The Dissertation Committee for Chintan Kishore Modi Certifies that this is the approved version of the following dissertation:

Evolution of Structure-function Relationships in the GFP-family of Proteins

Committee:

Mikhail V. Matz, Supervisor

George Georgiou

Kevin Dalby

Ron Elber

Rebekka Wachter

Evolution of Structure-function Relationships in the GFP-family of Proteins

by

Chintan Kishore Modi, B.S.

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

The University of Texas at Austin

August 2014

Dedication

To my beloved parents,

Dr. Kishore F. Modi and Bhagvati K. Modi

Acknowledgements

The dissertation presented here culminated not only from years of my research, but also from the support, advice and guidance from my family, friends and mentors over the course of my graduate years at The University of Texas at Austin. I hope this acknowledgement does justice to their contribution to my dissertation.

First and foremost, I would like to express my gratitude to my parents, Dr. Kishore Modi and Bhagavati Modi. My father, a chemist, inspired and encouraged my love of science and facilitated all my scientific endeavors starting from my childhood. They remain my biggest advocate in life and give me confidence to continually climb new heights in my scientific career. I will remain ever indebted to my parents for the decision to immigrate to the United States, seeking better education for my brother and me. Without their unwavering support and sacrifice, this dissertation would have been an impossible dream. They stood behind me over the years through the thick and thin of my academic career and for that, words cannot express my sentiment. I am grateful to have such a loving, supportive and talented parents.

My brother, Dr. Kavan Modi, has been my friend, confidante and advocate in life and in my scientific career. I feel very fortunate to have overlapped in our graduate careers at The University of Texas at Austin. I will forever cherish our late night conversations discussing science to philosophy at Dolce and at his very small apartment. Our culinary adventures were one of the most enjoyable things during those early years of my graduate school, and they kept me sane during some of those difficult times. I am also ever grateful to him for his support during the lowest point of my graduate career and for his financial support during the last year of my PhD.

I would like to thank my mentor, Dr. Mikhail (Misha) V. Matz, for introducing me to the exciting field of molecular evolution and the green fluorescent proteins, changing the trajectory of my scientific career. However, I reserve the height of my gratitude for Misha for helping me at the lowest point of my graduate career by accepting me into his lab six years ago; without his help this dissertation would not have been possible. His encouragement, support and guidance have been invaluable in the successful completion of this dissertation. I feel especially obliged to him for his advice on the scientific thinking behind and presentation of the data.

The protein oligomerization work presented in this dissertation would not have been possible without the support and advice of Dr. Austen Riggs and Mrs. Clair Riggs. Their support of my efforts to study the biophysical characteristics of fluorescent proteins using multi-angle laser light scattering was invaluable. I will remain indebted to them forever for all their help.

The work done on the Least Evolved Ancestor (LEA) was done in collaboration with Dr. Rebekka Wachter. I am especially thankful to Dr. Wachter and Dr. Kim for supplying purified LEA mutants for biophysical characterization of these proteins.

The contribution of Dr. Randall Hughes for synthesizing the degenerate gene library for the study of the evolution of oligomerization of Hydrozoan fluorescent proteins was invaluable. Without his help, the work presented in chapter four of this dissertation would not have been possible. I am also grateful to all my other collaborators, several laboratories and professors at The University of Texas at Austin for their material, advice and intellectual support over the years. I would especially like to thank Drs. Jeffrey Gross, Thomas Juenger and Hans Hofmann for their mentorship and advice over the course of my graduate career. I remain indebted to Dr. Gross for his advice and help as I changed my advisor at the end of third year of my graduate school.

I would also like to thank my committee, Drs. George Georgiou, Ron Elber, Kevin Dalby and Rebekka Wachter, for the research and career advice, especially on life in science after graduate school.

My gratitude to Ms. Courtney Chandler from the University of Michigan for her tenacious work, which was invaluable in the completion of all three projects described in this dissertation. I thank Ms. Haley Spotts and Ms. Ava Ibanez for their help in carrying out many of the experiments for this dissertation as well.

Dr. Emin Ulug has been a great mentor and friend throughout my graduate career. I remain indebted to him for his support and advice over the past seven years. Through teaching virology and immunology lab for him I learned a great deal in these subjects. I also cherish our many conversations covering the many topics and facets of life in academia. I thank him for his mentorship and friendship.

I very much appreciate the Department of Ecology, Evolution and Behavior at The University of Texas at Austin for welcoming me into their department and offering me their support in the past six years. I would especially like to acknowledge Professor Laurence (Larry) Gilbert, the director of the Brackenridge Field Laboratory, for the opportunity to study the parasitism working with fire ants and phorid flies as his research assistant. I am also grateful to the members of the Gilbert lab, Drs. Rob Plowes and Nathan Jones, for welcoming into their lab and for their support.

I am ever grateful to my girlfriend, Nicole, for her patience during the past couple of years as I geared up to finish my dissertation; her support has been invaluable in completing this work. I thank her for bringing a certain level of sanity and focus to an otherwise chaotic life as a graduate student. I am also very much appreciative of her advice in writing this acknowledgment.

I am especially grateful to Drs. Teofil Nakov, Sari Andoni and the Matz lab members for their help and input in writing this dissertation and my oral defense. I am especially indebted to Teo for his invaluable help with organizing my defense. I thank Nicole Vance, Dr. Carly Kenkel, Dr. Sarah Davies, Rachel Wright and Marie Strader for proof reading this dissertation and for their helpful suggestions.

I cannot forget the patience, dedication and encouragement of all my teachers over the years; I thank them from the bottom of my heart. Above all, I am obliged to Mr. Baskinger, Mrs. Puccio, Mrs. Mehta, Mrs. Munshi, Mrs. Oshima, Mr. Sabella, Drs. *Vander Voort,* and Zirkes, who were particularly influential during my primary and secondary level of education.

I am grateful to all my friends from Austin to anywhere and everywhere in the world; you have made these past nine years fun and enlightening. Chandan, I thank you for your friendship and support over these past nine years, and for being my conciliary. Chandan you have been there for me at every turn ever since we met back in 1998 at Penn State. I am also very grateful to all my friends (Chandan Sarkar, Dr. Saddad Toor, Dr. Nirav Mistry, Dr. Navpreet Minhas, Dr. Neal Gandhi, Ajay Devadanam, Gaurav Deshmukh, Jay Suthakaran, Amandeep Jangi, Sam Zachariah, Ramon Talwar, Janice Colaco, Preeti Shah, Jeni Sudhakaran, Shirin Trehan, Sumit Minhas, Sumita C. Minhas, Dawn Howard, Rachna Vohra, Anisha Jangi, Piya Gandhi, Alex Gociu and Catherine Veeraraj-Devadanam) from Penn State and Parsippany High School for their unwavering friendship and support during these past nine years, and that they maintained our deep friendship even with our busy schedules and physical distance. I am obliged to Sari Andoni for introducing me to a great music scene and fantastic musicians in Austin.

I am grateful to have a wonderful and supportive extended family. I am particularly thankful to my uncles, Drs. Arvind and Amrut Hazari, my aunts, Dr. Saroj

and Hina Hazari and my cousins: Dr. Pushpal Hazari, Dhyan Hazari, Vivek Hazari and Mukur Hazari.

Lastly, I must thank my scientific home for the past six years, the Matz Lab. Matz lab is a very active and dynamic place simply because of its highly talented lab members. Because of them, the lab is continuously evolving, and performing new and groundbreaking research. I have learned a great deal in the fields of evolution, ecology and population genetics simply through interactions with my lab mates. I am grateful to have known and worked with Dr. Carly Kenkel, Dr. Sarah Davies, Roxana Capper, Marie Strader, Rachel Wright and Groves Dixon. I thank Dr. Eli Meyer for his friendship and scientific insight. I am especially grateful to him for taking the time to help me with my programming questions and problems. Also, it was wonderful to engage Eli in spirited political and scientific conversation. Galina, I thank you for all your support and help over the years; especially for your patience for my old plates and tubes in the lab refrigerator.

To all of my family, friends and supporters above, I give you my unconditional gratitude. Each of you helped make this journey possible and this achievement a reality.

Evolution of Structure-function Relationships in the GFP-family of Proteins

Chintan Kishore Modi, Ph.D.

The University of Texas at Austin, 2014

Supervisor: Mikhail V. Matz

One of the most intriguing questions in evolutionary biology is how biochemical and structural complexity arise through small and incremental changes; however answering this question requires an explicit set of candidate residues and an experimental system in which to test them. This dissertation aims to understand how biochemical complexity evolves and assesses the structure-function relationship in the green fluorescent protein (GFP) protein family using an ancestral reconstruction approach. In the second chapter, I studied the evolution of biochemical complexity in Kaede-type red fluorescent proteins (FPs) from *Faviina* corals. An increase in biochemical complexity is represented by the emergence of red fluorescence because it necessitates the synthesis of a tri-cyclic chromophore from a precursor bi-cyclic chromophore through an additional autocatalytic reaction step. The autocatalytic reaction is fully enabled by as many as twelve historical mutations. Here, I showed that the red fluorescent chromophore evolved from an ancestral green chromophore by perturbing the ancestral protein stability at multiple levels of protein structure. Moreover, only three historical mutations are sufficient to initiate the selection-accessible evolutionary trajectory leading to emergence of red fluorescence. The third chapter investigates six mutations proximate to the chromophore in the Kaede-type FP that could have facilitated autocatalytic synthesis of

the red chromophore by enlarging the chromophore-containing cavity and modifying its microenvironment. Two of these six mutations were found to strongly affect the protein's stability and oligomeric tendency. Additionally, I showed that the dimeric least divergent Kaede-type FP, R1-2, evolved from the tetrameric green ancestor. Taken together the results of these studies indicate that the step-up in biochemical complexity in the Kaede-type FPs was achieved via disruption of the existing stable interactions at tertiary and quaternary protein structure levels. In the fourth chapter, I resurrected the common ancestor of all FPs cloned from the order Leptothecata (class Hydrozoa), which are characterized by the highest known homo-oligomeric diversity. I showed that the ancestor was a green monomeric FP with a large Stokes shift. The ancestral FP together with the extant Leptothecata FPs could server as a model system to study the evolution of function and homo-oligomerization, and the desirable photophysical characteristics would make this ancestral FP a useful bio-marker in bio-medical research.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

The advent of molecular biology over the later half of the 20th centaury has revealed increasingly complex molecular machines and biochemical reactions that are integral parts of crucial biological processes. However, the discovery of this complexity has led to one of the central questions in evolutionary biology - how does biochemical and biophysical molecular complexity evolve through small and incremental changes? Specifically, how can protein evolution bring about novel structures and functions despite the presence of ubiquitous biophysical constraints related to protein folding and stability? To answer these questions it is necessary to identify a set of mutations that are necessary and sufficient to generate observed large structure-function difference in the extant protein families (Harms and Thornton, 2010; Chakrabarti and Lanczycki, 2007; Capra and Singh, 2008; Donald and Shakhnovich, 2009).

The study of molecular evolution aims to elucidate processes leading to the emergence of novel functions observed in orthologous and paralogous extant proteins (Harms and Thornton, 2010; Datta, 2009; Maita et al., 2010). Comparing and identifying the difference in sequences and structures of divergent protein family members has yielded a list of candidate residues for experimental studies; however this approach is often inefficient at identifying the candidate residues that can help understand the appearance of novel function (Harms and Thornton, 2010, Gerlt and Babbitt; 2009). This approach, termed "horizontal" because it uses extant proteins at the tips of phylogenetic trees, may not be capable of identifying distant sites in proteins that interacted and played crucial roles in the evolution of novel structure or function (Noivirt-Brik et al., 2009; Harms and Thornton, 2010). Using horizontal comparative approaches to untangle trajectories responsible for molecular complexity has proven challenging, in part due to

the stepwise nature of the evolutionary process and acquisition of neutral mutations (Harms and Thornton, 2010). Another approach could be systemically exploring sequence diversity in mutant libraries through the use of high-throughput screening methods. However, this method would be time consuming and would only be capable of exploring a fraction of potentially relevant sequence diversity (Harms and Thornton, 2010; Romero and Aranold, 2009).

Ancestral reconstruction using statistical phylogenetic methods has yielded some of the most promising results in this research field. This approach enabled biochemical and biophysical analyses of ancient proteins to reconstruct the history of events responsible for the evolution of protein structure and function (Thornton, 2004; Harms and Thronton, 2010; Kacar and Gaucher, 2012; Studer et al, 2013; Tokuriki and Tawfic, 2009). The ancestral reconstruction approach is also called a vertical comparative approach because it allows for the comparison of resurrected molecular complexes to their extant counterparts (Harms and Thornton, 2010; Finnigan et al., 2012). The use of ancestral reconstruction has been successful in adding to our existing knowledge of the evolution of the complex molecular machines, emergence on novel enzymatic function, and parallel evolution of functions. For example, Finnigan and colleagues (2012) showed that two mutations in a protein-protein interaction surface initiated the evolution of a higher order oligomeric state in the hexameric transmembrane ring of the eukaryotic V-ATPase proton pump (Finnigan et al., 2012). These results also helped to propose an initial mutational trajectory and to identify the candidate residues in the evolution of paralogous proteins (Finnigan et al., 2012). Ingles-Prieto and colleagues (2013) provided another example by resurrecting seven ancestral thioredoxins and through a crystallographic study of these ancient thioredoxis showed them to be structurally highly conserved, even though the ancient and extant proteins only share 50% sequence identity

(Ingles-Prieto et al., 2013). Ancestral resurrection has also been able to address the question of functional evolution in beta-lactamase and in the estrogen receptor (Risso et al., 2013; Thornton et al., 2003). Resurrected beta-lactamase demonstrated that the ancient protein was hyper-thermostable and had substrate promiscuity (Risso et al., 2013). Furthermore, this work also proposed that the evolution of function might have been enabled by protein dynamics (Risso et al., 2013). Yokoyama and colleagues (2008) reported interaction between different mutations in their ancestral reconstruction study of the evolution of red and green opsins, a family of light-sensitive G-protein coupled receptors. Recently, ancestral reconstruction has also helped elucidate the repeated and parallel evolution of hemoglobin in Andean hummingbirds (Projecto-Garcia et al., 2013). Examples of structural and functional evolution have been also demonstrated in the evolution of yeast alcohol dehydrogenases, MHC in birds, SABATH enzymes in plants, and steroid receptors in vertebrate (Thomson et al., 2005; Burri et al., 2010; Haung et al, 2012; Bridgham et al., 2006; Ortlund et al., 2007). In summary, the study of molecular evolution through the ancestral reconstruction approach enables us to understand the underlying evolutionary mechanism and determinants of molecular structure and function. Furthermore, the evolutionarily important residues highlighted by the ancestral reconstruction studies could be promising modification targets for engineering proteins with desired functional and structural properties.

This dissertation focuses on the evolution of novel biochemical functions and increase in molecular complexity, using GFP-like fluorescent proteins as a study system. Chapters two and three focus on the evolution of red fluorescence in Kaede-type FPs from Faviina corals. Ancestral reconstruction study of these proteins has previously demonstrated that they evolved from a green ancestor (Ugalde et al., 2004), and that this functional transition required a minimum of 12 mutations (Field and Matz, 2010).

Importantly, in order for red fluorescence to evolve, it required the autocatalytic synthesis of a tri-cyclic red fluorescent chromophore from a precursor bi-cyclic green chromophore through an additional reaction step (Kim et al, 2013). Importance of 12 mutations of 37 mutational differences between the green ancestor and the least divergent Kaede-type extant FP, R1-2 (coral protein R1-2) in the functional transition was demonstrated by introducing these mutations in the green ancestor to generate a Least Evolved ancestor (LEA; Field and Matz, 2010). Therefore, the emergence of red fluorescence from the green ancestral state represents an increase in functional complexity of the protein. In Chapter two, I explored the initial mutational trajectory in the evolution of red fluorescent chromophore from a green ancestor using forward mutagenesis of the ancestral protein and spectroscopic analysis. In Chapter three I used reverse mutagenesis and biophysical techniques to establish that mutations necessary for the emergence of red fluorescence also contributed to diminished thermal stability and oligomeric tendency of the Kaede-like FPs over the course of evolution.

Chapter four of this dissertation is focused on the evolution of structural rather than functional diversity in GFP-like proteins, specifically, the diversity of native homo-oligomerization states observed among hydrozoan FPs (Prasher et al., 1992; Ormo et al, 1996; Xia et al., 2002; Gurskaya et al., 2003; Shagin et al., 2004; Markova et al., 2010; Hunt et al, 2012; Aglyamova et al., 2011). This study revealed that the ancestor of FPs from order Leptothecata (including the most famous GFP from *Aequorea victoria*) was monomeric and green-fluorescent with near-UV absorption and a large Stokes shift. This result indicates that the ancestral function of Leptothecata FPs was not related to color conversion in bioluminescence (Morise et al., 1974; Ward et al., 1980), since for this function the near-UV absorption characteristics are unsuitable. Most importantly, this study establishes a unique model system to investigate the evolution of homo-

oligomerization and the function of hydrozoan FPs. Additionally, the resurrected monomeric Leptothecata ancestor with unusual photophysical properties adds to the list of potential genetically encoded fluorescent labels for biomedical research.

In summary, the research presented in this dissertation explored several aspects of evolution of structural and functional diversity in paralogous proteins, taking advantage of the ancestral reconstruction methodology and the unique study system represented by GFP-like fluorescent proteins.

Chapter 2: Evolution takes baby steps: Emergence of red fluorescence from ancestral green fluorescent protein in *Montastrea cavernosa*

ABSTRACT

To explain how biological complexity arises through seemingly improbable combination of events is one of the greatest challenges of evolutionary biology. Emergence of red fluorescence in coral GFP-like fluorescent proteins (FPs) represents an increase in functional complexity because it requires an additional autocatalytic step in the synthesis of a tri-cyclic chromophore from a bi-cyclic precursor. In Kaede-like red FPs from Faviina corals, it takes as many as 12 historical mutations to fully enable this reaction. Here, by means of forward mutagenesis of the resurrected tetrameric ancestral green FP, we found that just three of these mutations are sufficient to initiate the autocatalytic synthesis of the novel chromophore type; moreover, one of them would be amenable to selection on its own since it red-shifts the whole emission spectrum without detectable red chromophore production. Therefore, only one mutation had to be introduced through mechanisms unrelated to selection for red fluorescence, which makes its evolution highly probable. Two of the three enabling mutations significantly destabilize the ancestral FP, as evidenced by reductions in its thermal stability and oligomerization. One of them appears to disturb the hydrogen-bonding network around the chromophore-bearing helix while another disrupts the protein's oligomeric structure. Taken together, these data provide a unique example of the evolution of molecular complexity that is enabled by perturbing existing stable interactions at multiple levels of biological organization: in this case, at the tertiary and quaternary protein structure levels.

INTRODUCTION

Complex molecular machines are integral part of crucial biological processes. The discovery of these molecular machines begs the question how such complex biochemical and biophysical adaptations could have arisen through small and incremental changes? The study of molecular evolution attempts to uncover processes leading to the emergence of novel functions protein families by analyzing the evolution of their sequence (Harms and Thornton, 2010; Chakrabarti and Lanczycki, 2007; Capra and Singh, 2008; Donald and Shakhnovich, 2009).

In the past, comparative studies have suggested that novel functions and molecular complexity do indeed arise through small and incremental changes (Finnigan et al., 2012; Harms and Thornton, 2010); however, these studies lacked experimental procedures to test hypotheses about the series of biophysical changes in the ancestral protein which selection acted on to produce the current function of the gene. Furthermore, presence of neutral mutations in the extant proteins makes it extremely difficult to deconvolute the historical changes involving large and small-effect mutations acting in concert (Harms and Thornton, 2010; Gerlt and Babbitt; 2009; Noivirt-Brik et al., 2009; Liberles et al., 2012). Resurrecting ancient proteins solves this conundrum. Computational phylogenetic analyses allow us to infer ancestral protein sequences, which can then be resurrected to explore the biochemical and biophysical properties of these ancient molecules (Thornton, 2004; Harms and Thornton, 2010). In addition, this method elucidates the effects of historical mutations along the evolutionary trajectory to the extant protein. Importantly, vertical comparative approaches - utilizing the resurrected molecules and their extant counterparts - enable the study of the evolution of structure-function relationships and the underlying epistatic interactions (Breen et al, 2012; Ortlund et al, 2007; Weinreich et al, 2006; Weinreich et al, 2013). This approach has been

successfully used to recapitulate changes in protein-protein interaction specificity leading to the evolution of higher order oligomeric state in the extant hexameric transmembrane ring of the eukaryotic V-ATPase proton pump (Finnigan et al., 2012). Similarly, crystallographic study of resurrected ancient thioredoxins and their extant counterparts help establish that these enzymes are highly conserved (Ingles-Prieto et al., 2013).

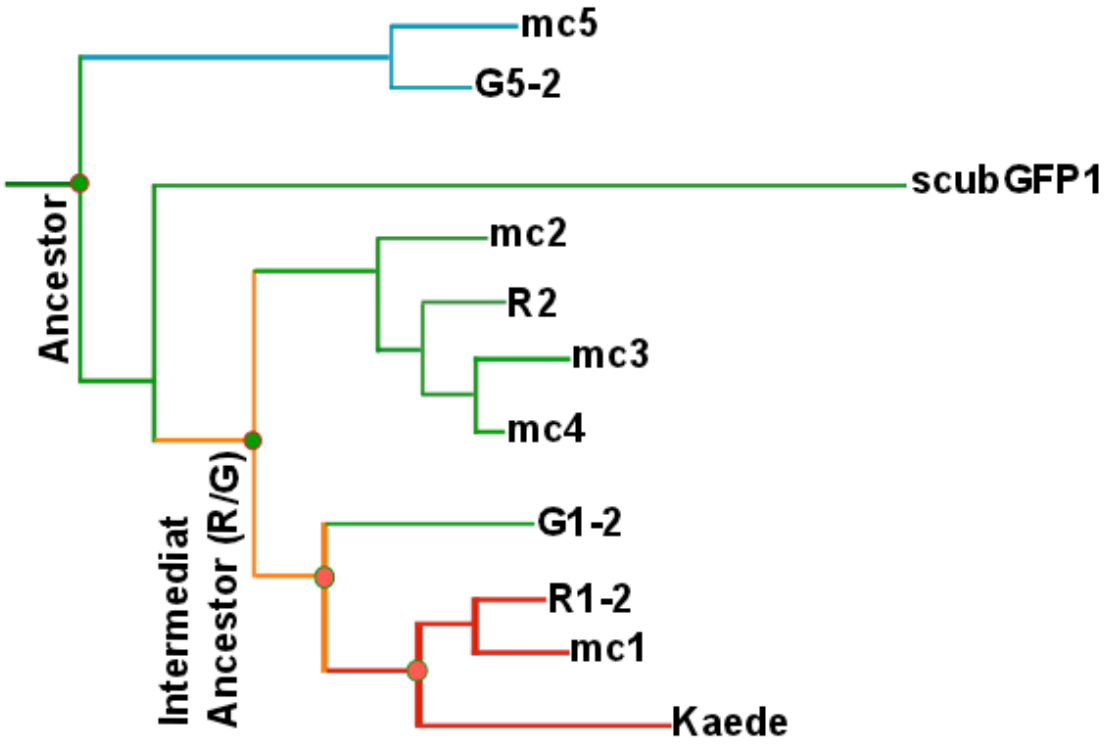


Figure 2.1: A diagram representing the relationships between Faviina coral FPs. Labeled branch tips are extant FPs. Two labeled internal nodes are ancestral proteins that were part of this study (from Ugalde, Chang, Matz 2004).

The GFP-family of proteins emerged as an ideal model to study the evolution of molecular complexity and biochemistry. Fluorescent proteins (FPs) exist across many

diverse marine organisms (Cnidaria, Arthropoda, Chordata) and exhibit a range of photophysical properties owing to diverse chromophore chemistries (Chudakov et al, 2010). Red fluorescence has emerged in many clades of Hydrozoa, Anthozoa and Chordata through convergent evolution, giving rise to varied chromophore chemistries (Chudakov, 2010; Pletnev et al., 2013). The origin of red fluorescence implies the increase in complexity of the autocatalytic process leading to the synthesis of the chromophore (Matz et al. 2002; Shagin et al., 2004; Ugalde et al, 2004). The great star coral *Montastrea cavernosa* has three different fluorescent colors (cyan, green and red) encoded by three different paralogous genes (Figure 2.1; Kelmanson and Matz, 2003). The red fluorescent Kaede-type FPs found in this clade exhibit unique irreversible green-to-red (G/R) emission transition when exposed to ultra violet radiation (UV-A) (Kelmanson and Matz, 2003). These genes have been shown to evolve from a green fluorescent common ancestor (Figure 2.1; Ugalde, 2004), whereby the red fluorescent FPs acquired the ability to synthesize a tri-cyclic chromophore from the ancestral bi-cyclic green chromophore via a beta-elimination reaction assisted by the UV-A (395 nm) light (Kim et al., 2013; Hayashi et al., 2007; Tsutsui et al., 2009). Subsequent work has determined that natural selection has facilitated at least some of the mutations responsible for color diversification (Field et al., 2006) and that out of 37 historical mutations separating the extant red FP (R1-2) from the green ancestor, twelve are necessary and sufficient to fully recapitulate the red fluorescence. (Field and Matz, 2010). The FP generated by introduction of these 12 mutations was labeled the Least Evolved Ancestor (LEA; Field and Matz, 2010).

For the natural selection to facilitate the assembly of such an extensive mutational combination each subsequent mutation should benefit the evolving function. The order of mutations satisfying this criterion is termed a “selection-accessible trajectory” (Weinreich

et al., 2006; Harms and Thornton, 2010; DePristo et al., 2005). Studies in other proteins have shown that selection-accessible trajectories can be very rare among all possible mutation combinations (Weinreich et al., 2006). In the case of red FP, a fully accessible trajectory does not even exist since none of the 12 mutations alone gives rise to even a trace of red fluorescence (Field and Matz 2010). This implies that a certain number of mutations should have been assembled by pure chance before natural selection could play a role. The purpose of this study was to determine which of the 12 mutations were sufficient to reach a selection-accessible trajectory, and which transitions within the protein accompanied this "unlocking" of the evolutionary potential. In the present we determined the minimal set of historical mutations that led to the emergence of red fluorescence from the ancestral green state by studying seven mutations (Q65H, T72A, S110N, V225E, R227H, Y227.1del and M228G) (Field and Matz, 2010). We propose that the green fluorescent protein, at the root of the *Montastraea cavernosa* fluorescent protein tree took the first steps of evolving towards the red chromophore through a combination of three mutations: Q65H, T72A and Y227.1 deletion. Our results show that a minimum of three changes in the green ancestor were sufficient to initiate the first steps towards evolution of red fluorescence observed in extent R1-2, but the evolution of full-fledged red chromophore was achieved through changes at an additional nine positions (Field and Matz, 2010). We found that just three mutations were sufficient to initiate a color change through diminished stability and oligomeric tendency of the protein and to make the subsequent mutational trajectory accessible for natural selection toward increased molecular complexity and novel molecular function.

MATERIALS AND METHODS

Collection of FP variants

We have taken advantage of a large collection of ancestral FPs and evolutionary intermediates that we have generated previously (Figure 2.1). As a template for site-directed mutagenesis described below we have used the common ancestor of FPs of all three coral colors (cyan, green, and red), the green fluorescent protein designated ALL, bearing a single mutation (Q65H). This mutation is essential for red fluorescence since the side chain of His65 is forming an integral part of the red-emitting chromophore. This mutation, however, has no effect on fluorescence phenotype when introduced alone (Field et al., 2006; Field and Matz, 2010). The properties of the generated mutants were compared to the common ancestor of red and green proteins (RG), the "least-evolved ancestor" (LEA, fully recapitulating the red fluorescent phenotype via 12 mutations to the ALL protein), and the extant red fluorescent protein R1-2, which was cloned from a coral (Kelmanson and Matz, 2003) and has 37 mutations compared to ALL.

Forward mutagenesis

All residue numbers are given following the sequence of the GFP from *Aequorea victoria* (Prasher et al, 1992). We have considered seven mutations (Q65H, T72A, S110N, V225E, R227H, Y227.1del and M228G) based on their early evolutionary appearance and their strong negative effect on the red fluorescence when mutated back to the ancestral state in the LEA (Field and Matz 2010). The site-directed mutagenesis PCR reactions were performed on the ancestral green FP gene in pGEM-T vector. Forward mutagenesis of the green ancestor targeting seven historical mutations was done using QuikChange II kit (Agilent, Santa Clara, CA). Each of the seven codons was mutated to

the extant state as in R1-2 (Field et al., 2010). To evaluate the combinatorial effect these mutations on the development of red fluorescence, further site-directed mutagenesis was carried out, introducing mutations sequentially. Through the process of sequential site-directed mutagenesis a collection of seven mutant FP variant were generated: (i) smQ65H: single mutant Q65H, (ii) dmT72A: double mutant Q65H, T72A, (iii) tmS110N: triple mutant Q65H, T72A, S110N, (iv) tmV225E: triple mutant Q65H, T72A, V225E, (v) tmR227H: triple mutant Q65H, T72A, V227H, (vi) tmV227.1del: triple mutant Q65H, T72A, Y227.1deletion, (vii) tmM228G: triple mutant Q65H, T72A, M228G (Figure 2.3A, B). All mutant clones were transformed into chemical competent One Shot Top-10 Escherichia coli cells (Invitrogen by Life Technologies; Carlsbad, CA, USA). Transformed cells were plated on Luria-Bertani-agar plates with isopropyl β-D-1-thiogalactopyranoside (IPTG; 0.5mM of final concentration) and 1X ampicillin (Amp; 100 μg/mL of final concentration), incubated at 37ºC for one day. All clones were screened for fluorescence using a Leica MZ FIII stereomicroscope (Leica; Wetzlar, Germany) with the filter set BL/VIO (# 11003, Chroma Technologies Corp., Bellows Falls, VT). All mutations were verified by Sanger sequencing (Figure 2.2). Since the emission spectral characteristics of the green ancestor and smQ65H reported to be the same, all mutants were evaluated for their functional and biophysical impacts through comparison of their spectral, stability and oligomerization characteristics to that of smQ65H and the extant FP (R1-2).

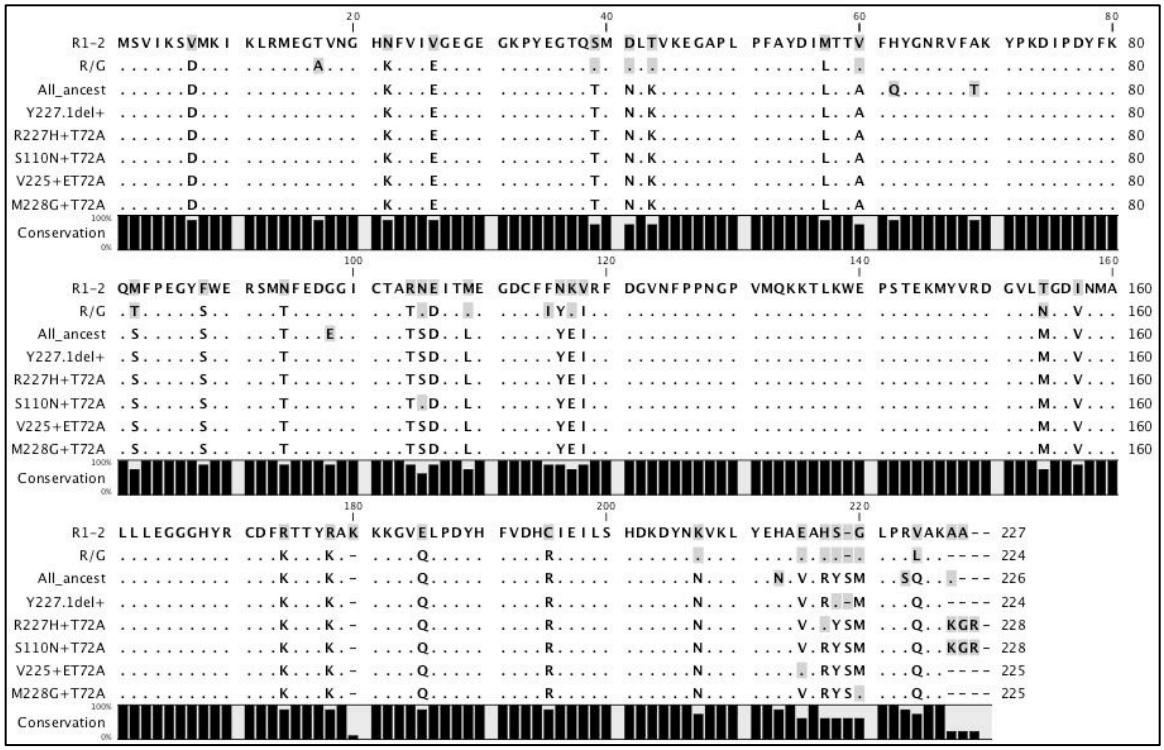


Figure 2.2: Multiple sequence alignment (amino acids) of fluorescent proteins. Multiple sequence alignment of the double mutant, triple mutants, intermediate ancestor (R/G) and the extant FP, R1-2.

