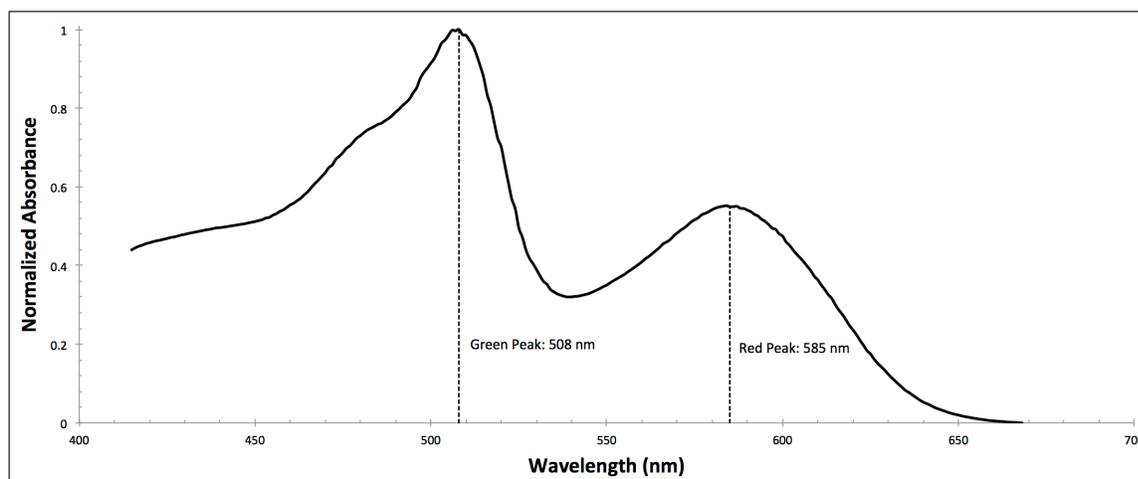
**Figure S2.** Sedimentation velocity analysis of mCherry and AQ143. mCherry, a monomer, sediments at its molecular weight, which is 27 kD. AQ143 sediments at 108 kD, very near its tetrameric weight.



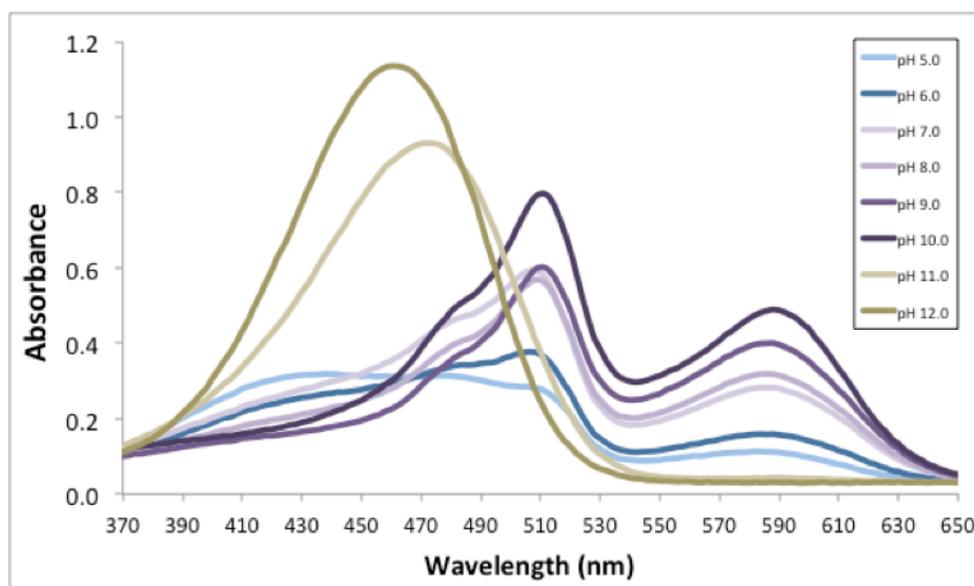
**Figure S3.** Excitation and emission spectra of the green chromophore of AQ1143. Emission was measured with an excitation wavelength of 465 nm, while excitation was measured at an emission wavelength of 560 nm. These spectra were taken in a Photon Technology International fluorometer.



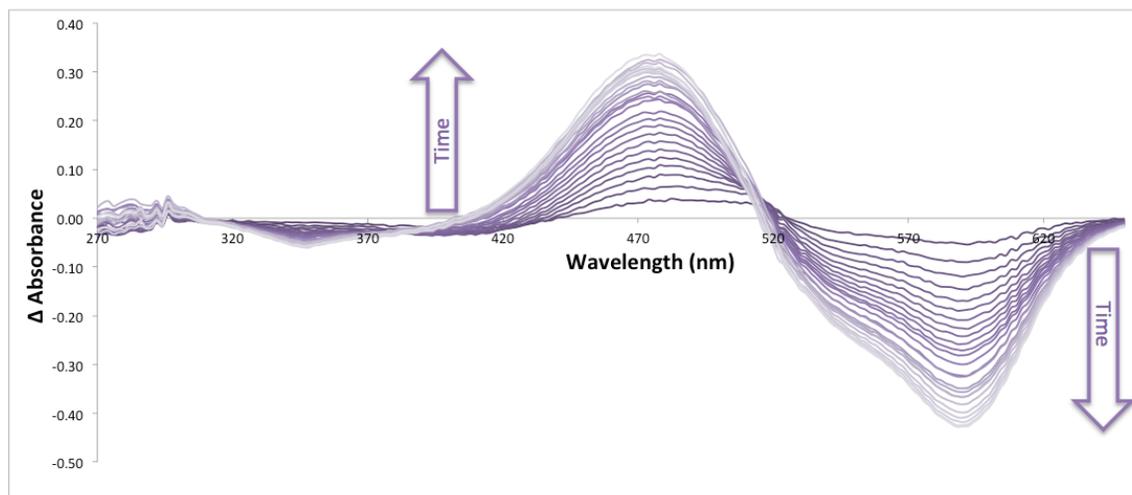
**Figure S4.** Excitation and emission spectra of the red chromophore of AQ1143. Emission was measured with an excitation wavelength of 550 nm, while excitation was measured at an emission wavelength of 660 nm. These spectra were taken in a Photon Technology International fluorometer.



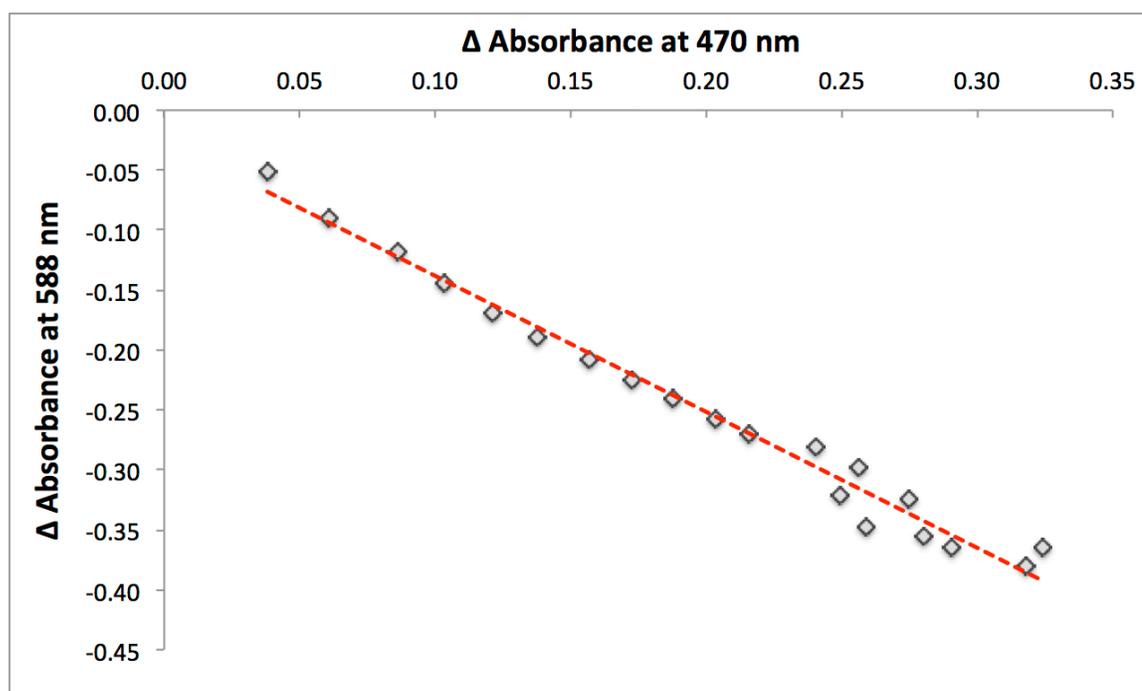
**Figure S5.** Absorbance spectrum of AQ143 taken on a Tecan Safire2 platereader.



**Figure S6.** pH profile of the absorbance of 30  $\mu\text{M}$  AQ143 between pH 5.0 and 12.0. The protein undergoes two distinct transitions. The first transition is between an acid-denatured absorbance profile at which there is little absorbance by either chromophore to the growth of both the green and red absorption peaks (510 nm and 588 nm, respectively). The second transition is a rapid base denaturation of the protein in which both the red and green chromophores are converted into a yellow, 450 nm-absorbing species. The extinction coefficient grows as the pH is increased, topping out near pH 10.0.



**Figure S7.** Time-dependent absorbance of AQ143 diluted 1:100 into alkaline buffer (100mM  $\text{Na}_2\text{HPO}_4$ ; 150mM  $\text{NaCl}$ ; pH 10.5). Time points are measured 5 minutes apart. Early time points are denoted with darker lines while late time points are denoted with lighter lines.



**Figure S8.** The loss in absorbance of the 588 nm (red chromophore) peak correlates linearly to an increase in the 470 nm (base-denatured chromophore) peak.

## 2.9 References

- 1 Tromberg, B., Shah, N., Lanning, R., Cerussi, A., Espinoza, J., Pham, T., Svaasand, L. & Butler, J. (2000) Non-invasive in vivo characterization of breast tumors using photon migration spectroscopy. *Neoplasia (New York, N.Y.)* 2:26-40.
- 2 Ikmi, A. & Gibson, M. (2010) Identification and in vivo characterization of NvFP-7R, a developmentally regulated red fluorescent protein of *Nematostella vectensis*. *PloS one* 5.
- 3 Wang, L., Jackson, W., Steinbach, P. & Tsien, R. (2004) Evolution of new nonantibody proteins via iterative somatic hypermutation. *Proceedings of the National Academy of Sciences of the United States of America* 101:16745-16749.
- 4 Shcherbo, D., Merzlyak, E., Chepurnykh, T., Fradkov, A., Ermakova, G., Solovieva, E., Lukyanov, K., Bogdanova, E., Zaraisky, A., Lukyanov, S. & Chudakov, D. (2007) Bright far-red fluorescent protein for whole-body imaging. *Nature methods* 4:741-746.
- 5 Kredel, S., Nienhaus, K., Oswald, F., Wolff, M., Ivanchenko, S., Cymer, F., Jeromin, A., Michel, F., Spindler, K.-D., Heilker, R., Nienhaus, G. & Wiedenmann, J. (2008) Optimized and far-red-emitting variants of fluorescent protein eqFP611. *Chemistry & biology* 15:224-233.
- 6 Chica, R., Moore, M., Allen, B. & Mayo, S. (2010) Generation of longer emission wavelength red fluorescent proteins using computationally designed libraries. *Proceedings of the National Academy of Sciences of the United States of America* 107:20257-20262.
- 7 Shcherbo, D., Shemiakina, I., Ryabova, A., Luker, K., Schmidt, B., Souslova, E., Gorodnicheva, T., Strukova, L., Shidlovskiy, K., Britanova, O., Zaraisky, A., Lukyanov, K., Loschenov, V., Luker, G. & Chudakov, D. (2010) Near-infrared fluorescent proteins. *Nature methods* 7:827-829.
- 8 Piatkevich, K., Malashkevich, V., Morozova, K., Nemkovich, N., Almo, S. & Verkhusha, V. (2013) Extended Stokes shift in fluorescent proteins: chromophore-protein interactions in a near-infrared TagRFP675 variant. *Scientific reports* 3:1847.
- 9 Shkrob, M., Yanushevich, Y., Chudakov, D., Gurskaya, N., Labas, Y., Poponov, S., Mudrik, N., Lukyanov, S. & Lukyanov, K. (2005) Far-red fluorescent proteins evolved from a blue chromoprotein from *Actinia equina*. *The Biochemical journal* 392:649-654.
- 10 Lin, M., McKeown, M., Ng, H.-L., Aguilera, T., Shaner, N., Campbell, R., Adams, S., Gross, L., Ma, W., Alber, T. & Tsien, R. (2009) Autofluorescent proteins with excitation in the optical window for intravital imaging in mammals. *Chemistry & biology* 16:1169-1179.

- 11 Morozova, K., Piatkevich, K., Gould, T., Zhang, J., Bewersdorf, J. & Verkhusha, V. (2010) Far-red fluorescent protein excitable with red lasers for flow cytometry and superresolution STED nanoscopy. *Biophysical journal* 99:5.
- 12 Jun, C., Russell, D. H., Stéphane, Y. C., Pengpeng, L., Emilio, G.-G., John, S. B., Niloufar, J. A., Amy, J. L., Paula, J. C., Michelle, A. B., Michael, W. D., Ho-Leung, N., Garcia, K. C., Christopher, H. C., Kang, S., Helen, M. B. & Michael, Z. L. (2014) Non-invasive intravital imaging of cellular differentiation with a bright red-excitable fluorescent protein. *Nature Methods*.
- 13 Merzlyak, E., Goedhart, J., Shcherbo, D., Bulina, M., Shcheglov, A., Fradkov, A., Gaintzeva, A., Lukyanov, K., Lukyanov, S., Gadella, T. & Chudakov, D. (2007) Bright monomeric red fluorescent protein with an extended fluorescence lifetime. *Nature methods* 4:555-557.
- 14 Chan, M. C. Y., Karasawa, S., Mizuno, H., Bosanac, I., Ho, D., Prive, G. G., Miyawaki, A. & Ikura, M. (2006) Structural Characterization of a Blue Chromoprotein and Its Yellow Mutant from the Sea Anemone *Cnidopus Japonicus*. *Journal of Biological Chemistry* 281.
- 15 Alieva, N. O., Konzen, K. A., Field, S. F., Meleshkevitch, E. A., Hunt, M. E., Beltran-Ramirez, V., Miller, D. J., Wiedenmann, J., Salih, A. & Matz, M. V. (2007) Diversity and evolution of coral fluorescent proteins. *PloS one* 3.
- 16 Krissinel, E. a. H., K. (2007) Inference of macromolecular assemblies from crystalline state. *Journal of Molecular Biology* 372:774-797.
- 17 Robert, E. C., Oded, T., Amy, E. P., Paul, A. S., Geoffrey, S. B., David, A. Z. & Roger, Y. T. (2002) A monomeric red fluorescent protein. *Proceedings of the National Academy of Sciences*.
- 18 Moore, M., Oteng-Pabi, S., Pandelieva, A., Mayo, S. & Chica, R. (2012) Recovery of Red Fluorescent Protein Chromophore Maturation Deficiency through Rational Design. *PloS one* 7.
- 19 Strack, R., Strongin, D., Mets, L., Glick, B. & Keenan, R. (2010) Chromophore formation in DsRed occurs by a branched pathway. *Journal of the American Chemical Society* 132:8496-8505.
- 20 Mudalige, K., Habuchi, S., Goodwin, P., Pai, R., De Schryver, F. & Cotlet, M. (2010) Photophysics of the red chromophore of HcRed: evidence for cis-trans isomerization and protonation-state changes. *The journal of physical chemistry. B* 114:4678-4685.
- 21 Pletnev, S., Shcherbo, D., Chudakov, D., Pletneva, N., Merzlyak, E., Wlodawer, A., Dauter, Z. & Pletnev, V. (2008) A crystallographic study of bright far-red fluorescent protein mKate reveals pH-induced cis-trans isomerization of the chromophore. *The Journal of biological chemistry* 283:28980-28987.
- 22 Adam, V., Lelimosin, M., Boehme, S., Desfonds, G., Nienhaus, K., Field, M., Wiedenmann, J., McSweeney, S., Nienhaus, G. & Bourgeois, D. (2008) Structural

- characterization of IrisFP, an optical highlighter undergoing multiple photo-induced transformations. *Proceedings of the National Academy of Sciences of the United States of America* 105:18343-18348.
- 23 Loos, D., Habuchi, S., Flors, C., Hotta, J.-I., Wiedenmann, J., Nienhaus, G. & Hofkens, J. (2006) Photoconversion in the red fluorescent protein from the sea anemone *Entacmaea quadricolor*: is cis-trans isomerization involved? *Journal of the American Chemical Society* 128:6270-6271.
- 24 Wall, M., Socolich, M. & Ranganathan, R. (2000) The structural basis for red fluorescence in the tetrameric GFP homolog DsRed. *Nature structural biology* 7:1133-1138.
- 25 Abbyad, P., Childs, W., Shi, X. & Boxer, S. (2007) Dynamic Stokes shift in green fluorescent protein variants. *Proceedings of the National Academy of Sciences of the United States of America* 104:20189-20194.
- 26 Shu, X., Wang, L., Colip, L., Kallio, K. & Remington, S. (2009) Unique interactions between the chromophore and glutamate 16 lead to far-red emission in a red fluorescent protein. *Protein science : a publication of the Protein Society* 18:460-466.
- 27 Strack, R. L., Hein, B., Bhattacharyya, D., Hell, S. W., Keenan, R. J. & Glick, B. S. (2009) A rapidly maturing far-red derivative of DsRed-Express2 for whole-cell labeling. *Biochemistry* 48:8279-8281.
- 28 Suto, K., Masuda, H., Takenaka, Y., Tsuji, F. I. & Mizuno, H. (2009) Structural basis for red-shifted emission of a GFP-like protein from the marine copepod *Chiridius poppei*. *Genes to cells : devoted to molecular & cellular mechanisms* 14:727-737.
- 29 Kredel, S., Oswald, F., Nienhaus, K., Deuschle, K., Röcker, C., Wolff, M., Heilker, R., Nienhaus, G. U. & Wiedenmann, J. (2008) mRuby, a bright monomeric red fluorescent protein for labeling of subcellular structures. *PloS one* 4.
- 30 Nienhaus, K., Nar, H., Heilker, R., Wiedenmann, J. & Nienhaus, G. U. (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* 130:12578-12579.
- 31 Adams, P., Afonine, P., Bunkóczi, G., Chen, V., Davis, I., Echols, N., Headd, J., Hung, L.-W., Kapral, G., Grosse-Kunstleve, R., McCoy, A., Moriarty, N., Oeffner, R., Read, R., Richardson, D., Richardson, J., Terwilliger, T. & Zwart, P. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta crystallographica. Section D, Biological crystallography* 66:213-221.
- 32 Emsley, P., Lohkamp, B., Scott, W. & Cowtan, K. (2010) Features and development of Coot. *Acta crystallographica. Section D, Biological crystallography* 66:486-501.

## CHAPTER 3

# Computational Design of the $\beta$ -sheet Surfaces of Red Fluorescent Proteins Allows Control of Protein Oligomerization

Timothy M. Wannier<sup>1</sup>, Matthew M. Moore<sup>1,2</sup>, Yun Mou<sup>2</sup>, Stephen L. Mayo<sup>1</sup>

<sup>1</sup> *Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, CA, USA*

<sup>2</sup> *Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, CA, USA*

*(This work is being submitted as a report for publication in 2015)*

### 3.1 Abstract

Computational design has been used with mixed success for the design of protein surfaces, with directed evolution heretofore providing better practical solutions than explicit design. Directed evolution, though, has the drawback of necessitating an easy, high-throughput screen because the random nature of mutation does not enrich for desired traits. Here we demonstrate the successful design of the  $\beta$ -sheet surface of a red fluorescent protein (RFP), enabling control over its oligomerization. To isolate the problem of surface design, we created a hybrid RFP from DsRed and mCherry with a stabilized protein core that allows for monomerization without loss of fluorescence. We designed an explicit library for which 93 of 96 (97%) of the protein variants are soluble, stably fluorescent, and monomeric. RFPs are heavily used in optical biology, but are natively tetrameric, and creating RFP monomers has proven extremely difficult. We show that surface design and core engineering are separate problems in RFP development and that the next generation of RFP markers will depend on improved methods for core design.

## 3.2 Introduction

Computational methods have heavily influenced protein design, but despite successes in core repacking, computational design of protein surfaces, especially those with a high  $\beta$ -sheet content, has lagged [1]. There are relatively few instances of demonstrated success in  $\beta$ -sheet design [2-4]. Directed evolution, having proven effective at tackling problems that computational protein design (CPD) is ill-equipped to address, has been used to some success to evolve soluble  $\beta$ -sheet surfaces [5,6]. Directed evolution, however, is lengthy, requiring high-throughput screening, and inefficient, as error-prone mutagenesis is used to randomly walk through sequence space. Here we present a CPD-driven library creation process that can efficiently search sequence space for soluble protein surfaces, facilitating surface design of proteins in situations that are not readily adaptable to high-throughput screening methods [7,8]. We demonstrate the successful design of fluorescent protein (FP)  $\beta$ -sheet surfaces, expediting monomerization of a core-stabilized RFP.

Oligomerization is a significant barrier to novel FP development. Most native FPs are oligomeric [9-11], and many engineered FPs that are thought to be monomeric exhibit dimerization in certain laboratory or biological contexts, complicating data interpretation, and even contributing to erroneous scientific findings [12,13]. Soluble, monomeric FP probes are needed to prevent FP-driven aggregation or FP-mediated assembly of linked protein targets, and to limit the cytotoxic effects of poorly soluble proteins [12,14]. A major challenge facing FP-engineering is to break oligomeric interfaces without negatively impacting the fluorescent characteristics of a wild-type FP. To do so means designing soluble, beta-barrel surfaces that are not aggregation prone. No standard technique has emerged for efficiently and effectively moving from a dimeric or tetrameric FP to a monomer without extensive intuition-based mutagenesis to disrupt oligomerization, followed by successive rounds of directed evolution to restore fluorescence [15].

The most challenging FPs to effectively monomerize have proven to be red FPs (RFPs). All known native RFPs are tetrameric, the vast majority of which have not been extensively used or characterized because of the difficulty of breaking their oligomerization without compromising fluorescence. Of the more than 50 native RFPs described to date, only four

have been successfully monomerized, as determined by a variety of in vitro methods, and in each case there has been significant mutation to the core of these proteins, often blue-shifting their fluorescent excitation and emission spectra and decreasing their brightness and photostability [5,12,16-18]. Efforts to improve brightness, engineer bathochromic shifts, or otherwise improve existing RFPs have focused on engineered monomers, and so have targeted only a small subset ( $< 10\%$ ) of known RFP biodiversity. It is poorly understood why monomerizing RFPs has negatively impacted their spectroscopic properties.

Here we explore the engineering of monomeric RFPs, attempting to deconvolute the design of a soluble  $\beta$ -barrel surface from any impacts that core mutations have on an RFP's spectroscopic properties. We show that CPD, used in conjunction with efficient library construction and screening, is an effective tool to engineer the surface of RFP-type  $\beta$ -barrel proteins. We tested our method in a hybrid protein engineered as a cross between DsRed, a native tetrameric RFP, and mCherry, a monomeric variant of DsRed, from which we took the evolved protein core (13 mutations from DsRed). Screening a small library of computationally designed variants of a tetrameric RFP, we found that 97% were bright, stable, monomeric, and little changed spectroscopically. This process represents a stark improvement to the speed and efficiency of RFP monomerization, and may facilitate the study of a much broader array of native RFPs, allowing researchers to target engineering efforts to residues in the protein core, as structural stabilization of the chromophore environment appears to be the primary bottleneck to RFP monomerization. This novel computationally driven method for the monomerization of fluorescent proteins should be applicable as a general technique for creating soluble monomeric protein variants.

### 3.3 Results

#### *A Hybrid RFP Demonstrates the Functional Distinction Between Core and Surface Domains*

To validate the computational design of  $\beta$ -barrel surfaces, we decided to work with DsRed, as it has been twice independently monomerized via directed evolution and there exist >60 characterized monomeric variants of the protein. We set out to create a DsRed mutant that would tolerate surface mutations – possessing a structurally sound core that would retain fluorescence upon monomerization. Such a variant would allow us to separate the problem of surface design from the effects of mutations to the chromophore environment. We hypothesized that stabilizing mutations to the core of DsRed were both responsible for changes to its spectroscopic properties and necessary for monomerization, as no monomerization of a native RFP has been successful without altering the protein core [5,12,16-18]. One of the most studied and thoroughly characterized monomeric variants of DsRed is mCherry, a less-bright but red-shifted 30-point mutant [19]. We created a hybrid RFP (DsRmCh) that is a 13-point mutant of DsRed, containing every mutation to a residue in the core of the protein that was introduced during the evolution of mCherry (Figure 1). Consistent with the hypothesis that core residues are determinant of the fluorescent properties of an FP, DsRmCh is spectroscopically mCherry-like, but remains tetrameric. Specifically, DsRmCh retains the bathochromic shift to its fluorescent emission, decreased brightness, and quickened maturation of mCherry, suggesting that the residues responsible for these properties are indeed among the 13 core mutations (Table 1). We measured the oligomerization of DsRmCh, which retains a wild-type DsRed surface, by analytical ultracentrifugation (AUC) and size exclusion chromatography (SEC) (Figure 2). DsRmCh remains tetrameric indicating that the core residues are not implicated in oligomerization. Conversely, the inverse of DsRmCh, mChDsR, comprising an mCherry exterior and a DsRed interior, is not fluorescent, and is in fact not solubly expressed (data not shown), indicating that stabilization to the core of DsRed is required to successfully monomerize the protein. To directly measure the stabilizing effect of DsRmCh's core mutations we ran thermal melts of DsRed, DsRmCh, and mCherry, in which fluorescence is tracked in real time as the temperature is ramped to 99 °C and the temperature at which the protein most

rapidly loses fluorescence (apparent  $T_m$ ) is calculated. These data show that DsRmCh is thermostabilized over DsRed by 4 °C, but that when its oligomerization is broken in moving from DsRmCh to mCherry, the protein is destabilized by 9 °C (Table 1 and Figure S1). The mutations that optimize DsRed's core for monomerization are also thermostabilizing.

