Structural and Functional Interrogation of Single Amino Acid Residues in Fluorescent Proteins

by

Jennifer L. Watkins

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved April 2012 by the
Graduate Supervisory Committee:

Rebekka Wachter, Chair
Giovanna Ghirlanda
James Allen

ARIZONA STATE UNIVERSITY

May 2012

ABSTRACT

Acquisition of fluorescence via autocatalytic processes is unique to few proteins in the natural world. Fluorescent proteins (FPs) have been integral to live-cell imaging techniques for decades; however, mechanistic information is still emerging fifty years after the discovery of the original green fluorescent protein (GFP). Modification of the fluorescence properties of the proteins derived from GFP allows increased complexity of experiments and consequently, information content of the data acquired. The importance of arginine-96 in GFP has been widely discussed. It has been established as vital to the kinetics of chromophore maturation and to the overall fold of GFP before post-translational self-modification. Its value during chromophore maturation has been demonstrated by mutational studies and a hypothesis proposed for its catalytic function. A strategy is described herein to determine its pKa value via NMR to determine whether Arg96 possesses the chemical capacity to function as a general base during GFP chromophore biosynthesis.

Förster resonance energy transfer (FRET) techniques commonly employ Enhanced Cyan Fluorescent Proteins (ECFPs) and their derivatives as donor fluorophores useful in real-time, live-cell imaging. These proteins have a tryptophan-derived chromophore that emits light in the blue region of the visible spectrum. Most ECFPs suffer from fluorescence instability, which, coupled with their low quantum yield, makes data analysis unreliable. The structural heterogeneity of these proteins also results in undesirable photophysical characteristics. Recently, mCerulean3, a ten amino acid mutant of ECFP, was

introduced as an optimized FRET-donor protein (*1*). The amino acids changed include a mobile residue, Asp148, which has been mutated to a glycine in the new construct, and Thr65 near the chromophore has been mutated to a serine, the wild-type residue at this location. I have solved the x-ray crystal structure of mCerulean3 at low pH and find that the pH-dependent isomerization has been eliminated. The chromophore is in the trans-conformation previously observed in Cerulean at pH 8. The mutations that increase the quantum yield and improve fluorescence brightness result in a stable, bright donor fluorophore well-suited for use in quantitative microscopic imaging.

ACKNOWLEDGMENTS

First and foremost, I would like to express my profound gratitude, respect, and appreciation for my advisor, Dr. Rebekka Wachter. She welcomed me into her lab and helped me make it my home. She let me make my graduate school experience what I needed it to be, and it was anything but traditional. I have grown personally and professionally with her help, encouragement and patience and I never would have arrived at this point without her guidance and mentorship. I'd also like to thank my committee, Drs. Giovanna Ghirlanda and Jim Allen, for their input and guidance along this journey. The Wachter Lab has been a wonderful resource for science and life lessons, so thank you to Alison, Alan, Dayna, Agnieszka, Suratna, Han, and Nathan. Also, thank you to my undergraduate researcher, Martine, for your invaluable assistance.

Next, I would like to thank my family, especially my husband, Josh, for helping me see this through. He is my safe place and my support, who frequently reminds me that the world is not, in fact, falling apart because I had a bad day, and who has been insistent that I'm not doing this alone. My parents, for always believing in me and pushing me to be best I can be. Also, to my little brother, Eric, for being so darn smart that I had to earn a PhD so I can prove we are related.

I also owe Dr. Scott Lefler a great debt of appreciation for hiring me as a teaching assistant every semester for nearly six years. Without his invaluable experience, scenic pictures, and sage advice, I may never have gotten past my third year.

Last, but by no means least, I would like to thank all of my friends, and coworkers (many of whom fall into both categories) who have seen me at my best and my worst and, for some reason, have never given up on me, **thank you all!** I could never have reached so high without such a solid foundation supporting me.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SCHEMES

Chapter 1

A PRIMER ON FLUORESCENT PROTEINS

Protein Structure and Function

The relationship between a protein's structure and its function lies at the core of biochemistry. Proteins with similar sequences and tertiary structures often perform similar functions in analogous manners, while sequentially divergent proteins often have no functional relationship. However, nature has engineered redundancy into its protein repertoire by allowing distinct structural groups to overlap in function. While the current interpretation of these relationships seems straightforward, a detailed understanding has not yet been achieved and thus accurate prediction of protein folding remains elusive. There is, however, enough knowledge about protein folding to guess at the functional implications of intentional mutations. This process is called rational design and is primarily based on the three-dimensional structure of a protein. As long as structural data exist, mutations can be planned and executed to manipulate the protein's function via sequential and therefore structural change. In the absence of a structural starting point, random mutagenesis can be performed on a protein construct to create a variety of unique proteins, each molecule containing random amino acid substitutions. To this end, scientists have been building an intellectual rulebook of the intra- and inter-protein interactions that lend themselves to function.

Green Fluorescent Protein

Structure of GFP. Nowhere have these methods of protein exploration been more prevalent or immediately gratifying than in the interrogation of the

functional amino acids in green fluorescent protein (GFP). This protein, first isolated in the early 1960s from the Pacific jellyfish *Aequorea Victoria,* exists in the ring around the jellyfish umbrella and acts as a light acceptor from a chemiluminescent protein, aequorin (*2*). GFP absorbs light at 475 nm and emits it at 503nm via the radiative energy conversion of an internal fluorophore. The structure of GFP was first described in 1996 (*3, 4*), but not before GFP had firmly embedded itself as a potential tool to change the way fluorescence microscopy was used (*5, 6*). The intrinsic ability of GFP to spontaneously form a light-emitting moiety without external enzymatic intervention made it an excellent tool for recombinant *in vivo* tagging technology. The *gfp* gene could easily be appended to that of a protein of interest and upon transformation into a live cell and subsequent expression of the gene fusion, the GFP fluoresces (*6*). The evolution of fluorescence transpires independent of cell line, and thus is widely applicable in cell biology applications. The colored light emission is readily seen by the naked eye, and when viewed microscopically, specific proteins can be monitored on a subcellular level to determine localization of the protein within cells, movement of organelles, and, once other colors had been developed, interactions between different cellular components.

The first published structures revealed that the overall protein fold consists of eleven beta-strands that assemble to form a rigid cylinder around a distorted helix which coaxially traverses the barrel-shaped structure (Figure 1.1A)(*3, 4*). This exterior construction not only houses the central helix, but also functions as a barrier, effectively partitioning the bulk solvent from the interior of the protein.

The barrel is capped at either end by short sections of α-helix, completing the segregation of the interior region from the exterior.

Within the GFP central helix is a chromophore derived from three consecutive amino acids: Ser65, Tyr66, and Gly67. This tripeptide undergoes a series of spontaneous, autocatalytic, post-translational modifications to form a pi-conjugated heterocycle with a phenol adjunct (Figure 1.1B). Positioned near the functional tripeptide are several key amino acids that confer various benefits toward the process of chromophore formation. However, before any of the chemical reactions that create the chromophore can occur, the protein must be folded with the immature chromogenic tripeptide encapsulated within the structure. The peptide backbone will not spontaneously cyclize unless the three residues are positioned appropriately in GFP's rather unconventional central helix(*7*, *8*).

Alpha helices fold in a very regular and predictable manner, with backbone amides forming hydrogen bonds with backbone carbonyls to reinforce the helical repeat. The definition of an alpha-helix requires that intra-helical hydrogen bonds occur between an amino acid carbonyl and the nitrogen of the amino acid four residues ahead ($i + 4$). GFP's central helix is imperfect, bending about 80 degrees near Tyr66, and bringing Ser65 and Gly67 into close contact(*8*–*10*) (Figure 1.2). Coincidentally, this conformation positions the functional tripeptide in the perfect arrangement to initiate cyclization and preorganizes the central helix for chromophore maturation. The crystal structures of precyclized GFP variants show that the helical distortion brings the carbonyl of Ser65 and the

amide nitrogen of Gly67 closer than their combined van der Waals radii and primes them for nucleophilic attack(Figure 1.2, B)(*10*).

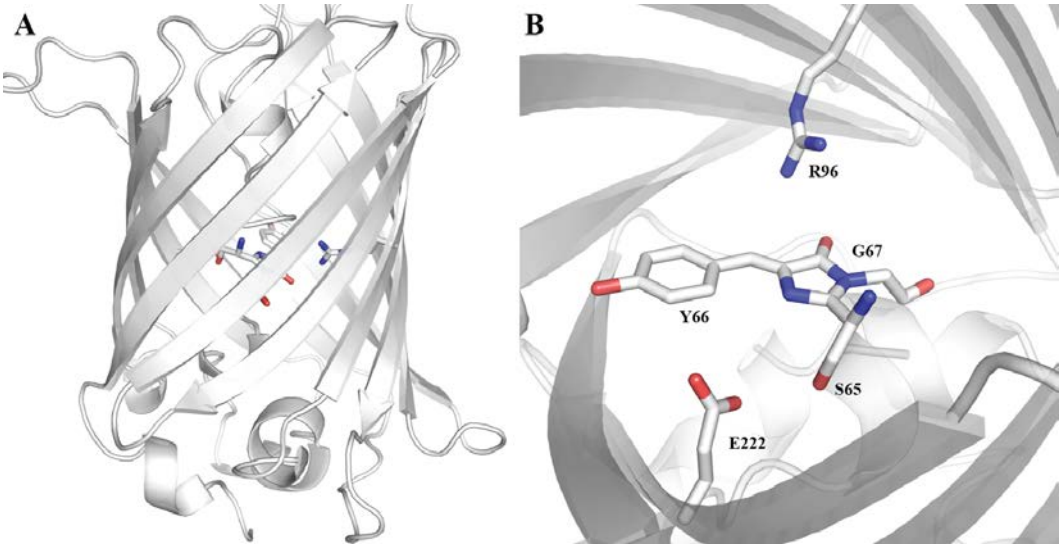


Figure1.1. Structure of wild-type GFP. A) The eleven-stranded β-barrel surrounding a coaxial central helix. B) The mature chromophore of wtGFP and the two highly conserved amino acids, Arg96 and Glu222 (PDB: 1EMB).

The presence of a hydrogen bond between an amino acid carbonyl and the nitrogen two residues ahead (i+2) as in this example is termed a "tight turn" conformation(*11*). Although by no means unique, this fold is rarely observed within proteins; more often they are found superficially, and if they are buried within the protein they tend to be rather flexible, not rigidly held as in GFP (*11*).

Besides the tight turn helical interaction, there are three additional intra-helical hydrogen bonds: 60-64 and 61-65 which are α-helical (*i*+4) and 68-71 which is a 3_{10} helical interaction (*i*+3)(Figure 1.2)(*9*, *10*). This observation shows that of the twelve possible hydrogen bonds in a twelve amino acid α-helix, the distortion of this one precludes all but three. Conveniently, the hydrogen bond network observed in the precyclized structures is identical to that in the mature

protein(*9*, *10*)(Figure 1.2). This represents an energetic advantage toward the formation of the chromophore in that GFP avoids the enthalpic penalties of hydrogen bond breaking during catalysis.

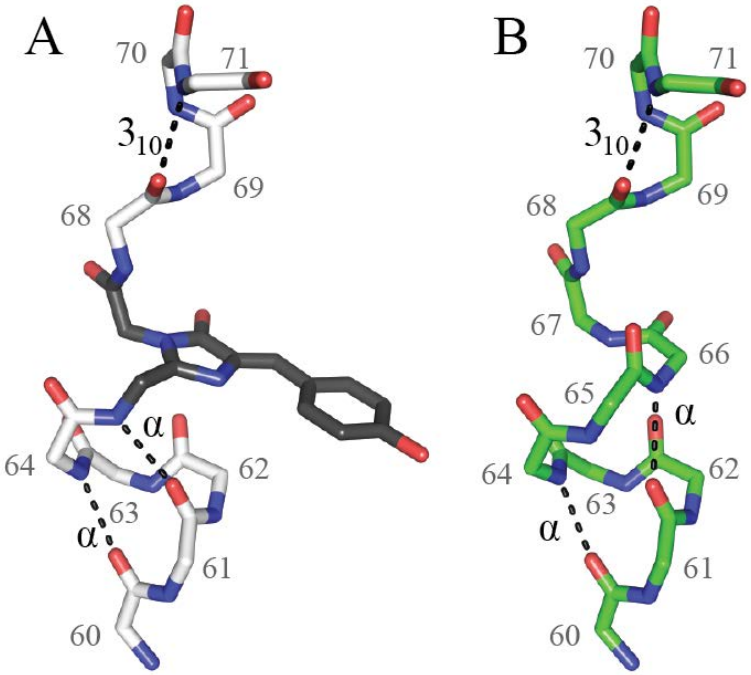


Figure 1.2. Hydrogen bonding within the GFP central helix. A) The structure of the central helix residues (60-71) in mature, wtGFP showing the light-active chromophore (dark grey) (PDB: 1emb)(*12*). B) The structure of the central helix of a GFPsol variant containing a compromised chromophore-forming tripeptide (GGG) showing the backbone conformation of the central helix in its precyclized state (PDB: 1qyo)(*10*). There are three main chain hydrogen bonds in each (dashed lines), two α helical, and one 3_{10} helical in the mature structure. Also of note is the presence of the tight-turn conformation in both structures, showing that the architecture of GFP positions the amino acids for backbone condensation.

Of the three contiguous amino acids that become the chromophore, positions 66 and 67 absolutely require conservation for fluorescence to occur(*13*, *14*). Gly67 is likely integral to the tight turn conformation by virtue of its flexibility(*11*), while at position 66, any aromatic amino acid will produce a

fluorescent product(*5*). Although changes to the immature, chromogenic tripeptide do not impact the kinetics of cyclization and ring oxidation, they do affect the spectral properties. This implies that the architecture of GFP is responsible for the autocatalysis, but the identities of the amino acids determine the fluorescent properties(*15*).

Mechanism of Chromophore Formation. The mechanism by which GFP fluoresces has been under scrutiny for as long as the protein has been studied. Although our present knowledge is extensive, the complexity of GFP's photophysics means fundamental questions remain unanswered. Studies on the mechanism of chromophore formation have given perspective into the chemistry and, in particular, the sequence of reactions relevant to chromophore biosynthesis from the immature tripeptide. The most compelling evidence points to a cyclization-oxidation-dehydration mechanism (Scheme 1.1) (*14*, *16–18*). After folding to form the principal structure of immature GFP, the first step in the formation of the chromophore is nucleophilic attack on the Ser65 carbonyl carbon by the Gly67 amide nitrogen (Scheme 1.1(a), Figure 1.3). This creates a 5-member imidazolone heterocycle from the backbone atoms of the functional tripeptide (Scheme 1.1(b)). This step is reversible, but the backbone condensation becomes trapped by the subsequent steps(*10*, *16*).