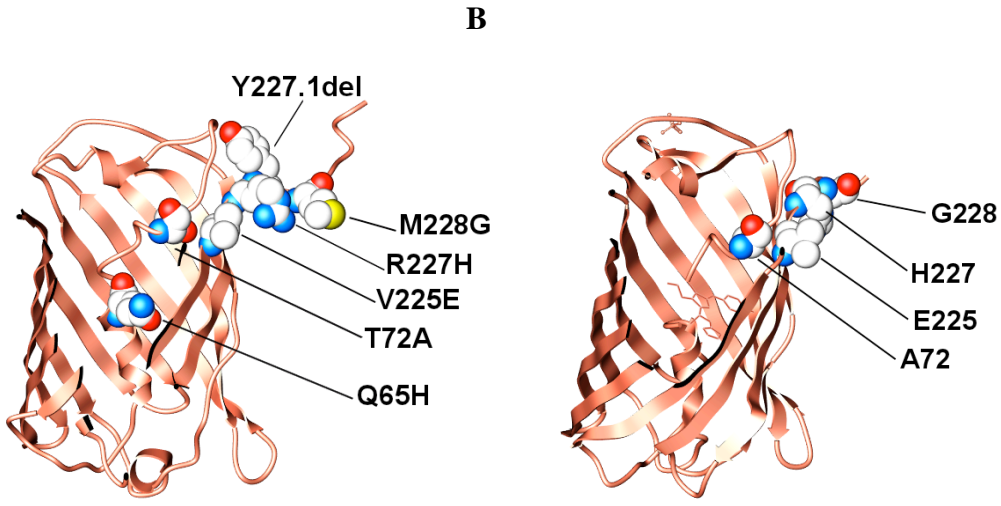


Figure 2.3: Residues targeted for the study of emergence of red fluorescence. **(A)** A structure of the ancestral FP (PDB: 4DXI) showing the positions of mutations. **(B)** A FP structure (PDB: 4DXN) showing the key residues after mutagenesis. Note the structure shows an unconverted green chromophore harboring histidine acquired through a mutation at position 65.

Expression and purification of fluorescent proteins

Fluorescent colonies were suspended in Amp-containing LB media with and streaked on large format plates containing 0.5mM IPTG and 100 μg/mL Amp, and incubated in the dark over-night at 37ºC and then in dark at room temperature for three days until sufficient bacterial biomass was accumulated. FPs were shielded from ambient light in all subsequent steps. Cells were harvested using 1X HEPES buffer (10mM HEPES, 100mM NaCL at pH 7.4). Harvested cells were sonicated for 20 minutes using 15-second pulse and 30-second rest protocol on the Misonix Sonicator 3000 instrument (Misonix Inc., Farmingdale, NY). Cells were kept on ice during the sonication procedure. Lysed cells were centrifuged at 4000RPM at 4ºC for 45minutes to obtain clarified supernatant.

The FPs were purified from the clarified supernatant using the tri-phase extraction protocol. Briefly, Ammonium sulfate (4M) was added to the clarified supernatants to the final concentration of 1.5M, followed by addition of tert-butanol in volume equal to the combined volume of Ammonium sulfate and clarified supernatant. Samples were mixed and then centrifuged to obtain phase separation. Aqueous phase with soluble FPs was collected and the organic phase was discarded. The phase separation step was repeated up to three more times by adding same volume tert-butanol, which resulted in an insoluble FP pellet to fully collect the protein (in between the aqueous and organic phases). Insoluble pellet was collected and re-dissolved in 1-5mL of 1X HEPES buffer. Dissolved FPs were centrifuged at 13000RPM for 5minutes to remove any insoluble proteins and were concentrated using Amicon®Ultra-15 centrifugal filter units with a 10kDa cutoff (Millipore; Billerica, MA, USA). The concentrated FPs were snap frozen in liquid nitrogen and then stored at -80ºC.

The extant FP (R1-2) with 6X His-tag was purified from the clarified supernatants by adding 2mL of Ni-NTA resin (Qiagen, Netherlands) and incubating overnight at 4ºC on a shaker. Bound FPs were washed 3-times with 3mL of 1X HEPES buffer containing 20mM Imidazole (pH 7.4) and then eluted with 5mL of 1X HEPES buffer containing 500mM Imidazole (pH 7.4). Eluted FPs were concentrated using Amicon®Ultra-15 centrifugal filter units with a 10kDa cutoff (Millipore; Billerica, MA, USA). Protein concentration was determined by BCA assay (Thermo Fisher Scientific, IL). Concentrated FPs were snap frozen in liquid nitrogen and stored at -80ºC.

Comparative spectroscopy of mutant FPs

To exclude the possibility that the observed spectral differences among mutants are due to variation during growth and protein purification, the effects of mutations on emission spectra were evaluated in bacteria expressing the compared mutant FPs grown on the same LB-agar plate. Bacterial spots were obtained by transforming 20uL of chemically competent Z-cells (Zymo Research, Irvine, CA) with ~30 ng of plasmid followed by plating 2.5uL of transformants in triplicate and growing them overnight at 37^0C. The plate was then exposed to UV-A ("black light") for 7 hours, since the maturation of these proteins requires exposure to 400nm light (Mizuno et al., 2003). The fluorescent spectra from bacterial spots and the agar background were collected at 0, 30, 60, 90, 120, 180, 240 and 420 minute time points using a USB 2000 spectrometer (Ocean Optics Inc., Dunedin, FL) attached to a Leica MZ FIII stereomicroscope (Leica, Wetzlar, Germany) with BL/VIO long-pass filter set (#11003, Chroma Technologies Corp., Bellows Falls, VT). Triplicate spectral measurements for each FP were averaged and the average agar background fluorescence was subtracted from each of the average fluorescence spectrum. Each of the average fluorescence spectra were normalized using the maximum fluorescence read in those spectra.

SDS-PAGE profiling of non-denatured and denatured FPs

Seven purified FPs (single mutant, double mutant and five triple mutants) were analyzed by non-denaturing SDS-PAGE. All FPs were shielded from ambient light during SDS-PAGE run. All FPs were diluted in 1x HEPES buffer to 250 μg/mL. Diluted FPs were mixed with the equal volume of 2x SDS-containing loading buffer. Pre-cast 4-20% gradient gels (Bio-Rad; Hercules, CA, USA) were loaded with 30uL of unheated protein samples (non-denatured) and heated protein samples with 2-mercaptoethanol

(denatured). Samples were run in 1x SDS-Tris-Glycine running buffer (Bio-Rad, Hercules, CA, USA) at 120v. Native FPs were imaged in the gels using a UltraViolet Transilluminator light board [UV-B: 305 nm] (UVP; Upland, CA, USA) followed by staining using Coomaisse G-250 BioSafe Dye (Bio-Rad; Hercules, CA, USA) and imaging under white light.

Determination of molar mass and oligomeric states by multi-angle laser light scattering

The molar mass and oligomeric state of FPs (at 150 μg/mL, except R1-2 at 250 μg/mL) were analyzed by HPLC size-exclusion chromatography with combined detection *via* multi-angle laser light scattering (MALLS) and refractometry. G4000PW$_{XL}$ or G3000PW$_{XL}$ columns were used (TosoHass, 300 x 7.8mm; Tosoh Bio-science LLC; King of Prussia, PA, USA) with buffer containing 10 mM HEPES and 100 mM NaCl adjusted to pH 7.4. EOS photometer and Optilab rEX differential refractometer detectors were used (Wyatt Technology; Santa Barbara, CA, USA). Data were analyzed using ASTRA software v. 4.9 and 5.3 (Wyatt Technology; Santa Barbara, CA, USA).

The static light scattering results are presented in terms of the Rayleigh ratio as a function of elution volume for each FP. The estimated molar masses for each peak were calculated in Astra 4.9 by taking a weighted-average value of all molecular species in a peak. The light scattering signal at the edges of a peak or at a low concentration of molecular species produces a low signal to noise ratio (S/N ratio), which decreases the accuracy of the calculated molar mass. In order to reduce the influence of molecular species on the edges of closely eluting peaks, we calculated molar masses for the majority of predominant peaks by using the trailing half of the peak. This approach of calculating molar masses was verified by accurately determining the estimated molar

mass of the control protein, BSA, within 3% of the predicted monomeric molar mass (66.4 kDa). By using the trailing half of the predominant peak of BSA (monomer) we were able to reduce the influence of closely eluting dimer peaks. Lastly, we determined the contribution of each oligomeric state in each FP sample based on the cumulative weight fraction analysis (Figure 2.9B). We based the oligomeric states corresponding to each peak on the calculated molar mass associated with that peak. The largest peak represents the predominant oligomeric form.

Differential scanning fluorimetry

Due to the innate fluorescence of the GFP-family of proteins, their thermal stability can be directly measured by differential scanning fluorimetry (DSF) without the use of additional fluorescent dyes (Dai et al., 2007; Kiss et al., 2009). Purified FPs were thawed on ice, mixed quickly by vortexing and centrifuged for 1min. FPs were diluted in HEPES buffer (50mM HEPES, 20mM NaCl, 1mM EDTA at pH 7.9). Three dilution series were prepared: (i) 1 mg/mL to 37.125 μg/mL, (ii) 50 μg/mL to 1.56 μg/mL and (iii) 200 μg/mL to 1.56 μg/mL. The dilutions in an optical 96-well plate (50 uL per well) were heated from 40ºC through 99ºC at a ramp rate of 0.01ºC second^{-1} while monitoring innate FP fluorescence in LightCycler 480 instrument (Roche Applied Science). Multi Color Hydrolysis Probe (the FAM filter combination: excitation-483nm and emission-533nm) detection format was used for the assays. The detection formats had 0.2sec integration time, with the melt factor set to 1 and the quant factor set to 10. FP fluorescence was measured during the continuous acquisition mode: (i) 50 acquisitions per second were made between 40ºC to 50ºC, and (ii) 75 acquisitions per second were made between 50ºC to 99ºC. The melting temperature (Tm) was determined as the

temperature at which the fluorescence amplitude dropped by half. Each FP was assayed in duplicate (independent dilution series).

Protein structure modeling

In order to better understand the effect of mutations on the structural arrangements of some of the key residues involved in chromophore synthesis, we generated structural models of all FPs using I-TASSER, which builds 3D models by generating multiple-threading alignments and iterative simulations (Zhang, 2008; Roy et al., 2010 and 2012). The structural modeling of all FP was done without setting any conditions. The algorithm was unable to model autocatalyzed FP chromophore; therefore, the three chromophore amino acids were modeled in non-catalyzed state. Threading and rendering of protein models and two crystal structures (PDB: 4DXI, 4DXN) was done using Swiss-PdbViewer and Molsoft (Guex, N. and Peitsch, M.C, 1997 and 1996; Abagyan et al., 1994; Molsoft, Sorrento Valley, CA).

RESULTS

Spectroscopic characterization of mutant FPs

A mutation at position 65 in the ancestor is essential for red fluorescence but by itself does not result in any change in spectroscopic properties relative to the green ancestor (Field and Matz, 2010). Therefore, the effect of combination of mutations on the emergence of red fluorescence was assessed relative to smQ65H. The dmT72A mutant had its whole emission spectrum shifted towards red by 15nm, but there was no trace of emerging red fluorescence peak around 578 nm signifying the tri-cyclic chromophore formation (Figure 2.4). The triple mutant with an additional deletion at position Y227.1

(tmY227.1del), in contrast, had a clearly discernible shoulder in its emission spectrum at 578nm after 4hrs of UV-A exposure. Data collected at 420 minutes was not included Figures 2.4 and 2.5, since there was no observable change detected in the emission curves after 240 minutes. This shoulder matches the red fluorescence peak of the evolutionary intermediate R/G protein (Figure 2.4). Two other triple mutants (tmV225E and tmR227H) also gained a red fluorescence shoulder after 240 minutes of exposure to UV-A (Figure 2.5); however, it was only barely detectable. There were no changes observed in the spectral characteristics of the remaining two triple mutants, tm M228G and tmS110N, when compared to the single and double mutant spectra (Figure 2.5).

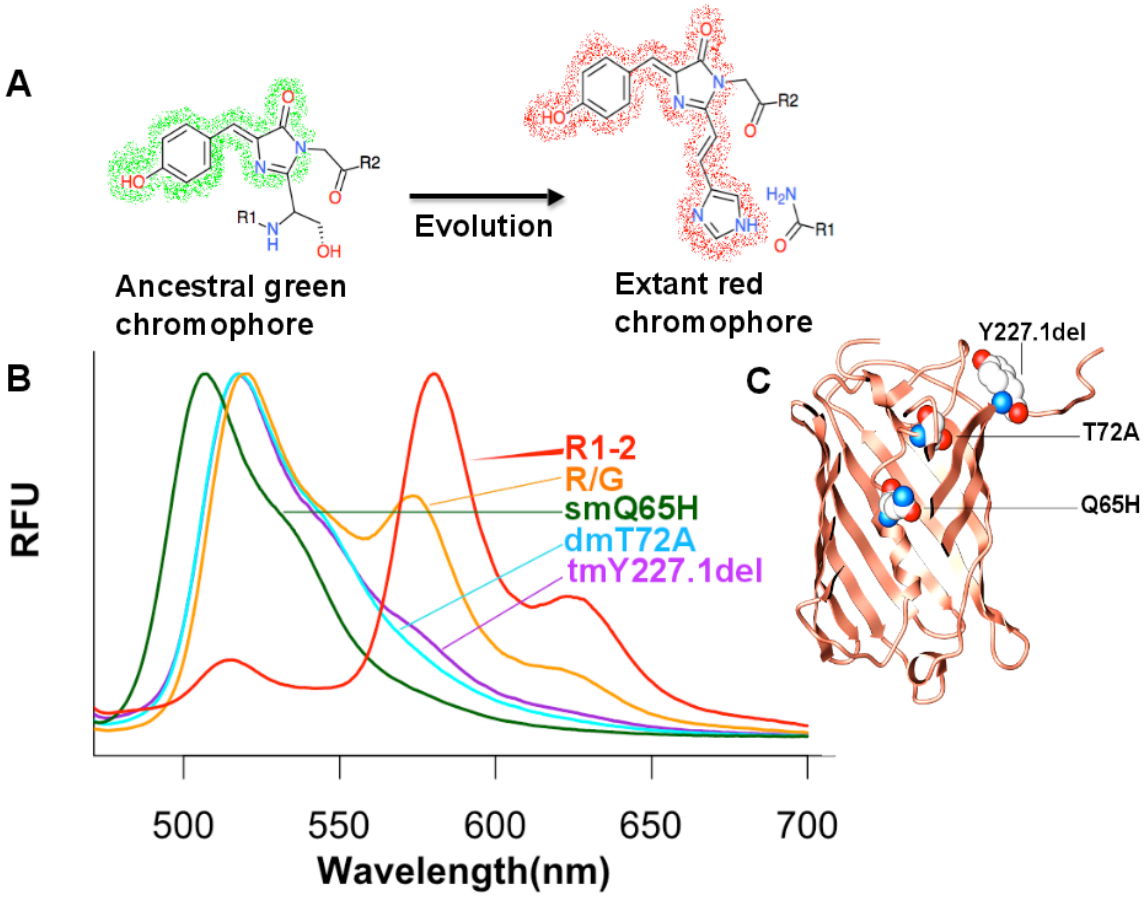


Figure 2.4: Three mutations are sufficient to initiate the evolution of red fluorescence from the ancestral green protein.
(A) Structures of the ancestral green chromophore and the extant red chromophore. **(B)** Spectroscopic characterization of the normalized emission spectra of mutants, intermediate ancestor (R/G) and extant FP (R1-2). The emergence of red fluorescence was detected as a shoulder in the emission spectra of tmY227.1del (green ancestor-Q65H, T72A, Y227.1del) and corresponded to the red emission peak of the intermediate ancestor (R/G). **(C)** The green ancestral FP structure depicting positions of the three mutations (as in tmY227.1del) required for the emergence of red fluorescence.

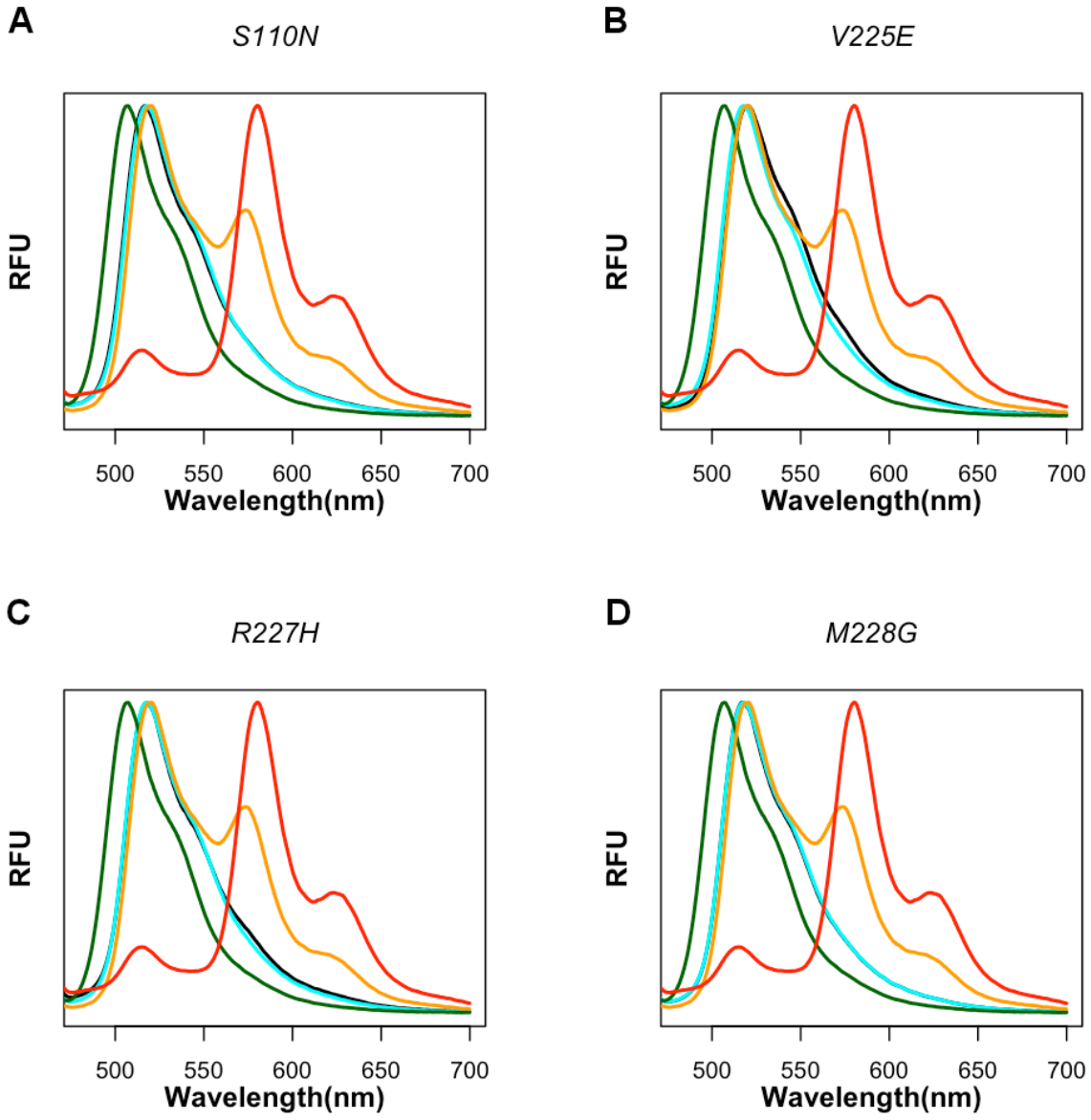


Figure 2.5: Normalized emission spectra of the four triple mutants.
(A) tmS110N, **(B)** tmV225E, **(C)** tmR227H and **(D)** tmM228G. The emission spectrum of the mutant is shown in black and compared to that of smQ65H (green), dmT72A (cyan), intermediate ancestor (R/G; orange) and extant FP (R1-2; red). M228G and S110N are identical to dmT72A, whereas V225E and R227H show a very subtle red emission shoulder at 570-580 nm.

The ratio of red fluorescence (578 nm) to the green fluorescence (518 nm) increased as a function of UV-A exposure time (420 minutes) in tmY227.1del compared to dmT72A (Figure 2.6A), indicating the UVA-dependent generation of the tri-cyclic red-emitting chromophore. The same analysis also confirmed UV-A-dependent production of a small amount of the red-emitting chromophore in tmV225E and tmR227H (Figure 2.6C, D), but none was observed for tmM228 and tmS110N (Figure 2.6B, E). Data collected at 420 minutes was included in calculating the ratio of red to green fluorescence even though there was no observable change detected in the emission curves after 240 minutes, because there was a small increase the red to green fluorescence ratio was observed for tmV225E and tmR227H (Figure 2.6C, D).

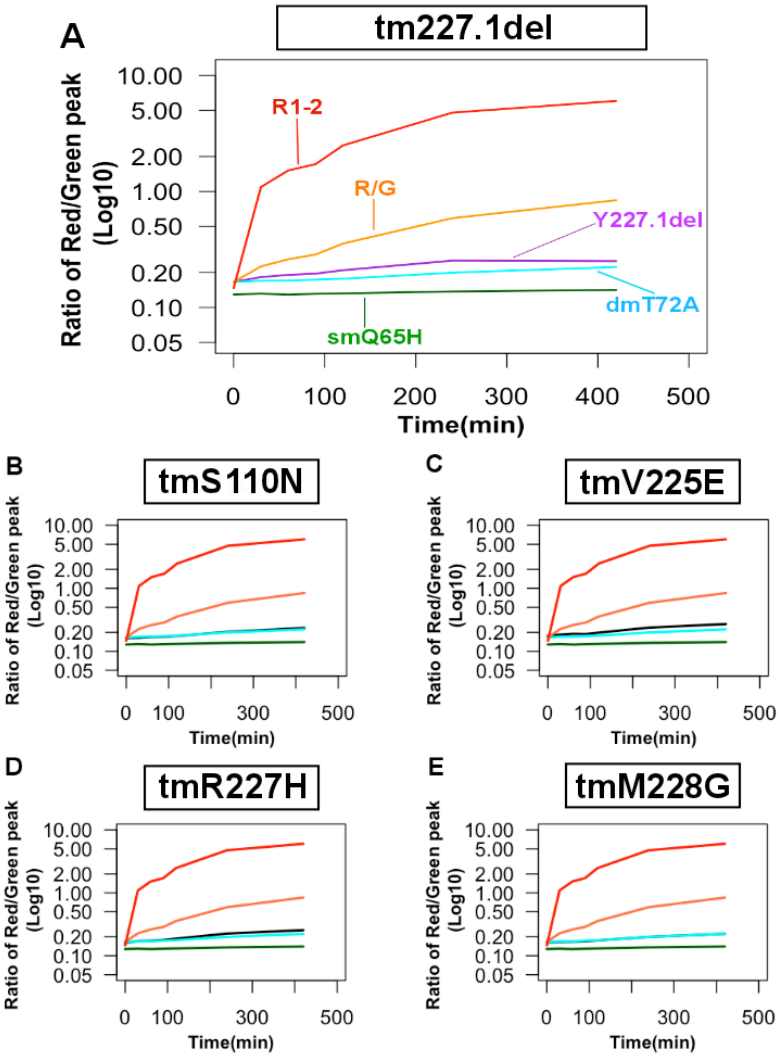


Figure 2.6: Change in ratio of red fluorescence (578nm) to green fluorescence (518nm) over time of UV-A exposure.
On each panel, the data for triple mutants are compared to the data for smQ65H (green), dmT72A (cyan), G/R ancestor (orange) and extant FP (red). **(A)** Three mutations of tmY227.1del exhibited the highest increase in red fluorescence with respect to green fluorescence. **(B, E)** No changes in the red to green ratio of tmS110N and M228G were detected with respect to that of dmT72A. **(C, D)** Slight increases in the red to green ratio were detected with respect to that of dmT72A for triple mutants, tmV225E and tmR227H.

To assess the initial increase red fluorescence during first 30 minutes of UV-A exposure, we calculated percent change in red to green fluorescence ratio between 0 and 30 minutes for each of the seven mutants relative to smQ65H and dmT72A. The double mutant, dmT72A, exhibited increase in fluorescence when compared to smQ65H. The most significant change in red fluorescence was observed the intermediate ancestor (R/G) and extant FP (R1-2). Of all the triple mutants, the results showed tmY227.1del had the most significant increase in red fluorescence after 30 minutes of UV-A exposure, especially when compared to the double mutant (Figure 2.7). The other two mutants, tmV225E and tmR227H, also showed increase in red to green fluorescence ratio over the double mutant (Figure 2.7). There was no increase in red to green fluorescence ratio detected for the tmM228G with respect to dmT72A (Figure 2.7). When compared to red to green fluorescence ratio of dmT72A after first 30 minutes of UV-A exposure, the triple mutant S110N showed decrease in it red to green fluorescence ratio (Figure 2.7).

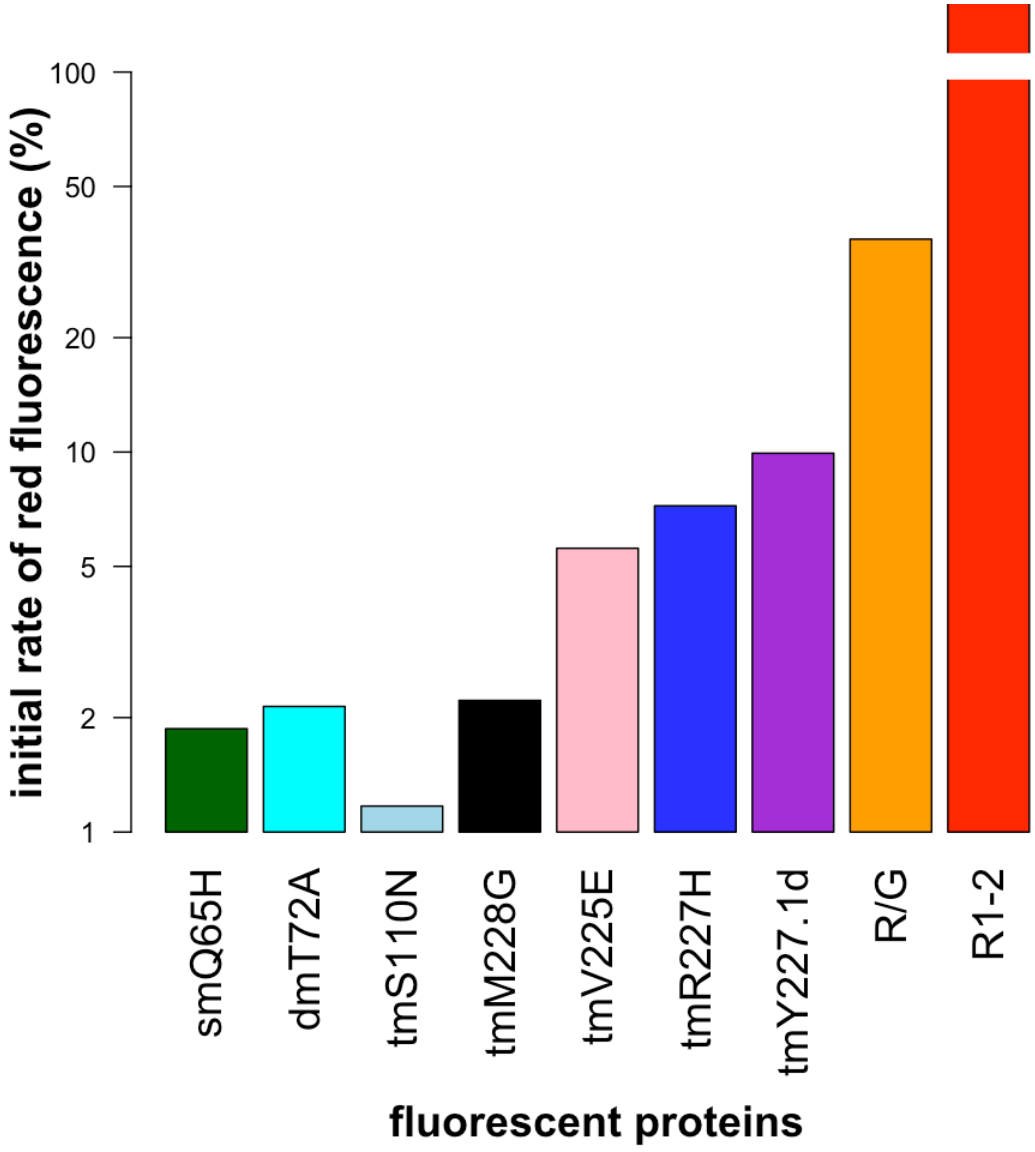


Figure 2.7: Percent increase in red to green fluorescence ratio after 30 minutes of exposure to UV-A.
Percent increase in the ratio of mutants relative to those of the extant red protein, R1-2 (red), and intermediate G/R ancestor (orange). Note that the Y-axis is on the logarithmic scale. The three mutations in tmY227.1del (purple) led to the highest increase in red fluorescence. Two other triple mutants, tmV225E (pink) and tmR227H (blue), also shows increase red fluorescence, albeit a small increase.

SDS-PAGE characterization of non-denatured and denatured mutant FPs

The results of the non-denaturing SDS-PAGE analysis showed that all mutants migrated the same distance except for the tmY227.1del, which migrated much further, suggestive of diminished oligomerization state (Figure 2.8). All denatured FP mutants migrated towards the bottom of a gel corresponding to the estimated molar masses of ~27kDa (data not shown). It was very surprising that a single mutation, Y227.1del, caused such a significant change; however, these results could not provide any definite information on the oligomeric state or the stability of tmY227.1del. The results showed likely dissociation and/or destabilization of the tetrmer as a possible cause for the distinct gel migration pattern.

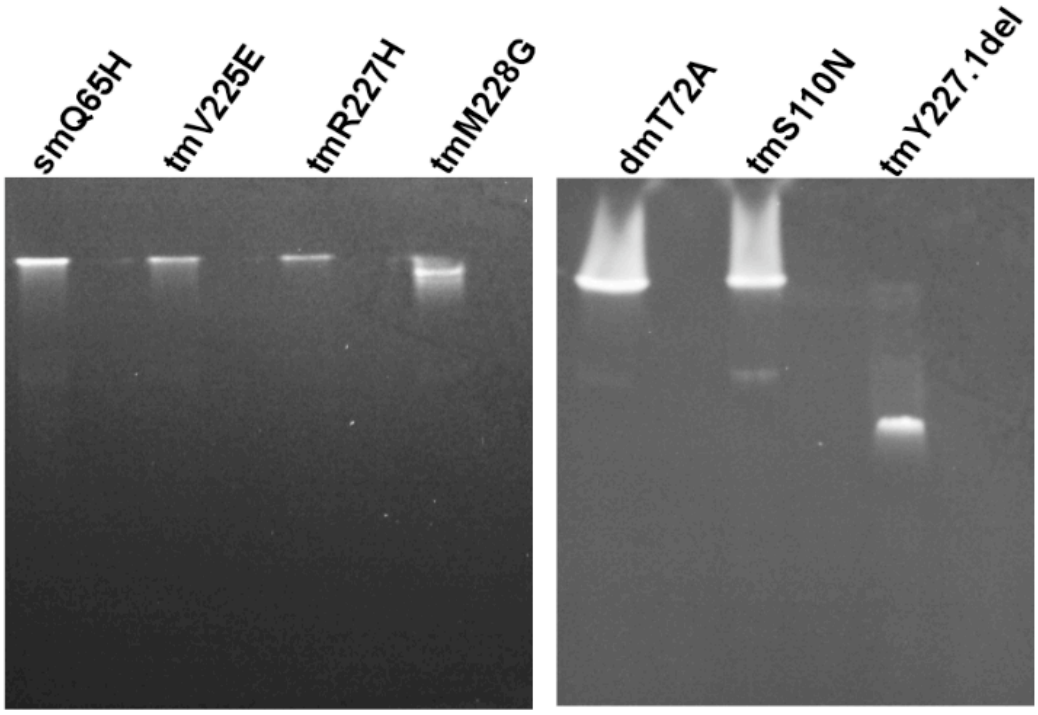


Figure 2.8: Non-denaturing SDS-PAGE of mutant FPs.
Native FPs imaged by exciting the native proteins' fluorescence by UV light. The mutant tmY227.1del shows increased mobility suggesting lower oligomerization tendency.

The three key mutations diminished oligomeric tendency of the ancestral FP

The SDS-PAGE results prompted us to assess the oligomeric state of Y227.1del triple mutant with using SEC-MALLS. In addition to tmY227.1del, we also determined the oligomeric states of smQ65H, dmT72A, and the extant FP (R1-2). The molar mass assigned to the major peaks of smQ65H (1.212×10^5 g/mol) and dmT72A (1.007×10^5 g/mol) corresponded to the canonical tetrameric structure (Figure 2.9A). However, a very small amount of monomer was also detected in the tails of main chromatographic peaks of both mutants. On the other hand, tmY227.1del had two distinct peaks; smaller of the

two peaks corresponding to the tetramer (~1x10^5 kDa), while the larger peak corresponding to a monomer (2.849 x 10^4 g/mol) (Figure 2.9A). The cumulative weight fraction analysis of these three mutant FPs indicated that all three FPs assume the same stoichiometric states, between tetramer and monomer (Figure 2.9B). Still, the cumulative weight fraction data taken together with the relative amounts of the stoichiometirc state (oligomeric states) determined based on the area under a peak, suggest that tmY227.1del favored the monomeric state, while smQ65H and dmT72A favored the tetrameric state. Surprisingly, we discovered that the extant FP, R1-2, was a dimer with molar mass of 4.867 x 10^4 g/mol (Figure 2.9A). The cumulative weight fraction analysis of R1-2 showed presence of a single stoichometric state corresponding to a dimer (Figure 2.9B).

