“Let us think the unthinkable, let us do the undoable, let us prepare to grapple with the ineffable itself, to see if we may not eff it after all.”

- Douglas Adams, *Dirk Gently’s Holistic Detective Agency*

University of Alberta

Development of cpRFPs for use as Ca^{2+} biosensors

by

Haley Jayne Carlson

A thesis submitted to the Faculty of Graduate Studies and Research
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Chemistry

Spring 2013
Edmonton, Alberta

I dedicate this thesis to my parents Deanna and Rodney

Abstract

The discovery of green fluorescent protein (GFP) from the *Aequorea victoria* jellyfish revolutionized many fields in the scientific community, including molecular biology, protein engineering, and neuroscience. The ability to genetically link a fluorescent protein to a protein of interest has allowed scientists to probe the exact structural localization of proteins. Another important application of FPs is their design for use in biosensors, whereby the fluorescence of the protein is intrinsically dependent on a small molecule of interest, such as calcium ion (Ca^{2+}) or a physiological process such as phosphorylation or caspase activity. In single FP-based biosensors of small molecules, the FP must be circular permutated, whereby the original N- and C-termini are linked together and new termini are introduced closer to the chromophore. At the start of the work described in this thesis a lot of work had gone into developing and improving GFP-based Ca^{2+} biosensors, but there were no reports of a red FP-based biosensor.

The work in this thesis describes the engineering of an RFP-based Ca^{2+} biosensor using a circulary permutated RFP, mCherry. The first step in this process was to engineer a cpmCherry variant with termini near the chromophore. mCherry required a lot of engineering and optimization in order to identify a fluorescent variant with termini near the chromophore. Ultimately a cpmCherry split at position 145 was found that, when fused to calmodulin (CaM) and M13, showed a response to Ca^{2+}. The initial construct had limited response and was subjected to several rounds of mutagenesis to improve both the brightness and fluorescence response. The final variant CH-GECO3.1 shows a 250% signal increase with Ca^{2+} and could be imaged successfully in mammalian cells to monitor Ca^{2+} fluctuations.

To further our understanding of this biosensor, site-directed mutagenesis was done to probe the structure-function relationship. After mutagenesis a few residues stood out as key residues that likely played a role in the mechanism of fluorescence increase, such as Gln163 and Glu61 (linker). Other mutations were

introduced into the protein to determine whether the excitation and emission wavelengths could be altered, while still retaining function.

The final section of this work describes the reconstitution of split green and red Ca^{2+} biosensors using intein technology. Split inteins will spontaneously splice together protein fragments that are genetically linked to themselves. To take advantage of protein splicing several different Ca^{2+} biosensors were split into and N-terminal and C-terminal fragments and attached to the N-terminal or C-terminal intein, respectively. These fragments were co-transfected into mammalian HeLa cells and imaged for fluorescence signal and response to Ca^{2+} fluctuations.

Acknowledgements

If it takes a village to raise a child, how many does it take to graduate a PhD student? I owe my success in graduate studies to not only the people I encountered along the way, but to the villages of people that came before and helped shape me into the person I am today.

Of course the most prominent village I was a part of is the Campbell group. I started as an undergraduate and have seen many people come and go. However, the overall generosity and helpfulness of the group has remained constant and I am very grateful for the opportunity to work with such a diverse and welcoming group. Key to my success was the support of my supervisor, Dr. Robert Campbell, who was enthusiastic about my successes, and perhaps more importantly, understanding about the many failures. I cannot forget one group member in particular, who insisted on 'whistling' Christmas carols, even in the middle of July. Spencer (D.B.) Alford helped me retain my sanity by offering an ear to listen to my rantings and ravings or giving technical advice when I had angered the Molecular Biology Gods.

I also owe a debt of gratitude to the people working behind the scences. To Gareth Lambkin in Biological Services, Wayne Moffat in Analytical and Instrumentation Laboratory, Bernie Hippel in the store room, and Cheryl Nargang and Troy Locke in MBSU; without whom I would never have finished.

Sometimes it hard to imagine my life before graduate school, but there have been teachers from K-12 and beyond that have encouraged me to succeed and believed in me more than I believed in myself. I hope one day I will be given the opportunity to pay it forward and inspire confidence in someone's, anyone's, ability to succeed. I cannot overstate the importance of good educators who challenge students to think outside the (particle in a) box and are invested in their students learning.

There is some sentimental clap trap out there on the internet that says 'friends are the family we choose for ourselves', but sometimes life leads you down roads less travelled and you are amazed at the rag tag bunch of individuals

you come out with at the end. To Lindsay, Ashley, and Amy from my youth, who are always up for a good gossip and catch-up session no matter how long it has been since we saw each other last. There are many people who were and are still in the graduate student trenches who offered invaluable opportunities to celebrate and commiserate the highs and lows of graduate student life. These people include Spencer, Garson, Jared, Ryan, Owen S., and Sarah. There are many more outside academia from the Whiskey Cats and Ice Beers (I sense a theme here…) that have kept me balanced and reminded me that there is a whole world out there beyond graduate school. The last group of people that instead makes me question my sanity, is my SOF family. Almost five years of 5:30 AM year-round, outdoor bootcamp would not be possible without the amazing group of individuals that also show up for fun. There are too many people to mention, but needless to say, they have been there to strengthen me physically, and more importantly, mentally.

There is one other person who is more to me than just friend or family. I may have started my PhD before I met Matt, but I cannot imagine how I could have finished it without him. We are two unrepentant peas in a pod and I am blessed to have you in my life. I look forward to whatever lies ahead for us in the future.

Finally, I must pay tribute to the original village that raised me, my immediate and extended family. My parents Deanna and Rod have loved me unconditionally and supported me in my academic and non-academic endeavors. Even if I had to explain to them on several occasions what it means to be a graduate student and what sort of research I was doing in the lab. Neither of them is quite sure where my desire for extra schooling comes from, but I am sure both of them would stake a claim that it stems from their side of the family. I love them both very much and am blessed to call them mom and dad. My grandparents Kay, George, Madeline, and Alan, and aunts, and uncles too numerous to mention, have supported me in their own way whether they realized it or not. To my brother Nate, I'll be sure to get you to call me Dr. at least once, and my future sister-in-law Lauren thank-you for your enthusiasm and support of my continued education.

What is out there waiting for me remains to be seen. But, I am confident with the continued love and support of the people in my life I will get there. “I may not have gone where I intended to go, but I think I have ended up where I needed to be.” – Douglas Adams, *The Long Dark Tea-Time of the Soul*

Table of Contents

List of Figures

List of Tables

List of Equations

List of Abbreviations

ADP	adenosine diphosphate
ATP	adenosine triphosphate
Ala	alanine
amp	ampicillin
Å	angstrom
Arg	arginine
Asn	asparagine
Asp	aspartate
*av*GFP	*Aequorea victoria* green fluorescent protein
β	beta
BFP	blue fluorescent protein
BiFC	bimolecular fluorescence complementation
BPER	bacterial protein extraction reagent
bps	base pairs
CaM	calmodulin
CCD	charge coupled device
CFP	cyan fluorescent protein
Ci-VSP	*Ciona intestinalis* voltage sensitive protein
cm	centimetre
cp	circular permutation or circular permutated

CPSARST	Circular permutation search aided by Ramachandran sequential transformations
cGMP	cyclic guanosine monophosphate
Cys	cysteine
d	dimer
dCTP	deoxycytosine triphosphate
ddGFP	dimerization dependent green fluorescent protein
ddRFP	dimerization dependent red fluorescent protein
ddYFP	dimerization dependent yellow fluorescent protein
DEVD	Asp-Glu-Val-Asp
DIC	differential interference contrast
DNA	deoxyribonucleic acid
dNTP	deoxyribonucleic acid triphosphate
DsRed	*Discosoma striata* red fluorescent protein
dTTP	deoxythymidine triphosphate
ε	extinction coefficient
EBFP	enhanced blue fluorescent protein
EBFP2	enhanced blue fluorescent protein version 2.0
ECFP	enhanced cyan fluorescent protein
EGFP	enhanced green fluorescent protein
EGTA	ethylene glycol tetraacetic acid
EPPCR	error-prone polymerase chain reaction
*eq*FP578	*Entacmaea quadricolor* fluorescent protein

ESPT	excited-state proton transfer
EYFP	enhanced yellow fluorescent protein
FBS	fetal bovine serum
FP	fluorescent protein
FRET	Förster resonance energy transfer
g	gravitational force
GECO	genetically encoded Ca^{2+} indicator for optical imaging
GFP	green fluorescent protein
Gln	glutamine
Glu	glutamate
Gly	glycine
h	hour
hrs	hours
HEPES	hydroxyethyl piperazineethanesulfonic acid
HHBSS	HEPES buffered hanks balanced saline solution
His	histidine
IDT	Integrated DNA Technologies
Ile	isoleucine
K_d	dissociation constant
kDa	kilodalton
KOAc	potassium acetate
LB	Lennox broth

Leu	leucine
LSS	long Stokes-shift
Lys	lysine
m	monomeric
M	molar (mol/L)
Met	methionine
min	minute
μL	microlitre
mL	millilitre
μM	micromolar
mM	milimolar
MOPS	3-morpholino propane-1-sulfonic acid
MRD	molecular recognition domain
mRFP1	monomeric red fluorescent protein
mTFP0.7	monomeric teal fluorescent protein version 0.7
mTFP1	monomeric teal fluorescent protein version 1.0
MWCO	molecular weight cut-off
NADH/NAD^+	nicotinamide adenine dinucleotide
nm	nanometre
nM	nanomolar
Npu	*Nostoc punctiform*
NTA	nitrilotriacetic acid

OE-PCR	overlap exchange polymerase chain reaction
PA	photo activation
PBS	phosphate buffered saline
PDB	Protein Database
Pfu	*Pyrococcus furiosus*
Phe	phenylalanine
pK_a	logarithmic measure of the acid dissociation constant or pH at which a fluorescent protein has 50% fluorescence signal
pM	picomolar
POI	protein of interest
PPI	protein-protein interaction
Pro	proline
PS	photoswitchable
Φ	quantum yield
RFP	red fluorescent protein
R-GECO1	red GECO version 1
rms	root mean square deviation
rpm	revolutions per minute
rsd	relative standard deviation
SDS-PAGE	sodium dodecyl sulphate polyacrylamide gel electrophoresis
sec	second

Ser	serine
sf	superfolder
Ssp	*Synechocytics sp.*
StEP	staggered extension process
TB	Terrific broth
TBS	Tris buffered saline
Thr	threonine
Trp	tryptophan
Tyr	tyrosine
Val	valine
VSFP	voltage sensing fluorescent protein

Chapter 1 General Introduction[1]

1.1 Overview and Premise

A variety of intrinsically fluorescent proteins (FP) from the superfamily of *Aequorea victoria* green FP (avGFP)-like proteins have proven to be popular and powerful tools for live cell fluorescence imaging applications. Their usefulness stems from the fact these proteins are *self-sufficient* to form a brightly fluorescent chromophore within the confines of their own tertiary structure (1). Accordingly, expression of a FP gene chimera in a wide variety of cell-types and tissue-types can provide a researcher with a non-invasive fluorescent marker for imaging of fusion protein localization and dynamics (Figure 1.1).

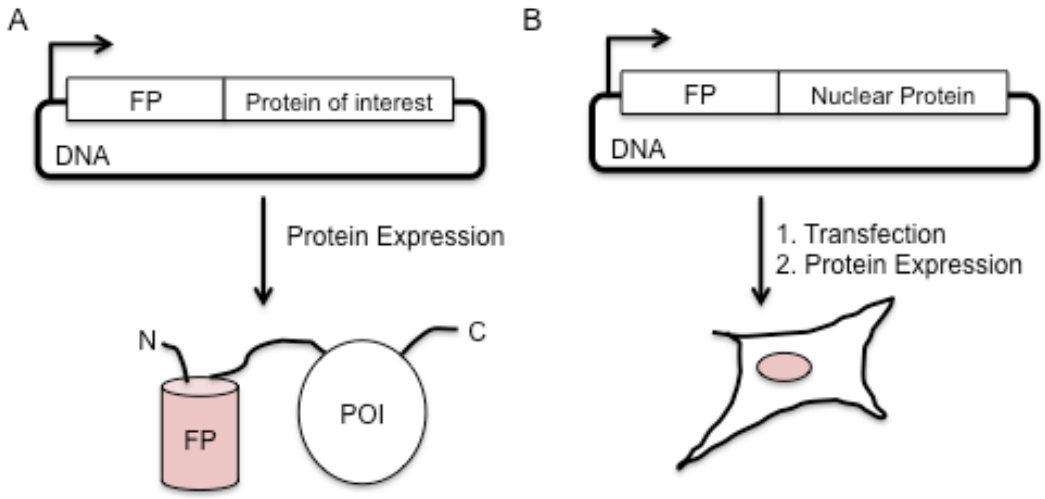


Figure 1.1 Schematic of FP tagged proteins. A) Genetically linking an FP gene with a protein of interest (POI) leads to a chimeric protein with a fluorescent protein attached. B) Example of an FP attached to a nuclear protein. Only the nucleus will be fluorescent when the protein is expressed.

[1] A portion of this Chapter has been published. Carlson, H.J., Cotton, D.W., Campbell, R.E. 2010 *Protein Science*, **19**: 1490-1499.

For many applications of FPs, the fact that the FP chromophore is fully encapsulated inside an 11-stranded 'β-can' tertiary structure (2) and thus effectively isolated from the environment of the bulk solvent, is advantageous. Since the chromophore is isolated, its inherent fluorescent properties (*e.g.*, quantum yield, extinction coefficient, lifetime, and rate of photobleaching) are presumably less susceptible to changes in the protein environment (*e.g.*, changes in pH, viscosity, and concentration of quenchers of fluorescence) than if the chromophore was exposed to the bulk solvent. The relative steadfastness of the intrinsic fluorescent properties of the FP allow researchers to confidently assume that fluorescent intensity is directly proportional to protein concentration in live cell imaging experiments.

There are some applications of FPs for which the protected nature of the chromophore is a distinct disadvantage. For example, there is substantial interest in using induced changes in the fluorescent properties of the FP chromophore to report on specific changes in the intracellular environment (3). FP constructs in which the fluorescent intensity or hue is dependent upon an external parameter are commonly referred to as single FP-type biosensors. An alternative strategy for creating FP-based biosensors is to exploit the modulation of Förster resonance energy transfer (FRET) between two different hues of FP (4). The advantages and disadvantages of each of these distinct strategies has previously been discussed (5).

In designing and constructing single FP-type biosensors, an effective 'line of communication' between the external variable of interest and the fluorescent properties of the chromophore must be created. That is, the external change in the environmental variable must produce a physical change in the chromophore environment that modulates its fluorescent hue, intensity, or lifetime. One of the simplest examples of an environmental variable that can be 'sensed' and imaged in live cells using FPs is pH. With few exceptions, the fluorescence of FPs is pH dependent and a number of variants that change fluorescence intensity at near-physiological pH values have been reported (6). Some FPs also exhibit an intrinsic sensitivity towards halide ions and have been used to image physiologically-relevant changes in chloride ion concentrations (7).

The creation of single FP-based biosensors that specifically respond to environmental variables other than pH or halide ions has required some relatively sophisticated engineering of avGFP chimeras. Specifically, researchers have found means of fusing avGFP with a second protein that provides the molecular recognition specificity for the target analyte of interest. Upon binding to (or being modified by) the target, the molecular recognition domain undergoes a conformational change that alters the chromophore environment and the fluorescent properties of the chromophore. However, the N- and C-termini of the FP, which are the most accessible locations for fusion of the molecular recognition domain, are distant from the chromophore. Accordingly, a conformational change in a protein fused to one of these termini is expected to have no substantial effect on the chromophore properties. One solution to this fundamental problem is to insert the molecular recognition domain into the FP at a location near the chromophore (8). Examples of this strategy include the insertion of calmodulin (CaM) (8) or a single EF-hand domain (9) into avGFP to create calcium ion (Ca^{2+}) biosensors. Similarly, biosensors for the protease trypsin have been created by introducing substrate sequence containing loops at various locations within an avGFP variant (10).

A second solution is to create a circular permutated (cp) variant of FPs in which the new N- and C- termini are located in close proximity to the chromophore (8, 11-13). Fusion of peptides or protein domains that undergo a conformational change or an interaction in the presence of the analyte of interest to the new termini can produce highly effective single FP-based biosensors (14-16). Perhaps the most important example of this second type of biosensor design are the Ca^{2+} biosensors (14-17) in which CaM and the M13 peptide are fused to the termini of a circular permutated avGFP variant. X-ray crystal structures of some of the most promising single FP-based Ca^{2+} biosensors, known as the GCaMP series, have recently been reported (18-20). Yet another variation on this design involves the insertion of a circular permutated FP into a second protein that undergoes a conformational change upon binding to its target analyte. For example, a circular permutated version of the red fluorescent protein mKate (21) has been converted into a voltage sensitive fluorescent protein by insertion into the voltage sensitive domain of a membrane protein (22). These FP applications

along with others such as bimolecular fluorescence complementation (BiFC) and a relatively unused technique of protein splicing using inteins will be discussed in more detail in Sections 1.3.6.3 and 1.3.6.2.

A potentially valuable application of single FP-based biosensors would be to use them in pair wise combinations for multi-parameter live cell imaging (11). At the time that this work had begun there was no reported red FP-based (RFP) Ca^{2+} biosensor. However, since all of the single FP-based biosensors reported to date are based on either green or yellow variants of avGFP, it is not possible to image multiple biosensors in a single cell. Recently, Zhao et al. published an impressive paper that described the design of several different FP-based Ca^{2+} biosensors ranging in colour from blue to red (14). It appears as though these Ca^{2+}-based biosensors function via the same mechanism as the previously reported green fluorescent biosensors. Specifically, the pK_a of the chromophore is lowered upon Ca^{2+} binding leading to a larger population of the protein in the fluorescent state. The work in this thesis undertook the development of circular permutated monomeric red FPs (cpRFPs) for their eventual use as Ca^{2+} biosensors (23). Our starting template is the engineered variant of *Discosoma* RFP (24) known as mCherry (25). Along the way several unique and intriguing mechanistic and structural differences were characterized for the mCherry-based sensor. The remainder of this chapter will provide an introduction to FPs , including their discovery, structure, and engineering, along with a more in-depth look at their use in single-FP based biosensors.

1.2 Circular permutated FPs in single-FP based biosensors

Circular permutation is the most common tool used to generate FPs with an exposed chromophore and the methodology is discussed in further detail in section 1.3.6. Baird et al. demonstrated that CaM or zinc finger (zinc binding domain) could be inserted into EYFP and used to detect Ca^{2+} or zinc, respectively (8). As mentioned in the previous section the original N- and C-termini are located on the periphery of the β-barrel structure, far-removed from the chromophore. Denatured GFP is non-fluorescent, as are chemically synthesized analogues of the chromophore structure (26-29). In order to take advantage of the chromophore fluorescence and use it as a tool to detect

different cellular events, the chromophore must be more exposed to the external environment. Attachment of external molecular recognition domains (MRD) would not be able to confer their conformation change to the chromophore environment and fluorescenct signal would remain unchanged (Figure 1.2A). Alternatively, if the termini are moved closer to the chromophore, there is a greater likelihood that the fluorescence will change upon analyte binding (Figure 1.2B).

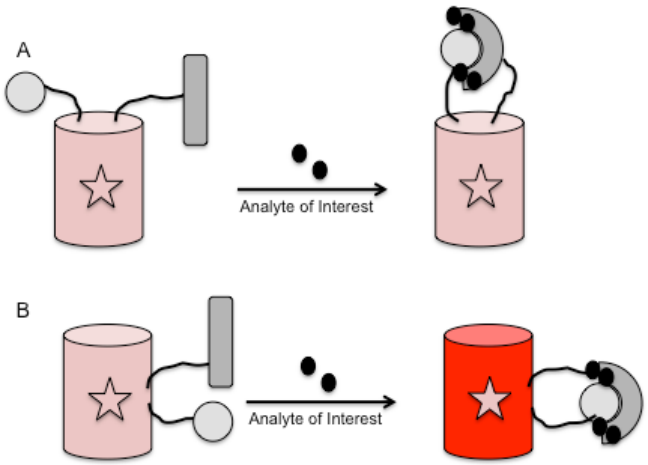


Figure 1.2 Schematic of single-FP based biosensor. A) Non-permutated FP does not change fluorescence upon analyte binding. B) cpFP fluorescence change due to close proximity of MRD to chromophore

At nearly the same time as the publication by Baird et al., another paper came out that systematically screened all positions of the enhanced GFP (EGFP) β-barrel for sites that tolerated circular permutation (12). Both papers reported positions 145 and 173 as tolerant of circular permutation. Preliminary work done by Baird et al. used circular permutated enhanced cyan FP (ECFP) and enhanced yellow FP (EYFP) as a FRET pair with the Ca^{2+} binding protein calmodulin (CaM) inserted as the external binding domain. They also generated a Ca^{2+}-sensitive cpEYFP by inserting CaM at position 145. Both preliminary sensor designs undoubtedly served as inspiration for many well known FRET-based and single-FP based biosensors (15, 16, 30).

Shortly thereafter both cpEGFP and cpEYFP permutated at position 145 were used to construct the first single-FP based Ca^{2+} sensors, GCaMP and Pericam, respectively (15, 16). Both sensors detect Ca^{2+}, with the notable

difference being the FP used. GCaMP uses cpEGFP permutated at position 145 with a few N-terminal residues deleted, while pericam has cpEYFP, permutated at position 145 as well. Significant work has gone into improving the dynamic range, chromophore maturation, and protein expression of GCaMP, resulting in several new variants, namely, GCaMP2, GCaMP3, GCaMP5, and the GECO series (14, 20, 31, 32).

In the decade since the development of these sensors, new biosensors for the detection of other small molecules as well as improved Ca^{2+} biosensors have been reported. Circular permutation has also been used to improve voltage sensitive FPs (VSFPs). Initial reports of successful single-FP VSFPs linked intact GFP or CFP (33-35) to voltage gated potassium channels such as the Shaker potassium channel, or the similar *Ci*-VSP domain from a voltage-sensor containing phosphatase. More recently Gautam et al. reported the first cpRFP-based voltage sensor, which was developed using the far-red emitting mKate (21) permutated at position 180 (22). This sensor shows improved fluorescence response to voltage changes.

For other small molecule detection you could argue that, without circular permutation, successful sensors could not have been designed. T-Sapphire, a GFP with excitation at 375 nm was circular permutated at position 145 and inserted into a tandem dimer of the T-Rex NADH binding domain to generate a biosensor that can report the NADH/NAD^+ ratio (36). A sensor for hydrogen peroxide, Hyper, was developed by inserting a cpYFP into the *E. coli* regulatory domain sensitive to hydrogen peroxide (OxyR-RD) (37). Berg et al. inserted cpmVenus, a bright YFP, into the bacterial regulatory protein, GlnK1, and the construct was engineered to produce a ratiometric FP-reporter of ATP/ADP ratio that was coined Perceval (38). Finally, cpECFP, cpEYFP, and cpEGFP permutated at position 145 were attached to a kinase substrate domain and a phosphorylation recognition domain to produce three distinct cpFP-based biosensors for the detection of kinase activity (39). A summary these sensors and others is provided below in Table 1.1. From Table 1.1 it is apparent that cpEGFP remains the protein of choice for single cpFP-based biosensors, but the discovery of new RFPs with improved characteristics and the desire for

multiparameter imaging will surely prompt an increase in the number of red-shifted cpFP sensors.

Table 1.1 Summary of cpFP based biosensors

Sensor name	cpFP	Small molecule/ Biological Process	Reference
GCaMP	cpEGFP (145)[1]	Ca^{2+}	(16)
Pericam	cpEYFP (145)	Ca^{2+}	(15)
Sinphos cyan	cpECFP (145)	phosphorylation	(39)
Sinphos green	cpEGFP (145)	phosphorylation	(39)
Sinphos yellow	cpCitrine (145)	phosphorylation	(39)
GCaMP2	cpEGFP (145)	Ca^{2+}	(31)
Hyper	cpEYFP (145)	H_2O_2	(37)
Case12/16	cpEGFP (145)	Ca^{2+}	(17)
FlincGs	cpEGFP (145)	cGMP	(40)
Perceval	cpmVenus (145)	ATP:ADP	(38)
GCaMP3	cpEGFP (145)	Ca^{2+}	(32)
VSFP(C)cpmKate (180)	cpmKate (180)	voltage	(22)
VSFP(D)cpmKate (180)	cpmKate (180)	voltage	(22)
Peredox	cpT-Sapphire (145)	NADH	(36)
G-GECO1	cpEGFP (145)	Ca^{2+}	(14)
G-GECO1.1	cpEGFP (145)	Ca^{2+}	(14)
G-GECO1.2	cpEGFP (145)	Ca^{2+}	(14)
B-GECO1	cpEGFP (145) Y66H	Ca^{2+}	(14)
GEX-GECO	cpEGFP (145)	Ca^{2+}	(14)
GEM-GECO	cpEGFP (145)	Ca^{2+}	(14)
R-GECO1	cpmApple (145)	Ca^{2+}	(14)
GCaMP5	cpEGFP (145)	Ca^{2+}	(20)
ElectricPk	cpEGFP (145)	voltage	(41)
CH-GECO3.1	cpmCherry (145)	Ca^{2+}	unpublished

[1] number indicates the amino acid position where the circular permutation was made

Despite these numerous examples of single-cpFPs it is apparent that there is a lack of RFP-based biosensors. With the exception of the one cpVSFP and red genetically encoded optical indicator (R-GECO1), all other sensors are based off of GFPs, CFPs, or YFPs. The lack of RFP biosensors is partly due to the limited

work published on cpRFPs (13, 22, 23, 42). Additionally, initially discovered RFPs formed tetramers, which delayed their use as single FP sensors until sufficiently bright monomeric versions could be engineered (21, 43). Two of the most commonly used monomeric RFPs are mCherry and mKate (21, 25), but both of these proteins are less tolerant of circular permutation than their GFP counterparts. Only recently was an RFP permutated at an analogous position to 145 reported (14). It appears that more extensive engineering must be done to generate cpRFPs, perhaps because the additional maturation step required for formation of the fully mature red chromophore can be more easily disrupted during permutation.

There are several reasons why RFP sensors are desirable including the ability to do multiparameter imaging with existing GFP-based sensors, lower levels of autofluorescence, less phototoxicity, and potential use of long-wavelength two-photon excitation (44). Given these potential benefits, despite the additional challenges of engineering RFPs, the design and optimization of cpRFP-sensors are worth the extra effort.

1.3 Fluorescent Proteins

Since the discovery of a squeezate-derived protein from the *Aequorea victoria* jellyfish by Osama Shimomura in 1962 (45), *av*GFP and it's related FPs from cnidarians and copepods, have revolutionized several facets of the scientific community by providing a means of genetically encoding fluorescence.

1.3.1 Amino acid sequence

It was 30 years after the discovery of *av*GFP that its amino acid sequence was determined (46). The protein is comprised of 238 amino acids and has a molecular weight of 27 kDa. It was also determined that the fluorescence of the protein came from an internal chromophore structure that formed autocatalytically from three residues in the amino acid sequence with no need of external co-factors. In *av*GFP a Ser65-Tyr66-Gly67 tripeptide undergoes an autocatalytic posttranslational modification to form the chromophore. A variety of residues are found at position 65 and 66 in different coloured FPs (47, 48), and Gly67 is strictly conserved across many FP families (49). A decade after *av*GFP was

sequenced, Matz et al. published the first paper characterizing several red-shifted proteins found in 5 different Anthozoa species (24). Surprisingly, the comparison of the protein sequence found only 20-30% sequence identity despite sharing similar protein structure (Figure 1.3). The chromophore forming tripeptide found in DsRed is Gln65-Tyr66-Gly67. There are a few other residues that are conserved across FPs, which also appear to play important roles in chromophore formation(50, 51).

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
*av*GFP	M	S	K	G	E	E	L	F	T	G	V	V	P	I	L	V	E	L	D	G
drFP583	M	R	S	S	K	N	V	I	K	E	F	M	R	F	K	V	R	M	E	G
	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
*av*GFP	D	V	N	G	H	F	K	S	V	S	G	E	G	E	G	D	A	T	Y	G
drFP583	T	V	N	G	H	E	F	E	I	E	G	E	G	E	G	R	P	Y	E	G
	41	42	43	44	45	46	47	48	49	50	51	52	-	53	54	55	56	57	58	59
*av*GFP	K	L	T	L	K	F	I	C	T	T	G	K	-	L	P	V	P	W	P	T
drFP583	H	N	T	V	K	L	K	V	T	K	G	G	P	L	P	F	A	W	D	I
	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79
*av*GFP	L	V	T	T	F	S	Y	G	V	Q	C	F	S	R	Y	P	D	H	M	K
drFP583	L	S	P	Q	F	Q	Y	G	S	K	V	Y	V	K	H	P	A	D	I	P
	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99
*av*GFP	Q	H	D	F	F	K	S	A	M	P	E	G	Y	V	Q	E	R	T	I	F
drFP583	-	-	D	Y	K	K	L	S	F	P	E	G	F	K	W	E	R	V	M	N
	100	101	102	103	104	105	106	107	108	109	110	111	112	113	114	115	116	117	118	119
*av*GFP	F	K	D	D	G	N	Y	K	T	R	A	E	V	K	F	E	G	D	T	L
drFP583	F	E	D	G	G	V	V	T	V	T	Q	D	S	S	L	Q	D	G	C	F
	120	121	122	123	124	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139
*av*GFP	V	N	R	I	E	L	K	G	I	D	F	K	E	D	G	N	I	L	G	H
drFP583	I	Y	K	V	K	F	I	G	V	N	F	P	S	D	G	P	V	M	Q	K
	140	-	141	142	143	144	145	146	147	148	149	150	151	152	153	154	155	156	157	158
*av*GFP	K	-	L	E	Y	N	Y	N	S	H	N	V	Y	I	M	A	D	K	Q	K
drFP583	K	T	M	G	W	E	A	S	T	E	R	L	Y	P	R	D	G	V	L	K
	159	160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175	176	177	178
*av*GFP	N	G	I	K	V	N	F	K	I	R	H	N	I	E	D	G	S	V	Q	L
drFP583	G	E	I	H	K	A	L	K	L	K	-	-	-	-	D	G	G	H	Y	L
	179	180	181	182	183	184	185	186	187	189	190	191	192	193	194	195	196	197	198	199
*av*GFP	A	D	H	Y	Q	Q	N	T	P	I	G	D	G	P	V	L	L	P	D	N
drFP583	V	E	F	-	-	K	S	I	Y	M	A	K	K	P	V	Q	L	P	G	Y
	200	201	202	203	204	205	206	207	208	209	210	211	212	213	214	215	216	217	218	219
*av*GFP	H	Y	L	S	T	Q	S	A	L	S	K	D	P	N	E	K	R	D	H	M
drFP583	Y	Y	V	D	S	K	L	D	I	T	S	-	-	-	H	N	E	D	Y	T
	220	221	222	223	224	225	226	227	228	229	230	231	232	233	234	235	236	237	238	239
*av*GFP	V	L	L	E	F	V	T	A	A	G	I	T	H	G	M	D	E	L	Y	K
drFP583	I	V	E	Q	Y	E	R	T	E	G	R	H	H	-	-	-	-	L	F	L

Figure 1.3 Sequence alignment of *av*GFP and DsRed (drFP538). Chromophore forming residues are highlighted in grey.