### ***DsRmCh is Monomerized with CPD***

Having determined that mutation to the core of mCherry is responsible for the spectroscopic alterations to the protein, we next sought to determine if the optimized mCherry core would indeed facilitate monomerization. As an initial test of the stability of the fluorescence in DsRmCh to surface perturbation, we partially destabilized the AC interface, the more stable of the two oligomeric interfaces, by deleting the protein's five-residue C-terminal tail. The C-terminal tail stabilizes the AC interface with a crucial intermolecular interaction (Figure 1B), and deletion of even the two C-terminal residues in DsRed completely knocks out fluorescence (data not shown). DsRmCh, however, tolerates a complete deletion of its C-terminal tail, comprising residues 221-225. This variant, which we call DsRmCh $\Delta$ 5, remains tetrameric (Figure 3), and there is no significant change to its fluorescent properties, although it does dim very slightly (Table 1). The tail deletion to DsRmCh does lower its apparent  $T_m$  to that of mCherry and there is a slight shift in its elution peak by SEC (Figure 3B), indicating that the oligomeric interaction is destabilized. As DsRmCh was mostly unperturbed by a deletion of its C-terminal tail, we set out to design, via CPD, the  $\beta$ -sheet surfaces of DsRmCh $\Delta$ c5 to fully monomerize the protein.

We analyzed DsRed's two oligomeric interfaces, the AB and AC interface (Figure 1), which are named for the crystallographic chain names from the original structure of DsRed [20], to determine which residues to target for design. To help narrow down the choices, we found ten instances of FP monomerization in the literature, and made an alignment of the native FPs and their engineered monomeric variants [5,16,18,21-27]. We found 17 positions at the two interfaces that were heavily buried, made significant intermolecular contacts, and that had been frequently mutated during the monomerization of previous FPs

(Figure 2). We then targeted these positions, split about evenly between the AB and the AC interface, for surface design using CPD, and allowed the software to sample 12 amino acids, mostly polar and charged residues. We ran a single state design calculation because in not allowing wild-type amino acids at many of the most buried, hydrophobic interface positions, we were confident that we would break oligomerization by force of the number of mutations made. Much like Hu and colleagues found in 2008, we did not need to use an explicit negative design to achieve a soluble  $\beta$ -sheet surface [2]. We constructed a monomeric library (mLib), which comprised the 96 designed variants with the lowest predicted free energy in silica. We then expressed and characterized mLib, and found 97% (93/96) of mLib to be measurably fluorescent. As an initial test of the success of the monomer design, the oligomerization of mLib was tested in high-throughput with a homo-FRET assay, which measures the loss of polarization due to non-radiative energy transfer between neighboring chromophores (Figure 4). Every fluorescent member of mLib was also monomeric, confirming the success of the design. To confirm the results of the homo-FRET assay, select members of mLib were verified by SEC and AUC to be monomeric (Figure 3). Importantly, members of mLib were significantly mutated, with 13-16 surface mutations per variant, but did not share significant mutational similarity to mCherry apart from the shared core mutations. Nine of the 17 non-core sites that were mutated during the evolution of mCherry were targeted for mutation in mLib, but in only one instance was the amino acid residue present in mCherry picked by the computational design software. As the protein core was unchanged between mCherry and the mLib variants, mLib variants were spectroscopically similar to mCherry and about equally as bright (Figure 5).

To objectively measure the success of the computationally designed library mLib, we created a 24-member random control library (rLib) in which we randomized amino acid mutations at the two oligomeric interfaces. We varied the same 17 surface positions in rLib as were targeted with computational design in mLib, and sampled only the amino acid residues that had been allowed in the computational design for mLib. Additionally, we weighted the frequency of occurrence of the amino acids at the randomized positions in rLib by their occurrence on the surface of bacterial mesophilic intracellular proteins [28]. We characterized rLib and when compared with mLib, which was 97% fluorescent, we

found only 8% (2/24) of rLib variants to be fluorescent enough to accurately measure. Most of rLib expressed poorly, with some variants displaying very faint, trace fluorescence (data not shown), and the two fluorescent variants that expressed well were both significantly dimmer than the median mLib variant (Figure 5).

We noticed that some mLib variants, although spectroscopically similar to mCherry when characterized *in vitro*, did not seem to express as robustly in culture. We reasoned that the C-terminal tail might have some effect on soluble protein expression. To gauge protein expression, we directly measured the fluorescence of induced bacterial cultures. We then chose four variants from mLib and added back a DsRed or an mCherry tail to see if these tails would aid protein expression. In every case, adding DsRed's hydrophobic C-terminal tail (His-His-Leu-Phe-Leu) worsened expression while adding mCherry's tail, which is a GFP mimic, improved expression to a level equivalent to that of mCherry (Figure S2). A culture expressing a representative variant, mLib77, was about 70% as bright as mCherry, but when mCherry's 11-residue C-terminal tail was added to the protein, it expressed slightly better than mCherry and was more thermostable (Table 1).

### ***Exploring the Core Mutations Found in mCherry***

To further assess the degree to which core stabilization was necessary in the engineering of mCherry, and to better understand why mChDsR is not fluorescent, we conducted a series of mutational analyses of mCherry. First we reverted each of the 13 positions that had been mutated in the core of mCherry (and which were mutated in DsRed to make DsRmCh) to the wild-type residue found in DsRed. Ten of these mutants were detectably fluorescent, eight of which were about equally fluorescent to mCherry (Table 2). All ten mutants showed similar excitation and emission peaks, shifted only by ~2-3 nm. One reversion mutation of note was from a glutamine to a lysine at position 163, which improved mCherry's quantum yield ( $\Phi$ ) and brightness. We named this variant mCherryR1. In DsRed and mCherry, residue 163 makes Van der Waals contact with the chromophore's phenolate group, but the mutation to glutamine in mCherry appears to disrupt a hydrogen bonded water molecule that lys163 stabilizes along with the backbone carbonyl of residue

144. Using mCherryR1 as a template, we then continued to revert core positions to their wild-type DsRed residues, beginning with the highest quantum yield single reversion variants we had characterized. No reversion mutations to mCherryR1 were found to be beneficial, and indeed as successive core reversions were made, even though they were not negatively impactful individually, in aggregate there was a steady loss of robust protein expression and brightness. After several rounds of mutation and screening, we found a minimally mutated mCherry core that contained only seven mutations from the wild-type DsRed core instead of the 13 present in mCherry. This protein, which we call mCherryR6, has six core residue reversion mutations: A44V, A71V, L124F, M150L, Q163K, and T179S. Any further reversion mutations to the remaining seven mutated core sites of mCherryR6 resulted in a protein that was too weakly expressed to characterize. Thermostability measurements of these variants help to explain mCherry's tolerance for the six core reversion mutations found in mCherryR6. The reversion mutation A217T had been the most puzzling, as it slightly improves mCherry's brightness, but was not tolerated in the mCherryR6 background. Thermostability data, however, showed that this mutation lowers mCherry's apparent  $T_m$  by 9%, which is more than double the impact of any of the six mutations present in mCherryR6. No mutation in mCherryR6 had more than a 4% impact on stability or a 9% impact on brightness.

### **3.4 Discussion**

The success of the computational design of an RFP surface shows both that properly implemented computational procedures can be successful in designing soluble protein surfaces, and that FP monomerization is a problem that is readily solved when separated from the more difficult problem of core optimization. We have provided a systematic demonstration of the capacity of CPD to design stable and soluble  $\beta$ -sheet surfaces, and more specifically the surface of a  $\beta$ -barrel protein. We allowed our computational procedure to design 17 surface residues on DsRmCh, or about one third of the total RFP surface, and designed variants had anywhere from 13 to 16 surface mutations. To our knowledge this is the first demonstration of the use of computational design to transform a

$\beta$ -sheet surface seamlessly from a hydrophobic intermolecular interface into a solvated surface. Designing the surface of a fluorescent protein presented us with a unique opportunity to easily diagnose the success or failure of designed variants, as the brightness and monomericity of each variant could be assayed in high throughput. In fact mLib, the library of 96 DsRmCh $\Delta$ 5 variants that we designed using CPD, is 97% fluorescent, entirely monomeric, and exhibits a mean brightness near to that of the heavily evolved mCherry. In mLib we replicated the monomerization of mCherry, reproducing with one small, explicitly designed library, the surface optimization that was conducted over numerous successive rounds of evolution by fluorescence-activated cell sorting (FACS) and colony screening [19]. An important confirmation of the value added from CPD is that a random control library, rLib, only produced two of twenty-four fluorescent variants, both of which were significantly dimmer than the median mLib variant.

In addition to showing the utility of CPD for designing the surface of globular  $\beta$ -sheet proteins, we can take away some interesting lessons as they pertain to RFP engineering more generally. Previous efforts at RFP engineering have struggled to decouple alterations to the spectroscopic properties of an RFP from oligomeric control of the protein. To our knowledge, significant perturbation to the oligomerization of RFPs by way of intuition-based mutation to RFP surfaces has negatively impacted fluorescence in every case. It has only then been through extensive directed evolution, involving heavy mutation to the RFP core, that fluorescence in a monomer has been restored. What has not been clear, however, is the degree to which surface positions are implicated directly in fluorescence as opposed to having more general structural roles. Here, we attribute RFPs' sensitivity to surface mutations, especially those targeted at disrupting oligomeric interfaces, to a general structural perturbation that ensues a loss or partial loss of oligomerization. In native RFPs, the chromophore core, being adapted to the oligomer, is presumably perturbed in some important way when oligomerization is disrupted even partially, leading to loss of protein stability or a disturbance of the active site geometry. This theory is supported by the fact that every successful RFP monomerization has involved significant restructuring of the immediate chromophore environment [18,29-32]. By contrast, in DsRmCh, we borrow a

previously adapted core, which then permits monomerization completely independent of any compensatory mutation to core residues.

It has also been proposed that a well-designed C-terminal tail is important to a monomeric FP, despite little structural evidence to support this theory. During the evolution of mCherry, for instance, GFP's tail was grafted onto the C-terminus of an early variant, a strategy that has subsequently been repeated in other FP engineering efforts [19]. We find that the C-terminal tail plays a critical structural role in oligomeric RFPs, but does not do so in monomeric RFPs, as an RFP that has a core that is adapted to a monomeric scaffold can tolerate a full tail deletion, whereas a strictly oligomeric RFP cannot. We deleted the entire C-terminal tail of DsRmCh (residues 221-225) before engineering its  $\beta$ -sheet surface to break oligomerization, with no significant impact to its brightness. DsRed, by contrast, being maladapted to monomerization, takes a significant hit to its fluorescence with the deletion of just one C-terminal residue, Leu225, and completely loses fluorescence with the deletion of a second (Table 1). We then find through the study of mLib variants that the C-terminal tail can improve the expression of monomeric RFPs, which most likely has to do with protein solubility effects. The addition of mCherry's C-terminal tail to mLib variants improves their expression and thermostability, whereas the addition of DsRed's hydrophobic tail decreases protein expression in mLib variants, although confusingly not thermostability. Neither tail, however, had any significant impact on brightness. We propose that a well-engineered C-terminal tail may help to prevent protein aggregation in some monomeric RFPs, but that further optimization may obviate the need for the tail altogether.

We describe the successful computational design of the surface of a soluble protein with a large  $\beta$ -sheet component, breaking homo-oligomerization. By designing the surface of an RFP we were able to both easily measure the success of the designed variants, and contribute to a better understanding of FP engineering. Our results suggest that the core of an FP determines its fluorescent characteristics, independent of serious structural perturbation, but that not every fluorescent core is viable in a monomeric scaffold. To design the next generation of brighter and bathochromically shifted RFPs, researchers will need to gain a more thorough understanding of the structural environment throughout the

protein core, especially in the vicinity of the chromophore. Specifically, it will be important to find red-shifted RFP cores that remain structurally sound, catalyze chromophore cyclization, and provide an environment shielded from bulk solvent in the context of a monomeric scaffold.

### **3.5 Materials and Methods**

#### ***Plasmids and Bacterial Strains***

DsRed and mCherry sequences were taken from their Genbank entries (accession numbers AF168419 and AY678264). Ten amino acids were added to the N-terminus of each protein, consisting of a methionine followed by a 6x histidine tag for protein purification, and then followed by a Gly-Ser-Gly linker sequence. All gene sequences were constructed with gene assembly PCR, oligonucleotides for the assembly were designed with DNAworks, and then ordered from Integrated DNA Technologies (IDT). Assembled genes were PCR-amplified and cloned into the pET-53-DEST expression plasmid (EMD Millipore) with PIPE cloning followed by CPEC. Constructs were sequence-verified (Laragen) with a primer specific to the T7 promoter, and then transformed into BL21-Gold(DE3) competent cells, a protein expression strain (Agilent).

#### ***Construction of Designed Libraries and Variants***

Explicitly-designed DsRmCh, mLib, and rLib protein sequences were input into DNAworks as “mutant runs” of the wild-type DsRed gene assembly. This allows explicit libraries of gene variants to be assembled and minimizes the number of oligonucleotides needed. Oligonucleotides were ordered from IDT and cloning was carried out as described above.

### ***Protein Expression and Purification***

Single bacterial colonies were picked with sterile toothpicks and inoculated into 300  $\mu$ l of Super Optimal Broth (SOB) supplemented with 100  $\mu$ g/ml ampicillin in 2 ml deep-well 96-well plates (Seahorse Bioscience). The plates were sealed with microporous film to facilitate gas exchange during growth. Cultures were grown overnight at 37 °C / 300 RPM. The next morning 800  $\mu$ l of fresh SOB with 100  $\mu$ g/ml ampicillin and 1mM Isopropyl  $\beta$ -D-1-thiogalactopyranoside (IPTG) was added to a total volume of 1 ml (evaporation losses overnight are approximately 100  $\mu$ l). Plates were then shaken 12 hours at 37 °C / 400 RPM. Cell cultures turn red if there is strong RFP expression. The 96-well plates were then centrifuged at 3,000 x g in a swinging-bucket rotor and the supernatant was decanted. Pellets were resuspended in 500  $\mu$ l of lysis buffer (50 mM sodium phosphate, 150 mM NaCl, 0.1% v/v Triton-X, 10% v/v 10x Cell Lytic B, pH 7.4) supplemented with 50 Units/ml Benzonase and 0.05 mg/ml Hen Egg Lysozyme. Plates were then shaken on a benchtop plate shaker with a 3 mm orbital stroke length at 1,000 RPM for 30 minutes. To pellet down the cellular debris, the plates were again centrifuged for 10 minutes at 3,000 RPM in a swinging-bucket rotor. The colored supernatant was then applied to a 96-well His-Select filter plate (Sigma), washed twice (50 mM sodium phosphate, 150 mM NaCl, 15 mM Imidazole, pH 7.4), and eluted with 500  $\mu$ l elution buffer (50 mM sodium phosphate, 150 mM NaCl, 250 mM Imidazole, pH 7.4). All His-Select purification steps were performed at 1,000 x g in a swinging bucket rotor.

### ***Fluorescent Protein Characterization***

Purified proteins were assayed in triplicate in Greiner UV-Star 96-well plates with a Tecan Safire2. An absorbance scan (260 – 650 nm), a fluorescence excitation scan (500 – 640 nm excitation / 675 nm emission), and a fluorescence emission scan (550 nm excitation / 575 – 800 nm emission) were run on 100  $\mu$ l of eluted protein to determine spectral peaks.

To measure the quantum yield we diluted each protein so that the absorbance for 200  $\mu$ l of protein at 540 nm was between 0.1 and 0.5. We then measured the A550 in triplicate (or duplicate if it was a poorly expressed protein), diluted the sample to an A550 of 0.04 and took an emission scan (540 nm excitation / 550 – 800 nm emission). The area under the

emission curve was calculated after fitting it to a 4th order Gaussian, and the quantum yield was calculated with the following formula:

$$\Phi_x = (A_s / A_x) (F_x / F_s) (n_x / n_s)^2 \Phi_s \quad (1)$$

Where  $\Phi$  is quantum yield, A is absorbance, F is total fluorescent emission (area under the curve), and n is the refractive index of the solvents used. Subscript X refers to the queried substance and subscript S refers to a standard of known quantum yield. It is important that the standard be excited with the same wavelength of light as the unknown sample. We use DsRed, which has a known quantum yield of 0.79 as the protein standard.

To measure extinction coefficient we took 100  $\mu$ l of the protein solution that had been diluted to an A550 of between 0.1 and 0.5 and measured absorbance between 400 nm and 700 nm in triplicate. We then added 100  $\mu$ l of 2M NaOH to each well and remeasured absorbance between 400 nm and 700 nm. The base-denatured chromophore, which peaks at approximately 450 nm has a known extinction coefficient of 44,000 M<sup>-1</sup>cm<sup>-1</sup>. Then to calculate the extinction coefficient is calculated with the following formula:

$$\varepsilon = A_{Chromophore} * 44,000 M^{-1} cm^{-1} / A_{450} \quad (2)$$

To measure thermal stability, purified proteins were diluted to an absorbance of 0.2 at the wavelength of maximum absorbance ( $\lambda_{abs}$ ) so that their fluorescence would not saturate the detector. 50  $\mu$ l of each purified protein was then loaded into a 96-well PCR plate and covered with clear optical tape. The proteins were incubated at 37°C for 10 minutes and then the temperature was ramped at 0.5°C every 30 seconds up to 99°C, with fluorescence measured every ramp step in a CFX96 Touch Real-Time PCR Detection System (Bio-Rad). We refer to this as a thermal melt. The derivative curve of the thermal melt finds the inflection point of the slope, which is the apparent temperature at which fluorescence is irrecoverably lost (apparent T<sub>m</sub>).

### ***Oligomeric Determination***

#### Size exclusion chromatography

100  $\mu$ l of each purified protein analyzed was run over a Superdex 75 10/300 size exclusion column with 25 ml bed volume on an AKTA from GE Life Sciences. Absorbance was measured after passage through the column at 575 nm, where the red chromophore absorbs.

#### Analytical ultracentrifugation

Purified protein samples were diluted to an  $A_{575}$  of 0.5 for a path-length of 1.25 cm. These samples were put into two-channel sedimentation velocity cuvettes with the blank channel containing PBS. Sedimentation velocity was run at 40,000 RPM overnight with full  $A_{575}$  scans collected with no pause between reads. Data was loaded into Sedfit and a  $c(m)$  distribution was run with default assumptions made for PBS buffer viscosity. After integration, the  $c(m)$  curve was exported to Excel. (C) Homo-FRET. 200  $\mu$ l of each purified protein was diluted to an Absorbance of 0.1 to 0.5 at 530 nm in 96-well Greiner UV-Star plates. Polarization scans were then taken with excitation at 530 nm and emission at 610 nm in a Tecan Safire2 plate-reader. Rose Bengal was used as a standard to calculate the instrument G factor ( $mP = 349$ ).

### **3.6 Acknowledgments**

We would like to acknowledge the help of two students, Melissa Lin and Shweta Bhatia for constructing some variants used in this study. Additionally we thank Alex Nisthal and Jan Kostecki for helpful discussions.

### 3.7 Tables and Figures

**Table 1. Properties of DsRed, mCherry, DsRmCh, and Variants.** This table lists the spectroscopic properties of the important RFP variants mentioned in this paper. “mLib Avg” gives the mean of all members of mLib, while “mLib Top” gives the top value of each attribute seen in mLib. “mLib Top” does not refer to one specific variant.

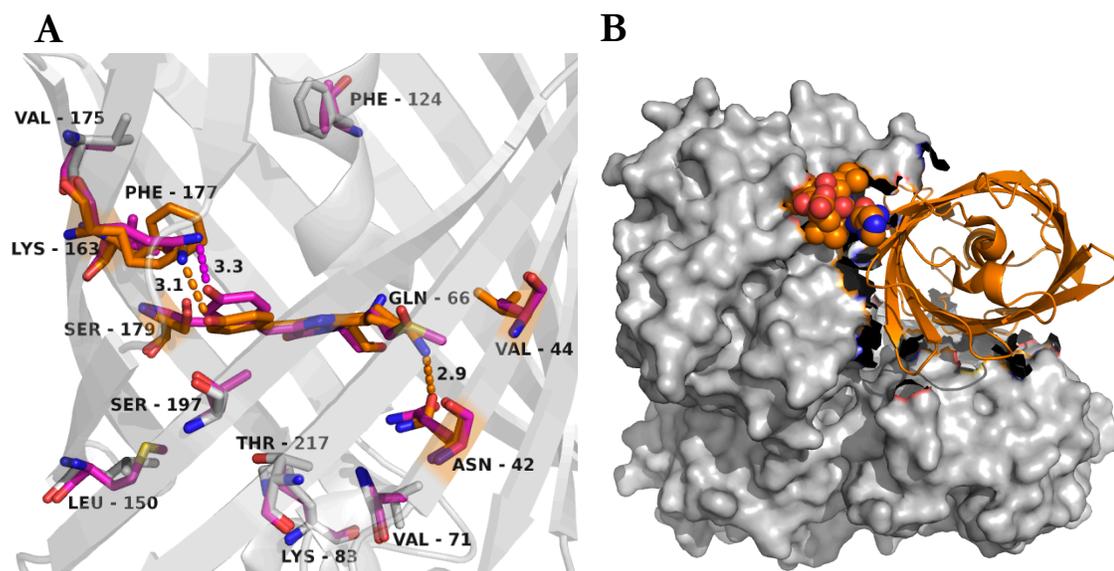
RFP Name	# Non-tail mutations ( $\Delta$ /+ tail length)	Excitation Max $\lambda_{ex}$ (nm)	Emission Max $\lambda_{em}$ (nm)	$\Phi$	$\epsilon$ ( $M^{-1} \text{ cm}^{-1}$ ) / 1000	Brightness ( $\Phi \times \epsilon$ ) / 1000	Apparent $T_m$ ( $^{\circ}\text{C}$ )	Fluorescence in Culture: 570 nm ex / 610 nm em (% mCherry)
DsRed	--	558	585	0.79	73	58	94.5	--
DsRed $\Delta$ 1	$\Delta$ 1 †	558	584	0.57	53	30	84.0	--
mCherry	30 +6 ‡	588	611	0.22	85	19	89.5	100
mCherry $\Delta$ 6	30	588	611	0.21	91	19	89.5	103
mCherry $\Delta$ 11	30 $\Delta$ 5 †	588	612	0.20	78	16	87.5	98
DsRmCh	13	585	611	0.23	97	22	98.5	133
DsRmCh $\Delta$ 5	13 $\Delta$ 5 †	586	612	0.20	92	21	89.5	113
mLib Avg	26 $\Delta$ 5 †	585	608	0.23	62	14	--	--
mLib Top	29 $\Delta$ 5 †	586	609	0.26	72	19	--	--
mLib77	27 $\Delta$ 5 †	586	609	0.23	72	17	89.0	72
mLib77 + DsRed Tail	27	584	609	0.23	72	17	91.5	45
mLib77 + mCherry Tail	27 +6 ‡	586	609	0.24	72	17	93.0	107

† -  $\Delta$  indicates a number of residues deleted from the C-terminal tail

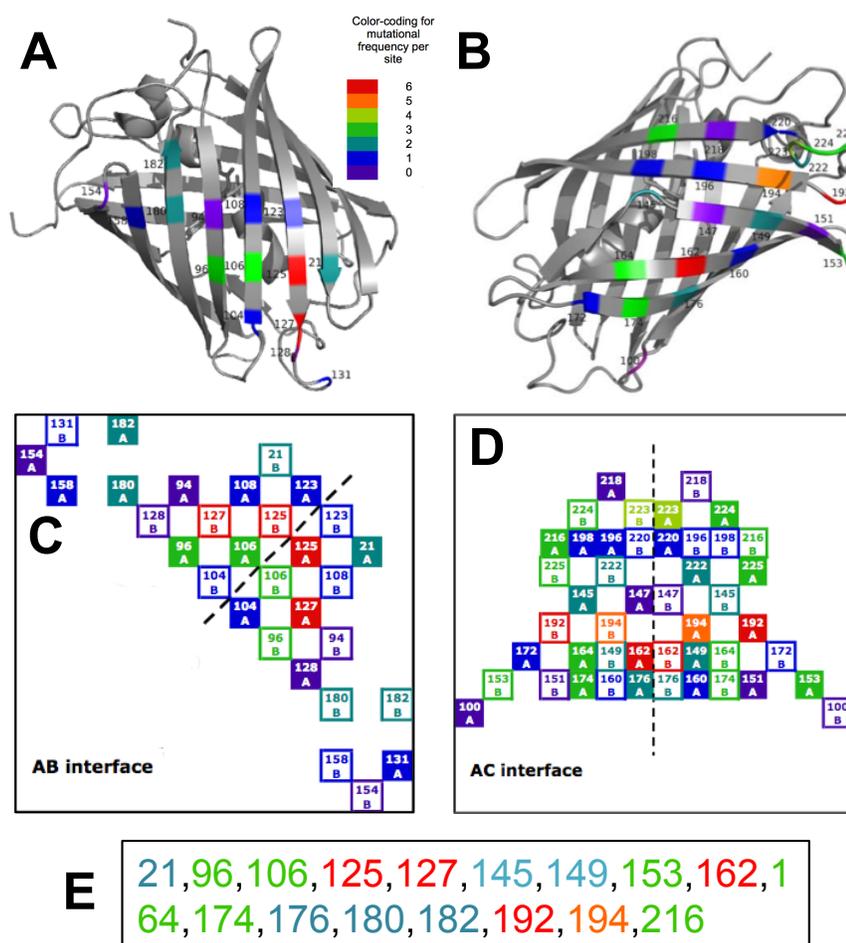
‡ - + indicates a number of residues added to the C-terminal tail

**Table 2. mCherry Core Reversion Variants** This table lists changes to peak emission wavelength, brightness, and thermostability as the result of reversion mutations to mCherry's core.