Structural investigation into the chromophore biosynthesis mechanism requires the use of trapped, immature variants of GFP. A popular background protein, mGFPsol (also called GFPtrix) (*16*), is a highly soluble, monomeric, and fast maturing GFP derivative. It contains the fluorescence-enhancing mutation

S65T, the folding-improvement substitutions F64L/F99S/M153T/V163A, and the monomerizing change A206K. Structural data from the GFPsol-S65G/Y66G mutant (dimeric, soluble GFP, containing a GGG chromophore-forming tripeptide) confirmed the hypothesis that intra-chain cyclization is the first step toward chromophore maturation(*10*). This variant, although folded correctly compared to wtGFP, did not exhibit main-chain cyclization when expressed in the absence of oxygen. However, in the presence of oxygen, it crystallized with a cyclized backbone(*10*). This was suggestive of a mechanism in which oxidation occurred before dehydration, since the oxidation event, catalyzed by the presence of molecular oxygen allowed the trapping of the cyclized backbone.

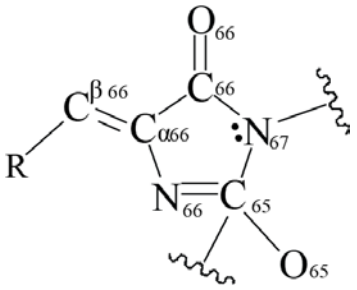
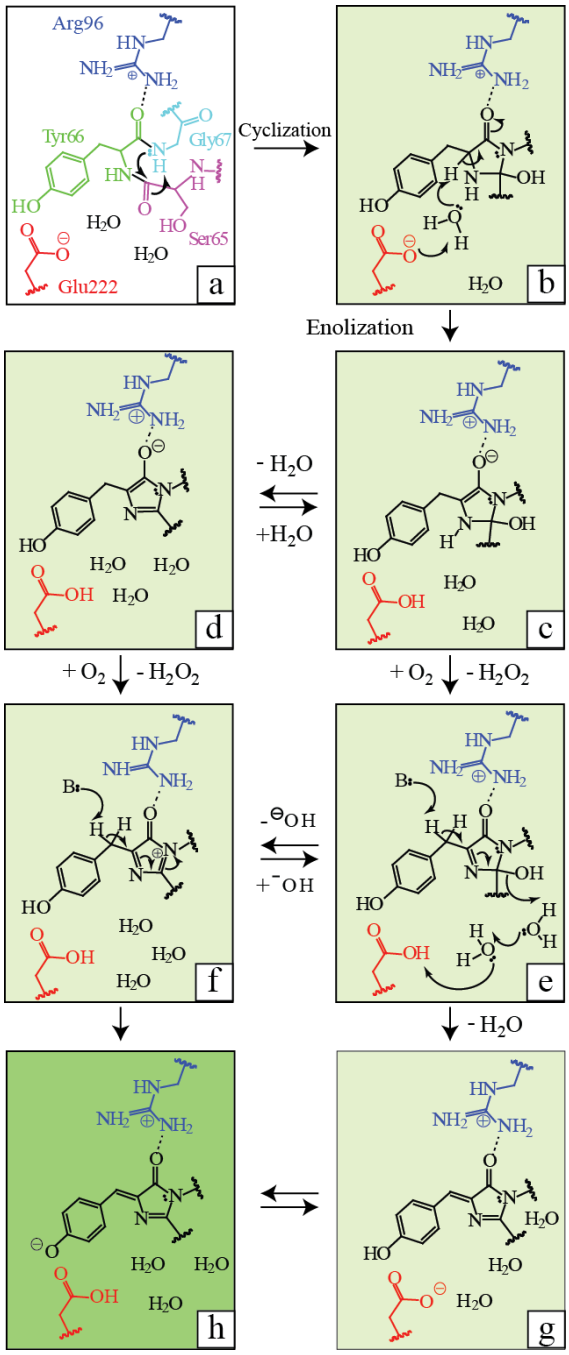


Figure 1.3. Atom labeling of the GFP heterocycle as it will be used in this thesis. R will either be representative of a phenol moiety (GFP) or an indole group (CFP and its derivatives).

Additionally, mutants that compromised the Arg96 side chain showed slow maturation, limited by a lack of backbone cyclization to initiate catalysis(*15*). These results show that the immature, functional tripeptide will not undergo any of the maturation reactions in the absence of a cyclization event and therefore, main chain condensation must be the first step after protein folding is complete.



Scheme 1.1. Mechanism of Chromophore maturation in GFP. The main chain cyclization (conversion of (a) to (b)) initiates the maturation process. The resulting cyclic α-enolate may exist in a hydrated (c) or dehydrated (d) state. Molecular oxygen oxidizes the ring creating hydrogen peroxide and the cyclic imine again either hydrated (e) or dehydrated (f). Proton abstraction from the tyrosyl β-carbon yields the mature chromophore, either neutral (g) or anionic (h). Figure adapted from ref (*17*)

Another GFP variant, EGFP-Y66L (GFP-F64L/S65T/Y66L), was used to investigate whether the presence of an aromatic amino acid at position 66 was absolutely required for chromophore maturation(*14*). The x-ray crystal structure revealed a trapped tetrahedral intermediate with the hydroxyl leaving group still attached to C65 (Scheme 1.1, species (c)). This finding confirmed that the aromatic side chain of the functional tripeptide was unnecessary for the backbone condensation reaction to occur(*14*).

In addition to the tetrahedral intermediate of EGFP-Y66L, $C_{\alpha 66}$ exhibited trigonal planar geometry, proving that the ring had already been oxidized, even though the $C_{\alpha 66}$-$C_{\beta 66}$ bond of the tyrosine residue had not yet been deprotonated (*14*). This oxidation event traps the heterocycle in its condensed state. Because the oxidation was apparent in the crystallized tetrahedral intermediate, the dehydration step can be assumed to be the final process in GFP chromophore maturation, requiring the presence of an aromatic residue at position 66 in the protein.

The last step in chromophore maturation is the deprotonation of the β-carbon of Tyr66 in the elimination of water from the heterocycle. This reaction is thought to be initiated by base catalysis, although the identity of the base is as yet unknown(*18*, *19*). Once the hydroxylated cyclic imine is subjected to dehydration, removal of the hydroxyl from the ring conjugates the heterocycle and completes the pi system affording GFP its signature fluorescence.

Kinetic isotope effect studies on the variant mGFPsol, coupled with mass spectroscopy (MS) interrogation of the reaction progress, have confirmed that the

sequence of events is as described herein (*16*, *18*). During the process of mGFPsol maturation, mass spectroscopy results showed a total of three peaks: one representing the native protein with an immature chromophore, one with m-2 and a peak with m-20. Had dehydration occurred before deprotonation of the tyrosyl α-β bond, an m-18 peak would have been observed (*16*, *18*). In tandem with the MS experiments an enzyme-coupled colorimetric assay was used to quantitate the production of hydrogen peroxide independently from the acquisition of green fluorescence(*16*). Hydrogen peroxide is released from the active site as a product of two electron oxidation by molecular oxygen (Scheme 1.1 (c) to (e) or (d) to (f)). The evolution of hydrogen peroxide occurred before the fluorescence appeared; a result in agreement with dehydration proceeding as the final step in fluorescence acquisition (Scheme 1.1 (g) or (h). The mature chromophore exists in equilibrium between two states, neutral A-form (Scheme 1.1 (g)) or anionic B-form (Scheme 1.1 (h)). The mature GFP chromophore typically emits green light from the B-form, but conversion from A- to B-form can occur via excited-state proton transfer (ESPT) upon excitation of the A-form.

The interior ligand binding pocket utilizes amino acid side chains and crystallographically conserved water molecules to position the mature chromophore appropriately within the interior of the barrel(*20*). The somewhat matched nature of the pocket to the ligand shape ensures limited rotational flexibility of the mature chromophore and encourages the dissipation of excited-state energy as light, rather than vibrational heat loss. The isolation of the fully-formed chromophore from the solution in which the protein resides confers a

resistance to quenching by small molecules and allows the protein to remain brightly glowing in a variety of chemical environments(*20*). It also forces the immature chromophore to participate in hydrogen bond interactions with neighboring amino acids which can then function as general bases or acids during chromophore maturation. Two of the most important such amino acids are Arg96 and Glu222, which are both strictly conserved in all known GFP-like proteins.

The Importance of Arginine-96. Arg96 in GFP is thought to perform a myriad of functions and thus becomes irreplaceable in GFP and its descendants. While GFP is folding, Arg96 is positioned above the chromophore, forming a hydrogen bond between one of its guanidinium nitrogens and the carbonyl oxygen of Tyr66 which after cyclization becomes the exocyclic imidazolinone oxygen, O_{66} (Scheme 1.1, Figure 1.3). The formation of two main chain hydrogen bonds are prevented by this interaction lest they require breaking during chromophore maturation: Thr62-Tyr66 and Thr65-Val68 (Figure 1.2) (*8*, *10*). This was proven in the crystal structure of GFPsol-R96A, where the two additional helix-enforcing hydrogen bonds were observed (*10*). These two hydrogen bonds are not present in the structure with the cyclized chromophore intermediate(*14*), nor in the precyclized mutant with the GGG chromophore sequence, which does exhibit the 80^{o} bend in the central helix, but has the same hydrogen bonding pattern as wtGFP (*7, 10*). Removal of the bulk of the side chain in GFP-R96A results in conformational compensation in the protein interior; Tyr66 undergoes a large conformational change within the chromophore binding pocket in the absence of the arginine side chain (*10*). Therefore, without Arg96 sterically contributing to

the positioning of the chromophore-bearing helix in the R96A mutant, these two hydrogen bonds require breaking before cyclization can occur, contributing to a higher activation energy of cyclization and the slower rate of cyclization observed in the Arg96-compromised mutant (*10*, *15*).

The removal of the positively charged side chain as in the GFP-R96A mutant (*10*) resulted in retarded maturation rates compared to wtGFP which were capable of being restored by the insertion of an arginine in a nearby location (Q183R) or reinstatement of the positive charge with a lysine (R96K)(*15*). In an effort to maintain steric interactions but eliminate the charge contribution of Arg96, GFPsol-R96M was designed (*13*). This change maintains the approximate volume of the arginine side chain, but lacks the positive charge. This mutation dramatically reduced the rate of chromophore maturation, and structural data was collected on both the immature and mature forms. The structures show that R96M precludes chromophore biogenesis and that the construct is predominantly trapped in a precyclized state (*15*) suggesting that Arg96 plays a role in the formation or stabilization of the cyclized, nascent chromophore. The requirement for a positive charge at or near position 96 during chromophore maturation highlights the electrostatic importance of this particular arginine in chromophore biosynthesis.

Likely, the hydrogen bond between the Arg96 $N_{\eta 2}$ and the carbonyl of Tyr66 stabilizes a buildup of negative charge on O_{66} and acidifies the proton on $C_{\alpha 66}$ for subsequent oxidation. Arg96 has also been proposed as a general base in

the final dehydration step of GFP chromophore biosynthesis (*18*). This hypothesis and its implications will be further discussed in Chapter 2.

Spectral Tuning of FPs

The mutagenic approaches used for the elucidation of chromophore maturation processes have also exposed the importance of key amino acids near the chromophore. The signature fluorescence of GFP originates from the chromophore, but its characteristics are closely related to the environment in which the chromophore resides. The simplest and most effective means of evolving spectral diversity in GFP-like proteins is to modify the pi-conjugated system itself. Changing the tyrosine to another aromatic amino acid reconfigures the light-bearing moiety and, as expected, the color. For example, the extension of the pi-system by substitution of an indole at the phenol location red-shifts the fluorescence relative to the neutral, A-form of the GFP chromophore. This substitution (Y66W) produces cyan fluorescent protein (CFP) while another, Y66H, creates blue fluorescent protein (BFP) (*5*).

However, more than the chromophore structure affects the fluorescent properties of the protein. Amino acids that interact with the chromophore directly or via bridging waters can affect changes in the optical properties of the protein. The mature GFP chromophore in solution, lacking non-covalent protein interactions, possesses a low quantum yield compared to the same chromophore structure encapsulated within the protein matrix; so the surrounding protein matrix can be assumed integral to fluorescence characteristics (*20*). When the contacts between the chromophore and protein are modified, the spectral

properties of the protein change (*5*). This is indicative of the chromophore structure in conjunction with the surrounding environment altering the fluorescent behavior of the protein. A prime example of this influence is the T203Y mutation which results in yellow fluorescent protein (YFP) (*21*, *22*). The introduction of the phenol side chain creates pi-stacking interactions with the wtGFP chromophore, modifying the fluorescent characteristics. Changes to the non-covalent protein contacts to the chromophore allow the fine-tuning of fluorescence emission in GFP-like proteins. The extensive hydrogen bond network in which the chromophore is situated determines both the excitation and emission energetics of the protein. From these seemingly minor protein modifications, an entire spectrum of FPs has evolved.

The effects of single amino acid substitutions are not limited to color; the photophysical properties such as the lifetime, quantum yield, extinction coefficient, and even fluorescence decay kinetics can be modified as well. The substitution of Ser65 with a threonine simplifies the fluorescence spectrum of wtGFP by breaking a hydrogen bond between the serine and Glu222 (*4*, *6*). The absence of this hydrogen bond destabilizes the anionic glutamate and shifts the equilibrium of the mature chromophore toward the presence of the fluorescent, anionic fluorophore.

The generation of colors spanning the entire visible spectrum from the original GFP construct has spurred the revolution of cellular imaging techniques (*23–26*). The ability to distinguish between colors and the capacity to follow energetic interactions between colored species offer researchers more options for

live-cell imaging than low molecular weight fluorophores. Since GFP autocatalytically generates its fluorescence, its research value *in vivo* lies in recombinant fluorescent gene tagging.

Attachment of the *gfp* gene directly to a gene of interest and subsequently expressing the hybrid allows researchers to monitor expression and lifetime of the protein in real-time. Additionally, since the fluorescent signal is readily viewed microscopically, localization of the protein chimera can also be followed on a subcellular level. GFP does not interact with any subcellular components and by virtue of the covalent link between the protein and the probe, the FP will act as a beacon to report on its partner protein (*24, 27*). The power of recombinant GFP-labeling is compounded with the concurrent use of a variety of FPs. The presence of multiple colors of FPs, each reporting on a unique cellular protein can lead to elucidation of protein-protein interactions in living cells. Ostensibly, the co-localization of FPs is only tenuously indicative of actual protein interaction, but upon deeper observation, the experimentally derived energetic parameters of the FPs can provide evidence for more than merely protein movement(*28*).