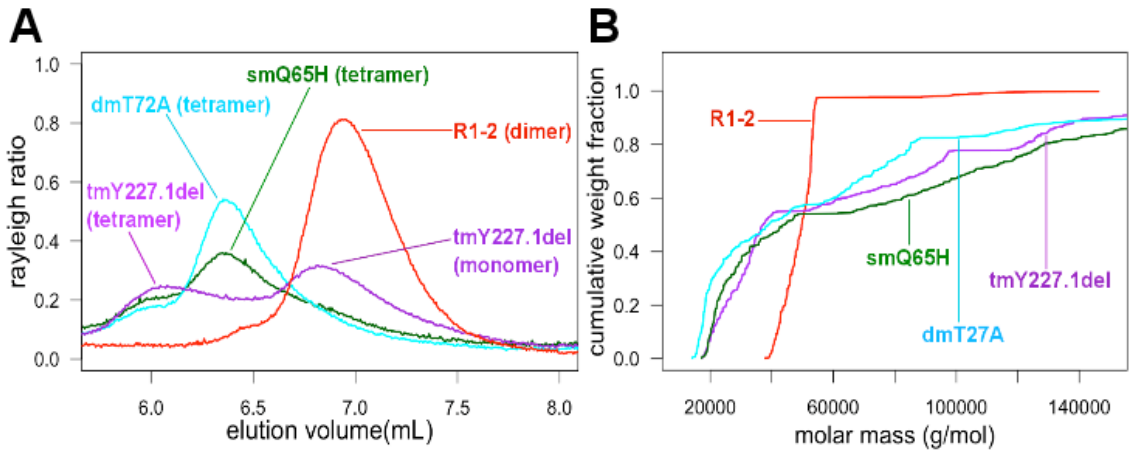


Figure 2.9: Molar mass and oligomeric states of the fluorescent protein complexes. **(A)** Oligomeric states separation of the proteins by size-exclusion chromatography and the molar mass detection by multi-angle laser light scattering (SEC-MALLS). The indicated oligomeric states of the mutants (smQ65H, dmT72A, tmY227.1del) and extant FP (R1-2) are inferred based the MALLS calculated molar mass. X-axis: elution volume (SEC), Y-axis: Rayleigh ratio (MALLS). The tmY227.1del is predominantly a monomer, with smaller amount as a tetramer. The extant FP (R1-2) is a dimer rather than a tetramer. **(B)** Oligomeric state stiochiometry in each of the proteins elucidated by the cumulative weight fraction analysis using the MALLS data. All three mutants are exploring same oligomeric states – predominantly between monomer and tetramer. The extant FP (R1-2) is in a single stiochiometric state – dimer.

Diminished stability along the evolutionary trajectory

The thermal stability of the single, double and all five triple mutants was measured using a DSF technique (Dai et al., 2007; Kiss et al., 2009; Niesen et al., 2007). In all Fps we have consistently observed two different Tm values for concentrations below 10 and above 100 μg/mL, suggestive of different oligomeric states (Figure 2.10). The difference in melting temperature of the single (smQ65H) and double (dmT72A) mutants was ~10ºC at high protein concentrations (1000-250 μg/mL) and ~3ºC at low

protein concentrations (7.81-1.95 μg/mL). There was no difference in the stability profiles of dmT72A and all five triple mutants at higher protein concentrations (1000-125 μg/mL). However, adding deletion at position 227.1 (tmY227.1del) to the double mutant led to additional ~10ºC loss of stability at lower protein concentrations (7.81-1.95 μg/mL) (Figure 2.10). There was also a slight reduction in the stability profiles of the three triple mutants with c-terminal (tmV225E, tmR227H and tmM228G) with respect to the stability profile of dmT72A at lower protein concentrations (100-1.52 μg/mL) (Figure 2.10). The stability of tmS110N remained unchanged when compared to the stability profile of dmT72A (Figure 2.10).

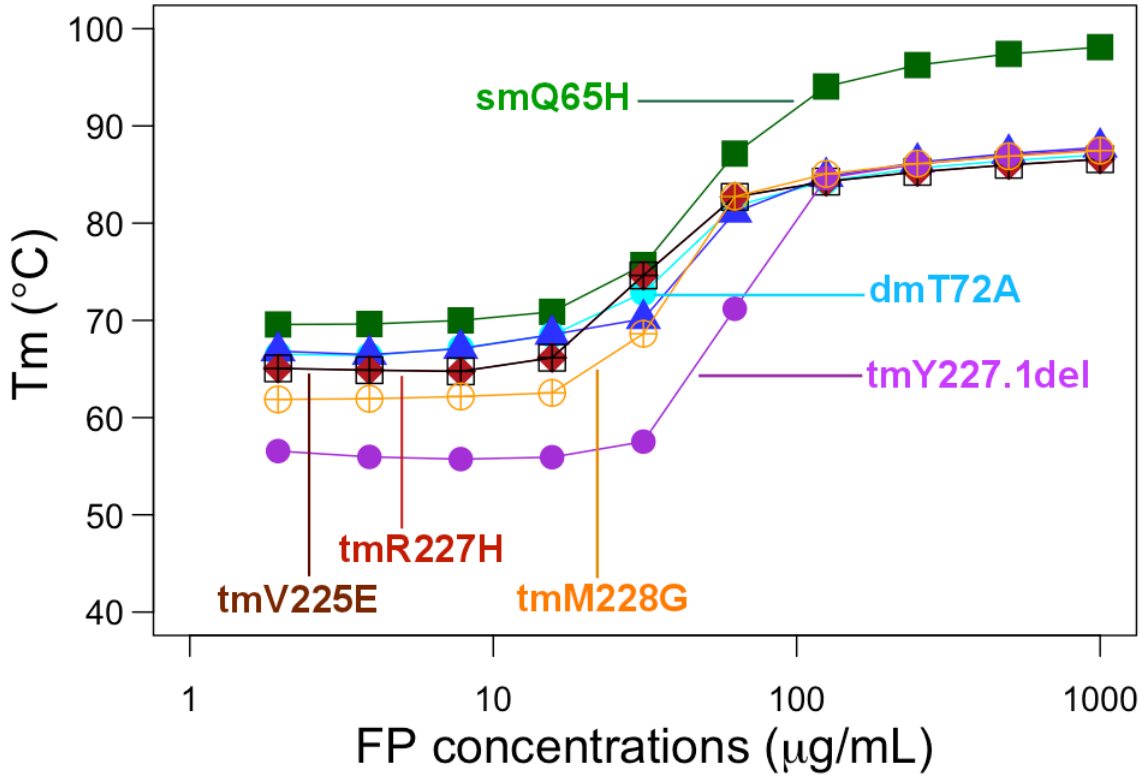


Figure 2.10: Thermal stability profiles of the mutant FPs.
Comparison of thermal stability profiles of the mutant FPs. X-axis: protein concentration; Y-axis: melting temperature (Tm). The protein is destabilized across the entire range of concentrations with introduction of T72A mutation; while, tmY227.1del specifically affects the stability at lower concentrations.

Structural modeling of tmY227.del

Comparison of tmY227.del model with the green ancestor crystal structure (PDB: 4DXI) suggested that the mutations considerably influence the microenvironment of the chromophore (Figure 2.11). The models showed that two ancestral hydrogen bonds between the backbone of the central helix and the backbone of the c-terminal residues, anchoring the central helix, were maintained in the mutants. Amine between positions 72-

73 and carbonyl oxygen between positions 224-225 formed one of the hydrogen bond, and the other hydrogen bond was between the carbonyl oxygen between positions 73-74 and amine between positions 225-226 (Figure 2.11B). Furthermore, the models suggested that the hydrogen bond network between the residues of the central helix might be affected by mutations at positions 72, 225, 227 and deletion at position 227.1 (data not shown). Similarly, the hydrogen bond network between the residues of c-terminal might be changing with mutations at same four positions (data not shown). Interestingly, The conformational arrangements of H65, H203 and E222 appear to be very different in tmY227.1del. These residues are known to play a crucial role in autocatalytic synthesis of the chromophore and influence the resulting spectral properties (Ormo et al., 1996; Barondeau et al., 2003; Sniegowski et al., 2005). Histidine at position 203 in tmY227.1del model has rotated by almost 90°, and the glutamic acid at position 222 in tmY227.1del model has a more extended conformation. From this comparative structural analysis we can speculate that the proximity of H203 and E222 to the chromophore and change in their orientation may have influenced the chromophore catalysis in tmY227.1del towards the evolution of red chromophore (Figure 2.11B).

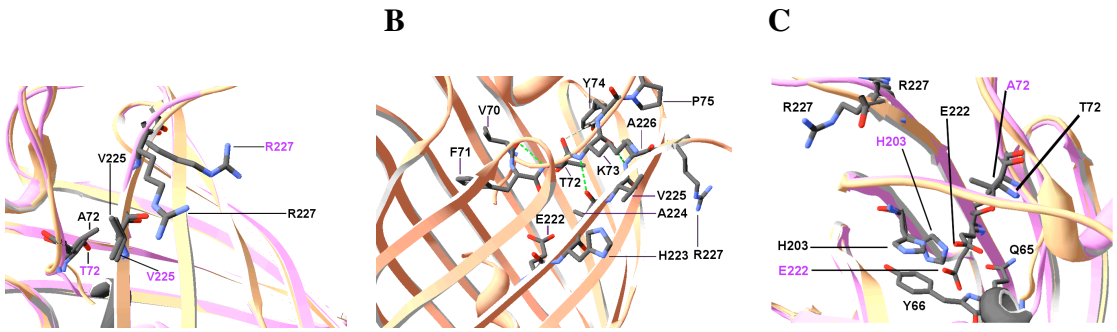


Figure 2.11: Comparison of FP structural models.
Structural comparison indicates changes in orientation of the key residues (Q65H, H203, E222) involved in chromophore synthesis. Structural alignment of the green ancestor (orange) and the tmY227.1del model (purple). The green ancestor residue labels in black; tmY227.1del residue labels in purple. **(A)** Residue alignment at positions 72, 225 and 227 in the green ancestor (PDB: 4DXI) to tmY227.1del. **(B)** Hydrogen bonds (green dashed lines) between the central helix and the c-terminal residues anchor the central helix (PDB: 4DXI). **(C)** Key residues (E222 and H203) involved in chromophore synthesis change orientation in tmY227.1del mutant.

DISCUSSION

Here, we identified the most likely initial evolutionary trajectory leading towards the extant red FP by combining a vertical comparative approach with biophysical techniques. The findings showed that three mutations are sufficient to initiate the autocatalytic synthesis of the tri-cyclic red chromophore from the bi-cyclic ancestral green chromophore. Surprisingly, these mutations also significantly destabilized the protein both by disrupting the interactions between residues within the globule and by reducing the protein's oligomerization tendency. Taken together the evidence suggested these mutations enabled the autocatalysis of a higher complexity chromophore by facilitating the evolution additional reaction step.

Mutations affecting fluorescence color

The mutation T72A was known to red shift the whole emission spectrum of the ancestral green protein by 15nm (Field et al., 2010) and our data confirmed that in the double mutant dmT72A. In addition, a subtle UV-A-dependent increase of the relative proportion of red fluorescence was observed in this mutant (Figure 2.6A), suggesting a possibility that tri-cyclic chromophore might be already forming at a very low efficiency. However, no 578-580 nm emission shoulder was observed in the double mutant (Figure 2.4B), raising an alternate possibility of the repositioning of the chromophore leading to the subtle reddening (Pakhomov et al, 2011; Remington et al., 2005). Such shoulder only appeared with the introduction of the third mutation, Y227.1del, along with a prominent UV-A-dependent accumulation of red fluorescence (Figures 2.4B and 2.6A), confirming that the Y227.1del mutation in addition to Q65H and T72A resulted in generation of detectable amounts of the red-emitting chromophore. In addition, a slight increase in red fluorescence in tmV225E and tmR227H (Figures 2.5B, C and 2.6B, C) raises a possibility that these two mutations could also have been among the early changes in the ancestral FP that furthered the evolution of red fluorescence, possibly through further diminishing stability and oligomeration tendency of the FP.

Other mutations

Despite being highlighted as a positively selected during red fluorescence evolution (Field et al., 2006), mutation M228G appears to have no effect on emission properties in our study (Figures 2.6 and 2.7). However, this mutation is clearly important for red fluorescence in the extant red protein (Field et al., 2006) and brings about a notable increase in red fluorescence when introduced on the background of more historical mutations (Field and Matz 2010). Therefore, mutation M228G likely played a

role later in the evolution of red fluorescence. Mutation S110N (tmS110N) also had no effect on the emergence of red fluorescence; instead it seems to have a deleterious effect by preventing the subtle reddening observed in the dmT72A mutant (Figures 2.6 and 2.7). Similarly to M228G, this mutation strongly contributes to red fluorescence in the later evolutionary stages (Field and Matz 2010), possibly after acquisition of a compensatory mutation (DePristo et al., 2005; Kondrashov et al., 2002; Kulathinal et al., 2004). The dependence of the polarity of S110N effect upon the mutational background raises a possibility of sign epistasis (Field and Matz, 2010; Weinreich et al., 2006; DePristo et al., 2005).

Destabilizing effects of Y227del and T72A

While enabling the detectable generation of red-emitting chromophore, mutations T72A and Y227.1del also have the strongest effect on the protein's stability of all the mutations tested: T72A destabilizes the protein at all concentrations but most prominently at higher concentrations, while Y227.del destabilizes it at lower concentrations (Figure 2.10). This concentration-dependent effect suggested that these mutations might affect the protein's oligomerizaton, which was indeed the case for Y227.del (Figure 2.9A, B). The majority of fluorescent proteins in the GFP-family are known to form a homo-tetrameric structure in which the c-terminal "clasp" between the subunits plays an important role (Chudakov et al., 2010). It is therefore not surprising that the deletion mutation in the c-terminal clasp at position 227.1 reduced oligomerization tendency of the FP (Figure 2.9A), and with the loss of oligomerization the overall stability of the protein also diminishes (Murugan and Hung, 2012; Banks and Gloss, 2003; Hashimoto et al., 2011; Ali and Imperiali, 2005; Liberles et al., 2012). Since the

extant red FP R1-2 is a dimer (Figure 2.9A), these mutations contributed not only to the functional (emission color) evolution, but also to the structural evolution. Mutation T72A does not affect oligomerization and must therefore destabilize the protein by disrupting the interactions within the protein's core (Figure 2.11A, B), which is likely since the 72 side chain is buried within the protein globule. Modeling of the structural effects of the studied mutations (Figure 2.11A, B) supports this idea: the mutations appear to perturb the hydrogen bonding network anchoring the central chromophore-bearing helix to the inner walls of the beta-can, which could have resulted in greater flexibility of the chromophore.

Likely evolutionary trajectory

The most parsimonious evolutionary trajectory towards red fluorescence and associated biochemical complexity involves three mutations, Q65H, T72A, and Y227del, interacting at a long-distance. Notably, assuming that the natural selection was acting to generally increase the fluorescence wavelength (Field et al 2006), one of these three mutations, T72A, could have been positively selected in the absence of others due to the 15-nm red shift of the emission peak. Therefore, only one of the three mutations must have appeared without the change of the protein's fluorescence phenotype, which is plausible even without involvement of natural selection (Kimura) but could potentially involve selection for some unknown aspect of the function of coral FPs (Liberies et al., 2012; Bloom et al., 2007; Zoldovich et al., 2007; Wylie et at., 2011). The most likely candidate for the "neutral start" mutation is Q65H, since it is the only one of the three that does not result a potentially deleterious loss of protein stability and therefore would be likely to persist in the coral population as a neutral mutation. Our results support the

idea that despite the apparent complexity of the novel protein function, there is a way to initiate its evolution by only a few highly probable mutations. Similarly, Finnigan and co-workers demonstrated through the ancestral reconstruction of the ring proteins (Vma3, Vma11 and Vma16) of the hexameric V_0 domain of the eukaryotic V-ATPase proton pump, that three paralogous ring proteins evolved from two ancestral proteins through two highly probable mutations leading to change in the ancestral protein-protein interactions (Finnigan et al., 2012).

Protein destabilization unlocking the evolutionary potential

The interplay between stability (robustness) and evolvability in highly interactive systems such as gene regulatory networks or ecological communities is a focus of active theoretical research (Chen and Ho, 2014; Torres-Sosa et al., 2012; Masel and Siegal, 2009), which has been extended to interactions between residues within an evolving protein (Tokuriki and Tawfik, 2009; Dellus-Gur et al., 2013). FPs could provide a convenient model system to study these issues experimentally. Notably, our results do not seem to agree with the theory that "permissive" mutations enabling the evolution of novel protein functions are likely to be stabilizing (Ortlund et al 2007, Gong and Bloom, 2014; Carter et al., 2014): in our case, two of the three permissive mutations (T72A and Y227del) were the most strongly destabilizing of all the mutations tested (Figure 2.10). It can be hypothesized that this destabilization was necessary to partially release the chromophore-forming residues from spatial constrains and permit alternative autocatalytic reactions. A circumstantial evidence of greater chromophore flexibility is the 15-nm red shift in emission as a result of T72A mutation, indicative of added vibrational states of the chromophore (Lakowicz, 2007). More evidence of the effect of

mutations on chromophore flexibility can be obtained in the future through modeling and experimental studies of protein dynamics (Liberles et al., 2012; Maguid et al., 2006, Pandini et al., 2007). As a consequence, the permissive destabilizing mutations would have broadened the phenotypic space that could be explored through mutation, thereby unlocking the evolutionary potential of the protein.

CONCLUSIONS

Three mutations are sufficient to enable detectable production of a higher complexity red-emitting chromophore in an ancestral coral FP. One of these mutations, although it does not result in detectable red chromophore production, could have been amenable to selection for red-shifted fluorescence since it induces the red shift of the whole emission spectrum. Hence, only a single mutation must be incorporated through mechanisms not related to selection for the new function, making the evolution of red fluorescence highly probable. In contrast to previous studies, in our case the mutations enabling evolution of the novel function strongly destabilize the protein, putatively releasing the chromophore-forming residues from spatial constraints and thereby permitting mutational exploration of alternative autocatalytic pathways. Finally, this is a unique example of the evolution of molecular complexity in which we observe the evolution of a novel chemistry, while observing a decrease in the oligomerization of a protein.

Chapter 3: Diminished oligomerization and destabilizing effect of historical mutations involved in changing the chromophore environment in the evolution of red fluorescence

ABSTRACT

The emergence of red fluorescence in coral GFP-like fluorescent proteins (FPs) represents an increase in functional complexity since it requires three consecutive autocatalytic reactions instead of two. In red FPs, designated as Kaede-type FPs from Faviina corals, 12 historical mutations were necessary and sufficient to evolve red fluorescence from the ancestral green state. An ancestral protein bearing just these twelve mutations, termed the least evolved ancestor (LEA), fully recapitulates the red fluorescence phenotype. In the present study, we examined six mutations that have the most pronounced effect on fluorescence and are located proximate to the chromophore that facilitated the change in the chromophore cavity and environment. . By means of reverse mutagenesis of the LEA, we demonstrate that they result in a diminished thermostability and reduced oligomerization tendency of the protein. These results were surprising given that the majority of these mutations are buried inside the protein and do not participate in the tetrameric interfaces. These data support the assertion that the evolution of biochemical complexity in the Kaede-type FP was enabled by perturbing the existing stable interactions at multiple levels of biological organization: in this case, at the tertiary and quaternary protein structure levels.

INTRODUCTION

How does molecular complexity; manifested at the protein level as an orchestrated system of interactions within and between proteins, evolve through small

and incremental changes? Specifically, how can sequence evolution lead to novel functionality over the course of protein evolution? The field of experimental molecular evolution attempts to recapitulate these processes through resurrection of ancestral molecules, retracing of possible evolutionary trajectories, and characterization of evolutionary intermediates (Thornton, 2004; Harms and Thornton, 2010; Kacar and Gaucher, 2012).

Comparative analysis of extant proteins has indicated that novel functions and the associated molecular complexity do indeed arise through small and incremental changes (Finnigan et al., 2012; Harms and Thornton, 2010, Datta, 2009; Maita et al., 2010). However, these studies lacked the capacity to experimentally test these predictions and could not reveal the underlying biochemical and biophysical changes. Physically resurrecting and characterizing the ancient proteins solves these problems (Harms and Thornton, 2010, Kacar and Gaucher, 2012; Studer et al, 2013; Tokuriki and Tawfic, 2009). Importantly, vertical comparative approaches - utilizing the resurrected molecular complexes and their extant counterparts - enable the study of the evolution of structure-function relationships and the underlying residue interactions (Harms and Thornton 2010, Yokoyama et al., 2008).

The Green Fluorescent Protein (GFP) family provides an ideal opportunity to probe the evolution of biochemical and biophysical properties and to study the evolution of molecular complexity. Diverse marine taxa (Cnidaria, Arthropoda, Chordata) express fluorescent proteins (FPs) with diverse photophysical properties enabled by varied autocatalytic pathways to synthesize different chromophores (Chudakov et al., 2010; Pletnev et al., 2013). The evolutionary origins of different chromophores remain largely unknown. Of the three red fluorescent chromophores, the tri-cyclic green-to-red (G/R) photo-switchable (upon exposure to UV-A) chromophore of the Kaede-type FPs from the

great star coral, *Montastrea cavernosa*, was shown to arise from a bi-cyclic chromophore of a green ancestral protein (Figure 3.1; Ugalde et al., 2004, Kim et al., 2013; Field and Matz, 2010). This work also made a case that the Kaede-type FPs represents an increase in molecular complexity since it implies the evolution of an additional autocatalysis step within the same protein. Of the 37 historical mutational changes detected between the green ancestor and the least divergent Kaede-type extant FP (a coral protein called R1-2), only 12 were necessary to fully recapitulate the red fluorescence phenotype (Field and Matz, 2010). The green ancestor with these 12 historical mutations was termed Least Evolved Ancestor (LEA; Figure 3.1; Field and Matz, 2010). A follow-up structural study involving the green ancestor, LEA and their mutants suggested that a subset of 12 historical mutations in LEA resulted in enlargement of the cavity surrounding the chromophore potentially enabling the transition to red fluorescent chromophore structure (Figure 3.2A, B; Kim et al., 2013).

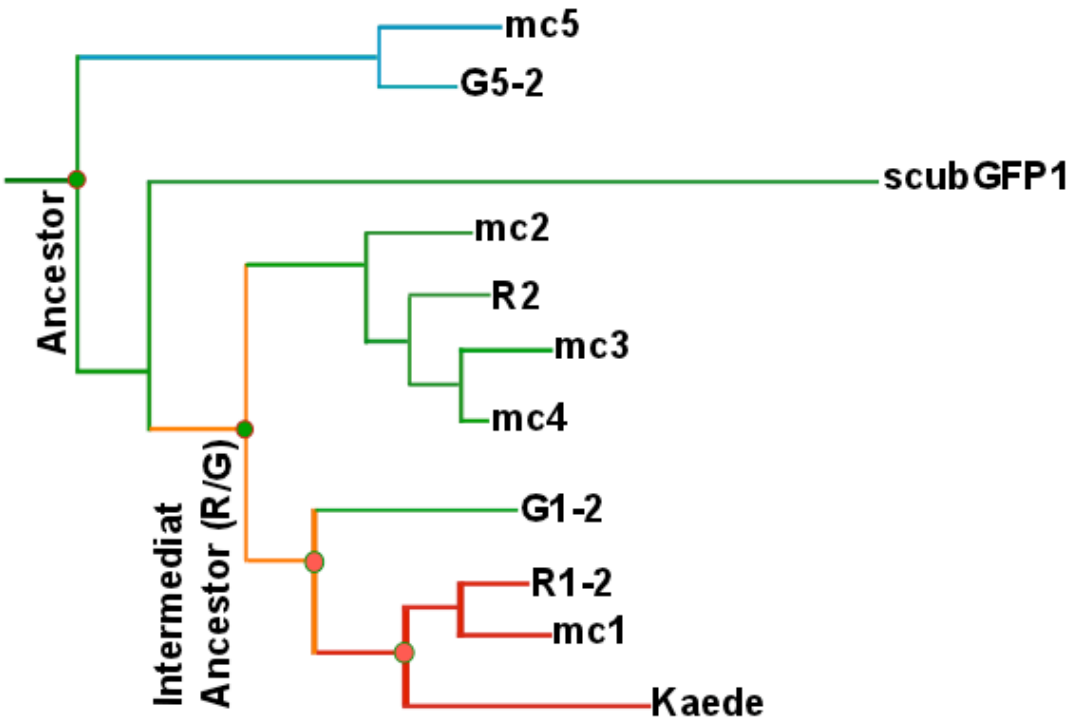


Figure 3.1: A diagram representing the relationships between Faviina coral FPs. Labeled branch tips represent the extant FPs. The green ancestor and intermediate ancestor (R/G) are two labeled nodes. Only 12 of 37 total mutational changes between the green ancestor and R1-2 were used to generate Least Evolved ancestor (LEA) with recapitulated red fluorescence phenotype of R1-2; therefore, LEA is shown at the halfway point on R1-2 branch (Ugalde et al., 2004; Field and Matz, 2010).

We have also subsequently discovered that only three mutations out of the 12 were sufficient to initiate the selection-driven evolution of red fluorescence from the green ancestor (Chapter 2: Modi et al., 2014). Notably, these three mutations strongly perturbed the stability and oligomeric tendency of the ancestral protein, which could have contributed to a lower mutational robustness of the ancestral green fluorescence phenotype and thereby broadened the phenotypic space accessible through random

mutations (Chapter 2: Modi et al., 2014). However, how the historical mutations located near the chromophore influenced the biochemical and biophysical properties of the evolving FP remained largely unexplored (Figure 3.2A, B).

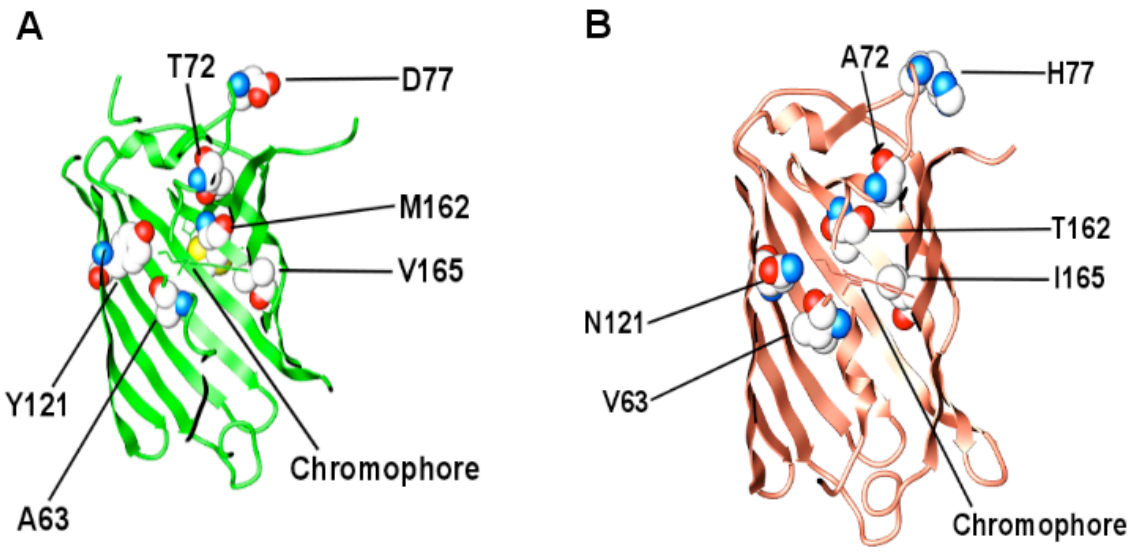


Figure 3.2: Six targeted residues proximate to the chromophore.
A study of the importance of six historical mutations in the evolution of structure-function relationship of FPs determined based on their impact on biophysical properties of the FP over the course of evolution of red fluorescence. **(A)** A structure of the green ancestor (All-New-GFP; PDB: 4DXI) showing six ancestral amino acids proximate to the chromophore. **(B)** A LEA structure (PDB: 4DXN) showing six historical mutations at same residue positions as in panel-A.

Based on these earlier observations, we hypothesized that the near-chromophore mutations did not only modify the immediate chromophore environment but also led to structural rearrangements on the scale of the whole LEA globule. The aim of the present study was two fold: (i) to assess the stability and oligomerization tendency of LEA relative to the ancestral, evolutionary intermediate and extant FPs, (ii) investigate how

stability and oligomerization are affected by near-chromophore mutations through the vertical comparative approach (Figure 3.2A, B). Two of the mutations chosen for this study (T72A and Y121N) have been shown to exert a large effect on red fluorescence, whereas the other four (A63V, D77H, M162T and V165I) are of moderate to small effect (Field and Matz, 2010).

MATERIALS AND METHODS

Reverse Mutagenesis

All residues are numbered based on the sequence of the GFP from *Aequorea victoria* (Prasher et al, 1992). The six historical mutations in LEA were reverse mutagenized using the site directed mutagenesis kit, QuikChange II (Agilent, Santa Clara, CA). The targeted six historical mutations were mutated back to the ancestral state either individually or in combination. Three LEA revertant FPs were generated: LEA-X72 (A72T), LEA-X121 (N121Y) and LEA-X6 (V63A, H77D, T162M and I165V). Similarly, a mutant FP was generated through forward mutagenesis by introducing a single mutation Q65H (smQ65H) in the green ancestor. The detailed methods on the site-directed mutagenesis are described in the chapter 2 of this dissertation.

Protein expression and purification

Purified green ancestor (All-New-GFP), smQ65H, LEA and LEA revertant mutants (LEA-X6, LEA-X72 and LEA-X121) were expressed in E. coli and purified by His-tag metal affinity chromatography in the absence of light as described in chapter 2 of this dissertation. Purified FP absorbance at 280nm was used to determine concentrations. Purified proteins were flash frozen in liquid nitrogen for long-term storage.

The intermediate ancestor (R/G) and extant FP (R1-2) were cloned into pGEM-T vector (Amp_r) (Invitrogen, CA) by means of TA-cloning. Cloned FPs were transformed into chemically competent One Shot Top-1 *Escherichia coli* cells (Invitrogen, CA). Cells were thawed on ice for 10 minutes. Approximately 40ng of plasmid containing the gene of interest was gently mixed with cells and incubated on ice for ~20-30 minutes. The chemically competent cells were transformed by heat shocking the cells for 45 seconds at 42ºC. Next, transformed cells were allowed to recover by incubating on ice for ~2-5 minutes. After adding 400µL of pre-warmed SOC media (Invitrogen, CA) cells were incubated at 37ºC for 45 minutes (on a shaker at 200RPM). Transformed cells were selected by plating on Luria-Bertani (LB) agar containing 100 μg/mL ampicillin and 0.5mM Isopropyl β-D-1-thiogalactopyranoside (IPTG). Plates were incubated over night at 37ºC.

Transformed clones were screened for FP expression using a Leica MZ FIII stereomicroscope (Leica, St Gallen, Switzerland) with the filter set BL/VIO (# 11003, Chroma Technologies Corp., Bellows Falls, VT). Fluorescent colonies were picked and grown over night in 2mL of LB broth with 50 μg/mL ampicillin. Cultures were used to isolate plasmid DNA using a mini-prep kit (Qiagen, Natherlands). Isolated plasmids were used to sequence both genes by means of Sanger sequencing using a T7 primer targeting T7-promoter site in the pGEM-T vector. Sequencing results analysis and alignment manipulations were done using BioEdit (v.7.0.9), CLC Sequence Viewer 6 and eBioX (v. 1.5.1).

The detailed methods for the FP expression and purification of his-tagged FPs using metal affinity chromatography are described in the chapter 2 of this dissertation. All FPs were shielded from ambient light throughout the entire expression and purification processes.

Non-denaturing SDS-PAGE

Four purified FPs (the green ancestor, intermediate ancestor (R/G), LEA and R1-2) were analyzed by non-denaturing SDS-PAGE. Two of the FPs, LEA and R1-2, were analyzed in both unconverted (green) and converted (red) forms. LEA and R1-2 were photo-converted by exposing the FPs (on ice) to UV-A light (400nm) for 90 minutes. All FPs were diluted in 1x HEPES buffer to 250 μg/mL. Diluted FPs were mixed with 2x loading buffer w/ βME in 1:1 ratio and loaded (30uL) on to 4-20% PAGE gels (Bio-Rad, CA). Samples were run in 1x SDS-running buffer ((Bio-Rad, CA) at 120v. The four FPs studied here are known to fluoresce when excited by UV-B; therefore, the FPs were detected in gels using a UV box.