1.3.2 Chromophore Formation

As mentioned in the previous section, the formation of the *av*GFP chromophore occurs from a post-translational modification of Ser65-Tyr66-Gly67, which occurs auto-catalytically in the presence of oxygen. The exact order of the chromophore formation steps is not completely determined due to several conflicting reports (52-57). However, the most recent paper by Strack et al. offers the most comprehensive mechanistic proposal, which also explains the formation of the red chromophore (58). This mechanism for GFP chromophore maturation is summarized in Figure 1.4. To initiate chromophore formation the peptide backbone of the central helix has an 80° bend to orient the three residues in the correct conformation and promote ring formation (54). In the first step the glycinyl nitrogen attacks the serine carbonyl carbon to form the five-membered imidazoline ring (intermediate I). This unstable intermediate is oxidized (58) to form intermediate II. The five-membered ring is trapped by a dehydration reaction to form the neutral phenol form. Reversible deprotonation forms the fluorescent anionic form (58).

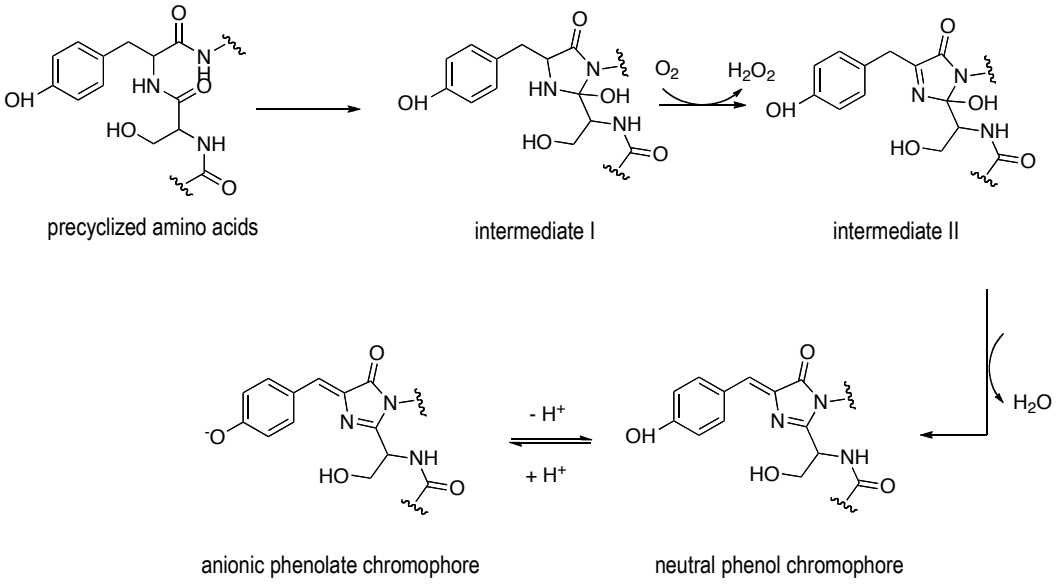


Figure 1.4 Mechanism of GFP chromophore formation

Initial reports on the formation of the red chromophore suggested a green intermediate because monitoring of fluorescence during drFP583 and DsRed maturation showed the disappearance of a green peak as the red form increased (49, 59-61). However, Verkhusha et al. found that in DsRed the amount of green intermediate formed increased throughout maturation and did not decrease, which should happen if it was a precursor to the red chromophore. They proposed that a transient blue intermediate forms, which can either irreversibly deprotonate to form the anionic green chromophore or undergo an oxidation and deprotonation to form the fully mature red chromophore (62). In their work with blue-to-red timers (63, 64) they discovered that the blue species was not the neutral green form, but instead an acylimine conjugated to the heterocyle ring (Figure 1.5).

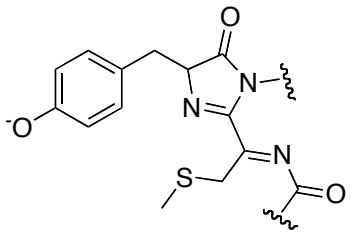


Figure 1.5 Structure of the blue intermediate proposed by the Verkhusha group (64).

The most up-to-date proposed mechanism of red chromophore formation is shown below in Figure 1.7 as proposed by Strack et al. Using a deuterated tyrosine residue they observed an increase in the red:green chromophore ratio, indicating that the dehydration step was slowed, but the oxidation branch towards the blue form was not impeded. This helped them confirm the branch in the pathway where the chromophore goes on to form either the green structure or red structure. Additionally, they determined the first oxidation step occurs before the dehydration by lowering the amount of O_2 available and recording a decrease in the red:green ratio. By comparing the appearance of H_2O_2 with the formation of the fully mature chromophores they determined that H_2O_2 production plateaus before the chromophore is fully mature, indicating oxidation occurs before dehydration. Finally, the presence of the acylimine blue form was supported by the continued presence of a blue form after oxidation was complete. If the blue form was the neutral green chromophore as initially suggested its production

should stop after oxidation is complete. Although it has been two decades since *av*GFP was first sequenced, we are only now are we gaining a clearer picture of the protein chromophore formation mechanism.

1.3.3 Three-dimensional protein structure

The fluorescence of FPs is influenced not only due by the internal chromophore structure, but also by the three-dimensional β-barrel structure. The arrangement and conformation of the tri-peptide promotes the formation of the chromophore (54) and the surrounding residues shields the chromophore from the external environment once formed. The crystal structure of *av*GFP was reported in 1996 and was revealed to be a β-barrel formed of 11 anti-parallel β-strands (2, 65). The protein forms a near-perfect cylinder that is 42 Å long with a diameter of 24 Å (Figure 1.6). There is a central alpha helix running through the centre of the barrel, which contains the chromophore forming tri-peptide. The chromophore is located in the centre of the helix and is approximately perpendicular to the barrel.

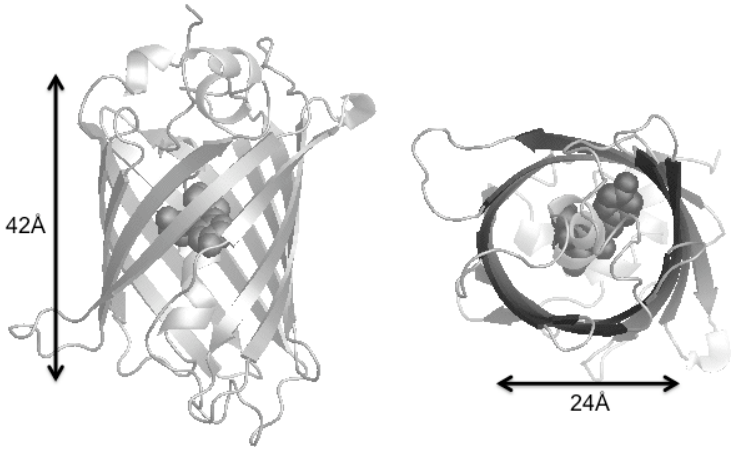


Figure 1.6 Crystal structure of *av*GFP. Protein Database (PDB) ID 1GFL (65).

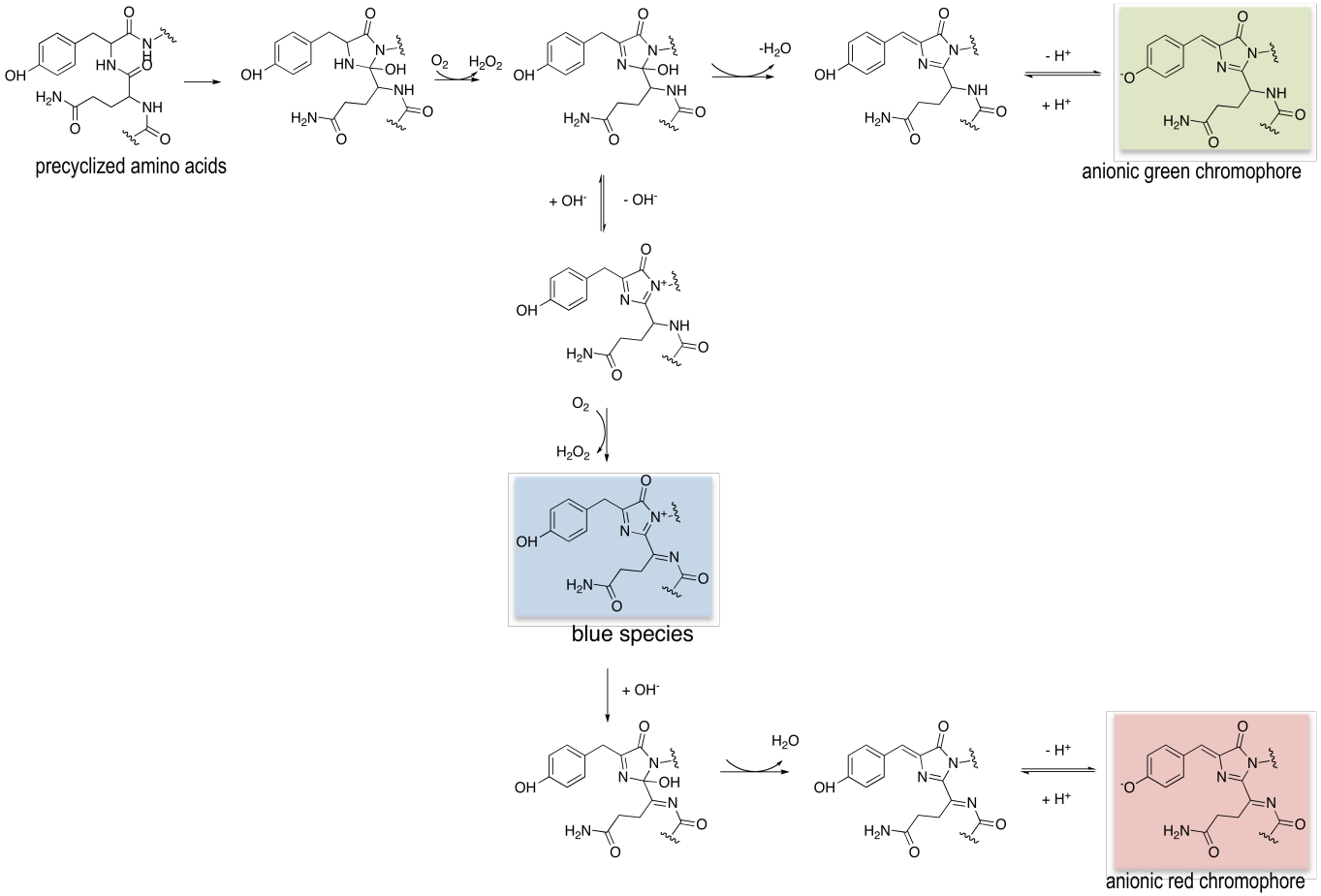


Figure 1.7 Branched pathway of red chromophore formation. Adapted with permission from Strack et al. 2010 Chromophore Formation in DsRed Occurs by a Branched Pathway. Journal of the American Chemical Society. 132, 8496-8505. Copyright 2012 American Chemical Society.

In 1999 the discovery of FPs from several species of Anthozoans was reported (24). The first representative crystal structure was that of DsRed, an obligate tetramer from *Discosoma striata* (66, 67). Subsequent work on DsRed confirmed the tetrameric structure as a requirement for fluorescence (59, 68, 69). Despite sharing only 20-30% sequence identity with *av*GFP, the structures of the barrels are remarkably similar, with the backbone α-carbons superimposing to ~1.0Å rms (66). The main structural differences between the proteins occur in the loop regions where DsRed has several truncated loop regions compared with *av*GFP.

1.3.4. Monomeric FPs

*av*GFP and its engineered variants are generally monomeric, except at protein concentrations in the high μM region. Most of the newly discovered green and red-shifted FPs were found predominantly as dimers or tetramers (70). DsRed is one of the more notable tetrameric RFPs, as it was the first RFP to be described in the literature (24, 66). Oligomerization of FPs can be problematic when targeting to specific organelles or when linked to proteins that form oligomers themselves (71-73). FPs can aggregate or cause unintended cross-linking when individual monomers associate to form higher order oligomers (Figure 1.8).

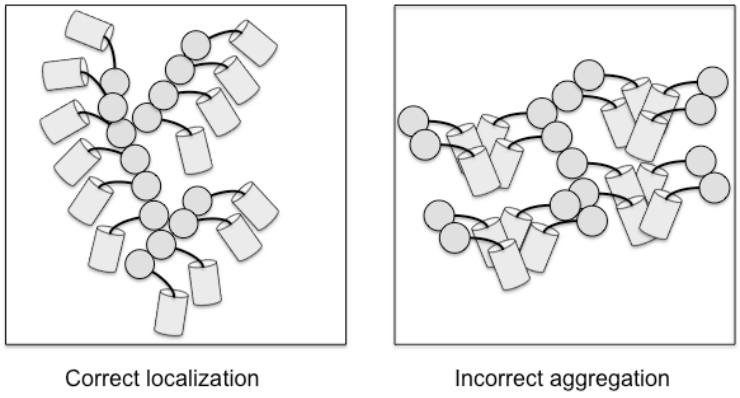


Figure 1.8 Schematic of correct localization and aggregation. Cross-linking can occur when using a tetrameric FP to label a filamentous protein.

For this reason a lot of work has gone into monomerizing newly discovered FPs. The monomerization of DsRed to generate monomeric RFP1 (mRFP1) was a herculean effort (43), which laid the groundwork for similar efforts including the

monomerization of *Clavularia* cyan fluorescent protein (74) and *Zoanthus* yellow fluorescent protein (Hiofan Hoi, unpublished results).

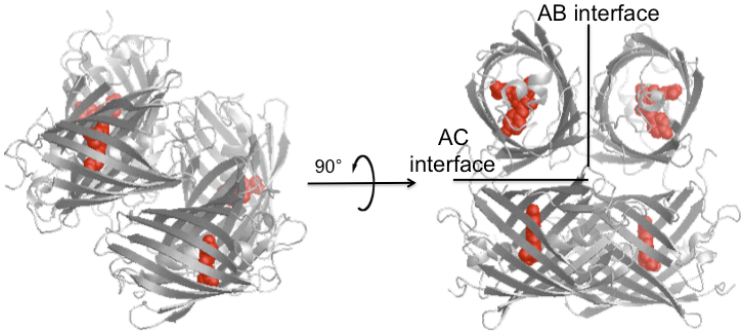


Figure 1.9 Tetrameric crystal structure of DsRed. The chromophore is shown as red spheres. PDB 1G7K (66).

DsRed forms a dimer of dimers with one pair arranged perpendicular to the other (Figure 1.9). In order to generate a monomeric RFP, two interfaces need to be broken, the AB interface comprised of mostly hydrophobic interactions, and the AC interface, which contains salt bridges and hydrogen bonds. Each interface was broken sequentially by introducing repulsive residue interactions and the fluorescence was restored through protein engineering until a red fluorescent monomer, mRFP1, remained. Since the initial publication of mRFP1 several more dimeric or tetrameric FPs have been engineered as monomers. A summary of several common monomeric FPs is shown below in Table 1.1.

Table 1.2 Summary of common monomerized FPs and their oligomeric precursors.

Monomer	Oligomer	λexc (nm)	λem (nm)	Reference
mTFP1	cFP484	462	492	(74)
mAG	Azami	492	505	(75)
mKO	KO	551	563	(76)
mKeima	Keima	440	620	(77)
tagRFP	eqFP578	555	584	(78)
mRuby	epFP611	558	605	(79)
mRFP1	DsRed	584	607	(43)
mKate	eqFP578	588	633	(21)
HcRed	hcCP	592	645	(80)
mPlum	aeCP597	590	649	(81)

Of particular importance to the work in this thesis is mRFP1, as it was the precursor to the mFruits, a series of monomeric FPs that spanned the visible range from green-yellow emission to red (25). mCherry is one of these mFruits and was used in the majority of the work presented in this thesis (25). The engineering of mRFP1 to generate various colours will be discussed in further detail in section 1.3.5.2 to follow.

1.3.5 Chemistry of the Chromophore

As mentioned in previous section 1.3.2, determining the exact mechanism of chromophore formation was not straightforward and it was nearly two decades before a consensus mechanism for both green and red chromophore formation was reported (58). Part of the reason for the uncertainty is the dynamic protein environment surrounding the chromophore. With the exception of a few key residues such as Gly67, Arg96, Glu222 (*av*GFP numbering) there can be a wide variety of sequences between FPs. This variable chromophore environment can have many effects on the chromophore. This includes changing the pK_a, excitation and emission wavelength, photostability, and other more complex behaviours such as photoswitching and photoactivation. All of these will be discussed in greater detail below.

1.3.5.1 pK_a and pH sensitivity

In Figure 1.4 and 1.6 the green and red chromophores are depicted as existing in equilibrium between a neutral phenol and an anionic phenolate structure. The distribution of the two forms depends on the pH of the surrounding environment as well as the pK_a of the chromophore. For fluorescent proteins the effective pK_a is defined as the pH where there is 50% of the maximum fluorescent intensity, which is typically assumed to correspond to a 50% mixture of neutral and anionic chromophore. A typical pH titration curve with a pK_a of 7 is illustrated below in Figure 1.10.

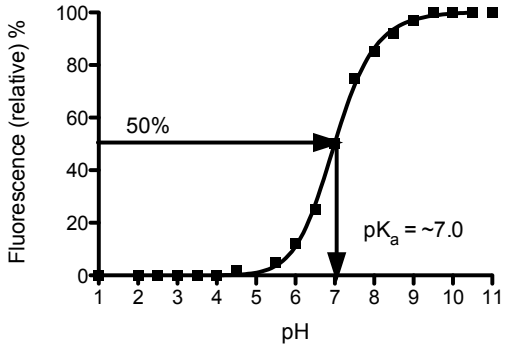


Figure 1.10 Typical pH titration curve. 50% fluorescence intensity, corresponding to pH 7.0, is the pK_a.

The fluorescence of *av*GFP is stable between pH 6 and 9, but the denatured chromophore has a pK_a of 8.1 (82). Tyrosine normally has a pK_a of 10.9, but the 2.8 unit decrease is attributed to the extended conjugated system, which increases the acidity of the phenol hydrogen because of increased stability of the negative charge. Additionally, neighbouring amino acids such as His148, Thr203, and Ser205 hydrogen bond to the phenolate and further stabilize the negative charge (65, 83). In *av*GFP two absorbance peaks at 395 nm and 475 nm, correspond to the neutral and anionic form, respectively.

EGFP contains the Phe64Leu and Ser65Thr mutations, which result in absorption only at 475 nm (2), but the protein is more pH-sensitive with a pK_a of ~6.1 (84-86). EGFP has much improved maturation times and efficiency, but at the cost of sensitivity to pH values only slightly lower than physiological pH. As with many FP engineering results, a comprise must be made as one characteristics improves at the expense of another.

pH-sensitivity of FPs can be problematic or advantageous depending on the application of use. Work to reduce the pH sensitivity of yellow FP (YFP) variants (87, 88) led to the generation of Citrine (pK_a 5.7) and Venus (pK_a 5.8) These YFPs are commonly used as FRET-pair acceptors with cyan-emitting FPs (CFPs). A pH-insensitive GFP, T-Sapphire, has also been generated that has a pK_a of 4.9 and has been used to generate a NADH:NAD^+ biosensor that is insensitive to pH (36, 89).

Alternatively, the pH sensitivity of EGFP and its mutants have been used to measure the pH of the cytosol, mitochondria, and Golgi (84, 85). An emission ratiometric GFP-variant, deGFP was used to monitor pH in mammalian cells with SNARF-1, a synthetic orange-red fluorescent pH-sensitive dye (90-92). A pH-sensitive RFP, mNectarine was also used to detect concentrative nucleoside transport in human embryonic kidney (HEK293) cells by fusing the protein to the human concentrative nucleoside transporter, hCNT3 (6).

Most RFPs have pK_a's lower than 6 (44), which is an added advantage for their use in cellular applications along with the longer excitation wavelength, and possibility of two-photon excitation. A low pK_a means that for all processes occurring around physiological pH (7.4) any small local fluctuations in pH will not affect the fluorescent signal of the RFP. DsRed and mRFP1 have pK_a's of 4.7 and 4.5, respectively, while the mCherry pK_a was measured at less than 4.5 (25, 43).

Despite the desire for most FPs to be pH insensitive, the design of single FP-based biosensors have resulted in proteins that are pH sensitive and in fact owe their function to a pK_a modulation mechanism. This pH sensitivity has been determined for several single FP-based biosensors, including those for Ca^{2+}, Cl^-, ATP:ADP, cGMP, and hydrogen peroxide (7, 14-16, 37, 38, 40, 93). The Ca^{2+} biosensor R-GECO1 has a pK_a of 8.9 in the absence of Ca^{2+} and a pK_a of 6.59 in the presence of Ca^{2+} (14). This means at physiological pH more of the protein will have a protonated chromophore and be non-fluorescent. However, upon Ca^{2+} binding the pK_a drops to 6.59 meaning that the amount of anionic, fluorescent protein increases. Figure 1.11 is a pH titration curve for R-GECO1 with and without Ca^{2+} that better illustrates the change in fluorescence output by including a curve that represents the dynamic range of the protein. Th dynamic ranges of R-GECO1 and other GECOs are determined by F_{max}/F_{apo} -1.

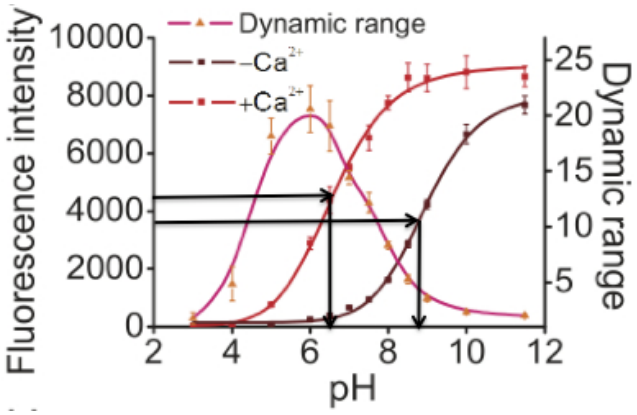


Figure 1.11 pH titration curve of R-GECO1 along with best-fit lines. The dark-red circles are R-GECO1 without Ca^{2+}. Red circles are R-GECO1 with Ca^{2+}. Indicated with black arrows are the approximate pK_a's of both forms. Copyright 2011 by the American Association for the Advancement of Science; all rights reserved. Zhao, Y. et al. An expanded palette of genetically encoded Ca2+ indicators, Science, 2011, 333, 1888. Published by Wiley Science.

R-GECO1 was engineered from mApple, which is a photostable red FP with a pK_a of 6.5. The contrast of pK_a's between mCherry and mApple have significant repercussions on the design and mechanism of action when used in single-FP based Ca^{2+} biosensors. This contrast will be discussed in more detail in later Chapters. It is apparent that the chromophore's protonation state plays a very important role in determining the fluorescence and can be either advantageous or disadvantageous depending on the proteins intended use. With this in mind, care must be taken when choosing an FP to do imaging experiments or when designing FP-based biosensors (94).

1.3.5.2 Colour-Structure relationship

The fluorescent colour of FPs is not only due to the conjugated chromophore structure found in the centre of the β-barrel protein, but also to the internal environment that promotes chromophore formation, prevents solvent access, and restricts vibrational modes of relaxation upon excitation. The first discovered FP, *av*GFP, was green, but thanks to advances in molecular biology such as error-prone polymerase chain reaction (EPPCR) and site directed mutagenesis (95-98), changes in the protein structure were introduced at the genetic level to alter the spectral properties (47, 99). Altering the colour (i.e., excitation and emission wavelengths) of a particular FP can be done in several ways. The colour can be

altered by directly mutating the chromophore forming residues or mutating amino acids in the chromophore environment. An alternative approach that has become popular in recent years but only works with certain variants, is to use high powered laser light to induce chemical changes in the chromophore structure. This latter approach, used for photoactivatable and photoswitchable FPs, will be discussed in further detail in the next section. The focus for this section will be on genetic mutations introduced into the chromophore and its surrounding environment.

Before discussing how the variety of FP colours were generated, a brief look at why they were made is pertinent. The ability to genetically link a fluorescent tag to virtually any protein of interest has helped to revolutionize the world of fluorescence microscopy (49). In order to differentiate between structures or monitor cellular interactions, two spectrally distinct fluorescent proteins need to be used. This means that the wavelengths of light used for excitation of one FP should not excite the second FP (Figure 1.12). Because FPs have broad excitation and emission bands, the use of bandpass filters is essential for limiting the amount of cross-talk between FPs. Additionally, the use of spectral unmixing can help to determine what percentage of the emission is due to unintended excitation of the second FP (100, 101). The desire to monitor several different cellular processes along with a healthy curiosity about *av*GFP lead researchers to generate several different colour mutants (47). Shortly thereafter, a more concerted effort went into developing new colours with ideal characteristics for imaging such as fast, complete maturation, spectrally distinct wavelengths, and increased photostability (48, 99, 102, 103).

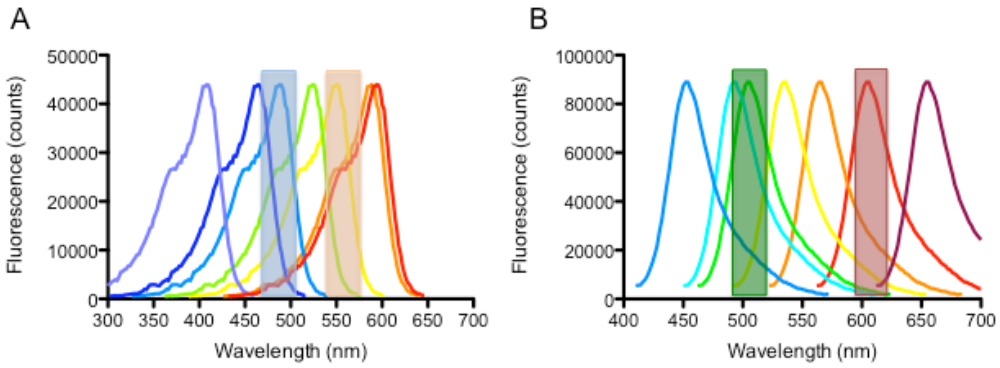


Figure 1.12 Excitation and emission spectra of different coloured FPs. A) Excitation spectrum indicating potential filter regions used to excite a green and red FP. B) Expected emission spectrum of several FPs. Coloured rectangles indicate bandpass filters wavelength ranges used to collect fluorescence from a green and red FP. In both excitation and emission spectra there is little overlap between the green and red peaks, which will limit the amount of background cross-talk signal.

Figure 1.13 is a brief summary of some of the mutagenesis that was done to *av*GFP to generate FPs with different colours. At a quick glance the FPs with alternate chromophore structures are blue FP (BFP), EBFP, and mKalama. EGFP has a more subtle change in chromophore structure, which is the Ser65Thr substitution, which influences the protonation state of the chromophore. Thr65 is too large to hydrogen bond with the phenolate group and as a result the protonation equilibrium shifts to the anionic state. Superfolder (sf) GFP contains the EGFP mutations, folding reporter GFP mutations, and six additional mutations that were found after extensive engineering, including Ser30Arg, Tyr145Phe, and Ala206Val (104). The introduction of Tyr66Trp decreases the chromophore's conjugation resulting in a CFP with shorter excitation and emission wavelengths with substantially reduced brightness. Introduction of the same folding mutations, Phe64Leu and Ser65Thr, found in EGFP resulted in a cyan protein, ECFP, with improved brightness albeit still dimmer than EGFP (47, 48, 105). Likewise with BFP, the Tyr66His mutation reduces the conjugation and blue shifts the excitation and emission peaks. Similar to ECFP, the initial construct had reduced brightness, but introduction of EGFP mutations lead to EBFP, and introduction of the sfGFP mutations with further engineering produced the very bright EBFP2 (47, 103, 106). mKalama is a BFP that contains the Tyr66 chromophore shifted towards the protonated state. It underwent several rounds

of mutagenesis to rescue the brightness although it is not as bright as EBFP2. Though mutation of position 66 can lead to hue-shifted variants, the surrounding environment also plays a critical role in shifting the fluorescence wavelengths of FPs. Mutation of the chromophore structure can lead to significant re-arrangements of neighbouring amino acids that are not optimal for fluorescence. Steric or electrostatic clashes can perturb the planarity and electron distribution of the chromophore; for this reason FPs mutants must undergo extensive engineering to rescue the fluorescence. However, once mutations to improve folding, maturation, etc. have been identified, as in the case of sfGFP, introduction of these same mutations into different FPs often confers similar benefits.

Alternatively, EYFP, Citrine, Venus, Sapphire, and monomeric teal FP (mTFP1) all contain mutations surrounding the chromophore that shift the excitation or emission wavelengths. mTFP1 was engineered from cFP484, a tetrameric protein from *Clavularia* coral (74). It shares the same chromophore structure as EGFP, but a salt bridge network keeps nearby His197 protonated. This protonation is thought to limit charge transfer in the excited state, resulting in the slight blue shift in excitation and emission wavelengths (74, 107). From the crystal structure of EGFP, Thr203 is in close proximity to the phenolate of the chromophore. Ormo et al. and Heim et al. found that introduction of Tyr in the place of Thr203 causes a large red-shift in excitation and emission to produce a yellow FP, due to π-π stacking of the aromatic rings (47, 65). Again the same EGFP folding mutations were introduced into YFP to generate EYFP (65). Unlike GFP, YFPs were also plagued by problems of pH sensitivity due to their lower pK_a and chloride sensitivity as mentioned in section 1.3.5.1. A lot of work went into improving brightness and reducing the pH and chloride sensitivity and two new variants Citrine and Venus were engineered (87, 108). In Citrine a Gln69Met mutation reduced the pK_a, improved expression at 37 °C, increased photostability, and reduced chloride sensitivity. Venus was also generated from EYFP, by introducing the GFP*mut3* and EGFP mutations (99, 103), while further screening found the Phe46Leu mutation, which vastly improved the maturation efficiency by promoting the oxidation step (15).