A process called Forster Resonance Energy Transfer (FRET)(*23–25*) has been revitalized by the explosion of new FPs. It relies on the resonant transfer of energy between two molecular probes when they are within a specific distance of each other. As a primary requirement, the donor probe is fluorescent although most commonly, both probes tend to have fluorescence capability. When they are close to one another in the cell, the excited state energy from the donor is transferred nonradiatively to the acceptor. Donor fluorescence is monitored, and

when they are able to participate in FRET interactions the donor fluorescence lifetime decreases and overall fluorescence is quenched (Figure 1.4). FRET partners are chosen based on several sets of paired characteristics. Most importantly the emission spectrum of the donor must overlap significantly with the excitation spectrum of the acceptor. Secondarily, the dipoles of emission must not be perpendicular to one another. Finally, the spatial separation of the two molecular probes cannot exceed a maximum distance, typically 10nm. If these criteria are met, the interaction between the two fluorophores can result in FRET signal(*24*).

In the past, FRET was carried out using small molecule fluorophores that were chemically linked to the proteins of interest. The ease of access to and recombinant labeling with FPs has breathed new life into FRET. The most commonly used FP pair in FRET experiments is CFP and YFP which, together, effectively meet the basic requirements for FRET partners. The use of dual fluorophores also engineers redundancy into experiments by allowing the fluorescence of the acceptor to be monitored as an alternative to the donor's(*23*). Excitation in this scenario utilizes the wavelength most relevant to the donor while monitoring the emission wavelength of the acceptor. When the acceptor lights up, it is presumed a result of FRET and then the frequency and duration of the protein interactions can be recorded. This method is not frequently used, however, due to inadvertent excitation of the acceptor fluorophore and the consequent noise in analysis (*24*).

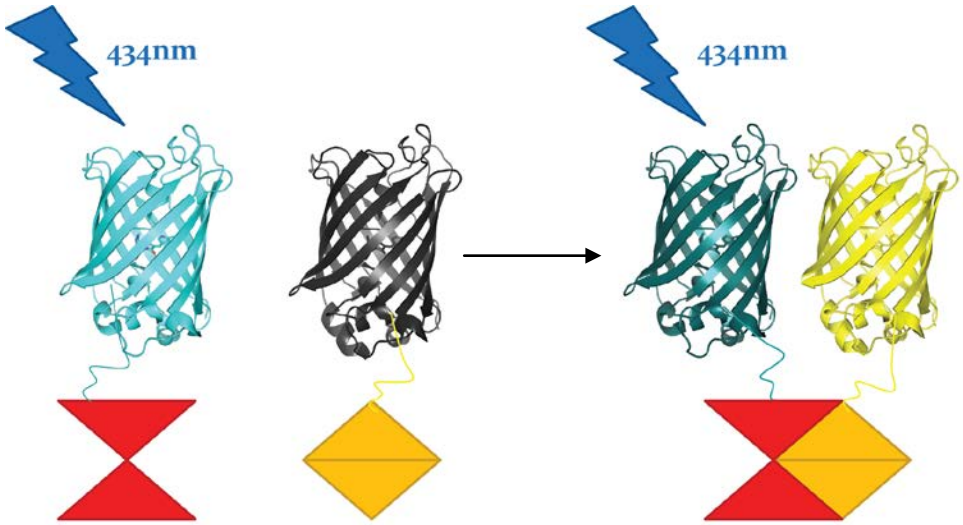


Figure 1.4. FRET energy transfer. Two proteins of interest (red and gold) are covalently tagged with the fluorescent proteins, in this case CFP (PDB: 2wso(*29*)) and VenusFP (PDB:1myw(*30*)). When the two tagged proteins interact with one another, donor fluorescence decreases while acceptor fluorescence increases.

CFP, although more advantageous in color than GFP in FRET experiments, has poor fluorescent characteristics(*23*, *31*). Its quantum yield and extinction coefficient are reduced compared to current GFP constructs due to protein isomerization and instability(*22*, *28*, *32*). Efforts are underway to design a better CFP for use in paired FP FRET experiments and the newest developments are discussed in Chapter 3.

Chapter 2

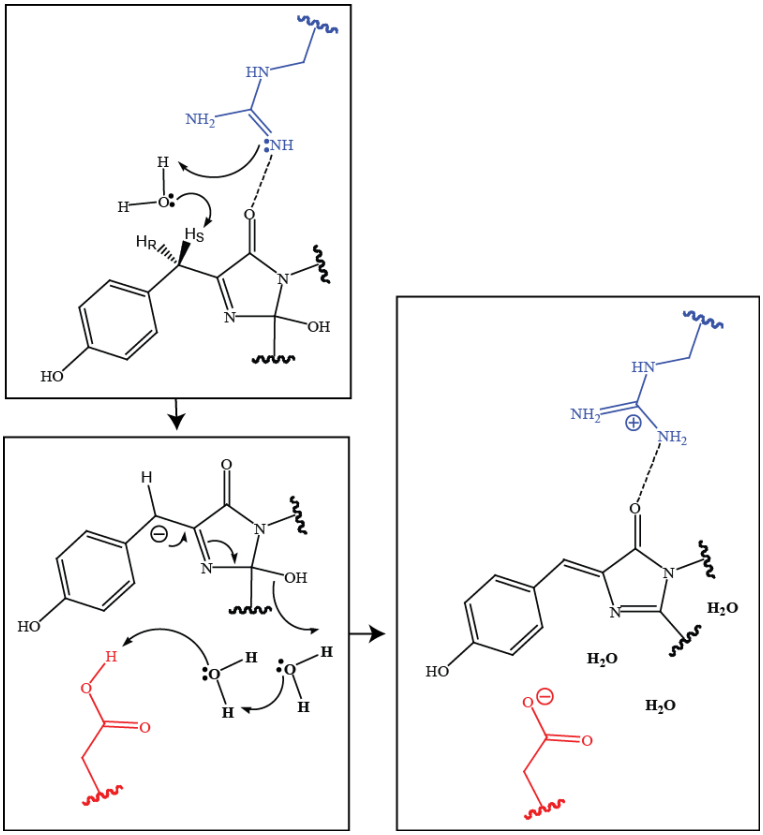
ARGININE-DERIVED BASE CATALYSIS

INTRODUCTION

Is Arginine-96 involved in GFP's chromophore maturation? GFP is a unique protein in many ways: it autocatalytically derives its fluorescence, uses complex architecture to enable its own post-translational modification, and even has the capacity to be spectrally tuned or spontaneously change color. But as extensively as researchers have studied it, GFP maintains some secrets about its functional properties. The chemical steps of the mechanism are known, and the order in which they occur has been demonstrated, but the identity of the species responsible for base catalysis remains unknown(*9*). Arg96, which is positioned within the GFP barrel, has been functionally demonstrated as an important amino acid in the acquisition of fluorescence with a high degree of conservation among the entire family of GFP-like proteins (*9*, *15*, *16*, *20*).

Probing the mechanistic importance of this basic amino acid has led researchers to believe that the importance of Arg96 is four-fold. Its steric effects within GFP's β-barrel help to position the central helix for cyclization(*8*, *10*, *15*, *33*). The positive charge on arginine's guanidinium group electrostatically forms or stabilizes the cyclized protein backbone facilitating chromophore maturation (*13*, *15*, *33*). The spectral properties of GFP variants are also affected by the substitution of Arg96, showing that the side chain plays a role in tuning the protein's color, chromophore pK_a, and fluorescence lifetime (*15*, *33*). Its positive charge may also help stabilize the B-form of the chromophore, lending itself to

the maintenance of the fluorescent state. More recently, Arg96 has been implicated in base catalysis of the final dehydration step of chromophore biogenesis (*18*, *33*).



Scheme 2.1. Proposed mechanism of general base action by Arg96. Arg96 has been proposed to stereospecifically deprotonate $C_{\beta 66}$ via a general base mechanism involving the $N_{\eta 2}$ of the guanidinium group. Adapted from (*18*)

Investigation into the mechanism of chromophore maturation showed that general base catalysis of the oxidized nascent chromophore was required to generate GFP's signature green color (*18*). Kinetic isotope effect experiments were performed to assess GFP's self-processing reactions and revealed that the species responsible for base catalysis had an apparent pK_a of about 9.4 in the EGFP-E222Q mutant (*18*). Previously attributed to the Gly67 amide (*33*), this

construct proved that theory impossible because, in this variant, most of the protein exists in the oxidized cyclic imine form. Based on the structural data available, one hypothesis is that Arg96 is activating a conserved water which is then stereoselectively abstracting a proton from $C_{\beta 66}$ (Scheme2.1)(*18*). The carbanion thus formed initiates dehydration of the heterocycle. This mechanism also proposes the donation of a proton by Glu222 during the dehydration process. The positively charged guanidinium group on Arg96 likely complements the negative charge on the mature, anionic chromophore, stabilizing the imidazolone ring and lowering the excited-state energy. This interaction is confirmed spectroscopically in Arg96 mutants where the loss of the positive charge on the guanidinium group (R96M or R96A) or altered position of the charge (R96K) blue-shifts the spectra(*13*).

Arginine is basic, but can it function as a base? At physiological pH, basic amino acids are protonated and cannot function as bases in catalytic reactions. In special circumstances, lysines in specific protein environments have been shown to exhibit perturbed pK_a values down to as low as 6, when lysine's pK_a is usually measured at 10.5 in aqueous environments (*34*, *35*). To affect such a large change in the protonation equilibrium, the titratable group usually exists in a region near other positive charges. This example of the environmental impact on the acid-base behavior of amino acid side chains invites the question of whether an even more basic amino acid, arginine (solution pK_a 12.5), could potentially function as a general base *in vivo.*

Typically, arginines are protonated at physiological pH, rendering them useless as bases in enzymatic reactions. Some very special arginine residues, it turns out, do exhibit perturbed pK_a values that can enable functional capacity. Inosine 5'-monophosphate dehydrogenase (IMPDH), pectate/pectin lyases, fumarate reductase, L-aspartate oxidase, tyrosine-phenol lyase are some examples of protein catalysts that utilize conserved, functional arginines in their active sites(*35*). Most of these catalytic arginines exist either in hydrophobic environments or near other charged side chains.

IMPDH is probably the most relevant example to the GFP situation. To perform its natural hydrolysis reaction, IMPDH uses an arginine to activate a water molecule during the transition state of proton transfer. Mutation of the arginine to lysine reduces the rate, but not as much as complete elimination of the positive charge in the active site. The general base has been shown experimentally to titrate with an acid dissociation constant of 8, implying that the arginine in the active site could, in fact function as a general base. Other examples like pectate/pectin lyase and fumarate reductase use arginines to deprotonate carbon acids. Furthermore, the catalytic arginine is estimated to have a pK_a of 9.5, lending credence to the hypothesis that Arg96 could behave as a general base during GFP chromophore maturation. It exists in an environment that is partitioned from the bulk solvent, and the side chain is engaged in an extensive hydrogen bond network with the chromophore and nearby residues. However, the interior of GFP is not hydrophobic, nor is Arg96 involved in a salt

bridge. So elucidation of the factors responsible for the potential perturbation of the acid dissociation constant of the buried arginine would be fascinating.

The guanidinium group of the Arg96 side chain directly interacts with the chromophore and surroundings via several hydrogen bonds. Its terminal nitrogens, $N_{\eta 1}$ and $N_{\eta 2}$, are involved in two hydrogen bonds, the most important of which exists between $N_{\eta 2}$ and O_{66} of the heterocycle (Figure 2.1). $N_{\eta 1}$ is also involved in a hydrogen bond with the carbonyl oxygen of Thr62. $N_{\eta 2}$ is positioned within van der Waals distance to $C_{\beta 66}$, in the mature GFP-S65T crystal structure (*4*), and slightly closer in the crystal structure of the trapped, oxidized intermediate EGFP-Y66L (*14*). This change in distance between the oxidized intermediate and the mature, green variant implies that there may be some conformational mobility in the transition between oxidized and dehydrated states. Structurally, there is not another chemical moiety near enough to $C_{\beta 66}$ that could potentially function as a general base catalyst with a pK_a of 9.4, so Arg96 is assumed responsible for the deprotonation event triggering dehydration of the heterocycle. The interaction between the guanidinium group and its surroundings is also relevant to this investigation because steric constraints can play a role in the perturbation of the pK_a of guanidinium groups (*35*).

To examine this side chain and learn whether it is responsible for the base catalysis of $C_{\beta 66}$ deprotonation, a trapped, immature variant must be used. By using a protein without a mature chromophore, the protein interior is measured in a state that is similar to that of the reaction. Conveniently, the EGFP-Y66L mutant that was crystallized with the oxidized, tetrahedral chromophore

intermediate represents a snapshot of the local environment in which the arginine would be functioning as a general base. The chromophore maturation has been interrupted after oxidation, but before dehydration; exactly when the activity of the side chain would be required. This environment is interesting because the arginine is not participating in a salt bridge interaction, nor does the intermediate cyclic imine structure have a charge or aromaticity. This intermediate state represents a very different chemical environment than the mature chromophore binding pocket.

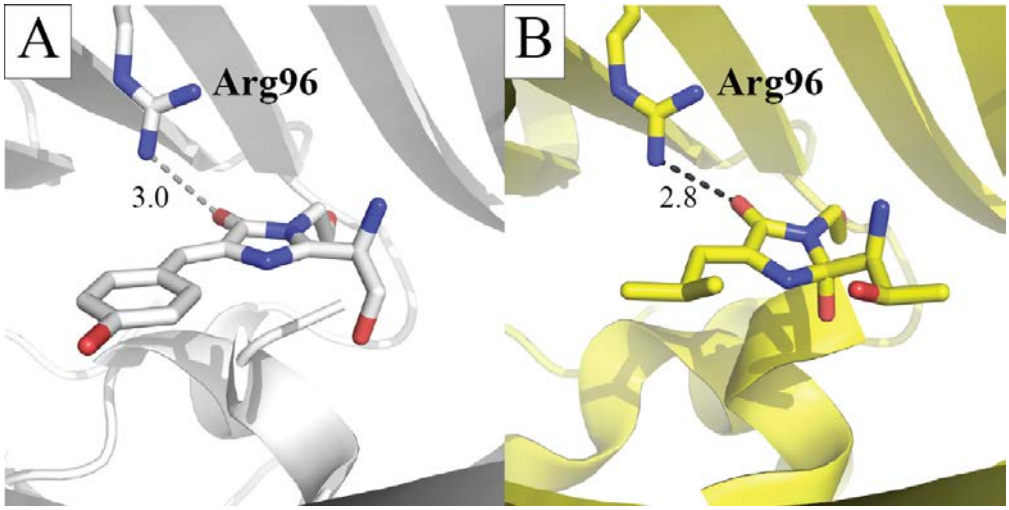


Figure2.1. Chromophore structure and position. A) wtGFP chromophore (white) and hydrogen bond between O_{66} and Arg96 (PDB: 1emb)(*12*) B) The EGFP-Y66L trapped cyclic imine chromophore intermediate (yellow) again with the hydrogen bond to Arg96 shown as a dotted line (PDB: 1s6z)(*14*)

NMR titration of Arg96 for pK_a determination. Acid dissociation constants are an excellent probe into the electrostatic environment within proteins. Direct measurement methods of protonation constants offer more insight into protein electrostatic interactions than indirect methods because they are not reliant upon secondary reporting, and are thus valuable techniques. Monitoring

individual residues provides important parameters for understanding ligand binding and catalysis. Acquisition of these data from experimental sources also aids the field of computational protein chemistry by dictating the range of pK_a values that a unique amino acid can experience and in which environments they can occur, enabling the calibration of electrostatic calculations.