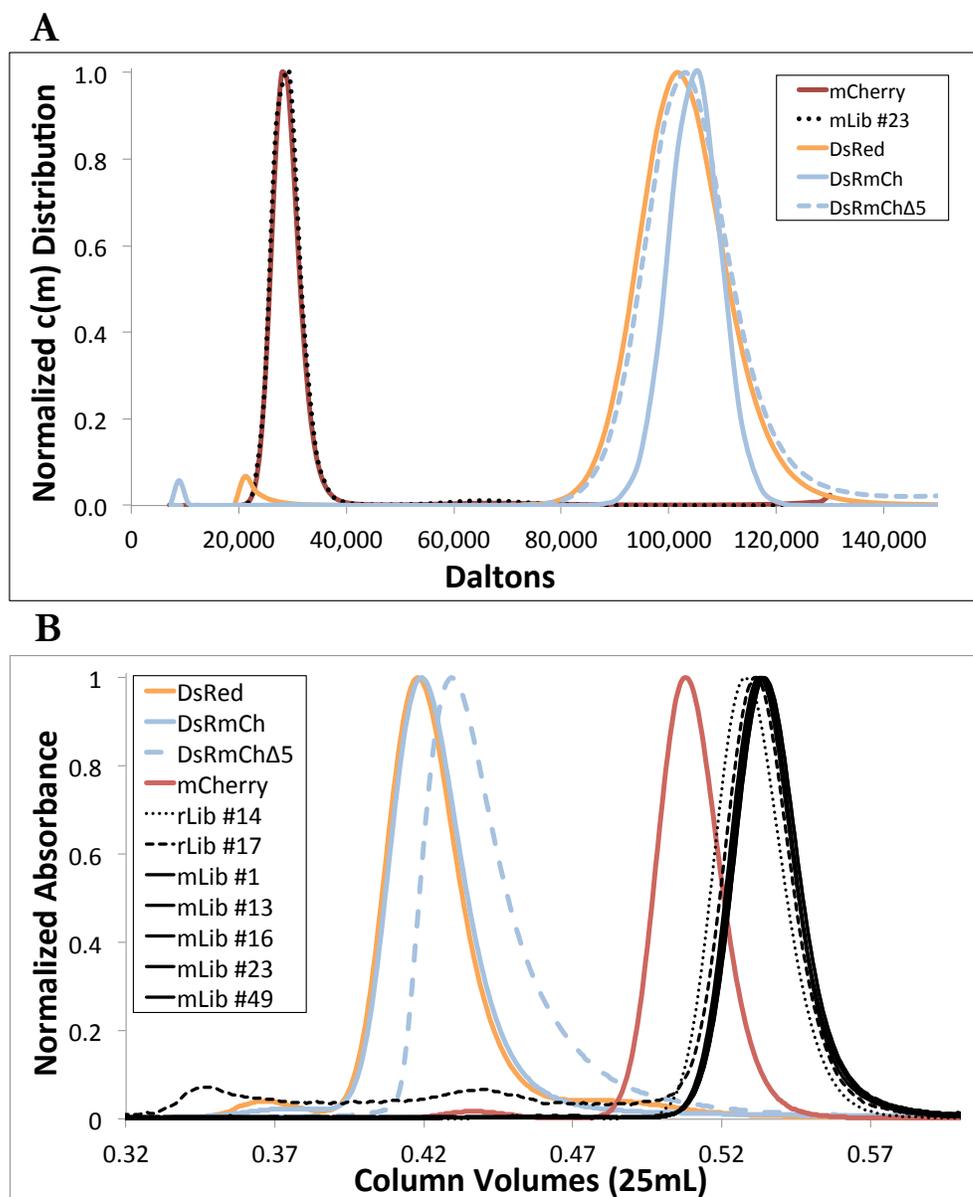
Mutation	Brightness (% mCherry)	Apparent Tm (% mCherry)	Shift in emission (nm)
WT mCherry	100	100	--
A44V	93	96	+2
M68Q	90	95	-3
A71V	100	100	-3
L124F	99	99	--
M150L	91	98	-1
Q163K	108	98	--
(mCherryR1)			
A175V	84	98	--
V177F	24	93	+4
T179S	95	98	+1
A217T	104	91	-2
M150L / Q163K	104	99	-2
Q163K / T179S	94	99	+1
A71V / Q163K / T179S	100	99	-1
A71V / M150L / Q163K / T179S	95	98	-3
A71V / L124F / M150L / Q163K / T179S	93	97	-3
A44V / A71V / L124F / M150L / Q163K / T179S (mCherryR6)	81	89	-1



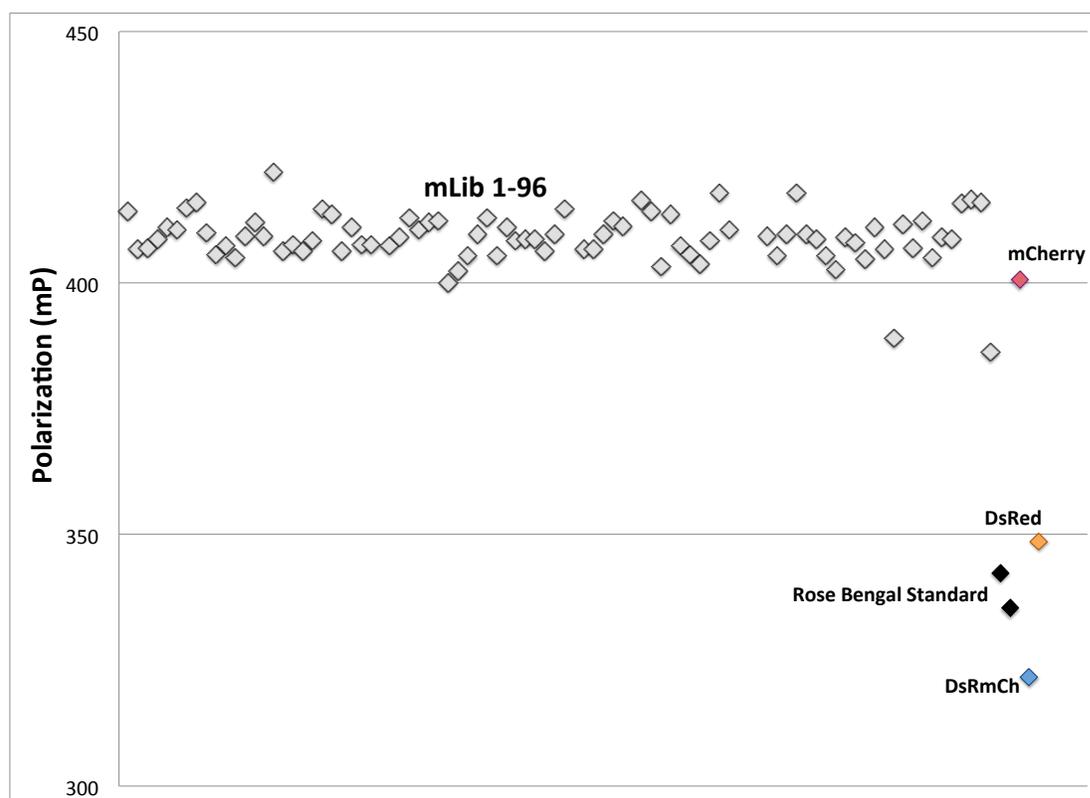
**Figure 1. Structure of DsRed (PDB ID: 1ZGO).** (A) Positions that were mutated in the core of the protein during the directed evolution of mCherry. DsRed is shown as a cartoon in gray, mutated residues are shown in sticks, and those within 5 Å of the chromophore are highlighted in orange. The aligned residues from mCherry (PDB ID: 2H5Q) are overlaid in pink. (B) To visualize the tertiary structure of DsRed chains B, C, and D from the crystal structure are shown as a gray surface, while chain A is shown as an orange cartoon. The C-terminal tail of chain A, shown as spheres, stabilizes the AC interface between chain A and chain C, to its immediate left in the image. Below chain A in the image is chain B, with which chain A forms the AB dimeric interface.



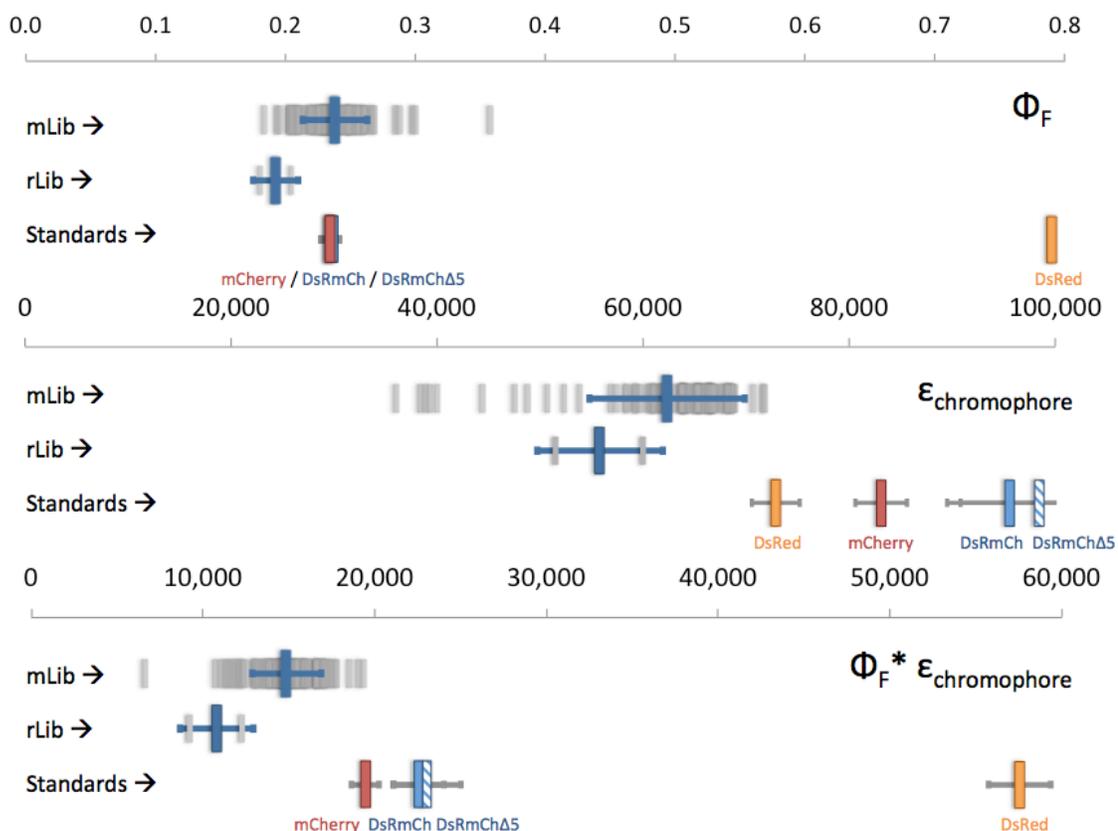
**Figure 2. Schematic Representation of the AB and AC Interfaces.** (*A + B*) Color-coded representation of mutational frequency during the monomerization of a select group of FP monomers (mRFP1, DsRed.M1, mTFP1, mAG, mKO, mEosFP, mKeima, TagRFP, Dendra, and Dronpa) mapped onto the AB and AC interfaces. (*C + D*) A representation of the two interfaces as a grid, with solid squares denoting residues from chain A and open squares denoting residues from the interacting chain. The proximity of solid to open residues represents inter-chain oligomeric interactions and the dashed lines are the lines of symmetry. Note that  $\beta$ -sheets from opposing subunits of the AC interface are stacked in an anti-parallel fashion, while those of the AB interface are offset by  $\sim 90^\circ$ . Note also that cleaving residues from position 221 onward (C-terminal tail) noticeably decreases the complexity of the AC interface. (*E*) The residues picked for computational design in mLib were chosen to be those that were highly mutated and made numerous intermolecular contacts.



**Figure 3. Low Throughput Oligomerization Analysis of Important FPs by AUC and SEC.** (A) The  $c(m)$  distribution was calculated for a sedimentation velocity run by AUC. This data clearly shows that DsRed, DsRmCh, and DsRmCh $\Delta$ 5 are tetrameric, while mCherry and mLib variant 23 are both monomeric. (B) SEC data was collected on a Superdex 75 column. DsRed and DsRmCh are clearly tetrameric and a shift can be seen in DsRmCh $\Delta$ 5 after the deletion of the five C-terminal residues. mCherry by contrast, runs as a clear monomer, and the designed mLib variants run even further shifted than mCherry, possibly reflecting their lack of C-terminal tail.

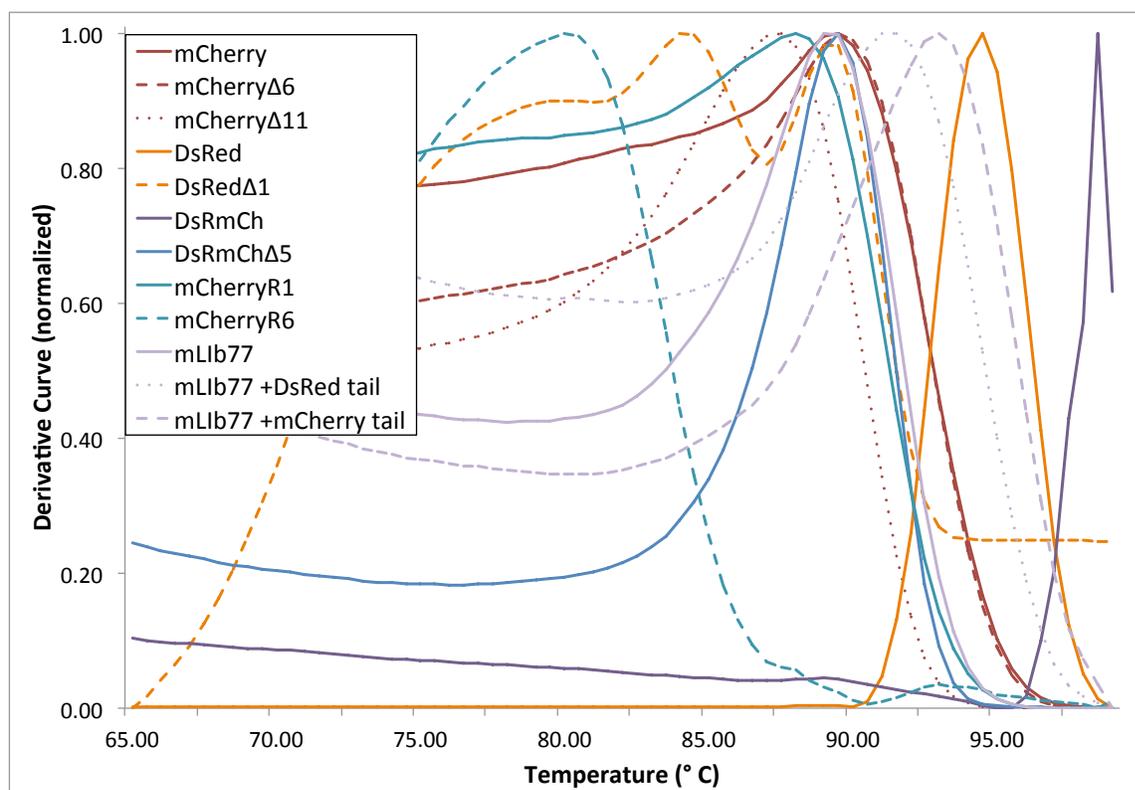


**Figure 4. High Throughput Oligomerization Analysis of mLib and Important FPs by Homo-FRET.** The polarization of the fluorescent emission of purified proteins was analyzed to determine their oligomeric state. Non-radiative transfer of energy between monomers in an oligomeric protein will cause a drop in the polarization of the fluorescent emission. DsRed and DsRmCh clearly show lower polarization than do mCherry and mLib, both confirming the results seen by AUC and SEC in Figure 3 and suggesting that Homo-FRET is a good high-throughput technique for accurately gauging the oligomerization of fluorescent proteins. Position along the x-axis just serves to separate individual variants for better visibility.

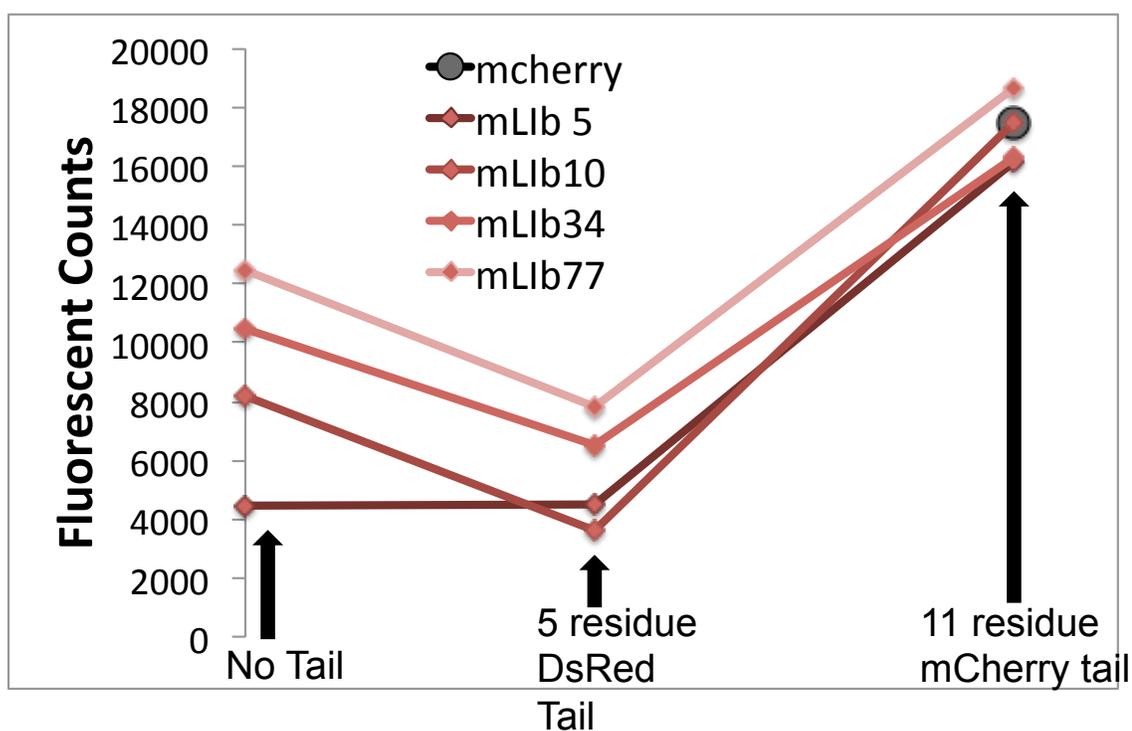


**Figure 5. Characterization of Designed Libraries and Important Variants.** The charts above graphically represent the (*Top*) quantum yield, (*Middle*) extinction coefficient, and (*Bottom*) brightness of mLib, rLib, and important standards. Only two of the twenty-four rLib variants expressed well enough to characterize, and both are significantly dimmer than the mLib mean. The brightest mLib variants are about equally as bright as mCherry, but with a higher quantum yield and lower extinction coefficient than mCherry. Note that DsRed is much brighter than DsRmCh, entirely due to its greatly increased quantum yield. DsRmCh by contrast has a higher extinction coefficient than DsRed. Library error bars represent one standard deviation of the average mLib or rLib member. Error bars on specific variants represent the standard deviation from three separate measurements.

### 3.8 Supplementary Tables and Figures



**Figure S1. Thermal Stability of Important Variants.** Important DsRed, DsRmCh, and mCherry variants were purified and thermally denatured in a real-time PCR machine. The decrease in fluorescence as the proteins unfolded was measured in real-time, and shown here are the normalized derivative curves. The peaks of these curves represent the temperature at which fluorescence dropped the most quickly (apparent  $T_m$ ). DsRmCh is the most stable variant, as it is tetrameric, has the C-terminal tail, and a stabilized mCherry core. Monomerization destabilizes mCherry, and the effects of core reversion mutations can be seen in the mCherry variants R1 and R6.



**Figure S2. mCherry Tail Improves Protein Expression.** We grafted either a 5-residue DsRed tail (HHLFL) or an 11-residue mCherry tail (HSTGGMDELYK) onto four mLib variants and measured their fluorescence in an induced bacterial culture. The DsRed tail negatively impacts protein expression, while the mCherry tail significantly improves it. mLib variants with grafted mCherry tails express as well as mCherry.

DsRed Residue:		T	V	T	I	V	A	R	R	H	A	L	E	I	M	Y	Y	R
mLib #		21	96	106	125	127	145	149	153	162	164	174	176	180	182	192	194	216
1	R	R	R	Q	T	T	E	N	Q	N	N	R	K	K	K	N	N	N
2	R	R	R	Q	T	T	E	N	Q	N	N	R	K	K	K	N	N	N
3	R	R	R	Q	T	T	E	N	Q	N	N	R	K	K	K	N	N	N
4	R	R	R	Q	T	T	E	N	Q	N	N	R	K	K	K	N	N	N
5	R	H	R	Q	T	T	E	N	Q	N	N	R	K	K	K	N	N	N
6	R	R	R	Q	T	T	E	N	Q	N	N	R	K	K	K	N	N	N
7	R	R	R	Q	T	T	E	N	Q	N	N	R	K	K	K	N	N	N
8	R	R	R	Q	T	T	E	N	Q	N	N	R	K	K	K	N	N	N
9	R	R	R	Q	T	T	E	N	Q	N	N	R	K	K	K	N	N	N
10	R	T	R	Q	T	T	E	N	Q	N	N	R	K	K	K	N	N	N
11	R	R	R	Q	T	T	E	N	Q	N	N	R	K	K	K	N	N	N
12	R	R	R	Q	T	T	E	N	Q	N	N	R	K	K	K	N	N	N
13	R	R	R	Q	T	T	E	D	Q	N	N	R	K	K	K	N	N	N
14	R	R	R	Q	T	T	E	N	Q	N	N	R	K	K	K	N	N	N
15	R	R	R	Q	T	T	E	N	Q	N	N	R	K	K	K	N	N	N
16	R	R	R	Q	T	T	E	N	Q	N	N	R	K	K	K	N	N	N
17	R	R	R	Q	T	T	E	N	Q	N	N	R	K	K	K	N	N	N
18	R	R	R	Q	T	T	E	N	Q	N	N	H	K	K	K	N	N	N
19	R	R	R	Q	T	T	E	N	Q	N	N	R	H	K	K	N	N	N
20	R	R	R	Q	T	T	E	Q	A	N	N	R	K	K	K	N	N	N
21	R	H	R	Q	T	T	E	N	A	N	N	R	K	K	K	N	N	N
22	R	R	R	Q	T	T	E	N	Q	D	D	H	K	K	K	N	N	N
23	R	R	R	Q	T	T	E	N	Q	D	D	R	K	K	K	N	N	N
24	R	H	R	Q	T	T	E	N	Q	N	N	R	K	K	K	N	N	N
25	R	R	R	Q	T	T	E	N	A	D	D	R	K	K	K	N	N	N
26	R	H	R	Q	T	T	E	N	Q	D	D	R	K	K	K	N	N	N
27	R	H	R	Q	T	T	E	N	Q	N	N	R	K	K	K	N	N	N
28	R	R	R	Q	T	T	E	K	Q	Q	N	R	K	K	K	N	N	N
29	R	R	R	Q	T	T	E	K	N	A	N	R	K	K	K	N	N	N
30	R	R	R	Q	T	T	E	N	A	N	N	R	K	K	K	N	N	N
31	R	T	R	Q	T	T	E	N	A	N	N	R	K	K	K	N	N	N
32	R	R	R	Q	T	T	E	N	Q	N	N	R	K	K	K	N	N	N
33	R	R	R	Q	T	T	E	N	Q	D	D	R	K	K	K	N	N	N
34	R	T	R	Q	T	T	E	N	Q	N	N	R	K	K	K	N	N	N
35	R	R	R	Q	T	T	E	N	Q	N	N	R	K	K	K	N	N	N
36	R	T	R	Q	T	T	E	N	Q	D	D	R	K	K	K	N	N	N
37	R	H	R	Q	T	T	E	N	Q	N	N	R	K	K	K	N	N	N
38	R	R	R	Q	T	T	E	N	Q	N	N	R	K	K	K	N	N	N
39	R	R	R	Q	T	T	E	N	Q	N	N	R	K	K	K	N	N	N
40	R	R	R	Q	T	T	E	N	A	N	N	R	K	K	K	N	N	N
41	R	H	R	Q	T	T	E	N	Q	N	N	R	K	K	K	N	N	N
42	R	R	R	Q	T	T	E	N	Q	N	N	R	K	K	K	N	N	N
43	R	R	R	Q	T	T	E	D	A	N	N	R	K	K	K	N	N	N
44	R	R	R	Q	T	T	E	N	A	N	N	R	K	K	K	N	N	N
45	R	R	R	Q	T	T	E	N	Q	D	D	R	K	K	K	N	N	N
46	R	T	R	Q	T	T	E	N	Q	N	N	R	K	K	K	N	N	N
47	R	R	R	Q	T	T	E	N	Q	N	N	R	K	K	K	N	N	N
48	R	R	R	Q	T	T	E	D	A	N	N	R	K	K	K	N	N	N
49	R	K	R	Q	T	T	E	N	Q	N	N	R	K	K	K	N	N	N
50	R	R	R	Q	T	T	E	D	Q	D	D	R	K	K	K	N	N	N
51	R	T	R	Q	T	T	E	N	Q	N	N	R	K	K	K	N	N	N
52	R	R	R	Q	T	T	E	N	Q	N	N	R	K	K	K	N	N	N
53	R	R	R	Q	T	T	E	N	N	N	N	R	K	K	K	N	N	N
54	R	R	R	Q	T	T	E	N	A	N	N	R	K	K	K	N	N	N
55	R	R	R	Q	T	T	E	K	N	Q	D	H	K	K	K	N	N	N
56	R	R	R	Q	T	T	E	N	A	N	N	H	K	K	K	N	N	N
57	R	R	R	Q	T	T	E	K	N	A	N	R	K	K	K	N	N	N
58	R	R	R	Q	T	T	E	K	N	Q	N	R	K	K	K	N	N	N
59	R	R	R	Q	T	T	E	N	S	N	N	R	K	K	K	N	N	N
60	R	H	R	Q	T	T	E	N	Q	N	N	H	K	K	K	N	N	N
61	R	R	R	Q	T	T	E	N	A	D	D	H	K	K	K	N	N	N
62	R	R	R	Q	T	T	E	K	N	A	D	R	K	K	K	N	N	N
63	R	T	R	Q	T	T	E	N	Q	N	N	H	K	K	K	N	N	N
64	R	H	R	Q	T	T	E	N	A	N	N	R	K	K	K	N	N	N
65	R	H	R	Q	T	T	E	K	N	A	N	R	K	K	K	N	N	N
66	R	H	R	Q	T	T	E	N	A	D	D	R	K	K	K	N	N	N
67	R	H	R	Q	T	T	E	K	Q	Q	N	R	K	K	K	N	N	N
68	R	R	R	Q	T	T	E	N	Q	N	N	R	K	K	K	N	N	N
69	R	H	R	Q	T	T	E	N	Q	N	N	R	K	K	K	N	N	N
70	R	R	R	Q	T	T	E	K	N	Q	D	R	K	K	K	N	N	N
71	R	R	R	Q	T	T	E	K	N	A	N	R	K	K	K	N	N	N
72	R	R	R	Q	T	T	E	N	A	D	D	R	K	K	K	N	N	N
73	R	H	R	Q	T	T	E	K	N	Q	N	R	K	K	K	N	N	N
74	R	H	R	Q	T	T	E	K	N	S	N	R	K	K	K	N	N	N
75	R	R	R	Q	T	T	E	N	Q	D	D	H	K	K	K	N	N	N
76	R	T	R	Q	T	T	E	K	Q	Q	N	R	K	K	K	N	N	N
77	R	T	R	Q	T	T	E	K	N	A	N	R	K	K	K	N	N	N
78	R	H	R	Q	T	T	E	K	N	Q	N	R	K	K	K	N	N	N
79	R	T	R	Q	T	T	E	N	A	D	D	R	K	K	K	N	N	N
80	R	T	R	Q	T	T	E	K	N	Q	D	R	K	K	K	N	N	N
81	R	T	R	Q	T	T	E	N	A	N	N	R	K	K	K	N	N	N
82	R	R	R	Q	T	T	E	N	Q	N	N	R	K	K	K	N	N	N
83	R	H	R	Q	T	T	E	K	N	Q	N	R	K	K	K	N	N	N
84	R	H	R	Q	T	T	E	N	A	N	N	R	K	K	K	N	N	N
85	R	T	R	Q	T	T	E	K	N	Q	N	R	K	K	K	N	N	N
86	R	T	R	Q	T	T	E	N	Q	D	D	R	K	K	K	N	N	N
87	R	H	R	Q	T	T	E	K	N	Q	N	R	K	K	K	N	N	N
88	R	T	R	Q	T	T	E	N	S	D	N	R	K	K	K	N	N	N
89	R	R	R	Q	T	T	E	K	D	A	N	R	K	K	K	N	N	N
90	R	T	R	Q	T	T	E	N	S	D	N	R	K	K	K	N	N	N
91	R	H	R	Q	T	T	E	K	N	S	N	R	K	K	K	N	N	N
92	R	R	R	Q	T	T	E	K	N	Q	D	R	K	K	K	N	N	N
93	R	R	R	Q	T	T	E	K	N	N	N	R	K	K	K	N	N	N
94	R	R	R	Q	T	T	E	N	Q	D	D	R	K	K	K	N	N	N
95	R	R	R	Q	T	T	E	N	Q	S	K	R	K	K	K	N	N	N