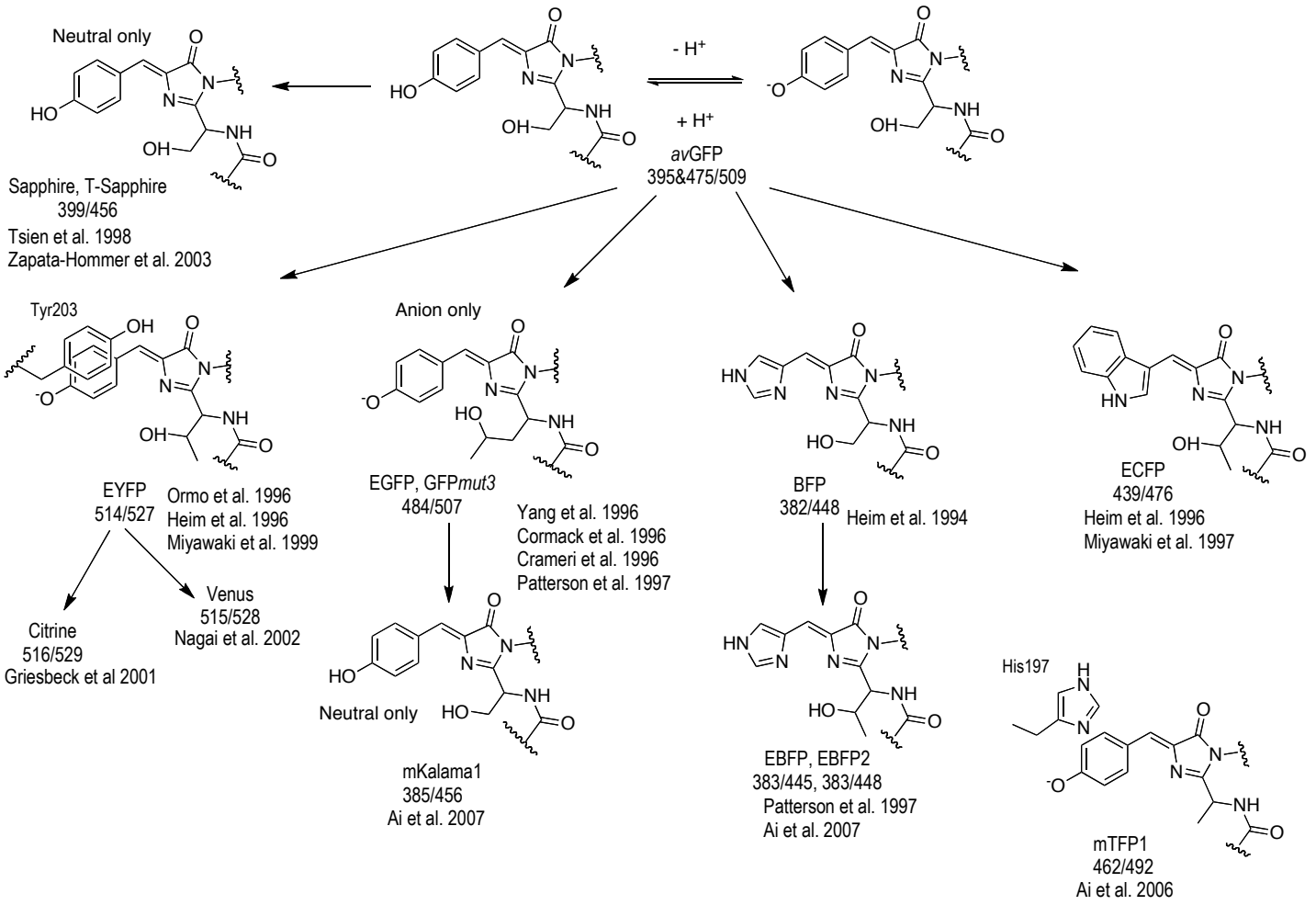


Figure 1.13 Flow-chart of *av*GFP colour evolution with corresponding chromophore structures. References for particular mutants are included below the FP name.

The importance of the chromophore environment in influencing the spectral wavelengths is more apparent in the evolution of the mFruit series from DsRed (25, 109). The generation of mRFP1 from DsRed required 33 mutations, of which 13 were internal in the β-barrel. mRFP1 has an excitation/emission shift of ~20 nm relative to DsRed, which has been attributed to several different structural changes (43). The Lys163Met and a few other substitutions remove the salt bridge with the phenolate of the chromophore, shifting the polarization of the chromophore. Additionally the Lys70Arg, Ser197Thr, and Thr217Ser, which are located in a plane under the chromophore, played an important role in generating mRFP1.

Subsequent to the initial report of mRFP1, Shaner et al. improved upon mRFP1 and generated new colour variants (25). mRFP1.1 contains a Gln66Met mutation and red-shifts excitation and emission another 5 nm. Continued screening for improved folding and brightness led to the discovery of the Met163Gln mutation and others (25). This work ultimately led to the production of mCherry, the protein used in this thesis. Replacement of Tyr66 with a Trp, analogous to CFP, resulted in a blue-shifted variant known as mHoneydew, with emission wavelengths of 537 and 562 nm. Position 66 within the chromophore was tolerant of several different amino acids and several blue-shifted variants were discovered. The Met66Cys mutant led to mTangerine, while the introduction of Glu at position 197 resulted in the yellow protein, aptly named mBanana. The Met66Thr mutation led to the development of mOrange and the orange-red protein mStrawberry, which emit at 562 and 596 nm, respectively. In a 2006 paper by Shu et al., the crystal structure of these two proteins along with mCherry provided structural insights into the varied emission wavelengths. Interestingly, there is a unique cyclization that occurs at Thr66 in mOrange that results in the blue shift. Both mStrawberry and mCherry contain a similar chromophore structure as mRFP1.1, with the key difference being Thr66 in mStrawberry. The reason for the difference in emission wavelengths is due to the surrounding residues such as Lys83Leu and Glu215. The Lys83Leu mutation causes a shift in Lys70 such that it now interacts with Glu148. Additionally, Glu215 in mCherry is protonated and hydrogen bonds with the imidazolinone ring

nitrogen. While in mStrawberry there may be a mixture of protonated and deprotonated Glu215.

Screening for a photostable mOrange lead to mApple, which shares the same chromophore structure as mRFP1.1, but has a slightly blue-shifted emission maximum of 592 nm (109). Mutation of Tyr67 to Phe in mCherry resulted in a BFP named mBlueberry. Improvements were made by including some mApple mutations to generate mBlueberry1 and mBlueberry2 (106).

An orange FP with a long Stokes-shift was recently reported. This protein absorbs at 430 nm and emits light at 572 nm (110). A long Stokes-shift is also present in *av*GFP as well; excitation of the protonated chromophore at 375 nm results in emission at 509 nm. The mechanism for this behaviour was determined to be an excited state proton transfer (ESPT). When the protonated chromophore is excited the phenol proton becomes very acidic and is transferred through a proton transfer network to ultimately end up on Glu222, leaving fluorescence emission to come from the phenolate form (83, 111, 112). T-Sapphire contains only the phenol chromophore, but its emission is slightly blue-shifted to 499 nm (51, 89). Prior to LSS-mOrange, two LSS-RFPs were engineered from mKate, a monomerized RFP engineered from *eq*FP578 (113). The authors found that either the Ser141Asp/Ser158Asp or Met160Lys/Glu/Asp mutations are sufficient to introduce the LSS phenotype. Introduction of some negatively charged residues near the chromophore also helped stabilize the neutral form of the chromophore. Other mutations were identified during the course of protein engineering. A variety of LSS RFPs were generated by screening libraries of mutations at Ser158 and Glu160 (mKate numbering) (114). Subsequently, a more comprehensive study of an LSSmOrange variant was reported (110).

Finally, more drastic chromophore chemistry changes were engineered into mOrange and mCherry to generate a photoswitchable (PS) mOrange variant, as well as a photoactivatable (PA) mCherry (115, 116). These variants will be discussed in the next section.

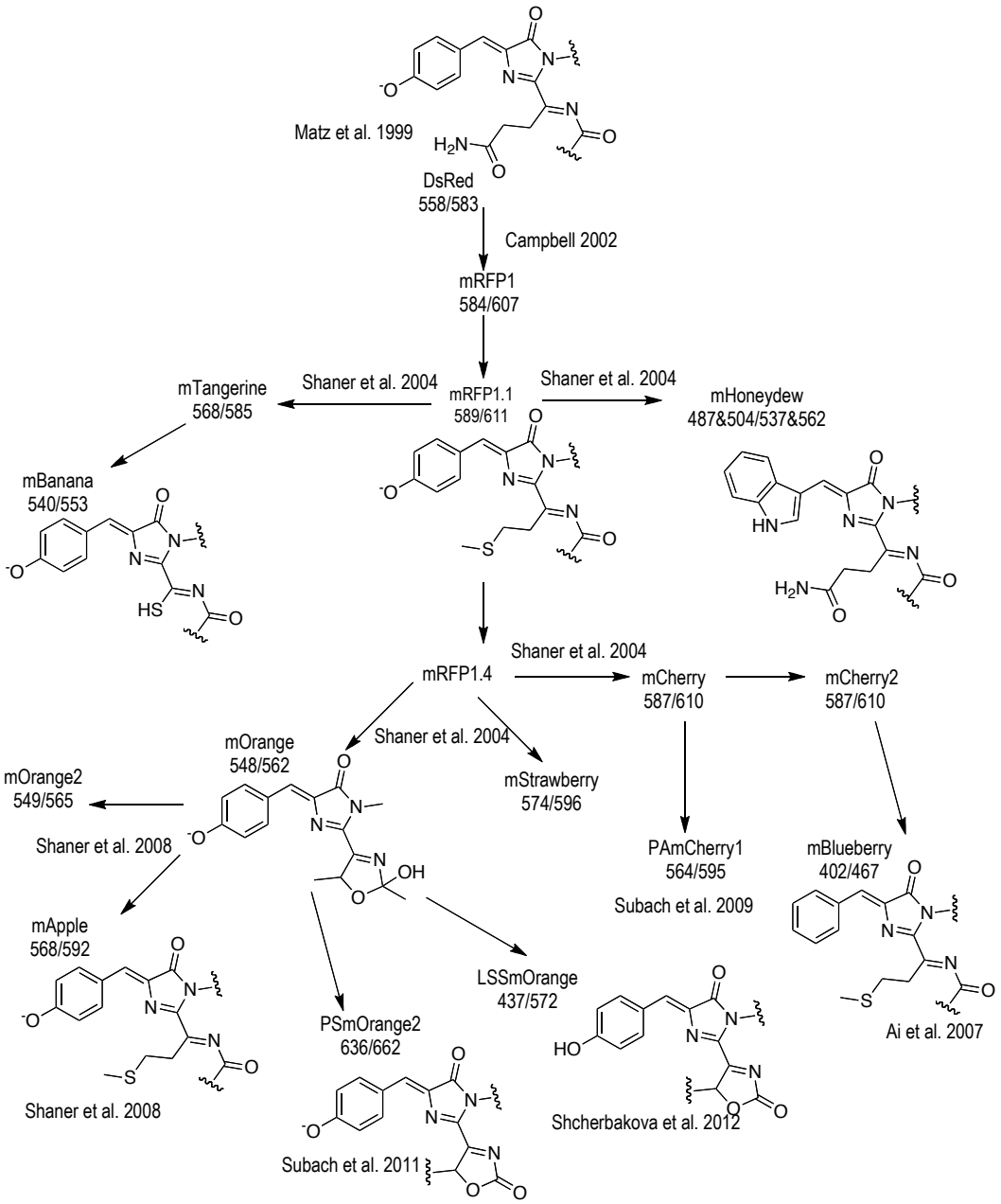


Figure 1.14 Alternate chromophore structures engineered from DsRed. Excitation and emission wavelengths are listed, separated by a backslash, along with appropriate references.

1.3.5.3 Photoswitchable and Photoactivatable FPs

A more recent addition to the battery of FPs was the development of photoswitchable and photoactivatable FPs. These types of proteins can be classified into three categories: (1) irreversible photoactivation from a dark to bright state; (2) irreversible photoswtiching from one colour to another; and (3) reversible photoswitching from a bright/dark state or between two colours (117).

Photoswitching behaviour of GFP mutants was first reported in 1997 where protein was immobilized in polyacrylamide gels and fluorescent blinking was monitored (118). Exposure of mTFP0.7 to high-powered laser light resulted in a reversible photobleaching of the protein (74, 119). Crystal structure determination of the dark and bright states revealed that isomerization from *cis* to *trans* occurred along the chromophore's Cα-Cβ double bond of Tyr67 (Fig 1.15) (120). In mTFP0.7 the chromophore normally has a *cis*-double bond, but flips into *trans* configuration when excited with 460 nm laser light (Figure 1.15). This structure is non-fluorescent and the protein will convert back into the *cis* form when left in the dark.

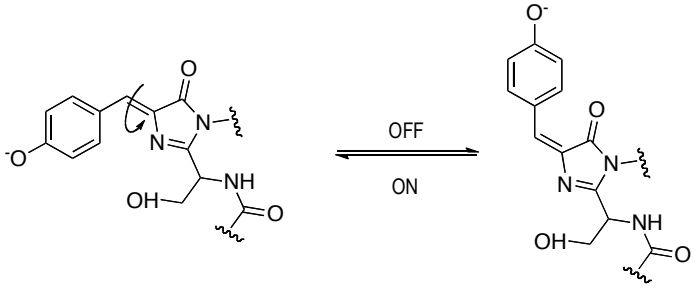


Figure 1.15 Reversible photoisomerization in mTFP0.7. The bright state contains the *cis* bond and the dark state contains the *trans*.

The first deliberately designed irreversibly photoactivable FP was PA-GFP, which contained the Thr203His mutation and was not green fluorescent until irradiated with intense violet light (390-415 nm) (121). The increase in green fluorescence is attributed to a decarboxylation of Glu222, which rearranges the chromophore environment to increase the proportion of chromophore in anionic form. There are now over two-dozen reported photoactivatable/switchable proteins, but the colours are somewhat limited. There are green and red reversible PAFPs, such as Dronpa, rsFasttime, Padron, KFP1, rsCherry, rsCherryRev, and rsTagRFP that have been summarized nicely in several reviews (44, 117). Irreversible PAFPs are either green (PA-GFP) or red, in the case of PAmCherry1, PATagRFP, and PAmKate (116, 121-123). More complex examples are mIrisFP and NijiFP, which are irreversibly photoswitched from green to red, but both the red and green forms can also be reversibly photoactivated from bright to dark states (124, 125). The most commonly used irreversible PSFPs are those that can be converted from green-to-red and

include Dendra2, mEos2, mKikGR, mClavGR, and recently reported mMaple from our lab (124, 126-129). Recently the colour palette was expanded by the introduction of PSmOrange, a PSFP that emits at 562 nm before irradiation with blue-green light, and emits far-red light at 662 nm after irradiation (115). The chromophore structures of PSmOrange, PAmCherry1, and mMaple are shown below in Figure 1.16A. PAmCherry1 was one of the first red PAFPs, which is in the dark state before irradiation with 405 nm light (Figure 1.16B) (116).

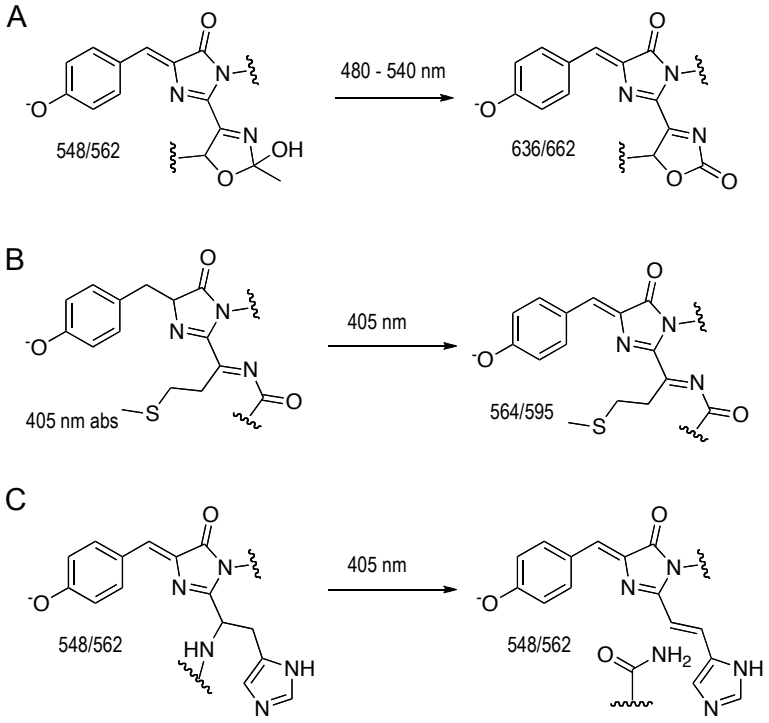


Figure 1.16 Photoswitchable and photoactivable RFPs. A) PSmOrange chromophore structure before and after photoswitching. Illumination with blue-green light extends the conjugation B) PAmCherry1 in the dark state, after illumination with 405 nm light the double bond is restored to generate an mCherry that emits at 595 nm. C) mMaple in the green form, after 405 illumination the chromophore is cleaved off the peptide backbone resulting in a new double bond.

All of the green-to-red PSFPs contain the His-Tyr-Gly tri-peptide, where cleavage of the peptide bond results in a red chromophore (Figure 1.16C). Aside from the addition of this mutation, additional beneficial mutations have been discovered using EPPCR mutagenesis and screening for photoconversion properties in colonies of *E. coli* (129). For PAFPs the complete maturation of the chromophore must be disrupted to form the dark form, and mutations of Ser148,

Ile165, Gln167, and Ile203 (GFP numbering) were screened to find PAmCherry1 (63).

Photoconversion in FPs is more widespread and can be detected in photobleaching experiments with cyan and yellow FPs (109, 130, 131). This means that care must be taken when imaging FPs for extended periods of time and it may be necessary to test possible photoconversion behaviour of FPs even if this behaviour has not been described previously. It was recently discovered that R-GECO1, a Ca^{2+} biosensor undergoes photoactivation/switching behaviour. Specifically, when illuminated with blue light the amount of red fluorescence increases (unpublished results). Despite the added precaution that must be taken, the seemingly widespread phenomenon can be used to develop PAFPs and PSFPs with different colours and properties.

1.3.6 Fluorescent Proteins as Biosensors

1.3.6.1 Circular Permutation

Circular permutation is a naturally occurring topological manipulation of a protein that can be thought of as a genetic rearrangement that effectively links a protein's N- and C-termini. New termini are generated at a split point elsewhere in the protein (Figure 1.17). This modification was first reported for the plant lectin, favin, which was found to be a circular permutation of Concanavalin A (Con A) (132). Circular permutations have now been identified across several different protein families including transaldolases, glucosyltransferases, and β-glucanases (133). The discovery of several more naturally occurring circular permutated proteins was aided by the development of sophisticated bioinformatics that allowed the comparison of protein sequence and structure (134-136). Recently, an extensive database of all naturally occurring circular permutations was established (137).

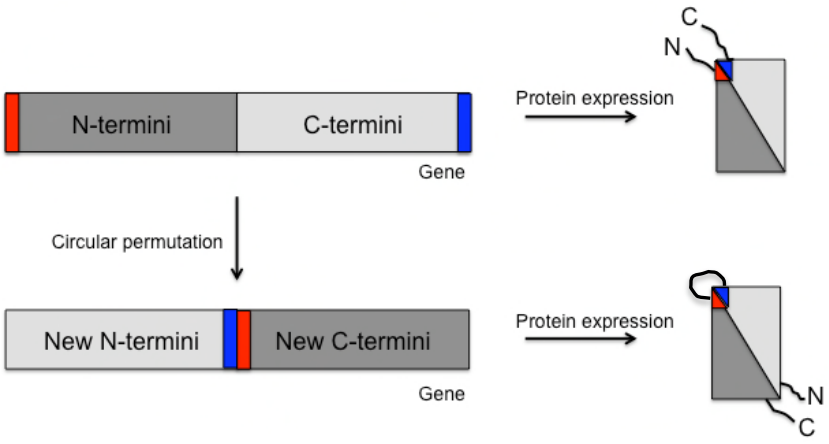


Figure 1.17 Schematic of circular permutation of a protein. The 5' (N-terminal) and 3' (C-terminal) portions of a gene encoding a protein are swapped, such that the C-terminal portion is now the new N-terminal fragment. The original amino and carboxy terminus, indicated in red and blue, respectively are now linked together.

Though the precise biological/evolutionary reason for circular permutation has remained elusive (138, 139) there are two main mechanisms by which natural permutation can occur. The two mechanisms are gene rearrangement involving duplication/deletion events or fission/fusion of gene fragments (Figure 1.18) (134, 135, 139, 140). The first examples of natural circular permutations thought to occur by a gene rearrangement were the saposins, where a saposin-like domain was found in a plant aspartic proteinase (141). Subsequent analysis of the protein sequence and genes determined the insert was a circular permutation of the saposins and was named a swaposin (139).

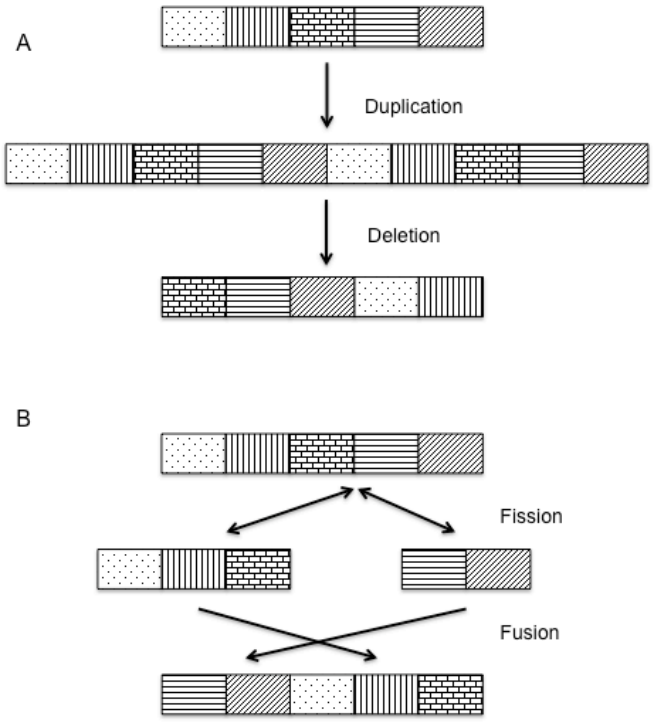


Figure 1.18 Mechanism of naturally occurring circular permutation. A) Duplication/deletion B) Fission/fusion model

The consequences of a circular permutated protein on structure and function can vary widely depending on the permutation location and protein in question. The presence of naturally occurring circular permutations must indicate these proteins fold appropriately and retain some of the their intended function, or perhaps gain a new, altered function. As it turns out, investigation of proteins related by circular permutation have determined that many are structurally similar (133) and retain a similar biological function (142, 143). In fact the basis for the development of the Circular Permutation Search Aided by Ramachandran Sequential Transformation (CPSARST) search tool was the premise that circular permutated proteins retain their native structure (136).

Advances in molecular biology enabled scientists to artificially construct circular permutated proteins in the laboratory using three different methods (144). When the site of permutation is known, fragment polymerase chain reaction (PCR) followed by overlap-extension PCR (OE-PCR) can be used (Figure 1.19A). Another method is the use of concatemers, whereby a tandem repeat of the gene is amplified or cloned into plasmid DNA (Figure 1.19B). Using specifically designed primers the desired circular permutated gene can be amplified. Finally,

limited DNAseI digestion is used for the generation of circular permutated libraries when the exact split site is unknown. Intramolecular ligation of the gene is done by introducing complementary 5' and 3' ends via PCR, which can be self-ligated to generate circular DNA (Figure 1.19C) (145). Once circularized the DNA is randomly linearized by digestion with DNaseI.

Common to all unnaturally generated circular permutated proteins is the need for a linker between the original N- and C-termini. A majority of protein termini are found in close proximity, <5 Å (146, 147), which means that relatively short amino-acid linkers should be sufficient to connect the two. As a first approximation Gly-rich linkers are used with the assumption that their flexible nature will not introduce any steric restraints between the two termini. Despite this favoured composition there have been reports of increased loop length and extended Gly-loop structures reducing the stability of proteins (148). Improvements in computational structure prediction have lead to the design of more optimal loop configurations including the use of starting loop compositions similar to those found in the native protein (149-151).

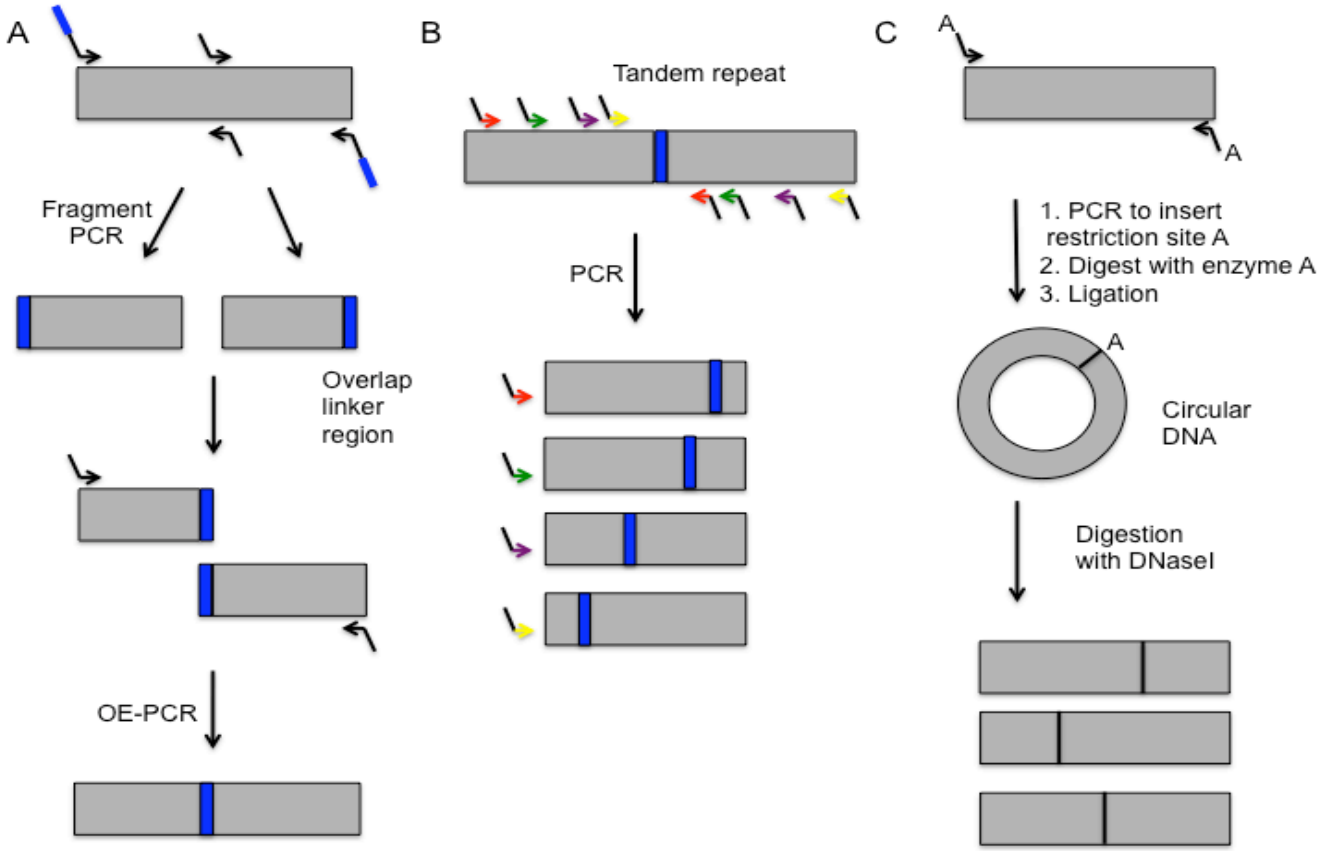


Figure 1.19 Methods for generating circular permutations in the laboratory. A) OE-PCR: primers have pre-determined split site and must include the linker between the N- and C-termini. B) Tandem repeats: two copies of the same gene are linked together with the pre-determined linker (can be cloned into a plasmid). Primers with various split sites can be used to generate multiple circular permutated variants. C) Random digestion: DNA has same restriction site added to 5' and 3' ends for blunt-end ligation, DNA is digested with the enzyme and self-ligation products are purified. Finally, limited DNaseI digestion generates circular permutated variants. All gene products must be cloned into desired plasmid to screen for viable protein.

The first application of circular permutation with fluorescent proteins was a happy accident. During mutagenesis of ECFP, Baird et al. discovered a variant with six residues inserted at position 145 that maintained its fluorescence (8). From this result they postulated that the protein halves split at position 145 could still fold into the correct FP structure. They tested this hypothesis by generating EYFP constructs permutated at position 145 and found the protein was still fluorescent. The linker they chose for linking the original termini was Gly-Gly-Thr-Gly-Gly-Ser, presumably to maintain linker flexibility. A systematic study of circular permutated GFP*uv* used a similar Gly-Ser-Gly-Gly-Thr-Gly linker to connect the original termini (12). Though work has been done to study the optimal linker length between a cpFP's N- and C-termini, the composition of the various linkers was a combination of Ser, Thr, and Gly residues (152). It is curious that with all of the work required to optimize a cpFP precious little is done to optimize this linker, while an inordinate amount of work is done to optimize linkers connecting protein domains in FP-based biosensors (14, 153-155). Mutations to the linker in cpFPs have been characterized during the normal course of random mutagenesis (13, 14) (unpublished results), which should indicate that some effort should go into the design.

Generation of cpFPs is predominantly done using either OE-PCR or by tandem repeats, with the former being used when circular permutation sites are known in advance, while the latter can be used to generate libraries of cpFPs. An additional screening method, transposon-based mutagenesis was used with mCherry to determine sites that tolerated insertion of a short peptide (23). From these results six sites were identified and circular permutations were made at each site, though only two of the six proteins could be expressed in soluble form. A benefit of generating cpFPs (as opposed to other types of proteins) is that screening for proteins that retain fluorescence is straightforward, though in some instances colony brightness does not translate to high protein expression levels (156).

After the first published work with cpEGFP, a variety of cpFPs were generated to study the stability and folding processes. As mentioned above circular permutated proteins tend to retain their native structure, but with altered

stabilities. *av*GFP and its mutants are very stable even in the presence of denaturants and unfolding kinetics are very slow, reaching equilibrium over a period of months (157, 158). Several studies looked at the mechanical unfolding of cpGFP (position 145 and 173) (159) and cpEYFP (position 145 and 174) (160), which both show reduced mechanical stabilities. Intact GFP has been used to report the unfolding behaviour of fibronectin. So the potential exists to use cpFPs to report the unfolding of other, less stable proteins (161).

Other work with cpFPs looked at the tolerance of specific FPs to circular permutation. Topell et al. made several cpEGFP constructs, while Pedelacq et al. compared the circular permutation tolerance between sfGFP and folding reporter GFP (12, 104). They both found there was greater tolerance for circular permutation in the loop regions and sfGFP tolerated permutation better than folding reporter GFP. From this work it was apparent that a more stable and efficient folding pathway would better tolerate introduction of circular permutation. Preliminary work has also been done to generate cpRFPs, with the initial efforts focussing on cpmCherry variants (23). Very few sites tolerated permutation and those that did had poor folding efficiency, even when located in the loops. Subsequent work with improved cpmCherry variants will be described in Chapter 2. More recently a comprehensive look at cpmKate and additional cpmCherry variants was published (42). Interestingly, neither effort led to the discovery of a fluorescent variant permutated at the position analogous to the common cp145EGFP and cp145EYFP. This could be due to slight differences in the structure of the β7 strand. In mCherry and mKate there is only a relatively short region of structure coil in β7 from residues 143-146, while *av*GFP and its mutants β7 have a region of structure coil that extends to residue 147 (Figure 1.20). The additional flexibility in the GFP β-strand may allow GFP to better tolerate permutation at positions within the strand.