Measurement of the pK_a values of titratable amino acid side chains has been standard practice for some time, though direct measurements are often difficult. Most commonly, pK_a values are derived via indirect methods or calculations (*34*). Mutagenesis of active site residues coupled with inhibition screens can provide enough information to extract a pK_a (*35*). However, since these methods do not directly observe the protonation state of the side chain, but rather the rate of catalysis or binding, care must be taken to control for changes to the local environment resulting from modification of the protein. One way to avoid such potential pitfalls is to use a direct method of measurement, like NMR.

One of the most obvious benefits of NMR is that it has the ability to monitor individual protons and therefore the capacity to directly report on the protonation state of an amino acid side chain in a site-specific manner while simultaneously reporting bulk information about the protein such as state of folding and other chemical changes resulting from pH change. This enables researchers to evaluate the catalytic competence of single amino acids. Basic amino acids are often difficult to monitor in NMR titrations because their pK_a values are so high that the protein begins to denature before the titration is complete.

André et al.(*34*) have developed a method in which they titrated uniformly labeled $^{13}C/^{15}N$ lysine residues in the *apo* calmodulin protein and extracted pK_a values for all eight lysines. They expressed $^{13}C/^{15}N$ uniformly-labeled calmodulin protein and dissolved it in 90%/10% $^{1}H_2O/^{2}H_2O$ (v/v) with small amounts of 2,2,-dimethylsilapentane-5-sulfonic acid (DSS) and sodium azide (NaN_3) for chemical shift calibration and contaminant growth prevention respectively. Beginning with protein at a concentration of 2mM in 100 mM KCl, aliquots of concentrated NaOH were added to the sample and the pH was recorded before and after data was collected to ensure consistency. The remaining volume in the NMR tube after the sample was added was filled with nitrogen gas to discourage dissolution of dissolved CO_2, which would affect pH.

Experimental spectra we collected on 500MHz or 600MHz Varian UNITY Inova spectrometers using triple resonance probes with pulsed-field gradient capability. *Apo* calmodulin data was collected at 27^{o}C in a variety of NMR experiments. 2D H2(C)N and H2C(N) correlation spectra were collected for Arg and Lys groups (80 min per experiment), 2D HD(CDNE)CZ correlation spectrum was used for Arg residues (90 min per experiment), and a 3D C-TOCSY-(CO)NH spectrum was used for assignment of the *apo* calmodulin arginine side chains (specifically, H^{δ}, C^{δ}, N^{ε}) in conjunction with the H2(CDNE)CZ and H2(C)N spectra (*34*).

These three pulse sequences were developed specifically for observing protonation states of lysine and arginine side chains. The H2CN experiment makes use of magnetization transfer from $^{1}H\delta \rightarrow {}^{13}C^{\delta} \rightarrow {}^{15}N^{\varepsilon}$ in arginine and $^{1}H^{\varepsilon}$

$\rightarrow {}^{13}C^{\varepsilon} \rightarrow {}^{15}N^{\zeta}$ in lysine via one-bond scalar couplings. In 2D mode, this pulse sequence allows for the detection of $^{1}H—^{15}N$ or $^{1}H—^{13}C$ correlation spectra of nonaromatic side chains containing labeled nitrogen atoms. In this experiment, the lysine $15N^{\zeta}$ chemical shifts were better resolved than those of $^{13}C^{\varepsilon}$ enabling observation of six of the eight residues, although only two of them were assigned during titration (*34*). The distinct arginine shifts were slightly easier to discern, with five of the expected six observed in the H2(C)N spectrum of *apo* calmodulin.

The protein was titrated across a range of pH values, from 6.5 to 11.84 with 2D $^{1}H^{\varepsilon}—^{15}N^{\zeta}$ correlations measured at each point. Six of the eight lysines began the titration with the same chemical shift in both the ^{15}N and ^{1}H dimensions (at 32ppm in the ^{15}N dimension and 3.1ppm in the ^{1}H dimension), but at titration points in the higher pH range (10.44 to 11.84), the unique chemical environments of each caused dispersion of the peaks relative to one another, allowing the derivation of acid dissociation constants for each. The authors postulate, although don't demonstrate, that this method could be applied to arginine residues as well, if their pK_a is affected enough by their local environment that the high pH range of the titration would not negatively affect protein folding for the duration of the experiments(*34*).

To determine the pK_a of an arginine in GFP using NMR, the protein must be labeled with a magnetically responsive isotope. Hydrogen is already present, but the number of signals observed in a protein would be overwhelming to interpret, so typically ^{15}N, ^{13}C or both are used as probes. The protein can be either uniformly labeled via introduction of $^{15}NH_4Cl$ or specifically labeled via

introduction of isotopically enriched amino acids. For observing Arg96, uniform labeling should suffice. GFP has seven arginines in its structure; however, only Arg96 has a side chain buried within the barrel (Figure 2.2). Therefore, it should be the only arginine side chain with the potential for a different pK_a than the solvent-exposed residues.

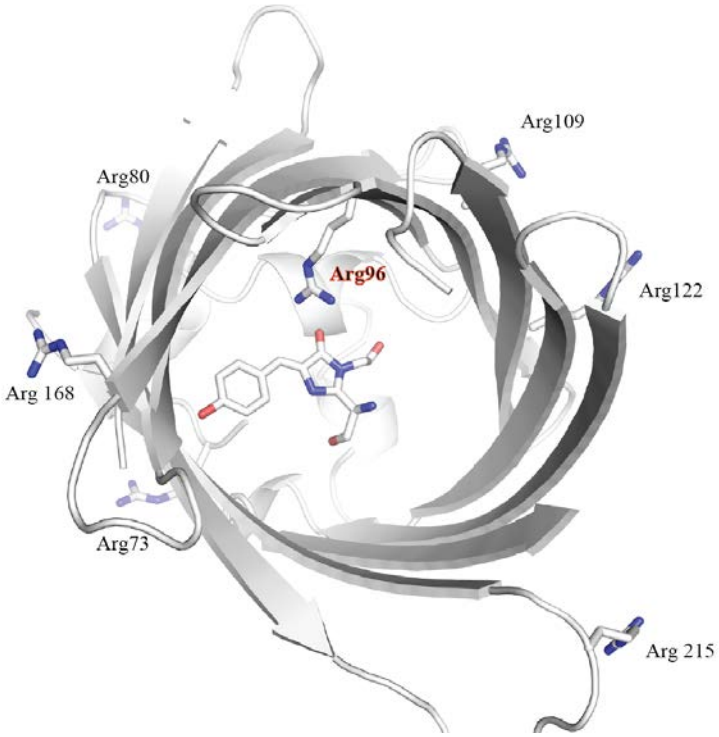


Figure 2.2. Arginines in GFP. Seven arginines are in various positions around the β-barrel. Only Arg96 (labeled in red) has a side chain buried in the interior of the barrel.

While ^{1}H NMR would be the most direct technique for measuring the acid dissociation constant of an arginine in GFP, it would not be very informative. In such a large protein, containing several charged groups, the ^{1}H response during pH changes can be highly convoluted, and difficult to extract relevant information. The ^{1}H shifts may experience multiple nearby protonation events,

and report on them, rather than the specific event of interest (*34*). For a more reliable approach, the ^{13}C or ^{15}N shifts within the side chain can be monitored directly. These groups are sensitive to the protonation state of the side chain, and will exhibit changes in chemical shift as the titration progresses. Indirect excitation of ^{1}H that are attached to or adjacent to the ^{13}C or ^{15}N atoms enhances the sensitivity of the experiment since the nonlabile protons in the side chain will be present along the full pH range of the experiment, reducing peak broadening and signal loss.

For an arginine titration via NMR, the N^{ε} or C^{ζ} shifts can be monitored even though the change in protonation will occur on one of the N^{η} groups. This is necessary due to rapid proton exchange between the solvent and guanidinium N^{η} atoms. Also, the $^{15}N^{\eta}$-$^{1}H^{\eta}$ cross-peaks tend to be very broad due to slow rotation of the guanidine head group(*34*). Because the head group charge is delocalized over the entire N^{ε}--C^{ζ}—$(N^{\eta 1}H_2)_2$ moiety, any of these atoms should be effective reporters for the protonation state of the guanidinium group of GFP's Arg96.

Identification of Arg96 should not prove too difficult. There are seven arginines in the GFP construct. All of them are solvent exposed and would therefore be expected to titrate with their intrinsic pK_a of 12. The only arginine in GFP that is buried within the protein, excluded from the bulk solvent is Arg96. It would be the only arginine side chain with an possible change in pK_a.

Observation of the chemical shift of the arginine side chain would directly correlate with the ratio of the protonated and deprotonated guanidinium groups on the arginine of interest. Plotting the chemical shift of the ^{15}N-labeled or ^{13}C-

labeled Arg96 side chain against the pH of the solution should yield a sigmoidal curve as in any titration experiment, with the point of inflection representing the observed pKa.

MATERIALS AND METHODS

mGFPsol-Y66L and EGFP-Y66L. The $pRSET_B$ vector containing the N-terminally 6xHis-tagged mGFPsol-Y66L (obtained from Di Wu)(*16*) or EGFP-Y66L(*14*) insert was transformed into BL21(De3) *E. coli* cells and expressed in LB media supplemented with 90ug/mL carbenicillin (carb). The media was inoculated with 5mL of overnight culture (LB/carb-90µg/mL, grown at 37°C with 250rpm for 15-18h) and grown at 37°C with shaking until it reached an OD_{600}= 0.8. Protein production was induced by the addition of 250 mg of solid IPTG (GoldBio, FW=238.29) following equilibration to 25°C and allowed to continue for 4 hours before the cells were harvested.

L64F Mutagenesis. The mGFPsol-Y66L construct was used to implement a reversion mutation, L64F, changes this amino acid to the wild-type residue. The PCR-based QuikChange® II site directed mutagenesis kit (Stratagene) was used to implement the L64F mutation using the DNA primers in Table 2.1. The mutagenized and subsequently digested plasmids were transformed into XL1-Blue Supercompetent cells and grown on LB containing carbenicillin (90ug/mL) overnight. The next day, eight colonies were picked and grown in LB-carbenicillin broth for 12 hours. The cells were pelleted and the plasmids extracted using the QiaPrep Spin MiniPrep (Qiagen) protocol. The

resulting DNA was sequenced using T7 primers and six of the eight clones contained the L64F amino acid substitution. This protein was grown the same way as the other two constructs (above).

Table 2.1. DNA primers for site directed mutagenesis of mGFPsol-Y66L to implement the L64F mutation. Mismatched nucleotides in each direction are colored red.

Primer name	**Sequence**	**Tm (^{o}C)**
L64F Trix For	5'- CCA ACA CTT GTC ACT ACT TTT ACT CTG GGT GTT CAA TGC	63.6
L64F Trix Rev	5'- GCA TTG AAC ACC CAG AGT AAA AGT GAC AAG TGT TGG	63.6

Protein Isolation. For all protein constructs, the cell pellet was resuspended and subsequently lysed in 50mL of buffer containing 50 mM HEPES, pH 7.9, 300 mM NaCl, 3 mM BME, 10% glycerol, 0.1 mM PMSF, and DNase. The sonicator was used at 60% power with output set to 6, for a total of 5 minutes cycling one minute on, one minute off with the cell suspension on ice throughout. The lysate was centrifuged for 45 min at 14,000 rpm and the clarified supernatant was applied to a Ni-NTA SuperFlow (Qiagen) column and run by gravity. The resin was washed with column wash buffer (50mM HEPES, pH 7.9, 300 mM NaCl) until the absorbance of the flowthrough returned to baseline values. Then, low imidazole buffer (50 mM HEPES, pH 7.9, 300 mM NaCl, 20 mM imidazole) was added to the column to wash off any nonspecifically bound proteins. Finally, the His-tagged protein was eluted using elution buffer (50mM HEPES, pH 7.9, 300 mM NaCl, 20 0mM imidazole) and the pooled protein was dialyzed against 50 mM HEPES, pH 7.9 with 1 mM EDTA overnight.

Multiple dialysis protocols were attempted in an effort to prevent the protein from aggregating. Attempts ranged from 300 mM NaCl in 50 mM HEPES, pH 7.9 overnight, to slower, stepwise buffer changes, designed to reduce the osmotic shock of reducing the salt concentration.

The EGFP-Y66L construct was purified in the same way as the mGFPsol-Y66L and mGFPsol-L64F/Y66L proteins with two additional steps. The dialyzed 6xHis tagged protein was cleaved from its affinity tag via the addition of α-chymotrypsin (1 mg α-chymotrypsin per 50 mg protein) and incubated overnight at 4^{o}C on the rocker. The next day, the protein was bound to a DEAE anion exchange column which had been equilibrated in 50 mM HEPES, pH 7.9. The column was washed with DEAE wash buffer (50 mM HEPES, pH 7.9, 50m M NaCl) and the pure protein eluted with DEAE elution buffer (50 mM HEPES, 300 mM NaCl). Protein was assayed for purity by SDS-PAGE. The DEAE elution was concentrated in Millipore Amicon YM-10 centrifugal filter units to about 1mL and frozen at -80^{o}C.

Isotopic Enrichment. To obtain uniformly-^{15}N labeled protein, minimal media stock solutions were prepared as follows: A 50% glucose stock (w/v), 20 mg/mL thiamine, and 1.0 M $MgCl_2$ were all prepared in water. The trace elements stock solution was made by combining 85mL of water with 0.60 g of $CaCl_2 \cdot 2H_2O$, 0.60 g of $FeSO_4 \cdot 7H_2O$, 0.115 g $MnCl_2 \cdot 4H_2O$, 0.08 g $CoCl_2 \cdot 6H_2O$, 0.07 g $ZnSO_4 \cdot 7H_2O$, 0.03 g $CuCl_2 \cdot 2H_2O$, 0.002 g H_3BO_3, 0.025 g $(NH_4)_6Mo_7O_{24} \cdot 4H_2O$, and 0.50 g EDTA-$Na_2$. The volume was subsequently adjusted to 100 mL, and filter sterilized to 0.22 μm. A 5x minimal salts solution

was prepared by adding 3 g KH_2PO_4, 11.32 g $Na_2HPO_4 \cdot 7H_2O$, 0.5 g NaCl and 1.0g $15NH_4Cl$ to 150 mL of water, adjusting the pH to 7.2, then adding water to a total volume of 200 mL (For protein preps greater than 1L, a corresponding volume of 5x minimal salts were prepared.) Each liter of minimal medium was assembled via the addition of 200 mL of 5x minimal salts, 8 mL of 50% glucose, 2 mL of thiamine, 1 mLof $MgCl_2$, 200 mg of carbenicillin, and 5 mL of trace elements stock then the total volume was adjusted to 1L with sterile water. The complete ^{15}N-enriched minimal medium was inoculated with 5mL overnight culture (LB/Carb at 37^{o}C, 300 rpm, 15-18h) and grown 1L per flask in 2.8L fernbach flasks. When optical cell density reached 0.6, the temperature was decreased to 23^{o}C and protein production was initiated by addition of 1 mM IPTG and incubation proceeded for 18 hours.