Differential scanning fluorimetry

Thermal stability of the FPs was assessed by differential scanning fluorimetry (DSF) using the LightCycler 480 System (Roche Diagnostics Corporation, Indianapolis, IN) as described in detail in chapter 2 of this dissertation. FPs were kept in the dark throughout the procedure unless otherwise specified.

Thermal stability of the FPs was studied employing two dilution series: (i) 1 mg/mL through 1.56 μg/mL and (ii) 150 μg/mL through 1.56 μg/mL. To determine the thermal stability profiles of the green ancestor, smQ65H, LEA, LEA revertant mutants, intermediate ancestor (R/G) and R1-2, we used fluorescence data for each FP at a given concentration to determine the temperature (Tm) at which it lost 50% of its fluorescence.

To determine the effect of photoconversion on the stability of the FPs, we investigated the stability of unconverted (green form) and converted (red form) of LEA, intermediate ancestor (R/G) and R1-2. Due to the oligomeric nature and the incomplete green-to-red conversion of LEA and R1-2, it was possible to excite the converted (red)

form of these FPs at 483nm, and through Förster Resonance Energy Transfer (FRET) measure the red emission at 568nm of the converted (red) form of these two FPs. Therefore, by exciting unconverted (green) and converted (red) forms of the intermediate ancestor (R/G), LEA and R1-2 at 483nm and by detecting the green emission (533nm) and red emissions (568nm) of both forms of these FPs, we are able assess the effect of photo-conversion on the thermal stability of these FPs. Furthermore, by detecting the emissions at two wavelengths (533nm and 568nm) for both forms of LEA and R1-2, we were able to control for any bleed over of green fluorescence into red fluorescence, by looking for any inconsistent changes in thermal stability across both detection modes (at 533nm and 568nm) of both forms of the intermediate ancestor (R/G), LEA and R1-2.

Two detection formats were used for the assays: (i) Multi Color Hydrolysis Probe (the FAM filter combination: excitation-483nm and emission-533nm), (ii) customized filter set-I (ex. 483nm and em. 568nm). Both detection formats had 0.2sec integration time, with the melt factor set to 1 and the quant factor set to 10.

Multi-angle Laser Light Scattering

All FPs were thawed on ice and then mixed by quick vortexing. FPs were then centrifuged for 1min and diluted using a pre-chilled HEPES buffer at three concentrations (5 mg/mL, 500 μg/mL and 250 μg/mL). All FP manipulations were done in the dark unless otherwise specified. We calculated the theoretical molar masses of the green ancestor, intermediate ancestor (R/G), LEA and R1-2 using Compute pI/Mw tool on ExPASy Bioinformatics Resource Portal (Bjellqvist et al., 1993; Bjellqvist et al., 1994; Gasteiger et al., 2005). We determined the oligomeric states of the FPs by first separating the oligomers by HPLC size-exclusion chromatography (SEC), and then determining

their molar mass with multi-angle laser light scattering (MALLS). Detailed methods for determining the molar mass and corresponding oligomeric state assignment of the FPs by SEC-MALLS and Astra 4.9 software (Wyatt Technology, Santa Barbara, CA) are described in chapter 2 of this dissertation. The cumulative weight fraction analysis to determine the relative contribution of each molar mass in a sample was done using Astra 4.9 software.

The static light scattering results presented here are in terms of Rayleigh ratio as a function of elution volume for each FP. The estimated molar masses reported here for each peak were calculated (in Astra 4.9) by taking a weighted-average value of all molecular species in a peak. The light scattering signal at the edges of a peak or at a low concentration of molecular species produces a low signal to noise ratio (S/N ratio), which decreases the accuracy of the calculated molar mass. Generally, in order to reduce the influence of molecular species on the edges of closely eluting peaks, we calculated molar masses for the majority of predominant peaks by using the trailing half of the peak. The detailed description of a method assessing the accuracy of molar masses determined by MALLS is described in chapter 2 of this dissertation. Lastly, we determined the percent contribution of each molecular species in each FP sample based on the cumulative weight fraction analysis (Table 3.1).

RESULTS

Differential band migration patterns of LEA and R1-2 on non-denaturing SDS-PAGE

The results of the non-denaturing SDS-PAGE analysis showed a varied gel migration pattern for all four FPs, the green ancestor, intermediate ancestor (R/G), LEA

and R1-2. The green ancestral band was observed near the top of the gel, while the extant FP, R1-2, band was observed near the bottom of the gel (Figure 3.3A). The migration pattern differences between the tetrameric ancestral FP and dimer R1-2 were expected based on their distinct oligomeric states. The intermediate ancestor (R/G) band migrated slightly below the green ancestral (Figure 3.3A). Both unconverted (green) and converted (red) forms of LEA bands migrated same distance slightly above the converted (red) form of R1-2. Surprisingly, the band migration patterns of both forms of LEA differed from that of R1-2 (converted red form), even though both FPs have same emission profile (Figure 3.3B) (Field and Matz, 2010). We also observed two faint bands in unconverted (green) and converted (red) R1-2 close to the major R/G and LEA bands (Figure 3.3). These bands likely represent a small fraction of higher-order oligomers we observed in MALLS study of R1-2 (Figure 3.6). Lastly, bands for the green (unconverted) form of R1-2 were very faint; loss of FP stability in presence of SDS could be a possible explanation for the loss of fluorescence.

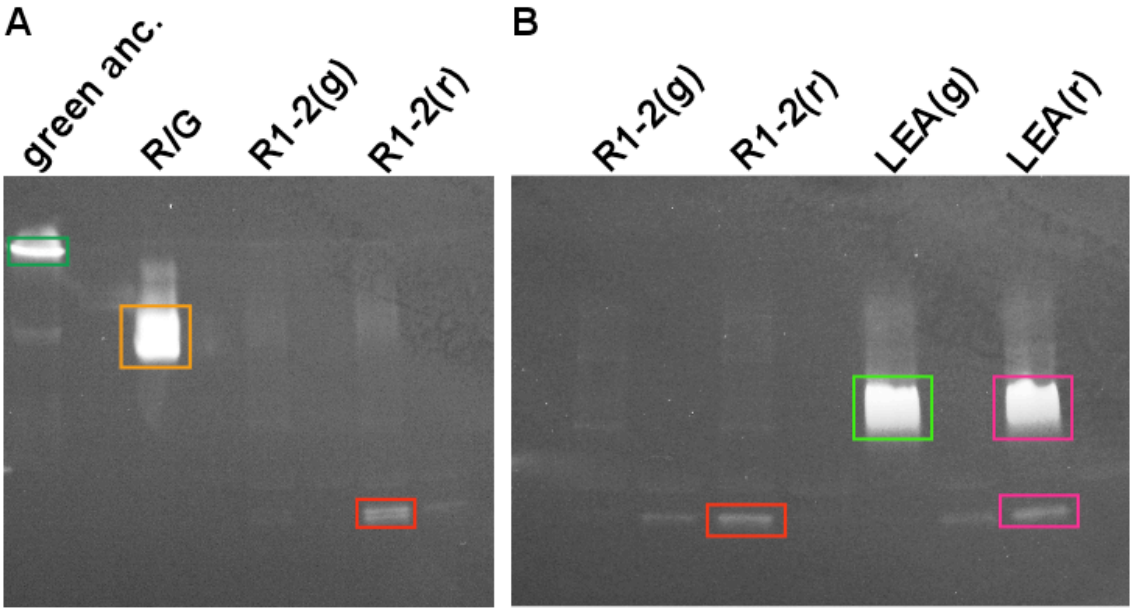


Figure 3.3: Non-denaturing SDS-PAGE of the ancestral FPs and extant FP. Native green ancestor, intermediate ancestor (R/G), extant FP (unconverted green form: R1-2(g), converted red form: R1-2(r)) and Least Evolved Ancestor (unconverted green form: LEA(g), converted red form: LEA(r); detection using the proteins' native fluorescence excited by UV light. **(A)** Band migration patterns showing the least migrated the green ancestor (green box), intermediate migration of R/G (orange box) and the farthest migration of converted R1-2(r) (red box). Bands for unconverted R1-2(g) are very faint. **(B)** Band migration patterns of unconverted (green) and converted (red) forms of R1-2 and LEA. Majority of converted R1-2(r) migrated towards the bottom of a gel. Unconverted LEA(g) and converted LEA(r) exhibited the same migration pattern, remaining in the middle of gel. The bottom band in converted LEA(r) could be the dissociated FP. Bands for unconverted R1-2(g) are very faint. Note: Two faint bands in unconverted (green) and converted (red) R1-2 corresponding to major R/G and LEA bands likely represents a small fraction of higher-order oligomers.

Relative stability and oligomeric tendency of LEA

The thermal stability profiles of three unconverted FPs (LEA, intermediate ancestor and R1-2) were determined at concentration ranging from 150 μg/mL to 1.56

μg/mL using DSF. The green ancestor thermal stability profile was used as a control. The green ancestor had the highest stability, the intermediate ancestor (R/G) and the extant FP (R1-2) were intermediately stable, and LEA had the lowest thermal stability of all four FPs (Figure 3.4).

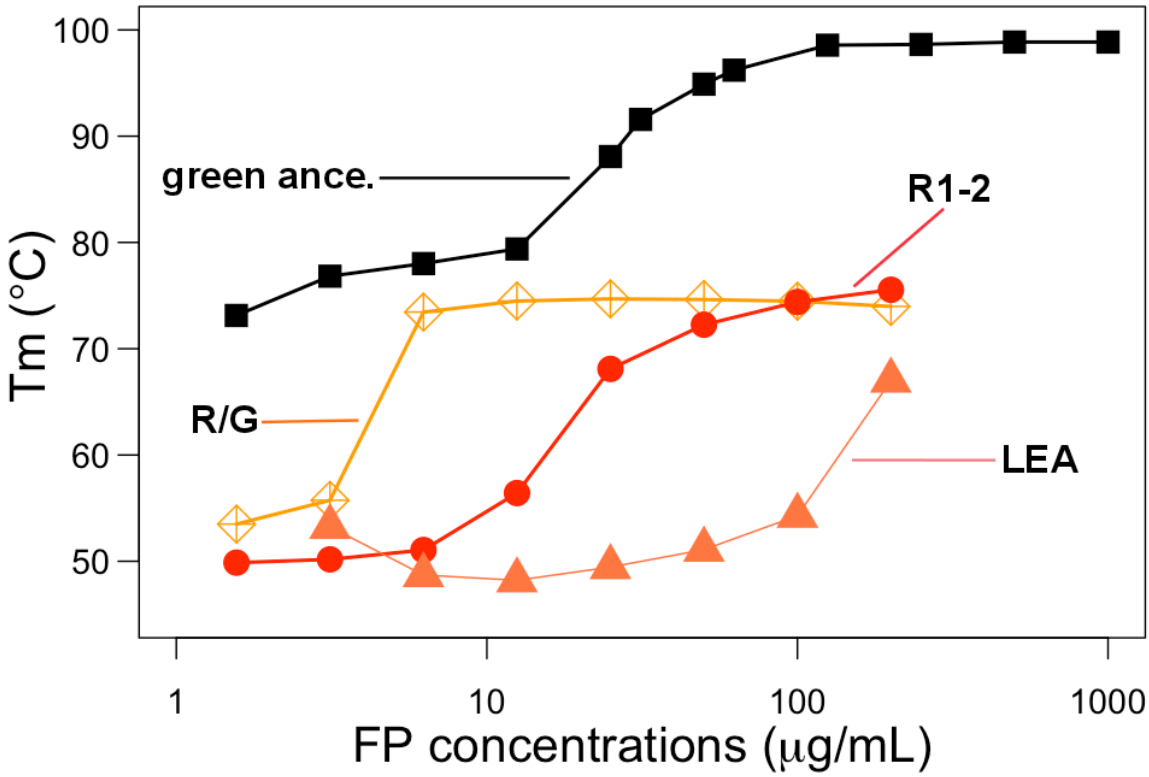


Figure 3.4: Thermal stability profiles of ancestral FPs and extant FP by differential scanning fluorimetry.
Comparison of thermal stability profiles of three ancestral FPs to each other and to the extant FP. X-axis: protein concentrations; Y-axis: melting temperature (Tm), defined as a temperature at which the protein loses 50% of its fluorescence amplitude. Thermal stability profiles of the converted intermediate ancestor, R1-2 and LEA were determined with fluorescence emission at 568nm. Thermal stability profile of the green ancestor was determined with fluorescence emission at 533nm.

Next, the stability profiles of unconverted (green) and converted (red) forms of LEA and R1-2 were determined. The unconverted (green) form of LEA had higher stability than the converted (red) form (Figure 3.5A). The LEA stability differences between the unconverted (green) and converted (red) forms were observable with either of the fluorescence emission detection modes, at 533nm and 568nm (Figure 3.5A). On the other hand, R1-2 stability profile remained unchanged between unconverted (green) and converted (red) forms regardless of the fluorescence emission detection modes (Figure 3.5B). The differences in intensities of the green and red fluorescence emissions of each FP were the most likely explanation for the dissimilarity in stability profiles of each FPs between the two different detection modes (at 533nm and 568nm) (Figure 3.5A, B).

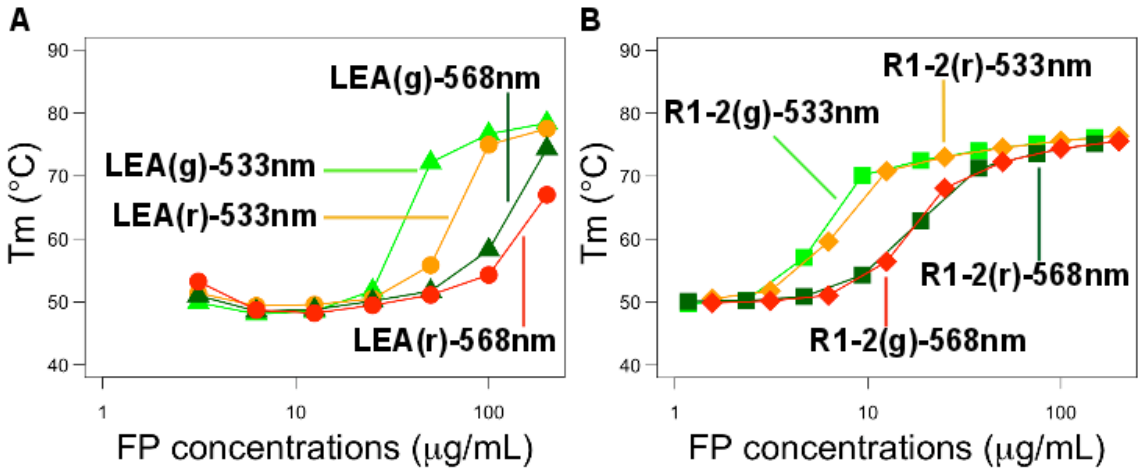


Figure 3.5: Thermal stability profile comparison of the unconverted (green) and converted (red) forms of R1-2 and LEA by differential scanning fluorimetry.
Thermal stability profiles of the green and red forms of R1-2 and LEA using two detection modes by measuring fluorescence emissions at 533nm and 568nm. X-axis: protein concentrations; Y-axis: melting temperature. **(A)** Regardless of the detection mode, unconverted LEA(g) is more stable than the converted LEA(r). **(B)** Unconverted R1-2(g) and converted R1-2(r) have the same stability profiles in spite of the detection mode.

A dramatic difference in the stability profiles of the green ancestor, intermediate ancestor (R/G), LEA and R1-2 prompted us to evaluate oligomeric states of these FPs (unconverted) through SEC-MALLS. Calculated molar mass of the green ancestor (26725.3 kDa), intermediate ancestor (26463.1 kDa), LEA (26412 kDa) and R1-2 (26875.7 kDa) were used in determining the oligomeric states of all four FPs. The results showed the green ancestor (1.21×10^5 g/mol), intermediate ancestor (1.179×10^5 g/mol) and LEA (9.294×10^5 g/mol) to be the tetramers, while R1-2 was a dimer (4.867×10^4 g/mol) (Figure 3.6). The data also highlighted that the elution profile of the predominant peak representing the tetramer of the green ancestor differed from that of the intermediate ancestor and LEA (Figure 3.6). The elution profiles of the predominant peak representing

the tetramer of the intermediate ancestor and LEA were similar (Figure 3.6). A change in hydrodynamics radius of the FPs could be the most likely explanation for the difference observed in the elution profiles of these peaks. A small amount of higher-order oligomers were observed in all four FPs, possibly due to the non-specific protein aggregation (Figure 3.6). There were no changes in the oligomeric state of LEA and R1-2 upon photo-conversion. Both unconverted (green) and converted (red) forms of LEA were tetramers, while the both forms of R1-2 were dimers (data not shown).

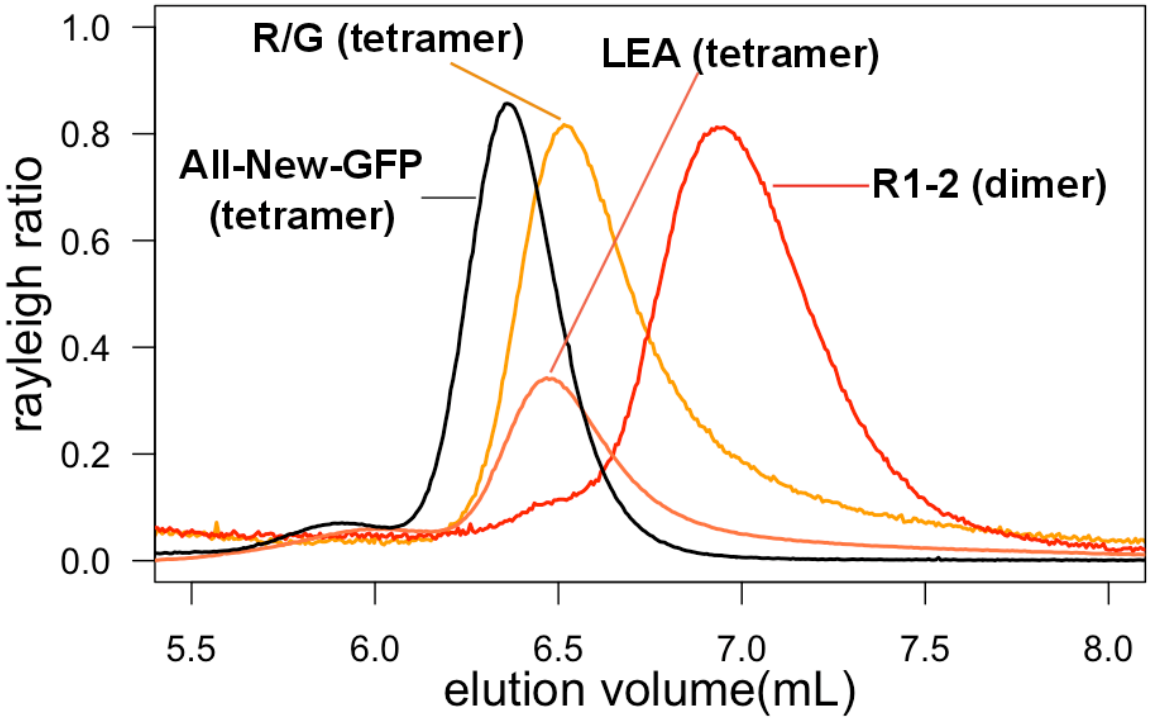


Figure 3.6: Oligomeric state determinations of the ancestral FPs and extant FP by SEC-MALLS.
Oligomeric states separation by size-exclusion chromatography and the molar mass detection by multi-angle laser light scattering (SEC-MALLS). The green ancestor, intermediate ancestor (R/G) and Least Evolved ancestor (LEA) are tetramers, while the control, extant FP (R1-2), is a dimer. Oligomeric states of the all FPs are inferred based the MALLS calculated molar mass. X-axis: elution volume (SEC), Y-axis: Rayleigh ratio (MALLS).

Contribution of the six historical mutations toward diminished stability of LEA and R1-2 FPs

In order to elucidate the relative contribution of mutations to the diminished stability of the LEA, we examined a set of six historical mutations thought to influence the chromophore environment by assaying the unconverted (green) form of three LEA revertant FPs (LEA-X6, LEA-X72 and LEA-X121) using DSF (Leung et al., 1996;

Senisterra and Finerty, 2008; Kiss et al., 2009; Dai et al., 2007). A reversion of mutation at position 72 (A72T) to the ancestral state increased the stability of the mutant LEA-X72. This single point mutation largely restored the stability profile of LEA-X72 back to that of the green ancestor (Figure 3.7). The reversion of five remaining mutations at positions 121, 63, 77, 162 and 165 to the ancestral state also increased the stability of other two LEA mutants, LEA-X6 and LEA-X121. However, the data showed that the contributions of these five mutations to FP destabilization were small compared to the mutation at position 72 (Figure 3.7). In general the data showed that all six historical mutations studied here had destabilizing effects on the FP in the course of evolution of red fluorescence.

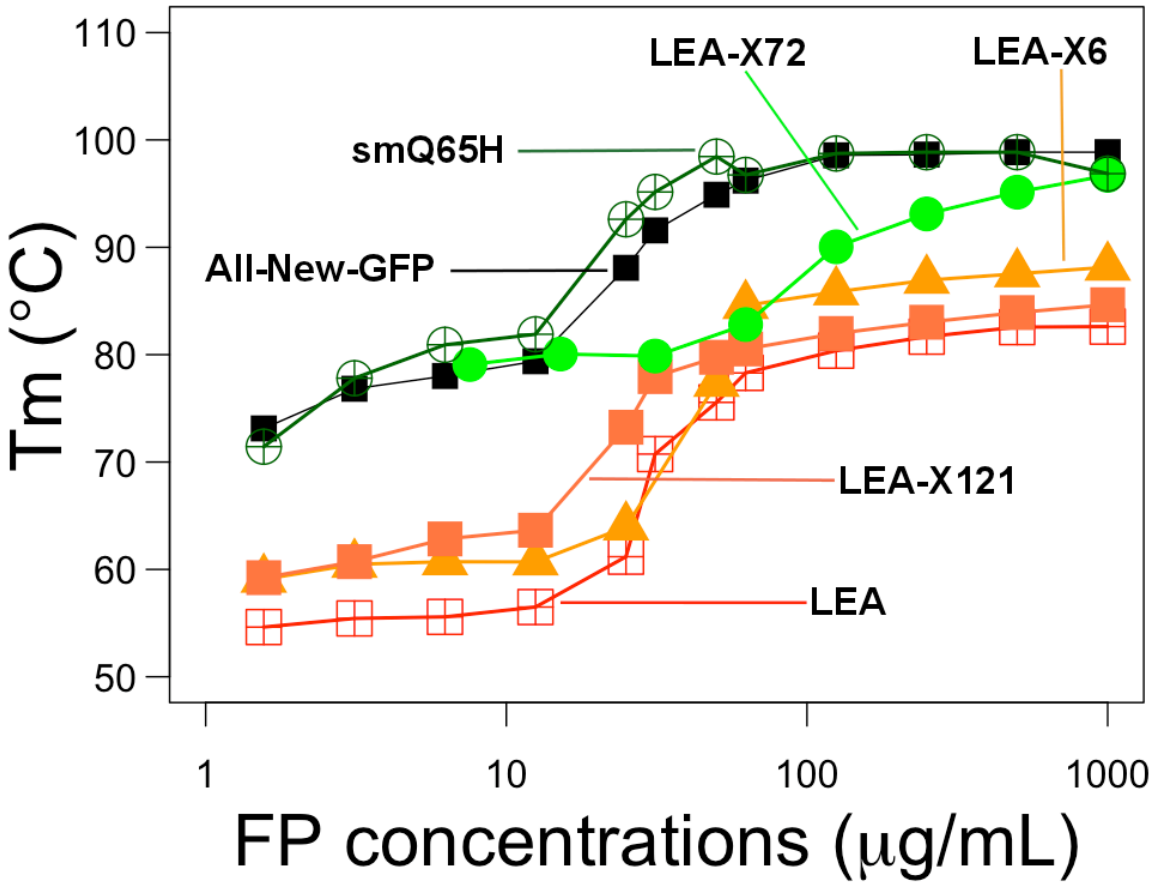


Figure 3.7: Thermal stability profile of unconverted (green) form of LEA and LEA revertant mutants by differential scanning fluorimetry.
Comparison of thermal stability profiles of revertant mutants to LEA and the green ancestor by monitoring fluorescence emission at 533nm as they are heated. X-axis: protein concentrations; Y-axis: melting temperature (Tm), The largest increase in thermal stability profile is observed with reverse mutagenesis at position 72 (A72A) in LEA-X72. The green ancestor and smQ65H demonstrate the highest thermal stability.

Contribution of historical mutations to the diminished oligomerization tendency

By examining all FPs at three different concentrations (5 mg/mL, 500 μg/mL and 250 μg/mL), the SEC-MALLS results highlighted the impacts of six historical mutations on the oligomeric states of the FP as a function of concentration. Table 3.1 summarizes

the results of the SEC-light scattering study. We determined the oligomeric states corresponding to each peak based on the calculated molar mass associated with that peak. The largest peak represents the predominant oligomeric form.

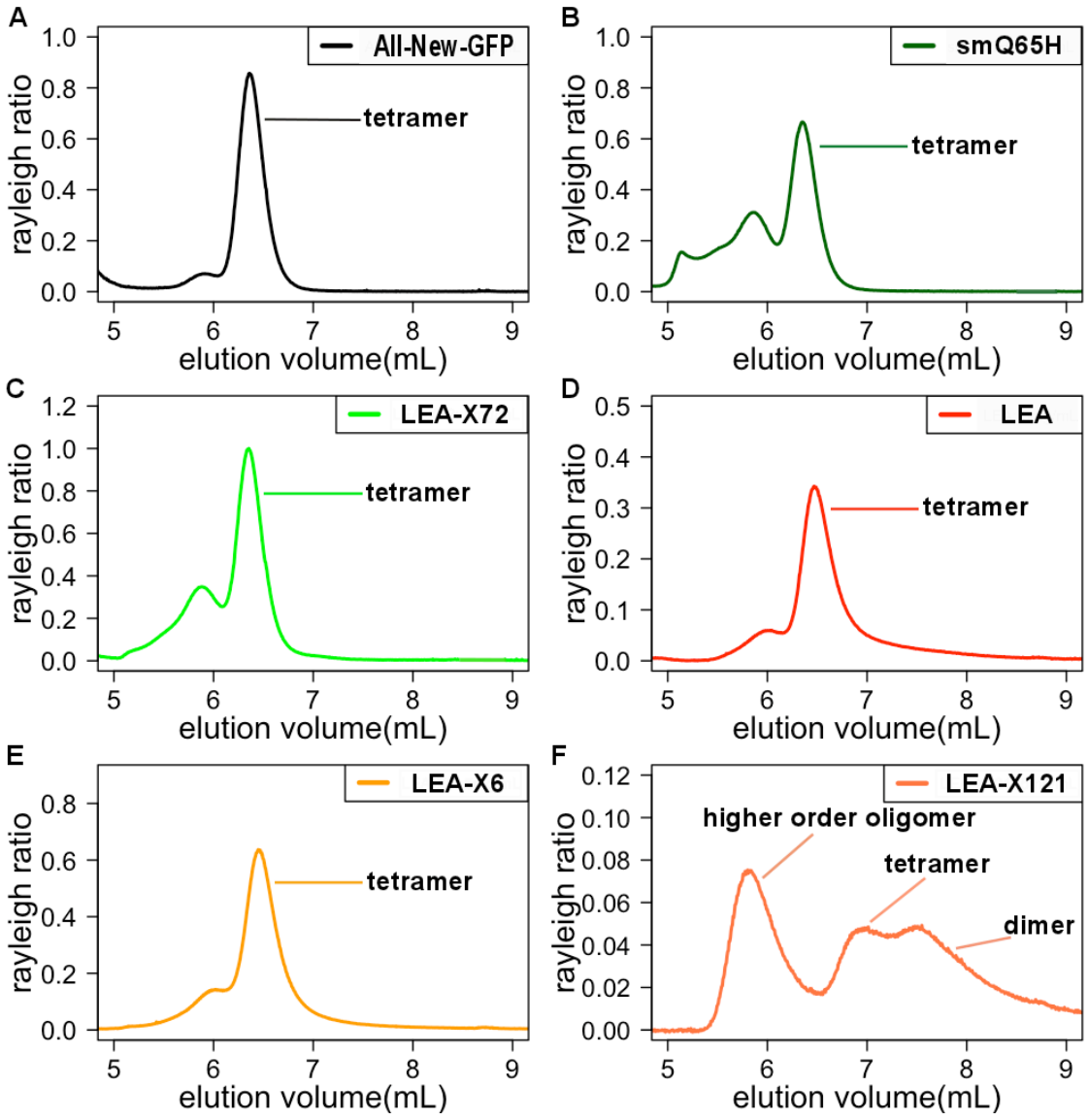


Figure 3.8: Impact of revertant mutations on the oligomeric state of the FP. Reversion to the ancestral state at position 121 in LEA-X121 (N121N) leads dissociation of tetramer into a dimer. Oligomeric states separation of the FPs at 5 mg/mL by size-exclusion chromatography and the molar mass detection by multi-angle laser light scattering (SEC-MALLS). Oligomeric states of the green ancestor (All-New-GFP), smQ65H, LEA and LEA revertant mutants (LEA-X6, LEA-X72 and LEA-X121) inferred based the MALLS calculated molar mass. X-axis: elution volume (SEC), Y-axis: Rayleigh ratio (MALLS). **(A)** The green ancestor, **(B)** smQ65H, **(C)** LEA-X72, **(D)** LEA, **(E)** LEA-X6: all five FPs are predominantly tetrameric. **(F)** LEA-X121: the FP has three oligomeric states: higher order oligomers, tetramer and dimer.

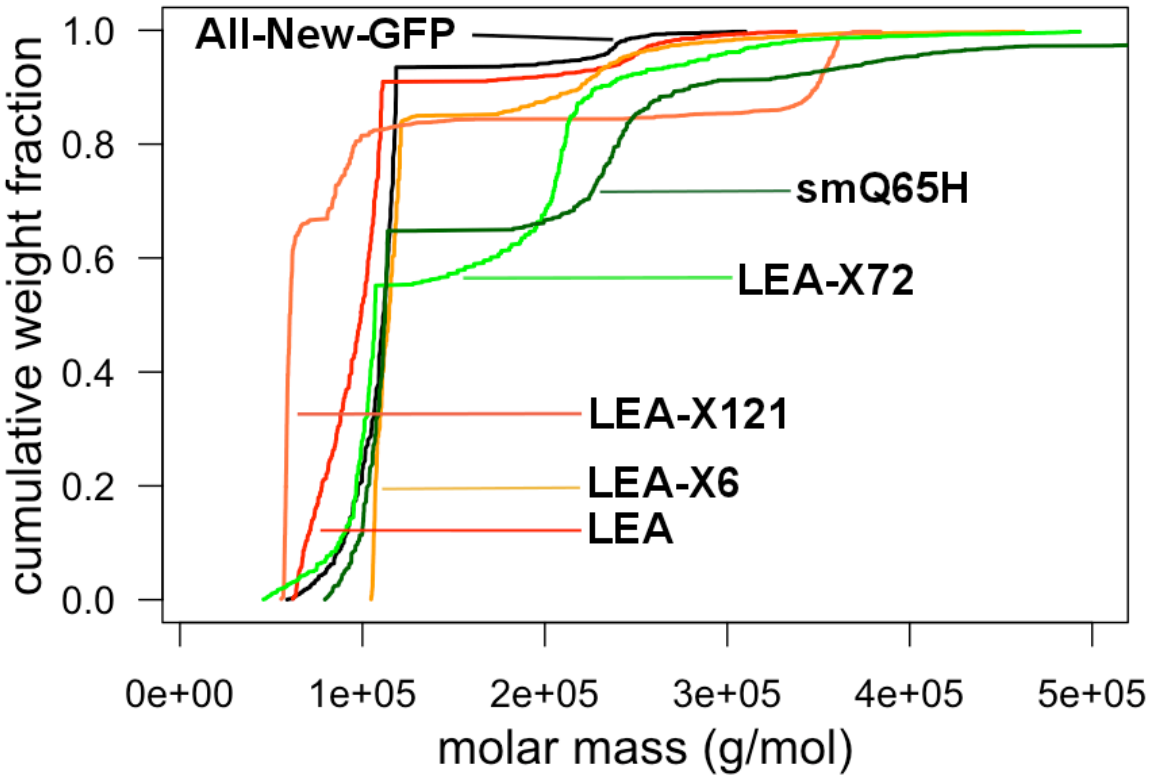


Figure 3.9: Oligomeric state stiochiometry analysis by the cumulative weight fraction.
Stiochiometry of between oligo forms in each of the proteins elucidated by the cumulative weight fraction analysis using the MALLS data. X-axis: molar mass (g/mol), Y-axis: cumulative sample fractions. LEA-X121 is predominantly a dimer, with two small fractions consisting of tetramer and higher order oligomer. The predominant LEA fraction is a tetramer, which also contains some dissociated FP. The green ancestor (All-New-GFP), smQ65H, LEA-X72 and LEA-X6 are tetramers, with smaller fractions representing the higher order oligomers.