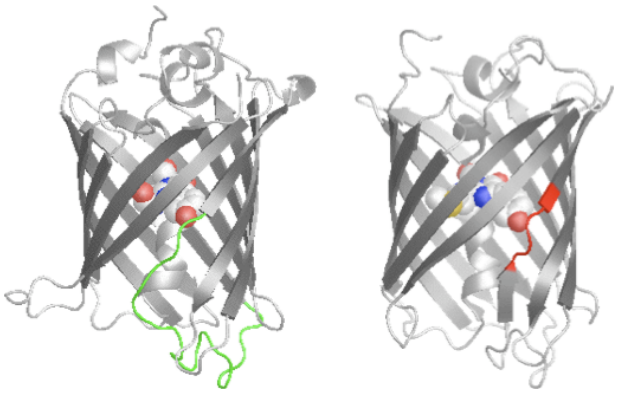


Figure 1.20 Crystal structure of *av*GFP and mCherry. On the left the unstructured loop and β7 of *av*GFP is shown in green. On the right the unstructured section of β7 is shown in red. PDB 1GFL and 2H5Q, respectively.

1.3.6.2 Inteins

Inteins were discovered over 20 years ago (46, 162). The term intein refers to internal protein, which is sandwiched by flanking polypeptide sequences termed the extein, for external protein. The post-translational modification excises the intein fragments and joins the N-terminal and C-terminal extein sequences, reconstituting an intact protein. Protein splicing is the term coined for this ligation event (163) and occurs via a four-step mechanism without the need for any external enzymes or cofactors. Since their initial discovery, over 400 inteins have been documented in Eukarya, Eubacteria, and Archea, as well as in viruses and phages (164).

Most inteins contain two domains, an endonuclease domain and the domain responsible for splicing (165-168). The splicing domain is of greater interest to scientists hoping to exploit protein splicing; so several inteins were generated that only contained the splicing domain (169-171). There are currently three classes of inteins (172), classified by the mechanism of splicing. For the purposes of this work only Class 1 inteins will be discussed. Class 1 inteins contain several conserved residues including a Cys or Ser at the N-terminus, along with either a Cys, Ser, or Thr directly following an Asn at the C-terminus. These residues play important roles in the splicing mechanism, which will be discussed shortly. Class

2 and Class 3 inteins contain alternate amino acids at these positions and use a different mechanism for protein splicing.

As mentioned above, protein splicing occurs via a four-step mechanism (Figure 1.21) (171, 173, 174). For intact inteins the splicing is referred to as *cis* splicing, while splicing of split inteins is referred to as trans splicing. The first step is either an N-S acyl re-arrangement or an N-O shift if the N-terminal residue is a Cys or Ser, respectively. Then the (thio)ester bond is attacked by either the OH- or SH-group of the C-terminal residue, which leads to transesterification, whereby the N-extein is transferred to the side chain of the C-extein. In the third step the conserved Asn residue cyclizes to cleave the peptide bond and release the extein product. Finally, the (thio)ester bond rearranges to form a peptide bond by another S-N or O-N acyl shift. The succinimide group on the intein can be hydrolyzed to re-generate the carboxy termini. *Cis* splicing of an intact intein occurs via the same mechanism except the N-intein and C-intein are already attached together and do not need to associate to initiate splicing.

The first discovered inteins were one intact protein, but several years later naturally occurring split inteins were identified that would self-associate and catalyze splicing in *trans* (175, 176). The most commonly used split intein was from *Synechocyctis sp.* (*Ssp*), until the discovery of the intein from *Nostoc punctiform* (Npu), which has an improved splicing efficiency *in vitro* of >98% (176), larger than those reported for *Ssp* (177, 178). Compared with *Ssp, Npu* splicing efficiency was more tolerant of alternate amino acids following the C-terminal Cys residue.

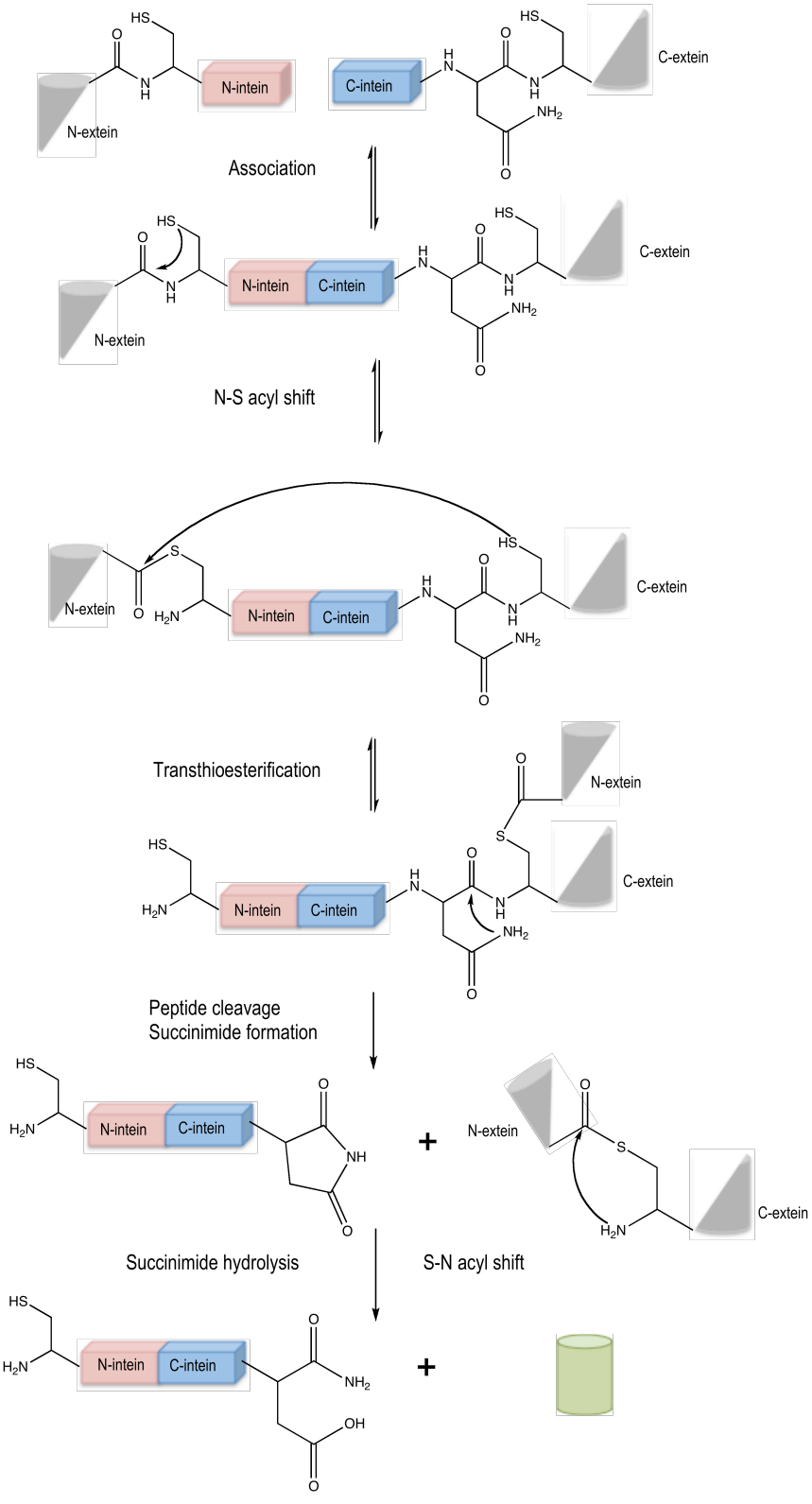

N-extein
N-intein
C-intein
C-extein
Association
N-S acyl shift
Transthioesterification
Peptide cleavage
Succinimide formation
Succinimide hydrolysis
S-N acyl shift

Figure 1.21 Mechanism of *trans* protein splicing. Adapted from David, R *et al.* Expressed Protein Ligation: Methods and Applications, *Eur J Biochem*, **2004**, 271, 663-677. With permission from John Wiley & Sons. © FEBS 2004.

Inteins were also split unnaturally for the purpose of *trans*-splicing, but many required de- and re-naturation to form the functional split inteins (179, 180). The benefit of split inteins is that greater control over splicing can be exerted and a wider variety of applications can be designed, such as detection of protein-protein interactions, protein cyclization, and isotope labeling (181-183). For intact inteins spatial control could be exerted if the construct was targeted to a particular location. However, *trans* splicing does not allow for temporal control, as the splicing would occur once the protein is fully folded. In fact, the intein could splice before reaching its targeted region. In split inteins, each half could be targeted to a location and only when the two parts are present would splicing occur. Additionally, temporal control could be exploited when monitoring protein-protein interactions, as splicing would occur only when the two exteins associate.

A Web of Science database search for “intein” brought up 1,102 results; when cross-referenced with “fluorescent protein” only 68 hits occurred. A further search for “GFP” brought up only 12 publications (as of January 2013). Despite both technologies being around for 20+ years, it appears as though little work has been done to combine the two.

Some of the oldest work done with inteins and GFP was the creation of a cyclized GFP*uv* from an artificially split intein, PI-*Pfu*I, from *Pyrococcus furiosus* (182). The extein (GFP) was cloned between the N-intein and C-intein halves, such that, after splicing there are three proteins created; cyclized GFP, N-intein, and C-intein. A His-tag was also included in the sequence of GFP for purification after cyclization. The newly cyclized GFP was used to study the unfolding and refolding kinetics in comparison with linear GFP. The cyclized protein was slower to unfold meaning that the cyclized protein was more thermodynamically stable than linear GFP.

Another body of work done with EGFP involved creating another artificially split intein VDE (VMA1-derived endonuclease), from the budding yeast *Saccharomyces cerevisiae*, where EGFP was split at position 125 (181). Without external binding partners the split intein fusions did not spontaneously splice

together to reform EGFP. However, when M13 and CaM were attached to the N-intein and C-intein, respectively, green fluorescence was observed in the presence of Ca^{2+}. One problem with the use of inteins with FPs is the potential incompatibility of amino acids required for fluorescence formation vs. those required for proper splicing. When reconstituting a non-permutated FP, the intein-specific mutations should be introduced into the FP to determine if these mutations will alter fluorescence signal. The VDE intein requires the C-terminal Cys residue as well as a hydrophobic residue three to five residues upstream. In the Ile124-Glu-Leu-Lys-Gly-Ile129 region of EGFP at least two mutations had to be introduced in order to promote splicing. After testing several different mutation positions Ile129Cys Glu125Ile were introduced, which did not alter the 488 nm excitation peak.

This problem could be overcome by using a circular permutated fluorescent protein. There are no examples in the literature using inteins in combination with a circular permutated protein for detecting protein-protein interactions. However, the amino acid requirements of the C-intein could replace part of the Gly-Gly-Thr-Gly-Gly-Ser linker such that the folding of the cpFP would not be disturbed. More recently Lee et al. described the validity of using circular permutations to design split points in proteins (184). The authors had previously developed the computer program CPSARST to search for cp-related proteins (136). They used their assessments to create CPred, a program to predict which circular permutation sites in a protein will result in a properly folded protein (185, 186). The authors ran CPred on the intein *Ssp* and potential permutation sites were compared with previously reported split sites. *Npu* inteins were tested for protein splicing efficiency at three predicted split sites and all three sites maintained protein splicing. From this work, circular permutation in not only the extein, but in the intein as well, can be beneficial to scientists looking to use split inteins in combination with circular permutated FPs.

More recent publications of inteins include using GFP as a marker for detecting protein-protein interactions (PPIs); use of inteins to purify GFP; and labeling a protein internally with GFP (187-189). From all of these examples of intein use with fluorescent proteins a common thread emerges; fluorescent

proteins are being used as passive markers or tags to detect the behaviour or localization of other proteins.

Until recently there were no publications reporting the coupling of FP-based biosensors and inteins. In a proof-of-concept paper Wong et al. demonstrated that the FRET-based Ca^{2+} sensor TN-XL and the single-FP Ca^{2+} sensor GCaMP2 could be spliced back together by the *Npu* intein when expressed in mammalian cells (190). Characterization of the sensor response in mammalian cells gave a dynamic range of spliced GCaMP2 that was statistically undistinguishable from original GCaMP2. The response of TN-XL was somewhat lessened, presumably from the increased linker length between FP domains that resulted from the insertion of the three required amino acids for splicing. The two gene constructs were cloned into the pharyngeal muscles of *C. elegans* and Ca^{2+} fluctuations that occurred as a result of serotonin stimulation to induce feeding behaviours were monitored. The authors suggest that one potential application of these inteins could be to place each half under control of distinct promoters to yield cell-type specific data. In addition, the possibility of targeting of interneurons, which are neurons that form connections between other neurons, were mentioned. At locations where two neurons are in close proximity (i.e., synapses) the expressed intein halves could splice together an active Ca^{2+} sensor, allowing Ca^{2+} signal transients to be measured.

In a recent review by Camiré and Topolnick the compartmentalization and regulation of postsynaptic Ca^{2+} signals in inhibitory interneurons was discussed (191). Krosnowski et al. looked for populations of cholinergic interneurons in mouse olfactory bulbs using a transgenic mouse in which expression of EGFP was being driven in neurons expressing choline acetyltransferase (192). Once spliced, the functioning sensor could be imaged not just for fluorescence location, but signal transients could be recorded as well. Once a robust splicing system is in place it needs only to be cloned into the gene(s) of interest to monitor PPIs or specific gene transcriptions.

1.3.6.3 Fluorescence Complementation

The final area of single-FP engineering to be discussed is that of bimolecular fluorescence complementation. This technology relies on using split fluorescent

protein fragments that can reconstitute the intact fluorescent protein and give a fluorescent signal when brought into close proximity. Another similar technology developed in our lab are the dimerization dependent (dd) FPs. DdFPs are heterodimeric FPs that are weakly or non-fluorescent when separated, but reconstitute fluorescence when dimerized (193, 194). Although this technology uses dimeric FPs, it is worth mentioning in the context of complementation technology as a comparison to BiFC.

The field of BiFC did not begin with FPs, but instead with the construction of split ubiquitin and several split enzymes (195-200). By fusing the two inactive protein fragments to other proteins that are known to interact, the interaction of the two proteins could be confirmed by monitoring enzyme activity. The field of BiFC has been widely reviewed (201-203), so only a few key points will be made in this section

Currently there is no algorithm or computer program that can be used to determine if a protein can be split, and if so, at which site. For fluorescent proteins a logical place to start is at sites of permutation as it has been shown that prediction of circular permutation sites using the CPred software can be translated into viable split sites (184). This roundabout method could be applied to the generation of new split FPs. Alternatively, there is a lot of literature describing cpFPs and split FPs (203) and these methods could be applied to novel split FPs owing to the similar 3D structure of FPs (204).

The first example of BiFC was published by Ghosh et al. who used a GFP split between residues 158 and 159 (205). The two halves were attached to antiparallel leucine zippers and green fluorescence was monitored in *E. coli* colonies doubly transformed with the DNA encoding both halves. In addition, denatured purified protein mixtures were also allowed to re-fold in the presence of each other (205). In other work, leucine zippers were used to demonstrate the reconstitution of EYFP split at position 159 (198). An important aspect to consider when using BiFC to monitor PPIs is whether the attachment of the fluorescent protein fragments disrupts the natural association of the two proteins. Additionally, BiFC cannot be used to monitor real-time kinetics as there is a delay between PPI and formation of a fluorescent signal. This is due to the time required for the

two FP halves to fold correctly and for complete chromophore maturation (198, 201). Hu and Kerppola undertook a comprehensive investigation of multicolour BiFC used EYFP, EGFP, ECFP, and EBFP split at position 155 and 173 (206). The bFos and bJun leucine zipper pairs were attached to the N- and C-terminal fragments and the fluorescence signal at various wavelengths was measured when pairs of proteins were expressed in mammalian cells. For some mixed FP pairs a unique excitation/emission profile emerged and could be distinguished from the native FP pairs. A split RFP has also been reported, which would expand the possible FP combinations as well as allow for better spectrally distinct BiFC pairs (207). Some of the more commonly used combination of split FPs have been summarized by Kerppola (203). Some challenges that still remain for BiFC including increasing fluorescence brightness (208-210), high background signal from spontaneous association (211), and stabilization of PPI leading to irreversible complex formation (198).

As an alternative to the traditional BiFC approach, the Campbell lab has recently developed ddFP technology. This technology is based upon the idea that naturally occurring dimeric FPs are non-fluorescent when in their separate monomeric units, which was also seen during the generation of monomeric FPs from their tetrameric precursors (43, 74). If the dimeric FPs could be mutated to generate heterodimeric partners then much like BiFC, PPIs could be monitored by the fluorescence signal created when the specific pair of FPs re-associate (Figure 1.22A). The first example of a heterodimeric fluorogenic FP pair was a ddRFP engineered from dimer Tomato, although one partner retained 10% of its fluorescence brightness (193). Attachment of the M13 peptide to one FP and CaM to the other, allowed the reversible detection of Ca^{2+} in mammalian cells. Additionally, Ca^{2+} oscillations could also be measured, which indicate that the association of the dimers is reversible, unlike BiFC, which is essentially irreversible (198, 203). ddRFP was also used successfully to detect activation of caspase-3 protease by linking the pair with a peptide containing the Asp-Glu-Val-Asp (DEVD) cleavage target (Figure 1.22A). For this application the initial state was fluorescent and loss of fluorescence was indicative of protease cleavage.

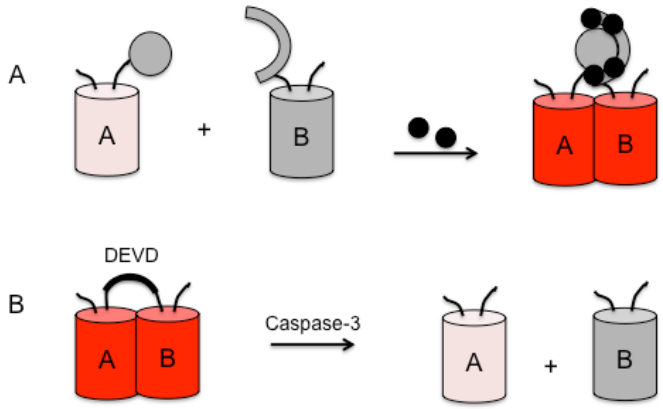


Figure 1.22 Schematic of fluorogenic ddRFP applications. The individual FPs are labeled 'A' and 'B' to differentiate the copies because they are heterodimers. For the ddRFPs one copy retains ~10% fluorescence, hence the light red colour. For ddGFP and ddYFP both copies are essentially non-fluorescent. A) Schematic of small molecule sensing where fluorescence increase indicated presence of analyte. B) Caspase-3/protease schematic where the initial construct is fluorescent. Cleavage by caspase-3 reduces the fluorescent signal.

Shortly thereafter a pair of ddYFPs and ddGFPs were also engineered from the ddRFP by introducing green or yellow inducing mutations (194). Both pairs had better contrast than the ddRFPs due to the negligible background fluorescence of the individual FPs. The K_ds for ddRFP, ddGFP, and ddYFP were 33 μm, 9 μm, and 14.5 μm, respectively. As with BiFC, self-association can lead to high background signals, which was what was observed for ddGFP and ddYFP. When used as caspase-3 sensors; the signal intensity change was much lower than expected based on *in vitro* contrast measurement. This result was attributed to a protein concentration within the cytosol that was comparable to the K_d value (194). At these higher concentrations some of the cleaved ddFPs would still associate and maintain fluorescence. Despite the potential problems related to larger K_d values, ddFPs offer a distinct advantage over BiFC. Because the chromophore is already formed, when the two copies associate the fluorescence signal increases almost instantaneously as compared with some BiFC experiments where fluorescence was seen to increase over 8 hours (211).

1.4 Engineering Fluorescent Proteins

In the previous sections the engineering, characterization, and application of FPs have been described in great detail. Molecular biology protocols were alluded to, but it would be remiss to not describe the multitude of methods and ingenious screening techniques scientists employ in order to generate FPs with widely varying characteristics.

Irrespective of the desired function of an FP, at some point during protein engineering, random or semi-random mutagenesis will be used to look for improved variants. Because of the dynamic chromophore environment, work done with *av*GFP and its mutants relied heavily on random mutagenesis to identify proteins with improved brightness, folding, maturation efficiency, and expression at 37 C° (47, 50, 52, 102, 212). The publication of FP crystal structures provided more insight into the specific amino acid interactions and allowed a more guided approach when planning mutagenesis (2, 24, 65), which has been most effectively exploited during the monomerization of oligomeric FPs (43, 74, 78)

There are several methods to consider when planning protein mutagenesis such as EPPCR, StEP mutagenesis, site-directed mutagenesis, and DNA shuffling (95-99, 213-216). EPPCR is the best first approach when structural information is missing or when characteristics such as folding or chromophore maturation efficiency need to be improved as increased fluorescence brightness can be screened easily in *E. coli*. The main objective of EPPCR is to create conditions that favour the introduction of mismatched base pairs during the extension phase of PCR with error rates around 1-2 mis-matched base pairs per 1000 base pairs (bps) (95, 96, 217). Too high of an error rate increases the likelihood that deleterious mutations will be introduced, such as a stop codon or a frame shift (218). For most FP engineering, improvements are made through the accumulation of small improvements over several rounds.

One benefit to FPs sharing a similar protein structure (204) is the ability to transfer beneficial mutations from one protein to another (49), which has been done for improved folding variants such as EGFP, sfGFP and YFP (103, 104, 219), LSS variants (110, 113), and PA/PS-FPs such as rsTagRFP, PSmOrange,

PAmCherry1, and PA-GFP (115, 116, 121, 220). Even so, for most of these efforts, specific sites of interest were partially randomized. The 20-30% sequence identity between *av*GFP and other RFPs suggests that the optimal amino acids may not be exactly the same for both proteins. Site-directed mutagenesis was first described by Hutchison in 1978 and is a semi-rational or rational approach to protein engineering and requires specifically designed oligos with the desired mutated codon(s) present (214). Semi-rational site-directed mutagenesis is accomplished by designing oligos with a degenerate codon that code for all amino acids or possibly just a limited subset. Alternatively, for a rational approach a codon corresponding to a specific position can be changed to code for a new amino acid. To introduce the mutation(s) OE-PCR can be used, which is the same method used to generate a circular permutation as described in Section 1.3.6.1.

Quikchange mutagenesis is a form of site-directed mutagenesis that allows the introduction of one or more mutagenic sites at the same time (221). Quikchange kits are commercially available and our lab uses Stratagene kits purchased from Agilent Technologies. The ability to introduce multiple mutagenic sites in one reaction and in a shorter time frame makes Quikchange mutagenesis an ideal method to use when constructing libraries composed of random mutations at multiple sites.

DNA shuffling, as first described by Stemmer et al., is advantageous when combining a pool of variants (216). StEP mutagenesis is a different method, but with a similar outcome as DNA shuffling (213). These methods are typically used at intermediate steps in the evolution process after EPPCR or site-directed mutagenesis when multiple variants are selected from the screening process. These methods allow for the recombination of DNA fragments from different clones and increase the chance that multiple beneficial mutations will be combined into one gene.

These methods of mutagenesis are excellent at generating mutated DNA, but the overall effectiveness of a directed evolution strategy hinges on the screening strategy employed. Once the mutated genes have been transcribed and translated into FPs, there needs to be an effective screen for the desired

output (222, 223). Common high-throughput screening methods are very desirable for not only fluorescent protein engineering, but for all evolutionary-based engineering where larger library sizes increase the odds of finding improved offspring (224-226). A few commonly used methods for FP screening are colony-based genetic screening and bacterial culture screening. Several reviews have covered high-throughput screening methods in depth (223, 227, 228), and a thorough discussion of each is beyond the scope of this thesis.

In our lab the predominant method of screening fluorescent proteins is in *E. coli* colonies on agar plates, using a home-built colony screening system, which will be described in Section 2.4.5. This is usually supplemented with *in vitro* testing of bacterial culture expressing the selected protein variants. Screening on plates and in culture is done in an iterative fashion such that one or more candidates from one round of mutagenesis are pooled together to serve as the template for the next round of mutagenesis (Figure 1.23). The characteristic of FPs that is most commonly being optimized is the brightness, which could be due to increased maturation efficiency, improved folding efficiency, or increased quantum yield or extinction coefficient. Using the home-built camera system, images of plates are taken with the appropriate excitation and emission filters allowing bright colonies to be differentiated from the dim ones.

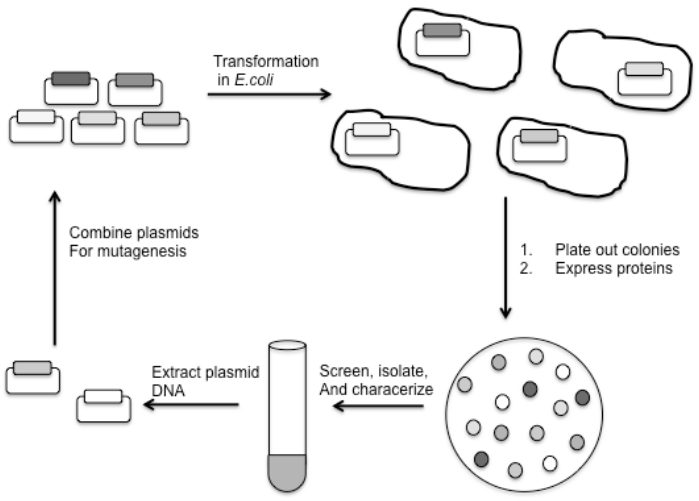


Figure 1.23 Schematic of iterative protein engineering.

It is also possible to screen for more complex signals such as photoactivation or photoswitching by monitoring more than one excitation/emission wavelength.

These multicolour screening techniques have been used to generate the green-to-red photoconvertible mMaple (129) and ddGFP along with ddYFP (194). It is also possible to screen biosensor constructs for response to the analyte of interest. In certain cases, a solution containing the analyte can be sprayed directly onto the plate (14, 154) and the ratio of the fluorescence signal can be calculated. Some problems associated with this technique are attaining uniform analyte distribution, effective diffusion of the analyte into the colonies, and preventing colony enlargement due to spreading, which all could affect the reproducibility of the signal intensity. More recently a group member has described a method of paper-based colony screening that involves blotting bacterial colonies onto filter paper, fixing the colonies, and lysing bacteria with freeze/thaw cycles, which can then be exposed to the analyte of interest via spraying (Yi Shen, unpublished work).

Secondary screening of bacterial cultures usually follows colony screening, especially when generating FP biosensors or FPs with shifted wavelengths as more quantitative data can be obtained compared with bright vs. dim on the plates. In our lab, this secondary screen is performed using a microplate fluorescence reader. Such instruments have been used to screen for other Ca^{2+} biosensors, YFP halide indicators, and are the workhorse of high throughput assays for drug screening in the pharmaceutical industry (14, 93, 229, 230).

An important aspect of FP engineering is the oft quoted First Law of Directed Evolution, “You get what you screen for” (222, 231). Mother Nature has optimized FPs for their intended purpose in sea creatures (232-234), but as scientists we are a bit more demanding and strive to generate bright, photostable, pH stable, monomeric FPs with quantum yields approaching unity, large extinction coefficients, and fast folding/maturation. When FPs are to be used as biosensors there is an additional layer of desired characteristics such as large dynamic range, ideal kinetics and K_d, as well as being highly sensitive and selective towards the analyte or process of interest. From these laundry lists of requirements it is apparent that compromises must be made when engineering an FP to best suit your particular scientific needs. Of course many of these characteristics can and have been screened for (44, 49, 235), but oftentimes mutations beneficial for one characteristic have detrimental effects on another. For example, mPlum is far-red

emitting FP engineered from mRFP1.2 that has emission shifted 59 nm to 649 nm, but a quantum yield that decreased from 0.25 to 0.10 (81). As another example, during the evolution of ddGFP and ddYFP from ddRFP both K_d's increased 2 to 5-fold to low μM values, and attempts to decrease the K_d resulted in a loss of brightness related to the extinction coefficient (194).

After the discovery of avGFP and the subsequent engineering efforts undertaken by scientists, it was postulated that FPs may exist in other sea creatures (236), but it was not clear whether the effort required to discover and engineer these putative proteins would be worth it (51). It is true that monumental amounts of work have gone into engineering FPs from *A. victoria* and numerous other sources, but no one could argue that the developments in this field of research were not worthwhile. In fact the awarding of the 2008 Nobel Prize in Chemistry to three key pioneers in this field further solidifies the impact the FPs have and continue to make in many aspects of the scientific community.

1.5 Research Objectives

The majority of the work in this thesis was inspired by some of the early work done to generate the single-FP based Ca^{2+} sensors GCaMP and Pericam (15, 16) and is a continuation of some preliminary work done by a Master's student in the Campbell lab (23). When this project started there were no reported RFP-based Ca^{2+} sensors, despite the availability of several promising monomeric RFPs (25, 78, 109, 237, 238). Ca^{2+} is the most ubiquitous second-messenger and is tightly regulated to keep cytosolic concentrations in the low nM range (239, 240). As such, the ability to monitor multiple cellular events with different coloured Ca^{2+} indicators or combinations with other small molecule sensors could provide new information regarding the interplay of different cellular processes (153, 241).

Previous work in the Campbell lab identified a cpmCherry variant which retains fluorescence when split at position 193, but with termini too far away from the chromophore to serve as a protein scaffold for the attachment of M13 and CaM (23). Instead this protein served as the template from which new permutation sites could be introduced. Once a cpmCherry with a suitable circular permutation site was generated the Ca^{2+} binding domains were attached and the Ca^{2+} response and spectroscopic behaviours was characterized. In addition,

protein splicing via split inteins was used to reconstitute several different split Ca^{2+} sensors in mammalian cells in order to determine the robustness of split biosensors.

The purpose of the work in this thesis was to generate a single RFP-based Ca^{2+} biosensor for use in imaging experiments and structural studies. More specifically, the objectives of the work were: (1) generate a series of cpmCherry constructs that tolerate new termini in the β10 strand closer to the chromophore; (2) use directed evolution to rescue the brightness and folding of the most promising variants; (3) attach M13 and CaM to a series of cpmCherry proteins and screen for Ca^{2+} response; (4) optimize and characterize behaviour of the protein *in vitro* and in mammalian cells; (5) use site-directed mutagenesis to determine the mechanism of fluorescence response as well as generate cpRFP sensors with altered spectral properties; and (6) generate a series of split RFP and GFP-based Ca^{2+} sensors to attach to the split intein, *Npu,* in order to test mammalian cell splicing and image reconstituted protein fluorescence response to Ca^{2+} dynamics.

Chapter 2 Generation of cpmCherry with new termini located in the β10 strand[2]

2.1 Introduction

The first naturally occurring circular permutated protein was discovered over 30 years ago and since then, circular permutated proteins have been found in several different protein families (132, 133). Thorough study and characterization of circular permutated proteins has determined that most circular permutated proteins retain an overall structure similar to the native proteins (242, 243). Despite the apparent absence of naturally occurring circular permutated fluorescent proteins, it has been shown that *av*GFP and its many related variants will tolerate circular permutation and fold into their native structure while still retaining fluorescence (8, 12, 244).