Expression Cell Line Tests for Maximum Protein Manufacture. As an alternative to BL21(De3) cell line expression, Rosetta(BL21)De3-pLysS (Novagen) cells were used to compare protein expression and overall yield. All protein expression protocols were the same as with BL21(De3) except that for each prep, freshly made chemically competent cells were transformed with the pRSetB plasmid containing the EGFP-Y66L insert the day before overnight cultures were started. JM109(De3) (Stratagene) was also used, although for only one preparation.

Algae-Derived Isotopic Enrichment Supplementation. Minimal media was supplemented with commercially available ISOGRO®-^{15}N, an algal lysate-derived complex labeling medium (Sigma). Competent *E.*

coli Rosetta(DE3)pLysS cells were freshly transformed with the pRsetB plasmid containing the gene encoding for EGFP-Y66L controlled by the T7 promoter, spread aseptically on LB plates containing carbenicillin and chloramphenicol, and grown overnight at 37°C. A single colony was transferred to 20 mL of LB/Carb/Cam, and grown at 37°C with shaking, overnight (16-18 hours). 1.5L of minimal media was prepared as above, and divided into two flasks of 0.75L each. 1.0 g of ISOGRO® powdered growth medium was added to one of the flasks and both flasks were inoculated with 7.5 mL of overnight culture. The cells were grown while monitoring absorbance at 600nm (OD_{600}) and grown at 37 °C with shaking. At an OD_{600} ~ 0.7-0.9, the protein expression was induced by addition of isopropyl β-D-1-thiogalactopyranoside (IPTG) to a final concentration of 0.1 mM. Protein production was analyzed by SDS-PAGE.

Circular Dichroism Measurement. Unlabeled EGFP-Y66L protein (concentration = 25.5 mg/mL) was added to 13 different buffers (HEPES: 8.0, 8.5; CHES: 8.5, 9.0, 9.5, 10.0; CAPS: 10.0, 10.5, 11.0, 11.5; $NaOH/PO_4$: 11.0, 11.5, 12.0) and allowed to equilibrate overnight. Final concentration of buffer was 25 mM containing protein at 1.275 mg/mL. Absorbance scans were run from 650 to 220nm to ensure that the protein samples were at equivalent concentrations, and then CD was run from 250 to 200nm at room temperature for each protein and buffer blank. Five scans were averaged and recorded for each sample. CD data is shown in Figure 2.3 (JW01-077).

NMR Measurement. First, uniformly ^{15}N-labeled EGFP-Y66L protein aliquots were thawed and pooled. This bulk protein solution was buffer

exchanged into 50 mM HEPES with 100 mM NaCl for the purposes of NMR. The protein was concentrated in an Amicon YM-3 centrifugal concentrator (Millipore) until it reached a volume of 800 µL. 100 µL of D_2O was added, and 300 µL of the uniformly ^{15}N-labeled protein at a concentration of 14.88 mg/mL (550 uM) was added to a D_2O matched Shigemi Advanced NMR Microtube (15 mm bottom length, Sigma). The sample was placed into a Varian 800MHz NMR, and a ^{1}H-^{15}N HSQC TROSY experiment was run at 310K for 1hr to look at peak dispersity and sample quality. The remaining concentrated protein was frozen at -80^{o}C for potential future experiments. The spectrum that was collected is shown in (Figure 2.4). A longer, 3D experiment was also attempted: 4 days under the same conditions.

Thermal Stability Test. Twelve buffers were prepared, ranging in pH from 5.5 to 11 (0.1M, increments of 0.5 pH units) each containing 0.1M NaCl (MES: 5.5, 6.0, 6.5; HEPES: 7.0, 7.5, 8.0; CHES: 9.0, 9.5, 10.0; CAPS: 10.0, 10.5, 11.0). A working stock of SYPRO orange dye was made by diluting the frozen 5000x stock (Molecular Probes) stored in DMSO to 300x in water. 12 master mix solutions were made by combining 8µL of 2.705mg/mL unlabeled mGFPsol-L64F/Y66L protein with 67.7µL of buffer and 4.3µL of the 300x SYPRO dye. The mixes were aliquoted to columns 1-3 in a 384-well plate so parallel measurements would be taken in triplicate, and the plate was covered with transparent film. It was placed in a RT-PCR instrument (Applied Biosystems, ASU DNA core facility) and the plate was allowed to equilibrate to 4^{o}C for 2 minutes. The temperature was ramped up to 95^{o}C at about three degrees

per minute, held at that temperature for 2 minutes, then returned to 4°C and equilibrated for another two minutes. Fluorescence measurements were taken throughout the course of the experiment and plotted vs. temperature. The midpoint of melting was assumed to be the temperature at which the fluorescence was halfway between the baseline and maximum values. The normalized data is shown in Figure 2.5, A.

RESULTS and DISCUSSION

The mGFPsol-Y66L protein proved difficult to purify in that it would precipitate soon after dialysis (Table 2.2). This phenomenon was assumed a result of the Y66L substitution, since mGFPsol is very soluble, itself. The substitution of the phenol ring with a smaller aliphatic group likely destabilized the protein by leaving a void in the center of the β-barrel.

Table 2.2. Protein preparations of mGFPsol-Y66L

mGFPsol-Y66L						
Date	**Notebook Reference**	**Media**	**Vol.**	**Cell Line**	**Yield**	**Storage/Use**
01/05/09	JW01-004	LB	2L	BL21(De3)	8.388g cells	Protein ppt
02/02/09	JW01-009	LB	2L	BL21(De3)	7.865g cells	Protein ppt
06/08/10	JW02-102	LB	2L	BL21(De3)	8.045g cells	Protein ppt

To try to circumvent the solubility problem experienced with mGFPsol-Y66L, the EGFP-Y66L construct was used since it had been crystallized previously (*14*) and its solution stability was known. The bulk of the work on this project was done using this construct. Concentrations of up to 25mg/mL were

obtained, and the protein was used for various NMR, CD, and UV/Vis experiments (Table 2.3, Figure 2.3, and Figure 2.4).

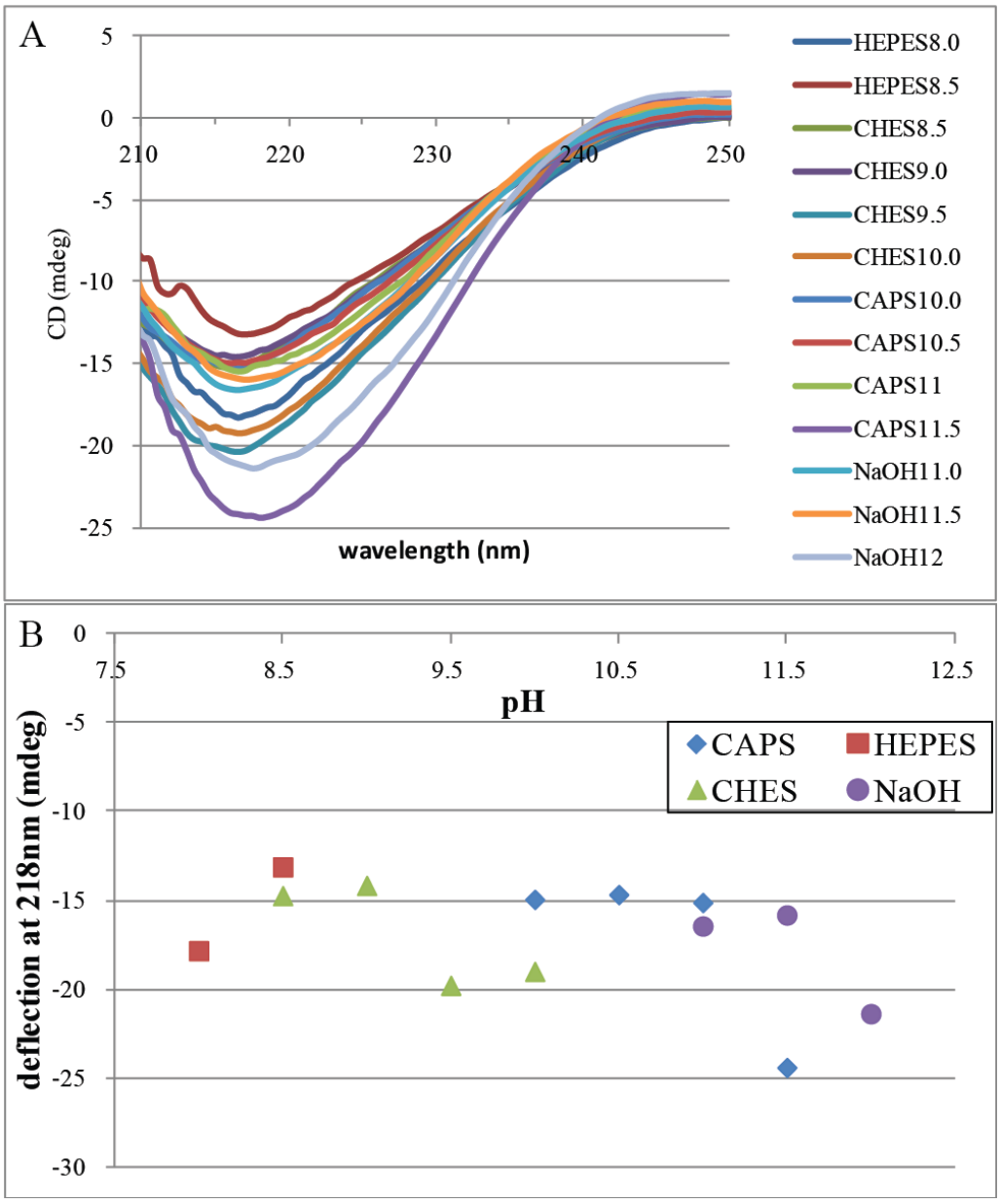


Figure 2.3. Circular Dichroism of EGFP-Y66L. A) The protein exhibits a very distinct negative peak at 218nm which is indicative of the β-character expected of GFP. B) There does not appear to be any correlation between pH of the solution and the degree of negativity of the peak.

Table 2.3. Protein preparations of EGFP-Y66L in BL21(De3) cells

EGFP-Y66L						
Date	**Notebook**	**Media**	**Vol.**	**Cell Line**	**Yield**	**Storage/Use**
02/21/09	JW01-015	LB	2x2L	BL21(De3)	24.334g cells/ 8.4 mg protein	Unlabeled, not used
03/20/09	JW01-023	LB	2L	BL21(De3)	14.492g cells/ 10.4 mg protein	4^{o}C
04/16/09	JW01-028	^{15}N-MM	1L	BL21(De3)	2.7g cells/ 4.76mg protein	NMR (JW01-033)
05/15/09	JW01-036	LB	4x1L	BL21(De3)	7.26g cells/ 7.4mg protein	UV/Vis titration/CD
05/20/09	JW01-040	LB	4x1L	BL21(De3)	10.2 g cells/ 5.73 mg protein	UV/Vis titration/CD
06/10/09	JW01-053	LB	4x1L	JM109 (De3)	20.37g cells/ 45mg protein	UV/Vis titration/CD
07/06/09	JW01-062	^{15}N-MM	4x1L	BL21(De3)	40.3g cells/ 2.4mg protein	Freeze
08/03/09	JW01-078	LB	4x1L	BL21(De3)	27.95g cells/ no protein	N/A
08/17/09	JW01-089	^{15}N-MM	2x1L	BL21(De3)	11.52g cells/ 10mg protein	NMR (JW01-092)
08/19/09	JW01-096	15N-MM	3L	BL21(De3)	12.7g cells/ 8.3mg protein	Freeze -80^{o}C
09/13/09	JW01-114	15N-MM	2L	BL21(De3)	7.35g cells/ low protein yield	Pool with below
10/06/09	JW02-006	15N-MM	2L	BL21(De3)	6.76g cells/ lost protein during conc	N/A
10/19/09	JW02-016	15N-MM	2L	BL21(De3)	7.48g cells/ no protein	N/A

Despite little success with mGFPsol-Y66L, EGFP-Y66L seemed to be a much better behaved protein. In the CD experiment, all measurements exhibited β-character by showing a negative peak at 218nm (Figure 2.3). Although data was collected to 200nm, the signal became unreliably noisy below 210nm so the data was shown in Figure 2.3A is truncated to 210nm. A plot of pH versus ellipticity at 218nm does not show any distinct trends, even within specific buffer groups. This is probably indicative of aggregation, or that the protein denatures while retaining its β-character. The protein concentration is low enough in the

CD experiments that aggregation would not necessarily extend beyond the formation of microparticles, so it is a distinct possibility even though no visible precipitate was observed.

Table 2.4. Protein preparations of EGFP-Y66L in Rosetta(De3)pLysS cells

EGFP-Y66L						
11/17/09	JW02-027	15N-MM	4L	Rosetta (De3)pLyS	No cell growth	N/A
11/30/09	JW02-038	15N-MM	10mL	Rosetta (De3)pLyS	No cell growth	N/A
12/07/09	JW02-040	15N-MM	5mL	Rosetta (De3)pLyS	Cell growth obtained	N/A
12/21/09	JW02-044	15N-MM	1L	Rosetta (De3)pLyS	3.14g cells/ 11.34 mg protein	NMR (JW02-086)
01/10/10	JW02-052	15N-MM	2L	Rosetta (De3)pLyS	7.2g cells/ 5mg protein	NMR (JW02-086)
02/16/10	JW02-069	15N-MM	2L	Rosetta (De3)pLyS	4.059g cells/ 2mg protein	NMR (JW02-086)
03/17/10	JW02-076	15N-MM	2L	Rosetta (De3)pLyS	4g cells/ 4mg protein	NMR (JW02-086)
04/20/10	JW02-090	15N-MM	2x0.75L	Rosetta (De3)pLyS	1.024g cells (-A)/2.2g cells (+A)/ 5mg protein	Freeze -80^{o}C
05/05/10	JW02-096	15N-MM	2L	Rosetta (De3)pLyS	4.393g cells/ no protein yield	N/A

The CD results did not explicitly indicate protein unfolding although it is assumed that it was occurring. There was no indication of the kind of protein

aggregation or solution properties that would preclude further experiments with the EGFP-Y66L construct. An hour-long NMR test run yielded the HSQC spectrum in Figure 2.5. This 2D NMR experiment on the EGFP-Y66L construct showed dispersed and well-formed peaks. Since the experiment was only run in two dimensions, it did not contain enough information for peak assignments to be made *ab initio* however, comparison with published spectra of the GFPuv construct does elucidate several similar trends in peak distribution and relative shifts (*36*, *37*) For example, the region where the glycine backbone amides reside is reminiscent of the GFPuv data previously published (*36*, *37*). The general dispersion of peaks is similar to the GFPuv NMR data, but there are some significant differences. First, all 238 peaks are not present. It is likely that the peak overlap in the unresolved regions between 8-9ppm in the ^{1}H dimension and 120-128ppm in the ^{15}N dimension contain more peaks than are clearly shown.