Table S1. mLib Variant Designed Positions. This table lists all of the 95 mLib variants. Each of the 17 designed positions is shown in a separate column. Mutations are listed next to the mLib variant number, with “-” indicating no mutation from the wild-type DsRed sequence.

### 3.9 References

1. Samish I, MacDermaid CM (2011) Theoretical and computational protein design. *Annual Review of Physical Chemistry*.
2. Hu X, Wang H, Ke H, Kuhlman B (2008) Computer-based redesign of a  $\beta$  sandwich protein suggests that extensive negative design is not required for de novo  $\beta$  sheet design. *Structure*.
3. Nanda V, Rosenblatt MM, Osyczka A (2005) De novo design of a redox-active minimal rubredoxin mimic. *Journal of the American Chemical Society*.
4. Street AG, Datta D, Gordon DB, Mayo SL (2000) Designing protein  $\beta$ -sheet surfaces by Z-score optimization. *Physical Review Letters*.
5. Agresti JJ, Antipov E, Abate AR, Ahn K (2010) Ultrahigh-throughput screening in drop-based microfluidics for directed evolution. *Proceedings of the National Academy of Sciences of the United States of America*.
6. Esvelt KM, Carlson JC, Liu DR (2011) A system for the continuous directed evolution of biomolecules. *Nature* 472: 499-503.
7. Alieva NO, Konzen KA, Field SF, Meleshkevitch EA, Hunt ME, et al. (2007) Diversity and evolution of coral fluorescent proteins. *PloS ONE* 3.
8. Shagin DA, Barsova EV, Yanushevich YG, Fradkov AF, Lukyanov KA, et al. (2004) GFP-like proteins as ubiquitous metazoan superfamily: evolution of functional features and structural complexity. *Molecular Biology and Evolution* 21: 841-850.
9. Wannier TM, Mayo SL (2014) The structure of a far-red fluorescent protein, AQ143, shows evidence in support of reported red-shifting chromophore interactions. *Protein Science* 23: 11481153.
10. Shemiakina II, Ermakova GV, Cranfill PJ, Baird MA, Evans RA, et al. (2012) A monomeric red fluorescent protein with low cytotoxicity. *Nature Communications*.
11. Wang S, Moffitt JR, Dempsey GT, Xie XS, Zhuang X (2014) Characterization and development of photoactivatable fluorescent proteins for single-molecule-based superresolution imaging. *Proceedings of the National Academy of Sciences of the United States of America*.
12. Strack RL, Strongin DE, Bhattacharyya D, Tao W, Berman A, et al. (2008) A noncytotoxic DsRed variant for whole-cell labeling. *Nature Methods* 5: 955-957.
13. Ai H-w, Baird MA, Shen Y, Davidson MW, Campbell RE (2014) Engineering and characterizing monomeric fluorescent proteins for live-cell imaging applications. *Nature Protocols*.
14. Takako K, Satoshi K, Toshio A, Kenta S, Masataka K, 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: 577-581.

15. Kredel S, Oswald F, Nienhaus K, Deuschle K, Röcker C, et al. (2008) mRuby, a bright monomeric red fluorescent protein for labeling of subcellular structures. *PLoS ONE* 4.
16. Campbell RE, Tour O, Palmer AE, Steinbach PA, Baird GS, et al. (2002) A monomeric red fluorescent protein. *Proceedings of the National Academy of Sciences*.
17. Strongin DE, Bevis B, Khuong N, Downing ME, Strack RL, et al. (2007) Structural rearrangements near the chromophore influence the maturation speed and brightness of DsRed variants. *Protein Engineering, Design & Selection* 20: 525-534.
18. Shaner NC, Campbell RE, Steinbach PA, Giepmans BN, Palmer AE, et al. (2004) Improved monomeric red, orange and yellow fluorescent proteins derived from *Discosoma* sp. red fluorescent protein. *Nature Biotechnology* 22: 1567-1572.
19. Wall MA, Socolich M, Ranganathan R (2000) The structural basis for red fluorescence in the tetrameric GFP homolog DsRed. *Nature Structural Biology* 7: 1133-1138.
20. Ai H-w, Henderson JN, Remington SJ, Campbell RE (2006) Directed evolution of a monomeric, bright and photostable version of *Clavularia cyan* fluorescent protein: structural characterization and applications in fluorescence imaging. *Biochemical Journal* 400: 531-540.
21. Wiedenmann J, Ivanchenko S, Oswald F, Schmitt F, Röcker C, et al. (2004) EosFP, a fluorescent marker protein with UV-inducible green-to-red fluorescence conversion. *Proceedings of the National Academy of Sciences of the United States of America* 101: 15905-15910.
22. Gurskaya NG, Verkhusha VV, Shcheglov AS, Staroverov DB, Chepurnykh TV, et al. (2006) Engineering of a monomeric green-to-red photoactivatable fluorescent protein induced by blue light. *Nature Biotechnology* 24: 461-465.
23. Shaner NC, Lin MZ, McKeown MR, Steinbach PA, Hazelwood KL, et al. (2008) Improving the photostability of bright monomeric orange and red fluorescent proteins. *Nature Methods* 5: 545-551.
24. Ando R, Mizuno H, Miyawaki A (2004) Regulated Fast Nucleocytoplasmic Shuttling Observed by Reversible Protein Highlighting. *Science* 306: 1370-1373.
25. Karasawa S, Araki T, Yamamoto-Hino M, Miyawaki A (2003) A Green-emitting Fluorescent Protein from *Galaxeidae* Coral and Its Monomeric Version for Use in Fluorescent Labeling. *Journal of Biological Chemistry* 278: 34167-34171.
26. Karasawa S, Araki T, Nagai T, Mizuno H, Atsushi M (2004) Cyan-emitting and orange-emitting fluorescent proteins as a donor/acceptor pair for fluorescence resonance energy transfer. *Biochemical Journal* 381: 307-312.
27. Fukuchi S, Nishikawa K (2001) Protein surface amino acid compositions distinctively differ between thermophilic and mesophilic bacteria. *Journal of Molecular Biology* 309: 835-843.
28. Akerboom J, Calderón N, Tian L, Wabnig S, Prigge M, et al. (2013) Genetically encoded calcium indicators for multi-color neural activity imaging and combination with optogenetics. *Frontiers in Molecular Neuroscience* 6.

29. Henderson JN, Osborn MF, Koon N, Gepshtein R, Huppert D, et al. (2009) Excited State Proton Transfer in the Red Fluorescent Protein mKeima. *Journal of the American Chemical Society* 131: 13212-13213.
30. Pletnev S, Shcherbo D, Chudakov DM, Pletneva N, Merzlyak EM, 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: 28980-28987.
31. Shu X, Shaner NC, Yarbrough CA, Tsien RY, Remington SJ (2006) Novel Chromophores and Buried Charges Control Color in mFruits<sup>†,‡</sup>. *Biochemistry* 45: 9639-9647.

## CHAPTER 4

### **mGinger: A Soluble Far-Red Fluorescent Monomer Derived from HcRed**

*(The following people contributed to this work: R. Scott McIsaac<sup>2</sup> provided helpful discussion, Kevin S. Brown<sup>3</sup> helped with MSA calculations, and Frances H Arnold<sup>2</sup> provided helpful discussion.)*

<sup>1</sup> *Division of Biology and Biological Engineering, Mail Code 114-96, California Institute of Technology, Pasadena, California, United States of America*

<sup>2</sup> *Division of Chemistry and Chemical Engineering, Mail Code 210-41, California Institute of Technology, Pasadena, California, United States of America*

<sup>3</sup> *Department of Biomedical Engineering, University of Connecticut, Storrs, Connecticut, United States of America*

#### **4.1 Abstract**

*Anthozoa* class red fluorescent proteins (RFPs) are heavily used as biological markers, with far-red emitting variants ( $\lambda_{em} \sim 600 - 900$  nm) sought for whole animal imaging because biological tissues are most permeable to light in this range. However, all known RFPs are natively tetrameric, which is not ideal for cell biological applications. Efforts to engineer monomeric variants for imaging applications have imposed a brightness cost and often produced blue-shifted monomers. Fluorescence is sensitive to structural changes in the protein, and is typically lost when oligomeric interfaces are broken apart. As a result only four native RFPs have been monomerized, leaving the vast majority of RFP biodiversity untapped in biomarker development. Here we report the first monomerization of HcRed, a far-red FP, and describe a comprehensive methodology for the rapid monomerization of novel red-shifted tetrameric RFPs. We first engineered HcRed7 ( $\lambda_{em} = 642$  nm), a dimeric core variant that is brighter, stabilized, and bathochromically shifted, and whose structure helps to shed light on far-red emission. The final designed monomeric variants are called mGinger0.1 ( $\lambda_{em} = 637$  nm) and mGinger0.2 ( $\lambda_{em} = 631$  nm).

## 4.2 Introduction

Monomeric fluorescent proteins (FPs) are preferred as fluorescent biomarkers because the oligomerization of an FP tag can artificially aggregate its linked protein target, altering diffusion rates and interfering with target transport, trafficking, and activity (1, 2). To the best of our knowledge, ~50 native red fluorescent proteins (RFPs) and ~40 chromoproteins (CPs) with peak absorbance in the red or far-red ( $\lambda_{\text{abs}} > 550 \text{ nm}$ ) have been described to date, but most have not been extensively characterized because they are as a class tetrameric (3, 4). These proteins exhibit spectroscopic properties that would be of use in monomeric RFP markers (e.g. far-red excitation/emission and enhanced brightness), but breaking tetramerization without disrupting fluorescence has proven difficult. An underlying biological reason for the observed obligate tetramerization of native RFPs is not well understood, but it has been suggested to be related to proposed photoprotective functions of RFPs (5, 6), and it has been noted that chromoproteins (CPs) and RFPs, which are all obligate tetramers are protected better against reactive oxygen species than green and cyan FPs (7, 8). It is possible, however, that the tetramerization of *Anthozoa* class RFPs plays a basic structural role in stabilizing a chromophore environment that allows for bright and stable fluorescence at red wavelengths. This possibility is supported by the observation that of the five known independent instances in which an RFP has been monomerized (9-12), there has always been either a hypsochromic shift to the maximum intensity emission wavelength ( $\lambda_{\text{em}}$ ) or a decrease in the brightness of the protein, and sometimes both (Table 1). Furthermore, in each of these cases significant mutagenesis of the RFP core and chromophore environment has been necessary to permit fluorescence in the engineered monomer. We recently showed that a protein core that has been optimized to be fluorescent in a monomeric scaffold is sufficient to enable efficient monomerization, which we accomplished via computational design of the RFP  $\beta$ -sheet surface. Changes to the fluorescent properties of the protein were conveyed entirely by mutations to the protein core, which suggested that in future RFP engineering, core optimization will need to be a separate and carefully conducted part of any monomerization process.

Beyond monomericity, RFPs that excite and emit in the far-red or near infrared are in high demand. These wavelengths penetrate tissue with minimal absorption by biological molecules such as hemoglobin, melanin, and water, making bright infrared molecular probes ideal for a variety of imaging applications in live vertebrate hosts (13). RFPs have been useful in this context, and engineered monomeric RFP variants have seen the most widespread adoption because of the benefits that a monomer confers. Efforts to monomerize RFPs though have invariably led to detrimental effects on their fluorescent properties, in addition to the aforementioned loss of brightness or hypochromic shift to fluorescence emission, many engineered monomeric RFPs are less photostable or show disrupted chromophore maturation (14, 15). The current slate of monomeric RFPs consists of a cadre of red to far-red variants that has not been extensively or uniformly characterized and that – when compared with their green- or cyan-emitting cousins – is extremely dim (16, 17). Furthermore, the brightest of these RFPs tend to exhibit the lowest  $\lambda_{em}$ , because as they have been targeted for brightness gains via directed evolution there has been a tendency to accumulate mutations that hypsochromically shift fluorescent emission. In fact there is a clear negative correlation between brightness and bathochromic fluorescent emission among both native and engineered RFPs regardless of oligomeric state (Figure 1). Native RFPs, which are all tetrameric, mostly emit between 575 and 615 nm, and engineered monomeric and oligomeric variants within this spectral range have been optimized to approach the native proteins' brightness. When proteins are engineered for far-red fluorescent emission, however, brightness drops off significantly faster among monomeric proteins than among higher-order oligomers, although there is a clear decline in brightness among the latter group as well. The biophysical reasons for this trend are not understood, and many of these far-red proteins have been poorly characterized, with oligomeric state in particular subject to frequent mischaracterization (1, 11).

RFP engineering has mostly focused on the continued development of previously targeted proteins, leaving the vast majority of known RFP biodiversity untapped. Previous efforts to monomerize native RFP tetramers relied on directed evolution, with intuition-based disruption of oligomeric interfaces followed by random mutagenesis to recover fluorescence. Though partially successful, these processes are long and labor-intensive, and

furthermore the spectroscopic character of the parental RFP has never been preserved in the monomeric derivative. Fluorescence is knocked out in native RFPs after the introduction of mutations that break a tight dimeric interface, necessitating recovery of the lost fluorescence by error-prone mutagenesis, which has invariantly introduced protein core and chromophore-proximal mutations. By requiring mutation to the chromophore environment for successful monomerization, it has not been possible to exert any significant degree of control over the spectroscopic properties of the resultant monomer. To illustrate this point, the furthest red parental protein of any successful monomerization attempt to date is mKate2 (18), with a  $\lambda_{em}$  of 630 nm. But in monomerizing the protein, 7 core positions were mutated, leaving the monomer FusionRed significantly hypsochromically shifted, with a  $\lambda_{em}$  of 608 nm (11) (Table 1).

Here we present a comprehensive engineering strategy for the monomerization of novel RFPs that focuses on separating core optimization from surface design. This allows the screening of a diverse set of RFP cores for optimized variants that better tolerate monomerization. This strategy allows us to fully monomerize an RFP for the first time without mutation to the protein core. We chose to engineer HcRed, a far-red engineered dimer/tetramer (19) ( $\lambda_{em} = 633$  nm) with the dual goal of monomerization and retention of far-red  $\lambda_{em}$ . We generated mGinger0.1, the most red-shifted first generation monomeric RFP to date; however, unlike our previous work in which an optimized RFP core allowed monomerization with little change to the parent tetramer's spectroscopic properties, the monomerization of a core-optimized HcRed caused a hypsochromic shift to the  $\lambda_{em}$  and a significant loss of brightness. Still, mGinger0.1 and mGinger0.2, named for their bright red coloring, fit along the monomeric RFP brightness frontier (Figure 1), and are among the brightest and furthest red-shifted monomeric RFPs engineered to date (mGinger0.1:  $\lambda_{em} = 637$  nm,  $\Phi = 0.02$ ; mGinger0.2:  $\lambda_{em} = 631$  nm,  $\Phi = 0.04$ ). To contribute to a better understanding of RFP engineering, and learn some lessons from the design process, we conducted a structural study of the monomerization of HcRed. Here we present the structures of a red-shifted, core-optimized dimer; we are close to obtaining a structure of the final monomers. We successfully engineer red-shifting chromophore interactions into the dimer that we believe are partially maintained after monomerization. This structural

study will help to shed some light on the reasons behind hypsochromic shifts and brightness loss in RFP monomers and suggests strategies for designing next generation markers.

### 4.3 Results

#### *Small Perturbations to Oligomerization Knock Out Fluorescence in HcRed*

HcRed is a 9-point mutant of the chromoprotein HcCP (19). Two mutations were made to its core while engineering the protein: Cys143Ser is a known *cis* chromophore stabilizing mutation that induced fluorescence in HcCP, and Leu173His was found during error prone mutagenesis. Additional mutations were made to the surface to disrupt oligomerization, but HcRed remains dimeric with a slight tetrameric proclivity (Figure S1). Its AC interface remains intact, while its AB interface is partially broken; these interfaces are named for the ‘A’, ‘B’, and ‘C’ chains from the first structure of DsRed (20). The AC interface is the more stable and difficult to break of the two dimeric interfaces, and it buries a defect in the RFP beta barrel: a protrusion in the surface of the  $\beta$ -barrel at  $\beta$ -strand 7 is caused by the phenolate ring of the chromophore (Figure 2B). To test the ability of HcRed to sustain perturbation to its AC interface, we made successive deletions to its C-terminal tail, which stabilizes the interface with an intermolecular loop- $\beta$ -sheet interaction (Figure 2A). Like other native RFPs, HcRed does not tolerate a C-terminal tail deletion, losing significant brightness with the deletion of just one C-terminal residue, and completely losing detectable fluorescence with a deletion of five C-terminal residues (data not shown). This loss of fluorescence mirrors the loss of fluorescence seen during efforts to break HcRed oligomerization through perturbations to the surface of the AC interface. The intolerance of HcRed to a tail deletion suggested that it was not optimized for fluorescence in a monomeric scaffold, and so optimization of the chromophore environment would be necessary prior to monomerization.

### ***Optimization of the HcRed Core Leads to a Brighter RFP Dimer***

To pick sites to target for mutagenesis in the protein core we constructed an alignment of well-characterized, bright, and far-red monomeric and dimeric RFPs along with their native tetrameric parents (Table S2). Residues that showed diversity in the alignment or that were frequently mutated in engineered monomers and dimers were selected for mutagenesis. We then designed a library that sampled amino acids that appeared frequently in the alignment, or that were of interest because they contributed to known red-shifting chromophore interactions.

The first mutational hotspot we identified was a group of residues that surrounds the two alternative conformations of the phenolate side chain of the chromophore, which are visible in HcRed's crystal structure (Figure 3A) and were confirmed to exist in single-molecule studies. HcRed was engineered from a non-fluorescent chromoprotein (HcCP) with the introduction of a cysteine to serine mutation at position 143 that stabilizes an alternative, fluorescent conformation of the chromophore with a hydrogen bond to its phenolate oxygen. This alternative conformation is called the *cis* chromophore, as the phenolate group sits *cis* to the proximal nitrogen on the imidazolinone ring rather than *trans* to it. In most known RFPs and red-absorbing chromoproteins (CPs), the *cis* chromophore is the fluorescent species, whereas the *trans* chromophore is non-fluorescent. As HcCP is a native chromoprotein with a chromophore pocket naturally evolved to stabilize the non-fluorescent *trans* chromophore, we designed a first core library (cLibA) to increase brightness in HcRedAB1 by comparatively stabilizing the *cis* over the *trans* chromophore. We targeted *trans*-stabilizing residues for mutation (Asn158 and His173 are hydrogen bond donors to the phenolate oxygen), and hoped to place bulkier side chains in the *trans* pocket. A total of 14 mutations were allowed at Ser143, Asn158, Met160, His173, and Thr175, for a theoretical total library size of 448 (Figure S1).

We used a second core library (cLibB) to target a structural region that lies above the chromophore, between the central  $\alpha$ -helix and the unbroken AC oligomeric interface (Figure 3B). The reason for the high occurrence of mutations found in this region when monomerizing RFPs is unclear. An important structural feature may provide a clue, in that

there is a channel populated by structural waters that stretches from the protein surface at the top of the  $\beta$ -barrel significantly into the center of the protein. Thus instability in this region caused by a break in the AC oligomeric interface may open up a water channel from the bulk solvent to the chromophore, which would quench the chromophore cyclization reaction. Alternatively, the break in oligomerization may interfere with the placement of arg67, which is a key catalytic residue found in this region that is responsible for abstracting a proton from the bridging carbon of the phenolate side chain during chromophore maturation. A total of 18 mutations were allowed at Gly28, Met41, Arg67, Thr68, Phe80, Asn158, Ile196, and Ser215, for a theoretical total library size of 432 (Figure S1). Two chromophore-backing positions (Gly28 and Met41) were included in this library, because they have been shown to be important in maturation and far-red fluorescence (14, 21, 22).

We screened each library to >95% coverage on large LB agar plates supplemented with IPTG, which produced colonies of various colors ranging from pink to red to purple. We picked colonies that were brightly colored, sequenced them and expressed and characterized 16 cLibA variants and 21 cLibB variants *in vitro*. The variants had a surprisingly large diversity of spectroscopic properties including improved brightness of up to 10-fold over HcRed and a range of  $\lambda_{em}$ 's from 606 to 647 nm. To determine whether the collection of improved core variants included cores that were stabilized in the context of a monomer, we deleted the five C-terminal residues from each characterized core variant. HcRed does not tolerate this tail deletion, but six of the core variants tolerated the deletion well, and of these a double mutant R67K and H196Y, which we call HcRed7, was the most red-shifted ( $\lambda_{em} = 642$  nm). The core mutations in HcRed7 bathochromically shift its emission by 9 nm, improve  $\Phi$  by 60%, and thermostabilize the protein by 6 °C over HcRed. This stabilized core allows it to partially tolerate a C-terminal tail deletion. HcRed7 maintains most of its brightness with a deletion of up to 5 C-terminal residues (HcRed7 $\Delta$ 5), but loses significant brightness with the deletion of a sixth C-terminal residue (HcRed7 $\Delta$ 6), indicating that the core is partially but not wholly optimized for monomerization. With no obvious rational design strategy for improving the fluorescence of the tail-deleted variants,

we used a directed evolution strategy to recover the dim but detectable fluorescence of a 6-residue tail-deleted HcRed7 mutant.