Fluorescent proteins are 11-stranded β-barrels with the chromophore responsible for fluorescence located near the center of the protein where it is largely protected from the external environment. The isolated nature of the chromophore is advantageous when they are used for protein localization microscope experiments, as the fluorescence signal is stable in the complex cellular environment. However, if we want to exploit the chemistry of the chromophore for biosensing applications, we need to expose it to the surrounding environment without crippling its ability to fold and properly mature. Circular permutation addresses these concerns and several successful FP-based biosensors have been designed on these types of scaffolds (14-16, 22, 38). In general, these biosensors are constructed by introducing new termini close to the chromophore and genetically linking external binding domains that respond to the analyte of interest onto the new N- and C-termini. This is done with the expectation that the external binding event will trigger a conformational change in the vicinity of the chromophore. This conformational change should be detected as a change in the fluorescence emission/excitation of the protein.

[2] A version of this Chapter has been published. Carlson, H.J., Cotton, D.W., Campbell, R.E. 2010 *Protein Science*, **19**: 1490-1499. Section 2.2.1 was carried out by a former undergraduate student, Darrel Cotton.

At the time when the work described in this thesis was initiated, several cpGFP and cpYFP biosensors had been described in the literature. Glaringly absent were the cpRFP counterparts. With this in mind we set out to develop a series of circular permutated red fluorescent proteins. We expected that these new cpRFPs could potentially serve as FP scaffolds for the creation of analogous biosensors. Our group has previously reported the discovery of three variants of mCherry (cp22, cp184, and cp193) that were permutated such that the new termini were located in loop regions at either end of the β-barrel and that retained detectable red fluorescence when expressed in colonies of *E. coli* (23). The method undertaken by Li was to first identify positions within the protein that tolerated insertion of a peptide, and then created the corresponding circular permutation at that position, akin to the first reported cpFP, whereby the accidental insertion of a small peptide within the protein sequence led the authors to generate a cpFP permutated at the point of insertion (8). The greatly diminished brightness of these variants was attributed to a diminished ability to fold and undergo the post-translational modifications necessary for formation of the red fluorescent chromophore. The intrinsic brightness of the chromophore (*i.e.,* the product of quantum yield and extinction coefficient for the fraction of proteins that did form the chromophore) was essentially unperturbed in the circular permutated variants. The directed evolution of the most compliant of these circular permutated variants (cp193), for improved folding and chromophore maturation efficiency was completed by Darrel Cotton, an undergraduate in the Campbell lab. This effort has led to the discovery of a number of new permutation sites, including some that are within the β-sheet region of the β-barrel and in relatively close proximity to the chromophore. These sites represent promising locations for the insertion of binding domains with the aim of constructing single RFP-type biosensors.

2.2. Results and Discussion

2.2.1 Directed Evolution of cp193 for Brighter Fluorescence

A key assumption in our previous efforts to generate circular permutated variants of mCherry was that the linking of the C- to N- termini was a relatively benign modification compared to the introduction of the new termini elsewhere in

the protein (23). Accordingly, we had expected that the amino acid substitutions that would 'rescue' the properties of the cp193 variant would tend to cluster in close proximity to the new termini. This assumption was thought to be valid due to the seemingly generous length of polypeptide sequence joining the last and first residues to be resolved in the x-ray crystal structure of mCherry (245). Indeed, models of the cp193 circular permutation loop with reasonable conformations (*i.e.,* as shown in Figure 2.2A) reveal that the circular permutation loop region extends well away from the β-can portion of the protein. The role of circular permutation linker length has previously been investigated for avGFP (152). The researchers found a minimal linker length of 3 residues was sufficient and longer linker did not provide further substantial advantageous or deleterious effects.

Libraries of randomly mutated genes derived from the previously reported(23) circular permutated mCherry variants cp22, cp26, cp136, cp183, and cp193, were expressed in *E. coli* and colonies were screened for improvements in the brightness of the red fluorescence. Our expectation that beneficial substitutions would occur near to the site of the new termini was, ostensibly, borne out in the first generation of directed evolution in which a substitution (Asn196Asp) at a residue just a few positions removed from the location of the new termini resulted in substantially brighter red fluorescence in colonies of *E. coli* (Figure 2.2C). The overall effect of this substitution is to place a negative charge into a patch on the surface dominated by basic amino acids including Lys198, Arg 216, and Arg220. In the x-ray crystal structure of mCherry(245), the basic guanidinium group of Arg 220 is just 3.3 Å from the amide group of the side chain of Asn196. We speculate that the carboxylate side chain of the Asp is participating in a salt-bridge interaction with the Arg 220 guanidinium group. It is interesting to note that Arg 220 is located within the extended loop that links together the 11th and 1st strands of the protein. This loop is composed of the original C- and N- terminal sequences fused by the Gly-Gly-Thr-Gly-Gly-Ser linker sequence.

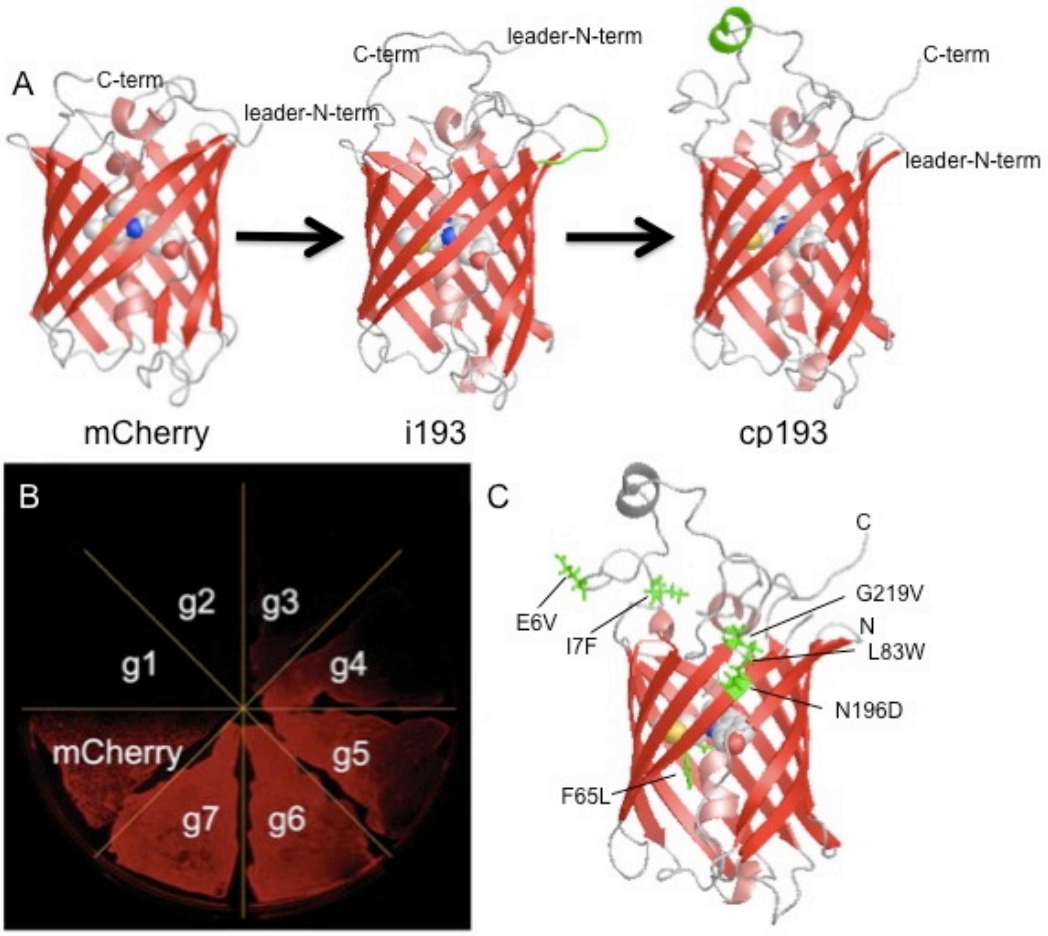


Figure 2.1 Engineering of a bright circular permutated mCherry. (A) Schematic of the steps taken to construct cp193 from i193. (B) Fluorescent image of *E. coli* streaked on LB/agar plate fluorescent image of cp193g1 through cp193g7 in comparison with mCherry. (C) Image of cp193g7 depicting the location of six mutations found after seven rounds of mutagenesis.

Our subsequent directed evolution efforts focused on this lineage, as variants of comparable brightness were not found after one round of screening libraries based on the other 4 permutation sites. An additional 6 generations of directed evolution were undertaken, with the brightest variant identified in each generation serving as the template for the subsequent cycle of library creation by error prone PCR (Figure 2.1B). The directed evolution process was stopped at the seventh generation, as no substantial improvements arose in subsequent efforts. DNA sequencing revealed that the seventh generation variant (cp193g7) had six substitutions relative to the original variant and was equivalent to cp193 +

Glu6Val, Ile7Phe, Phe65Leu, Leu83Trp, Gly219Val, Asn196Asp (Figure 2.2C and Figure 2.3). The cp193g7 variant has 61% of the intrinsic fluorescent brightness of mCherry and 69% of the in vivo brightness following expression for 24 h in colonies of *E. coli* (Table 2.1). The absorbance and fluorescence emission maxima of cp193g7 ($\lambda_{abs,\ max}$ = 580 nm, $\lambda_{em,\ max}$ = 602 nm) are slightly blue-shifted relative to mCherry ($\lambda_{abs,\ max}$ = 587 nm, $\lambda_{em,\ max}$ = 610 nm) (25).

Of the five additional substitutions accumulated during the remaining rounds of the directed evolution process, three of them (Glu6Val, Ile7Phe, and Gly219Val) are also located on the extended loop that connects the 11th and 1st strands of the protein (Figure 2.2C). Notable among these is Gly219Val that is immediately adjacent to Arg220, the residue discussed in the previous paragraph. Yet another of the 5 substitutions, Leu83Trp, lies on the loop between the central helix and the 4th strand of the protein, and is effectively sandwiched between the circular permutation loop and the bulk of the β-can structure. Overall, it is apparent that there is a clustering of the beneficial mutations either on or near the introduced circular permutation linker that ties together the original C- and N-termini. These results suggest that, in contradiction to our original expectations, mutations in the vicinity of the circular permutation loop were of greater importance than mutations in the vicinity of the newly introduced termini for rescuing the fluorescence of cp193. We speculate that these mutations are helping to stabilize the circular permutation loop in a conformation that is more favourable for overall protein folding which in turn leads to a greater proportion of the proteins having a properly formed chromophore.

The last of the 5 substitutions that arose during the directed evolution process is Phe65Leu, which is structurally analogous to the Phe64Leu substitution that dramatically improves the folding efficiency of avGFP (102). Of all the substitutions in cp193g7, Phe65Leu is in the closest proximity to the chromophore and is the most likely candidate for altering the chromophore environment to produce the observed blue-shift in the absorbance and emission maxima.

In the original cp193 variant, three residues (Ala-Tyr-Asn) had been repeated at the N- and C-termini (Figure 2.3). These were originally included in order to

increase the odds of success for identifying brightly fluorescent circular permutated variants. To further explore the role of the overlapping Ala-Tyr-Asn sequence in cp193g7, we created two additional variants that retained the Ala-Tyr-Asn sequence at either the N- or C-terminus, but not both. These variants are designated as cp193g7N (with permutation site at 191/192) and cp193g7C (with permutation site at 194/195), respectively. Characterization of these variants revealed that cp193g7N retained 97% of the intrinsic brightness of cp193g7, but cp193g7C retained only 39% (Table 2.1). This result indicates that it is the N-terminal repeat of the Ala-Tyr-Asn sequence in cp193g7 that is participating in stabilizing interactions and most likely present in a conformation that mimics the same sequence of residues in non-permutated mCherry. In contrast, the C-terminal repeat is likely displaced from the folded structure and not participating in stabilizing interactions.

Table 2.1 Spectral characteristics of cp193mCherry variants

protein	Φ ±0.01	ε ($mM^{-1}cm^{-1}$) rsd. ±4%[a]	Intinsic Brightness ($mM^{-1}cm^{-1}$)	relative brightness in *E. coli*[b] (%)	relative brightness in *E. coli*[c] (%)
mCherry[d]	0.22	72000	16	100	100
cp193	0.20	12000	2.4	4[e]	8
cp193V1.0	0.20	29000	5.8	2	47
cp193g7	0.23	42000	9.7	69	100
cp193g7N	0.23	41000	9.4	81	100
cp193g7C	0.21	18000	3.8	3	22

[a] Average rsd
[b] Measured after 24 hrs at 37 °C
[c] Measured after 72 hrs at 4 °C
[d] From (25)
[e] colony brightness ±5-15%

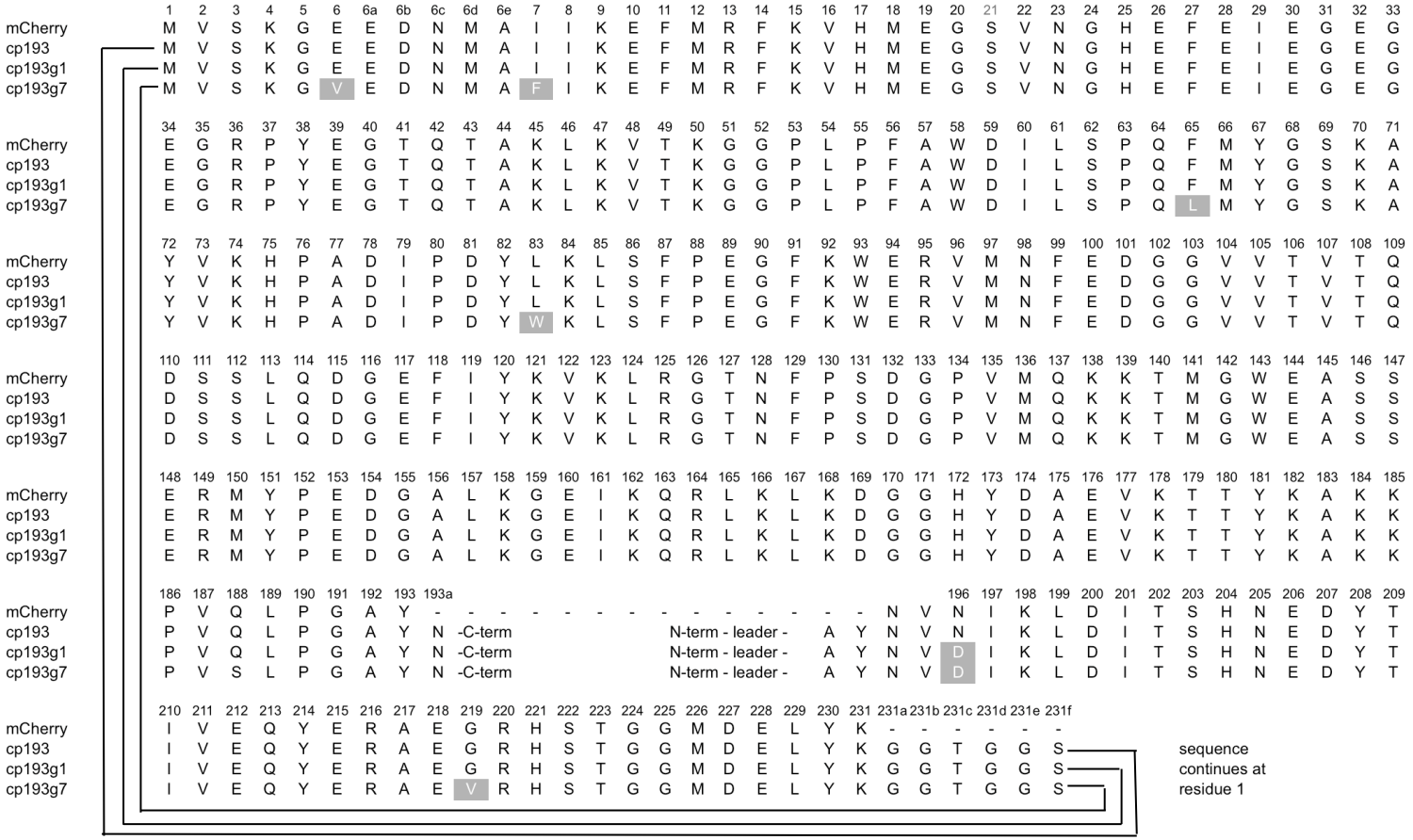


Figure 2.2 Sequence alignment for mutagenesis of cp193 variants. Cp193 is the original circular permutated protein; cp193g1 is the new variant after round 1,cp193V1.0; cp193g7 was the final variant found after the 7th round of mutagenesis. White text on the gray background indicates the mutations found within the amino acid sequence.

2.2.2 Identification of New Circular Permutation Sites in β-strand 10.

It has previously been demonstrated that, when starting with template FPs with destabilizing loop insertions, directed evolution for efficient folding can produce FPs with extreme thermal stability (246). Furthermore, FPs optimized for folding efficiency are more tolerant of circular permutation (104, 247). Similarly, we found that the optimized cp193g7 is generally more tolerant of circular permutation than the progenitor protein, mCherry. We suggest two possible explanations for the greater tolerance of cp193g7 to circular permutation. The first explanation is that, due to the additional rounds of optimization, cp193g7 has improved folding efficiency and overall stability relative to mCherry. The second explanation is that the original introduction of the loop that connected the C- and N-termini in circular permutated mCherry had a destabilizing effect that had prevented the discovery of all but the most favourable variants. In cp193g7, point mutations have compensated for the destabilizing effect of the circular permutation linker and thus it is possible to generate additional permutated variants with new termini in less favoured sites.

We interpret the observation that the beneficial substitutions in cp193g7 tended to occur in close proximity to the circular permutation linker region as evidence in favour the second explanation. If these substitutions were generally beneficial to folding efficiency, as opposed to specifically compensating for the circular permutation linker, no particular bias would be expected in terms of their localization within the protein structure. Furthermore, since mCherry is the result of extensive optimization for folding efficiency (25), it is unlikely that the substitutions identified in this work (which did not require the screening of particularly large libraries), could have produced a protein with an overall improvement in folding efficiency. Indeed, cp193g7 has an ensemble ε that is only 58% of mCherry (Table 2.1), indicating it is less efficient at folding into the exact tertiary structure required for chromophore formation. We have previously demonstrated that the inherent spectral properties of the fraction of FP molecules that form chromophores are essentially identical, and differences in ensemble ε are attributable to differences in chromophore formation efficiency (23).

Starting from the cp193g7 template, we created all possible circular permutations with new termini between each pair of consecutive residues from positions 192 through 207 (15 in total) (Figure 2.4). To minimize ambiguity, no overlapping sequences were included at the termini of the new circular permutated variants. To increase the odds of success, we created a series of libraries in which the two residues immediately flanking the 'break' in the circular permutated variants (*i.e.*, the first and last residues of the protein) were both randomized to all possible amino acid residues. Libraries were expressed in *E. coli* and, for each permutation thousands of colonies were screened for bright red fluorescence. For each permutation site that produced red fluorescent colonies, individual red fluorescent clones were cultured and the plasmid DNA isolated. DNA sequencing of the permutated RFP genes was used to determine the identity of the residues immediately flanking the permutation site. These results are summarized in Figure 2.4B. For each permutation site, the variants that exhibited the brightest fluorescence in colonies were purified and their quantum yields (Φ) and extinction coefficients (ε) determined (Table 2.2). Overall, we identified red fluorescent variants for 10 of the 15 permutation sites (Figure 2.4B). Of the five permutation sites that did not result in any red fluorescent variants, 4 of them were the sites in closest proximity to the chromophore (196/197, 197/198, 198/199 and 199/200). The fifth site was 193/194, which is only two residues removed from the presumed 191/192 permutation location in the optimized template cp193g7. This characterization revealed a general trend that positions closer to the loops at each end of β-strand 10 tended to be more tolerant of introduction of new termini. With respect to both in vitro and in vivo brightness, the 4 brightest variants had new termini either close to the N-terminal end of β-strand 10 (192/193) or close to the C-terminal end of β-strand 10 (203/204, 205/206, and 206/207). Variants with new termini closer to the middle of β-strand 10, and thus closer to the chromophore, tended to be substantially dimmer (*i.e.*, 194/195, 195/196, 200/201, 201/202, and 202/203).

We attribute the high rate at which we successfully identified circular permutated RFPs derived from cp193g7 to the fact that we generated libraries of all possible residues at the two positions immediately flanking the site of the new permutation. As illustrated in Figure 2.4B, in no case did we identify the

combination of both wild-type residues as being among the preferred combinations at the randomized positions. We did observe that wild-type residue V195 was strictly required when 195 was the N-terminus of the protein, but not when 195 was the C-terminus of the protein. Similarly, wild-type residue E206 was strictly required when 206 was the C-terminus of the protein, but not the N-terminus. Yet other circular permutation positions showed a strong requirement for one particular non-wild-type residue type at either the N-terminus (192/193 and 194/195) or both termini (201/202 and 203/204). For some permutation sites (192/193, 195/196, 204/205, and 205/206), 5 or more different combinations of flanking residues were observed, suggesting that these locations do not have stringent requirements regarding the identity of the flanking residues. For other permutation sites, no red fluorescent colonies were observed, indicating that no combination of substitutions at the flanking residues could rescue the protein folding and/or chromophore formation. In the absence of further structural characterization of the variants, we are unable to rationalize these amino acid preferences. However, these results do suggest that other efforts to circular permute proteins could benefit from the creation and screening of libraries in which the first and last residues of the protein are randomized.

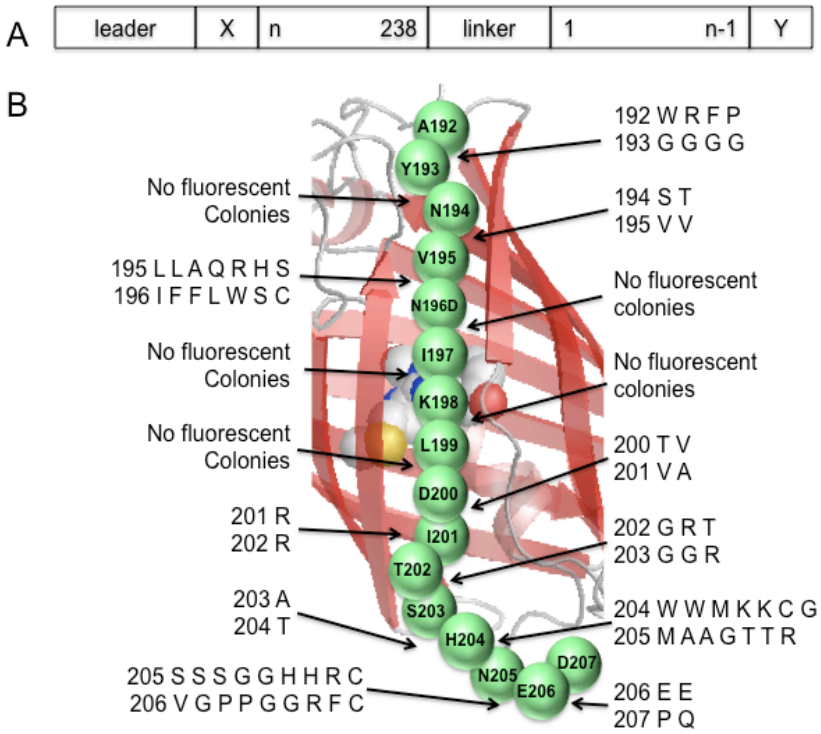


Figure 2.3 Identification of new circular permutation sites along β10 strand. (A) Schematic of the protein structure with new termini introduced between residues n and n-1. 'Leader' is the purification epitope found in pBAD/His B (MGGSHHHHHHGMASMTGGQQMGRDLYDDDDKDPSSR); 'X' is the sequence SSSGALGG; 'linker' is the sequence GGTGGS; and 'Y' is the sequence GGDQLTEE. (B) Overview of the new permutations attempted in this work. The a carbons of residues 192-207 are represented as green spheres. For each consecutive pair of residues shown, a library of circular permutated variants was created in which the chain was split at the peptide bond and both residues were randomized with an NNK codon (1024 members/library). Pairs of residues that produced appreciable red fluorescence when expressed in *E. coli* are shown for each position in vertical alignment. For example, for n= 195 the mutation Val195Val was found at the N-terminus while the C-terminus had the mutations Asn194Ser,Thr.

Table 2.2 Properties of circular permutated variants with new termini between positions 192 and 207

N-terminal residue[a]	C-terminal residue	ϕ ±0.01	ε ($mM^{-1}cm^{-1}$) rsd ±4%[b]	Intrinsic brightness ($mM^{-1}cm^{-1}$)	relative brightness in *E. coli*[c]	relative brightness in *E. coli*[c]
M1	K231	0.22	72000	15.8	100	100
Y193G	A192F	0.23	20000	4.6	2[d]	13
V195V	N194S	0.23	6300	0.14	3	8
D196F	V195A	0.15	4600	0.78	2	2
I201V	D200T	0.22	2600	0.57	2	6
T202R	I201R	0.24	3500	0.86	2	4
S203G	T202G	0.20	1500	0.3	2	4
H204T	S203A	0.21	54000	11	9	100
N205M	H204W	0.22	9300	2	2	9
E206G	N205G	0.21	47000	9.9	39	70
D207Q	E206E	0.20	19000	3.8	3	10

[a] For each position only the brightest variant was characterized
[b] Average rsd
[c] Defined in Table 2.1
[d] ±5-15%

One important application of circular permutated FPs is in the generation of single-FP based biosensors sensitive to not only Ca^{2+}, but other biologically relevant small molecules such as Cl-, ATP/ADP, cGMP, H_2O_2, and NADH(7, 36-38, 40, 93). The key component of a successful sensor is having termini as close to the chromophore as possible so that the chromophore and its environment feel external conformational changes internally. Considering only the circular permutated variants described here, we now have two new mCherry topologies cp196V1.0 and cp201V1.0 (*i.e.,* Asp196Phe/Val195Ala and Ile201Val/Asp200Thr) with spatially distinct termini that are relatively close to the chromophore that may serve as templates for single-FP based Ca^{2+} biosensors. Two drawbacks would be their relatively slow maturation times and their low Φ values that would, like with mCherry itself, lead to poor sensitized emission.

2.3 Conclusion

In conclusion, we engineered a brightly fluorescent circular permutated variant of mCherry with new termini located at the N-terminal end of β-strand 10. With this optimized variant in hand, we created a series of circular permutated variants with new termini along the length of β-strand 10. Of those positions within the strand that tolerated the introduction of the new termini, most resulted in variants that were quite dim relative to mCherry. However, one variant with new termini between residues 203 and 204 did retain 67% of the inherent fluorescent brightness of mCherry itself. Our next goal is to fuse CaM and M13 to the N- and C-termini of each of these variants in an attempt to create red fluorescent Ca^{2+} biosensors analogous to the existing Pericam (15) and GCaMP (16) type biosensors.

2.4 Methods and Materials

2.4.1 General Methods and Materials.

PCR products and products of restriction digestion were routinely purified using agarose gel electrophoresis and the GenCatch kit method (Epoch Biolabs). Plasmids were purified using GeneJet plasmid miniprep kit (Fermentas) according to the manufacturer's protocol. Restriction enzymes and Taq DNA polymerase were purchased from New England Biolabs and used with the supplied buffers. Ligations were performed using T4 DNA ligase (Invitrogen) according to the manufacturer's instructions. FP genes were cloned into either the *XhoI/HindIII* or *Xho1*/*EcoR1* sites of pBAD/His B (Invitrogen) for expression in *E. coli*. Synthetic DNA oligonucleotides (oligos) were purchased from Integrated DNA Technologies. A summary of all of the oligos used in this work with unique numerical identifiers and sequence can be found in Table A1 (Appendix A). DNA was sequenced by dye terminator cycle sequencing using the DYEnamic ET kit (Amersham Biosciences) and oligos 2.1 (forward) and 2.2 (reverse). Sequencing reactions were analyzed at the University of Alberta Molecular Biology Service Unit (MBSU). Protein samples for *in vitro* spectral characterization experiments were first dialyzed into 10 mM Tris-HCl, pH 7.4.

2.4.2 Construction of cp193 Libraries by Error Prone PCR

Error prone PCR was performed using oligos that annealed to the pBAD/His B plasmid outside of the cp193 gene sequence. Error prone PCR reactions were done using TAQ polymerase (New England Biolabs) and with a deficit of each one of the four dNTPs as previously reported (87). DNA was purified using agarose gel electrophoresis, doubly digested using the restriction enzymes XhoI and HindIII (New England Biolabs), and ligated into appropriately digested pBAD/His B.

2.4.3 Construction of cp193g7N and cp193g7C.

The two variants cp193g7N and cp193g7C with the Ala-Tyr-Asn amino acid sequence at the N- or C-termini, respectively, were constructed with cp193g7 as template. Oligos 2.9 and 2.12 used for cp193g7N deleted the Ala-Tyr-Asn sequence at the C terminus, while the cp193g7C oligos (2.10, 2.11) deleted Ala-Tyr-Asn at the N-terminus. Similar to the β10 variants, *XhoI* and *EcoRI* restriction sites were added to the 5' and 3' ends of the DNA constructs, respectively. The two variants were ligated into pBAD/His B after digestion with the appropriate restriction enzymes. The sequence was confirmed with DNA sequencing.

2.4.4 Construction of Libraries of FP Variants with Termini in β-strand 10

Circularly permutated variants of mCherry were constructed with new termini at each peptide bond between positions 192 and 207. The starting template DNA was the cp193g7 construct that resulted from the directed evolution of cp193. Oligos were designed such that the first and last residues of each circular permutated variant were encoded by a degenerate codon (NNK) either preceded by (at the N-terminus), or followed by (at the C-terminus), a Gly-Gly linker. Oligos 2.13 to 2.44 generated cp196 through cp207. Accordingly, a library of 1024 genetically distinct variants was generated for each circular permutation. Furthermore, gene sequence corresponding to the last 5 residues of M13 (Ser-Ser-Gly-Ala-Leu) were appended before the Gly-Gly linker at the N-terminus and the first 6 residues of CaM (Asp-Gln-Leu-Thr-Glu-Glu) was appended immediately after the Gly-Gly sequence at the C-terminus. *XhoI* and *EcoRI* restriction sites were introduced at the 5' and 3' ends of the DNA construct,

respectively. The PCR reactions were carried out in series to limit the length of the oligos. That is, the cp193 variant served as the template for cp194, and cp194 was the template for cp195, and so on. After digestion with *XhoI* and *EcoRI* the gene libraries were ligated into pBAD/His B that had been digested with the same two restriction enzymes.

2.4.5 Plasmid Library Screening.

Ligation products were introduced into *E. coli* DH10B cells via electroporation and then plated out on LB/agar plates supplemented with 0.01% ampicillin and 0.02% arabinose and then left to incubate at 37°C overnight. Colonies were screen for red fluorescence using a previously described colony-imaging system (248). Briefly, plates were illuminated with 535-550 nm excitation light and emission was collected by a monochromatic CCD camera and a 630-660 nm emission filter. If no red colonies were observed, the plates were allowed to sit on the bench top or 4°C fridge for several days. Colonies that exhibited red fluorescence were picked and used to inoculate 5 mL volumes of LB media supplemented with ampicillin (amp). After overnight growth, 100 µL of the culture was pipetted into a microwell plate. The fluorescence emission of each well was measured over 560 nm to 700 nm using at an excitation wavelength of 540 ± 5 nm in a Safire II microplate reader (Tecan). The plasmid DNA for the most brightly fluorescent variants was purified and sequenced as described above. For most permutation positions there were several red variants found. In order to determine the brightest, streak plates were made of each variant and the appearance of red fluorescence was monitored using the same imaging system described above. Only the brightest variant at each position was subsequently purified and characterized.