Unresolved regions in this spectrum remain, and can be attributed to the dimeric nature of EGFP-Y66L construct. An attempt was made to collect data on the same sample over a two day time frame (NOESY) to resolve the overlapping peaks. Upon examination, the data was unusable due to an extremely low signal-to-noise ratio. The protein had aggregated during the extended period of time during which it had been subjected to a temperature of 310K.

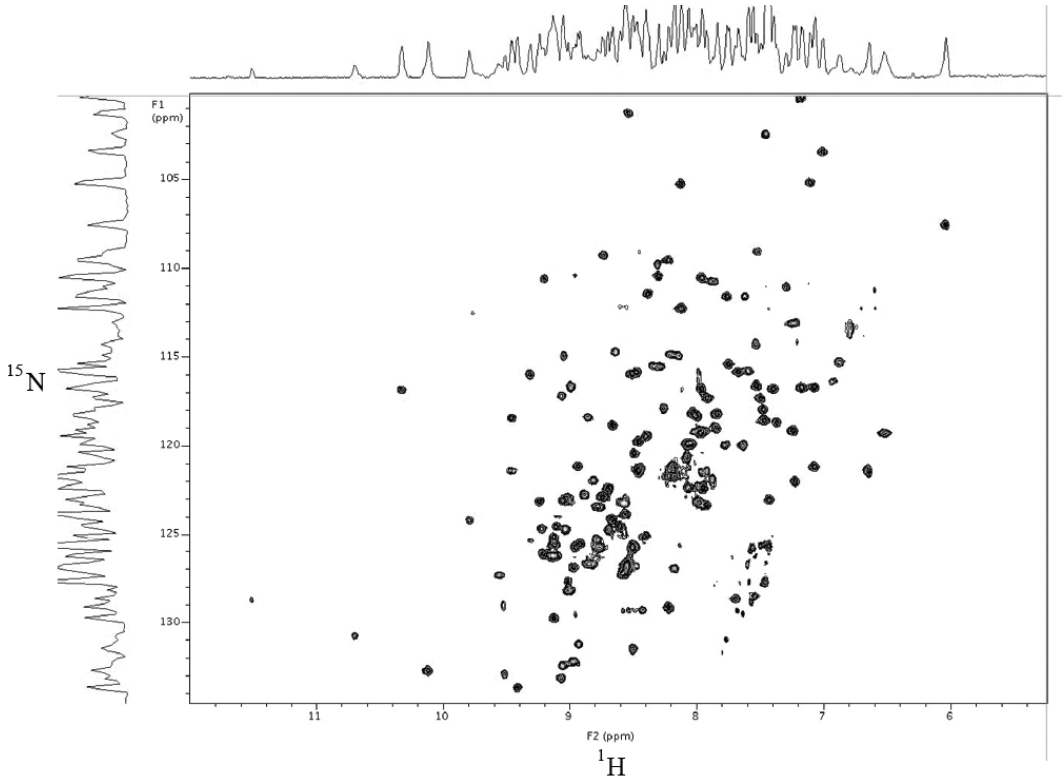


Figure 2.4. HSQC TROSY NMR spectrum of EGFP-Y66L at 310K for 1 hr in the Varian 800MHz NMR (JW02-087)

Molecules with large radii have slower tumbling speeds in NMR experiments. Since signals are an average of all the observed magnetic interactions within the protein, slow tumbling rates will cause peak broadening and therefore overlap. The resolution can be increased by increasing the number of data collection intervals and increasing the temperature to amplify the kinetic energy of the molecules, but this protein aggregated during the duration of the experiment.

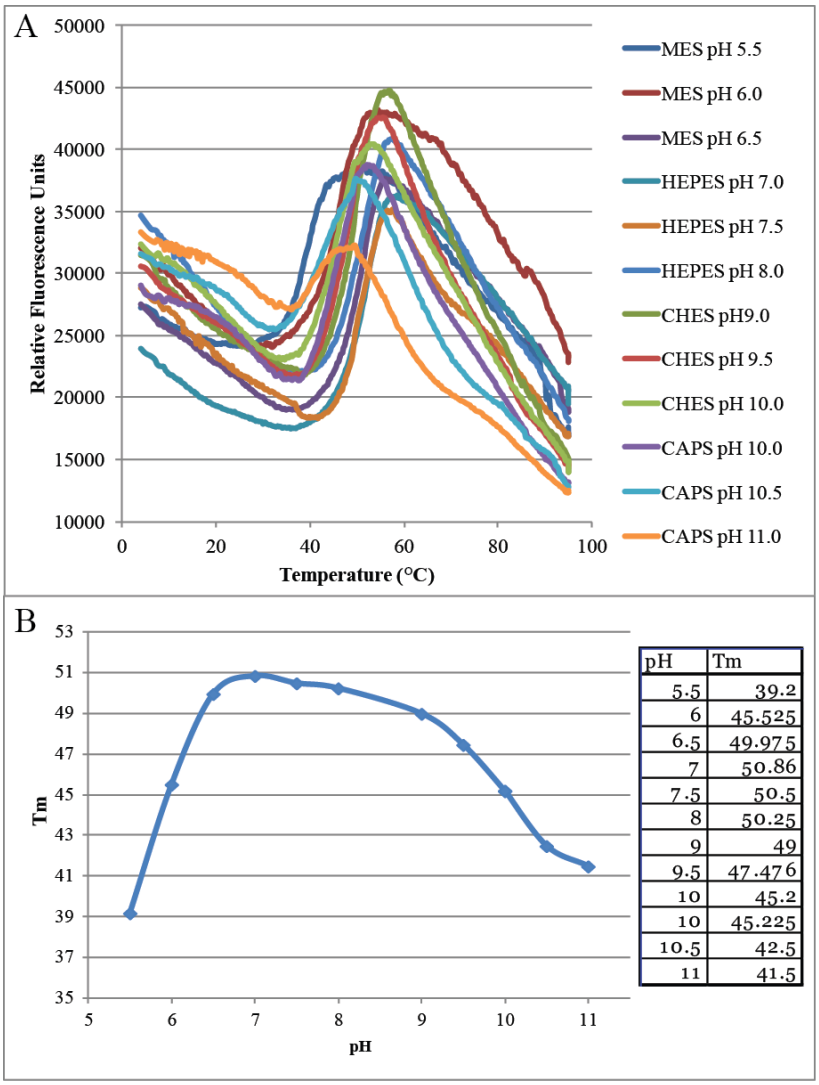
In an attempt to remedy this problem, the monomeric protein (mGFPsol-Y66L) was again examined because the smaller protein would theoretically require less time in NMR for reasonable data acquisition. The assumption for the previously observed instability was that the Y66L mutation, in the absence of the

stabilization provided by the dimer interface, left too much empty space in the center of GFP's beta-barrel. To recover some of the structural stability lost with the Y66L mutation, the wild-type reversion mutation L64F was introduced into mGFPsol-Y66L. The new double mutant of mGFPsol was, in fact, more stable; remaining in solution throughout the purification process. However, its stability characteristics were as yet unknown, specifically whether it could withstand the temperatures required for NMR.

The thermal stability was assessed using a fluorescence-based thermal shift assay with the thermofluor dye, SYPRO orange. This molecule is hydrophobic and only fluoresces when it is bound in a likewise hydrophobic environment (*38*). As temperature increases, the protein begins to melt, exposing its buried hydrophobic regions to the solvent. The SYPRO orange molecule binds to these regions and via hydrophobic interaction, becomes fluorescent. The fluorescence due to protein denaturation is directly correlated with the temperature and the midpoint of protein melting (T_m) can be estimated. The raw data shown in Figure 2.5A is evaluated as such and the temperature at which the fluorescence is midway between the minimum and maximum values can be assigned as the apparent T_m.

This method is very effective for use in assaying protein stability under different buffer or salt conditions, ligand binding, and small molecule interactions. This assay was performed in triplicate with the protein in twelve separate conditions sampling eleven unique pH values. A plot of pH vs. T_m yielded a bell-shaped curve Figure 2.5B) showing a maximum thermal stability

for mGFPsol-L64F/Y66L at pH 7 (T_m= 50.86^{o}C). These data show that the monomeric double mutant should be thermally stable for the planned NMR experiments at 310K. The purification procedure places the protein in HEPES buffer with a pH near 8. Even in those conditions, the T_m is greater than 50^{o}C.



pH	Tm
5.5	39.2
6	45.525
6.5	49.975
7	50.86
7.5	50.5
8	50.25
9	49
9.5	47.476
10	45.2
10	45.225
10.5	42.5
11	41.5

Figure 2.5. Thermal stability of mGFPsol-L64F/Y66L. A) Representative fluorescence intensity curves as related to temperature in the RT-PCR heat block. B) calculated melting temperatures of each representative protein solution versus pH.

Table 2.5: Protein preparations of the reversion mutant, mGFPsol-L64F/Y66L

mGFPsol-L64F/Y66L						
Date	**Note book**	**Media**	**Vol.**	**Cell Line**	**Yield**	**Storage/Use**
08/06/10	JW04 -002	LB	1L	BL21 (De3)	5.903g cells/ very low protein yield	N/A
08/20/10	JW04 -006	LB	2x2L	BL21 (De3)	Dirty protein	N/A
09/02/10	JW04 -010	^{15}N-MM	2x1L	BL21 (De3)	Low yield of somewhat pure protein	Freeze at -80^{o}C
10/20/10	JW04 -016	LB	4x1L	BL21 (De3)	Dirty protein. No overexpressed band	Pool with 12/29/10 Ni Wash and try to get purer
12/22/10	JW04 -022	^{15}N-MM	2x1L	BL21 (De3)	6.273g cells/ could not get protein pure	N/A
01/12/11	JW04 -030	^{15}N-MM	2x1L	BL21 (De3)	7.194g cells/ no protein visible on gel	N/A
03/10/11	JW04 -044	^{15}N-MM	2x1L	BL21 (De3)	6.64 g cells/ not enough pure protein to use	Freeze at -80^{o}C
05/12/11	JW04 -056	LB	2x1L	BL21 (De3)	0.27mg concentrated protein	Thermofluor assay
06/21/11	JW04 -063	^{15}N-MM	4x1L	BL21 (De3)	12.0g cells/ 10mg protein	Freeze at -80^{o}C

When preparing the mGFPsol-L64F/Y66L protein for NMR experiments (Table 2.4), the concentration of the protein when ready for freezing was low relative to what had been observed with the EGFP variant. It was unclear whether this was due to low protein yield during growth and expression or due to loss of protein during concentration. Unlabeled protein was purified and concentrated while regular UV absorbance scans were recorded to see what the maximum

achievable concentration would be. This exercise showed that the highest concentration I could reach was 4.7 mg/mL. This concentration is about 1/3 of that required for protein NMR experiments.

The protein could potentially be engineered further in an attempt to remedy the solubility issues. The protein is thermally stable as is, but higher concentrations are needed for the planned experiments. The protein constructs used in the preceding experiments have different stability parameters due to the different mutations possessed in each. EGFP-Y66L exhibits weak dimerization, which although good for stability, is bad for NMR experiments since the dimer is a larger unit. mGFPsol-L64F/Y66L is more stable in solution than mGFPsol-Y66L, but still exhibits some stability issues resulting in protein loss arising during concentration. For future work, another mutation could be made near the chromophore-binding pocket which may restore stability. A T203Y mutation could potentially help fill the gap that has been created by the Y66L substitution. These colorless, intermediate protein constructs may yet prove useful in determining the pK_a of the buried arginine inside of GFP.

Chapter 3

FLUORESCENT PROTEINS IN FRET

INTRODUCTION:

The aggressive expansion of the fluorescent protein (FP) repertoire has revolutionized live-cell imaging techniques (*23*, *24*, *27*, *32*). The multitude of available colors makes it possible to visualize dynamic protein activity on a subcellular level in real time. One of the most informative and therefore extensively used of these techniques is Förster resonance energy transfer (FRET), which measures the nonradiative transfer of excited-state energy between two adjacent molecules, a donor and an acceptor. This method is predicated upon the probability of radiationless energy transfer when the two probes are situated within a specific distance of one another. Only the donor molecule must be fluorescent, although both often are (*24*). For resonance transfer to occur, the two proteins must have overlapping emission and absorption spectra for the donor and acceptor respectively, where the coincident wavelengths of the two energetic states represent the optimal region of energy transfer. Consideration of this basic tenet of FRET is integral when choosing partner probes.

The most common FP pair in FRET experiments usually involves variants of ECFP and yellow FP (YFP)(*23*), both engineered descendents of green fluorescent protein (GFP) carrying tryptophan-derived chromophores. ECFP emits its maximum light at 475nm while YFP absorbs light in a similar region (λ_{max} =515nm). Although the two proteins fulfill the requirement for spectral overlap, other factors complicate the FRET experiments. ECFP has very dim

fluorescence compared to EGFP and this reduces the maximum signal to noise measurement that can be obtained. Additionally, ECFP's fluorescence lifetime is multi-exponential which indicates the presence of multiple excited states and which reduces FRET success and is therefore an undesired trait(*39*). In 2003, the ECFP crystal structure was solved, exposing two competing protein conformations(*40*). The majority of the structural heterogeneity centered on the His148 and Tyr145 side chains which moved rather substantially. In one structure, Tyr145 pointed into the beta-barrel while His148 was solvent exposed. In the other, the positions were reversed, with His148 inside and Tyr145 on the outside of the barrel(*40*).

An effort to improve ECFP for more effective use in FRET began, utilizing several methods including rational design and random mutagenesis. First, the H148D mutation was introduced to stabilize the protein in a single conformation(*32*). The rationale behind this mutation was that a charged side chain would most likely reside in the bulk solvent, and be disinclined to intrude upon the chromophore binding pocket. H148D did successfully simplify the fluorescence decay curve(*32, 41*), implying the stabilization of a structure that would shift the relaxation equilibrium to preferentially favor fluorescence emission over competing nonradiative mechanisms. However, the fluorescence kinetics remained complex, again indicating several fluorescent excited states (*41*). Rizzo et al. (*32*) set about improving the fluorescent properties further by incorporation of the GFP folding-enhancement mutation, S72A(*42*). This mutation would theoretically minimize the impact of environmental temperature

on the fluorescence emission by helping the protein mature more quickly. Additionally, a Y145A mutation was discovered through random mutagenesis experiments in which this substitution improved the quantum yield(*32*). The triple mutant, Cerulean, was 2.5 times as bright as ECFP and its fluorescence decay was best characterized by a single exponential fit(*32*), mostly due to the single amino acid substitution H148D(*41*).