At 5 mg/mL, the predominant peak represented the tetramer in all FPs, except LEA-X121. Regardless of the protein concentration the tetramer was the most predominant oligomeric state in the green ancestor and smQ65H (Figures 3.8A-B, 3.10A-B, 3.11A-B). There also exist small amounts of higher-order oligomers at all three

concentrations of both of these FPs (Figures 3.8A-B, 3.9, 3.10A-B, 3.11A-B). The results also showed that LEA is a tetramer at a higher concentration (5 mg/mL). However, we observed a slightly lower than expected molar mass of LEA (9.294 x 10^4 g/mole), suggesting the presence of lower molar mass oligomers in the main peak. This observation was supported by the cumulative weight fraction data showing presence of small amount of dissociated FP in the main peak (Figures 3.8D and 3.9). The observed dissociated oligomeric states were confirmed with the 500 μg/mL and 250 μg/mL runs, which clearly showed the presence of the lower order oligomers, 7.431 x 10^4 g/mole and 4.980 x 10^4 g/mole (Figures 3.10D, 3.11D). We observed the LEA revertant mutant, LEA-X72 to be tetrameric at all three concentrations; however, a small fraction of LEA-X72 dissociated into a dimer at 500 μg/mL (5.027 x 10^4 g/mole) and at 250 μg/mL (4.441 x 10^4 g/mole) (Figures 3.8C, 3.9, 3.10C, 3.11C).

Another revertant mutant, LEA-X121 has multiple oligomeric states at the highest protein concentration (5 mg/mL), these peaks corresponded to a higher-order oligomer (3.469 x 10^5 g/mole), a tetramer (9.897 x 10^4 g/mole) and a dimer (5.934 x 10^4 g/mole) (Figures 3.8F, 3.9). However at 500 μg/mL-protein concentration, we were unable to detect a dimer peak, and at 250 μg/mL-protein concentration the tetramer peak also disappeared. At 250 μg/mL-protein concentration, we were only able to detect a higher-order oligomer (8.053 x 10^5 g/mol; 23%). The absence of tetramer or dimer at two lower protein concentrations could be explained by complete dissociation and denaturing of the FP (Figures 3.10F, 3.11F). The last of the three revertant LEA mutants, LEA-X6, was observed predominantly as a tetramer at 5 mg/mL (1.128 x 10^5 g/mole) with two higher-order oligomers (4.837 x 10^5 g/mole and 2.359 x 10^5 g/mole) (Figures 3.8E, 3.9). The dissociated oligomeric states of LEA-X6 were not detected at either of the three protein concentrations (Figures 3.10E, 3.11E).

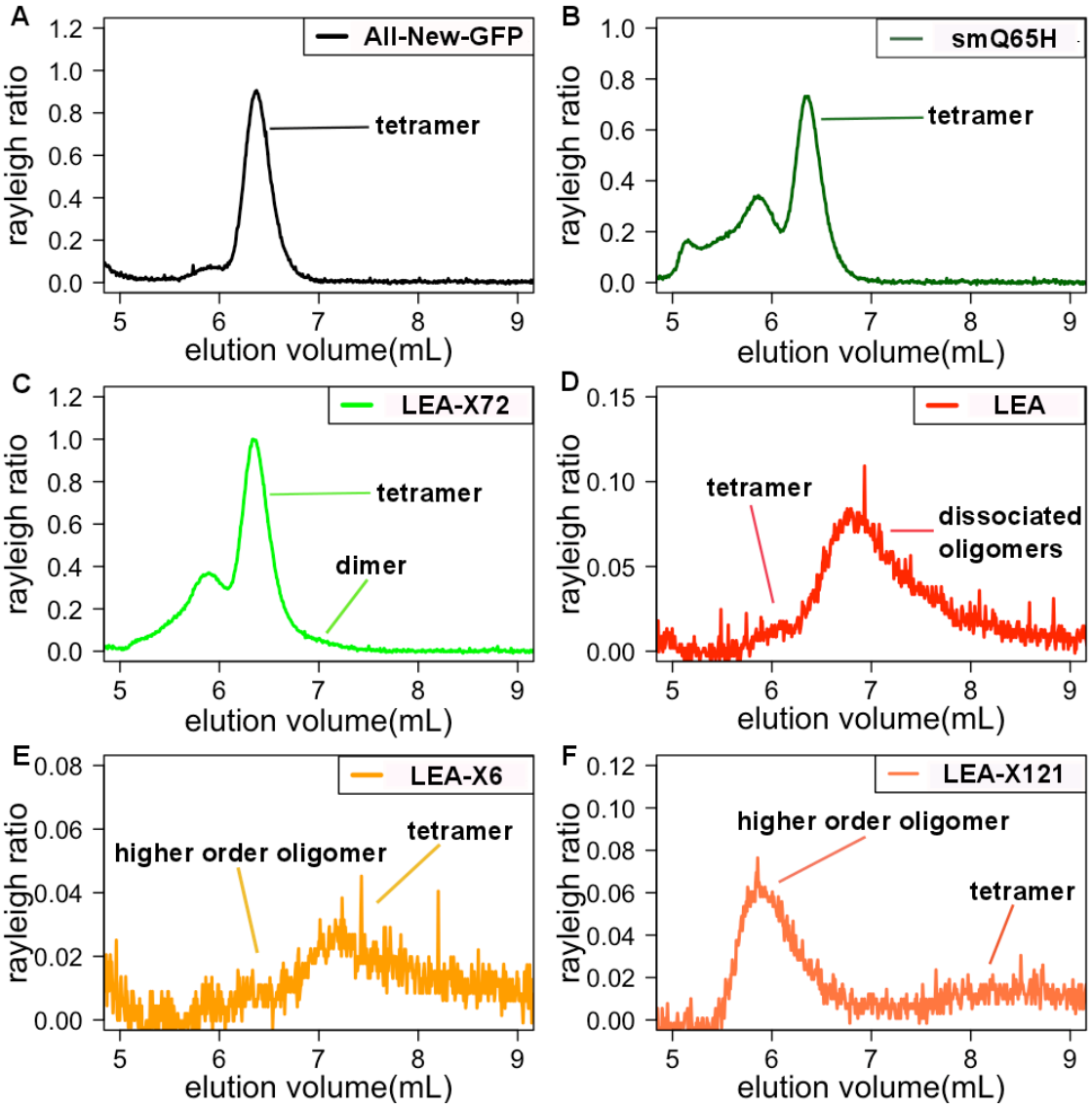


Figure 3.10: Molar mass and oligomeric state determination at 500 µg/mL FP concentration.

The mutant and ancestral FPs analyzed at 500 µg/mL concentration by SEC-MALLS. X-axis: elution volume (SEC), Y-axis: Rayleigh ratio (MALLS). **(A)** The green ancestor, **(B)** smQ65H: both of these FPs are predominantly tetramers. **(C)** LEA-X72: majority of the FP is a tetramer, with a small amount of dissociated dimer. **(D)** LEA: majority of the FP is a mixture of dissociated oligomers, with a small amount of tetramer. **(E)** LEA-X6: majority of the FP is a tetramer, with a small amount of higher order oligomer. **(F)** LEA-X121: majority of the FP is a higher order oligomer, with a small amount of tetramer.

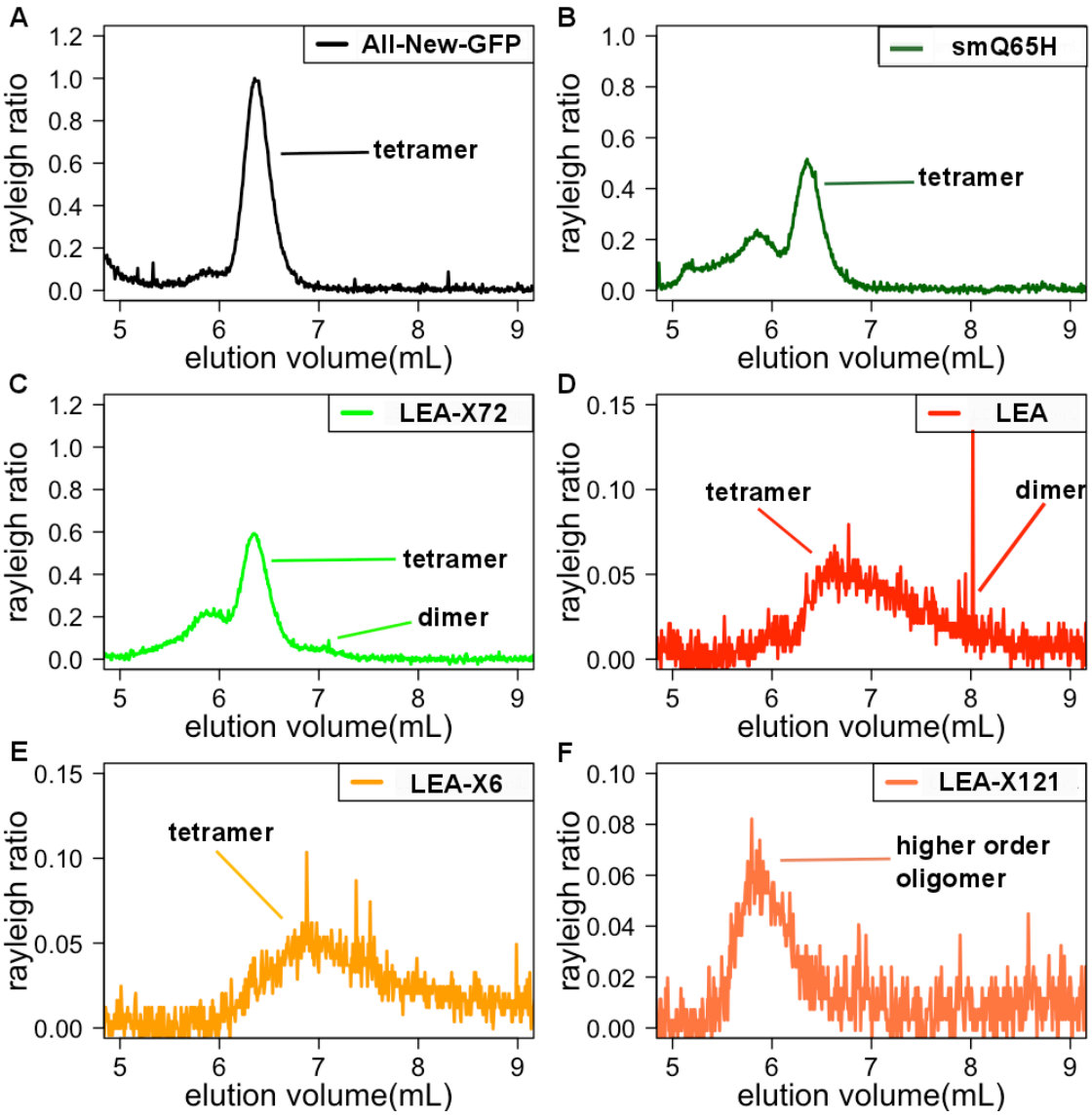


Figure 3.11: Molar mass and oligomeric state determination at 250 μg/mL FP concentration.

The mutant and ancestral FPs analyzed at 250 μg/mL concentration by SEC-MALLS. X-axis: elution volume (SEC), Y-axis: Rayleigh ratio (MALLS). **(A)** The green ancestor, **(B)** smQ65H: both of these FPs are predominantly tetramers. **(C)** LEA-X72: majority of the FP is a tetramer, with a small amount of dissociated dimer. **(D)** LEA: majority of the FP is a tetramer, with a small amount of dissociated dimer. **(E)** LEA-X6: majority of the FP is a tetramer. **(F)** LEA-X121: only higher order oligomer is detected.

At lower concentrations (500 μg/mL and 250 μg/mL) the light scattering signal was too weak to reliably determine the oligomeric states of LEA, LEA-X6 and LEA-X121 (Figures 3.8D, E, F; 3.10-11(D, E, F)). The static light scattering results for the LEA and LEA mutant showed that these FPs exist predominantly as tetramers at a higher concentration (5 mg/mL), but readily dissociate at lower protein concentrations; except for LEA-X72, which did not dissociate with lowering of the protein concentration.

At two lower FP concentrations (500 μg/mL and 250 μg/mL) we observed LEA, LEA-X6 and LEA-X121 as the higher-order oligomers and/or the dissociated FPs, with the higher-order oligomers being the predominant fractions. The detection of the higher-order oligomers at lower concentrations (500 μg/mL and 250 μg/mL) can be attributed to the fact that molecular species with higher molar masses tend to have a higher light scattering signal; because the product of molar mass and the concentration of a molecular species is directly proportional to the light scattering signal. Overall, at lower concentrations MALLS signal to noise ratio was very low of the three FPs, LEA, LEX-X6 and LEA-X121, which is reflected by the high percent error associated with the estimated molar masses of these three FPs. (Table 3.1).

Table 3.1: LEA FP molar mass (g/mol) by light scattering.

Proteins (Cal. M.W.)	**Peaks**	**Molar Mass (g/mol) at 5 mg/mL (%error) (Oligomeric state) (Elution vol.) Stoichiometry, Cumulative MW fractions**	**Molar Mass (g/mol) at 500 µg/mL (%error) (Oligomeric state) (Elution vol.) Stoichiometry, Cumulative MW fractions.**	**Molar Mass (g/mol) at 250u µg/mL (%error) (Oligomeric state) (Elution vol.), Stoichiometry, Cumulative MW fractions.**
All-GFP-New (26675.3)	1	2.382×10^5 ++ (0.7%) (higher order oligomer) (5.595-6.038mL) 8.92, ~5%	2.194×10^5 ++ (7%) (higher order oligomer) (5.729-6.038mL) 8.22, ~10%	3.991×10^5 ++ (22%) (higher order oligomer) (5.702-5.971mL) 14.96, ~5%
	2	1.065×10^5 ** (0.4%) (tetramer) (6.367-7.106mL) 3.99, ~95%,	1.150×10^5 ** (5%) (tetramer) (6.361-6.898mL) 4.31, ~90%	1.271×10^5 ** (9%) (tetramer) (6.367-6.999mL) 4.76, ~90%

Table 3.1: Continued.

LEA (26412)	1		1.165 x 10^6 ++ (57%) (higher order oligomer) (5.870-6.220mL) 44.11, ~5%	
	2	2.507 x 10^5 ++ (1.1%) (higher order oligomer) (5.427-6.166mL) 9.49, ~10%		
	3	9.294 x 10^4 ** (0.3%) (tetramer) (6.475-7.314mL) 3.52, ~90%		1.081 x 10^5 @@ (15%) (tetramer) (6.199-6.770mL) 4.09, ~20%
	4		7.431 x 10^4 ++ (13%) (trimer) (6.435-8.510mL) 2.81, ~95%	
	5			4.980 x 10^4 ** (13%) (dimer) (6.851-7.563mL) 1.89, ~70%

Table 3.1: Continued.

LEA-X6 (26383.9)	1	4.837 x 10^5 ++ (3%) (higher order oligomer) (5.011-5.353mL) 18.33, ~5%		
	2	2.359 x 10^5 ++ (0.5%) (higher order oligomer) (5.461-6.139mL) 8.94, ~10%	3.107 x 10^5 ++ (68%) (higher order oligomer) (6.240-6.609mL) 11.87, ~10%	
	3	1.128 x 10^5 ** (0.4%) (tetramer) (6.455-7.959mL) 4.28, ~85%	1.204 x 10^5 ++ (46%) (tetramer) (6.670-8.476mL) 4.56, ~85%	1.326 x 10^5 ++ (37%) (tetramer) (6.132-8.396mL) 5.03, ~95%

Table 3.1: Continued.

LEA-X121 (26461.1)	1			8.053 x 10^5 ++ (23%) (higher order oligomer) (5.541-6.354mL) 30.43, ~95%
	2	3.469 x 10^5 ++ (0.8%) (higher order oligomer) (5.373-6.421mL) 13.11, ~10%	4.845 x 10^5 ++ (18%) (higher order oligomer) (5.407-6.797mL) 18.31, ~30%	
	3	9.897 x 10^4 @@ (2.6%) (tetramer) (6.535-6938mL) 3.74, ~15%	1.055 x 10^5 ++ (35%) (tetramer) (7.630-8.396mL) 3.99, ~55%	
	4	5.934 x 10^4 ** (1.1%) (dimer) (7.503-8.537mL) 2.24, ~65%		

Table 3.1: Continued.

LEA-X72 (26442)		4.485 x 10^5 ++ (1.4%) (higher order oligomer) (5.064-5.293mL) 16.96, ~2%		
		3.044 x 10^5 ++ (0.4%) (higher order oligomer) (5.333-5.68mL) 11.51, ~5%		
		2.040 x10^5 ++ (0.19%) (higher order oligomer) (5.608-6.085mL) 7.72, ~30%	2.513 x 10^5 ++ (2.9%) (higher order oligomer) (5.064-6.085mL) 9.50, ~15%	2.643 x 10^5 ++ (11%) (higher order oligomer) (5.091-6.065mL) 10.00, ~25%
		9.551 x 10^4 ** (0.4%) (tetramer) (6.347-6.999mL) 3.61, ~55%	1.067 x 10^5 ** (1.1%) (tetramer) (6.347-6.777mL) 4.04, ~60%	9.706 x 10^4 ** (3%) (tetramer) (6.327-6.737mL) 3.67, ~50%
			5.027 x 10^4 ++ (15%) (dimer) (6.878-7.496mL) 1.90, ~15%	4.441 x 10^4 ++ (20%) (dimer) (6.784-7.180mL) 1.68, ~30%

Table 3.1: Continued.

All_Q65H (26684.3)		7.699 x 10^5 ++ (2.5%) (higher order oligomer) (4.964-5.306mL) 28.85, ~2%	6.412 x 10^5 ++ (4%) (higher order oligomer) (5.011-5.521mL) 24.03, ~2%	1.482 x 10^6 ++ (17%) (higher order oligomer) (5.037-5.333mL) 55.54, ~2%
		3.818 x 10^5 ++ (0.6%) (higher order oligomer) (5.353-5.602mL) 14.31, ~5%		3.197 x 10^5 ++ (8%) (higher order oligomer) (5.380-6.085mL) 11.98, ~30%
		2.375 x 10^5 ++ (0.29%) (higher order oligomer) (6.649-6.079mL) 8.90, ~25%	2.633 x 10^5 ++ (1.7%) (higher order oligomer) (5.642-6.112mL) 9.67, ~20%	
		1.058 x 10^5 ** (0.5%) (tetramer) (6.354-7.039mL) 3.97, ~65%	1.299 x 10^5 ** (4%) (tetramer) (6.347-6.999mL) 4.87, ~65%	1.269 x 10^5 ** (10%) (tetramer) (6.361-6.979mL) 4.76, ~65%

**Molar mass was calculated based on the right half of the peak in order to minimize the influence of a higher order oligomeric peak. Therefore, the elution volume corresponds to the right half of the peak.

@@ Molar mass was calculated based on the left half of the peak in order to minimize the influence of a lower order oligomeric peak. Therefore, the elution volume corresponds to the left half of the peak.

++Molar mass was calculated based on the whole peak, since these peaks were minor peaks or their molar mass was not significantly influenced by surrounding peaks.

DISCUSSION

The red converted LEA has a comparable emission profile (emission maxima at 578nm) to that of the extant FP, R1-2 (Field and Matz, 2010). However, it remained unclear whether the LEA with 12 historical mutations has a similar stability and oligomeric profiles as that of the extant FP, R1-2. Other extant Kaede-type FPs have been reported to be tetramers, while R1-2 evolved into a dimer from a tetrameric green ancestor (Chapter 2: Modi et al, 2014; Ando et al., 2002; Hayashi et al, 2007). Furthermore, we have previously shown that the three key historical mutations involved in the emergence of red fluorescence from the green ancestor diminished oligomeric tendency and destabilized the FP. Given these observations and based on the gel migration patterns of non-denatured LEA, we hypothesize that a loss of stability may be prerequisite for evolving red fluorescence. Therefore, we anticipated lower stability profiles for LEA, intermediate ancestor (R/G), and extant FP relative to the stability profile of the green ancestor. To better understand the impact of 12 historical mutations on the stability of the LEA, we used DSF to determine the relative stabilities of FPs (Chapter 2: Modi et al., 2014; Senisterra and Finerty, 2009; Dai et al., 2007; Kiss et al., 2009). Relative stabilities of these FPs were consistent with our hypothesis – the diminishing stability of FPs followed the evolutionary trajectory when compared to the previously reported stability profile of the green ancestor (Kelmanson and Matz, 2003; Ugalde et al., 2004; Chapter 2: Modi et al., 2014): the green ancestor has the highest stability, followed by the intermediate ancestor (R/G), while the extant FP was the least stable of the three (Chapter 2: Modi et al, 2014).

The data showed that the photo-conversation did not impact the R1-2 stability profile (Figure 3.5B); however, the LEA stability profile differed between the unconverted (green) and converted (red) forms. Remarkably, the stability of the LEA was

even lower than that of the extant FP (R1-2), which suggests that some of the 25 mutational changes between the green ancestor and extant FP (R1-2) that did not have an effect on the color of fluorescence compensated for the destabilizing effect of 12 color-changing mutations. Therefore, difference in the stability profiles of LEA and R1-2 could be explained by the absence of compensatory mutations in LEA (DePristo et al, 2005; Field and Matz, 2010). We only observed difference in stability profiles of the unconverted (green) and converted (red) forms of LEA, but not that of unconverted (green) and converted (red) forms of R1-2, suggesting some of the 25 remaining mutations likely have a role influencing the stability of unconverted (green) and converted (red) forms (Liberles et al., 2012; DePristo et al., 2005). These observations suggest that the mutations involved in changing the chromophore environment also destabilized and diminished the oligomeric tendency of the FP over the course of evolution through long-range interactions (Chapter 2: Modi et al., 2014; DePristo et al, 2005).

Using SEC-MALLS here we showed that the green ancestor, intermediate ancestor and LEA are tetramers, while the extant FP, R1-2, is a dimer (Chapter 2: Modi et al., 2014). These results were unexpected because generally a tetramer such as LEA would have higher thermal stability than a dimer such as R1-2, because of the higher likelihood of a larger number of interactions between tetrameric interfaces (Chapter 2: Modi et al, 2014, Murugan and Hung, 2012; Banks and Gloss, 2003; Hashimoto et al., 2011; Ali and Imperiali, 2005; Liberles et al., 2012). Taken together these results demonstrate that while introduction of 12 historical mutations in the green ancestor is sufficient to recapitulate the red fluorescence phenotype, they are not sufficient to recapitulate the biophysical characteristics of the extant FP, R1-2 (Field and Matz, 2010; Kim et al., 2013).

Previous work suggests that a subset of historical mutations in LEA increased the space around the chromophore, allowing it to change conformation and making the green-to-red transition possible (Kim et al., 2013). Additionally, a significantly diminished stability of LEA relative to the green ancestor and extant FP (R1-2) suggested that the mutations modifying the chromophore cavity might also contribute to the evolution of general biophysical properties relationship of the extant FP, R1-2 (Figure 3.4). We showed here that two buried mutations proximate to the chromophore at position 72 and 121 influenced the stability and oligomeric tendency of the FP most significantly (Figure 3.2); at the same time, these two historical mutations are among the most crucial for the evolution of red fluorescence (Field and Matz, 2010; Chapter 2: Modi et al., 2014).

A reversion of historical mutation at position 72 to the ancestral state (A72T) led to a large increase in stability of LEA-X72, restoring the stability back towards the ancestral state (Figure 3.7). These results supported the earlier findings using forward mutagenesis in the green ancestor, showing the significant contribution of historical mutation at position 72 (T72A) in reducing the FP stability, while contributing to red-shifting the ancestral fluorescence spectra by 15nm and the emergence of red fluorescence (Chapter 2: Modi et al, 2014; Field and Matz, 2010). Position 72 is located in the central helix, which is buried inside the fluorescent protein. Since the residue at position 72 is not involved in forming the tertrameric structure, it must be destabilizing the protein through unknown long-range interactions (Kim et al., 2013; Chapter 2: Modi et al., 2014, Liberles et al., 2012; Harms and Thornton, 2010; Weinreich et al., 2006; DePristo et al., 2005).

We also observed the LEA revertant mutant, LEA-X72, to be tetrameric at all three concentrations; however, a small fraction of LEA-X72 dissociated into a dimer at

two lower concentrations (Figures 3.8C, 3.9, 3.10C, 3.11C). The dissociative tendency of LEA-X72 at lower protein concentrations suggests influence of remaining 11 historical mutations in LEA on its oligomeric state. Therefore, a small observed difference between the stability profiles of LEA-X72 and the green ancestor could be attributed to destabilizing effect of the remaining 11 historical mutations on 3° and 4° structures of the FP.

When five other historical mutations are reverted to the ancestral state (LEA-X121: N121Y; LEA-X6: V63A, H77D, T162M and I165V), resultant slight increase in the stability profile of LEA-X121 and LEA-X6 once again suggest a small contribution of these five historical mutations toward evolutionary destabilization of the FP (Figure 3.7). However, LEA-X121 was observed to have multiple oligomeric states: a higher-order oligomer, tetramer and dimer (Figures 3.8, 3.9). The data also showed that with the exception of LEA-X121, all FPs (the green ancestor, smQ65H, LEA, LEA-X6 and LEA-X72) formed a tetramer at a high protein concentration of 5 mg/mL (Figures 3.8, 3.9). The dissociation of LEA-X121 was unexpected because the DSF revealed that the historical mutation at position 121 (Y121N) contributed slightly toward destabilizing the ancestral FP. Furthermore, the structural data (PDB: 4DXI and 4DXN) showed the position 121 to be buried, and suggests the likely role of historical mutation Y121N) in enlarging the chromophore cavity (Kim et al., 2013). Therefore, the dissociation of revertant LEA-X121 can be attributed to possible unknown long-range interactions between the ancestral residue at position 121 (Y121) and the interface residues.

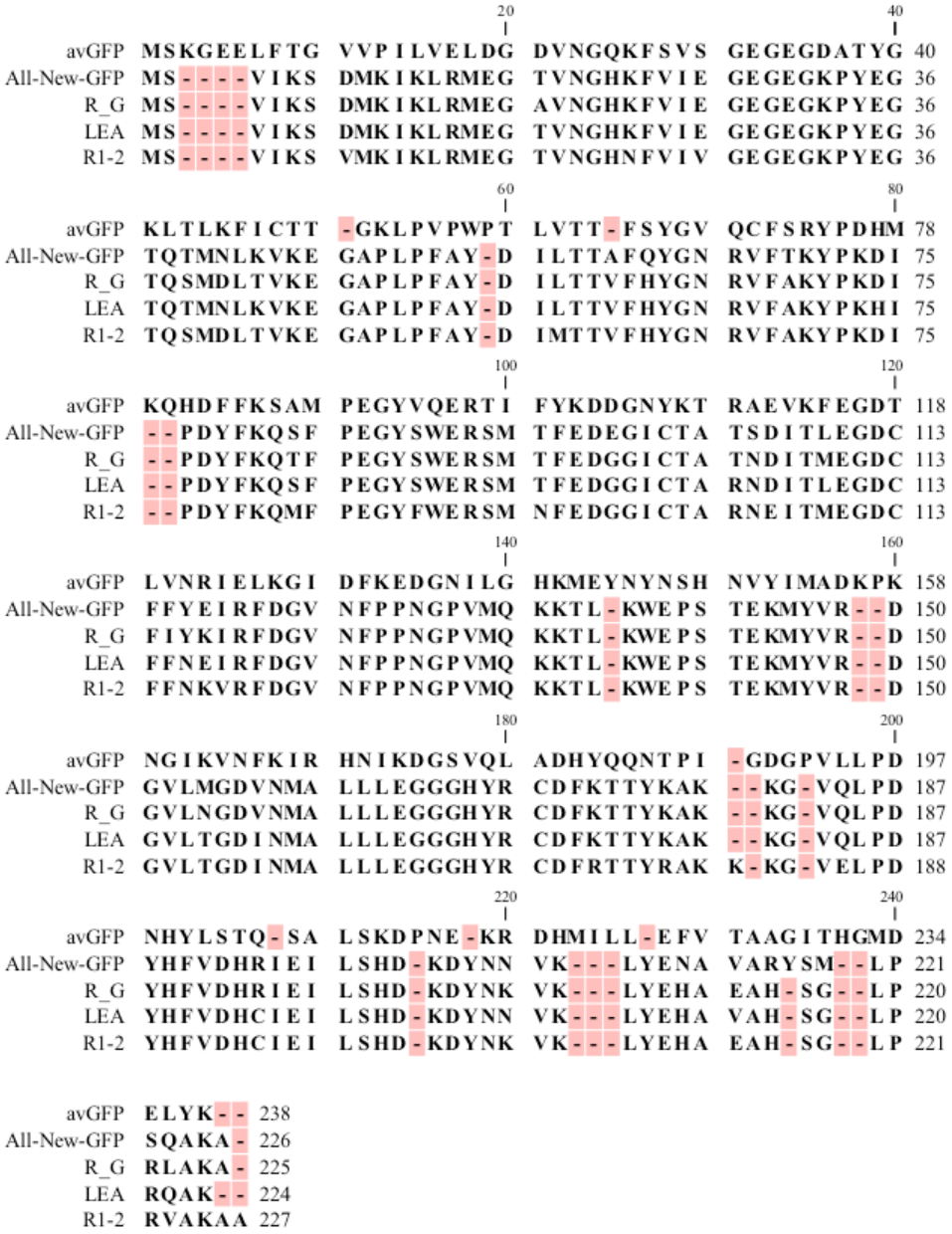


Figure 3.12: Multiple sequence alignment (amino acids) of ancestral FPs and the extant FP.

Amino acid sequences of the green ancestor (All-New-GFP), intermediate ancestor (R/G), extant FP (R1-2) and Least Evolved Ancestor (LEA) were aligned using Promales3D, and then manually adjusting the alignment (Henikoff S and Henikoff, 1992; Pei and Grishin, 2007).

Based on the predicted sequences of ancestral proteins at the nodes of FP phylogeny, the mutational event at position 121 (Y121N) occurred after the evolution of intermediate R/G ancestor (Field and Matz, 2010; Figures 3.1, 3.12). Given the dimeric extant FP (R1-2) evolved from the tetrameric intermediate ancestor (R/G) and with the detection of a dimer in revertant LEA-X121, we propose that the evolution of extant dimeric state happened first followed by the historical mutational event at position 121. If so, other historical mutations involved in functional evolution and the compensatory mutations favoring the dimeric state must have appeared prior to the mutation at 121 (Y121N), which was important for facilitating the synthesis of the red chromophore. This observation is further supported by the reported tetrameric nature of G/R photoconvertible Kaede FP, one of the closest homolog of R1-2 (Figure 3.1) (Ando et al., 2002; Hayashi et al, 2007). Both of these FPs evolved from the intermediate ancestor (Ugalde et al., 2004).

The dissociation of LEA and LEA-X6 at two lower protein concentrations, 500 μg/mL and 250 μg/mL, suggested that their lower oligomeric tendencies might have contributed to loss of stability (Figures 3.7, 3.10, 3.11). This observation was further supported for not just LEA and LEA-X6, but also for LEA-X121. These three FPs (LEA, LEA-X6 and LEA-121) that readily dissociated with lowering of protein concentration also have the lowest thermal stability profiles (Figure 3.7), while the tetrameric state of the other three FPs (the green ancestor, smQ65H and LEA-X72) with relatively high thermal stability remained unaffected upon lowering of protein concentration.

CONCLUSIONS

Taken together, our results highlight the role of the six historical mutations in evolution of the present-day biophysical characteristics of the Kaede-type red FP. Mutations affecting the immediate chromophore environment contributed in both a large and small way to destabilize the FP toward the lower stability of extant FP, R1-2. Additionally, the near-chromophore mutations contributed to diminishing the oligomerization tendency of the protein as it evolved towards the dimeric extant FP, R1-2, ostensibly via some long-range interactions with the interface-forming residues on the surface. The dramatically reduced stability of LEA both in comparison to the ancestral and extant states could be attributed for the most part to the mutation at position 72 (T72A). The results reaffirmed our earlier findings showing evolutionary importance of T72A mutation in destabilizing the FP. Relatively minor effect of the four other near-chromophore mutations on the protein's stability mirrors their relatively minor effects on red fluorescence (Matz and Field, 2010; Liberles et al., 2012; DePristo et al., 2005). The 25 historical mutations that did not affect fluorescence color played a crucial role in compensating for the destabilizing effect of the functionally relevant mutations, and also might have contributed to the evolution of dimeric extant FP from the tetrameric ancestor. Overall, our results support the hypothesis that evolution of complex novel function requires initial destabilization of the protein and highlight the role of long-range interactions and compensatory mutations in facilitating this process (Liberles et al., 2012; Harms and Thornton, 2010; Weinreich et al., 2006; DePristo et al., 2005; Yokoyama et al, 2008).