2.4.6 Protein Expression and Purification

For all protein production, *E. coli* strain DH10B was first transformed with the gene of interest in the pBAD/His B vector by electroporation. A single colony was used to inoculate 5 ml LB/amp cultured overnight (37 °C, 250 rpm) before being diluted into 500 mL of LB/amp supplemented with arabinose (0.02%). Again the culture was shaken overnight before harvesting the cells (37°C, 250 rpm). For the slowly maturing constructs, the temperature was decreased to 28°C once the

culture reached an OD of 0.6, and protein production was induced with 0.02% arabinose and shaken overnight at 250 rpm. The bacteria were pelleted by centrifugation and resuspended in PBS before being lysed by a cell disrupter (Constant Systems). The lysates were then centrifuged at 14,000 g for 45 min, and the proteins were purified from the supernatants using Ni-NTA chromatography (Amersham). Purified proteins were then dialyzed into 10 mM Tris-HCl, pH 7.4. Proteins were stored at 4°C prior to spectral characterization.

2.4.7 Protein Characterization.

Absorbance spectra were recorded on a DU-800 UV-visible spectrophotometer (Beckman). Fluorescence emission and excitation spectra were acquired on a Quantamaster spectrofluorometer (Photon Technology International). Determination of quantum yields (Φ) for the circular permutated mCherry variants utilized mCherry as the reference standard (25). Extinction coefficients (ε) were measured by UV-visible absorbance spectroscopy on purified proteins. The ensemble ε of the proteins was determined by comparing the intensity of the 587 nm absorbance peak for each protein to that of a solution of mCherry with matched absorbance at 280 nm. The ε was calculated by multiplying the ratio of absorbance intensities at 587 nm by the ε of mCherry (72 000 M cm^{-1}) (25).

2.4.8 Molecular Modeling and Figures

Molecular models of mCherry (245) variants with complete termini, insertion, or circular permutation (as shown in Figure 2.1A) were assembled using the Modloop server (249) and the loop modeling application of Rosetta 3.0 (250). MacPyMOL(251) was used for molecular rendering and structure viewing.

Chapter 3 Construction of GCaMP-like Biosensors

3.1 Introduction

In Chapter 2 the construction and characterization of several circular permutated mCherry variants with new termini in the β10 strand was discussed. As mentioned in Section 2.3 the end goal was to generate an RFP-based Ca^{2+} biosensor. At the time when this work was initiated, there had been no published report of such a sensor. Our initial plan was to generate a sensor with termini in the β10 strand rather than the β7 strand, which is the location for the successful GFP-based Ca^{2+} sensors (Figure 3.1) (15, 16).

The discovery that EGFP could tolerate permutation at position 144/145 was reported by two separate groups in 1999 (8, 12). Baird et al. discovered that ECFP could tolerate a 6 amino acid insertion at position 145 and went on to generate fluorescent EGFP, EYFP, and ECFP variants permutated at position 145. In addition, they screened all possible permutation locations within EGFP and found 10 positions that tolerated permutation, one of them being position 145, which validated their initial decision to permute at position 145. Their version of cp145EGFP retained the absorption peak at 488nm, while the 375nm peak was still diminished. They demonstrated that YFP was fluorescent when CaM, a Ca^{2+} binding domain, was inserted at position 145 and also showed a fluorescence response to Ca^{2+}. A Zn^{2+}-binding zinc finger was also inserted at position 145 and the resulting construct had a fluorescence increase of 1.7 times when in the presence of Zn^{2+}. The design of cpFPs that tolerated insertion of the Ca^{2+} binding domain CaM undoubtedly inspired the initial GFP-based Ca^{2+} sensors mentioned in the previous paragraph.

In contrast to the work of Baird et al., Topell and coworkers used GFP*uv* (GFP-Thr99Ser, Met153Thr, Val163Ala) to generate 20 circular permutated proteins at specified positions. They also obtained a green fluorescent cp145GFP*uv* variant with slightly different spectral characteristics from that reported by Baird et al. In contrast to the results of Baird et al., their protein had

enhanced absorption at 375 nm, while the 480 nm peak was almost completely abolished. This difference is most likely due to the different starting proteins.

Given the success of permutation at position 145 in β7 for GFPs, the analogous position of mCherry would seem to be the most promising position for engineering useful circular permutated variants for use in biosensors. However, since this position was not identified in the previous screen of circular permutation positions, we instead focused our attention on the adjacent β10 strand of mCherry (Figure 3.1). Yankun Li, a former master's student in our group determined that the most successful site at tolerating permutation was position 193 (23), which is located in the loop directly preceding the β10 strand. Subsequent work described in Chapter 2 located several more positions in the β10 strand closer to the chromophore. Li and coworkers used a strategy where all possible positions were screened to determine which sites tolerated a 6 amino acid insertion, from which circular permutations were there constructed (23). The closest sites to the β7 strand were 137 and 138, but initial attempts to rescue the fluorescence of the six permutated proteins resulted in cp193V1.0 mCherry showing the greatest improvements (13). Given the low sequence identity and structure of the β7 strand (Figure 1.20) it seemed unlikely that mCherry would be tolerant of permutations in the same locations. As such, future work was done with cp193V1.0 as the template based on the assumption that new termini in the β10 strand would be better tolerated rather than relocating the termini into the adjacent β7 strand.

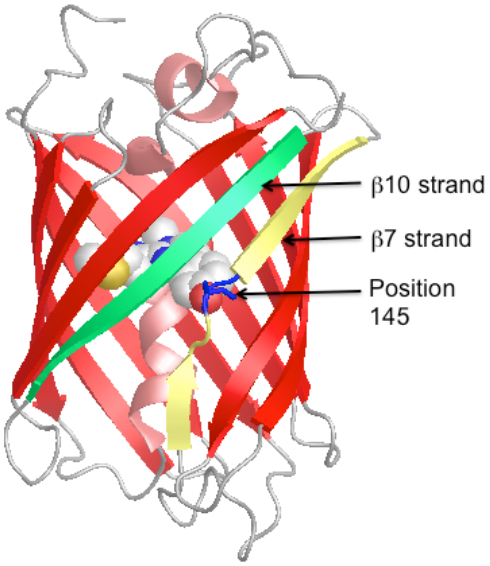


Figure 3.1 Crystal structure of mCherry where the β10 strand is shown in green, β7 strand in yellow, and position 145 highlighted in blue.

We hypothesized that a split point would need to be as close to the chromophore as possible in order for an external conformational change of the Ca^{2+} binding domains to influence the chromophore. My previous work showed the two closest permutation sites were cp196V1.0 and cp201V1.0, which were also the dimmest and poorest folding variants. To increase the chances of discovering a variant sensitive to Ca^{2+} we wanted cpRFPs permutated at sites in the range of 197 to 200, which are even closer to the chromophore. This design meant that the fluorescence and folding efficiency of cp196V1.0 and cp201V1.0 needed be rescued to ensure permutation sites at 197 to 200 would not significantly disrupt the folding or maturation of the protein.

In this Chapter we describe our efforts to improve the fluorescence of our initial circular permutated variants, as well as initial attempts to construct GCaMP-type Ca^{2+} sensors. Towards the end of my mutagenesis efforts another lab member concurrently generated a mApple-based Ca^{2+} sensor (14), which was permutated in the β7 strand with much success. Accordingly, we then chose to also explore some variants with permutation sites in β7. The mutagenesis and optimization of its fluorescence response is also presented.

3.2 Results and Discussion

3.2.1 Directed evolution of cp196V1.0 and cp201V1.0 for increased fluorescence

Much like the work with the initial cpmCherry we hypothesized that the mutagenic strategies we employed would lead to mutations at or near the new split points. Most of the mutations found in the original cp193g7 were near the linker region between the original N- and C-termini. We started with both cp196V1.0 and cp201V1.0 to increase the likelihood of finding a template for constructing GCaMP-type sensors. Libraries of randomly mutated genes of cp196V1.0 and cp201V1.0 were expressed in *E. coli* and colonies were screened for improved brightness of red fluorescence as described in Section 3.4.3. Alternate EPPCR and StEP mutagenesis reactions during evolution was used to accelerate recombination in the hopes that more useful variants would be found in a shorter amount of time (213, 224). After one round of EPPCR a pool of plasmid variants were mixed to serve as the template using this recombination strategy. After the third and fourth rounds 17 cp196V1.0 and 32 cp201V1.0 variants were sequenced and a summative list of all mutation positions found is in Table 3.1. Comparison of mutations indicates there are commonalities between circular permutated variants at the same position, indicated in bold, as well as common sites between cp196V1.0 and cp201V1.0, indicated by italics.

Table 3.1 Mutations found after EPPCR and StEP mutagenesis of cp196V1.0 and cp201V1.0 and their locations in the β-barrel

starting template	Location of residue (inside/outside the barrel)	Mutation Position[a,b]
cp196V1.0	Outside	N6c, ***M6d***, E19, H25, **T43**, T49, D78, T108, S131, D132, ***S147***, *E160*, H172, D174, **Q188**, P190, Y193, ***N194***, Y214, H221, T223, *K231*
	Inside	*M18*, **M150**
cp201V1.0	Outside	K4, E6a, N6c, *M6d*, A6e, R13, E34, T43, G52, **A77**, E117, **R125**, **K139**, A145, ***S147***, Y151, G155, A156, *E160*, K168, E176, K178, ***N194***, K198, T202, E206, I210, D227, L229, Y230, ***K231***
	Inside	*M18*, I29, **L65**, V135, **Q213**

[a]Positions are marked in bold if more than one clone had a mutation at this position.
[b] Positions are italicized if a common mutation position was found between cp196V1.0 and cp201V1.0

For cp196V1.0 variants 24 mutations were identified in 17 clones and a summary of all 17 clones can be found in Table B1 in Appendix B. Only 2 mutations are pointed inside the barrel, and both are located in hydrophobic pockets at either end of the chromophore. We had anticipated that the majority of the mutations we found would be located on the surface of the protein, presumably located near the new termini and the amino acid linker. Based on colony brightness and screening of bacterial protein expression, the brightest variant of these 17 clones contained mutations Met150Leu, Tyr193His, Thr223Ile, and a Gly-to-Val mutation in the first Gly of the linker between M13 and the RFP and was designated cp196V1.1 (Figure 3.2).

Looking only at the cp201V1.0 results there is a total of 36 unique mutations out of 32 sequences (Appendix B, Table B2). Only 5 mutations are pointing inside the barrel, including Met18Leu, Ile29Val, Leu65Phe, Val135Ile, and

Gln213Leu. The Leu65Phe mutation is the analogous position for the Phe64Leu mutation that was found to significantly increase the folding efficiency of GFP, giving it the name EGFP (102). Alternatively, a rearrangement of the chromophore due to the new termini may also better tolerate the added bulk of Phe65. The brightest variant of these 32 clones was determined the same way as for cp196V1.1 and had mutations Leu65Phe, Ser147Thr, Lys178Thr, and a Gly-to-Ser mutation in the first Gly of the linker between RFP and CaM and was designated cp201V1.1 (Figure 3.2).

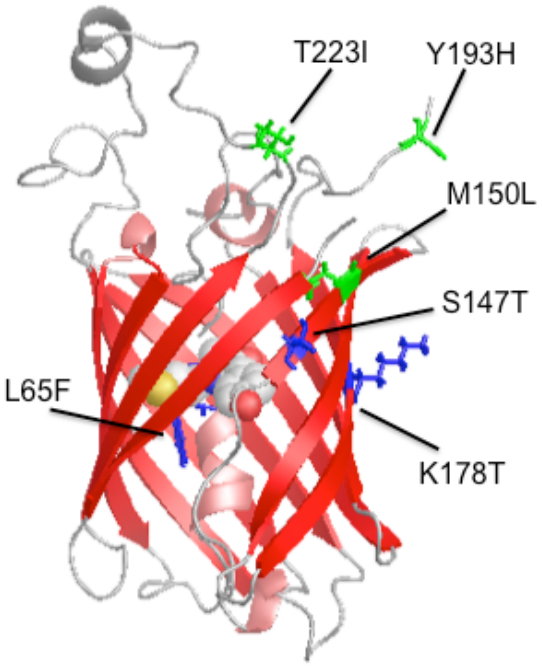


Figure 3.2 Structure of modeled cp193mCherry showing the locations of the mutations found in cp196V1.1 (green), and cp201V1.1 (blue). The cp196V1.1 mutations were Met150Leu, Tyr193His, Thr223Ile. The cp201V1.1 mutations were Leu65Phe, Ser147Thr, and Lys178Thr. Due to the unstructured N-terminus the Gly to Val linker mutation is absent.

As a preliminary test for Ca^{2+} response the M13 peptide and CaM were genetically linked to the N- and C-termini, respectively, onto the cp196V1.1 and cp201V1.1. The addition of the extra domains did not negatively affect the protein folding as the relative brightness of cp196V1.1 and cp201V1.1 remains nearly unchanged (Table 3.2). Since the BPER extracted proteins did not show any fluorescence response upon addition of Ca^{2+} we narrowed the focus to only the cp196V1.1 variant, since it had termini closer to the chromophore. Specifically,

further directed evolution was undertaken to improve its brightness, as described in the following section.

Table 3.2 Relative colony brightness of cp196 and cp201 FPs and GCaMP-type proteins

Variant	Relative Brightness in *E. coli* Day 1	Relative Brightness in *E. coli*[a] Day 3
cp196V1.0	2	2
cp196V1.1	9	84
cp196V1.1-GECO	10	100
cp201V1.0	2	6
cp201V1.1	1	98
cp201V1.1-GECO	3	83

[a] Standard deviation in day three measurement 5-15%

3.2.2 Directed evolution of cp196V1.1 for increased fluorescence

Comparison of the cp196V1.0 and cp196V1.1 sequences revealed four mutations, Met150Leu, Tyr193His Thr223Ile, and a Gly to Val mutation at the appended sequence at the N-terminus (Figure 3.3). At both the N- and C-terminus of cp196V1.0, a Gly-Gly was added linker region to separate from M13 and CaM, which would later be fused to the cpFP gene to create the GECO-type construct.

In the absence of a crystal structure, it is difficult to know why these particular mutations are beneficial in cp196V1.1. However, the crystal structure of mCherry can act as a guide (245). Met150Leu is a mutation that is found in the original progenitor to mCherry, mRFP1.0, and is in a hydrophobic pocket near the chromophore. Mutation to a less bulky Leu could play a role in the rearrangement of the phenolate of the chromophore (Figure 3.2) (43). Tyr193His points inwards, but is still solvent exposed in the mCherry crystal structure (Figure 3.2) (245), especially now that it is near the new C-termini of the protein. The pK_a of a histidine is around pH 6 and given Tyr193His more exposed position it is likely deprotonated at physiological pH and may be better solvated by surrounding water molecules. Thr223Ile is near the original C-termini and mutation to the

hydrophobic Ile could re-orient the linker such that a portion of the loop is pulled more tightly into the top of the β-barrel, rather than be fully extended into solvent, as depicted in Figure 3.2. The first Gly in the linker at the N-termini is mutated to Val. The slightly larger hydrophobic residue could help to fill the gap left by the new permutation site. As summarized in Table 3.3 the quantum yield of cp196V1.1 remains similar to cp196. However, the extinction coefficient increases 10-fold from 4,600 to 46,500 $M^{-1}cm^{-1}$, likely due to improved folding and maturation. Since we measured the ensemble extinction coefficient we are taking the total protein concentration into account, including those copies of the protein that are not fully matured. If the fraction of properly matured fluorescent protein increases the A_{sample} in Equation 3.2 (Section 3.4 6) would also increase leading to a larger calculated ε_{sample}.

Pedelacq and coworkers reported the generation of a superfolder GFP engineered from folding-reporter GFP, in which several circular permutations retain greater than 50% solubility compared with the original superfolder (104). In contrast, analogous permutations of folder reporter-GFP led to variants that retained only 0-15% of the solubility of the unpermutated variant. Comparison of common mutations between cp196V1.1 and cp201V1.1 variants led to six sites chosen for saturation mutagenesis, namely Ser147, Met150, Gln188, Tyr193, Asn194, and His221. These mutation locations were clustered in three main areas. Tyr193, Asn194, and Gln188 are near the new C-termini and mutations found here may help the protein better tolerate the new termini. This prediction follows that of the original cpmCherry engineering results in Chapter 2, whereby mutagenesis of the N- and C-terminal residues allowed for the discovery of circular permutations with non-native amino acids at the split point that better tolerated permutation. Positions Ser147 and Met150 are located on the adjacent β7 strand near the new split point and may help to accommodate the previously mentioned mutation sites as well as the now exposed residues of the β10 strand. Finally, His221 is near the linker between the original N- and C-termini and, as previously discussed during cp193g7 mutagenesis, may facilitate a more stable conformation of the linker.

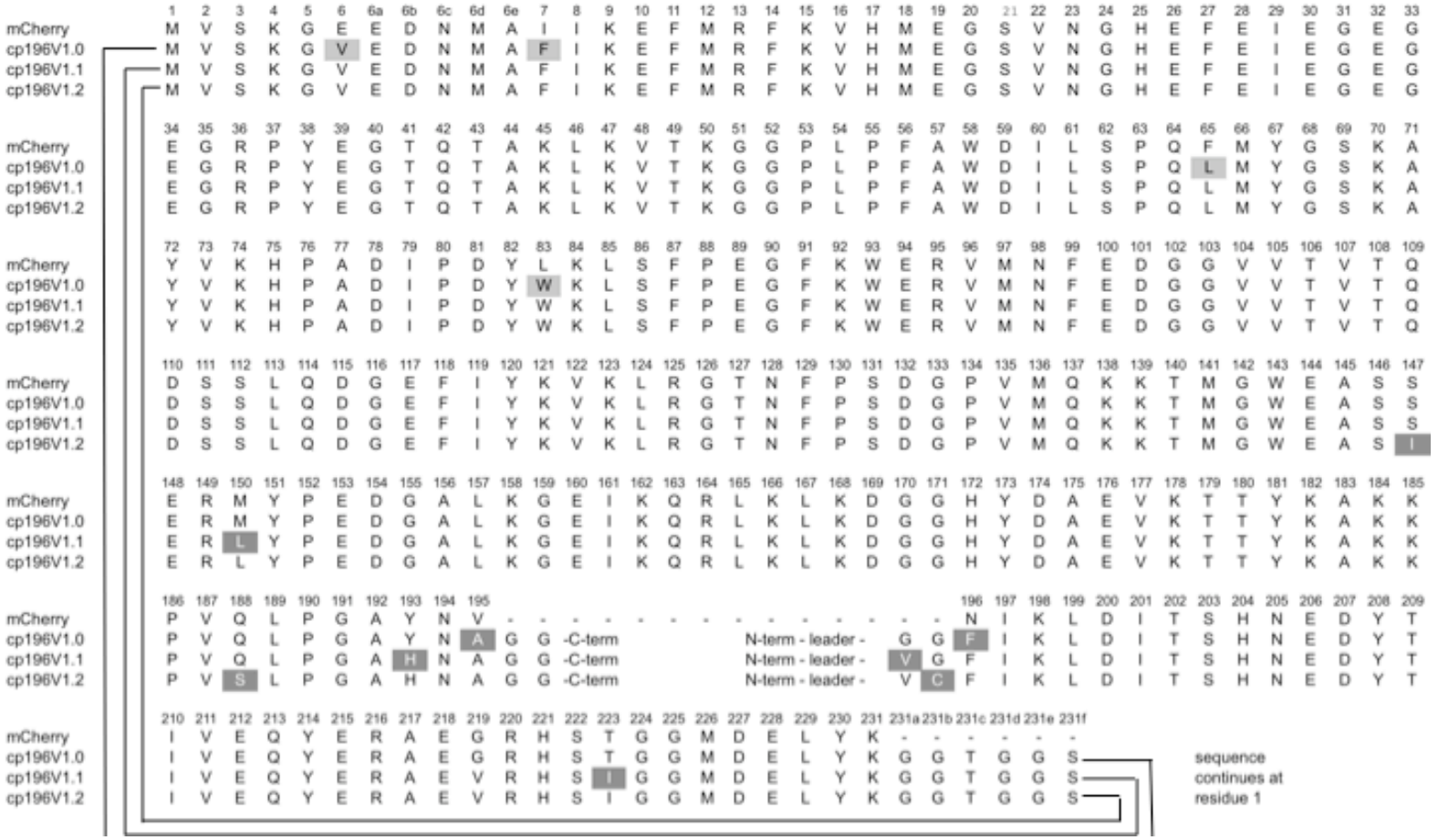


Figure 3.3 Sequence alignment of cp196 variants. Cp196V1.0 is the original cp196 described in Chapter 2; cp196V1.1 was found after the alternating rounds of EPPCR and StEP mutagenesis; cp196V1.2 is a result of the site-directed saturation mutagenesis. Black font on a gray background denotes the mutations that were found in the mutagenesis of cp193g7, described in Chapter 2. White text on a dark gray background are the mutations found in each subsequent round of mutagenesis of cp196V1.0.

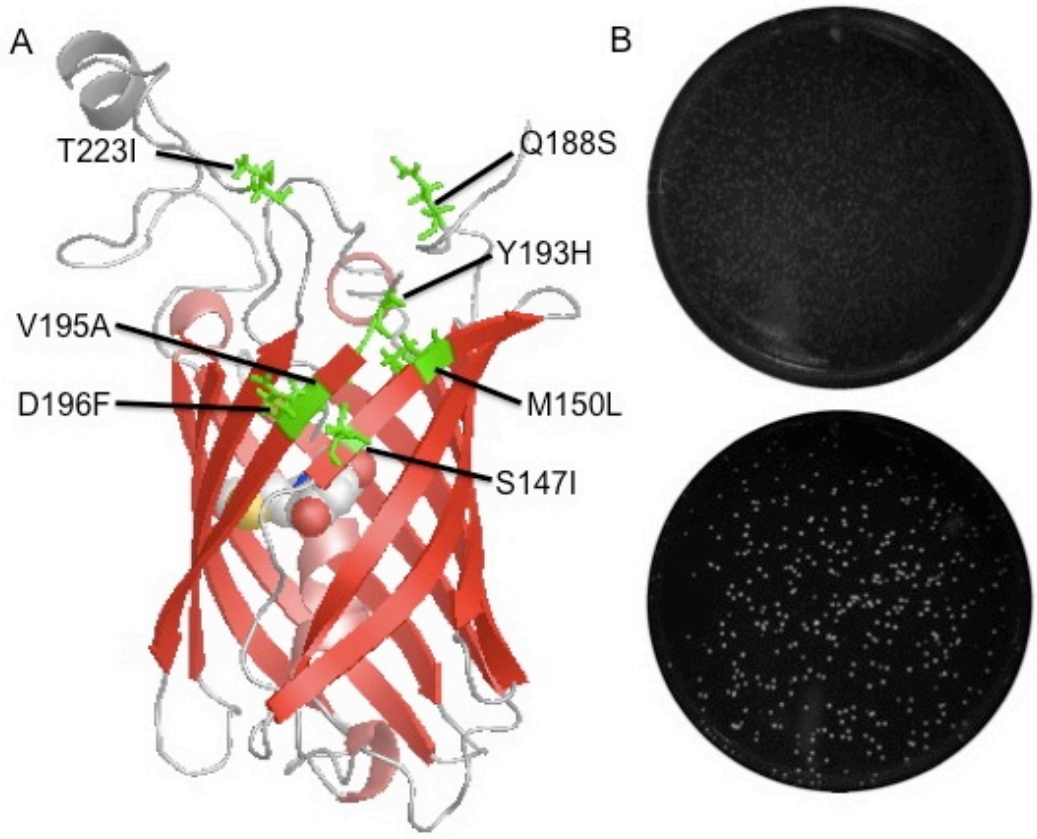


Figure 3.4 Locations of mutations and improved brightness of cp196V1.2. A) Location of seven mutations found after 4 rounds of random mutagenesis and 6 rounds of site-directed mutagenesis resulting in protein cp196V1.2. Three mutations are located near the new split site (Gln188Ser, Tyr193His, Val195Ala, Asp196Phe); one near the linker (Thr223Ile); and two are located near the chromophore in β-strand 7 (Ser147Ile, Met150Leu). The Gly to Cys and Gly to Val mutations are not shown because the N-termini is disordered. B) Fluorescent images of *E. coli* colonies expressing cp196V1.1 (top) and cp196V1.2 (bottom).

After 6 rounds of site-directed saturation mutagenesis followed by fluorescence-based screening for improved brightness, we arrived at cp196V1.2 with three new mutations: Ser147Ile, Gln188Ser, and an additional mutation of the second glycine in the N-terminal linker to cysteine. The location of all mutations, including those from previous generations, is depicted in Figure 3.4A. The possible roles of Met150Leu, Tyr193His, and Thr223Ile have already been discussed, and these mutations remained unchanged from cp196V1.1. Ser147Ile is located in the β7 strand and points outside, but is on the adjacent strand next to the split point. This residue is next to Ser146, which is analogous to the His148

in avGFP. His148 is responsible for hydrogen bonding the phenolate of the chromophore (65). Mutation to a bulkier Ile may help to fill the gap generated by the new split point and could help orient Ser146 towards the chromophore. Gln188 is located in the loop between β9 and β10. The Ser mutation might stabilize the loop in a conformation that promotes more efficient folding of the protein (Figure 3.2, 3.4). The mutation of Gly to Cys is similar to the Gly to Val mutation in that the cysteine is also a bulkier residue that could be oriented towards the hydrophobic core of the protein.

Table 3.3 Spectral Characterization of cp196 variants

Variant	ϕ ±0.01	ε ($mM^{-1}cm^{-1}$) rsd 4%[a]	Intrinsic brightness ($mM^{-1}cm^{-1}$)	relative brightness in *E. coli*[b]	relative brightness in *E. coli*[c]
mCherry	0.22	72000	15.8	100	100
cp196V1.0	0.15	4600	0.78	2[d]	2
cp196V1.1	0.14	46500	6.5	8	81
cp196V1.2	0.21	49500	10.4	40	98

[a] Average rsd
[b] Measured after 24hrs in 37°C incubator
[c] Measure after 72hrs in 4°C fridge
[d] ±5-15%

The extinction coefficient and quantum yield of cp196V1.2 were determined (Table 3.3). Compared with cp196V1.1 the quantum yield of cp196V1.2 increased from 0.14 to 0.21, while the extinction coefficient improved from 46500 to 49000 $M^{-1}cm^{-1}$. The relative colony brightness also improved significantly from 8% to 40% (relative to mCherry). After day three the relative colony brightness was comparable to mCherry. Fluorescent images of cp196V1.1 and cp196V1.2 expressed in *E. coli* are shown in Figure 3.4B and the difference in colony brightness is readily apparent. The intrinsic brightness from cp196V1.1 to cp196V1.2 increased from 6.5 to 10.4, while the colony brightness increased 80% on Day 1 and 17% by Day 3. The vastly improved colony brightness must not be completely due to intrinsic brightness, but also from improved protein folding and maturation. These improvements would be expected, as the intial

cp196V1.0 did not express well in bacterial colonies. An increase in correctly matured protein would be reflected in an increase in ensemble extinction coefficient, while an increase in quantum yield is due to the subtle adjustments in the chromophore planarity and its surrounding environment (245).

3.2.3 Generation of GCaMP type sensors

Using cp196V1.2 as template, a series of potential GCaMP-type Ca^{2+} sensors were generated at various positions within the β-barrel. Positions 197 to 200 were chosen in β10 because they are closest to the chromophore (Figure 3.5). These positions initially did not tolerate permutation as discussed in Chapter 2, but we hoped that the improvements in the starting template would allow insertion of termini closer to the chromophore. Sites 144 to 147 in the β7 strand were also chosen, as these are analogous to the split points in the original GCaMP and GECO series (14-16). The reason for including more termini in the β7 strand was a result of successful cpmApply-GECO constructs a fellow lab member constructed. Since cpmApple-GECO constructs in the β7 strand showed a fluorescence response to Ca^{2+}, we hoped by expanding the pool of variants to include these positions we would more likely find a construct that also had a fluorescence response to Ca^{2+}. The Gly-Gly linkers between M13 and CaM were retained along with the randomized amino acid at the N- and C-termini of cpFP. This strategy was included due to the initial success in generating cpmCherry variants in the β10 strand, described in Chapter 2 (13).

The libraries were screened in bacterial colonies using the TorA export system, initially used to screen for the GECO-series (14). The protein was exported into the periplasm, where the Ca^{2+} concentration (100 μM), is much higher than the cytosolic concentration of 100 nM. Since the Ca^{2+} concentration is much higher than the expected K_d of these indicators (~0.1 to 1 μM), the potential sensors are expected to be in the bright Ca^{2+} bound state rather than the dim Ca^{2+} free state (252).

After imaging the libraries on LB/agar/arabinose plates, bright red fluorescent colonies were picked and grown overnight in 5 mL culture tubes. The BPER extracts were screened using 96-well plates and a microplate reader. The

fluorescence response was measured in the presence of 1 mM EGTA or 1 mM Ca^{2+}. After the preliminary screens new culture tubes were inoculated with clones of interest. Protein was then extracted from the bacterial pellet using the osmotic shock protocol described in Section 3.4.8. Percentage response was calculated by using the equation:

$$\% \text{ Change} = (F_{Ca^{2+}}/F_{EGTA}-1)*100$$

Equation 3**.1** Determination of fluorescence response for protein constructs

The results from the screening have been summarized in Table 3.4. Surprisingly, more fluorescent colonies were found from the positions in the β7 strand rather than the β10 strand. In the β10 strand, only cp197 had many fluorescent colonies, but the remaining positions were less successful. The colony plates were monitored for several days, but no fluorescent colonies above the background signal were detected. We had expected that, since the cpRFP has been optimized at a position in the β10 strand, extension of the termini further into the strand should have been less disruptive. Additionally the introduction of M13 and CaM might have been expected to introduce extra 'bulk' near the split point that might help shield the chromophore from external solvent. The permutation of mCherry has also been described by Shui and coworkers, where they reported permutation sites up to position 198 being tolerated, which appears to match our results (42). Another explanation could be that the short fragment of the N-terminal β10 strand is not long enough to form the secondary structure and instead forms a long extended loop between M13 and the β11 strand. The long unstructured loop could leave an exposed chromophore and hinder the proper maturation of the chromophore. Interestingly red fluorescent colonies were identified at all positions in the β7 strand, which is in closer contact with the chromophore. Other attempts to introduce circular permutation in the β7 strand of mCherry have been unsuccessful (22, 42), but that work started with the original mCherry gene, rather than the extensively engineered circular permutated mCherry. Consistent with our hypothesis, the introduced mutations that improved folding efficiency and optimized the linker orientation have also helped the protein tolerate new termini in an alternate β strand.