Cerulean, although brighter than its cyan predecessor, does not emit very intense light compared to many FPs. Additionally, it exhibits reversible photobleaching when exposed to excitation light. These are common problems among blue FPs, which through *in vitro* evolution and design, are being improved with every successive generation. Data collection relies heavily on donor fluorescence so FRET experiments are heavily impacted by dim fluorescence and photoswitching(*24*). The donor is monitored because cross-talk between the probe pair and inadvertent direct excitation of the acceptor during donor excitation can muddle the results (*23*). When the donor fluorophore, in this case Cerulean, exhibits unstable fluorescence and photoswitching the signal-to-noise ratio decreases, leading to decreased resolution and quality of data. To maximize the FRET signal, the donor's quantum yield needs to be as high as possible while the acceptor's absorbance is likewise maximal.

One of the ways to improve quantum yield is to examine the structure of the fluorophore with an eye toward optimization. The fluorescent characteristics of pi-conjugated ring systems are heavily affected by the co-planarity of adjacent rings (*20*). The rigidity and planarity of the molecule is directly related to the

quantum yield (*43*), so any torque on the ring system can negatively impact fluorescence. GFP's chromophore is a *para*-hydroxybenzylidene-imidazolinone formed autocatalytically from a Serine-Tyrosine-Glycine tripeptide buried in the center of the protein. The chromophore in wild-type GFP experiences hula twisting phenomena which permit fluorescence quenching(*20*). Since Cerulean is a descendant of GFP, its chromophore pocket is optimized for a tyrosine-based chromophore, while the chromophore of Cerulean, (Threonine-Tryptophan-Glycine) is bulkier than GFP's and may be crowded within the protein. This is further influenced by structural contributions from nearby β-strands. The amino acid side chains of the β-strands 7 and 8 can point toward the interior of the barrel affecting one of two outcomes: stabilizing the chromophore packing within the barrel or contributing unfavorable steric interactions to the chromophore pocket, increasing torsional strain on the ring system. The tryptophan derivative introduces a large amount of steric interaction with the substitution of the indole moiety for the phenol ring as it is, so the additional crowding and Van der Waals interactions from the surrounding protein could force twisting between the rings, negatively impacting the fluorescence yield. Steric pressure may be alleviated by the introduction of structural modifications to improve the complementarity of the chromophore and the pocket which would encourage radiative excited state relaxation.

Crystallization of Cerulean revealed the existence of two different chromophore conformations dependent upon the pH of the drop. At low pH, the chromophore of Cerulean existed in an unexpected "cis" conformation (Figure

3.1A, purple) where the indole group was rotated around the C_β-C_γ bond of Trp66, positioning it unexpectedly close to the imidazolinone ring(*44*). This conformation explained the observations made by Malo et al. whereby the fluorescence properties changed coincident with change in pH. In solution at pH 5.0, the fluorescence was observed to decrease four-fold relative to that at pH 8.0.

Later, a structure obtained at a pH of 8.0 presented a chromophore structure that very closely matched that of ECFP (Figure 3.1, grey)—the expected conformation of Cerulean (*29*). These two chromophore conformations corresponded to structural changes in β-strand 7, similar to those observed in ECFP; namely, movement of the Asp148 side chain from inside the barrel (low pH, *cis*-chromophore) (*44*) to outside the barrel (neutral pH, *trans*-chromophore) (*29*). The H148D mutation in Cerulean had been expected to improve the cyan construct (*41*) . However, even when coupled with the S72A and Y145A mutations to form Cerulean, the protein exhibited a strong environmental sensitivity, as the structural data clearly revealed distinct pH-dependent isomerization. The conformational change in the chromophore led to a less radiatively efficient *cis*- conformation.

Efforts to consolidate the isomers into one stable conformation for the improvement of Cerulean's fluorescence properties and quantum yield ultimately resulted in the recently published mCerulean3(*1*). Random, pair-wise mutations along the seventh and eighth strands of the β-barrel of Cerulean were introduced in an attempt to alleviate steric stress on the chromophore (S147H/ D148G/ K166G/ I167L/ R168N/ H169C). Redesign of the chromophore binding pocket

replaced the bulky and mobile residues in the β-strands near the indole ring and produced a mutant with 25% higher quantum yield that was 1.3-times as bright as Cerulean(*1*).

The six mutations, although improving the optical statistics, seemed to decrease the photostability of the construct as evidenced by the substantial increase in photoswitching behavior (*1*). To relieve some of the conformational strain on the planarity of the chromophore by improving the fit in the pocket, the T65S wild-type reversion mutation was introduced. Early Cyan variants had Ser65 in the chromophore and were substantially brighter than the EGFP-derived ECFPs (*5*), though their utilization was somewhat obscured by the ease and popularity of the EGFP descendents. In GFP-S65T, Thr65 stabilizes the anionic form of the chromophore, improving fluorescence by breaking a hydrogen bond to Glu222 (*3*, *4*, *12*). In the crowded interior of the Cerulean barrel, the removal of a methyl group on the chromophore may have allowed the structure to more adequately fit into the central cavity of the β-barrel. This small change, a methyl group deletion, dramatically improved the brightness, quantum yield and, at the same time, almost entirely eliminated the photoswitching behavior in mCerulean3 (*1*). The fluorescence properties of mCerulean3 far surpassed those of Cerulean; this variant is monomeric, brighter than previous CFPs, and possesses a vastly improved quantum yield; comparable to that of EGFP.

MATERIALS AND METHODS:

Crystallization. The mCerulean3 protein was expressed and purified as previously described by (*1*). The protein stock contained 20mM HEPES, pH 7.9 with 20mM NaCl and about 12mg protein per mL buffer. Crystallization screens were assembled using the hanging-drop vapor-diffusion method. Diffraction quality crystals were obtained from wells with mother liquor consisting of 0.08 M sodium acetate trihydrate at pH 4.7, 0.16 M ammonium sulfate, 9% (w/v) PEG4000, and 19% (v/v) glycerol. Hanging drops contained 2µL protein solution and 1µL mother liquor.

Data Collection and Processing. The mounted crystal was flash frozen in liquid nitrogen before exposure to x-rays. Diffraction data was collected using the Advanced Light Source beam line 8.2.1 and an ADSC 315 detector (3072x3072 pixels) (Berkeley, CA). Data were indexed and scaled in HKL2000(*45*).

Structure Solution and Refinement. Spot intensities were converted into structure factors and 5% of the reflections were flagged for R_{free} calculations using DENZO and SCALEPACK (*45*). PHASER v2.3.0 (*46*) from the CCP4 program suite 6.2.0 (*47*) was used for molecular replacement with the structure of Cerulean Fluorescent Protein (PDB: 2wso,(*29*)) as the search model. This model was selected because of the high resolution of the solved structure and it is the protein model with the highest sequence identity to mCerulean3. Two protein chains were searched for within the asymmetric unit. Iterative rounds of refinement were performed using REFMAC5 v5.6.0117 (*48*) and Coot v0.6.2(*49*)

without the use of non-crystallographic symmetry (NCS) restraints. The Wilson B factor was calculated using CTRUNCATE v1.0.12 (*47*) Average B factors were calculated using the temperature factor analysis function in the structure analysis folder within the CCP4 program suite. The chromophore was modeled based on the CRF ligand in the 2wso Cerulean structure (*29*). It was modified to reflect the T65S mutation by removal of the CG1 atom and bonds therein. Figures were created using PyMOL. Data collection and refinement statistics are shown in Table 3.1.

Table 3.1: X-Ray Data collection and refinement statistics
Values in parentheses are for the highest resolution bin

Data collection	
space group:	$P2_12_12_1$
unit cell dimensions	a=79.41, b=88.73, c=94.74 α= β=γ=90
wavelength (Å):	0.97950
Resolution limits(Å):	50.0-1.62 (1.65-1.62)
Total reflections:	250772
Unique reflections:	84496
Redundancy:	3.0 (3.0)
Completeness (%):	98.7 (99.7)
{I/σ(I)}	14.5 (2.04)
R_{merge} (%)*:*	7.1 (47.7)
Wilson B Factor (Å):	18.4
Refinement	
Resolution range(Å):	29.983-1.620
# reflections:	79895
R_{cryst} (%):	17.77
R_{free} (%):	20.55
RMS deviation :bond lengths(Å)	0.0176
Bond angles (^{o})	2.3797
Average B factor, all protein atoms (Å):	20.27
Average B factor, all water atoms (Å):	29.917
Total atoms:	4213
Model residues:	829

RESULTS AND DISCUSSION:

The crystal structure of mCerulean3 has been solved to a resolution of 1.62Å. Two protein molecules are observed in the asymmetric unit and both are accompanied by various solvent molecules. The overall fold is the standard 11-stranded "β-can" typical of GFP-derived fluorescent proteins. The two monomers interface along strands 9, 10, and 11, while orienting themselves approximately ninety degrees offset, rotated along that interface. Overlay of the two protein molecules results in a RMSD of 0.169. The structure has been deposited in the Protein Data Bank as entry 4en1.

The crystal structure of mCerulean3 provides insight into the complex interactions between amino acids within a protein. The chromophore, although crystallized at low pH, exists in the same orientation as in the structure obtained at neutral pH regarding the relationship between the indole moiety and the imidazolinone ring(*29*). (FIGURE 3.1, grey) This conformational stability was enabled by the six mutations (S147H/D148G/K166G/I167L/R168N/H169C) made to β-strands 7 and 8. It is possible that the mutations belonging to mCerulean3 are able to both allow the chromophore to relax somewhat in the pocket and to prevent rotation around the methylene bridge. One of the most likely contributors to the enhanced stability is the mutation in mCerulean3 at position 148 from aspartate to glycine. A major difference observed in the two previous Cerulean structures was the mobility of the D148 side chain from inside the barrel near the chromophore to the exterior of the barrel in bulk solvent driven by pH increase (Figure 3.1A). Elimination of the entire side chain has removed

the pH-sensitive isomerization. In addition, the removal of the side chain from within the barrel allows more space for the chromophore, (Figure 3.1B) possibly reducing unfavorable steric interactions and quenching by hula twist. The I167L mutation pushes the leucine closer to the top of the chromophore, which should favor planarity over isomerization.

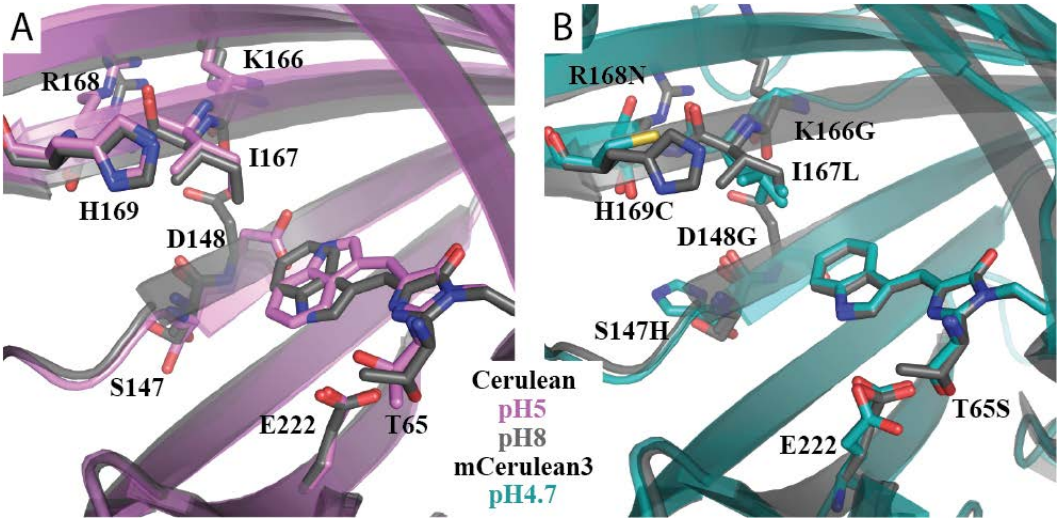


Figure 3.1. Chromophore Environments of Cerulean Fluorescence Proteins: A) Cerulean at physiological pH (PDB: 2wso, grey(*29*)) showing the Cyan-like trans-coplanar chromophore and Cerulean at pH 5.0 (PDB: 2q57, purple(*44*)) showing the cis-coplanar chromophore conformation. The Asp148 side chain transitions from outside to inside the barrel as the pH decreases. B) mCerulean3 (teal) overlaid with the high pH Cerulean structure (grey). mCerulean3 shows the Cyan-like chromophore conformation although it was crystallized at low pH. The substitutions of glycines for Asp148 and Lys166 in strands 7 and 8 have introduced flexibility into the region while generating more room for the chromophore within the barrel.

To evaluate the planarity of the chromophore in mCerulean3, the restraints in the library file were released and the model subjected to ten rounds of refinement. The resulting chromophore model demonstrated no significant deviation from the constrained chromophore. This indicates that the model of the

planar chromophore fits the density well (Figure 3.2) in the absence of planarity constraints during the refinement process.

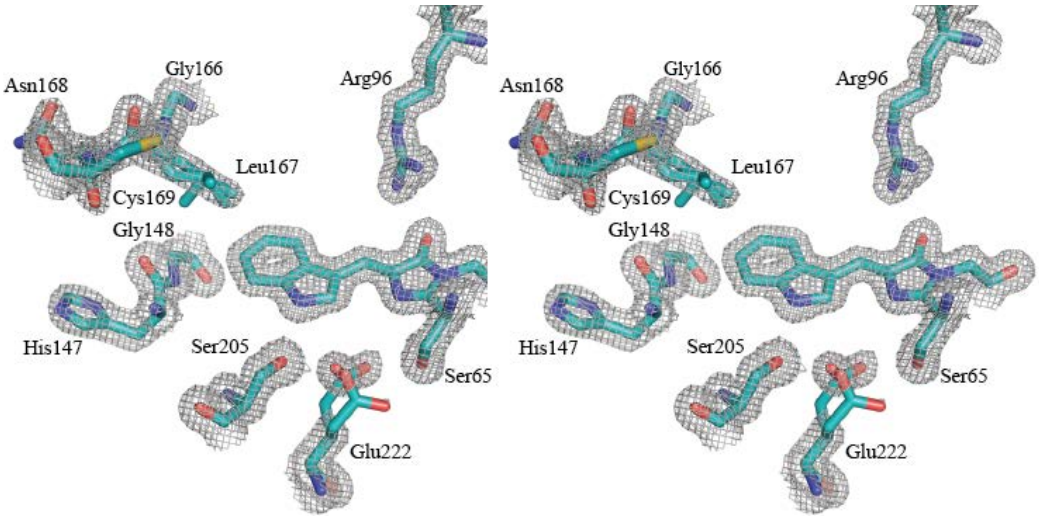


Figure 3.2. A stereo view of the mCerulean3 chromophore environment with map density modeled at 1.5 sigma and modified amino acid side chains shown as sticks.