Chapter 4: Evolution of Oligomerization in Hydrozoan Fluorescent Proteins

ABSTRACT

In the past decade, the resurrections of ancestral proteins have provided an elegant solution to the study of molecular evolution. Employing fluorescent protein models and statistical phylogenetic reconstruction, we showed that the resurrected ancestral Hydrozoan FP is a monomeric green FP with a large Stokes shift, exhibiting an excitation peak at 395nm and an emission peak at 510nm. Since several of the extant FPs used in this study are excited through interactions with biomiluminescent proteins with emissions at 450-460nm, excitation at 395nm in the ancestor suggested that the extant molecular phenotype appeared after the divergence of Leptothecata. The neutral and anionic chromophore is responsible for the unique dual excitation maxima of *Aquoria victoria* GFP; however, a complete absence of the second excitation peak in the ancestral FP suggested that the unique dual excitation peaks associated with bioluminescence observed in *Aquoria victoria* GFP likely evolved during the time since the divergence of Leptothecata. Additionally, the presence of a large diversity of homooligomeric states (monomers, dimers, tetramers and higher order oligomers) in Hydrozoan fluorescent proteins (FPs; the majority of native FPs form a highly stable canonical tetrameric structure) provided a unique opportunity to study the molecular evolution of homo-oligomerizaiton through a comparison between the ancestral FP and the extant FP sequences. The present study of FP oligomerization presents the opportunity to greatly enhance our ability to engineer many more novel, genetically encoded monomeric biomarkers needed for the biomedical fields.

INTRODUCTION

A recent report (Hashimoto et al., 2011) identified that close to 60% of proteins in a protein database formed some level of homo-oligomers, which are fundamental to driving cellular functions. The presence of neutral mutations, structure-function convergence or divergence makes it difficult to teasing apart long-range interactions based sequence and structural comparison of extant orthologs and paralogs proteins (Harms and Thornton, 2010; Gerlt and Babbitt; 2009). In the past decade, the resurrections of ancestral proteins through the statistical phylogenetic method have provided an elegant solution to the study of evolution of oligomerization and the evolution of molecular complexity (Finnigan et al., 2012).

The presence of a large diversity of homo-oligomeric states (monomers, dimers, tetramers and higher order oligomers) in Hydrozoan fluorescent proteins (FPs) (Prasher et al., 1992; Ormo et al, 1996; Xia et al., 2002; Gurskaya et al., 2003; Shagin et al., 2004; Markova et al., 2010; Hunt et al, 2012; Aglyamova et al., 2011) provided a unique opportunity to study the molecular evolution of homo-oligomerizaiton through the resurrection of an ancestral FP gene. In the present study, we resurrected three ancestral FPs using four different amino acid substitution models and degenerate gene synthesis (Dayhoff et al., 1972; Dayhoff et al, 1978; Jones et al., 1992; Whelan and Goldman, 2001; Le and Gascuel, 2008; Thornton, 2004; Ugalde et al., 2004), accounting for the uncertainty associated with predicting an ancestral sequence. We hypothesized a tetrameric structure for the resurrected ancestors owing to the fact it was the most common structure observed in the Green Fluorescent Protein (GFP)-family of proteins (Chudakov et al., 2010). Furthermore, our hypothesis was based on previously reported tetrameric structure for the Faviina ancestral FP (Field and Matz, 2010; Chapter 2: Modi et al., 2014).

Here we report that one of the three ancestral FPs synthesized was functional. The functional ancestor was for the Leptothecata sub-clade in the Hydrozoan clade harboring all homo-oligmerization states discovered in the GFP-family of proteins., Importantly, the most common variant of the ancestor was a monomeric and not a tetramer as per our hypothesis.

MATERIALS AND METHODS

Multiple sequence alignment, phylogeny and computational phylogenetic analysis

To retrace the evolutionary steps leading to diverse oligomeric states we decided to resurrect ancestral FPs (Ugalde et al., 2004; Field and Matz, 2010; Field et al., 2006) positioned at the key nodes within the hydrozoan FPs phylogeny (Figure 4.2). All known Hydrozoan FPs were aligned using a structure-based multiple sequence alignment algorithm, PROMALS3D using default settings (Figure 4.1; Pei et al., 2008). The alignment was manually corrected in CLC Sequence Viewer 6 and eBoiX. Corrected alignment was used to generate a phylogenetic tree using MrBayes 3.2 (Huelsenbeck and Ronquist, 2001). The prediction of the ancestral protein sequences was done using FASTML algorithm based on four different amino acid substitution models: LG, WAG, Dayhoff and JTT (Ashkenazy et al., 2012; Pupko et al., 2002; Pupko et al., 2000; Pupko et al., 2000; Dayhoff et al., 1972; Dayhoff et al, 1978; Jones et al., 1992; Whelan and Goldman, 2001; Le and Gascuel, 2008;). FASTML algorithm was run using the default settings for each of the four models with and without input from a phylogenetic tree. The variations in the model's predictions were incorporated into the synthesized genes in the form of degeneracy at the variable positions (Figure 4.3; Ugalde et al., 2004). Genes

coding for three ancestral proteins have been synthesized expressed in *E. coli* (Figure 4.2).

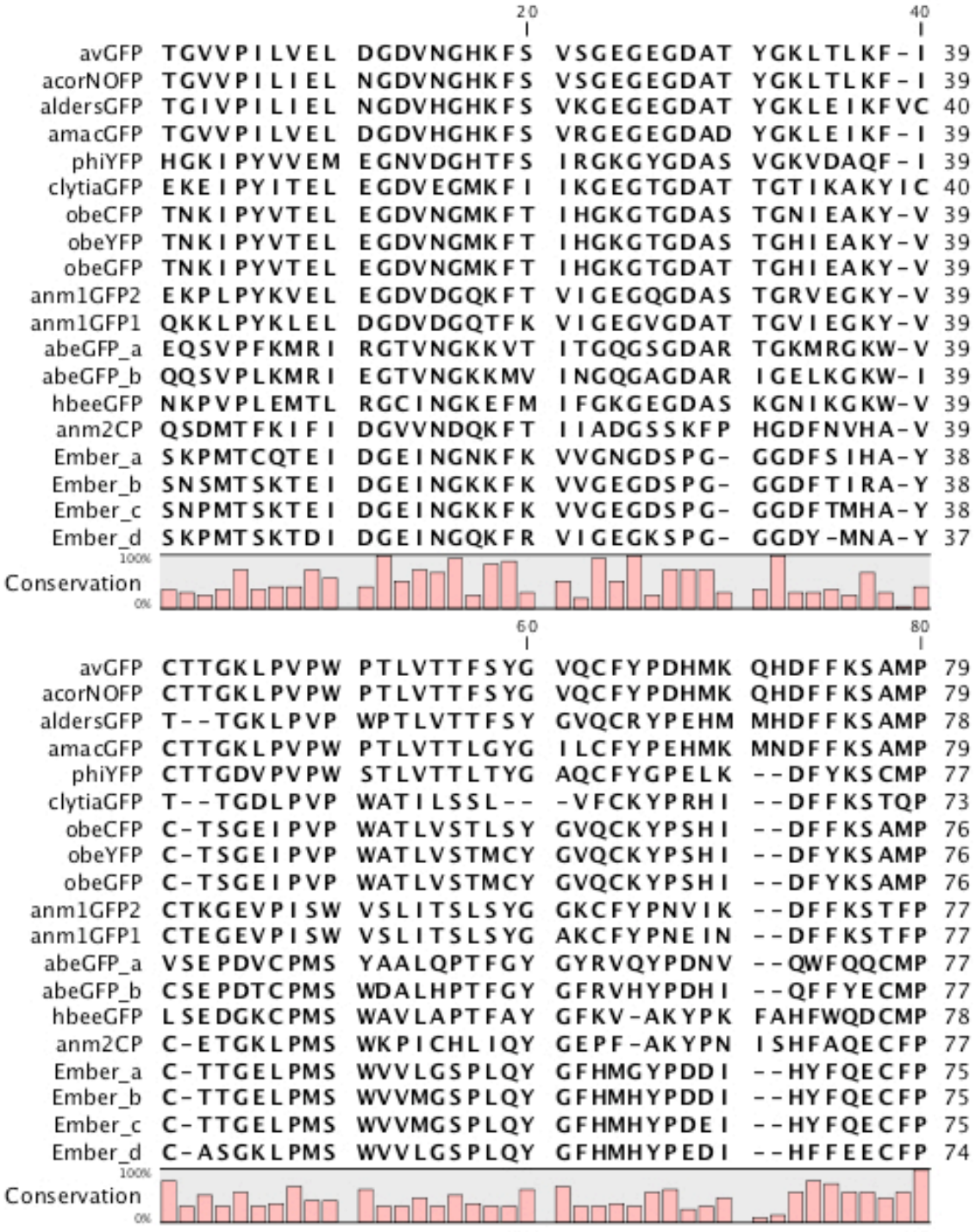


Figure 4.1: Continued.

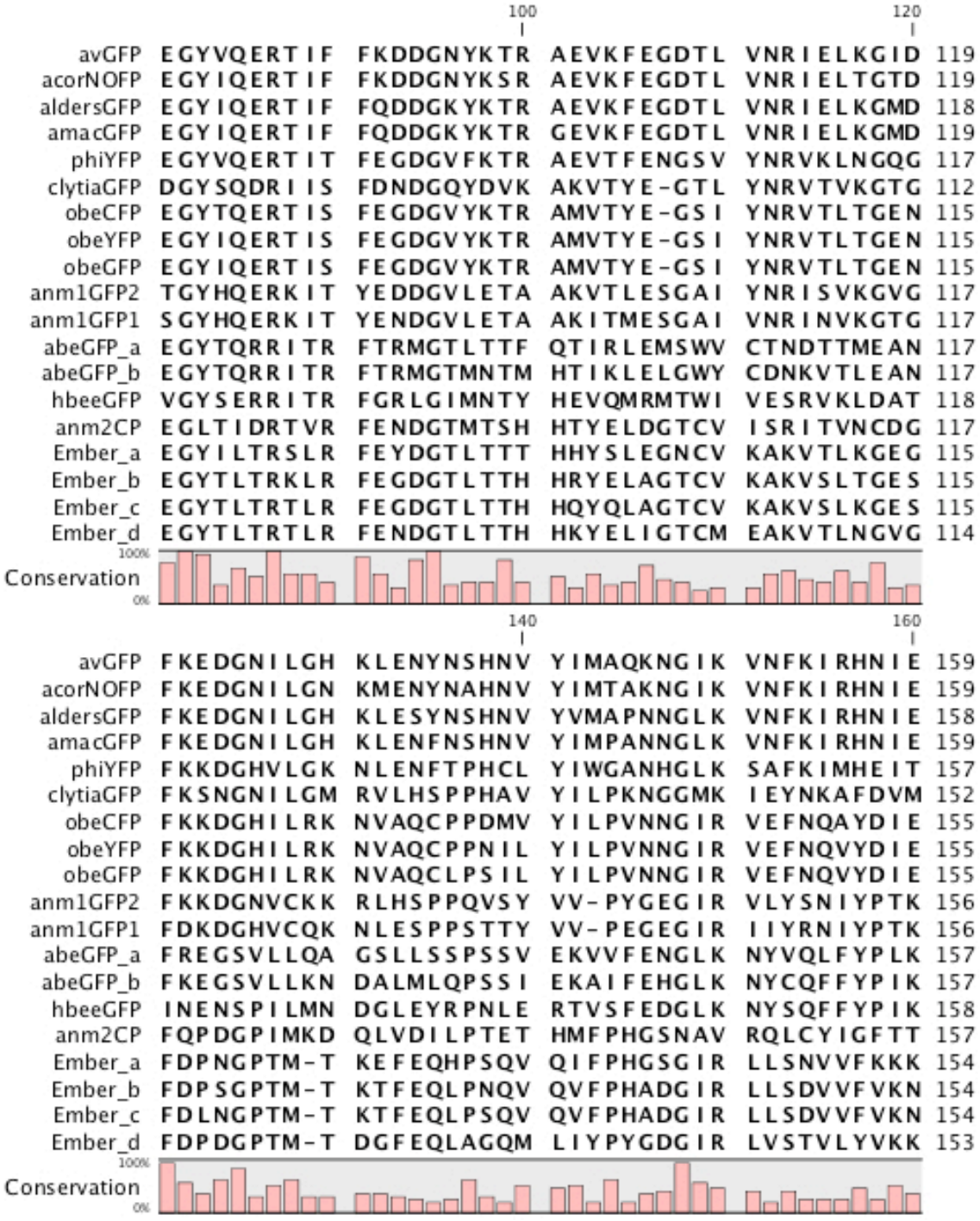


Figure 4.1: Continued.

Figure 4.1: Continued.

Figure 4.1: Multiple sequence alignment (amino acids) of the known hydrozoan FPs. Extant hydrozoan FP amino acid sequence alignment constructed first by PROMALS3D and then manually adjusted. The alignment has all known Hydrozoan FPs representing the oligomeric state and molecular functional diversities. The alignment was used to generate the phylogenetic tree using MrBayse, and then used to determine the ancestral genes using FASTML algorithm.

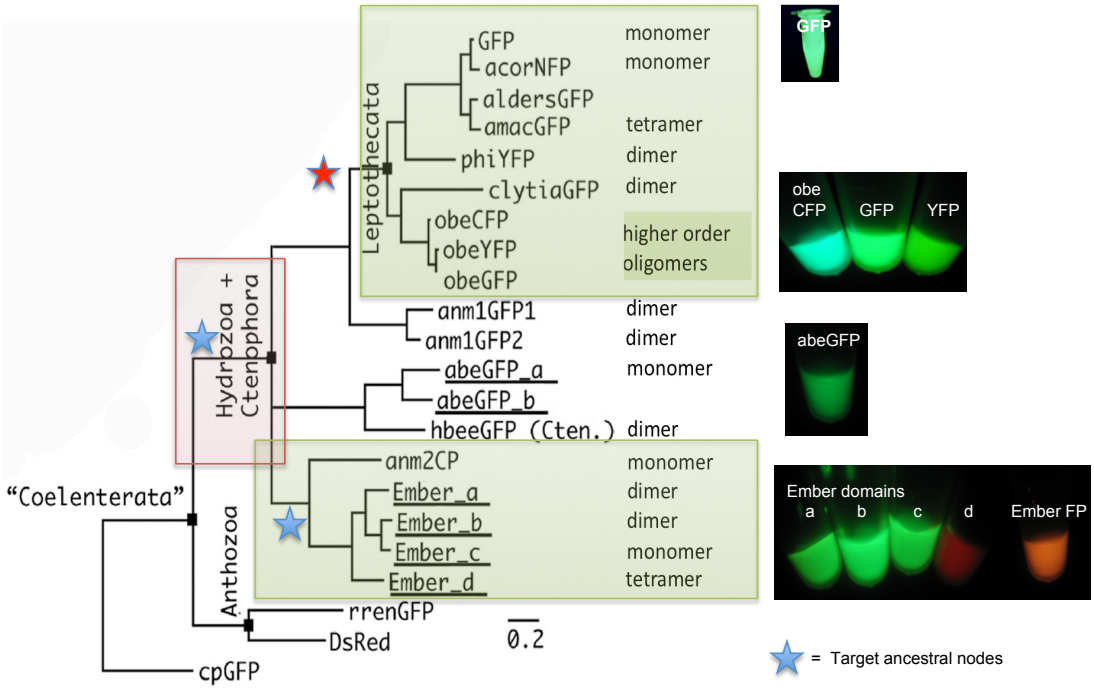


Figure 4.2: Hydrozoan fluorescent proteins phylogeny and corresponding oligomeric states.

Phylogenetic relationships and oligomeric states of the Hydrozoan fluorescent proteins. A rooted tree illustrating phylogenetic relationships between known extant Hydrozoan fluorescent proteins (FPs). The diversity of oligomeric states in the Hydrozoan FPs is described on the right hand side of each protein's name. The diversity of the color pallet in this clade is represented by a select few FPs; exemplifying, cyan, green, yellow and red FPs. The resurrected ancestral genes at three nodes are marked by stars: (i) the ancestor of all Leptothecata FPs is designated as Node-13, (ii) the ancestor of Ember FPs and anm2CP is designated as Node-3, (iii) the ancestor of all Hydrozoan and tenophara FPs is designated as Node-2. The ancestral gene that resulted in a functional FP (Node-13) and successfully expressed in bacteria is indicated by the red star. Three FPs, two anthozoan and one bilaterian were used as outgroups. The ctenophoran FP, hbeeGFP, is not completely verified and it is suspected to be a Hydrozoan FP (Chudakov et al., 2010).

Synthesis of Degenerate Fluorescent Protein Libraries

The fluorescent protein libraries were synthesized in-house at the University of Texas at Austin-Applied Research Laboratories. The oligonucleotides necessary for gene assembly were synthesized on a MERMADE 192 oligonucleotide synthesizer (Bioautomation; Plano, TX) using standard phosphoramidite synthesis conditions. Table 4.1 shows the oligos used for synthesis of degenerate Node-13 ancestral gene. Locations of sequence degeneracy within the gene library were denoted using IUPAC nucleotide code and introduced into the assembled gene library via degenerate oligonucleotides synthesized using custom phosphoramidite mixtures normalized for the coupling efficiencies and molecular weights of the individual phosphoramidites to minimize synthesis bias (Hall et al., 2009). The gene synthesis scheme and the resulting substituent oligonucleotide sequences were designed using GeneFab software (Cox et al., 2007) setting the maximum oligonucleotide length at 80 nucleotides and the melting temperature of the oligonucleotide overlap sequences fixed at 60°C ± 3°C. The resulting oligonucleotides were inspected and manually manipulated to avoid placing degenerate regions in oligonucleotide overlap regions to minimize assembly problems. The gene libraries encoding the fluorescent protein variants were robotically assembled according to the protein fabrication automation methodology (Cox et al., 2007; Miklos et al., 2012) using a Tecan EVO liquid handling robotic platform guided by custom software.

Table 4.1: Oligos used in synthesis of degenerate gene library for Node-13.

Degenerate Oligos		
Oligo Name	**Synthesis Oligo**	**Sequence**
Dengen1300001	1a.A	AAATCTTTAATATGGKMCGGATATTTCGCAAA GCACTGCACGCCATAGCTCAGGGTGGTCACCA GGGTCGCCCACGGCAC
Dengen1300002	1a.B	GGTATGGGTGCTCAGATGATGATAWYKCGGAA TATGCACCGGGCYCGCGCCCAGCGGGCGGTTCT GCTGATGATGATCCGCCAGCTG
Dengen1300003	1s.A	GGCGATGCGAGCACCGGCAAAATTGAAGCGAA ATATATTTGCACCACCGGCGAAVTKCCGGTGCC GTGGGCGACCCTG
Dengen1300004	1s.B	GGATAAAGCGAACAACGGCATTAAAGTGGAAT TTAACATTGCGTATGATATTGAAGGCGGCGGCT ATCAGCTGGCGGATCATCATCAG
Dengen1300005	2a.A	TTCAAAGCTAATGGTGCGTTCCTGGGTATAGCC TTCCGGCATCSMGCTTTTAAAAAAATCTTTAAT ATGGMKCGGATATTTCGCA
Dengen1300006	2a.B	ACCACTTCCACCAGCACCATATGATCGCGGCGT TCGTTCGGATCTTTGCTCAGKBTGGTATGGGTG CTCAGATGATGAT

Table 4.1: Continued.

Dengen1300007	2s.A	AAATTCCGTATGTGGTGGAACTGGAAGGCGAT GTGAACGGCCAKAAATTTACCATTCGCGGCGA AGGCRMCGGCGATGCGAGCACCGGC
Dengen1300008	2s.B	GCCATATTCTGGGCAAAAACCTGGAATTTAAC WRCCCGCCGCATASCGTGTATATTCTGCCGGAT AAAGCGAACAACGGCATTAAAG
Dengen1300009	3a.A	TCACGCGGTTATAAATGCTGCCGTTTTCAAAGG TCACTTTCGCGCGGGTTTTATACACGCCATCGY CTTCAAAGCTAATGGTGCGTTCCTGGGT
Dengen1300010	3a.B	TCATCAGTGATGGTGATGGTGATGTTTATACAG TTCGGTCATGGTAATATCCACCGCTTTCACCAC TTCCACCAGCACCATATG
Dengen1300011	3s.A	TTGATTGATTGAAGGAGAAATATCATGAGCRSC GGCGCGGAACTGTTTACCAAAAAAATTCCGTAT GTGGTGGAACTGGA
Dengen1300012	3s.B	AACGGCAGCATTTATAACCGCGTGACCCTGAA AGGCRMGGGCTTTAAAAAAGATGGCCATATTC TGGGCAAAAACCTGGA

Cloning, screening and sequencing of the ancestral FPs

Synthesized degenerate sequences were cloned with 5'-end Shine-Dalgarno sequence, ribosome binding sequence (RBS) and 3'-end His-tag (6x His) by TA cloning using forward primer [5'-TTT GAT TGA TTG AAG GAG AA TA-3'] and reverse

primer [5'-TTA TTA GTG ATG GTG ATG GTG ATG-3'] (IDT; San Jose, CA, USA). The amplification products were ligated into pGEM-T vector (50 nm/ml, Promega; Madison, WI, USA) and transformed into chemically competent one-shot TOP10 and BL21 Star™(DE3)PLysS *E. Coli* cells (Invitrogen by Life Technologies; Carlsbad, CA, USA). Fluorescent clones from the resulting plates were detected and collected under a Leica MZFLIII fluorescent stereomicroscope (Leica; Wetzlar, Germany) and streaked onto new Luria-Bertani-agar plates with added isopropyl β-D-1-thiogalactopyranoside (IPTG; 0.5mM of final concentration) and ampicillin (Amp; 100 μg/mL of final concentration). Plates were incubated at 37ºC for one day and then at 4ºC an additional two to three days to allow chromophore maturation for sufficient fluorescence. The fluorescent clones from the expression library were sequenced using the Sanger method. The sequences were analyzed using eBoiX v. 1.5.1 (eBioinformatics.org) and CLC Sequence Viewer 6 (CLC bio; Cambridge, MA, USA).

Protein expression and purification of the ancestral FP clones

Clones of interest streaked on large format plates, incubated at 37ºC for a 24h and then at room temperature for an additional three to four days to accumulate sufficient biomass. Cells were harvested using HEPES buffer (10mM HEPES, 100mM NaCl at pH 7.4) and lysed by sonication on the Misonix Sonicator 3000 (Misonix Inc., Farmingdale, NY). The expression product was purified from the bacteria using metal-affinity chromatography on Ni-NTA agarose (Invitrogen by Life Technologies; Carlsbad, CA, USA) following the manufacturer's protocol. A detailed description of fluorescent protein expression and purification is in chapter 2 of this dissertation.

Spectroscopy analysis of the ancestral FP clones

The excitation and emission spectra of all bacterial expression products were measured using a SpectraMax M2 plate reader (Molecular Devices; Sunnyvale, CA, USA). Molar extinction coefficient (ME), quantum yield (QY), and relative brightness were measured over a range of 395-400 nm, as described previously (Aglyamova et al., 2011; Matz et al., 2006) *Aequorea victoria* GFP (*Av*GFP) (Creative BioMart; Shirley, NY, USA) was used as a standard (ME = 21,000 M^{-1} cm^{-1}, QY= 0.77) (Heim and Tsien, 1996). A range of dilutions of both native and denatured proteins was used to collect absorption and fluorescence spectra, and the data were plotted as regressions. The slope differences between the new proteins and *Av*GFP were used to determine QY (total fluorescence *versus* absorption) and ME (maximum absorption *versus* concentration). The product of QY and ME of each FP relative to *Av*GFP is relative brightness.

SDS-PAGE analysis of the ancestral FP clones

The oligomeric states of the expressed FPs were initially assessed by SDS-PAGE. FPs were analyzed under non-denatured (unheated protein samples) and denatured (heated protein samples) with added 2-mercaptoethanol (βME; 715mM final concentration) n SDS-running buffer. Protein samples were run in a pre-cast 4-20% gradient gel with SDS-Tris-Glycine buffers (Bio-Rad; Hercules, CA, USA). In gel detection of non-denatured FPs was done using a UV box (UV-B). Denatured FPs were detected by coomassie brilliant blue (CBB) stain.

Oligomerization state analysis of the ancestral FP clones by SEC-MALLS

The molar mass and oligomeric state were further evaluated by HPLC size-exclusion chromatography with combined detection *via* multi-angle light scattering and refractometry. The buffer used for chromatography contained 100 mM NaCl and 10mM HEPES adjusted to pH 7.4, and G4000PW$_{XL}$ or G3000PW$_{XL}$ columns were used (TosoHass, 300 x 7.8mm; Tosoh Bio-science LLC; King of Prussia, PA, USA). The detectors used were the EOS photometer and Optilab rEX differential refractometer (Wyatt Technology; Santa Barbara, CA, USA). Data were analyzed using ASTRA software v. 4.9 and 5.3 (Wyatt Technology; Santa Barbara, CA, USA). A detailed method for determining the FP oligomeric states using SEC-MALLS is in chapter 2 of this dissertation.

RESULTS

Ancestral gene prediction through statistical phylogenetic analysis

Phylogenetic tree constructed using the hydrozoan extant FP sequences recapitulated the previously reported phylogenetic relationships (Figures 4.1, 4.2; Aglyamova et al., 2011; Hunt et al., 2012). Two Anthozoan FPs and a single copepod FP were used as out groups (Figure 4.2). Each of the four substitution models used to determine the ancestral sequence at three nodes by computational statistical method, FASTML, predicted four ancestral amino acid sequences with slight variations for all three nodes (Figure 4.3). A multiple sequence alignment of the four ancestral sequences representing Node-13 showed variation at 13 positions (Figure 4.3A). Similarly, amino acids varied at 18 positions when the four ancestral sequences representing the Node-3 were aligned (Figure 4.3B). Lastly, an alignment of the four ancestral sequences

representing the Node-2, all ancestor of all hydrozoan FPs, had degeneracy at 19 positions (Figure 4.3C).

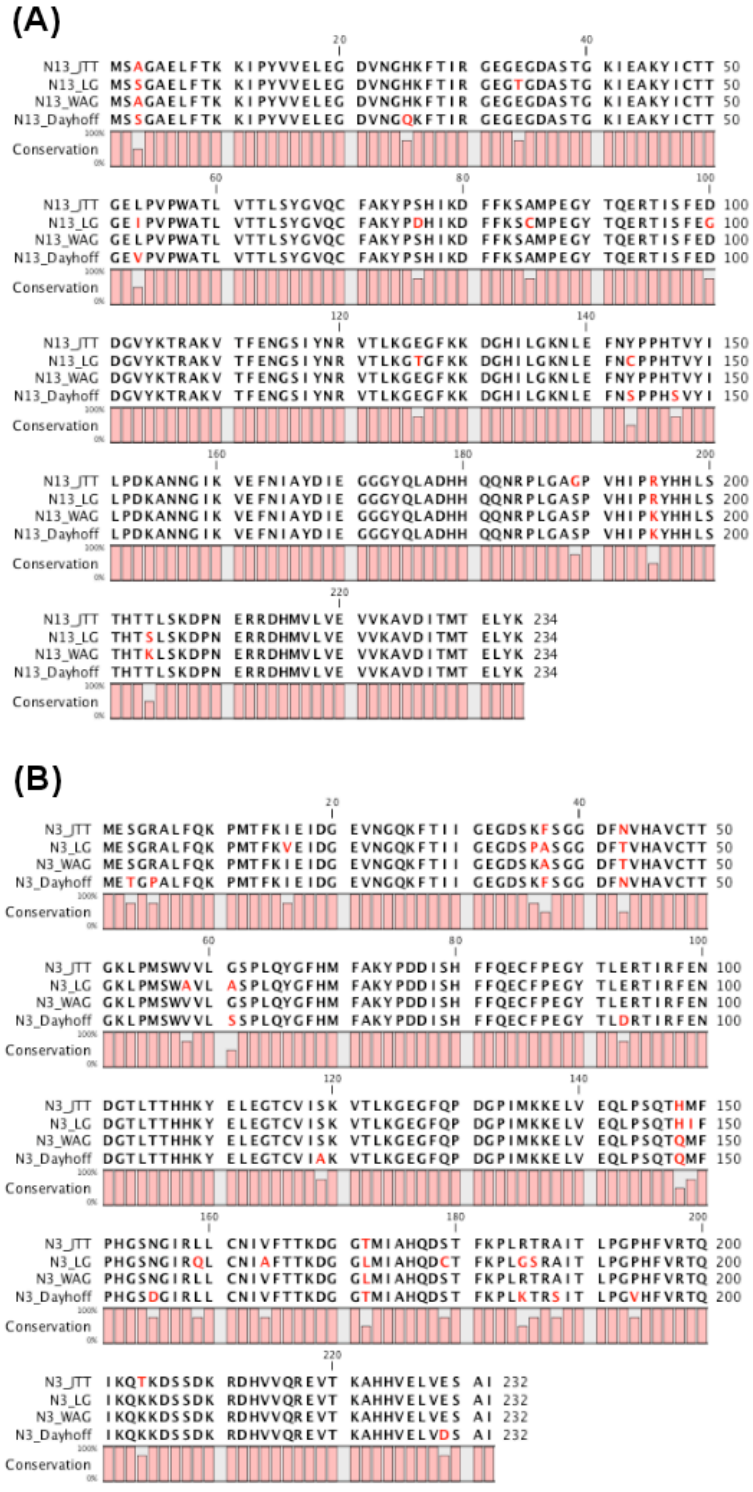


Figure 4.3: Continued.

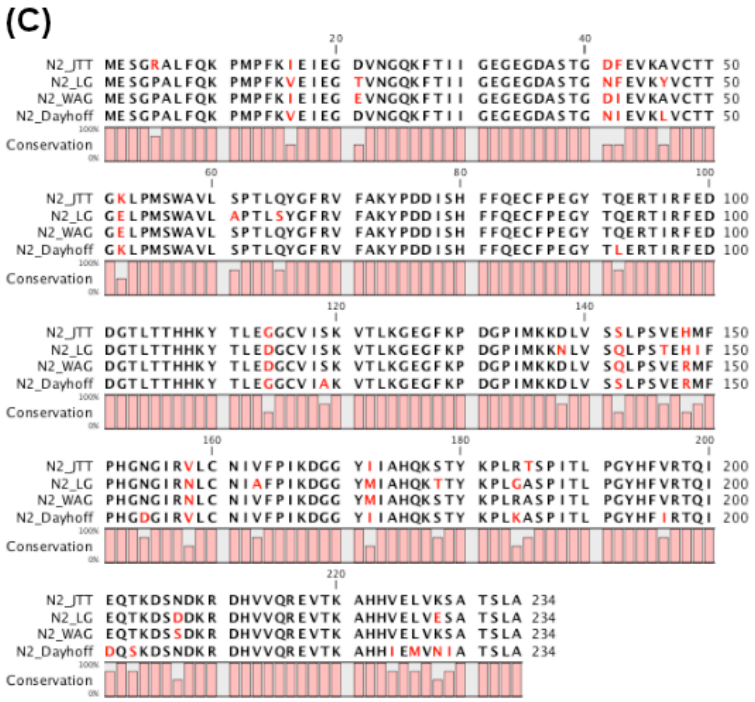


Figure 4.3: Multiple sequence alignment (amino acids) ancestors predicted by four substitution models.

For each of the three ancestral sequences were determined using four different amino acid substitution models. Four different ancestral sequences for each ancestral node are aligned to highlight degenerate positions. **(A)** Node-13 designated as N13 represents the ancestral FP for the Leptothecata clade of the hydrozoan phylogeny. **(B)** Node-3 designated as N3 represents the ancestor FP of the EmberFPs and anm2CP. **(C)** Node-2 designated as N2 represents the ancestral FP of the all hydrozoan FPs.

Degenerate gene synthesis and sequence verification

The sequencing results of the ancestral sequences corresponding to the three nodes showed that the synthetic degenerate genes for each of the three nodes did accommodate the amino acid variability predicted based the four different substitution model. The results also showed that the method used to synthesize the ancestral genes also introduced a small number of unintended mutations. However, the significance of these unintended mutations remains unclear at the present moment.

Upon screening multiple degenerate gene libraries for each of the three ancestral genes for fluorescence, we were able to isolate fluorescent clones only for Node-13. An extensive screening of degenerate gene libraries representing Nodes 2 and 3 yielded only non-fluorescent clones.

Forty fluorescent clones representing Node-13 were sequenced, confirming the planned extent of sequence diversity due to degenerate gene synthesis (Figure 4.4). Of the forty fluorescent clones sequenced we found 28 unique sequences. The library screen also showed a large of variation in brightness of these clones suggesting sequence variability did have some impact on the molecular function of these proteins. Resurrected ancient amino acid sequence representing a range of brightness did not reveal any pattern or assort into specific categories when analyzed by a phylogenetic approach. Overall the results confirmed that the molecular function was not restricted to a very small number of unique sequences, rather the resurrected ancient gene representing Node-13 was able to accommodate a considerable amino acid sequence variability to yield a functional FP.