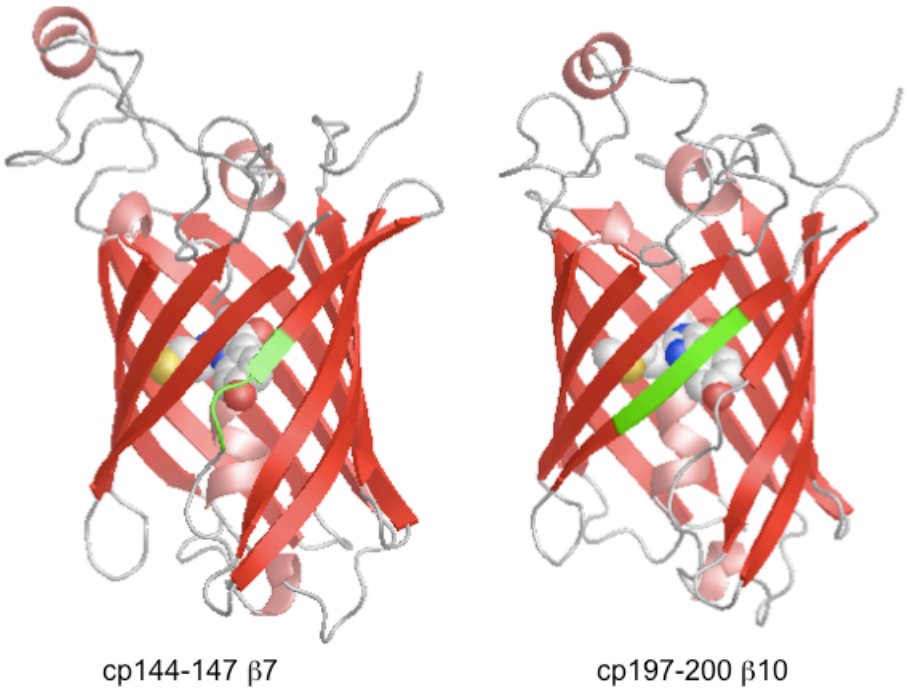


Figure 3.5 Location of the permutation sites in the β7 and β10 strands for generating GCaMP-type constructs.

Table 3.4 shows there were a wide range of responses, including fluorescence decreases upon binding Ca^{2+}, fluorescence increases upon binding Ca^{2+}, and no change in fluorescence upon binding Ca^{2+}. Furthermore, some GCaMP-type constructs exhibited no fluorescence when expressed in bacterial culture. The constructs with the largest fluorescent changes all had circular permutation sites in the β7 strand, specifically at positions 145 and 146. Red colonies with circular permutation at positions 197 and 198 were identified, but none of the selected constructs increased fluorescence in response to Ca^{2+}. At position 197, predominantly hydrophobic residues were found at both termini such as Phe, Val, and Ile. In unpermutated mCherry, position 197 is an Ile directed inside the barrel, which was a mutation originally found during the evolution of mRFP1. The Ile197 side chain, in conjunction with Met163, sandwiches the phenolate moiety of the chromophore and was thought to shift electron density away from the phenolate group and towards the imidazolinone ring (43). In mCherry position 163 is occupied by Gln rather than Met, and the carboxamide nitrogen is likely contributing to the lower pK_a by stabilizing the negative charge on the phenolate. Due to the structurally important role position 197 plays it is possible that even when permutated at position 197 it may still play

the same role. At the C-terminus (position 196), Phe or Ile is preferred in all versions.

The few cp198-GECO constructs that were fluorescent contained mostly small polar side chains at the N- and C-termini such as Gly and Thr. An Arg was found at the N-termini and a Val at the C-termini for variant 198C-GECO. With the cp198 constructs there was fluorescence decrease in the presence of Ca^{2+}. In the absence of further characterization or a crystal structure it is difficult to predict the exact role that these terminal residues play in decreasing the fluorescence.

Table 3.4 Summary of GCaMP-type sensors and the response

Position identifier	N-terminal residue	C-terminal residue	% Change $[Ca^{2+}]$
145_7_9	L	L	72% (+++)
145_3_22	G	L	37% (++)
145_3_33	V	K	35% (++)
145_5_25	W	Y	20%(+)
145_8_2	R	I	29% (++)
145_2_20	F	T	-36% (- -)
145_3_43	W	C	0
145_7_7	P	N	-18% (-)
146_1_39	H	S	-12%(-)
146_1_38	S	K	-18%(-)
146_1_37	G	S	24%(+)
147_2_23	A	S	0
196_4_18	F	A	0
197	V/I	F/I	0
198C	R	T	-20% (-)
198B	G	G	-12% (-)
198E	X[a]	V	-15% (-)
200 II	R	Q	N/A
200 III	E	T	N/A
200 IV	R	X	N/A
200 VI	C	D	N/A
200 VII	W	X	N/A

[a] X indicates that the amino acid at this position was not determined

Several non-fluorescent cp200-GECO variants were picked off ligation plates and their identify confirmed by sequencing. The identity of the first and last residues are listed in Table 3.4. Our expectation is that the identity of the residues found in the first and last positions would be representative of the library diversity introduced by saturation mutagenesis of these positions. Based on this small sample set, there is no discernable pattern, though there could be a bias against hydrophobic amino acids. As several hundred colonies were screened for fluorescence, it is very likely that we did interrogate these sequences even if they are more rare than expected within the library.

It is likely that the same strategy used for the original cp196V1.0 could have been used for cp198-GECO. That is, we could probably have embarked on an effort to further improve the fluorescence until it was a suitable template for making even more aggressive circular permutations such as those at positions 199 and 200. However, since the constructs in the β10 strand did not show a fluorescence increase in the presence of Ca^{2+} those permutation sites were not pursued further. Rather, we chose to refocus our efforts on β7, since these circular permutation variants were giving substantial responses to Ca^{2+} in the context of GECO-type constructs.

Of the constructs with permutation sites in β7, those with new termini at position 145 appeared to be particularly promising for the generation of new Ca^{2+} sensors. Indeed, this is the permutation site used in a variety of other Ca^{2+} sensors including those based on GFP(16), YFP (15), and mApple (14). For the GFP-based and mApple-based sensors, a few C-terminal residues have been deleted whereas my constructs contain the complete sequence along with additional amino acid linkers. In my cp145 variants with fluorescence increase there was a wider range of acceptable amino acids. The results of the screening along the β7 and β10 strand are summarized in Figure 3.6. For cp145 variants, hydrophobic Leu and Val, bulky Trp, and positively charged Lys coupled with Lys, Tyr, and Ile, respectively, produced constructs with a fluorescence increase. For cp146 the small hydrophilic residues Gly and Ser yielded a 24% fluorescence increase, while the His/Ser and Ser/Lys combination resulted in a negative fluorescence response of -12% and -18%, respectively.

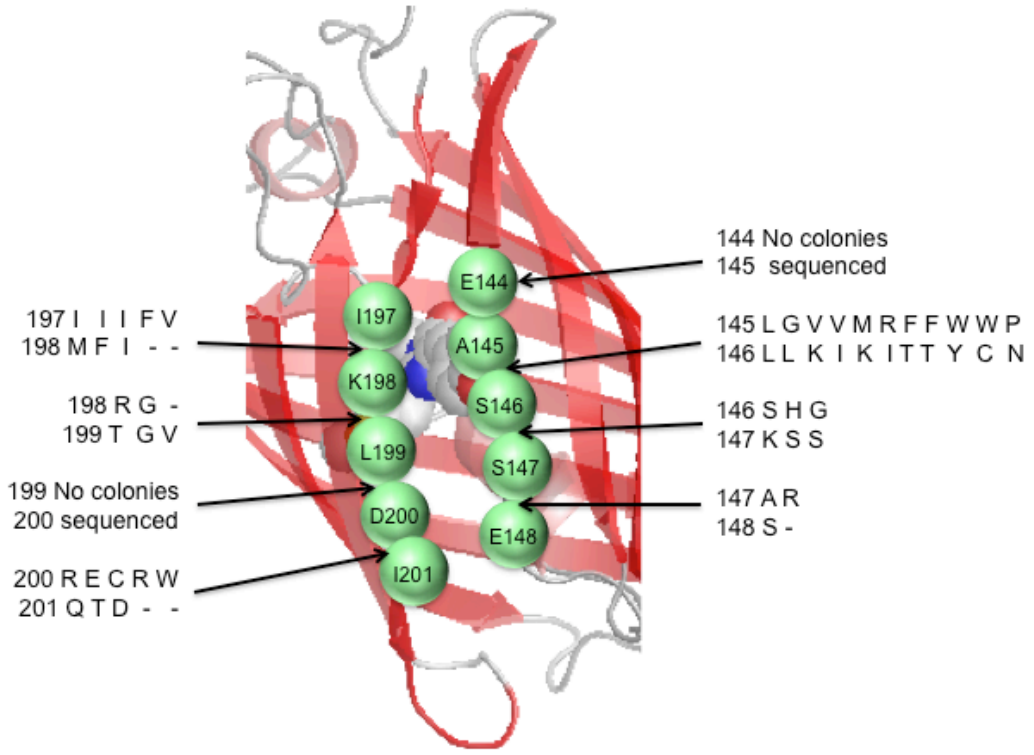


Figure 3.6 Location of the permutation sites for the GCaMP-type Ca^{2+} constructs. Where residues are replaced with a '-', the amino acid was not determined with sequencing.

In each of the cp145-GECO variants with a positive response there is a hydrophobic residue such as Ile, Leu, Trp, or Val at one of the termini. With the exception of 145_7_9 the other amino acid is hydrophilic and possibly solvent exposed, while the extra Leu in 145_7_9 may also be pointed inwards. For the cp145-GECO and cp146-GECO variants that have a negative response, the N- and C-terminal residues have one polar uncharged side chain such as Ser, Thr, Asn, while the other residue is a charged His, Lys, or Tyr. In the case of 145_7_7 there is a Pro and Asn, which might be also be expected to have a positive response, but the kink in the amino acid chain introduced by a Pro might negate the effect of the hydrophobic ring group. For all of the cp145-GECO variants the original Glu/Ala termini combination was not identified, which reinforces the benefit of randomizing terminal residues when generating circular permutations. The cp146-GECO constructs have both positive and negative responses, but they all contain the native Ser residue. This position is structurally analogous to the His148 found in GFP and Glu61 in GCaMP2, which is responsible for

hydrogen bonding to the phenolate and plays an important role in determining protonation state, rotational freedom, and polarization (19, 253). None of the cp147-GECO constructs had any response to Ca^{2+}, probably because the split point was too far away from the phenol of the chromophore. Notably, the one sequenced fluorescent cp147-GECO variant still contained the important Ser146 residue, reinforcing the idea that this residue is important for chromophore maturation.

Based on the results from the β10 and β7 permutation variants, there are multiple factors that play a role in modulating the response including both the proximity of the split point to the chromophore and the identity of the N- and C-terminal residues. Of the variants discovered in this screen, the best response was a 75% increase, which is very small compared to the 1600% signal increase of R-GECO1. Accordingly, further optimization to improve the response to Ca^{2+} will be required to create an mCherry-based Ca^{2+} that offers comparable performance.

3.2.4 Directed Evolution of an mCherry-based Ca^{2+} indicator

In an effort to improve the performance of the most promising Ca^{2+} indicators, variants 145_3_22, 145_8_2, 145_7_9, 146_1_37 were subjected to further rounds of evolution to improve fluorescence response as well as overall brightness. After one round of EPPCR followed by screening for brighter fluorescence of the Ca^{2+}-saturated state, brighter variants of 145_7_9 responsed between 50-100%, and a couple had completely abolished response. None of the other templates gave rise to variants with a larger response than that of the 145_7_9 variants, and in most cases the response had been abolished.

To keep track of new generations of sensors, the GCaMP-type variant 145_7_9 was renamed CH-GECO2.0. The most improved variant identified after one round of EPPCR and screening was CH-GECO2.1 with the mutation Asp23Ala of CaM and a fluorescence response of 100%. Colony brightness was similar to that of CH-GECO2.0 (Table 3.5).

Asp23 is located in the first EF hand of CaM, and is the fifth residue at the designated position Z, in the twelve residue Ca^{2+} binding loop depicted in Figure

3.7 (254). The Asp is one of the seven coordinating groups that bind the Ca^{2+} and mutation to Ala means that this binding position is now vacated, likely filled with a water molecule. The alteration of this EF hand loop composition would likely result in a reorganization of the loop structure, which could be transferred into the first α-helix and the linker between the RFP and CaM.

Table 3.5 Summary of relative colony brightness and fluorescence signal change of cpmCherry-based Ca^{2+} sensors

Variant	Relative Brightness in *E. coli* Day 1	Relative Brightness in *E. coli* Day 3	% Change $[Ca^{2+}]$
cp196V1.2	40	98	N/A
CH-GECO2.0	4	25	72%
CH-GECO2.1	4	14	100%
CH-GECO3.0	14	49	170%
CH-GECO3.1	13	75	250%

The crystal structures of GCaMP and GCaMP2 revealed that the Leu-Glu linker between M13 and cpEGFP plays an important role forming hydrogen bonds with residues in cpEGFP and CaM (18, 19). Comparing the sequence between R-GECO1 and CH-GECO2.1 in Figure 3.8, the original linker Leu-Glu in R-GECO1 is mutated to Pro-Val-Val. This hydrophobic linker stands in contrast to the Glu residue found in the analogous linker of GFP-based Ca^{2+} sensors. Additionally, the crystal structure of R-CaMP, an mRuby based Ca^{2+} sensor, has an Ala-Ile linker between M13 and mRuby. This bulkier hydrophobic linker may occlude the opening of the FP in the Ca^{2+} bound state allowing amino acids such as Glu148 to form a hydrogen bond network with Thr179, Tyr150 and the nitrogen on the imidazolinone ring moiety. With this alternate design in mind Quikchange mutagenesis was used to mutate the Gly-Gly linker to a few hydrophobic or small polar residues such as Thr, Ser, Ala, Val, Pro, Leu, Ile, Phe, and Met. Since the linker in R-GECO1 is shorter than the one in CH-GECO2.1 Quikchange mutagenesis was used to either mutate the Gly-Gly linker between M13 and the cpFP or to delete the Gly-Gly linker using oligos 3.37 and 3.38, respectively. After screening several hundred colonies, one bright colony had a response of ~170% to Ca^{2+}. Upon sequencing I discovered that the construct

was a hybrid between R-GECO1 and CH-GECO2.1 (Figure 3.8), which was named CH-GECO3.0. Many of the mutations occurred in the β10 strand at position 188, 193, 194, 195, 196, and 202, along with 219 and 223. There were also two mutations near the chromophore at position 65, 73, and 132. Given the mix of mutation positions the hybrid occurred due to a recombination similar to DNA shuffling or StEP mutagenesis occurred. The likely cause of this hybrid is that a small amount of R-GECO1 DNA had contaminated the plasmid DNA, which was then transferred to the Quikchange reaction.

There are a total of 16 mutations between CH-GECO2.1 and CH-GECO3.0 including the Gly-Gly linker mutation to Ser-Leu and the Asp132Asn mutation that is not present in R-GECO1. All other mutations that occurred were conversions of CH-GECO2.0 amino acid sequence to R-GECO1 residues. Despite the numerous mutations, the protein still retains the excitation/emission characteristics of CH-GECO2.1 with an increased response of 170% (Table 3.5 and Figure 3.9A). CH-GECO3.0 has 91% sequence identity with R-GECO1, but completely different spectral behavior. There must be only a few amino acids playing a direct role in controlling the chromophore environment. This assertion is supported by previous reports of engineering new colours of *av*GFP and DsRed using a relatively small number of mutations (245, 253, 255).

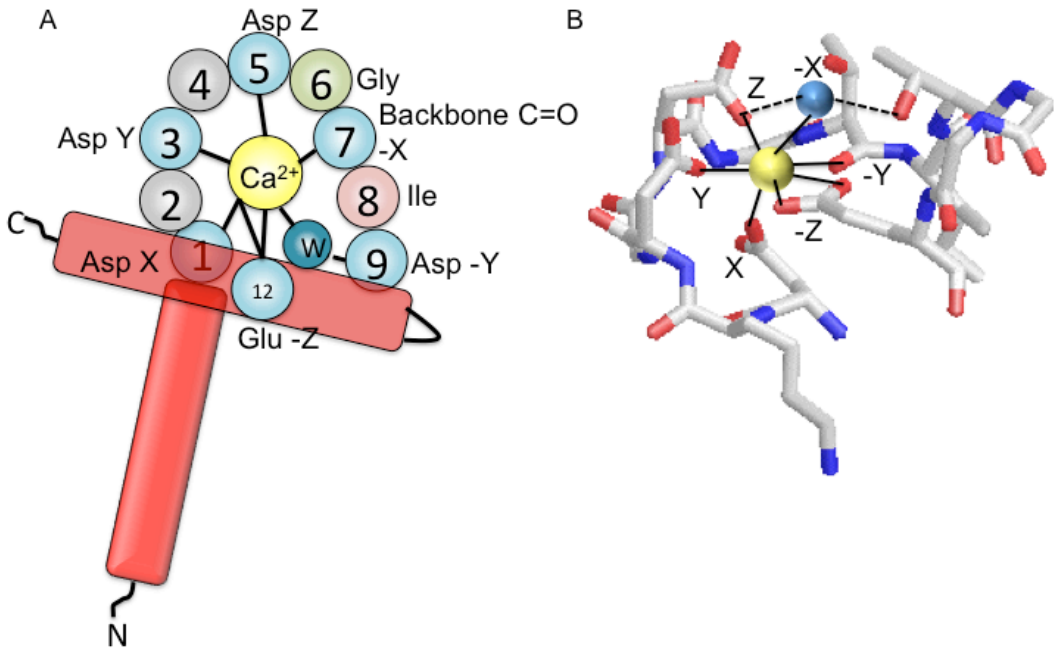


Figure 3.7 Structure of a typical EF-hand depicting the 12aa involved in Ca^{2+} binding. A) Depiction of an EF hand with the N-terminal helix, EF-hand loop, and exiting C-terminal helix. The 12 amino acids involved in the loop are shown as coloured spheres with the appropriate coordinating bonds as thick black lines. Spheres in light blue are the 6 residues that form the coordination sphere around a Ca^{2+} ion including the X, Y, Z letter notation. The green sphere is the conserved Gly residue that is responsible for forming the turn in the loop structure. The pink sphere is the highly conserved hydrophobic residue that forms contacts with the partnered EF-hand. Finally the dark turquoise sphere is a water molecule that forms a contact with Ca^{2+} and residue 9. B) A stick-format structure of the first EF-hand from the crystal structure of CaM (PDB 1EXR). The coordination sphere forms a pentagonal bipyramidal structure around the Ca^{2+} ion in yellow, including a water molecule shown in blue. Adapted with permission from Gifford J.L., Walsh, M.P., Vogel, H.J, 2007, *Biochemical Journal*, **405 (2)**, 199-221. © The Biochemical Society.

Another round of EPPCR and screening yielded a mutant, CH-GECO3.1 (Figure 3.9B), with 250% response to Ca^{2+} and five mutations compared with the previous generation CH-GECO3.0. Three additional mutations are introduced into CaM (Asp21Gly, Phe61Leu, and Thr77Ser) and two mutations are in the FP domain (Ile147Thr and Gly191Asp). Both Asp21Gly and Phe61Leu are in EF hands of CaM, while Thr77Ser is in the loop between α-helix 4 and α-helix 5. Asp21 is position Y in the EF hand and directly helps to chelate a Ca^{2+} ion. Mutation to Gly would render this position unable to bind and the space would likely be filled with a water molecule, similar to the Asp23Ala mutation found in

CH-GECO2.1. The Phe61Leu mutation occurs at position 8 in the EF hand, which is a strictly conserved hydrophobic residue where the carbonyl and amine groups of its peptide backbone face away from the binding site and point towards the loop of the paired EF hand. In some structures of EF-hand pairs position 8 helps to form an anti-parallel β sheet with the groups of paired loops in the eighth position (254). The two mutations could be compensatory, as mutation to the more extended side chain of Leu may better tolerate any structural arrangements that occur in the first EF hand as a result of the additional Asp21Gly mutation. The role of Thr77Ser is less obvious as position 77 is located in a loop between the second and third EF hand. Mutation to Ser may re-organizes this loop to allow Ser77 of CaM to form a better hydrogen bond with His172 of the RFP. Without a crystal structure of the Ca^{2+}-bound protein these suggestions are purely speculative. Ile147Thr is located directly at the split point and undoubtedly mutation to Thr, which is able to form hydrogen bonds, could cause a rearrangement of the electrostatic environment near the opening of the protein. The subtle rearrangement could be responsible for the increased response to Ca^{2+}. Finally Gly191Asp is located in the loop between the β9 and β10 strand, but based on its distance from the chromphore, the role of this mutation is unclear.

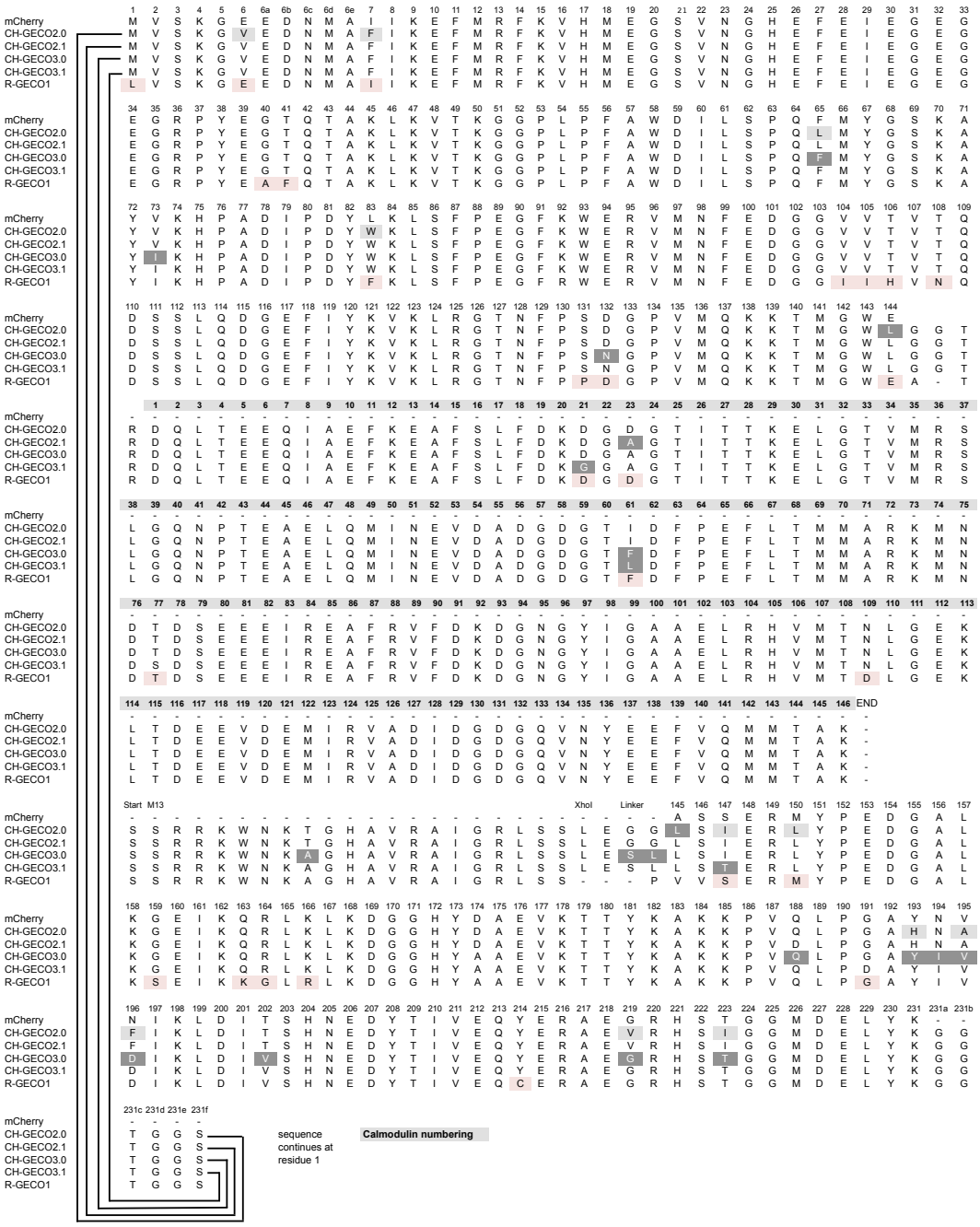


Figure 3.8 Sequence alignment for CH-GECO constructs. CH-GECO2.0 is the original cp145_7_9, CH-GECO2.1 is after one round of EPPCR; CH-GECO3.0 is the R-GECO1-CH-GECO2.1 hybrid found after Quikchange; CH-GECO3.1 is the final construct identified after another round of EPPCR. Black text on a gray background are mutations carried forward from cp196V1.2. White text on a dark gray background are mutations found during EPPCR and Quikchange saturation mutagenesis. Black text on a pink background are the mutations that differentiate R-GECO1 from CH-GECO3.1.

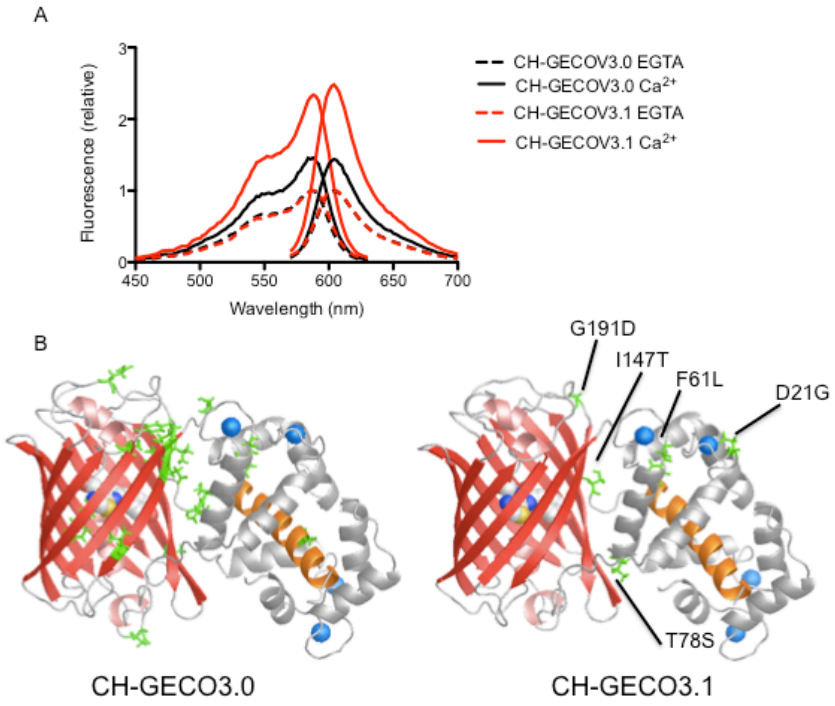


Figure 3.9 Comparison of CH-GECO3.1 and CH-GECO3.0 excitation and emission, along with relevant mutation locations. A) Excitation and emission of CH-GECO3.0 and CH-GECO3.1. Excitation maximum is 582 nm and emission maximum is 602 nm. B) Location of mutations found in CH-GECO3.0 and CH-GECO3.1. Most of the mutations found in CH-GECO3.0 are located at the interface between cp145FP and CaM. For the mutations found in CH-GECO3.1 three are located in CaM (Gly21Asp, Phe61Leu, Thr77Ser) while the other two; Ile147Thr and Gly191Asp are located at or near the interface.

3.3 Conclusion

The results presented in this chapter represent our efforts to achieve several of the primary goals of this project including: rescuing the fluorescence of cpmCherry variants; generating a first generation mCherry-based Ca^{2+} sensor; and improving the response of the first generation mCherry-based Ca^{2+} sensor.

The strategy of generating an RFP-based Ca^{2+} sensor was the same as those used for the GCaMP-type sensors. The circular permutation site of mCherry must be near the chromophore for the conformational change of M13/CaM to be transferred to the chromophore environment. Unfortunately mCherry initially did not tolerate permutation beyond the periphery of the β-barrel

and the variants had to undergo a substantial amount of directed evolution to rescue the fluorescence of the protein (13, 23). Specifically, a combination of EPPCR, saturation mutagenesis, and site-directed mutagenesis, coupled with high throughput image-based screens, were used to improve the fluorescence of cp196V1.0. The variants had improved folding and were more tolerant of permutation at positions closer to the chromophore.

Using optimized cp196V1.2 I generated a variety of GCaMP-like constructs with new termini in the β10 and β7 strand, including some positions analogous to GCaMP (16). Not all of the constructs showed a response to Ca^{2+}, with the only positive responses belonging to proteins permutated at positions 145 and 146. The initial cp145 construct, designated CH-GECO2.0, had a 70% signal increase. After a round of EPPCR a construct was found with one mutation (Asp23Ala) in CaM, and an increased response of 100%. Since the linker between M13 and the RFP plays an important role in the mechanism of R-GECO1, mutagenesis of the Gly-Gly linker in CH-GECO2.1 and subsequent library screening resulted in the identification of an R-GECO1/mCherry hybrid, CH-GECO3.0, where the Gly-Gly linker was mutated to Ser-Leu and the response increased to 170%. A final round of directed evolution led to the identification of a further improved variant, CH-GECO3.1 with several more mutations and a 250% increase in fluorescence upon binding to Ca^{2+}.

Considering the extensive amount of engineering that DsRed underwent to generate mCherry, it should not be surprising that further disruption in the protein structure would require extensive mutagenesis to generate an RFP that tolerated new termini within the barrel structure. However, despite the difficulty of identifying circular permutated mCherry variants that retained red fluorescence, I was ultimately successful in engineering variants with new termini in the β10 and β7 strands. Furthermore, several of these variants retained red fluorescence when fused to the Ca^{2+} binding domain CaM at the C termini and M13 at the N-termini. In the next chapter, I describe a thorough characterization of the CH-GECO3.1's performance *in vitro* and in mammalian cells, and my efforts to obtain a better understanding of the mechanism of action.

3.4 Methods and Materials

3.4.1 General Methods and Materials

PCR products and restriction digest products were purified by gel electrophoresis and the Gel Extraction kit method (Fermentas). Plasmids were prepared using Fermentas Miniprep Kits. Restriction digest enzymes were Fast Digest purchased from Fermentas. TAQ DNA polymerase was purchased from New England Biolabs while Pfu polymerase was from Fermentas. T4 ligase was obtained from Invitrogen and ligations were performed according to the manufacturers protocol. The initial cpFP genes were cloned between *XhoI/EcoRI* sites. When the GCaMP-type constructs were designed cpFP genes were digested and cloned between *XhoI/MluI* sites. The M13/CaM portions were between *XbaI/XhoI* and *MluI/HindIII* sites, respectively. All genes were cloned into a modified pBAD/His B plasmid with a TorA export sequence at the 5' end, described elsewhere (ref).

DNA sequencing was done using the BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosciences). Sequencing reactions were analyzed by the on-campus MBSU. Protein samples used for characterization were buffer exchanged with Amicon MWCO 10,000 centrifuge tubes with 10 mM Tris, pH 7.4.

3.4.2 Mutagenesis of cp196V1.0 and cp201V1.0

Error-prone polymerase chain reaction (EPPCR) was carried out using the established method with $MnCl_2$ and dCTP/dTTP deficiency (96). Oligos 3.1 and 3.2 were used to amplify the EPPCR products. StEP mutagenesis was run using a modified thermocycler protocol (213). EPPCR and StEP were carried out in alternating rounds for a total of 4 rounds. Several bright variants were sequenced to confirm the correct construct and to determine which sites to be chosen for saturation mutagenesis. Oligos for saturation mutagenesis contained the degenerate codon NNK at the desired amino acid position (oligos 3.3-3.12). Oligos with two degenerate sites were used for Tyr193 and Asn194, as well as Ser147 and Met150. Overlap extension PCR was used to generate two DNA fragments, which were recombined using the 5' and 3' PCR oligos 3.1 and 3.2.