While GFP structures have shown compelling evidence for the spectroscopic phenomena caused by the S65T mutation (*12*)**,** the Cerulean proteins don't seem constrained by the same chemical parameters. In GFP, the observation is that the S65T mutation favors chromophore ionization by causing the hydroxyl group to rotate about 90 degrees clockwise, away from Glu222, breaking the hydrogen bond and favoring the neutral carboxylic acid (*3*, *4*). This change reduces electrostatic competition in the chromophore binding pocket, encouraging the chromophore to exist in the anionic state. The previously published low pH structure of Cerulean shows the same Thr65 conformation as GFP-S65T(*44*), while the neutral structure (Figure 3.1grey) shows the hydroxyl pointing toward Glu222 as in wtGFP (*29*). The mCerulean3 structure, however, not only shows the hydrogen bond between Ser65 and Glu222, but also reveals a

second conformation of Glu222 which preserves the hydrogen bonds to Ser65 and Ser205 (Figure 3.3). The second glutamic acid conformation has a 40% occupancy and the functional group is rotated about 70 degrees around the β-carbon. The removal of the methyl group in the mutation T65S must alleviate some steric stress and therefore increase fluorescence. To compensate for the extra space, Glu222 appears to adopt a second, minor, conformation which is present in both monomers within the asymmetric unit.

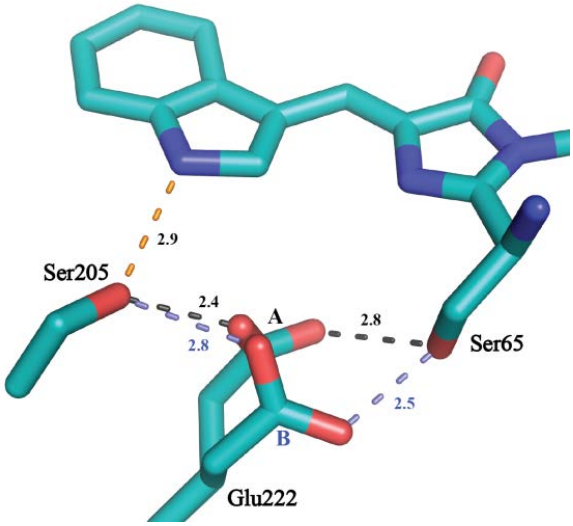


Figure3.3. Glu222 conformations in mCerulean3. Conformation A is the dominant one with 60% occupancy. Both conformations maintain hydrogen bonds to Ser65 and Ser205. Bond lengths between the side chains and the chromophore are noted in Å.

Despite the variety of changes to the amino acid sequence enhancing photostability, quantum yield, and overall brightness, mCerulean3 has nearly identical absorption and emission spectra compared to Cerulean. Its increased efficiency in fluorescence and photostability make mCerulean3 an optimal protein for use as a donor fluorophore in quantitative FRET experiments.

REFERENCES

1. Markwardt, M. L., Kremers, G.-jan, Kraft, C. A., Ray, K., Cranfill, P. J. C., Wilson, K. A., Day, R. N., Wachter, R. M., Davidson, M. W., and Rizzo, M. A. (2011) An Improved Cerulean Fluorescent Protein with Enhanced Brightness and Reduced Reversible Photoswitching, *PLoS ONE 6*.

2. Shimomura, O., Johnson, F. H., and Saiga, Y. (1962) Extraction, purification and properties of aeqourin, a bioluminescent protein from the luminous hydromedusan, Aequorea, *J. Cell. Comp. Physiol. 59*, 223-239.

3. Yang, F., Moss, L., and Phillips, Jr., G. N. (1996) The molecular structure of green fluorescent protein, *Nature 14*, 1246-1251.

4. Ormö, M., Cubitt, A. B., Kallio, K., Gross, L. A., and Tsien, R. Y. (1996) Crystal Structure of the Aequorea victoria Green Fluorescent Protein, *Science 273*, 1392-1395.

5. Heim, R., and Tsien, R. Y. (1996) Engineering green fluorescent protein for improved brightness, longer wavelengths and fluorescence resonance energy transfer, *Current Biology 6*, 178-182.

6. Cubitt, A., Heim, R., Adams, S. R., Aileen, E., Gross, L. A., and Tsien, R. Y. (1995) Understanding, improving, and using green fluorescent proteins, *Techniques in Molecular Biology 20*, 2967-2971.

7. Craggs, T. D. (2009) Green fluorescent protein: structure, folding and chromophore maturation, *Chemical Society reviews 38*, 2865-75.

8. Lemay, N. P., Morgan, A. L., Archer, E. J., Dickson, L. a, Megley, C. M., and Zimmer, M. (2008) The Role of the Tight-Turn, Broken Hydrogen Bonding, Glu222 and Arg96 in the Post-translational Green Fluorescent Protein Chromophore Formation., *Chemical physics 348*, 152-160.

9. Wachter, R. M. (2007) Chromogenic cross-link formation in green fluorescent protein., *Accounts of chemical research 40*, 120-7.

10. Barondeau, D. P., Putnam, C. D., Kassmann, C. J., Tainer, J. A., and Getzoff, E. D. (2003) Mechanism and energetics of green fluorescent protein chromophore synthesis revealed by trapped intermediate structures, *October 100*, 12111-12116.

11. Branchini, B. R. (1998) A Computational Analysis of the Unique Protein-Induced Tight Turn That Results in Posttranslational Chromophore Formation in Green Fluorescent Protein, *Journal of the American Chemical Society 120*, 1-6.

12. Brejc, K., Sixma, T. K., Kitts, P. a, Kain, S. R., Tsien, R. Y., Ormö, M., and Remington, S. J. (1997) Structural basis for dual excitation and photoisomerization of the Aequorea victoria green fluorescent protein., *Proceedings of the National Academy of Sciences 94*, 2306-11.

13. Sniegowski, J. a, Phail, M. E., and Wachter, R. M. (2005) Maturation efficiency, trypsin sensitivity, and optical properties of Arg96, Glu222, and Gly67 variants of green fluorescent protein, *Biochemical and biophysical research communications 332*, 657-63.

14. Rosenow, M. a, Huffman, H. a, Phail, M. E., and Wachter, R. M. (2004) The crystal structure of the Y66L variant of green fluorescent protein supports a cyclization-oxidation-dehydration mechanism for chromophore maturation, *Biochemistry 43*, 4464-72.

15. Wood, T. I., Barondeau, D. P., Hitomi, C., Kassmann, C. J., Tainer, J. A., and Getzoff, E. D. (2005) Defining the Role of Arginine 96 in Green Fluorescent Protein Fluorophore Biosynthesis †,‡, *Biochemistry* 16211-16220.

16. Zhang, L., Patel, H. N., Lappe, J. W., and Wachter, R. M. (2006) Reaction Progress of Chromophore Biogenesis in Green Fluorescent Protein, *JACS 128*, 4766-4772.

17. Wachter, R. M., Watkins, J. L., and Kim, H. (2010) Mechanistic diversity of red fluorescence acquisition by GFP-like proteins., *Biochemistry 49*, 7417-27.

18. Pouwels, L. J., Zhang, L., Chan, N. H., Dorrestein, P. C., and Wachter, R. M. (2008) Kinetic isotope effect studies on the de novo rate of chromophore formation in fast- and slow-maturing GFP variants., *Biochemistry 47*, 10111-22.

19. Rosenow, M. a, Patel, H. N., and Wachter, R. M. (2005) Oxidative chemistry in the GFP active site leads to covalent cross-linking of a modified leucine side chain with a histidine imidazole: implications for the mechanism of chromophore formation., *Biochemistry 44*, 8303-11.

20. Maddalo, S. L., and Zimmer, M. (2006) The role of the protein matrix in green fluorescent protein fluorescence., *Photochemistry and photobiology 82*, 367-72.

21. Miyawaki, A., Llopis, J., Heim, R., Mccaffery, J. M., Adams, J. A., Ikurak, M., and Tsien, R. Y. (1997) letters to nature Fluorescent indicators for Ca 2 + based on green fluorescent proteins and calmodulin, *Nature 388*, 882-887.

22. Tsien, R. Y. (1998) The green fluorescent protein., *Annual review of biochemistry 67*, 509-44.

23. Piston, D. W., and Kremers, G.-J. (2007) Fluorescent protein FRET: the good, the bad and the ugly., *Trends in biochemical sciences 32*, 407-14.

24. Vogel, S. S., Thaler, C., and Koushik, S. V. (2006) Fanciful FRET., *Science's STKE : signal transduction knowledge environment 2006*, re2.

25. Shaner, N. C., Patterson, G. H., and Davidson, M. W. (2007) Advances in fluorescent protein technology., *Journal of cell science 120*, 4247-60.

26. Pakhomov, A. a, and Martynov, V. I. (2008) GFP family: structural insights into spectral tuning., *Chemistry & biology 15*, 755-64.

27. Takanishi, C. L., Bykova, E. a, Cheng, W., and Zheng, J. (2006) GFP-based FRET analysis in live cells., *Brain research 1091*, 132-9.

28. Sarkar, P., Koushik, S. V., Vogel, S. S., Gryczynski, I., and Gryczynski, Z. (2011) Photophysical properties of Cerulean and Venus fluorescent proteins., *Journal of biomedical optics 14*, 034047.

29. Lelimousin, M., Noirclerc-Savoye, M., Lazareno-Saez, C., Paetzold, B., Le Vot, S., Chazal, R., Macheboeuf, P., Field, M. J., Bourgeois, D., and Royant, A. (2009) Intrinsic dynamics in ECFP and Cerulean control fluorescence quantum yield., *Biochemistry 48*, 10038-46.

30. Rekas, A., Alattia, J.-R., Nagai, T., Miyawaki, A., and Ikura, M. (2002) Crystal structure of venus, a yellow fluorescent protein with improved maturation and reduced environmental sensitivity, *The Journal of biological chemistry 277*, 50573-8.

31. Schultz, C. (2009) Fluorescent revelations., *Chemistry & biology 16*, 107-11.

32. Rizzo, M. a, Springer, G. H., Granada, B., and Piston, D. W. (2004) An improved cyan fluorescent protein variant useful for FRET., *Nature biotechnology 22*, 445-9.

33. Sniegowski, J. a, Lappe, J. W., Patel, H. N., Huffman, H. a, and Wachter, R. M. (2005) Base catalysis of chromophore formation in Arg96 and Glu222 variants of green fluorescent protein., *The Journal of biological chemistry 280*, 26248-55.

34. André, I., Linse, S., and Mulder, F. a a. (2007) Residue-specific pKa determination of lysine and arginine side chains by indirect ^{15}N and ^{13}C

NMR spectroscopy: application to apo calmodulin., *Journal of the American Chemical Society 129*, 15805-13.

35. Guillén Schlippe, Y. V., and Hedstrom, L. (2005) A twisted base? The role of arginine in enzyme-catalyzed proton abstractions., *Archives of biochemistry and biophysics 433*, 266-78.

36. Khan, F., Stott, K., and Jackson, S. (2003) Letter to the Editor: 1 H, 15 N and 13 C backbone assignment of the Green Fluorescent Protein (GFP), *Journal of Biomolecular NMR* 281-282.

37. Seifert, M. H. J., Ksiazek, D., Azim, M. K., Smialowski, P., Budisa, N., and Holak, T. a. (2002) Slow exchange in the chromophore of a green fluorescent protein variant., *Journal of the American Chemical Society 124*, 7932-42.

38. Ericsson, U. B., Hallberg, B. M., Detitta, G. T., Dekker, N., and Nordlund, P. (2006) Thermofluor-based high-throughput stability optimization of proteins for structural studies., *Analytical biochemistry 357*, 289-98.

39. Demachy, I., Ridard, J., Laguitton-Pasquier, H., Durnerin, E., Vallverdu, G., Archirel, P., and Lévy, B. (2005) Cyan fluorescent protein: molecular dynamics, simulations, and electronic absorption spectrum., *The journal of physical chemistry. B 109*, 24121-33.

40. Hyun Bae, J., Rubini, M., Jung, G., Wiegand, G., Seifert, M. H. J., Azim, M. K., Kim, J.-S., Zumbusch, A., Holak, T. a., Moroder, L., Huber, R., and Budisa, N. (2003) Expansion of the Genetic Code Enables Design of a Novel "Gold" Class of Green Fluorescent Proteins, *Journal of Molecular Biology 328*, 1071-1081.

41. Villoing, A., Ridhoir, M., Cinquin, B., Erard, M., Alvarez, L., Vallverdu, G., Pernot, P., Grailhe, R., Mérola, F., and Pasquier, H. (2008) Complex fluorescence of the cyan fluorescent protein: comparisons with the H148D variant and consequences for quantitative cell imaging., *Biochemistry 47*, 12483-92.

42. Cubitt, A. B., Woolenweber, L. A., and Heim, R. (1999) Understanding structure-function relationships in the Aequorea victoria green fluorescent protein, *Methods Cell Biology 58*, 19-30.

43. Nijergorodov, N.I., Downey, W. S. (1994) The Influence of Planarity and Rigidity on the Absorption and Fluorescence Parameters and Intersystem Crossing Rate Constant in Aromatic Molecules, *Journal of Physical Chemistry N 98*, 5639-5643.

44. Malo, G. D., Pouwels, L. J., Wang, M., Weichsel, A., Montfort, W. R., Rizzo, M. A., Piston, D. W., and Wachter, R. M. (2007) X-ray Structure of Cerulean GFP : A Tryptophan-Based Chromophore Useful for Fluorescence Lifetime Imaging, *Biochemistry 46*, 9865-9873.

45. Otwinowski, Z., and Minor, W. (1997) Processing of X-Ray Diffraction Data Collected in Oscillation Mode, *Methods in Enzymology; Macromolecular Crystallography, part A 276*, 306-315.

46. McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007) Phaser crystallographic software., *Journal of applied crystallography 40*, 658-674.

47. Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G. W., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. a, Powell, H. R., Read, R. J., Vagin, A., and Wilson, K. S. (2011) Overview of the CCP4 suite and current developments., *Acta crystallographica. Section D, Biological crystallography 67*, 235-42.

48. Murshudov, G. N., Skubák, P., Lebedev, A. a, Pannu, N. S., Steiner, R. a, Nicholls, R. a, Winn, M. D., Long, F., and Vagin, A. a. (2011) REFMAC5 for the refinement of macromolecular crystal structures., *Acta crystallographica. Section D, Biological crystallography 67*, 355-67.

49. Emsley, P., and Cowtan, K. (2004) Coot: model-building tools for molecular graphics., *Acta crystallographica. Section D, Biological crystallography 60*, 2126-32.