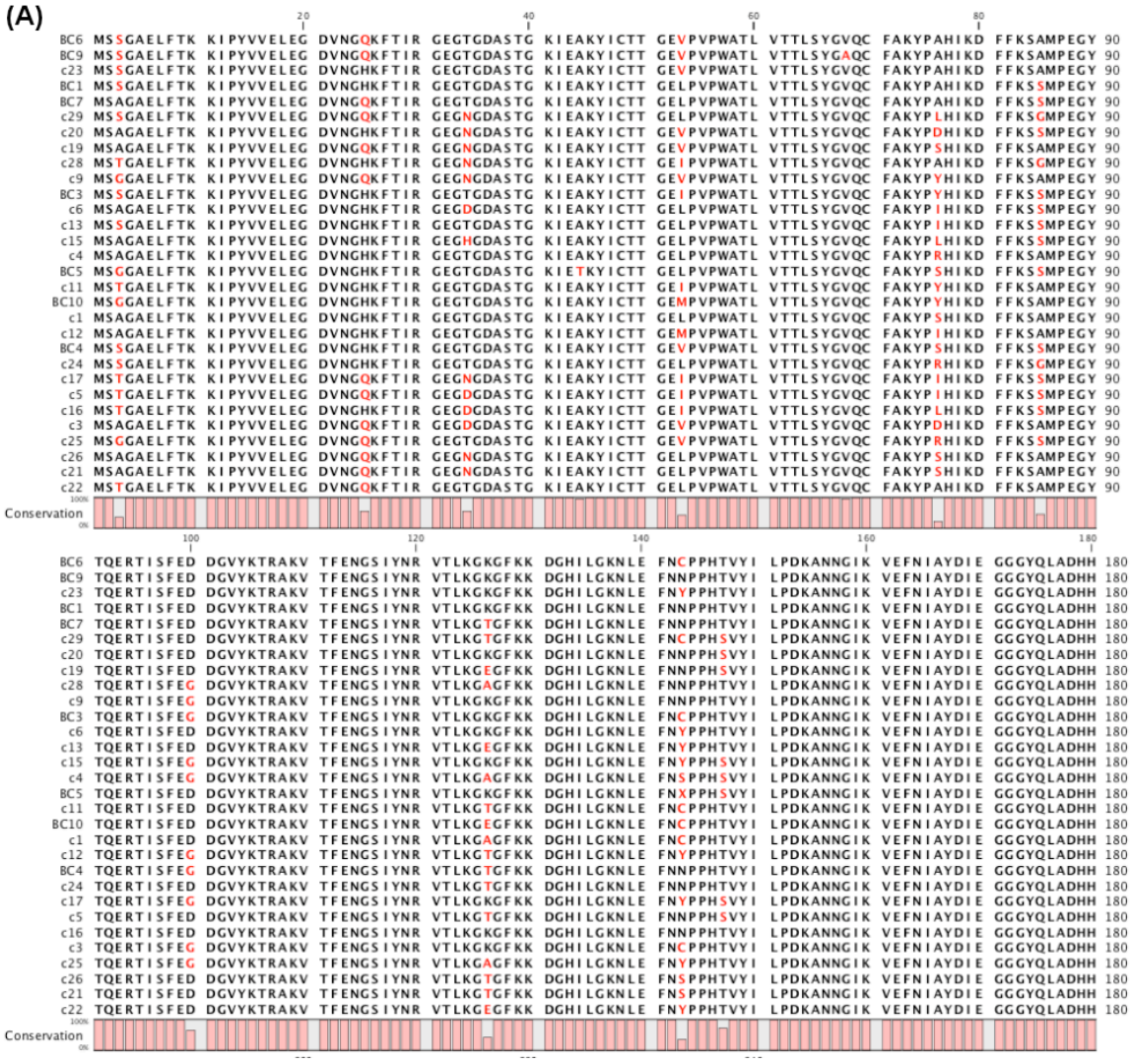
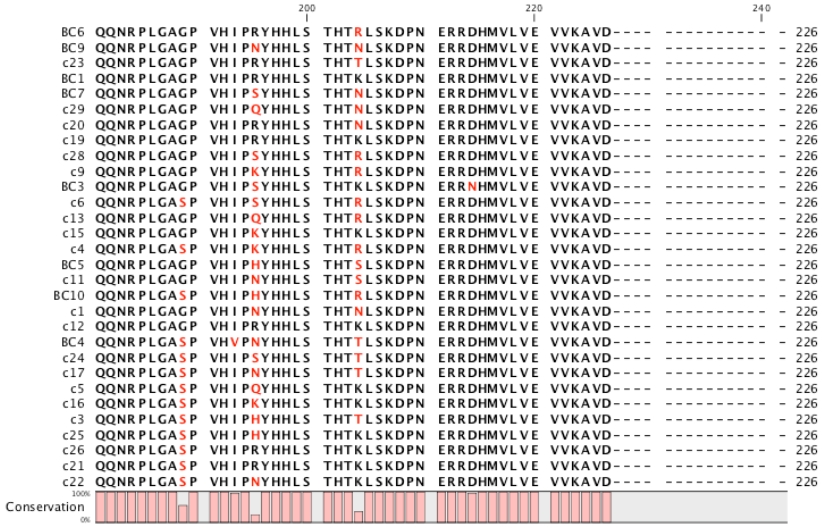


Figure 4.4: Continued.



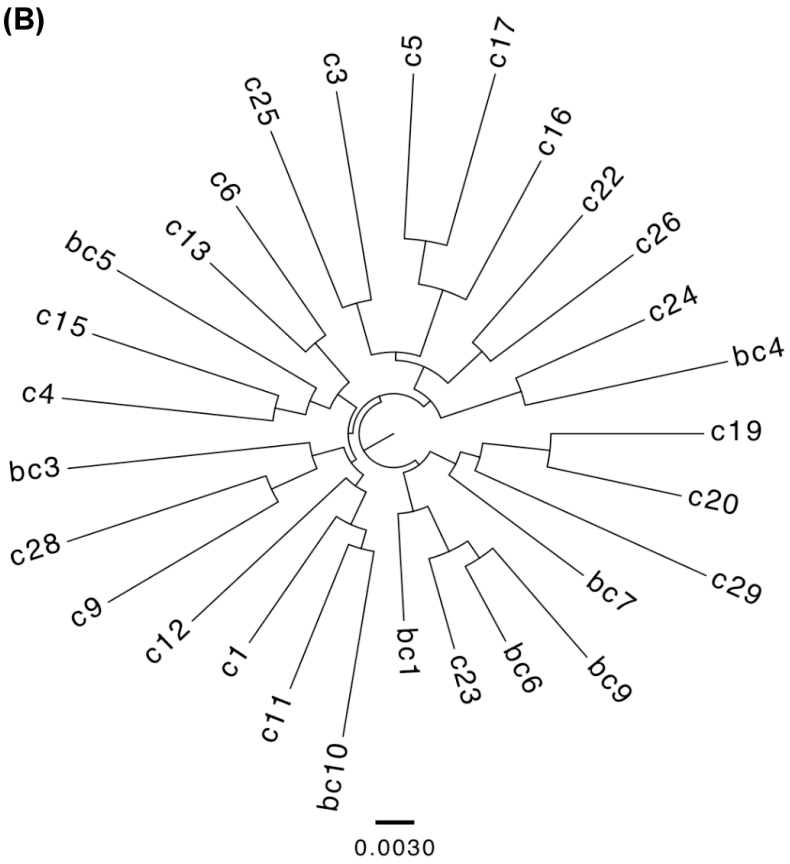


Figure 4.4: Continued.

Figure 4.4: Multiple sequence alignment (amino acids) and the relatedness of fluorescent Node-13 clones.

(A) Amino acid sequence alignment of Node-13 unique clones. Residues in red are shows 13 degenerate positions where residues differ from each other. There are three sequences with single unplanned point mutations. **(B)** A neighbor-joining tree of fluorescent clones from the ancestral FP expression library. Clones labeled "bc" correspond to "bright clones" that were characterized spectroscopically.

Spectroscopic study of the Node-13 clones

The results showed that the degeneracy introduced to account for the uncertainty associated with the prediction of an ancestral gene did not have a major impact on the spectroscopic characteristics of the Node-13 ancestral clones. Despite the sequence differences, each of the twenty-eight unique clones exhibited the same spectroscopic characteristics, with excitation maxima at 400nm and emission maxima at 510nm (Figure 4.5).

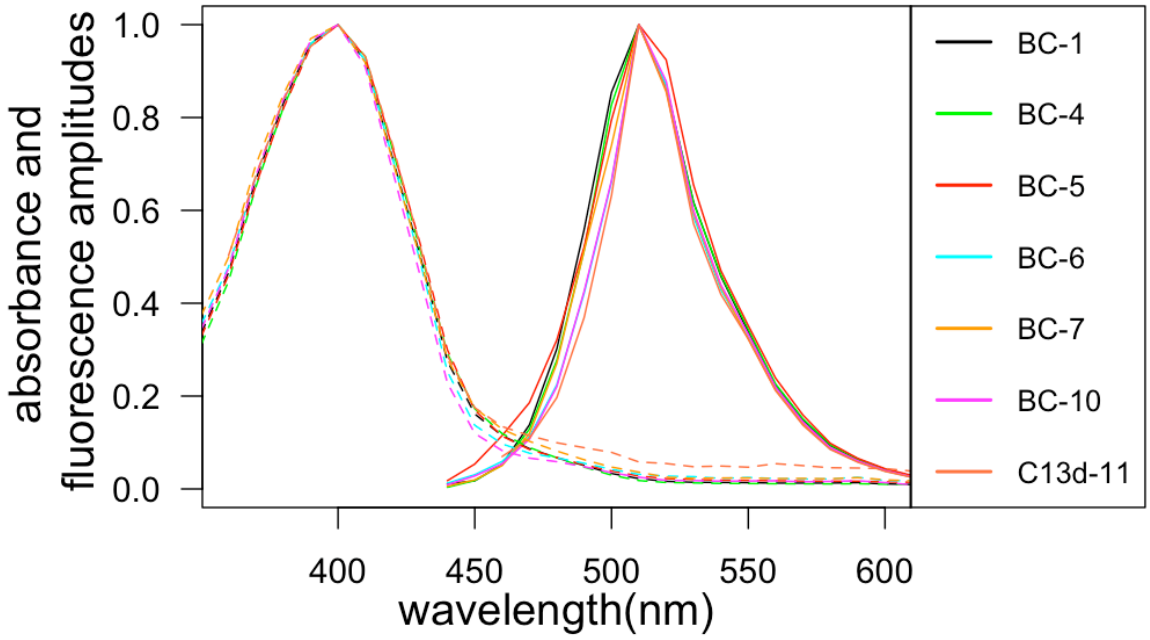


Figure 4.5: Spectroscopic study of the Node-13 brightest fluorescent clones. Spectroscopic properties of selected seven brightest ancestral FP variants designated at "BC". The clones-4, 5 and 7 were bc4, bc5 and bc7 (Figure 4.2), while the clone-11 was isolated during the 1st round of screen (designated as c11 in Figure 4.2). The horizontal axis denotes the wavelength in nanometers. The normalized absorbance profiles are depicted by DASHED CURVES, with maximum absorbance at 400nm. The normalized emission profiles are SOLID CURVES, with emission maxima at 510nm.

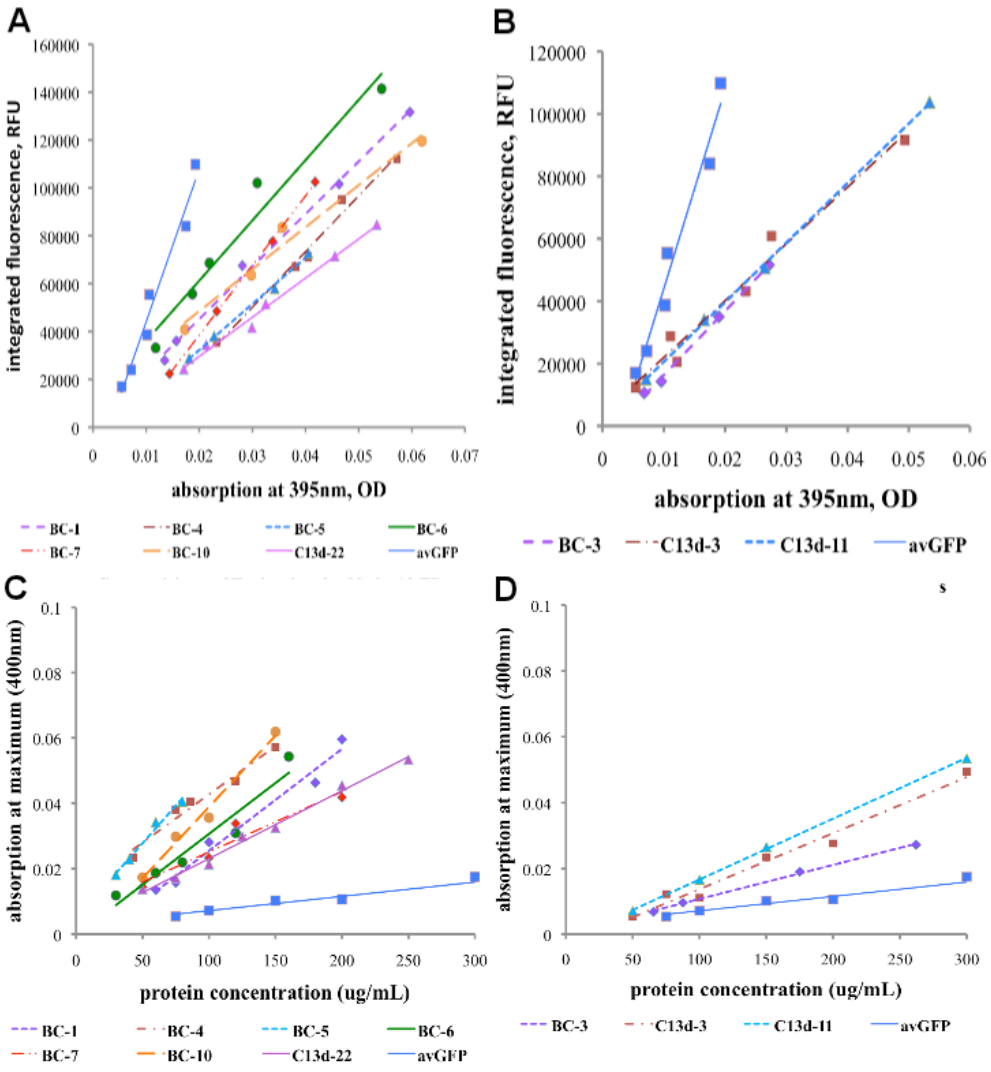


Figure 4.6: Photophysical study of the Node-13 clones.
Quantum Yield (QY) comparisons and Molar Extinction (ME) comparisons between Node-13 clones and the wild-type avGFP. Node-13 clones isolated in the 1st screen are designated as "C," while the bright clones isolated in the 2nd screen are designated as "BC." **(A, B)** X-axis: absorption at 395nm, Y-axis: integrated fluorescence under an emission curve. Comparison of slopes show relative QY; FPs with higher slopes have a higher QY with respect to ME of the other FPs. QY analysis shows that avGFP has highest QY in comparison to the all ancestral clones. **(C, D)** X-axis: protein concentration (μg/mL), Y-axis: absorption at 395nm. Comparison of slopes show relative ME; FPs with higher slopes have a higher ME with respect to the other FPs. ME analysis shows that avGFP has lowest ME in comparison to the ME of all ancestral clones.

Even though these close exhibited similar excitation and emission spectra, the range of brightness varied from very bright to very dim among these Node-13 ancient clones. The differences in relative quantum yield (QY) and relative molar extinction coefficient (ME) of these twenty-eight ancient clones was the likely explanation for the observed range of brightness (Figure 4.6 A, B). The results showed that relative to the QY of wild-type *Av*GFP (slope: $6x10^{6}$), all clones studied here had poor QY (Figure 4.6A; Table 4.2). However, the relative ME of each of these ancient clones was higher than the ME of wild-type *Av*GFP (slope: $4x10^{-5}$)(Figure 4.6B; Table 4.2). Since the product of QY and ME determines the brightness of a FP, some of the ancient clones were as bright as the wild-type *Av*GFP. The brightest clone, Node-13 BC-5, was five-fold brighter than the dimmest clone, Node-13 BC-3, and 4.2-fold brighter than the wild-type *Av*GFP (Table 4.2).

Table 4.2: Relative brightness of the Node-13 ancestral clones.

Fluorescent proteins	QY slope (RFU/OD at 395nm)	ME slope (Abs.395nm/(ug/mL)	FP brightness = (QY * ME)	Normalized brightness
Wild-type avGFP	$6x10^{6}$	$4x10^{-5}$	240	1
Node-13 BC-1	$2x10^{6}$	$3x10^{-4}$	600	2.5
Node-13 BC-3	$2x10^{6}$	$1x10^{-4}$	200	0.83
Node-13 BC-4	$2x10^{6}$	$3x10^{-4}$	600	2.5
Node-13 BC-5	$2x10^{6}$	$5x10^{-4}$	1000	4.167
Node-13 BC-6	$3x10^{6}$	$3x10^{-4}$	900	3.75
Node-13 BC-7	$3x10^{6}$	$2x10^{-4}$	600	2.5
Node-13 BC-10	$2x10^{6}$	$4x10^{-4}$	800	3.33
Node-13 C-3	$2x10^{6}$	$2x10^{-4}$	400	1.67
Node-13 C-11	$2x10^{6}$	$2x10^{-4}$	400	1.67
Node-13 C-22	$2x10^{6}$	$2x10^{-4}$	400	1.67

Non-denaturing SDS-PAGE study of the Node-13 clones shows variation in the band patterns

Each of the brighter ancient Node-13 clones had higher expression; presumably both of these characteristics were due to better stability of these FPs, were purified and analyzed by non-denaturing and denaturing SDS-PAGE. Surprisingly, the ancestral clones showed difference in their migration patterns under non-denaturing condition suggesting influence of the small differences in their amino acid composition (Figure 4.7 A, B, C, D). Denaturing SDS-PAGE results showed that the ancestral clones had the molar mass of ~27kDa (results not shown).

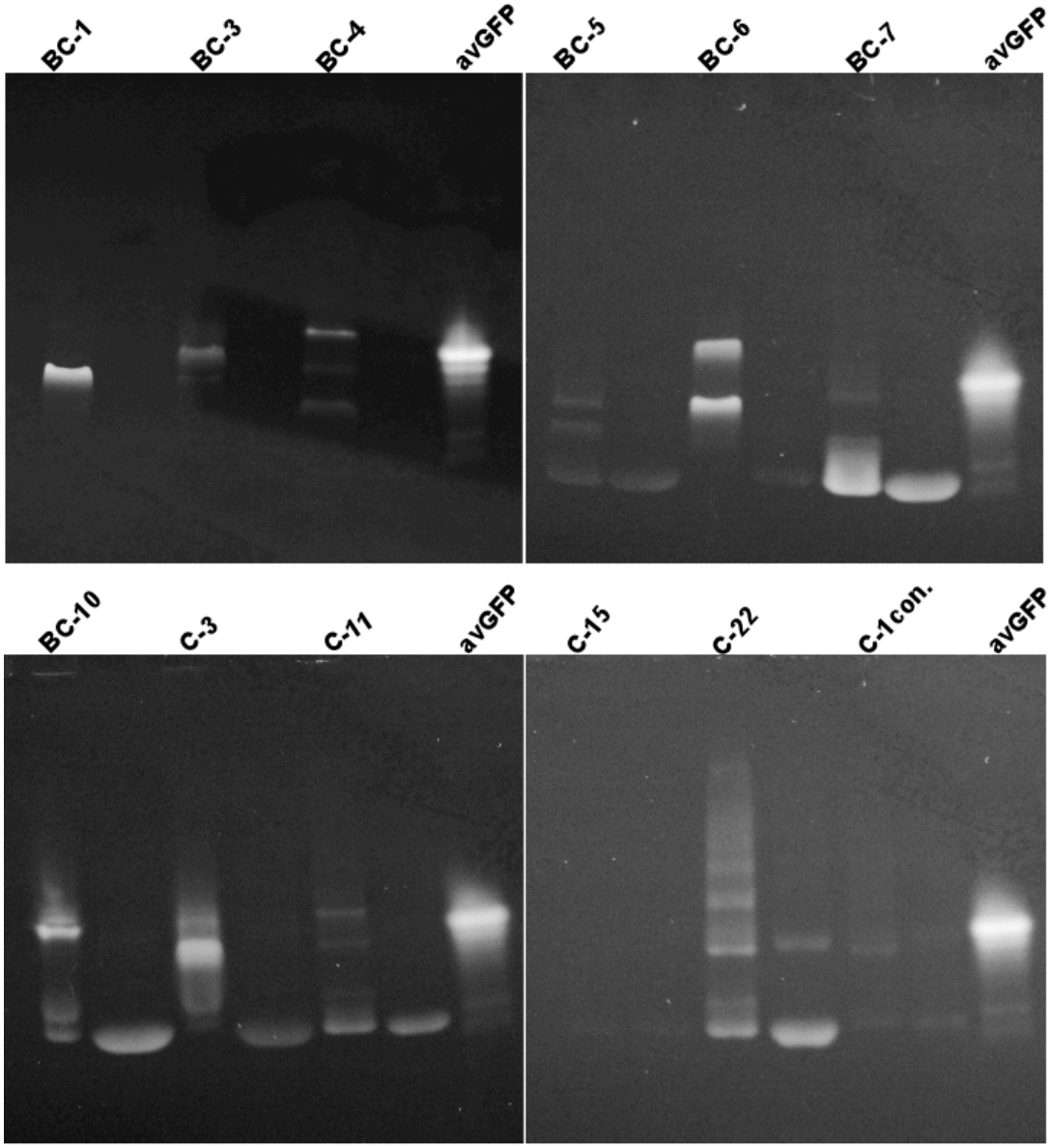


Figure 4.7: Continued.

Figure 4.7: Non-denaturing SDS-PAGE analysis of the select Node-13 ancestral clones.

Fluorescence of native FPs was detected using UV-B excitation. Wild-type Aequorea Victoria GFP (avGFP) was used as control to assess the relative bad migration of non-denatured FP clones. The clones isolated in the 1st screen are designated with "C," and the bright clones designed by "BC" were isolated in the 2nd screen. A consensus sequence clone created based the four amino acid substitution model is designated as C-1con. Band migration patterns showing all the three bright clones, BC-1, 3, and 4, migrated about same distance relative to the avGFP band. Small sequence differences between the clones appear to impact the band migration patterns. The three bright clones, BC-5, 6, and 7, migrated different distances, and there are multiple bands present with different intensities. Observed low band intensities of BC-5 could be due to destabilizing effect of SDS on the FP. The band patterns show that two FPs, BC-10 and C-3, have major bands with similar migration as that of avGFP. All three clones, BC-10, C-3 and C-11, have multiple bands. Of the three clones, bands for two clones, C-22 and C-1con., are visible, with multiple bands. Destabilizing effect of SDS on the FP, C-15, is a likely reason for lack of fluorescence detection.

Oligomeric states of Node-13 clones determined by Multi-angle laser light scattering analysis

Purified Node-13 ancient clones analyzed by SEC-MALLS showed that the majority these ancient clones existed as a monomer (molar mass: $2.7x10^4$ g/mol) (Figure 4.8A). However, the study also discovered that several clones had noteworthy multiple oligomeric states, i.e. Node-13 BC-10 molar masses: *1.701 x 10^5 (tetramer), 5.335 x 10^4 (dimer), 3.190 x 10^6 (monomer)* (Figure 4.8B). The cumulative weight fraction analysis confirmed that the monomeric state was the predominant oligomeric state (Figure 4.9). The data also showed that the small fraction of FP from these ancient clones also formed higher order oligomers (Table 4.3).

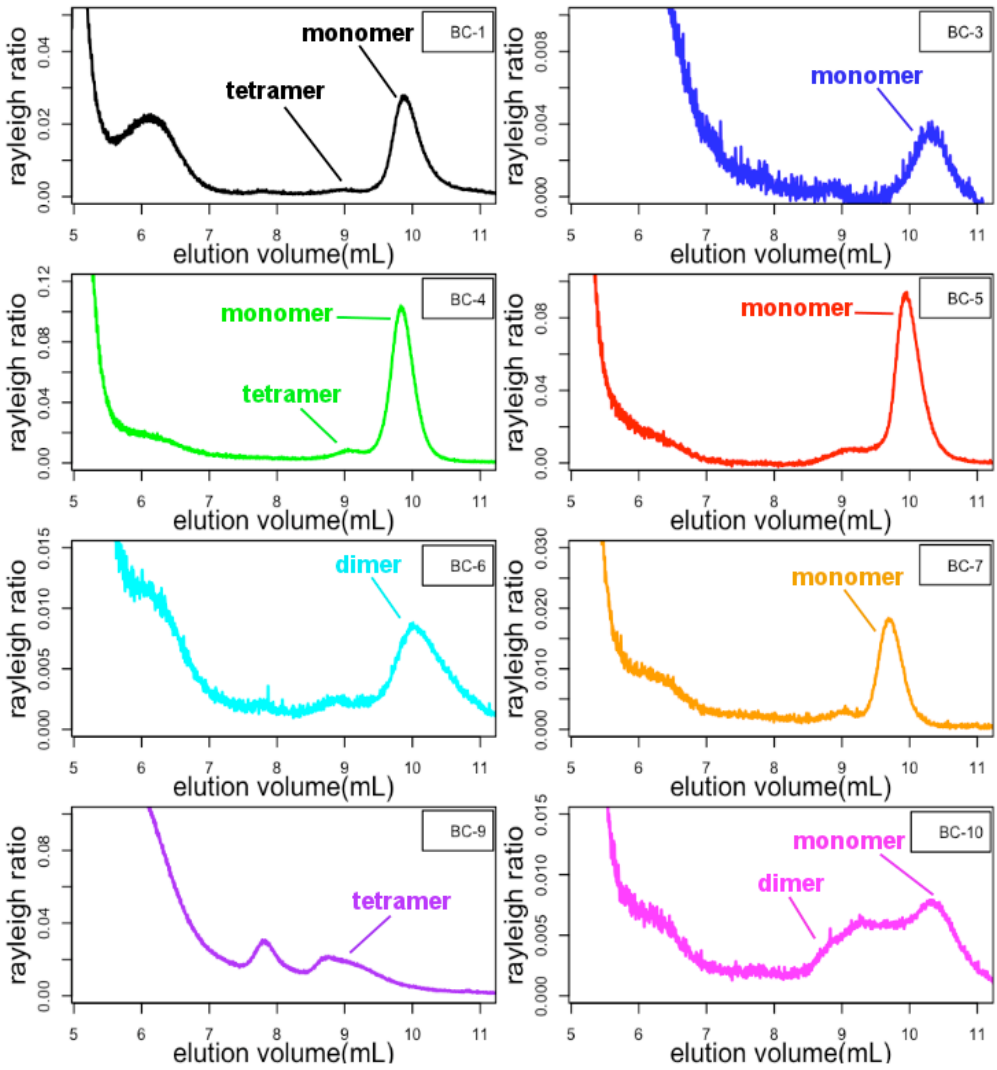


Figure 4.8: Molar mass and oligomeric state determinations of the select Node-13 clones.

Oligomerization analysis of the selected eight Node-13 ancestral FP clones by a tandem size-exclusion HPLC and light scattering (SEC-MALLS). These selected clones were designated as "BC" for bright clone. Horizontal axis is the elution volume (mL), left, vertical axis is the Rayleigh ratio (the *ratio* of intensities of incident and scattered light at a specified distance; left vertical axis), and right vertical axis is the estimated molar mass corresponding to thick lines above each curve. Seven of the eight clones show the presence of monomeric oligomeric state. BC-9 was observed as a canonical tetramer and not as the monomer, which was the predominant state observed for the other clones. Higher order oligomeric states were observed in all eight clones.

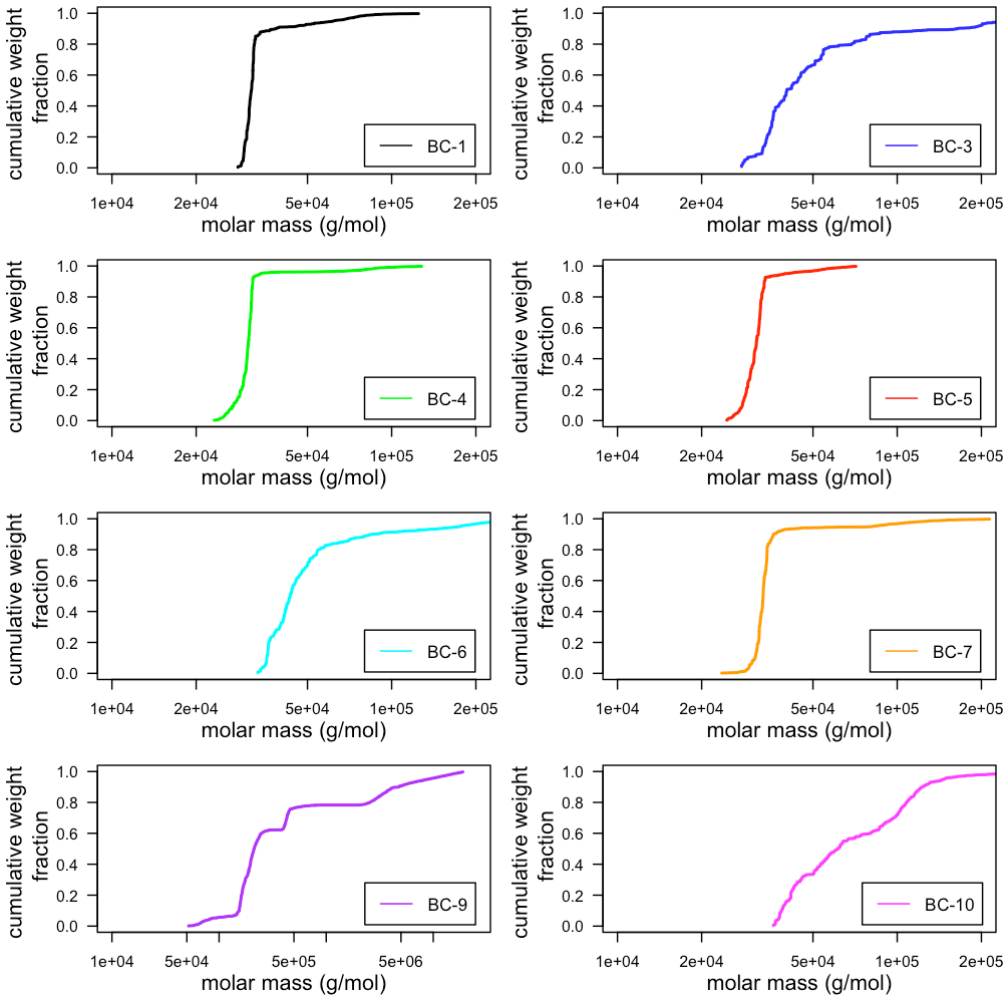


Figure 4.9: Oligomeric state stiochiometry analysis by of the select Node-13 clones by the cumulative weight fraction.
Oligomeric states of selected clones analysis using the MALLS data. Eight Node-13 ancestral clones designed as the bright clones ("BC") elucidating the cumulative weight fraction showing the monomeric oligomeric state as the predominant state. X-axis: molar mass (g/mol), Y-axis: cumulative sample fractions. BC-1, BC-4, BC-5 and BC-7 are predominantly a monomer, consisting of approximately 90% of the total protein. These four clones also have small fractions consisting of the canonical tetramer and higher order oligomer. The BC-3, BC-6 and BC-10 are monomers, with a considerable presence of other oligomeric states: dimer, tetramer and higher order oligomers. Predominant BC-9 fraction is a tetramer, which also contains small amount of smaller and higher order oligomers. FP.

Table 4.3: Molar masses of the Node-13 ancestral clones by SEC-MALLS.

Fluorescent proteins	**Peak-1 Molar mass (g/mol) (% error) Oligo. state**	**Peak-2 Molar mass (g/mol) (% error) Oligo. state**	**Peak-3 Molar mass (g/mol) (% error) Oligo. state**	**Peak-4 Molar mass (g/mol) (% error) Oligo. state**	**Peak-5 Molar mass (g/mol) (% error) Oligo. state**
Node-13 BC-1	1.418×10^5 (5%) tetramer	2.749×10^4 (1.3/%) monomer			
Node-13 BC-3	3.977×10^5 (2.7%)	1.650×10^5 (5%)	3.942×10^4 (2.6%) monomer		
Node-13 BC-4	2.494×10^5 (2.1%)	9.656×10^4 (1.9%) tetramer	2.789×10^6 (1.5%) monomer		
Node-13 BC-5	7.161×10^5 (1.3%)	1.586×10^5 (6%)	2.988×10^4 (1.0%) monomer		
Node-13 BC-6	5.501×10^5 (1.4%)	1.72×10^5 (1.8%)	5.052×10^4 (1.3%) dimer		
Node-13 BC-7	2.694×10^5 (3%)	3.277×10^4 (1.1%) monomer			

Table 4.3: Continued.

Node-13 BC-9	9.03×10^6	3.01×10^6	4.64×10^5	2.02×10^5	1.104×10^5 tetramer
Node-13 BC-10	2.849×10^6 (1.9%)	1.701×10^5 (6%)	7.653×10^4 (3%)	5.335×10^4 (10%) dimer	3.190×10^6 (2.4%) monomer
Node-13 C-3	3.22×10^5 (16%)				
Node-13 C-11	4.116×10^6 (4%)	6.889×10^5 (9%)	1.171×10^5 (11%) tetramer	2.883×10^4 (12%) monomer	
Node-13 C-22	1.094×10^5 (1.5%) tetramer	5.563×10^4 (2.1%) dimer			

DISCUSSION

Over the past several years we have described several novel fluorescent proteins (FPs) from Hydrozoa representatives possessing unusual oligomeric structures (Figure 4.2; Hunt et al., 2012; Aglyamova et al., 2011). A large diversity of oligomeric states in currently known hydrozoan FPs prompted the present study of the evolution of FP oligomerization with the goal of gaining an insight into the molecular basis for the diversity of oligomeric states (Finnigan et al., 2012). This knowledge could help in

designing novel monomeric fluorescent genetically encoded labels and would advance our general understanding of evolution of protein-protein interactions.