### ***Enhancing HcRed7 $\Delta$ 6 Fluorescence via Directed Evolution.***

HcRed7 $\Delta$ 6 is significantly thermally destabilized compared to HcRed7, losing 16°C of thermal stability with the tail deletion (Figure 5). We reasoned that improving the thermal stability of HcRed7 $\Delta$ 6 might increase its brightness. To this end we screened two libraries in parallel. The first was a library built with error-prone PCR (mLibEP), which randomly mutagenized the gene, while the second was a consensus library (mLibC) in which we mutated positions to the consensus sequence from an FP alignment. The consensus mutation strategy has been shown to be effective for engineering thermostable protein variants, but requires a large, well-constructed multiple sequence alignment (MSA) and a good genetic distance algorithm to be effective. To generate the consensus FP sequence we made an MSA that consisted of every *Aequorea victoria* class FP, a total of 741 sequences. From this MSA, a consensus sequence was generated (see supplemental Methods), and we designed mLibC to sample all 105 non-consensus positions in HcRed with the consensus amino acid. The error rate of mLibC and mLibEP were carefully tuned to a rate that would allow significant variation while retaining fluorescence in most variants. We screened mLibC at 1.2 mutations per variant and mLibEP at 1.8 mutations per variant. Induced bacterial cultures from each library were screened in 96-well format for bright fluorescent emission at 675 nm to give a selective advantage to mutants that maintained a red-shifted  $\lambda_{em}$ . Screening at 675 nm allowed us to maximally differentiate between HcRed7 mutants peaking around 630 nm and those hypsochromically-shifted variants that often peaked between 605 nm and 620 nm. mLibC was screened to 40x coverage (~4300 clones) and ~8600 clones were screened from mLibEP. We isolated 14 unique hits from mLibC, and 26 from mLibEP. Guided by the variations in the brightness and  $\lambda_{em}$  of each variant, and considering the location of each mutation on the protein, we made four chimera constructs that combined hits from each library. We assembled consensus chimeras that included either four (HcRed74) or seven (HcRed77) of the top mutations isolated from the consensus library, and error prone chimeras that split the 29 isolated error prone mutations

between a fourteen-fold variant (EPchimera1) and a fifteen-fold variant (EPchimera2). The brightest of these synthetic proteins was HcRed77. All of the brightness lost in deleting the tail from HcRed7 was recovered in HcRed77, without any hypsochromic shift to the  $\lambda_{em}$  of HcRed7 $\Delta$ 6, although HcRed77 remained hypsochromically shifted from HcRed7.

### ***A Fluorescence-Stabilized HcRed77 is Monomerized***

Despite the partial sensitivity of HcRed7's core to tail deletion, we investigated if HcRed77 would tolerate full monomerization. By size exclusion chromatography (SEC), HcRed77 is still dimeric (Figure S1), although it appears to have lost the trace tetramerization visible in HcRed7, indicating that the tail deletion did indeed partially destabilize oligomerization. To fully monomerize HcRed77, we designed the remaining intact dimeric AC interface with a computational protein design (CPD) procedure that we had used successfully in previous work with RFPs. We targeted a set of five residues (Val146, Val159, Ile170, Phe191, and Phe193) in the heart of the AC interface that make extensive intermolecular contacts and which we found were frequently mutated in past FP monomerizations (Figure 2C and 2D). We decided to limit the design to these five crucial residues both for ease of library construction and screening and to limit synthesis costs. We made a small combinatorial library to sample the top design hits at each of the five positions using degenerate codons, requiring just four custom degenerate primers. The library was screened on large agar plates supplemented with IPTG, and colored colonies were picked after two days of expression at room temperature. We saw faint color in only one colony from the screen, but this colony proved to be the first generation HcRed monomer, which we called HcRedm1. HcRedm1 was verified to be monomeric by analytical ultracentrifugation (AUC) (Figure S1). The protein is very dim, and poorly expressed, but doesn't see a large hypsochromic shift to its fluorescence. The extremely poor fluorescence of HcRedm1 shows that the core of HcRed77 is not perfectly adaptable to monomerization without additional stabilizing mutations. However, as HcRedm1 is further red-shifted than any other first generation RFP monomer, we thought it would be a good candidate for improvements to its brightness through further directed evolution. We had thus far made only one additional mutation to the core of HcRed7 (A59S), and a reversion of this mutation did not significantly impact

the fluorescence or expression of HcRedm1. To keep the core of HcRed7 intact so that we could directly ask what impact monomerization had on far-red emission and brightness, we moved forward with a variant of HcRedm1 with the S59A reversion.

### ***The Brightness of HcRedM1 is Improved with DNA Shuffling and Error-Prone Mutagenesis***

We sought to improve the brightness of HcRedm1 with a DNA shuffling library that incorporated mutations identified from the error prone library of HcRed7 $\Delta$ 6. This represented a pool of mutations that had been shown to be beneficial, but that had not been incorporated into HcRedm1. The error prone library had identified 26 mutations, some of which occurred to the same position on the protein, and many of which were close together in primary structure. The proximity of the mutations did not allow them to be efficiently shuffled if they were on the same parental strand of DNA as our DNA shuffling procedure cuts DNA into ~30bp fragments. We therefore synthesized two chimeric HcRed variants and distributed the mutations between the two constructs, which we named chimeraA and chimeraB. Following DNA shuffling, we screened ~750 clones by 96-well plate expression, and isolated 18 variants that were brighter and red-shifted. We chose to fully characterize two variants, one of which had significantly recovered brightness (HcRedm13), and one of which was only marginally brighter, but slightly red-shifted (HcRedm14). HcRedm13 has seven mutations, three of which are on external  $\beta$ -strands, with the rest on loops at the two ends of the protein. HcRedm14 has 12 mutations, one of which, Ile43Val, is a chromophore backing mutation in the protein core, four are on surface  $\beta$ -strands, and the remainder are on loops. The core mutation in HcRedm14 is responsible for the bathochromic shift in emission, as a reversion of this mutation does not exhibit the same spectral shift (data not shown).

HcRedm13 and HcRedm14, however display indications of inhibited chromophore maturation. Both proteins exhibit multiple absorbance peaks, a sign that they do not fully mature to the red chromophore. Additionally, once purified into PBS, neither protein is stable for long at room temperature, losing pigment and yellowing. We reasoned that we may have lost some stability by screening protein variants at 30°C instead of 37°C, and so

both variants were then subjected to a subsequent round of directed evolution by error-prone mutagenesis with screening at 37°C. Large libraries of ~4,000 clones were screened for both HcRedm13 and HcRedm14 by 96-well protein expression. Ten mutations were found between the two proteins that increased their fluorescent emission at 675 nm in culture. A final round of DNA shuffling was then done to incorporate in these new mutations as well as to resample the diversity in HcRedm13 and HcRedm14 at 37°C. We synthesized a chimeric variant of HcRedm13 that included the ten new error-prone mutations, which we called chimeraC. We then shuffled HcRedm14 and chimeraC, and screened at 37°C. After screening about 750 clones we isolated two variants that we called mGinger0.1 and mGinger0.2. The mGingers are first generation monomeric RFPs that express well at 37°C, mature completely to the red chromophore, and have a  $\lambda_{em}$  of 637/631 nm and a brightness of 1.2/1.5 when excited with far-red light. The two proteins fall right along the brightness- $\lambda_{em}$  boundary for monomeric proteins.

### ***Engineered HcRed Variants Show That Thermal Stability is Linked to Brightness***

As we engineered HcRed, we measured the thermal stability of the designed variants. Fluorescence was repeatedly knocked down and improved while engineering HcRed monomers, and we noticed that the thermal stability of HcRed variants (the degree to which the fluorescence of a purified protein changes with an increase in temperature) was well correlated to brightness. We measure the thermal stability of the RFP variants by ramping up the temperature while measuring fluorescence in real-time, to determine an apparent temperature at which fluorescence is irrecoverably lost (apparent  $T_m$ ). Interestingly, the mutations during the engineering process that improved brightness also improved the apparent  $T_m$ , including HcRed  $\rightarrow$  HcRed7, HcRed7 $\Delta$ 6  $\rightarrow$  HcRed77, and HcRedm1  $\rightarrow$  mGinger0.1, while those that decreased brightness also decreased the apparent  $T_m$ , including HcRed7  $\rightarrow$  HcRed7 $\Delta$ 6, and HcRed77  $\rightarrow$  HcRedm1 (Figure 5A). We further observed that there is a positive correlation between quantum yield and apparent  $T_m$  for all of these variants ( $R^2$  of dimers = 0.94; monomers = 0.14) (Figure 5B). The correlation appears to divide into two distinctly correlated groups: the dimeric species have higher quantum yields at higher apparent  $T_m$ 's than do the monomeric species. Significantly,

mGinger0.1 and mGinger0.2 are  $\sim 5$  °C thermostabilized over the parental protein, HcRed7, but are both dimmer. It is not clear why this is the case, but clearly monomerization has an important impact on RFP fluorescence.

## 4.4 Discussion

### *The RFP Monomer Design Process*

Native oligomeric proteins have most likely had selective pressure to oligomerize, and their functionality has evolved in the context of an oligomeric scaffold. It is an oversimplification to think that monomerizing a protein is as simple as slipping large polar or charged residues into a hydrophobic interface. We present here a comprehensive design process for the monomerization of RFPs that can equally well be applied to other oligomeric enzymes or signaling proteins. Past monomerization efforts have ignored the role that core-stabilization plays in engineering a soluble monomer. Together with previous work, we show here that an evolved oligomeric protein does not necessarily function well as a monomer without significant optimization. We suggest that this could be due to a loss of overall structural integrity, as breaking apart tightly bound oligomers can leave an individual monomeric subunit without the evolved structural support from its oligomeric mate. Here, with HcRed, we first attempted to evolve a stabilized protein core and subsequently tested a diverse HcRed core library with perturbation to its oligomerization. A two-pronged rational design strategy was used to improve HcRed's core: first, by comparatively stabilizing the *cis* vs the *trans* chromophore we hoped to increase the apparent quantum yield of the protein, and second, through repacking of a key internal structural region, we expected to create HcRed variants with improved structural integrity. This strategy worked reasonably well, and the improved core was enough to get us to a monomeric protein, but we did not fully succeed in designing a core solid enough to tolerate full monomerization without additional stabilizing mutations.

As it may be difficult *a priori* to fully optimize a protein to maintain its functionality as it is monomerized, we find that it is instrumental to take modest steps toward monomerization while maintaining a functional handle for screening and optimization. Importantly, we

maintained some minimal fluorescence in the HcRed variants throughout monomerization as a handle to guide the evolution of variants that were more stable and showed improved brightness. This allowed us to be stringent in our selections and maintain HcRed7's far-red  $\lambda_{em}$ . Additionally, consensus design has been described as a method to add thermostability to proteins, and we found that it was a useful tool for stabilizing the HcRed variants, allowing recovery of fluorescence of the tail-deleted HcRed7 $\Delta$ 6, and significantly outperforming random mutagenesis (Figure S2). We also found temperature to be an important consideration during evolution and screening, as evolution at low temperatures (30°C) significantly thermally destabilized the HcRed variants, interfering with cyclization, while evolution at 37°C quickly recovered thermal stability. Finally, a very successful strategy for us was to use successive rounds of error-prone mutagenesis interspersed with DNA shuffling (23, 24), which served to efficiently sample every error-prone hit. This strategy allowed us to use only two rounds of error-prone mutagenesis in the engineering of mGinger0.1. We screened ~4,000 variants in each round of error-prone and ~750 variants in each round of DNA shuffling. This is a comparatively efficient directed evolution approach, as we incorporated 15 error-prone mutations after only four rounds (two error-prone, two shuffling) and shuffled variants were marked improvements over individual error-prone hits.

### ***The Structure of HcRed7 Reveals the Mechanism For Its Increased Brightness and Bathochromically Shifted $\lambda_{em}$***

We solved an x-ray crystal structure of HcRed7, which shows that the mutation from histidine to tyrosine at position 196 serves to add a  $\pi$ -stacking interaction with the chromophore phenolate ring (Figure 3C). Tyr196  $\pi$ -stacks with the fluorescent *cis* orientation of the phenolate, serving to both stabilize the fluorescent chromophore over the *trans* phenolate (in wild-type HcRed the chromophore occupies both *cis* and *trans* conformations), and to red-shift the  $\lambda_{em}$ , as a  $\pi$ -stacking phenolate interaction has been shown to reduce the energy of the excited state of the chromophore (25-27). In turn, position 67 is a key catalytic residue that functions as a base, abstracting a proton from the bridging carbon of the phenolate side chain during cyclization (28, 29). This residue is

almost invariably a lysine or arginine in the alignment of RFPs, and we propose that the mutation from arginine to lysine here allows room for the  $\pi$ -stacking interaction and the bulkier tyrosine side chain. A bathochromic shift from this  $\pi$ -stacking interaction has been demonstrated, but here we note that a 6 °C improvement to apparent  $T_m$  and a 60% improvement in quantum yield accompany this engineered interaction, which has not been shown before.

### ***Monomerization of mGinger0.1 Hypsochromically Shifts the $\lambda_{em}$ Without Mutation to the Protein Core***

We had expected to stabilize HcRed's protein core enough that it would tolerate monomerization with little to no change in spectroscopic properties, as we had demonstrated in previous work. This did not happen, however, and even with minimal mutation to the protein core, we saw a significant loss of brightness and a hypsochromic shift to  $\lambda_{em}$  in moving from the dimer to the monomer. This has been seen in past monomerization efforts, but because these past efforts have all involved significant mutation to the core of the protein (9, 11, 30-32), it had been difficult to separate the effects of core mutations from the true effects of monomerization, which may include a loss of scaffold rigidity or increased hydration of the protein core (33).

One clue into scaffold rigidity is to measure the thermal stability of HcRed variants. It has been suggested that quantum yield is linked to the structural rigidity of an FP's excited-state chromophore (16, 34). The more rigid the excited-state of the chromophore, the higher the quantum yield, as thermal motion can lead to non-radiative decay via other atomic interactions. This is supported by the fact that the quantum yield of small molecule fluorophores increases with decreased temperature (35), and holds true even in proteins, as a cyan FP with a quantum yield of 0.93 was engineered with rational design by stabilizing a  $\beta$ -strand near the chromophore (16). The apparent  $T_m$  of mGingers were improved by ~5 °C over HcRed7, ~10 °C over HcRed, and ~13 °C over HcRed77, their nearest dimeric ancestor. The cause of this thermal stabilization is unknown, but we presume that it had a large impact on the brightness of these variants as far-red monomers. It is puzzling though that a less-thermally stable dimer should be brighter than its thermostabilized monomeric

derivative, despite sharing an almost identical protein core, and this could indicate that thermostabilizing mutations were permissive to the evolution of monomerization (36).

### ***Monomerizing Mutations Accumulate in Key Structural Regions***

The mutations found during the design/evolution of mGinger0.1 and mGinger0.2 help further our understanding of the important structural regions in RFPs that might be targeted during future RFP monomerization. The core of the mGingers was not mutated except for the chromophore backing position Ile43Val (The C $\gamma$  of Ile43 lies 6.5 Å from the chromophore glutamate), which was responsible for maintaining a bathochromic shift to emission in the HcRedm1-chimeraA/B shuffling library, and three positions found on the periphery of the  $\beta$ -barrel. One of these peripheral core mutations, His72Tyr (12.3 Å from the chromophore), was a spontaneous mutation that appeared in the only fluorescent monomer hit from the surface design of HcRed77 to HcRedm1. This position was reverted, but found to be beneficial. The last two core mutations: Ser137Thr and Met200Ile (13.5 and 8.3 Å from the chromophore) are part of a structural deformity in the  $\beta$ -barrel that was mentioned earlier, and is buried by the AC interface (Figure 2B). Residues 135-139 are part of the deformed  $\beta$ -strand punched out by the *cis* chromophore and residues 199-203 are on a neighboring  $\beta$ -strand that is distorted by an inter-molecular interaction with the neighboring monomer's C-terminal tail. These deformities leave a wide gap between the two  $\beta$ -strands that is populated by 11 stationary water molecules in the structure of HcRed7 (Figure 6A). Of the ten residues that make up the distorted  $\beta$ -strands, eight are mutated in the mGingers. Three neighboring residues on the surface of the  $\beta$ -barrel that interact with the C-terminal tail of the neighboring monomer are also mutated (Arg197, Tyr212, and Ala214), two of which, Arg197 and Ala214 were identified by the grid-map of the AC interface but not designed as part of the AC surface design of HcRed77 to HcRedm1. This large structural region is a mutational hotspot, and is clearly strongly tied to improving the integrity of the monomeric scaffold and protecting the chromophore environment from bulk solvent in the absence of the stabilizing effects of oligomerization. Three more positions mutated (Cys155, Tyr174, and Ser176) are on the outside of the  $\beta$ -barrel near the fringe of the AC interface. The rest of the positions mutated in the mGingers are on loop

regions and the surface of the  $\beta$ -barrel away from the two oligomeric interfaces, and probably affect protein solubility. The clearest conclusion from this mutational data is that a larger proportion of the AC interface would benefit from explicit design during future monomerization attempts, and special focus should be given to the  $\beta$ -barrel defect between  $\beta$ -strands 7 and 10.

## 4.5 Conclusion

We monomerized HcRed, a far-red fluorescent protein that had been the target of previous unsuccessful monomerization attempts. The rational approach that we lay out in monomerizing HcRed should be repeatable in the future with other novel RFP oligomers that have interesting spectroscopic properties. We did not successfully stabilize the core of HcRed enough so that it tolerated monomerization unperturbed, but did find a core that was stabilized enough to be adaptable to monomerization with the addition of beneficial surface mutations. The design process that we lay out includes elements of intuition-based design, computational design, and directed evolution, which are all tools that skilled protein engineers will need to be able to incorporate to push the field of protein engineering. We notice an interesting correlation between brightness and thermal stability among HcRed variants, which has been suggested by other work, but to our knowledge never explicitly shown. Finally, we identify a key structural region whose stabilization greatly impacts monomerization and conclude that a more extensive design of the surface of the RFP AC interface may be helpful in future RFP monomerization efforts. Future engineering of RFP cores will be necessary to determine how to significantly improve brightness post-monomerization.

## **4.6 Materials and Methods**

### ***Plasmids and Bacterial Strains***

The HcRed sequence was taken and modified from the HcCP Genbank entry (accession number AF363776). Ten amino acids were added to the N-terminus, consisting of a Methionine followed by a 6x Histidine tag for protein purification, and then followed by a Gly-Ser-Gly linker sequence. All gene sequences were constructed with gene assembly PCR, oligonucleotides for the assembly were designed with DNAworks, and then ordered from Integrated DNA Technologies (IDT). Assembled genes were PCR-amplified and cloned into the pET-53-DEST expression plasmid (EMD Millipore) with PIPE cloning followed by CPEC. Constructs were sequence-verified (Laragen) with a primer specific to the T7 promoter, and then transformed into BL21-Gold(DE3) competent cells, a protein expression strain (Agilent).

### ***Construction of HcRed Designed Libraries***

To construct explicitly-designed HcRed core libraries, protein sequences were input into DNAworks as “mutant runs” of the wild-type HcRed gene assembly. This allows explicit libraries of gene variants to be assembled and minimizes the number of oligonucleotides needed. The AC surface library was designed using degenerate codons that code for the amino acids found in the designed sequence variants. The triplet codon “VRN” was used to code for all residues found in the sequence design. Oligonucleotides were ordered from IDT and cloning was carried out as described above.

### ***Error Prone Mutagenesis***

Error prone mutagenesis of HcRed variants was performed by addition of manganese chloride to Taq DNA polymerase PCR reactions. 10 $\mu$ M, 15 $\mu$ M, and 20 $\mu$ M MnCl<sub>2</sub> were tested and cloned with PIPE cloning into pET-53-DEST for sequencing. Twelve colonies from each library were picked and sequenced, and the library with a mutation rate closest to but not more than 2.0 mutations per gene was selected for further screening.

### ***DNA Shuffling***

The variants that were to be shuffled together were PCR-amplified run on gel electrophoresis, visualized with Gel Red, cut out of the gel and then purified via a standard spin-column gel purification kit (Qiagen). 5 µg of the purified DNA fragments were then cut with 0.5 U of DNaseI (NEB) in a 50 µl reaction. The reaction was allowed to sit for 7.5 minutes at room temperature and then quenched with 5 µl of 100 mM EDTA (4x the concentration of MgCl<sub>2</sub> in the reaction buffer). The reaction was further heat-inactivated for 10 minutes at 90°C in a thermocycler and run on gel electrophoresis. Bands were cut out of the gel that ran at ~30 bp. We used a 30 bp primer (IDT) and a 100 bp DNA ladder (NEB) as standards. These gel slices were frozen and then purified using a Freeze 'N Squeeze gel purification kit (BioRad) because a typical spin column will not efficiently bind 30bp DNA. Purified digested fragments were mixed together at a 1:1 ratio and assembled via Gene Assembly PCR as discussed above.

### ***Protein Expression and Library Screening***

Single bacterial colonies were picked with sterile toothpicks and inoculated into 300 µl of Super Optimal Broth (SOB) supplemented with 100 µg/ml ampicillin in 2 ml deep-well 96-well plates (Seahorse Biosciences) The plates were sealed with microporous film (Denville Scientific) to facilitate gas exchange during growth. Cultures were grown overnight at 37 °C / 300 RPM. The next morning 800 µl of fresh SOB with 100 µg/ml ampicillin and 1mM Isopropyl β-D-1-thiogalactopyranoside (IPTG) was added to a total volume of 1 ml (evaporation losses overnight are approximately 100 µl). Plates were then shaken 12 hours at either 30°C or 37 °C and 400 RPM. Cell cultures turn red if there is strong RFP expression. After overnight expression, plates were screened with a Tecan robotic liquid handling robot with robotic arm linked to a Tecan platereader Saffire 2. 200 µl of each culture was added to Greiner UV-Star 96-well plates and imaged for fluorescence emission at 675 nm after excitation at 600 nm. Controls were included on each plate to account for plate to plate variability. To re-screen potential hits from the initial screen, a sterile toothpick was dipped into the bacterial culture and the potential hit was streaked out onto a fresh LB plate supplemented with ampicillin. The plate was grown overnight at 37°C, and four colonies were picked for each potential hit. These were then grown again and screened

as detailed above, with hits then ranked on their significant variation from the parent or control.

### ***Protein Purification***

To further characterize important variants, 1 L of SOB in Fernbach flasks were supplemented with ampicillin were inoculated with overnight growth of bacteria and induced at 37°C for 12 hours with 1mM IPTG. After protein expression, the broth was then transferred to centrifuge flasks and spun at 5,000 x g in a fixed angle rotor for 10 min, whereupon the supernatant was decanted. Pellets were resuspended in 25 ml of lysis buffer (50 mM sodium phosphate, 150 mM NaCl, 0.1% v/v Triton-X, pH 7.4) supplemented with 50 Units/ml Benzonase and 0.05 mg/ml Hen Egg Lysozyme. Resuspended pellets were then run over a microfluidizer to fully lyse the bacteria. To pellet down the cellular debris, the lysed cultures were again centrifuged for 10 minutes at 15,000 x g in a fixed angle rotor. The colored supernatant was then applied to His-Select resin (Sigma) in a hand-poured column, washed twice (50 mM sodium phosphate, 150 mM NaCl, 15 mM Imidazole, pH 7.4), and eluted with 500 µl elution buffer (50 mM sodium phosphate, 150 mM NaCl, 250 mM Imidazole, pH 7.4). Proteins were then polished by running over a Superdex 75 10/300 column on an AKTA fast protein liquid chromatography (FPLC) instrument and in the process buffer exchanged into PBS.

### ***Fluorescent Protein Characterization***

Purified protein variants were assayed in triplicate in Greiner UV-Star 96-well plates with a Tecan Saffire 2. An absorbance scan (260 – 650 nm), a fluorescence excitation scan (500 – 640 nm excitation / 675 nm emission), and a fluorescence emission scan (550 nm excitation / 575 – 800 nm emission) were run on 100 µl of eluted protein to determine spectral peaks.

To measure the quantum yield we diluted each protein so that the absorbance for 200 µl of protein at 540 nm was between 0.1 and 0.5. We then measured the  $A_{550}$  in triplicate (or duplicate if it was a poorly expressed protein), diluted the sample to an  $A_{550}$  of 0.04 and took an emission scan (540 nm excitation / 550 – 800 nm emission). The area under the emission curve was calculated after fitting it to a 4<sup>th</sup> order Gaussian, and the quantum yield was calculated with the following formula:

$$\Phi_X = (A_S / A_X)(F_X / F_S)(n_X / n_S)^2 \Phi_S$$

Where  $\Phi$  is quantum yield, A is absorbance, F is total fluorescent emission (area under the curve), and n is the refractive index of the solvents used. Subscript X refers to the queried substance and subscript S refers to a standard of known quantum yield. It is important that the standard be excited with the same wavelength of light as the unknown sample. We use DsRed, which has a known quantum yield of 0.79 as the protein standard.

To measure extinction coefficient we took 100  $\mu$ l of the protein solution that had been diluted to an  $A_{550}$  of between 0.1 and 0.5 and measured absorbance between 400 nm and 700 nm in triplicate. We then added 100  $\mu$ l of 2M NaOH to each well and remeasured absorbance between 400 nm and 700 nm. The base-denatured chromophore, which peaks at approximately 450 nm has a known extinction coefficient of 44,000  $M^{-1}cm^{-1}$ . Then to calculate the extinction coefficient is calculated with the following formula:

$$\epsilon = A_{Chromophore} * 44,000 M^{-1} cm^{-1} / A_{450}$$

### ***Thermal Stability***

Purified proteins were diluted to an absorbance of 0.2 at the wavelength of maximum absorbance ( $\lambda_{abs}$ ) so that their fluorescence would not saturate the rtPCR detector. 50  $\mu$ l of each purified protein was then loaded into a 96-well PCR plate and covered with clear optical tape. The proteins were incubated at 37°C for 10 minutes and then the temperature was ramped at 0.5°C every 30 seconds up to 99°C, with fluorescence measured every ramp step in a CFX96 Touch Real-Time PCR Detection System (Bio-Rad). We refer to this as a thermal melt. The derivative curve of the thermal melt finds the inflection point of the slope, which is the apparent temperature at which fluorescence is irrecoverably lost (apparent  $T_m$ ).