Based on the phylogenetic relationships between the extant FPs and the diversity of oligomeric states of these FPs, here we reconstructed three ancestral FPs within the Hydrozoan clade of the GFP-family of proteins to study the diversity oligomeric states in these extant FPs. Of the three ancestral genes, only the common ancestor of *Aequorea victoria* GFP and *Obelia* FPs resulted in a fluorescent FP. Sequenced clones revealed the planned extent of sequence diversity was able to address some of the uncertainty associated with the ancestral sequence predictions (Figures 4.3, 4.4); however, regardless of these sequence differences, all of the ancestral FPs demonstrated identical spectroscopic characteristics (Figure 4.5), suggesting that the reconstructed ancestral FP must harbor its chromophore in neutral ground state, similar to the GFP from *Aequorea victoria* (Ward and Bokman, 1982; Chattoraj et al., 1996; Brejc et al., 1997). The molar extinction (ME) and quantum yield (QY) analysis of the highly fluorescent Node-13 variants suggested that the observed differences in these values between these clones could be attributed to the small differences in their amino acid sequence influencing the chromophore environment (Figure 4.6; Bomati et al., 2014; Ong et al., 2011; Shaner et al., 2004; Heim and Tsein, 1996; Cambell et al., 2002, Pakhomov et al., 2011; Pletneva et al., 2007; Remington et al., 2005). Furthermore, the data showed that the majority of these ancient clones were brighter than wild-type *Av*GFP.

An oligomeric state analysis by size exclusion chromatography coupled with light scattering (SEC-MALLS) of the Node-13 ancestral variants isolated from highly fluorescent bacterial colonies indicated that they are monomeric: based on light scattering data, these highly expressing variants have approximate molar mass 27kDa (Figures 4.8A, 4.9). These results were corroborated by data from SDS-PAGE under denaturing

conditions. Furthermore, our observation of predominantly monomeric oligomeric state by SEC-MALLS suggested that the small amino acid sequence differences in these ancient clones were responsible for the observed variation in the band migration pattern of these FPs under non-denaturing conditions (Figure 4.7). Ancestral proteins purified from less bright bacterial colonies demonstrated a pronounced signal for higher order oligomers (Figures 4.8B, 4.9). The link between higher oligomerization and lower heterologous expression is expectable since oligomerizing proteins are typically more toxic to the expressing bacterial cells. This result indicates that within the library of ancestral sequences certain sequence variants result in changes in oligomeric state, raising a possibility of an association study as in (Field and Matz, 2010) to identify structural determinants of oligomerization. Furthermore, observed differences in oligomeric states of these ancient clones due to the small amount of variation in their amino acid sequences suggested that there maybe a long-range interaction between some of these degenerate sites and with the residues responsible for FP oligomerization.

Ancestral FPs corresponding to the two other nodes (Nodes-2 and 3) were also synthesized and cloned successfully; however, upon extensive screening we have not been able to identify any fluorescent clones, though we were able to verify the presence of the target sequences in our library through Sanger sequencing and SDS-PAGE. These results confirm our initial apprehensions with the predicted genes at these two nodes, because both of these genes contained close to 20 degenerate sites. In the future, using different amino acid substitution models and introducing less degenerate sights on these two nodes could potentially lead to these ancestral proteins being resurrected successfully.

Ancestral gene prediction by statistical phylogenetics assumes each residue evolves independently and does not account for interactions between the residues. Not

accounting for the impact residue-residue interactions over the course of protein evolution creates certain uncertainty with the prediction of ancestral gene (Pollock et al., 2012; Williams et al., 2006; Gruenheit et al., 2008) Here by using four different amino acid substitution models and degenerate gene synthesis we attempted to account for the uncertainty associated with the ancestral gene prediction. The results of this study raises a possibility that Leptothecata FP oligomerization states are likely impacted by just a small number of mutational changes, and same changes are simultaneously impacting ME and QY of these proteins. By using four substitution models and degenerate gene synthesis we have a set of candidate residues for a future study in the functional ancestor that are either individually or through interactions with other residues impacting the structure and function.

Hydrozoan FPs have been known to associate with chemiluminescent proteins forming a hetro-oligomer. Leptothecata extant FPs, *Aequorea victoria* GFP, *Clytia*GFP and *Obelia* FPs (Figures 4.1, 4.2; Gorokhovatsky et al., 2004; Morise et al., 1974; Inouye and Tsuji, 1993; Inouye, 2008; Illarionov et al., 1995) are excited through interactions with bioluminescent proteins with emissions at 450-460 nm. In light of our observation of 395nm excitation of the ancestral FP, the presence of dual excitation peaks at 395nm and 475nm in *Av*GFP suggests that appearance of 475nm excitation peak is an evolutionary event since the divergence from the common ancestor of Leptothecata (Ward et al, 1982; Morise et al., 1974). Similarly, two of the *Obelia* FPs, *Obe*GFP and *Obe*YFP have excitation peaks at 502nm and 514nm, respectively, suggesting evolutionarily these molecular phenotypes appeared since the common ancestor of Letothecata (Aglyamova et al., 2011). On the other hand, *Obe*CFP has an excitation peak at 400nm, an evolutionarily conserved molecular phenotype corresponding to the ancestral excitation peak at 395nm (Aglyamova et al., 2011). Our observation here raises a possibility that the function of

the ancestral Leptothecata might not be associated with color conversion in bioluminescence as its near-UV absorption characteristics are incompatible.

The model system developed in this study opens up a distinct prospect of studying the evolution of hetro and homo oligomerization, and examination of the underlying structure-function changes over the course of evolution. The results presented here suggest the importance of analyzing the biophysical and biochemical properties of resurrected ancestral variants in revealing candidate residues. Most importantly, this research project can serve as a model for future investigations to elucidate the role of evolutionary mechanisms involved homo-oligomerization of proteins. The model presented here for the study of FP of oligomerization could prove valuable in enhancing our ability to engineer many more novel, genetically encoded monomeric biomarkers needed for the biomedical fields. Finally, the green ancient FP reported here is not only the first ancient bio-marker with desirable properties such as monomeric oligomeric state and a large stoke shift, but adds the ever growing color palette of novel bio-marker wanted in various fields of biomedical research.

CONCLUSION

Taking advantage of the diversity of homo-oligomeric states represented in the hydrozoan clade we reconstructed one functional of three ancestors attempted. The functional ancestor of Leptothecata FPs provides an opportunity to study the evolution of homo-oligomerization. The synthetic ancestor has green fluorescence and a large stokes shift. Moreover, the ancient FP was a monomer as oppose to the expected tetramer, suggesting that the diversity of oligomeric states observed in the Leptothecata clade must have evolved recently. We also showed that the evolution of dual excitation peak

observed in the *Aequorea victoria* GFP must be a novel biochemical event since the divergence of Leptothecata as the secondary excitation peak at 475nm was absent in the ancestral FP (Morise et al., 1974; Ward et al., 1982). Use of multiple amino acid models and degenerate ancestral gene synthesis enabled us to not only discover the dominant functional and structural phenotype of the ancestral FP, but we also discovered the diversity of molar extinction, quantum yield and oligomeric states associated with the ancestral FP, providing a set of candidate residues involved in the evolution of oligomerization and functional characteristics of these FPs. In future studies, the candidate residues indentified based on the functional and structural diversity arising from our approach to ancestral FP reconstruction could provide insight into the evolution of Leptothecata FPs and the evolution of homo-oligomerization in general. Finally, the monomeric green ancestral FP with a large stokes shift synthesized in this study was the first ancient bio-marker with the desirable properties required for the biomedical research.

References

1. R. Abagyan, M. Totrov, Biased probability Monte Carlo conformational searches and electrostatic calculations for peptides and proteins. *J Mol Biol* 235, 983-1002 (1994).

2. G. V. Aglyamova, M. E. Hunt, C. K. Modi, M. V. Matz, Multi-colored homologs of the green fluorescent protein from hydromedusa Obelia sp. *Photochem Photobiol Sci* 10, 1303-1309 (2011).

3. M. H. Ali, B. Imperiali, Protein oligomerization: how and why. *Bioorg Med Chem* 13, 5013-5020 (2005).

4. R. Ando, H. Hama, M. Yamamoto-Hino, H. Mizuno, A. Miyawaki, An optical marker based on the UV-induced green-to-red photoconversion of a fluorescent protein. *Proc Natl Acad Sci U S A* 99, 12651-12656 (2002).

5. H. Ashkenazy *et al.*, FastML: a web server for probabilistic reconstruction of ancestral sequences. *Nucleic Acids Res* 40, W580-584 (2012).

6. D. D. Banks, L. M. Gloss, Equilibrium folding of the core histones: the H3-H4 tetramer is less stable than the H2A-H2B dimer. *Biochemistry* 42, 6827-6839 (2003).

7. D. P. Barondeau, C. D. Putnam, C. J. Kassmann, J. A. Tainer, E. D. Getzoff, Mechanism and energetics of green fluorescent protein chromophore synthesis revealed by trapped intermediate structures. *Proc Natl Acad Sci U S A* 100, 12111-12116 (2003).

8. J. D. Bloom, F. H. Arnold, C. O. Wilke, Breaking proteins with mutations: threads and thresholds in evolution. *Mol Syst Biol* 3, 76 (2007).

9. J. D. Bloom, A. Raval, C. O. Wilke, Thermodynamics of neutral protein evolution. *Genetics* 175, 255-266 (2007).

10. J. D. Bloom, P. A. Romero, Z. Lu, F. H. Arnold, Neutral genetic drift can alter promiscuous protein functions, potentially aiding functional evolution. *Biol Direct* 2, 17 (2007).

11. E. K. Bomati, J. E. Haley, J. P. Noel, D. D. Deheyn, Spectral and structural comparison between bright and dim green fluorescent proteins in Amphioxus. *Sci Rep* 4, 5469 (2014).

12. M. S. Breen, C. Kemena, P. K. Vlasov, C. Notredame, F. A. Kondrashov, Epistasis as the primary factor in molecular evolution. *Nature* 490, 535-538 (2012).

13. K. Brejc *et al.*, Structural basis for dual excitation and photoisomerization of the Aequorea victoria green fluorescent protein. *Proc Natl Acad Sci U S A* 94, 2306-2311 (1997).

14. J. T. Bridgham, S. M. Carroll, J. W. Thornton, Evolution of hormone-receptor complexity by molecular exploitation. *Science* 312, 97-101 (2006).

15. R. Burri, N. Salamin, R. A. Studer, A. Roulin, L. Fumagalli, Adaptive divergence of ancient gene duplicates in the avian MHC class II beta. *Mol Biol Evol* 27, 2360-2374 (2010).

16. R. E. Campbell *et al.*, A monomeric red fluorescent protein. *Proc Natl Acad Sci U S A* 99, 7877-7882 (2002).

17. J. A. Capra, M. Singh, Characterization and prediction of residues determining protein functional specificity. *Bioinformatics* 24, 1473-1480 (2008).

18. C. W. Carter *et al.*, The Rodin-Ohno hypothesis that two enzyme superfamilies descended from one ancestral gene: an unlikely scenario for the origins of translation that will not be dismissed. *Biol Direct* 9, 11 (2014).

19. S. Chakrabarti, C. J. Lanczycki, Analysis and prediction of functionally important sites in proteins. *Protein Sci* 16, 4-13 (2007).

20. M. Chattoraj, B. A. King, G. U. Bublitz, S. G. Boxer, Ultra-fast excited state dynamics in green fluorescent protein: multiple states and proton transfer. *Proc Natl Acad Sci U S A* 93, 8362-8367 (1996).

21. B. S. Chen, S. J. Ho, The stochastic evolutionary game for a population of biological networks under natural selection. *Evol Bioinform Online* 10, 17-38 (2014).

22. D. M. Chudakov, M. V. Matz, S. Lukyanov, K. A. Lukyanov, Fluorescent proteins and their applications in imaging living cells and tissues. *Physiol Rev* 90, 1103-1163 (2010).

23. J. C. Cox, J. Lape, M. A. Sayed, H. W. Hellinga, Protein fabrication automation. *Protein Sci* 16, 379-390 (2007).

24. M. Dai *et al.*, The creation of a novel fluorescent protein by guided consensus engineering. *Protein Eng Des Sel* 20, 69-79 (2007).

25. A. K. Datta, Comparative sequence analysis in the sialyltransferase protein family: analysis of motifs. *Curr Drug Targets* 10, 483-498 (2009).

26. E. R. Dayhoff MO, Park CM, in *Atlas of protein sequence and structure,* M. Dayhoff, Ed. (National Biomedical Research Foundation., Silver Spring, Md., 1972), vol. 5, pp. 89-99.

27. S. R. Dayhoff MO, Orcutt B., in *Protein segment dictionary 78: from the Atlas of protein sequence and structure, volume 5, and supplements 1, 2, and 3,* M. Dayhoff, Ed. (National Biomedical Research Foundation; Georgetown University Medical Center, Silver Spring, Md. Washington, D.C., 1978), vol. 5, pp. 345–352.

28. E. Dellus-Gur, A. Toth-Petroczy, M. Elias, D. S. Tawfik, What makes a protein fold amenable to functional innovation? Fold polarity and stability trade-offs. *J Mol Biol* 425, 2609-2621 (2013).

29. M. A. DePristo, D. M. Weinreich, D. L. Hartl, Missense meanderings in sequence space: a biophysical view of protein evolution. *Nat Rev Genet* 6, 678-687 (2005).

30. S. Dey, A. Pal, P. Chakrabarti, J. Janin, The subunit interfaces of weakly associated homodimeric proteins. *J Mol Biol* 398, 146-160 (2010).

31. J. E. Donald, E. I. Shakhnovich, SDR: a database of predicted specificity-determining residues in proteins. *Nucleic Acids Res* 37, D191-194 (2009).

32. T. F. Fagan, Y. Ohmiya, J. R. Blinks, S. Inouye, F. I. Tsuji, Cloning, expression and sequence analysis of cDNA for the Ca(2+)-binding photoprotein, mitrocomin. *FEBS Lett* 333, 301-305 (1993).

33. S. F. Field, M. Y. Bulina, I. V. Kelmanson, J. P. Bielawski, M. V. Matz, Adaptive evolution of multicolored fluorescent proteins in reef-building corals. *J Mol Evol* 62, 332-339 (2006).

34. S. F. Field, M. V. Matz, Retracing evolution of red fluorescence in GFP-like proteins from Faviina corals. *Mol Biol Evol* 27, 225-233 (2010).

35. G. C. Finnigan, V. Hanson-Smith, T. H. Stevens, J. W. Thornton, Evolution of increased complexity in a molecular machine. *Nature* 481, 360-364 (2012).

36. J. A. Gerlt, P. C. Babbitt, Enzyme (re)design: lessons from natural evolution and computation. *Curr Opin Chem Biol* 13, 10-18 (2009).

37. L. I. Gong, J. D. Bloom, Epistatically interacting substitutions are enriched during adaptive protein evolution. *PLoS Genet* 10, e1004328 (2014).

38. A. Y. Gorokhovatsky *et al.*, Fusion of Aequorea victoria GFP and aequorin provides their Ca(2+)-induced interaction that results in red shift of GFP absorption and efficient bioluminescence energy transfer. *Biochem Biophys Res Commun* 320, 703-711 (2004).

39. N. Gruenheit, P. J. Lockhart, M. Steel, W. Martin, Difficulties in testing for covarion-like properties of sequences under the confounding influence of changing proportions of variable sites. *Mol Biol Evol* 25, 1512-1520 (2008).

40. N. Guex, M. C. Peitsch, SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis* 18, 2714-2723 (1997).

41. N. G. Gurskaya *et al.*, A colourless green fluorescent protein homologue from the non-fluorescent hydromedusa Aequorea coerulescens and its fluorescent mutants. *Biochem J* 373, 403-408 (2003).

42. B. Hall *et al.*, Design, synthesis, and amplification of DNA pools for in vitro selection. *Curr Protoc Mol Biol* Chapter 24, Unit 24.22 (2009).

43. M. J. Harms, J. W. Thornton, Analyzing protein structure and function using ancestral gene reconstruction. *Curr Opin Struct Biol* 20, 360-366 (2010).

44. K. Hashimoto, H. Nishi, S. Bryant, A. R. Panchenko, Caught in self-interaction: evolutionary and functional mechanisms of protein homooligomerization. *Phys Biol* 8, 035007 (2011).

45. I. Hayashi *et al.*, Crystallographic evidence for water-assisted photo-induced peptide cleavage in the stony coral fluorescent protein Kaede. *J Mol Biol* 372, 918-926 (2007).

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

47. R. Huang *et al.*, Enzyme functional evolution through improved catalysis of ancestrally nonpreferred substrates. *Proc Natl Acad Sci U S A* 109, 2966-2971 (2012).

48. J. P. Huelsenbeck, F. Ronquist, MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* 17, 754-755 (2001).

49. M. E. Hunt *et al.*, Multi-domain GFP-like proteins from two species of marine hydrozoans. *Photochemical & Photobiological Sciences* **11**, 637-644 (2012).

50. I. G. Ianushevich *et al.*, [Spectral diversity among the members of the family of Green Fluorescent Protein in hydroid jellyfish (Cnidaria, Hydrozoa)]. *Bioorg Khim* 31, 49-53 (2005).

51. R. P. Ilagan *et al.*, A new bright green-emitting fluorescent protein--engineered monomeric and dimeric forms. *FEBS J* 277, 1967-1978 (2010).

52. B. A. Illarionov, V. S. Bondar, V. A. Illarionova, E. S. Vysotski, Sequence of the cDNA encoding the Ca(2+)-activated photoprotein obelin from the hydroid polyp Obelia longissima. *Gene* 153, 273-274 (1995).

53. A. Ingles-Prieto *et al.*, Conservation of protein structure over four billion years. *Structure* 21, 1690-1697 (2013).

54. S. Inouye, Cloning, expression, purification and characterization of an isotype of clytin, a calcium-binding photoprotein from the luminous hydromedusa Clytia gregarium. *J Biochem* 143, 711-717 (2008).

55. S. Inouye, F. I. Tsuji, Cloning and sequence analysis of cDNA for the Ca(2+)-activated photoprotein, clytin. *FEBS Lett* 315, 343-346 (1993).

56. D. T. Jones, W. R. Taylor, J. M. Thornton, The rapid generation of mutation data matrices from protein sequences. *Comput Appl Biosci* 8, 275-282 (1992).

57. B. Kacar, Towards the Recapitulation of Ancient History in the Laboratory: Combining Synthetic Biology with Experimental Evolution. *Artificial Life* 13, 11-18 (2012).

58. I. V. Kelmanson, M. V. Matz, Molecular basis and evolutionary origins of color diversity in great star coral Montastraea cavernosa (Scleractinia: Faviida). *Mol Biol Evol* 20, 1125-1133 (2003).

59. H. Kim *et al.*, Acid-base catalysis and crystal structures of a least evolved ancestral GFP-like protein undergoing green-to-red photoconversion. *Biochemistry* 52, 8048-8059 (2013).

60. C. Kiss, J. Temirov, L. Chasteen, G. S. Waldo, A. R. Bradbury, Directed evolution of an extremely stable fluorescent protein. *Protein Eng Des Sel* 22, 313-323 (2009).

61. I. M. Klotz, D. L. Hunston, Protein interactions with small molecules. Relationships between stoichiometric binding constants, site binding constants, and empirical binding parameters. *J Biol Chem* 250, 3001-3009 (1975).

62. F. A. Kondrashov, I. B. Rogozin, Y. I. Wolf, E. V. Koonin, Selection in the evolution of gene duplications. *Genome Biol* 3, RESEARCH0008 (2002).

63. R. J. Kulathinal, B. R. Bettencourt, D. L. Hartl, Compensated deleterious mutations in insect genomes. *Science* 306, 1553-1554 (2004).

64. J. R. Lakowicz, *Principles of fluorescence spectroscopy*. (Springer, New York, ed. 3rd, 2006), pp. xxvi, 954 p.

65. S. Q. Le, O. Gascuel, An improved general amino acid replacement matrix. *Mol Biol Evol* 25, 1307-1320 (2008).

66. S. M. Leung *et al.*, Thermal activation of the bovine Hsc70 molecular chaperone at physiological temperatures: physical evidence of a molecular thermometer. *Cell Stress Chaperones* 1, 78-89 (1996).

67. D. A. Liberles *et al.*, The interface of protein structure, protein biophysics, and molecular evolution. *Protein Sci* 21, 769-785 (2012).

68. M. Lynch, The evolution of multimeric protein assemblages. *Mol Biol Evol* 29, 1353-1366 (2012).

69. M. Lynch, Evolutionary diversification of the multimeric states of proteins. *Proc Natl Acad Sci U S A* 110, E2821-2828 (2013).

70. J. P. Mackay, M. Sunde, J. A. Lowry, M. Crossley, J. M. Matthews, Protein interactions: is seeing believing? *Trends Biochem Sci* 32, 530-531 (2007).

71. S. Maguid, S. Fernández-Alberti, G. Parisi, J. Echave, Evolutionary conservation of protein backbone flexibility. *J Mol Evol* 63, 448-457 (2006).

72. N. Maita, J. Nyirenda, M. Igura, J. Kamishikiryo, D. Kohda, Comparative structural biology of eubacterial and archaeal oligosaccharyltransferases. *J Biol Chem* 285, 4941-4950 (2010).

73. G. D. Malo *et al.*, Crystal structure and Raman studies of dsFP483, a cyan fluorescent protein from Discosoma striata. *J Mol Biol* 378, 871-886 (2008).

74. N. J. Marianayagam, M. Sunde, J. M. Matthews, The power of two: protein dimerization in biology. *Trends Biochem Sci* 29, 618-625 (2004).

75. S. V. Markova *et al.*, Green-fluorescent protein from the bioluminescent jellyfish Clytia gregaria: cDNA cloning, expression, and characterization of novel recombinant protein. *Photochem Photobiol Sci* 9, 757-765 (2010).

76. J. Masel, M. L. Siegal, Robustness: mechanisms and consequences. *Trends Genet* 25, 395-403 (2009).

77. J. M. Matthews, M. Sunde, Dimers, oligomers, everywhere. *Adv Exp Med Biol* 747, 1-18 (2012).

78. M. V. Matz, N. J. Marshall, M. Vorobyev, Are corals colorful? *Photochem Photobiol* 82, 345-350 (2006).

79. A. E. Miklos, R. A. Hughes, A. D. Ellington, Design and assembly of large synthetic DNA constructs. *Curr Protoc Mol Biol* Chapter 3, Unit3.23 (2012).

80. H. Morise, O. Shimomura, F. H. Johnson, J. Winant, Intermolecular energy transfer in the bioluminescent system of Aequorea. *Biochemistry* 13, 2656-2662 (1974).

81. S. Murugan, H. C. Hung, Biophysical characterization of the dimer and tetramer interface interactions of the human cytosolic malic enzyme. *PLoS One* 7, e50143 (2012).

82. G. U. Nienhaus, J. Wiedenmann, Structure, dynamics and optical properties of fluorescent proteins: perspectives for marker development. *Chemphyschem* 10, 1369-1379 (2009).

83. F. H. Niesen, H. Berglund, M. Vedadi, The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. *Nat Protoc* 2, 2212-2221 (2007).

84. O. Noivirt-Brik, R. Unger, A. Horovitz, Analysing the origin of long-range interactions in proteins using lattice models. *BMC Struct Biol* 9, 4 (2009).

85. W. J. Ong *et al.*, Function and structure of GFP-like proteins in the protein data bank. *Mol Biosyst* 7, 984-992 (2011).

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

87. E. A. Ortlund, J. T. Bridgham, M. R. Redinbo, J. W. Thornton, Crystal structure of an ancient protein: evolution by conformational epistasis. *Science* 317, 1544-1548 (2007).

88. M. Pagel, A. Meade, D. Barker, Bayesian estimation of ancestral character states on phylogenies. *Syst Biol* 53, 673-684 (2004).

89. A. A. Pakhomov, V. I. Martynov, Probing the structural determinants of yellow fluorescence of a protein from Phialidium sp. *Biochem Biophys Res Commun* 407, 230-235 (2011).

90. A. Pandini, M. S. Denison, Y. Song, A. A. Soshilov, L. Bonati, Structural and functional characterization of the aryl hydrocarbon receptor ligand binding domain by homology modeling and mutational analysis. *Biochemistry* 46, 696-708 (2007).

91. A. Pandini, G. Mauri, A. Bordogna, L. Bonati, Detecting similarities among distant homologous proteins by comparison of domain flexibilities. *Protein Eng Des Sel* 20, 285-299 (2007).

92. J. Pei, B. H. Kim, N. V. Grishin, PROMALS3D: a tool for multiple protein sequence and structure alignments. *Nucleic Acids Res* 36, 2295-2300 (2008).

93. V. Z. Pletnev *et al.*, Structure of the red fluorescent protein from a lancelet (Branchiostoma lanceolatum): a novel GYG chromophore covalently bound to a nearby tyrosine. *Acta Crystallogr D Biol Crystallogr* 69, 1850-1860 (2013).

94. N. V. Pletneva *et al.*, [Three-dimensional structure of yellow fluorescent protein zYFP538 from Zoanthus sp. at the resolution 1.8 angstrom]. *Bioorg Khim* 33, 421-430 (2007).

95. D. D. Pollock, G. Thiltgen, R. A. Goldstein, Amino acid coevolution induces an evolutionary Stokes shift. *Proc Natl Acad Sci U S A* 109, E1352-1359 (2012).

96. D. C. Prasher, V. K. Eckenrode, W. W. Ward, F. G. Prendergast, M. J. Cormier, Primary structure of the Aequorea victoria green-fluorescent protein. *Gene* 111, 229-233 (1992).

97. J. Projecto-Garcia *et al.*, Repeated elevational transitions in hemoglobin function during the evolution of Andean hummingbirds. *Proc Natl Acad Sci U S A* 110, 20669-20674 (2013).

98. T. a. P. e. I. Pupko, in *Currents in Computational Molecular Biology,* S. Miyano, Shamir, R, and Takagi, T., Ed. (Universal Academy Press, Tokyo, Japan, 2000), pp. 184-185.

99. T. Pupko, I. Pe'er, M. Hasegawa, D. Graur, N. Friedman, A branch-and-bound algorithm for the inference of ancestral amino-acid sequences when the replacement rate varies among sites: Application to the evolution of five gene families. *Bioinformatics* 18, 1116-1123 (2002).

100. T. Pupko, I. Pe'er, R. Shamir, D. Graur, A fast algorithm for joint reconstruction of ancestral amino acid sequences. *Mol Biol Evol* 17, 890-896 (2000).

101. S. J. Remington *et al.*, zFP538, a yellow-fluorescent protein from Zoanthus, contains a novel three-ring chromophore. *Biochemistry* 44, 202-212 (2005).

102. V. A. Risso, J. A. Gavira, D. F. Mejia-Carmona, E. A. Gaucher, J. M. Sanchez-Ruiz, Hyperstability and substrate promiscuity in laboratory resurrections of Precambrian β-lactamases. *J Am Chem Soc* 135, 2899-2902 (2013).

103. P. A. Romero, F. H. Arnold, Exploring protein fitness landscapes by directed evolution. *Nat Rev Mol Cell Biol* 10, 866-876 (2009).

104. D. A. Shagin *et al.*, GFP-like proteins as ubiquitous metazoan superfamily: evolution of functional features and structural complexity. *Mol Biol Evol* 21, 841-850 (2004).

105. N. C. Shaner *et al.*, Improved monomeric red, orange and yellow fluorescent proteins derived from Discosoma sp. red fluorescent protein. *Nat Biotechnol* 22, 1567-1572 (2004).

106. J. A. Sniegowski, J. W. Lappe, H. N. Patel, H. A. Huffman, R. M. Wachter, Base catalysis of chromophore formation in Arg96 and Glu222 variants of green fluorescent protein. *J Biol Chem* 280, 26248-26255 (2005).

107. O. V. Stepanenko, I. M. Kuznetsova, V. V. Verkhusha, K. K. Turoverov, Beta-barrel scaffold of fluorescent proteins: folding, stability and role in chromophore formation. *Int Rev Cell Mol Biol* 302, 221-278 (2013).

108. R. A. Studer, B. H. Dessailly, C. A. Orengo, Residue mutations and their impact on protein structure and function: detecting beneficial and pathogenic changes. *Biochem J* 449, 581-594 (2013).

109. J. M. Thomson *et al.*, Resurrecting ancestral alcohol dehydrogenases from yeast. *Nat Genet* 37, 630-635 (2005).

110. J. W. Thornton, Resurrecting ancient genes: experimental analysis of extinct molecules. *Nat Rev Genet* 5, 366-375 (2004).

111. J. W. Thornton, E. Need, D. Crews, Resurrecting the ancestral steroid receptor: ancient origin of estrogen signaling. *Science* 301, 1714-1717 (2003).

112. N. Tokuriki, D. S. Tawfik, Protein dynamism and evolvability. *Science* 324, 203-207 (2009).

113. N. Tokuriki, D. S. Tawfik, Stability effects of mutations and protein evolvability. *Curr Opin Struct Biol* 19, 596-604 (2009).

114. N. Tokuriki, D. S. Tawfik, Stability effects of mutations and protein evolvability. *Curr Opin Struct Biol* 19, 596-604 (2009).

115. C. Torres-Sosa, S. Huang, M. Aldana, Criticality is an emergent property of genetic networks that exhibit evolvability. *PLoS Comput Biol* 8, e1002669 (2012).

116. H. Tsutsui *et al.*, The E1 mechanism in photo-induced beta-elimination reactions for green-to-red conversion of fluorescent proteins. *Chem Biol* 16, 1140-1147 (2009).

117. J. A. Ugalde, B. S. Chang, M. V. Matz, Evolution of coral pigments recreated. *Science* 305, 1433 (2004).

118. J. M. Walker, *The proteomics protocols handbook*. (Humana Press, Totowa, N.J., 2005), pp. xviii, 988 p.

119. W. W. Ward, S. H. Bokman, Reversible denaturation of Aequorea green-fluorescent protein: physical separation and characterization of the renatured protein. *Biochemistry* 21, 4535-4540 (1982).

120. D. M. Weinreich, N. F. Delaney, M. A. Depristo, D. L. Hartl, Darwinian evolution can follow only very few mutational paths to fitter proteins. *Science* 312, 111-114 (2006).

121. S. Whelan, N. Goldman, A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Mol Biol Evol* 18, 691-699 (2001).

122. P. D. Williams, D. D. Pollock, B. P. Blackburne, R. A. Goldstein, Assessing the accuracy of ancestral protein reconstruction methods. *PLoS Comput Biol* 2, e69 (2006).

123. C. S. Wylie, E. I. Shakhnovich, A biophysical protein folding model accounts for most mutational fitness effects in viruses. *Proc Natl Acad Sci U S A* 108, 9916-9921 (2011).

124. N. S. Xia *et al.*, Bioluminescence of Aequorea macrodactyla, a common jellyfish species in the East China Sea. *Mar Biotechnol (NY)* 4, 155-162 (2002).

125. S. Yokoyama, H. Yang, W. T. Starmer, Molecular basis of spectral tuning in the red- and green-sensitive (M/LWS) pigments in vertebrates. *Genetics* 179, 2037-2043 (2008).

Vita

Chintan Modi was born in 1977 in Gujarat, India, and then grew up in Staten Island and New Jersey after immigrating to the United States. He completed his high school in Parsippany, New Jersey. Upon graduation, he attended The Pennsylvania State University, where from he received his Bachelor of Science degree in Biochemistry and Molecular Biology in 2000. During his undergraduate years, Chintan contributed to the astronomy and breast cancer research projects working under the guidance of Dr. Maria Womack and Dr. Andrea Mastro, respectively. His contribution to these projects was recognized with co-authorship in a peer reviewed journal and an abstract at a scientific meeting. In 2001, Chintan joined Merck & Co., Inc. as a junior scientist, where he was involved in vaccine and therapeutic antibody projects aimed to treat some of the major diseases afflicting the world. Chintan left the public sector in 2005 to pursue his doctorate at The University of Texas at Austin (UT) to advance his education and to further explore his interest in protein biology and chemistry. After joining the laboratory of Dr. Mikhail Matz in 2008 at UT, he contributed to the study of novel Hydrozoan multi-domain green fluorescent proteins (GFP). His dissertation work focused on the emergence of biochemical complexity and the evolution of structure-function relationships in the GFP-family of proteins. In addition to his career in science, he enjoys working with community outreach programs aimed at teaching and communicating current scientific work. Also while at UT, he started a salsa dance club with the aim of introducing the dance to a wider community in a friendly atmosphere. Chintan has a deep appreciation for international affairs and philosophy, and enjoys in-depth conversations on them with his friends and family. He enjoys cooking and traveling, and leads an active social life.

Permanent Email Address: Chintan.Modi@utexas.edu

This dissertation was typed by the author.