### ***Oligomeric Determination***

(A) Size exclusion chromatography. 100  $\mu$ l of each purified protein analyzed was run over a Superdex 75 10/300 size exclusion column with 25 ml bed volume on an AKTA from GE Life Sciences. Absorbance was measured after passage through the column at 575 nm, where the red chromophore absorbs. (B) Analytical ultracentrifugation. Purified protein

samples were diluted to an  $A_{575}$  of 0.5 for a path-length of 1.25 cm. These samples were put into two-channel sedimentation velocity cuvettes with the blank channel containing PBS. Sedimentation velocity was run at 40,000 RPM overnight with full  $A_{575}$  scans collected with no pause between reads. Data was loaded into Sedfit and a  $c(m)$  distribution was run with default assumptions made for PBS buffer viscosity. After integration, the  $c(m)$  curve was exported to Excel. (C) Homo-FRET. 200  $\mu$ l of each purified protein was diluted to an Absorbance of 0.1 to 0.5 at 530 nm in 96-well Greiner UV-Star plates. Polarization scans were then taken with excitation at 530 nm and emission at 610 nm in a Tecan Safire2 plate-reader. Rose Bengal was used as a standard to calculate the instrument G factor ( $mP = 349$ ).

### *Crystallography*

Rectangular plate crystals of HcRed7 grew in 7 days by the sitting-drop vapor diffusion method in 100 mM Bis-Tris pH 6.5 with 200 mM ammonium sulfate and 25% w/v PEG 3350. Crystals were flash frozen in 2-Methyl-2,4-pentanediol (MPD) and shipped to beamline 12-2 at the Stanford Synchrotron Radiation Lightsource, where a 1.63 Å data set was collected. Phases were obtained through molecular replacement using the crystal structure of HcRed (PDB ID 1YZW).

Following molecular replacement, model building and refinement were run with COOT and PHENIX (37, 38). NCS restraints were applied to early refinement steps and removed at the final stages of refinement. TLS parameters were used throughout. The chromophore was initially left out of the refinement and added at a later stage when clear density became evident for it. Coordinates were deposited in the Protein Data Bank with the code XXXX. Data collection and refinement statistics are listed in Table S1.

## **4.7 Acknowledgments**

We thank Jens Kaiser and Pavle Nikolovski at the Molecular Observatory at Caltech for help with protein crystallization and discussions about refinement. We had helpful conversations about the project with Kurt Mou, Alex Nisthal, and Jan Kostecki, who provided good insight. We would like to further thank the Gordon and Betty Moore Foundation and the National Institute of Biomedical Imaging and Bioengineering for funding this work.

## 4.8 Tables and Figures

Table 1 – Previously monomerized first generation RFPs.

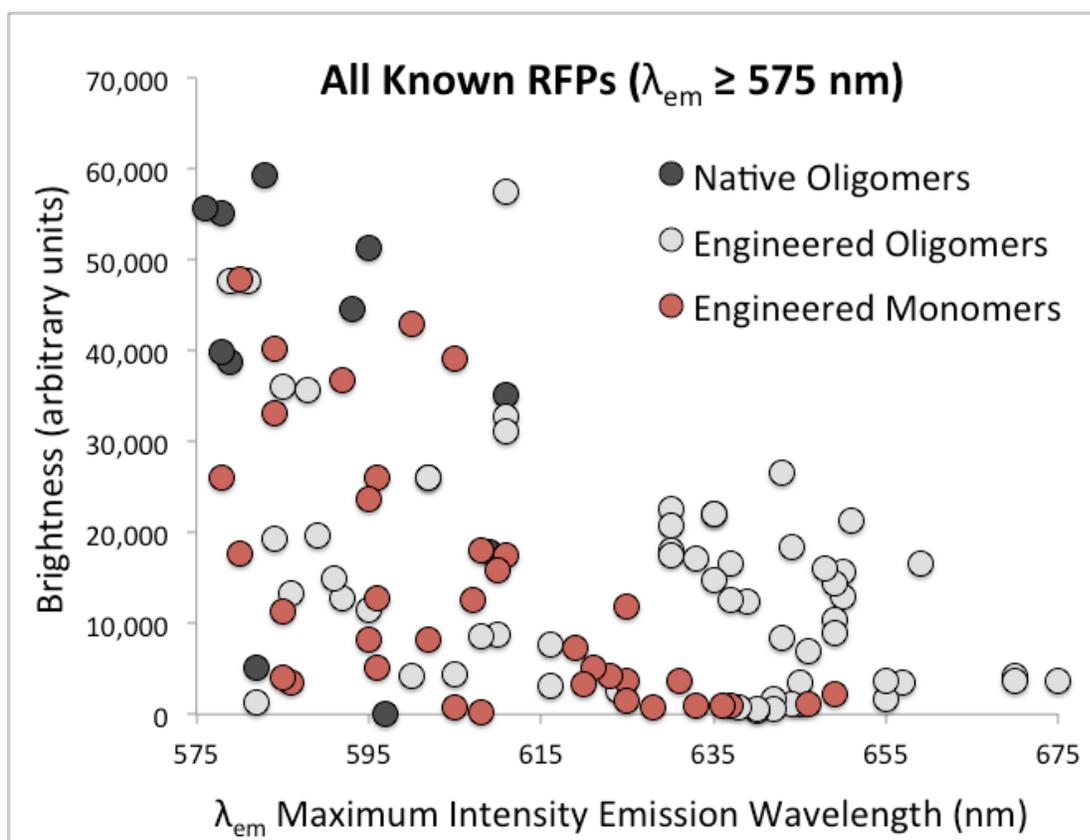
Monomeric RFP	Brightness ( $\Phi \times \epsilon$ ) / 1000	$\lambda_{em}$ (nm)	Mutations to Core	Total Mutations	Immediate Parent (dimer/tetramer)	Brightness ( $\Phi \times \epsilon$ ) / 1000	$\lambda_{em}$ (nm)	Mutations to Core	Total Mutations	Ancestral Protein (dimer/tetramer)	Brightness ( $\Phi \times \epsilon$ ) / 1000	$\lambda_{em}$ (nm)
mRFP1	12.5	607	13	33	DsRed (T)	59.3	583	0/a	0/a	DsRed (T)	59.3	583
DsRed.M1	3.5	586	10	45	DsRed (T)	59.3	583	0/a	0/a	DsRed (T)	59.3	583
FusionRed	18.0	608	9	45	mKate2 (D)	18.0	630	7	27	eqFP578 (T)	55.1	578
mRuby	39.2	605	6	40	eqFP611 (T)	35.1	611	0/a	0/a	eqFP611 (T)	35.1	611
mKeima	3.5	620	7	17	dKeima (D)	7.6	616	5	13	COCP (T)	--	--
mGinger0.1	1.2	637	8	47	HcRed7 (D)	4.8	643	4	11	hChCP	--	--

**Table 2 – Spectroscopic characterization of HcRed variants**

RFP	$\Phi$	$\epsilon$ ( $M^{-1} \text{ cm}^{-1}$ ) / 1000	Brightness ( $\Phi \times \epsilon$ ) / 1000	$\lambda_{\text{ex}}$ (nm)	$\lambda_{\text{em}}$ (nm)	Apparent $T_m$ ( $^{\circ}\text{C}$ )
HcRed	0.05	70	6.0	585	633	69.0
HcRed7	0.08	75	8.4	592	645	75.0
HcRed7 $\Delta$ 5	0.06	69	4.3	592	643	70.5
HcRed7 $\Delta$ 6	ND	ND	ND	582	635	65.0
HcRed77	0.05	†	†	ND	ND	67.5
HcRedm1	0.01	†	†	ND	ND	64.0
HcRedm13	0.03	†	†	ND	ND	65.0
HcRedm14	0.02	†	†	ND	ND	58.0
mGinger0.1	0.02	58	1.2	587	637	79.0
mGinger0.2	0.04	36	1.5	578	631	80.0

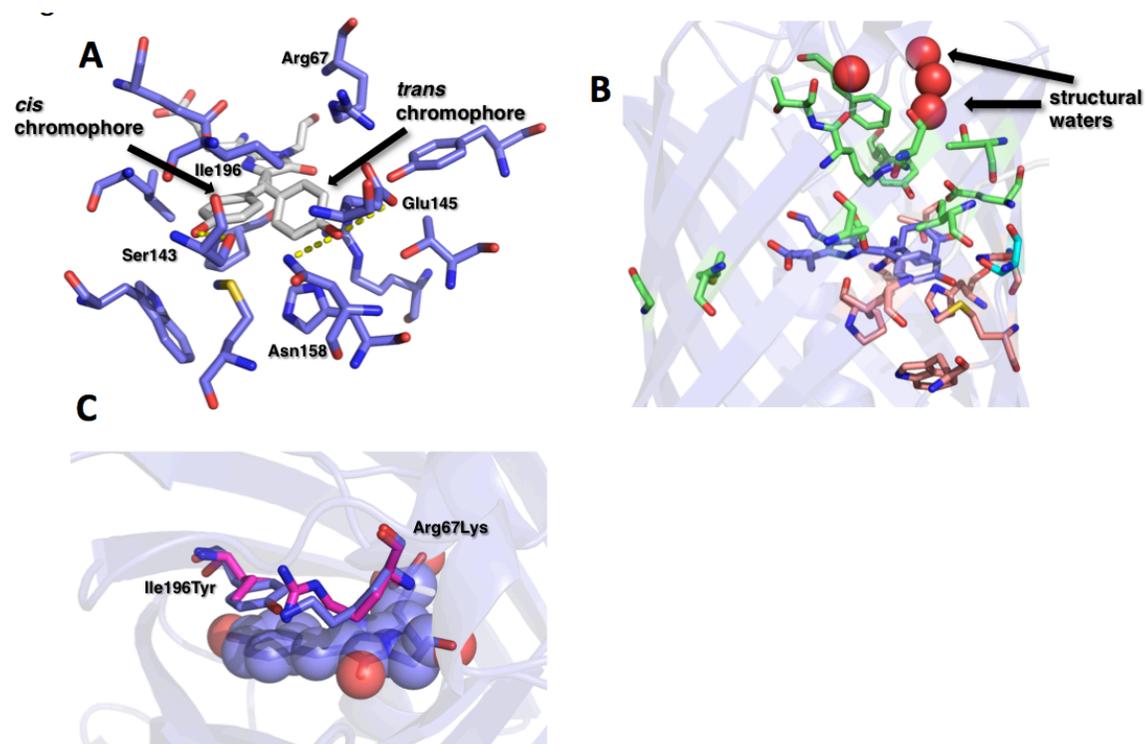
ND – Not determined

† - Extinction coefficient (and therefore brightness) could not be measured because of multiple chromophore species present.

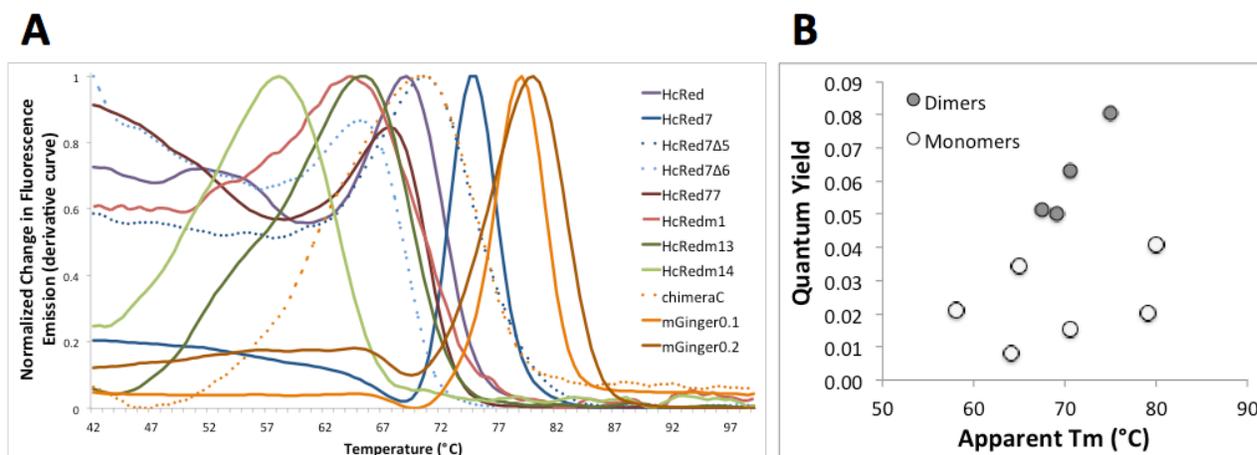


**Figure 1 – There is a negative correlation between brightness and  $\lambda_{em}$  among RFPs.** All known RFPs whose brightness and  $\lambda_{em}$  have been measured are plotted. Among *Aequorea victoria* class RFPs there is a noticeable trend that the further bathochromic the shift to fluorescent emission, the dimmer the fluorescence. Native proteins are all tetrameric and exhibit the brightest fluorescence. Monomeric proteins are the dimmest and their brightness drops off quickly at longer wavelengths.

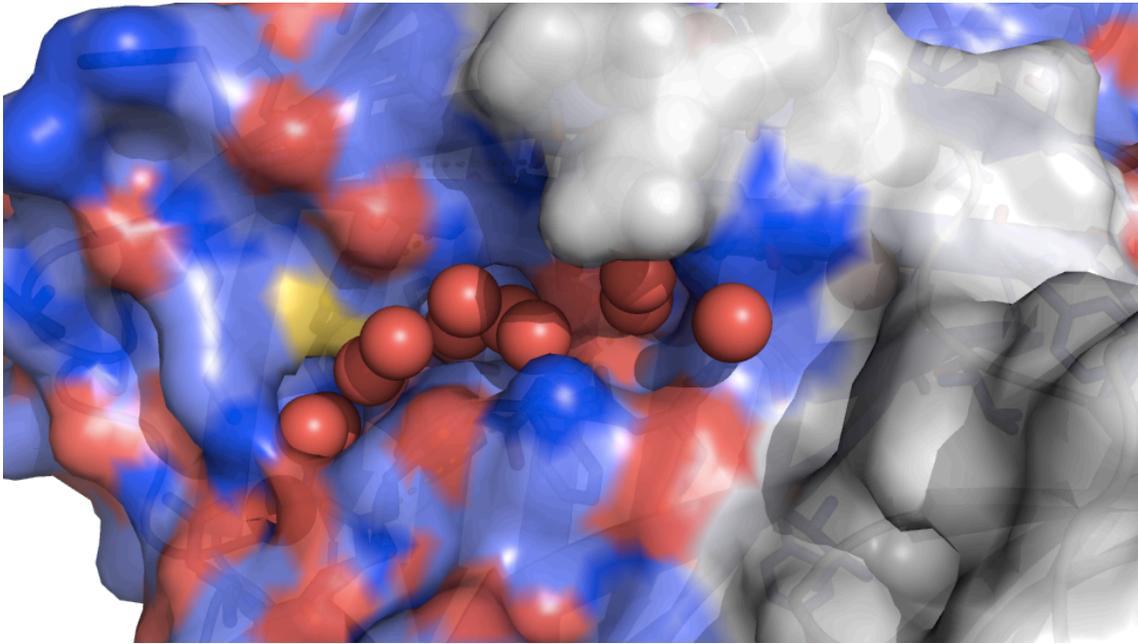




**Figure 3 – The design of HcRed7.** (A) The crystal structure of HcRed (PDB: 1YZW) shows dual occupancy of the chromophore's phenolate group. The *cis* chromophore (*cis* to the imidazolinone nitrogen) is stabilized by a Cys143Ser mutation from parent protein HcCP. The *trans* chromophore is stabilized by two hydrogen bonds from Glu145 and Asn158. Chromophore is shown in gray, side chains are shown in blue, and hydrogen bonds in yellow. (B) Two core libraries targeted unique structural regions of the protein core. The first region (pink) surrounds the phenolate side chain of the chromophore. The second (green), is a very highly mutated region in RFP monomer evolution. This region holds an internal water channel, key catalytic residues, and abuts the AC interface. (C) A crystal structure of HcRed7 (blue), isolated from core library B is overlaid onto HcRed (magenta). HcRed7 has mutations Arg67Lys and Ile196Tyr. Tyr196 provides a  $\pi$ -stacking interaction with the chromophore, red-shifting emission and stabilizing the chromophore in the *cis* orientation. Lys67 plays a key catalytic role, but the mutation away from arginine frees up room for Tyr196.



**Figure 5 – Thermal stability of HcRed variants.** (A) We measured the stability of the various HcRed variants with a ramped temperature melt in a qPCR machine. Thermal stability decreased with significant engineering events such as the deletion of HcRed7's tail and the monomerization of HcRed. Directed evolution at 30°C improved expression of the protein, but decreased its thermostability (HcRed77, HcRedm13, HcRedm14). When the expression temperature for screening was increased to 37°C, the thermostability jumped rapidly (mChimeraC, mGinger0.1, mGinger0.2). (B) Thermostability during the evolution of HcRed (the same ten variants from (A) are plotted) appears to be correlated to quantum yield. Monomers exhibit weaker quantum yield than tetramers. Extinction coefficient was not measurable for some variants because there were multiple chromophore species present.



**Figure 6 – A structural comparison of the mGingers to HcRed7.** (A) There is a deformity in the structure of the  $\beta$ -barrel that takes the form of a large gap between  $\beta$ -strands 7 and 10. Eleven water molecules crystalize in this gap, forming a long channel that appears to be structurally stabilized by the nearby AC dimeric interactions. Molecule A is shown as a blue surface with nitrogens and oxygens colored blue and red respectively. Molecule B is shown as a gray surface. Water molecules are shown as red spheres. (B,C,D) More sub-figures expected with the mGinger structure..

## 4.9 Supplementary Tables and Figures

	Library A - Chromophore phenolate pocket									A / B	Library B - Above chromophore, near structural waters + chromophore backing positions 28 and 41											Chromophore		
	63	95	143	161	163	177	179	199	146		31	44	70	71	83	148	181	195	197	215	217	66	67	68
Position (DsRed)	60	92	140	158	160	173	175	198	143	28	41	67	68	80	145	177	194	196	213	215	63	64	65	
hcCP	P	R	W	N	M	L	T	L	C	G	M	R	T	F	E	Y	T	I	E	S	E	Y	G	
HcRed	--	--	--	--	--	H	--	--	S	--	--	--	--	--	--	--	--	--	--	--	--	--	--	
aeCP597	P	R	W	S	M	L	T	I	C	G	E	K	T	F	E	Y	G	H	E	A	M	Y	G	
AQ143	--	--	--	A	--	--	--	--	S	--	--	--	--	--	--	--	--	--	--	--	--	--	--	
#20CP	P	R	W	N	M	F	S	L	N	G	V	I	P	V	E	Y	V	R	E	S	Q	Y	G	
mKeima	--	--	--	D	--	--	--	--	S	--	--	--	--	F	--	--	I	--	A	--	--	--		
anmCP	H	R	I	C	I	F	L	I	T	A	V	P	F	A	T	F	V	V	E	T	Q	Y	G	
J-Red	--	--	--	A	--	--	M	T	N	--	--	--	--	--	--	--	--	I	--	--	--	--	--	
DsRed	P	R	W	I	K	F	S	L	S	G	V	K	V	K	E	Y	V	S	E	T	Q	Y	G	
DsRed.M1	--	--	--	--	H	--	T	--	--	--	A	--	A	M	--	--	--	--	A	--	--	--		
mRFP1	--	--	--	--	M	V	T	--	--	--	A	--	A	L	--	--	T	I	--	A	--	--	--	
mCherry	--	--	--	--	Q	V	T	--	--	--	A	--	A	L	--	--	--	I	--	A	M	--	--	
mGrape3	--	--	--	--	V	M	V	T	--	--	A	--	A	M	--	--	L	Y	--	A	--	--	--	
mPlum	--	--	--	--	M	M	V	T	--	--	A	--	A	L	--	--	T	I	--	A	M	--	--	
mRaspberry	--	--	--	--	M	M	V	T	--	--	A	--	G	L	--	--	T	I	--	A	M	--	--	
E2-Crimson	--	--	--	--	N	M	--	S	--	--	A	--	A	L	--	--	--	Y	--	A	F	--	--	
eqFP611	T	R	W	S	M	F	T	L	N	G	M	K	T	F	E	Y	V	H	E	A	M	Y	G	
mRuby	--	--	--	T	--	--	--	--	--	--	--	R	--	--	--	--	--	--	--	--	--	--	--	
mRuby2	--	--	--	T	--	--	--	--	--	--	--	R	--	--	--	--	--	--	--	--	--	--	--	
RFP637	--	--	--	A	--	--	--	--	S	--	--	--	--	--	--	--	--	--	--	--	--	--	--	
RFP639	--	--	--	C	--	--	--	--	S	--	--	--	--	--	--	--	--	--	--	--	--	--	--	
eqFP578	T	R	W	S	M	F	T	L	N	S	M	K	T	F	E	Y	V	H	E	A	M	Y	G	
FusionRed	--	--	--	C	--	L	--	--	S	--	--	R	--	--	--	--	--	--	--	--	--	--	--	
TagRFP	--	--	--	--	--	--	--	--	--	--	--	R	--	--	--	--	--	--	--	--	--	--	--	
Katushka	--	--	--	--	--	L	--	--	S	--	--	--	--	--	--	--	--	R	--	--	--	--	--	
mKate	--	--	--	--	--	L	--	--	S	--	--	--	--	--	--	--	--	R	--	--	--	--	--	
mKate2	--	--	--	A	--	L	--	--	S	--	--	--	--	--	--	--	--	R	--	--	--	--	--	
Neptune	--	--	--	C	--	L	--	--	S	--	G	--	--	--	--	--	--	R	--	--	--	--	--	
mNeptune	--	--	--	C	--	L	--	--	S	--	G	--	--	--	--	--	--	R	--	--	--	--	--	
TagRFP657	--	--	--	T	L	--	--	--	H	--	Q	H	--	--	--	--	--	--	Y	--	--	--	--	
eqFP650	--	--	--	S	--	L	--	--	S	--	A	--	--	--	--	--	--	--	R	--	--	--	--	
eqFP670	--	--	--	N	--	L	--	--	--	--	C	--	--	--	--	--	--	--	R	--	--	--	--	
Positions allowed in Library				I	H	L	S		N	S	A	K	A	L				H	A					
				A	L	F				A	C	H	P	M				R						
				C	Q	V					E							Y						
				S							Q													
				T							V													
				M																				

**Table S1 – Mutations to select core residues during RFP engineering.** Alignment of far-red engineered proteins of interest and the tetrameric RFPs they were derived from. Dots on the far left indicate monomers. Residue 143 was part of both HcRed core libraries. Allowed residues in both core libraries are listed at the bottom. Total theoretical library size for library A was 448, and for library B was 432.

	$\lambda_{\text{ex}}$	$\lambda_{\text{em}}$	$\Phi$	$\epsilon$	Brightness
core1	574	616	0.36	96,403	34,414
core2	596	632	0.03	79,614	2,301
core3	586	628	0.09	64,094	5,631
core4	580	618	0.20	50,323	9,813
core5	596	636	0.04	94,905	3,657
core7	594	642	0.07	76,670	5,367
core9	588	626	0.02	63,298	1,038
core10	590	638	0.04	48,913	2,005
core11	592	640	0.04	70,564	2,990
core12	596	646	0.04	76,537	3,328
core13	596	638	0.03	87,410	2,302
core14	584	644	0.04	88,840	3,402
core15	588	612	0.25	73,240	18,056
core17	592	628	0.03	77,821	2,389
core18	588	632	0.02	89,906	2,011
core19	592	630	0.02	89,220	2,199
core20	586	632	0.09	75,300	7,108
core21	596	634	0.03	81,281	2,738
core24	584	624	0.19	108,173	20,014
core25	574	618	0.26	84,953	22,046
core26	564	614	0.43	68,643	29,312
core27	586	622	0.14	49,196	6,994
core28	568	610	0.37	101,824	38,009
core29	586	616	0.02	95,910	1,745
core30	594	616	0.00	92,130	222
core31	594	640	0.07	98,159	7,173
core34	592	620	0.01	95,290	577
core35	592	620	0.01	102,455	1,504
AC_core1	579	617	0.30	73,312	22,250
AC_core2	594	632	0.14	61,430	8,333
AC_core3	580	619	0.33	64,566	21,531
AC_core4	587	628	0.23	64,526	14,800
AC_core5	584	618	0.24	73,402	17,724
AC_core6	590	628	0.22	65,283	14,057
AC_core7	572	614	0.39	59,841	23,159
AC_core8	583	620	0.21	64,220	13,216
AC_core9	582	621	0.16	64,915	10,409
AC_core10	594	632	0.13	55,972	7,124
AC_core11	586	624	0.14	50,955	7,208
AC_core12	586	624	0.16	36,297	5,716
AC_core13	583	620	0.17	62,469	10,826
AC_core14	577	616	0.33	65,210	21,386
AC_core15	580	618	0.01	108,944	1,051
AC_core16	588	624	0.07	56,968	3,822
AC_core17	592	630	0.12	70,650	8,594
AC_core18	585	624	0.15	56,009	8,450
AC_core19	584	626	0.19	39,874	7,730
AC_core20	594	633	0.13	78,293	9,820
AC_core21	577	620	0.31	71,419	22,456
AC_core22	580	618	0.24	71,205	16,843
AC_core23	589	624	0.18	61,844	10,880
AC_core24	582	614	0.19	59,592	11,063
AC_core25	578	613	0.14	66,129	9,523
AC_core26	583	612	0.01	63,028	431

**Table S2 – Spectroscopic characterization of hits from two HcRed core libraries.**

Variant core7 was called HcRed7 in this work.

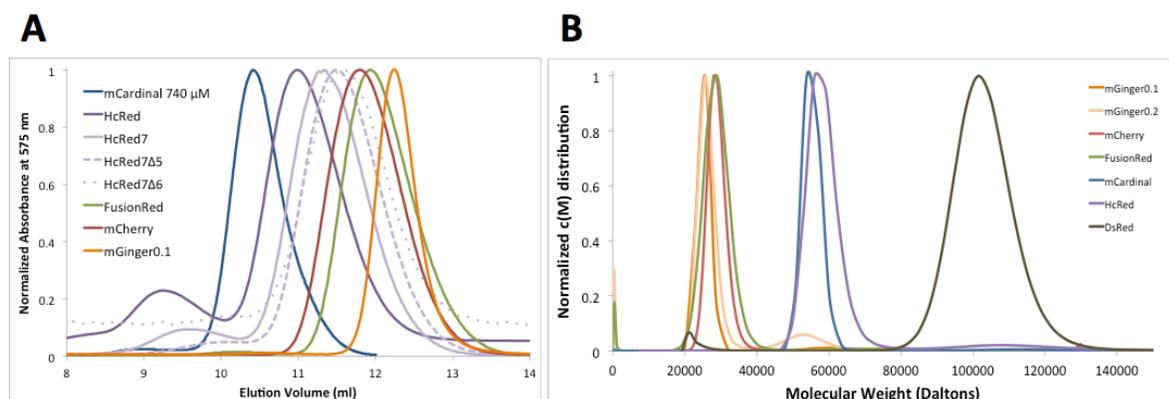
Position	Library A				A / B	Library B						
	158	160	173	175	143	28	41	67	68	80	196	215
WT Residue	N	M	H	T	S	G	M	R	T	F	I	S
core1	-	-	-	-	-	-	-	-	-	-	H	A
core5	-	-	-	-	-	-	-	-	A	-	-	-
core6	-	-	-	-	-	-	-	H	-	-	Y	-
core7	-	-	-	-	-	-	-	K	-	-	Y	-
core8	-	-	-	-	-	-	-	K	A	-	-	-
core9	-	-	-	-	-	-	-	K	P	-	-	-
core10	-	-	-	-	-	-	A	K	-	-	Y	A
core11	-	-	-	-	-	-	A	K	A	-	Y	A
core12	-	-	-	-	-	-	C	K	-	-	Y	-
core13	-	-	-	-	-	-	C	K	A	-	-	-
core14	-	-	-	-	-	-	Q	K	A	-	-	-
core15	-	-	-	-	-	-	V	K	-	-	-	-
core16	-	-	-	-	-	A	-	K	-	-	-	-
core17	-	-	-	-	-	A	-	K	-	-	-	A
core18	-	-	-	-	-	A	-	K	A	-	-	-
core19	-	-	-	-	-	A	-	K	A	-	-	A
core2	-	-	L	-	-	-	-	-	-	-	-	-
core3	-	-	V	-	-	-	-	-	-	-	-	-
core20	-	-	V	-	-	-	-	-	-	-	Y	A
core4	-	Q	V	S	-	-	-	-	-	-	-	-
core21	A	H	-	-	-	-	-	-	-	-	Y	A
core22	A	H	-	-	N	-	-	-	-	-	Y	-
core23	A	H	-	-	N	-	-	-	-	-	Y	A
core24	C	-	-	-	-	-	-	-	-	-	-	-
core25	C	-	-	-	-	-	-	-	-	-	-	A
core26	C	-	-	-	-	-	-	-	-	-	H	-
core27	C	-	-	S	-	-	-	-	-	-	-	-
core28	C	-	V	-	-	-	-	-	-	-	-	-
core29	C	-	V	-	N	-	-	-	-	-	-	A
core30	C	-	V	S	N	-	-	-	-	-	-	-
core31	C	H	-	-	-	-	-	-	-	-	Y	A
core32	M	Q	-	-	-	-	-	-	-	-	H	A
core33	S	-	V	-	N	-	-	-	-	-	-	A
core34	S	H	-	-	-	-	-	-	-	-	Y	A
core35	T	-	V	S	-	-	-	-	-	-	-	-
core36	T	L	F	-	N	-	-	-	-	-	R	A
core37	A	H	-	-	-	-	-	-	-	-	-	-

**Table S3 – Mutations to the HcRed core in the bright, characterized hits.** Variant core7 was called HcRed7 in this work.

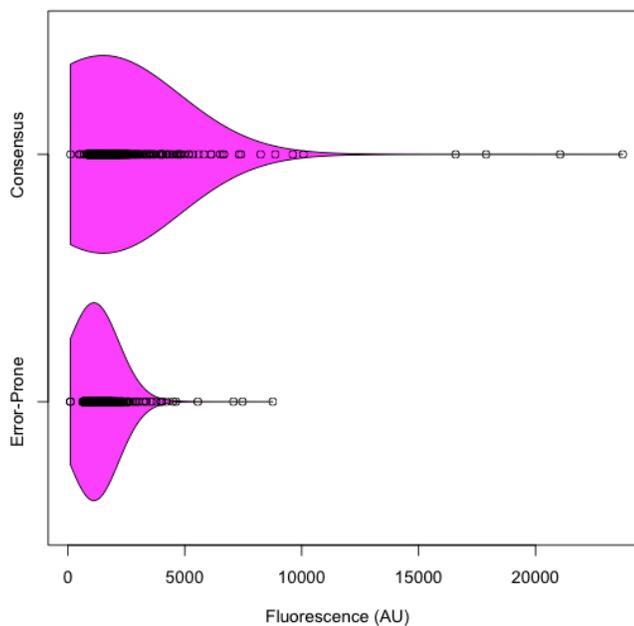


**Table S4.** X-ray data reduction and crystallographic refinement statistics for HcRed7

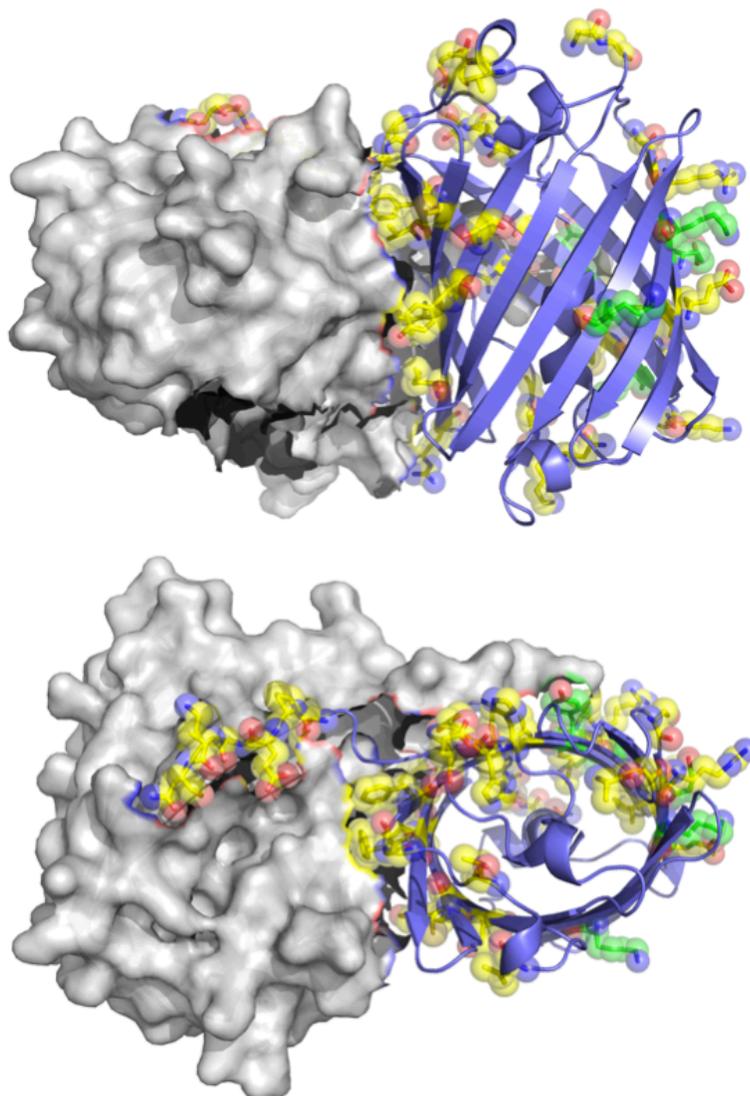
(A) X-ray data reduction statistics	
Space group	P1 21 1
Unit cell dimensions ( <i>a</i> , <i>b</i> , <i>c</i> )	54.3 Å, 122.1 Å, 108.8 Å
Resolution	39.3 Å – 1.63 Å
(last shell)	1.72 Å – 1.63 Å
Total measurements (last shell)	761,418 (83,674)
Number of unique reflections (last shell)	111,329 (13,784)
Wavelength	
<i>R</i> -merge (last shell)	0.093 (1.554)
<i>I</i> / $\sigma$ ( <i>I</i> ) (last shell)	11.6 (1.1)
Completeness (last shell)	0.957 (0.797)
Multiplicity (last shell)	6.8 (6.1)
(B) Crystallographic refinement statistics	
Resolution	33.7 Å – 1.63 Å
(last shell)	1.83 Å – 1.63 Å
No. of reflections (working set)	111,232
No. of reflections (test set)	5,487
<i>R</i> -factor (last shell)	0.190 (0.312)
<i>R</i> -free (last shell)	0.221 (0.338)
No. of amino acid residues	893
No. of atoms	7,154
No. of solvent molecules	655
Average <i>B</i> -factor	
Protein	30.1 Å <sup>2</sup>
Solvent	36.1 Å <sup>2</sup>
<i>R.m.s.d.</i> from ideal geometry	
Bond lengths	0.010 Å
Bond angles	1.331°



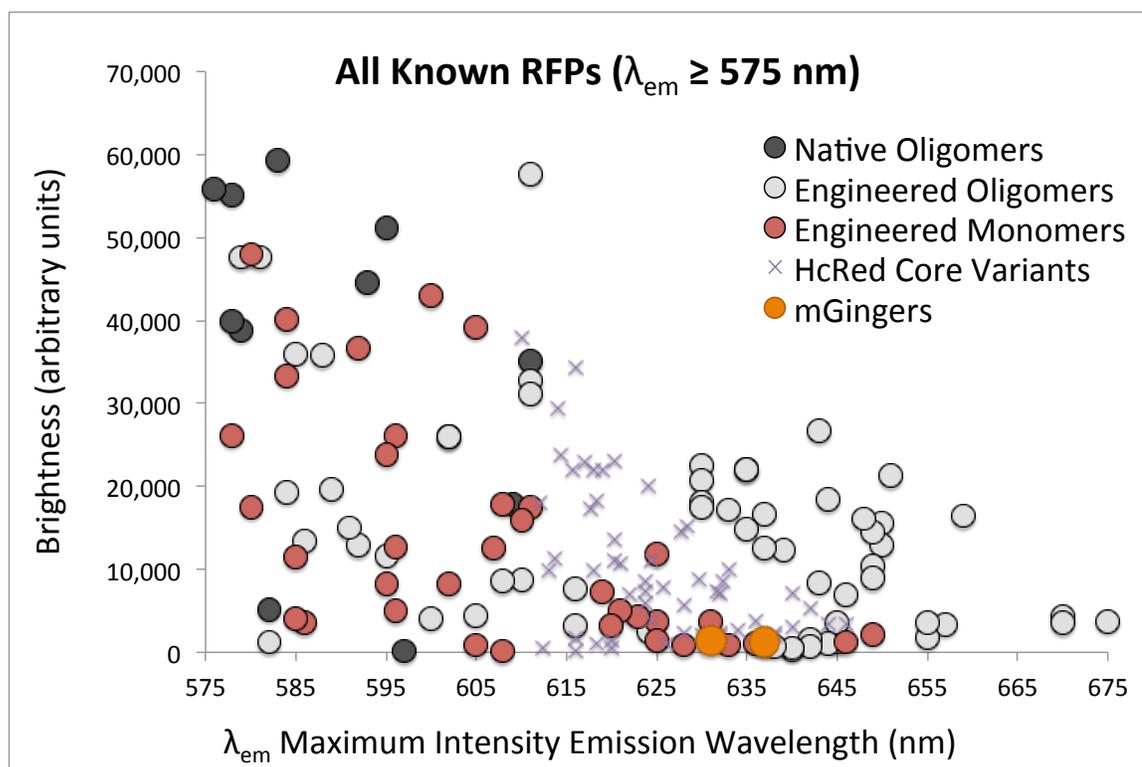
**Figure S1 – Oligomeric characterization of HcRed variants.** (A) As a first test of oligomerization we ran size exclusion chromatography (SEC). This method gives a rough idea of molecular size, and the trend is clear that throughout the engineering process from HcRed to mGinger0.1, there has been a steady shift towards later elution from the column, indicating that smaller molecules are present. (B) We confirmed SEC data with analytical ultracentrifugation (AUC). We ran sedimentation velocity at 50,000 RPM and performed a  $c(M)$  analysis of the sedimentation data. DsRed was run as a tetrameric standard, HcRed as a dimeric standard, and mCherry and FusionRed as monomeric standards. The mGingers sediment even slightly below mCherry and FusionRed, an effect that is explained by their lack of C-terminal tail. mCardinal, in contrast with the mGingers, is not actually monomeric.



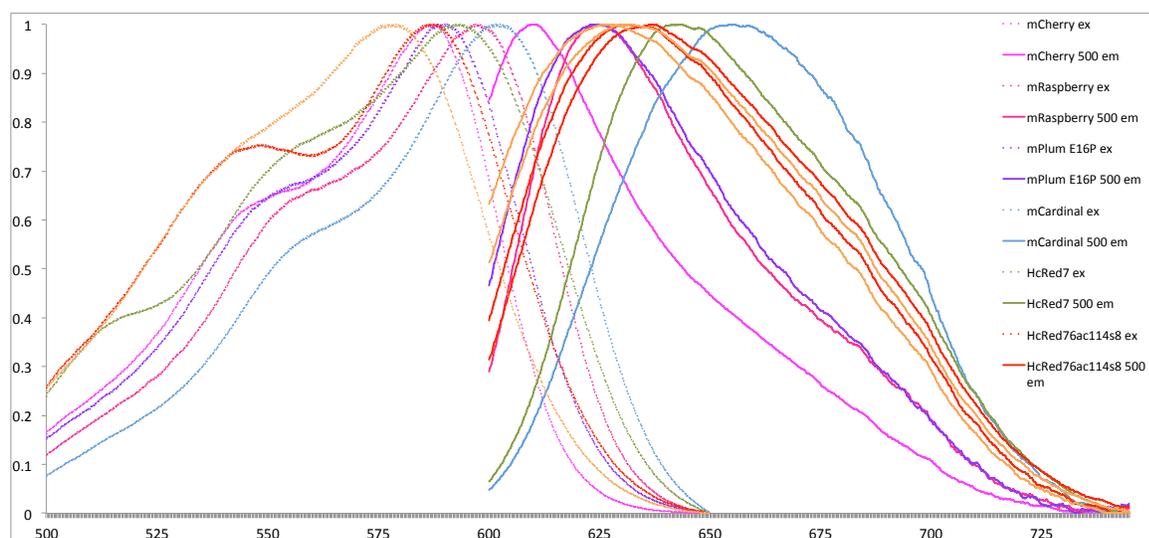
**Figure S2 – A comparison of directed evolution techniques.** Consensus design and error-prone mutagenesis were used concurrently to try to recover the fluorescence of a destabilized HcRed7 $\Delta$ 6. Consensus design produced many more bright variants, and a greater percentage of the overall library was fluorescent. Error-prone mutagenesis pulled out mutations that complemented the consensus positions.



**Figure S3 – All mutations made to HcRed7 during monomerization.** These mutations are overlaid onto the structure of HcRed7. In yellow are the mutations shared between mGinger0.1 and mGinger0.2. In green are the five positions mutated in mGinger0.2 from mGinger0.1.



**Figure S4 – A brightness vs. emission plot that includes HcRed variants.** HcRed core variants and the mGingers are plotted along with the rest of the known characterized fluorescent proteins. The HcRed core library spanned a large number of wavelengths and produced some bright variants. The mGingers fit right onto the apparent brightness frontier for monomeric RFPs.



**Figure S5 – Fluorometer readings of far-red FP and HcRed variant spectra.** HcRed7 is 9 nm red-shifted from HcRed. mGinger is red-shifted from mRaspberry and mPlum E16P, which is a variant of mPlum with full maturation to the red chromophore, in contrast to mPlum, which has a mixed green/red chromophore.

## 4.10 Supplemental Methods

### *Generating a multiple sequence alignment (MSA) and computing a consensus sequence.*

We searched various resources including GenBank, SwisProt, UniProt, NCBI-BLAST, and patent databases for reported FP sequences. We found 741 unique fluorescent protein sequences and aligned them with MAFFT, which we then hand-curated with the use of a 163-member structural alignment. Phylogenetic distances were estimated and then used to weight a consensus sequence calculation.

Specific technique:

Computes a consensus sequence for the MSA, by doing the following:

1. Henikoff weight the alignment
2. Compute symbol counts (singlets are used here)
3. Compute the column entropy, IGNORING GAPS
4. Compute uncertainty reduction for each column, using

$$R_c = \log_2(20) - H_c$$

5. Reduce  $R_c$  by the gap fraction ( $R_{c'} = (1 - \text{gaps}[c]) * R_c$ )
6. Calculate a score for each position and each non-gap character via

$$S_{c(A)} = p_{c(A)} * R_{c'}$$

7. Find  $\max(S_{c(A)})$  over A, for each c. That's the consensus character.

For a rough score interpretation, a nongapped column that is partitioned equally among the 20 amino acids will have a score of 0.0.

This function returns an alignment-position indexed dictionary of consensus AAs and score, along with a 20 x Npos matrix of character scores, and a corresponding key to identity of the rows.

No pseudocounting is used in determining character frequencies.

## 4.11 References

1. Wang S, Moffitt JR, Dempsey GT, Xie XS, & Zhuang X (2014) Characterization and development of photoactivatable fluorescent proteins for single-molecule–based superresolution imaging. *Proceedings of the National Academy of Sciences of the United States of America*.
2. Wenqin W, Gene-Wei L, Chongyi C, Xie XS, & Xiaowei Z (2011) Chromosome Organization by a Nucleoid-Associated Protein in Live Bacteria. *Science* 333(6048):1445-1449.
3. Alieva NO, *et al.* (2007) Diversity and evolution of coral fluorescent proteins. *PLoS ONE* 3(7).
4. Shagin DA, *et al.* (2004) GFP-like proteins as ubiquitous metazoan superfamily: evolution of functional features and structural complexity. *Molecular Biology and Evolution* 21(5):841-850.
5. Salih A, Larkum A, Cox G, Kühl M, & Hoegh-Guldberg O (2000) Fluorescent pigments in corals are photoprotective. *Nature* 408(6814):850-853.
6. Smith EG, D'Angelo C, Salih A, & Wiedenmann J (2013) Screening by coral green fluorescent protein (GFP)-like chromoproteins supports a role in photoprotection of zooxanthellae. *Coral reefs*.
7. Palmer CV, Modi CK, & Mydlarz LD (2009) Coral Fluorescent Proteins as Antioxidants. *PLoS ONE* 4(10).
8. Bou-Abdallah F, Chasteen ND, & Lesser MP (2006) Quenching of superoxide radicals by green fluorescent protein. *Biochimica et Biophysica Acta* ( ...
9. Campbell RE, *et al.* (2002) A monomeric red fluorescent protein. *Proceedings of the National Academy of Sciences*.
10. Kredel S, *et al.* (2008) mRuby, a bright monomeric red fluorescent protein for labeling of subcellular structures. *PLoS ONE* 4(2).
11. Shemiakina II, *et al.* (2012) A monomeric red fluorescent protein with low cytotoxicity. *Nature Communications*.
12. Takako K, *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. Tromberg BJ, *et al.* (1999) Non-invasive in vivo characterization of breast tumors using photon migration spectroscopy. *Neoplasia (New York, N.Y.)* 2(1-2):26-40.
14. Moore MM, Oteng-Pabi SK, Pandelieva AT, Mayo SL, & Chica RA (2012) Recovery of Red Fluorescent Protein Chromophore Maturation Deficiency through Rational Design. *PLoS ONE* 7(12).
15. Wang L, Jackson WC, Steinbach PA, & Tsien RY (2004) Evolution of new nonantibody proteins via iterative somatic hypermutation. *Proceedings of the*

- National Academy of Sciences of the United States of America* 101(48):16745-16749.
16. Goedhart J, *et al.* (2011) Structure-guided evolution of cyan fluorescent proteins towards a quantum yield of 93%. *Nature communications* 3:751.
  17. Ilagan RP, *et al.* (2010) A new bright green-emitting fluorescent protein--engineered monomeric and dimeric forms. *The FEBS journal* 277(8):1967-1978.
  18. Lin MZ, *et al.* (2009) Autofluorescent proteins with excitation in the optical window for intravital imaging in mammals. *Chemistry & biology* 16(11):1169-1179.
  19. Gurskaya NG, *et al.* (2001) GFP-like chromoproteins as a source of far-red fluorescent proteins1. *FEBS Letters* 507(1):16-20.
  20. 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.
  21. Piatkevich KD, *et al.* (2012) Extended Stokes shift in fluorescent proteins: chromophore-protein interactions in a near-infrared TagRFP675 variant. *Scientific reports* 3:1847.
  22. Wannier TM & Mayo SL (2014) The structure of a far-red fluorescent protein, AQ143, shows evidence in support of reported red-shifting chromophore interactions. *Protein Science* 23(8):1148-1153.
  23. Cramer A, Whitehorn EA, Tate E, & Stemmer WP (1996) Improved green fluorescent protein by molecular evolution using DNA shuffling. *Nature biotechnology* 14(3):315-319.
  24. Zhao H & Arnold FH (1997) Optimization of DNA Shuffling for High Fidelity Recombination. *Nucleic Acids Research*.
  25. Chica RA, Moore MM, Allen BD, & Mayo SL (2010) Generation of longer emission wavelength red fluorescent proteins using computationally designed libraries. *Proceedings of the National Academy of Sciences of the United States of America* 107(47):20257-20262.
  26. Morozova KS, *et al.* (2010) Far-red fluorescent protein excitable with red lasers for flow cytometry and superresolution STED nanoscopy. *Biophysical journal* 99(2):5.
  27. Grigorenko BL, Nemukhin AV, Polyakov IV, & Krylov AI (2013) Triple-decker motif for red-shifted fluorescent protein mutants. *The Journal of Physical Chemistry Letters* 4(10):1743-1747.
  28. Strack RL, Strongin DE, Mets L, Glick BS, & Keenan RJ (2010) Chromophore formation in DsRed occurs by a branched pathway. *Journal of the American Chemical Society* 132(24):8496-8505.

29. 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<sup>†,‡</sup>. *Biochemistry* 44(29):9833-9840.
30. Strongin DE, *et al.* (2007) Structural rearrangements near the chromophore influence the maturation speed and brightness of DsRed variants. *Protein Engineering, Design & Selection* 20(11):525-534.
31. Henderson JN, *et al.* (2009) Excited State Proton Transfer in the Red Fluorescent Protein mKeima. *Journal of the American Chemical Society* 131(37):13212-13213.
32. Kredel S, *et al.* (2009) mRuby, a Bright Monomeric Red Fluorescent Protein for Labeling of Subcellular Structures. *PLoS ONE* 4(2).
33. Regmi C, K. , Bhandari YR, Gerstman BS, & Chapagain P, P. (2013) Exploring the Diffusion of Molecular Oxygen in the Red Fluorescent Protein mCherry Using Explicit Oxygen Molecular Dynamics Simulations. *The journal of physical chemistry. B*.
34. Lelimosin M, *et al.* (2009) Intrinsic dynamics in ECFP and Cerulean control fluorescence quantum yield. *Biochemistry* 48(42):10038-10046.
35. Rurack K & Spies M (2011) Fluorescence quantum yields of a series of red and near-infrared dyes emitting at 600-1000 nm. *Analytical chemistry* 83(4):1232-1242.
36. Bloom JD, Labthavikul ST, Otey CR, & Arnold FH (2006) Protein stability promotes evolvability. *Proceedings of the National Academy of Sciences of the United States of America* 103(15):5869-5874.
37. Adams P, *et al.* (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta crystallographica. Section D, Biological crystallography* 66(Pt 2):213-221.
38. Emsley P, Lohkamp B, Scott W, & Cowtan K (2010) Features and development of Coot. *Acta crystallographica. Section D, Biological crystallography* 66(Pt 4):486-501.

## CHAPTER 5

### Accurate Characterization of Some Common *Anthozoa* Class Red Fluorescent Proteins

#### 5.1 Abstract

Red fluorescent proteins (RFPs) are broadly used across biology, and are of special interest to protein engineers because of the allure of *in vivo* imaging of vertebrate hosts at near-infrared wavelengths. ~200 RFP variants have been reported in the literature, but we have found that drawing on the literature for RFP engineering or choosing an RFP marker for biological studies is hampered by a lack of reliable data. Most proteins are not thoroughly characterized, and for those that are, the data are often marred by significant errors or inconsistencies in measurement. Here we detail a thorough characterization of some of the most heavily used RFPs. We find that key attributes such as oligomerization, peak emission, and brightness are often misreported or have significant error. Because all RFPs in this study are expressed, purified, and assayed in parallel, as opposed to comparing values calculated across different instrumentation, we are confident in the comparative story that these results tell. To guide future measurement techniques, this study attempts to lay out a standard methodology for characterizing RFPs so that future variants can be effectively compared to the existing cadre of engineered RFPs.

## 5.2 Introduction

*Aequorea victoria* class red fluorescent proteins (RFPs) are heavily used biological markers. They contribute to a broad color palette that is used for multi-wavelength spectroscopy and are of crucial interest because of the possibility of engineering bright, near infrared (NIR) variants that would allow imaging through biological tissue (1). Since 2000 when the first structures of DsRed were published, ~50 native and ~130 engineered RFPs have been reported, and characterized to some degree. The most common characterization of these proteins is to report the wavelengths at which the RFP maximally excites and emits light. Quantum yield and extinction coefficient, which are the two determinants of brightness, are then reported for most of the engineered variants. Apart from brightness and wavelength, there is little uniformity in the characterization of RFPs. Additionally, many of the assays that measure biophysical attributes are not standardized or conducted with any uniformity. Even some of the more standardized assays that measure basic biophysical attributes can produce very different results depending on subtle differences in the instruments or protocol used (2-4). The problems with RFP characterization can in turn complicate any efforts to draw conclusions about RFP engineering. Pooling data from multiple sources that are subject to large experimental variation is an exercise that can be easily swamped with error. Even more importantly, imprecisely determined FP attributes can lead to the misinterpretation of experimental data when these proteins are used to visualize biological processes.

The most commonly reported fluorescent characteristics of RFPs are their maximum intensity absorbance ( $\lambda_{\text{abs}}$ ), excitation ( $\lambda_{\text{ex}}$ ), and emission ( $\lambda_{\text{em}}$ ) wavelengths, and their brightness, which is the product of a protein's quantum yield ( $\Phi$ ) and its extinction coefficient ( $\epsilon$ ). Other commonly reported parameters are the pKa of the chromophore, its photostability, its fluorescence lifetime, and the oligomeric state of the protein. Together these are thought to be the most important fluorescent characteristics for common biological imaging applications. There are other properties of these proteins that are less often reported such as the green to red ratio of the chromophore, the half-time of chromophore maturation, and the thermal stability of the protein. These attributes can

provide insight into the biophysical basis for fluorescence, and we believe they are important for RFP design. Thorough and accurate characterization of a key set of important and commonly used RFP variants would be an invaluable data set that we believe would help to instruct future RFP design.

Here we focus on a subset of fluorescent characteristics that we believe are tied to the design of bright, far-red monomeric RFP variants. We measure quantum yield, extinction coefficient,  $\lambda_{em}$ , thermal stability, and oligomericity in some of the most commonly used and a group of recently reported far-red FPs. This data set is ripe for querying with questions about the correlations amongst RFP fluorescent properties, and for drawing conclusions about some important structural underpinnings of fluorescence.

### **5.3 Results and Discussion**

#### ***RFP Oligomericity is Poorly Characterized in the Literature***

RFP monomerization has been one of the principal goals for RFP engineers, as all known native RFPs are tetrameric (5, 6). Monomeric markers are much preferred for biological applications, as the fusion of a dimeric or tetrameric RFP to a target cellular protein can drive aggregation or clustering of the linked protein target (7). Five RFPs have been claimed to have been monomerized: DsRed, dKeima, eqFP578, eqFP611, and HcRed recently by our group (3, 8-11). The means by which the oligomericity of the designed variants has been measured, however, are inconsistent between these studies. Recent *in vitro* and *in vivo* studies have hinted that many so-called “monomeric” RFPs are in fact relatively high-affinity dimers (7, 8). A large family of engineered RFPs derived from eqFP578, which includes some of the brightest and most red-shifted variants, has long been claimed to be monomeric. This family was engineered from a first generation variant that called TagRFP. It was recently shown, however, that two key members of this family, mNeptune and mKate2, do not behave monomerically by HPLC. This is problematic for the claim that any member of this family, which includes mKate, mKate2, mNeptune, mNeptune2.5, and mCardinal, is in fact monomeric. There are very few mutations to either of the dimeric protein-protein interfaces between any of the family members; the vast

majority of the mutations are elsewhere on the surface or in the protein core, effecting fluorescence. In addition to *in vitro* data that seems to question the monomericity of one family of RFPs, mCherry has been implicated alongside TagRFP as being aggregation prone in live bacterial cells when fused to an *Escherichia coli* nuclear associated protein, H-NS.

We sought to directly test the oligomericity of the TagRFP family of fluorescent proteins and mCherry, both of which had been hinted to not be completely monomeric. We first attempted to assay the oligomericity via Homo-FRET, which we had used as a high-throughput test for a library of DsRed variants, but because of the high degree of variability between the various RFP cores, the read-out from this assay was not easily interpretable. As a slightly lower-throughput option, we then ran size exclusion chromatography (SEC) as an initial test of oligomericity. We ran mCherry; mCardinal – a recently reported far-red variant of mNeptune; FusionRed – a reportedly fully monomerized mKate variant; and mGinger0.1 – an HcRed variant that we had recently engineered. SEC analysis was not conclusive, but strongly suggested that mCardinal was dimeric as it eluted earlier than any other dimeric standard we ran (Figure 1A). mCherry, FusionRed, and mGinger0.1 ran close together in order of their molecular weight. SEC is not quantitative, but it does appear that FusionRed and especially mCherry run significantly earlier than mGinger0.1, enough so that it there may be an indication that mCherry has some slight oligomeric tendency, which would confirm *in vivo* results.

To test if mCardinal would run differently at lower concentrations, perhaps behaving dimerically only at high protein concentrations, we analyzed varying concentrations of mCardinal by SEC (Figure 2). The elution profile of mCardinal at 740 $\mu$ M, 7.4 $\mu$ M, and 740nM, which neared the signal-to-noise limit for our instrument, all perfectly overlap, showing that its oligomeric state is not concentration-dependent within that range. This is biologically relevant, as the average concentration of an expressed protein in a budding yeast cell is between 0.4 and 1.4  $\mu$ M, suggesting that expression of mCardinal as a fusion protein *in vivo* would necessarily cluster its linked protein target.

For further confirmation of these results, we ran a sedimentation velocity study by analytical ultracentrifugation (AUC), which is one of the most reliable biophysical assays for determining molecular size in solution. We ran DsRed as a tetrameric standard, HcRed as a dimeric standard, and mGinger0.1 as a monomeric standard. Indeed, mCardinal appears dimeric by AUC, sedimenting at ~60 kD, or about twice the molecular weight of the monomer, confirming our earlier study by SEC (Figure 1B). FusionRed and mCherry neatly overlay with mGinger0.1, contrasting with SEC data that had shown both proteins, but especially mCherry, eluting much earlier than mGinger0.1. Further study of these proteins is needed; specifically, it would be helpful to determine the dissociation constants of oligomerization, possibly with accurate fluorescence anisotropy or isothermal calorimetry.

### ***Noisy measurement of Extinction Coefficient and Quantum Yield Affects Reported Brightness***

A critical measure of the usefulness of an RFP is its brightness. Brightness is dependent on two parameters. The first, quantum yield ( $\Phi$ ), is a unitless measure of the efficiency of a chromophore at emitting absorbed light, and can be thought of as the amount of light emitted as a percentage of the total light absorbed by the chromophore. Second, extinction coefficient ( $\epsilon$ ) is a measure of how readily a chromophore absorbs light. It is usually expressed in per molar, per centimeter units, meaning that it is a measure of the absorbance at a particular wavelength of light over a one-centimeter path length by a one molar solution of the chromophore. Reported brightness values can vary widely between publications, but most publications do not remeasure previous work, which can lead to errors being propagated through the literature.

Here, we measure  $\Phi$  and  $\epsilon$  in parallel for a number of widely used red and far-red FPs. We find that the values we measure vary by as much as 50% from reported values in the literature. We tried to follow the most precise methods and used careful technique, but even more importantly, these measurements were made in parallel on the same instruments. This means that if the absolute values are not accurate, the comparative value of these measurements will allow insights from the variance of  $\Phi$ ,  $\epsilon$ , and brightness as they relate to

protein structure and sequence. We have fairly high confidence in our numbers because reported literature values cluster pretty uniformly both above and below our measured values (Figure 3). For two RFPs, mCardinal and mNeptune2.5 the reported values are significantly higher than our measured values: 63% and 53% respectively, while for mPlum and FusionRed the reported values are lower than our measured values by 41% and 21% respectively (Table 1). All other brightness measurements we made are within 20% of the reported values.

These measurements were made with a Tecan Saffire 2 platereader. We used the red dye Rose Bengal (12) as a standard and came very close to the known values for the heavily studied proteins mCherry and DsRed. The values that have only been reported once or twice in the literature tended to be much less accurate.

### ***Peak Excitation and Emission Wavelengths are Dependent on the Instrument Used***

The wavelengths at which an RFP maximally excites and emits are of high interest, as RFPs are sought that both excite and emit further into the near infrared. Instrumentation, however, can give very different measurements dependent on a number of factors. Principally, optical instrumentation is very precise and needs to be tuned regularly. Of specific note is that many photomultiplier tubes do not detect far-red photons as efficiently as shorter-wavelength photons, meaning that some instruments will undercount the intensity of far-red light. We use a fluorometer from Photon Technology International with a Xenon arc lamp that has a correction for the dampened detection at long wavelengths to measure the spectra of various RFPs (Figure 4).

For the most part we found that the reported literature values were fairly accurate. mCardinal and mNeptune2.5 were slight exceptions, with reported values overstated by 2-4 nm (13). A major outlier, however, which was ironic considering the amount of work we had put into it, was HcRed (14). The reported literature value for its peak fluorescent emission is 645 nm, reported by two groups. We found it to have maximum emission at 633 nm, which is a very large error.

***There is a Correlation Between Thermal Stability and Quantum Yield.***

Quantum yield ( $\Phi$ ) is one of the most important RFP attributes, as it is a measure of the efficiency of the chromophore at emitting fluorescent light. A major goal of NIR FP engineering is to improve  $\Phi$  in far-red monomers, as RFP monomerization causes a significant hit to  $\Phi$ , and there is also a steep negative relationship between quantum yield and  $\lambda_{em}$ . We have reported on the negative correlation between  $\Phi$  and  $\lambda_{em}$  in a previous chapter. This negative correlation is different for monomers and for oligomers, with monomers having significantly lower quantum yields at high wavelengths than oligomeric RFPs. It is known that quantum yield is correlated to the stability of the excited state of the chromophore, and because monomerization would seem to destabilize the protein scaffold, there seemed to be a good explanation for why quantum yield drops with monomerization. We showed with HcRed variants that thermal stability was a good metric for predicting the drop in quantum yield, with unstable proteins more likely to take significant hits to quantum yield than monomers that have been thermo-stabilized, usually via core optimization.

Here we measure the thermal stability of some monomeric and dimeric variants of DsRed, HcRed, and eqFP578, and show that this relationship holds (Figure 5A). As with the correlation between  $\Phi$  and  $\lambda_{em}$ , there appears to be a diagonal barrier, above which lies space unpopulated by FP engineering. Essentially, the current relationship between  $\Phi$  and apparent  $T_m$  implies that protein engineers have been working within a framework that limits the  $\Phi$  of a protein to some value defined by its  $T_m$ . However DsRed, a native tetramer, easily cruises through this barrier, with a quantum yield of nearly four times the brightest monomer, FusionRed (Figure 5B). The measurement of a broader array of RFPs is needed to further characterize this relationship. It is possible though, in the context of an *Anthozoa* class RFP, that a three-parameter analysis of  $\lambda_{em}$ ,  $\Phi$ , and apparent  $T_m$  could predict a maximum brightness for a given stability and emission wavelength.

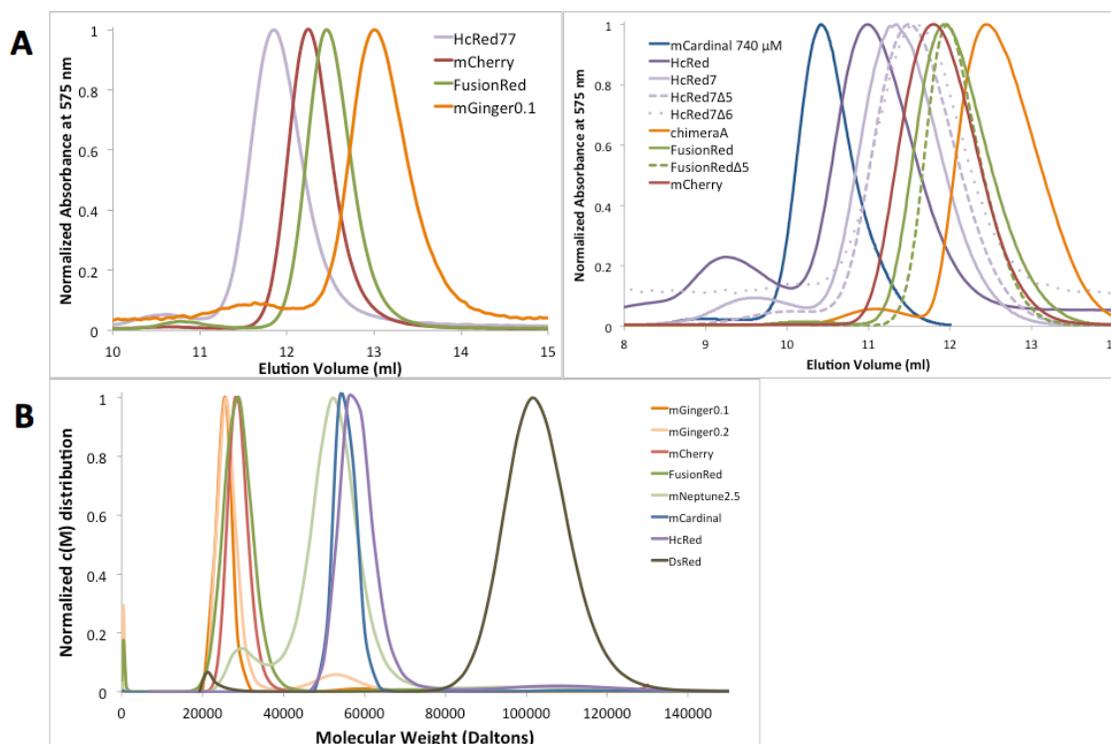
## 5.4 Conclusion

This has been a preliminary study of the characterization of RFPs. More standardization and accuracy is needed in the field for data to be useful for FP engineers. We attempt to lay out best practices for measuring  $\epsilon$ ,  $\lambda_{em}$ ,  $\Phi$ , apparent  $T_m$ , and oligomericity. Of these parameters, the most disturbingly misreported parameter is oligomericity, with a whole family of RFPs having been categorically miscategorized as monomers. Brightness values are very noisy as well, so better techniques are needed to improve accuracy.  $\lambda_{em}$  seems to be the one parameter that is fairly accurately reported.

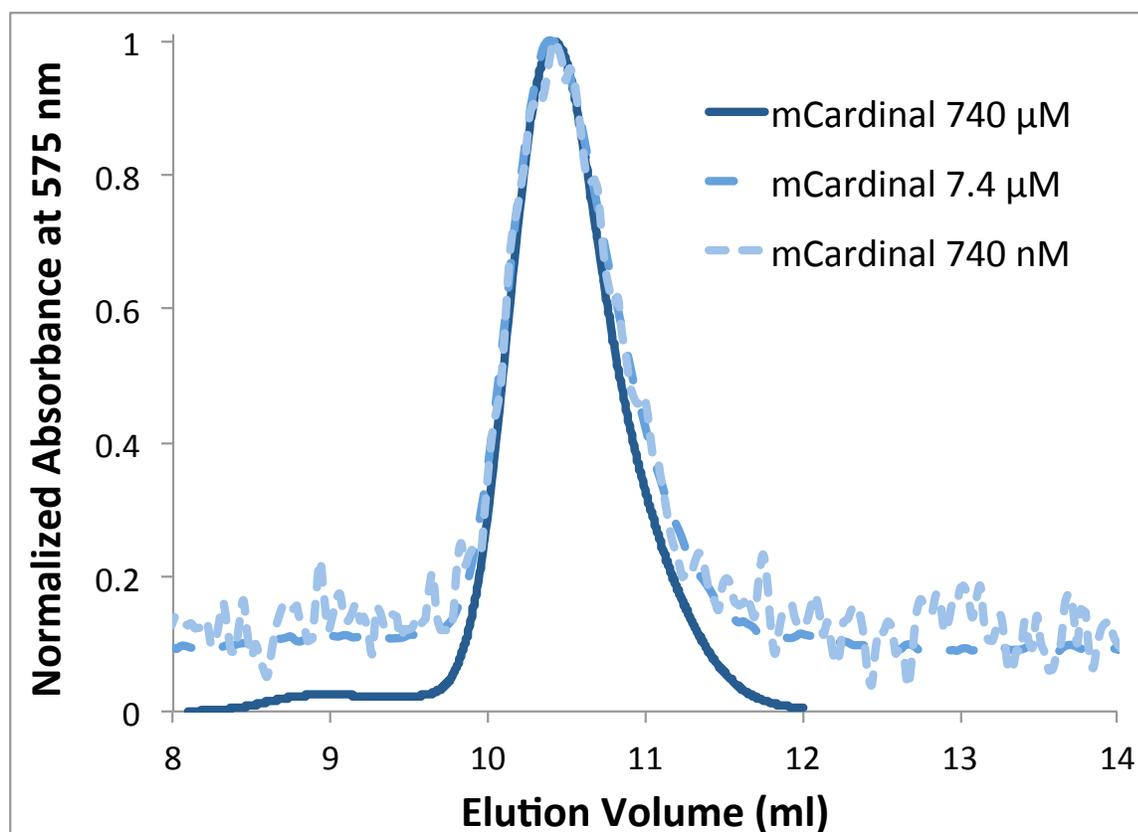
## 5.5 Materials and Methods

Materials and Methods have been described previously in chapters III and IV.

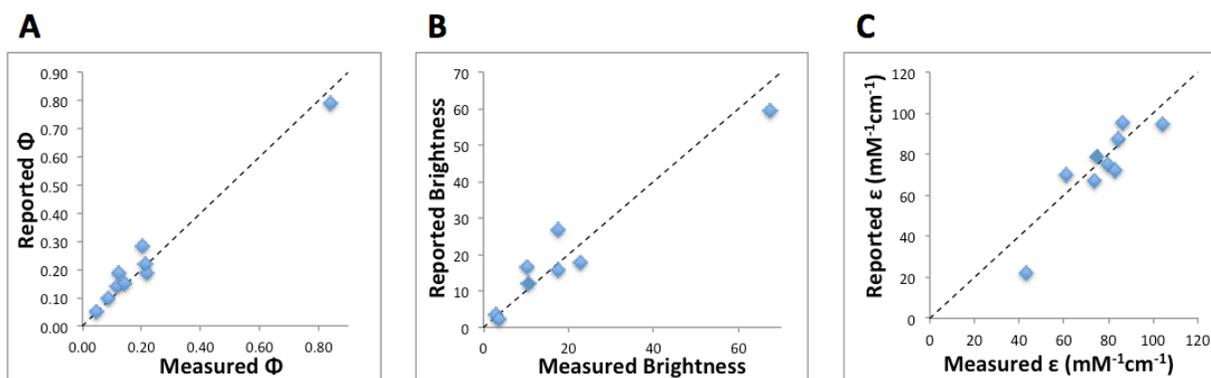
## 5.6 Tables and Figures



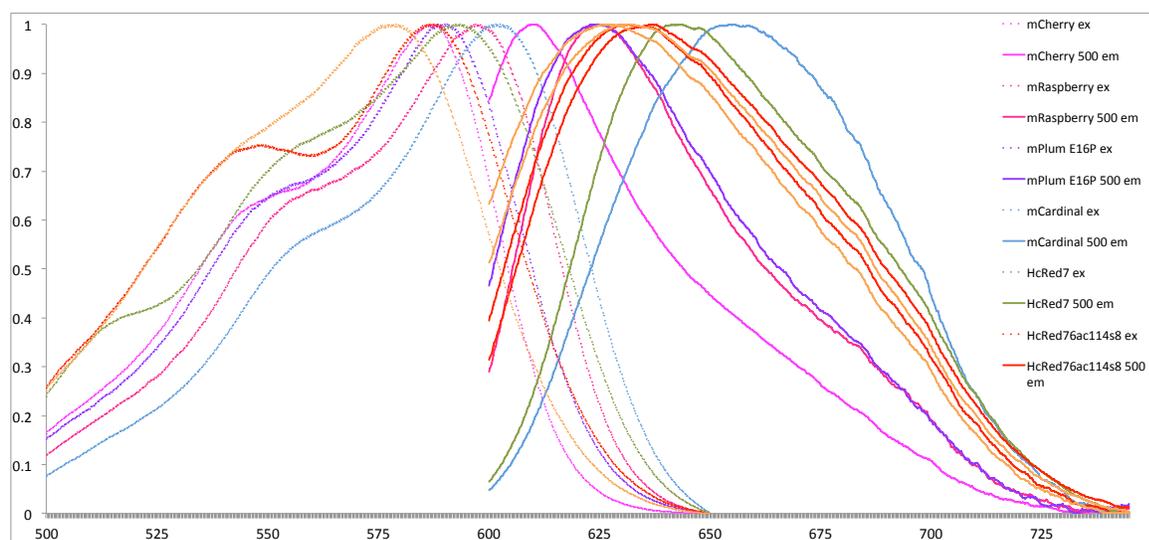
**Figure 1 – Measuring the oligomericity of RFPs.** (A) We first ran RFPs over SEC to gauge their oligomericity. SEC provides a qualitative output that shows that mCardinal is an apparent dimer, while mCherry and FusionRed are likely monomers, and mGinger0.1 is a definite monomer. (B) We then measured the apparent molecular weight by sedimentation velocity on an analytical ultracentrifuge. There is a clear distribution into monomers: mGingers, mCherry, FusionRed; dimers: mNeptune2.5, mCardinal, HcRed; and tetramers: DsRed.



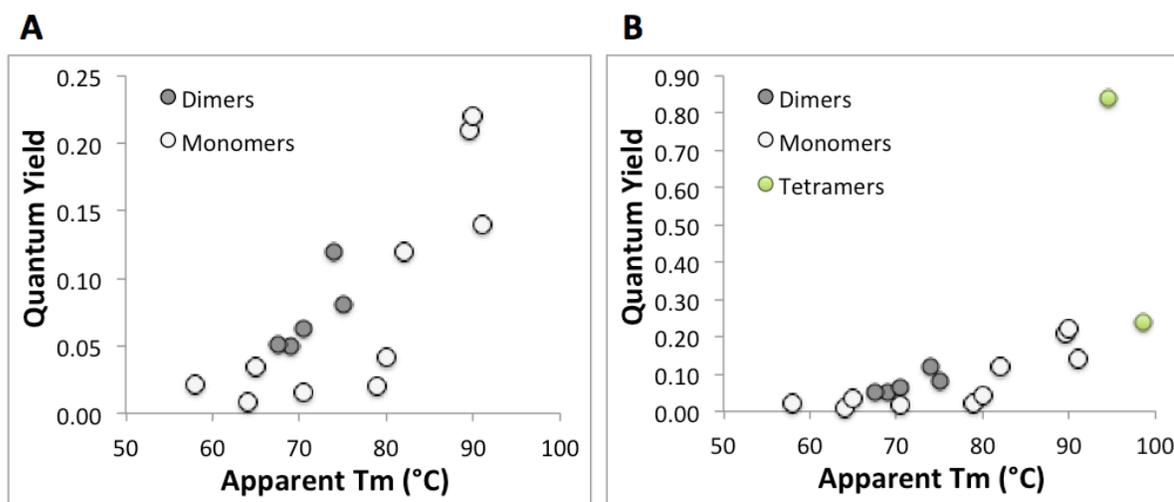
**Figure 2 – The oligomericity of mCardinal.** We analyzed three concentrations of mCardinal, spanning three orders of magnitude via SEC. We see that the peaks perfectly overlay one on top of the other, indicating that mCardinal’s oligomericity does not change over this concentration gradient, confirming that it is an obligate dimer at concentrations as low as 740 nM.



**Figure 3 – The deviation of (A) quantum yield ( $\Phi$ ), (B) extinction coefficient ( $\epsilon$ ), and (C) brightness between measured and reported values. Values seemed to cluster around equality, with reported values showing errors in both directions.**



**Figure 4 – Spectra of fluorescent proteins measured in a fluorometer and corrected for the decreased sensitivity of the photomultiplier tube at longer wavelengths.**



**Figure 4 – There is a positive correlation between  $\Phi$  and apparent Tm.**

(A) We plot this relationship for dimeric and monomeric variants of DsRed and eqFP578. (B) DsRed (bright tetrameric point), a native tetramer and DsRmCh (thermostable tetrameric point), an engineered tetramer are added to the plot. DsRed clearly breaks out of the Tm to  $\Phi$  correlation.

## 5.7 References

1. Tromberg BJ, *et al.* (1999) Non-invasive in vivo characterization of breast tumors using photon migration spectroscopy. *Neoplasia (New York, N.Y.)* 2(1-2):26-40.
2. 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.
3. Campbell RE, *et al.* (2002) A monomeric red fluorescent protein. *Proceedings of the National Academy of Sciences*.
4. Bevis BJ & Glick BS (2001) Rapidly maturing variants of the *Discosoma* red fluorescent protein (DsRed). *Nature biotechnology* 20(1):83-87.
5. Alieva NO, *et al.* (2007) Diversity and evolution of coral fluorescent proteins. *PloS ONE* 3(7).
6. Shagin DA, *et al.* (2004) GFP-like proteins as ubiquitous metazoan superfamily: evolution of functional features and structural complexity. *Molecular Biology and Evolution* 21(5):841-850.
7. Wang S, Moffitt JR, Dempsey GT, Xie XS, & Zhuang X (2014) Characterization and development of photoactivatable fluorescent proteins for single-molecule-based superresolution imaging. *Proceedings of the National Academy of Sciences of the United States of America*.
8. Shemiakina II, *et al.* (2012) A monomeric red fluorescent protein with low cytotoxicity. *Nature Communications*.
9. Kredel S, *et al.* (2008) mRuby, a bright monomeric red fluorescent protein for labeling of subcellular structures. *PloS ONE* 4(2).
10. Strongin DE, *et al.* (2007) Structural rearrangements near the chromophore influence the maturation speed and brightness of DsRed variants. *Protein Engineering, Design & Selection* 20(11):525-534.
11. Takako K, *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.
12. Rurack K & Spieles M (2011) Fluorescence quantum yields of a series of red and near-infrared dyes emitting at 600-1000 nm. *Analytical chemistry* 83(4):1232-1242.
13. Jun C, *et al.* (2014) Non-invasive intravital imaging of cellular differentiation with a bright red-excitable fluorescent protein. *Nature methods* 11(5):572-578.
14. Gurskaya NG, *et al.* (2001) GFP-like chromoproteins as a source of far-red fluorescent proteins. *FEBS letters* 507(1):16-20.