The sites were screened sequentially, with a pool of the brightest variants at one position serving as the starting library template for the next saturation site. All PCR products were doubly digested with *XhoI/EcoRI* and ligated into pBAD/His B plasmid. Plasmid DNA was sequenced after EPPCR/STEP mutagenesis and after completed saturation mutagenesis.

3.4.3 Screening for improved variants

The ligation products are expressed in DH10B cells on LB agar plates supplemented with ampicillin and 0.2% arabinose. The colonies were screened with a home-built colony screening system previously described (74). Plates were illuminated with 535-550 nm light and emission collected with a CCD camera and 630-660 nm emission filter. If no red fluorescence developed overnight the plates were left on the bench top for several days and rescreened.

The brightest red colonies were picked and grown overnight in 5 mL LB culture tubes (37 °C, 250 rpm). A GenCatch Miniprep DNA kit was used to isolate the plasmid DNA. The plasmid DNA was pooled and used as the template DNA for subsequent EPPCR/step mutagenesis rounds. After 4 rounds several of the brightest constructs were sequenced using the Big Dye V3.1 kit. Sequencing reactions were processed by Molecular Biology Services Unit at the University of Alberta.

3.4.4 Measuring Colony Brightness

After the third and fourth rounds of mutagenesis the brightest variants were individually expressed on LB/agar/Arabinose plates. Images of the plates were taken with the home-built imaging system and analyzed with a built in macros to measure the colony brightness. Briefly, a fluorescent image of an empty LB agar plate is subtracted from an image of the cpFP and the brightness and size of each spot is exported to an Excel file. mCherry was used as a control and imaged on plates concurrently with the cpRFPs. Percentage brightness was calculated relative to the mCherry colonies.

3.4.5 Protein Expression and Purification

Two improved versions cp196V1.1 and cp196V1.2 picked after EPPCR/StEP and saturation mutagenesis, respectively, were transformed into the *E. coli* strain DH10B. A single colony inoculated 5 mL of LB/amp and incubated overnight with shaking (37°C, 250 rpm). The 5 mL culture tubes were used to inoculate 500 mL of LB/amp and the flasks were shook for 3-4 hours until the OD was 0.6. Arabinose was then added to 0.02% and shook for another 16 hours. The next day the bacteria were pelleted by centrifugation at 10,000 rpm for 10 minutes and resuspended in PBS before being lysed by a cell disrupter (Constant Systems). The lysates were then centrifuged at 14,000 g for 45 min, and the proteins were purified from the supernatants using Ni-NTA chromatography (Amersham). 300 mM imidazole was used to elute the proteins off the columns, which was then removed by buffer exchange with 10 mM PBS, pH 7.4 using Amicon centrifuge tubes (MWCO 10,000). Proteins were stored at 4°C prior to spectral characterization.

3.4.6 Determination of extinction coefficient and quantum yield

mCherry was used as a control to determine the extinction coefficient and quantum yield of cp194V1.1 an cp196V1.2. Protein fluorescence was measured using a Quantamaster spectrophotometer (PTI) and the absorbance was collected on a DU800 UV-Vis spectrometer (Beckman). Solutions of the proteins were diluted to three absorbance values between 0.01 and 0.05 at 550 nm and prepared in triplicate. The proteins were excited at 550 nm ± 5 nm and the fluorescence emission was integrated under the curve. Measurements were made in triplicate Using equation 3.2 from Lackowicz (256) the quantum yield of the protein can be determined:

$$\Phi_{sample} = \Phi_{control}\left(\frac{A_{control}}{A_{sample}}\right)\left(\frac{I_{sample}}{I_{control}}\right)$$

Equation 3.2 Determination of quantum yield where A is absorbance and I is the integrated intensity of fluorescence emission

The extinction coefficients were determined as previously described. Briefly, the protein absorbance was matched at 280 nm such that absorbance at 587nm was between 0.2 and 0.6. Three separate absorbance values were measured in

triplicate. Using equation 3.3 the extinction coefficient of the proteins can be determined:

$$\varepsilon_{sample} = \varepsilon_{control} \frac{A_{sample}}{A_{control}}$$

Equation 3.3 Determination of extinction coefficient where A is absorbance and ε is the extinction coefficient

3.4.7 Construction of preliminary GCaMP-type sensors in the β10 and β7 strand

To construct the GCaMP-type sensors cp196V1.2 was used as the template DNA. In the β10 strand positions 196 through 201 were chosen as well as 144 through 147 in the β7 strand. Oligos 3.13 to 3.24 were used to introduce *XhoI/MluI* sites at either end. For the β7 variants OEPCR was needed to generate the constructs from cp196V1.2. Oligo 3.34 is used with 3.26, 3.28, 3.30, and 3.32 for the C-terminal portion of cp144, cp145, cp146, and cp147, respectively. Oligo 3.35 is used with 3.25, 3.27, 3.29, and 3.31 for the N-terminal portion of cp144, cp145, cp146, and cp147, respectively. To regenerate the intact DNA constructs the external oligos 3.25 through 3.32, are used. Similar to the oligos for generating cp196-cp207mCherry as well as cp196-201GCaMP a degenerate codon NNK was used for the first and last amino acid of mCherry at the N-terminus and C-terminus, respectively. The oligos 3.25 through 3.32 also introduce *XhoI/MluI* sites to the end of the DNA so that they may be ligated into the appropriate plasmid. After successful OEPCR the DNA fragments were excised from the agarose gel and extracted using the GeneJet gel extraction kit. The purified DNA was then digested with *XhoI/MluI*, agarose gel electrophoresis separated, and purified with the GeneJet gel extraction kit.

The G-GECO plasmid containing the TorA peptide was used for ligation of the constructs. The GFP portion of the plasmid was digested out using *XhoI/MluI* Fast Digest restriction enzymes and the DNA was purified in the same manner of as my cpRFPs. The ligation products were plated out onto LB/agar plates with 0.02 % arabinose and incubated overnight at 37°C.

3.4.8 Screening for Ca^{2+} response

The plates containing the cpRFP-GCaMP libraries were imaged using the home-built system and the presence of red fluorescence was monitored. If no red fluorescence was immediately apparent the plates were stored at 4 °C and checked daily. Bright colonies were picked off of plates and used to inoculate 5 mL LB/amp/arabinose culture tubes. The culture tubes were shaken overnight at 37 °C and the following day a modified osmotic shock protocol was used to extract any of the expressed protein (14). The bacteria was pelleted by spinning at 12,000 rpm and 4 °C. The pellet was resuspended with 500 μL of pH 8.0, 30 mM Tris-Cl, 1 mM EDTA, 20% sucrose and shook on ice for 10 minutes. The pellet is spun down at 9000 *g*, 4 °C and then resuspended with 500 μL of 5 mM $MgSO_4$, 1 mM EGTA, 10 mM Tris-Cl, pH 7.4. For pellets where osmotic shock extraction was inefficient the remaining bacterial pellet was incubated with BPER (Pierce) and the resulting supernatant was collected.

100 μL of the protein extracts were aliquoted into three replicate wells in a 96-well microplate (Corning) along with the addition of either 1 mM EGTA or 1 mM $CaCl_2$. The fluorescence emission was measured with 550nm excitation on the Tecan Safire2 microplate reader. The fluorescence emissions with EGTA or Ca^{2+} were compared and the percentage change was determined.

3.4.9 Increasing the Ca^{2+} response

Oligos 3.35 and 3.36 were used for EPPCR on the variant with the largest fluorescence response, the cp145 construct CH-GECO2.0. The PCR products were digested with *XbaI/HindIII*, purified, and ligated into the previously digested plasmid. The ligation products were again electroporated into *E. coli* DH10B cells and the colonies were screened in the same manner as the initial constructs.

Oligos used for Quikchange mutagenesis were designed using the Agilent Technologies Quikchange primer design software. Quikchange lightning mutagenesis (Agilent) was used to randomize the Gly-Gly linker between M13 and the cpRFP as well as delete the Gly-Gly linker and randomize the *XhoI* restriction site. Oligo 3.35 and 3.36 contained two *nyn* codons to restrict the amino acids to Ser, Thr, Ala, Ile, Val, Leu, Met, and Pro. The Quikchange

reactions were run with a modified thermocycler program as follows: 95°C for 90 seconds; 95°C for 20 seconds; 55°C for 30-40 seconds; 65°C for three to five minutes; Go to step 2 35 times; 65°C for 5 minutes; 12°C at the end. The Quikchange reaction mixture was as follows: 1.3 μL 10x buffer; 1 μL of each oligo 100 pmol/μL; 0.5 μL dNTPs; 0.4 μL of Quiksolution; 100 ng of plasmid DNA; ddH2O to adjust volume to 13μL; and 0.5 μL of enzyme. Afterwards 0.5 μL of DpnI was added to the reaction tube and digested at 37°C for one hour. 2 μL of the reaction mixture was electroporated into DH10B *E. coli* cells and spread onto LB/agar/0.02% arabinose plates. After incubating overnight at 37°C the colonies were imaged using our home-built screening system as described above.

Chapter 4 Characterization of mCherry-based Ca^{2+} sensor

4.1 Introduction

Despite having a wide array of protein engineering technologies available to us, the ability to rationally design proteins with desired characteristics remains very challenging. Current computational and modeling software have made many improvements, but we are long ways off being able to plug in an amino acid sequence and receive an accurate protein structure prediction (257-261). As such, x-ray crystallography is still one of the golden standards for interpreting the link between protein structure and function. This standard remains despite the knowledge that proteins do not exist in static crystalline-like states and are instead undergoing macro- and micro-scale conformational changes in response to the ever-changing cellular dynamics. Regardless of this drawback to crystallography, analysis of the protein structure/function relationship can often be understood by using mutagenesis to analyze the new spectroscopic characteristics, and correlate these results with the crystal structure.

It is fortuitous that the somewhat rigid 11-stranded β-barrel structure of the green fluorescent protein and its family members enables the use of the crystal structure to portray a fairly accurate picture of the environment surrounding the chromophore. The chromophore is packed in a well-defined environment and shielded from external solvent. Various researchers have used rational mutagenesis to alter the environment or the chromophore itself leading to alternate colours (25, 43, 48, 63, 79, 106, 114, 204). Another way FP engineering has benefited is from the use of irrational and semi-rational mutagenic strategies. Mutations are introduced into the gene encoding the FP and bacterial colonies expressing the mutants are screened such that those clones with the desired properties are isolated and further characterized. Irrational and semi-rational mutagenesis have been used to improve the folding, expression levels and, in some cases, to monomerize oligomeric fluorescent wild-type proteins (25, 43, 216, 224, 262, 263). By comparing the mutations with the known crystal

structures of close homologues, we can draw conclusions about the role of individual amino acids on the protein's structure and function.

The structural and functional analysis of fluorescent protein-based biosensors is more complicated because there are usually binding domains attached to one or both termini and as a result of an external conformational change the internal environment is altered. The dynamics of the biosensor as a whole can complicate the interpretation of crystal structures and pose problems obtaining crystal structures that accurately portray the protein. In the case of GCaMP2 there was significant dimerization of the protein in Ca^{2+} saturating conditions, including domain-swapped Ca^{2+} binding domains. The M13 of one copy of the protein was bound to the CaM of the other, and vice versa (18, 19). A Ca^{2+} free structure could only be generated when mutations were introduced into each of the four EF hands to abolish the Ca^{2+} binding sites. To construct the monomeric crystal structure, the CaM structure from a mutated variant was substituted into the monomer because the CaM in the monomeric protein was unstructured (19). Due to these difficulties there is a chance that some unintended variations in amino acid positions could come about. Different versions of the GFP-based Ca^{2+} sensor GCaMP have been crystallized and conclusions about the mechanism of actions have been made based on their crystal structures and accompanying mutagenic studies; namely GCaMP (19), GCaMP2 (18), and GCaMP5 (20). More recently a crystal structure of RCaMP, an mRuby-based Ca^{2+} sensor has been deposited in the Protein Database (PDB, http://www.pdb.org), but with no accompanying publication (PDB # 3UOK). Without a publication of the mechanism it is difficult to determine which structural features may play an important role.

In the case of CH-GECO3.1, there is no crystal structure so structural interpretation must rely on several different protein structures such as mCherry, cp193g7mCherry modeled with a linker, an R-GECO1 model, the RCaMP crystal structure, and an as of yet unpublished crystal structure of R-GECO1 (Eric R. Schreiter, personal communication). The lack of a true crystal structure made interpreting the results of my mutagenesis efforts complicated, especially considering that this protein did not behave in a manner typical to other published single FP-based Ca^{2+} sensors. In this Chapter I describe the analysis of the

spectroscopic characteristics along with the determination of the pK_a and Ca^{2+} K_d of the protein. The kinetics of the protein response to Ca^{2+} were also characterized, as all three variables will play a role in determining if the protein can be used to monitor Ca^{2+} in cells. Since there is no crystal structure, a barrage of single-site mutants were generated and the resulting pH and Ca^{2+} binding behaviour was used to rationalize the beneficial mutations accumulated during directed evolution. In addition, several unique mutations were introduced into CH-GECO3.1 to determine if the excitation/emission spectra could be altered while still maintaining function.

4.2 Results and Discussion

4.2.1 Characterization of CH-GECO3.0 and CH-GECO3.1

As mentioned in Chapter 3 there are five mutations between CH-GECO3.0 and CH-GECO3.1 (Asp21Gly (CaM), Phe61Leu (CaM), Thr77Ser (CaM), Ile147Thr, Gly191Asp) and 26 mutations between R-GECO1 and CH-GECO3.1 (Leu1Met, Glu6Val, Ile7Phe, Ala40Gly, Phe41Thr, Phe83Trp, Ile104Val, Ile105Val, His106Thr, Asn108Thr, Pro131Ser, Asp132Asn, Glu144Leu, Ser147Thr, Met150Leu, Ser159Gly, Lys183Gln, Gly164Arg, Arg166Lys, Gly191Asp, Cys214Tyr, Asp21Gly (CaM), Asp23Ala (CaM), Phe61Leu (CaM), Thr77Ser (CaM), Asp109Asn (CaM)). Initial characterization of excitation/emission peaks along with fluorescence response seems to indicate that the CH-GECO sensors are not behaving the same way as R-GECO1. Several different spectroscopic characteristics were studied in this first section including pH behaviour, K_d determination, and the absorbance/excitation/emission spectra.

4.2.1.1 pH behaviour

The pH behaviour of R-GECO1 is the same as the GCaMP Ca^{2+} sensor series (14, 20). That is, the pK_a of the chromophore decreases in the presence of Ca^{2+}, which results in a larger portion of the chromophore in the fluorescent deprotonated state, which is responsible for the fluorescence increase (Figure 4.1A). In the case of GCaMP2, the pK_a shift occurs because of an intricate hydrogen-bond network that is formed between residues of cpEGFP, CaM, and

the M13-cpEGFP linker (18, 19). Both Akerboom et al. and Wang et al. found that Glu61 in the linker between M13 and cpEGFP hydrogen bonds to the protonated chromophore in the Ca^{2+}-free state. In the presence of Ca^{2+}, Glu61 hydrogen bonds with Arg81, which in turn also hydrogen bonds with Glu387 (CaM residue). This stabilizing interaction allows Arg377 in CaM to hydrogen bond to the phenolate of the chromophore, thereby stabilizing the deprotonated fluorescent state. The shift in the H-bond network drops the pK_a of the chromophore and we see a fluorescence signal increase when exposed to Ca^{2+}. Based on the pH titration of R-GECO1 (Figure 4.1A), it is very likely that a similar mechanism is responsible for the increase in fluorescence upon binding Ca^{2+}. The simple sigmoidal shape of this curve stands in stark contrast to the substantially more complex pH curves of CH-GECO2.0, CH-GECO3.0, and CH-GECO3.1 (Figure 4.1B-D).

Comparison of the pH vs. fluorescence curves indicates that the mCherry-based sensors do not operate using a similar mechanism to the previously developed Ca^{2+} sensors. In the absence of Ca^{2+}, the pK_a of the R-GECO1 chromophore is 9.3. When bound with Ca^{2+} the pK_a shifts to 5.8 resulting in an increase in fluorescence signal due to increased proportion of chromophore in the anionic form. However, in all mCherry constructs the pK_a of the chromophore remains relatively constant at around 5 in both the absence and presence of Ca^{2+}. Closer examination of the CH-GECO3.1 pH titration curve reveals three separate pK_a values, the chromophore at ~5.1, Glu215 at ~9.5, and a third unknown residue at ~ pH 6. Glu215 is a strictly conserved residue found in *av*GFP proteins as well as RFPs, which is believed to act as the catalytic base during chromophore maturation (264). In the crystal structure of mCherry Glu215 appears to be protonated and forms a hydrogen bond with the imidazoline nitrogen. Above pH 10 there is a blue-shift in emission corresponding to the deprotonation of Glu215 and subsequent rearrangement of the electron density of the chromophore (245). This blue shift in emission is also observed for my mCherry-based GECO proteins.

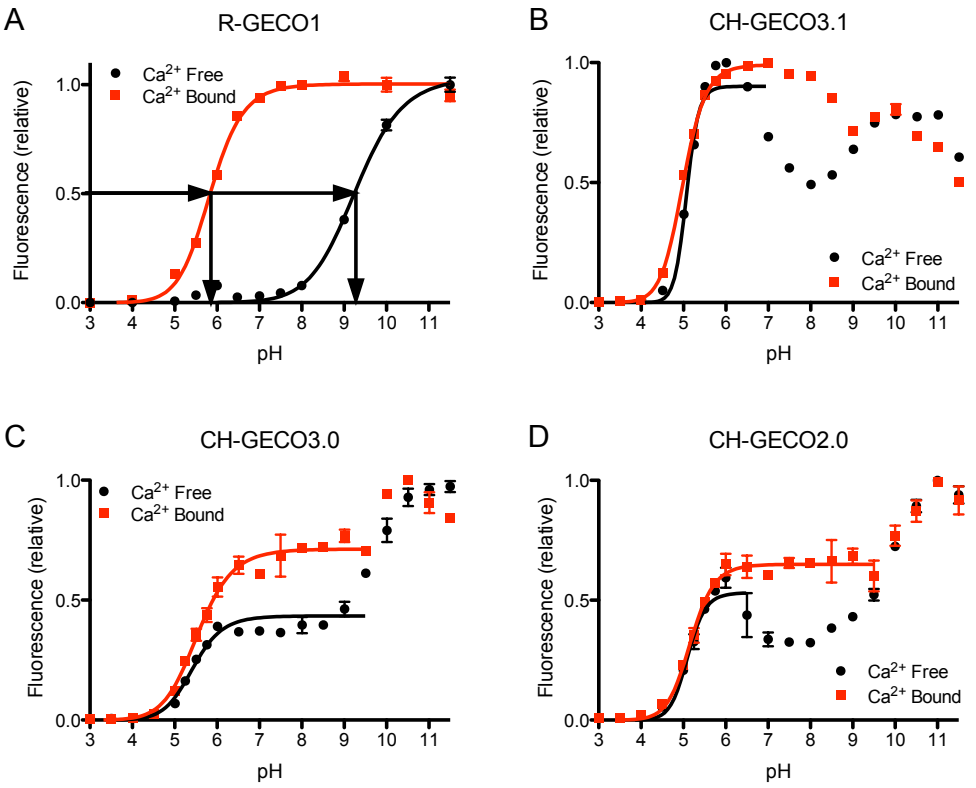


Figure 4.1 pH titration curves for RFP-based Ca^{2+} sensors. In the Ca^{2+}-free state the protein is mixed with 10 mM EGTA, while in the Ca^{2+}-bound state the protein is mixed with 10 mM CaEGTA. The best fit curves for Ca^{2+}-free and Ca^{2+} bound are shown in black and red, respectively. Data points without Ca^{2+} are black circles, with Ca^{2+} red squares. A) R-GECO1 titration curve indicates a pK_a shift upon Ca^{2+} binding. B) CH-GECO3.1 C) CH-GECO3.0 D) CH-GECO2.0

To my knowledge all single-FP-based Ca^{2+} sensors operate under the pK_a modulation mechanism and the apo-state pK_a values of the chromophore range from 7-9.5 (14, 15, 17, 20). mCherry, the template for my Ca^{2+} sensor, has a pK_a of <4.5 so it is improbable that the pK_a of the chromophore could be decreased further in the presence of Ca^{2+}(25). Additionally, Ca^{2+} measurements are routinely made at physiological pH, which would result in a fully deprotonated chromophore even before Ca^{2+} addition.

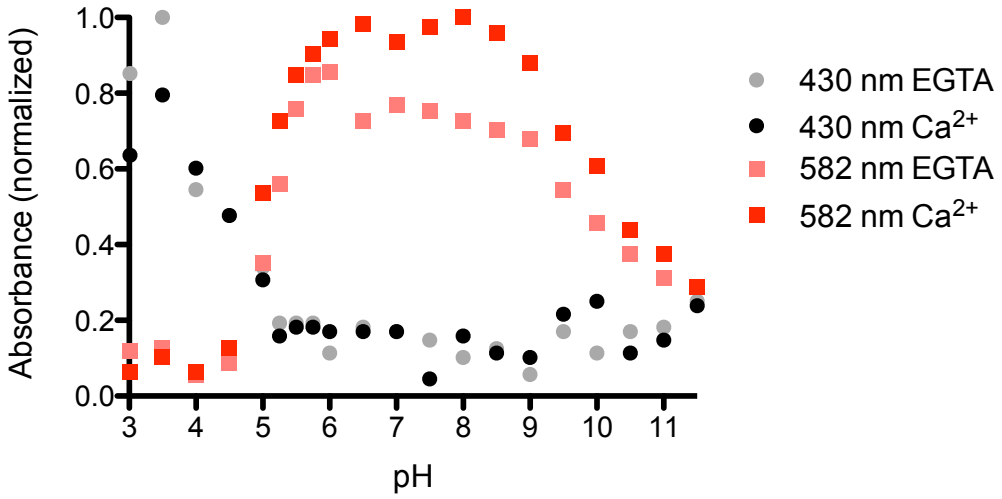


Figure 4.2 Absorbance of CH-GECO3.1 at 430 nm and 582 nm at various pH values in a Ca^{2+}-free buffer (30 mM MOPS, 10 mM EGTA) or Ca^{2+}-saturated buffer (30 mM MOPS, 1 mM CaEGTA). The grey circles indicate 430 nm absorbance with EGTA, while black circles are 430 nm absorbance with Ca^{2+}. Pink squares are 582 nm absorbance with EGTA and red squares are 582 nm absorbance with Ca^{2+}.

4.2.1.2 Quantum yield and extinction coefficients

Since there was no appreciable shift in the pK_a that could explain the increase in fluorescence the next two logical explanations would be changes in either the quantum yield and/or the extinction coefficient. Accordingly, these values were measured for both CH-GECO3.0 and CH-GECO3.1 in the absence and presence of Ca^{2+}. From Table 4.1 it is apparent that both an increase in quantum yield and extinction coefficient are responsible for the increased fluorescence in the presence in Ca^{2+}. The quantum yield for CH-GECO3.0 increases 2.3 times from 0.07 to 0.16, while the extinction coefficient increases 10% from 21700 to 23800 $M^{-1}cm^{-1}$. For CH-GECO3.1 the quantum yield increases 3.5 times from 0.05 to 0.17. There is a 30% increase in extinction coefficient from 37000 to 48000 $M^{-1}cm^{-1}$. Despite the increased quantum yield and extinction coefficient they are both smaller than those of mCherry (0.22, 72000 $M^{-1}cm^{-1}$) (25). If the fluorescence increase is solely due to quantum yield and extinction coefficient then there remains substantial room for improvement as the best sensor of this type could reasonably have a quantum yield and extinction coefficient in the presence of Ca^{2+} that is at least as high as that of mCherry.

Despite the low pK_a it could be possible that Ca^{2+} binding shifts the equilibrium between the protonated and deprotonated state. To address these concerns the absorbance of the protein at 430 nm and 582 nm was measured in the absence and presence of Ca^{2+} at pH values between 3 and 11.5. If we look at the absorbance curve at 430 nm in Figure 4.2 the curves in the presence (black circles) and absence (grey circles) of Ca^{2+} are nearly identical with the decrease in absorbance occurring near the pK_a of the chromophore, 5.1. This decrease in absorbance is coupled with an increased absorbance of the 582 nm peak, indicating deprotonation of the chromophore with (red squares) or without (pink squares) Ca^{2+}. However, from pH 6 through 8 there is an increase in 582 nm absorbance in the presence of Ca^{2+}, because the 430 nm absorbance stays constant, this absorbance increase must be due to the increased extinction coefficient of the deprotonated chromophore rather than a shift in the equilibrium between protonated and deprotonated chromophore.

The protein environment surrounding the chromophore helps prevent fluorescence quenching by molecular oxygen and introduces steric bulk that prevents the chromophore from undergoing radiationless decay in the excited state (65, 265). However, the mCherry chromophore is less planar than its DsRed ancestor, which could play a role in mCherry's lower quantum yield of 0.22 (245). Taken together, one might expect that the fluorescence of a protein circular permutated near the chromophore would suffer to some degree from all three effects; fluorescence quenching from solvent, an increased rate of radiationless decay, and a non-planar chromophore conformation. For CH-GECO3.1 in the absence of Ca^{2+}, CaM and M13 are not associated and the low quantum yield is likely due to fluorescence quenching and radiationless relaxation. The crystal structures of different Ca^{2+} sensors in the Ca^{2+}-free state confirm this assumption (19). Without a crystal structure of CH-GECO3.1 in the Ca^{2+} free state or more examination of fluorescence dynamics it is not possible to determine if the chromophore is less planar than mCherry or if it might be able to rotate more freely in the excited state. In the presence of Ca^{2+}, CaM binds four molecules of Ca^{2+} and wraps around the M13 peptide as shown in the modeled structure of R-GECO1 (14), as well as crystal structures for GCaMP2 (PDB ID 3EK4), GCaMP3 (PDB ID 3SG3,G4,G5), GCaMP5 (20), and RCaMP (PDB ID

3U0K). This conformational change effectively closes up the hole at the permutation site which could do two things; block solvent access to limit fluorescence quenching and crowd the chromophore environment restricting the movement of the chromophore to reduce radiationless relaxation.

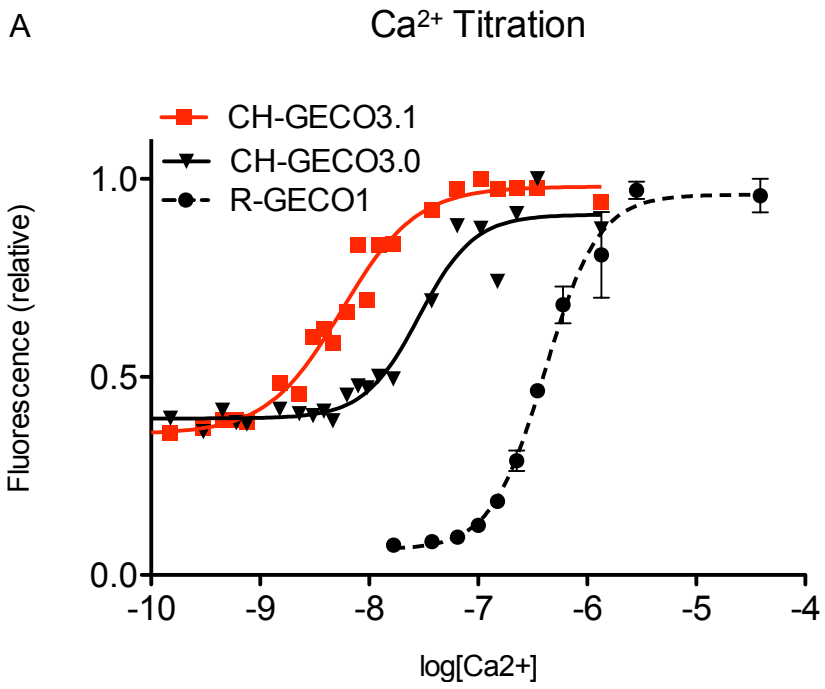
Table 4.1 Spectroscopic characterization of CH-GECO variants

Protein	Solution	ϕ	ε ($mM^{-1}cm^{-1}$)	Brightness ($mM^{-1}cm^{-1}$)	pK_a
CH-GECO3.0	EGTA	0.07	21700	1.5	5.32±0.03
	Ca^{2+}	0.16	23800	3.8	5.45± 0.04
CH-GECO3.1	EGTA	0.05	37000	1.8	5.05 ± 0.02
	Ca^{2+}	0.17	48000	8.2	4.97± 0.01

4.2.1.3 K_d Determination

Comparison of Ca^{2+} titration results of CH-GECO3.0, CH-GECO3.1 and R-GECO1 revealed that the mCherry-based sensors have 15 to 75-fold smaller K_d values than R-GECO1 (Figure 4.3). CH-GECO3.0 has an unexpectedly low K_d of 28 nM, while CH-GECO3.1 has an even lower K_d of 6 nM. The Hill coefficients of R-GECO1 and CH-GECO3.0 are similar at 1.8 and 1.7, while CH-GECO3.1 has a smaller Hill coefficient of 1.3. There are three mutations in CaM between CH-GECO3.0 and CH-GECO3.1 and two additional mutations in CaM compared with R-GECO1 as discussed in Section 3.3.3. The effect of each mutation is explored in the next section to determine the influence on the K_d. We hypothesized that the mutations found in CaM could account for the differences in K_d values between proteins.

A



B

Construct	K_d (nM)	Hill Coefficient	Dynamic Range
R-GECO1	408	1.8	20
CH-GECO3.0	28	1.7	1.5
CH-GECO3.1	6	1.3	2.5

Figure 4.3 Ca^{2+} titration data for CH-GECO3.1, CH-GECO3.0, and R-GECO1. A) Ca^{2+} titration curves for R-GECO1 (black circles), CH-GECO3.0 (black triangles), and CH-GECO3.1 (red squares). B) K_d and Hill coefficient data extracted from the best-fit lines along with the calculated dynamic range.

4.2.2 Altering the K_d

One major problem that genetically encoded Ca^{2+} sensors face is the interference due to binding of endogenous CaM and CaM-binding peptides. Imaging of the sensors in neurons and animals can be problematic when the concentration of endogenous CaM is very high (266-268). Work has been done

to genetically engineer unnatural CaM/M13 pairs with better specificity to limit the potential interference with endogenous CaM (269). The CaM/M13 pair used in CH-GECO3.1 originated from GCaMP3, but now contains seven mutations in CaM and one in M13. With this many alterations in sequence it is possible that the M13/CaM pair in CH-GECO3.1 have increased specificity to each other, though without *in vitro* titration of exogenous CaM the extent of endogenous CaM interference cannot be determined.

As mentioned in the previous section there are 3 mutations in CaM of CH-GECO3.1 compared to CH-GECO3.0, while an additional two mutations separate CaM in CH-GECO3.1 and R-GECO1. To determine the affect of each individual mutation as well as selected combinations, mutations in CaM of CH-GECO3.1 were changed to match the CH-GECO3.0 construct. If the CaM mutations play a role in determining the K_d then we would expect the new K_d to match that of CH-GECO3.0.

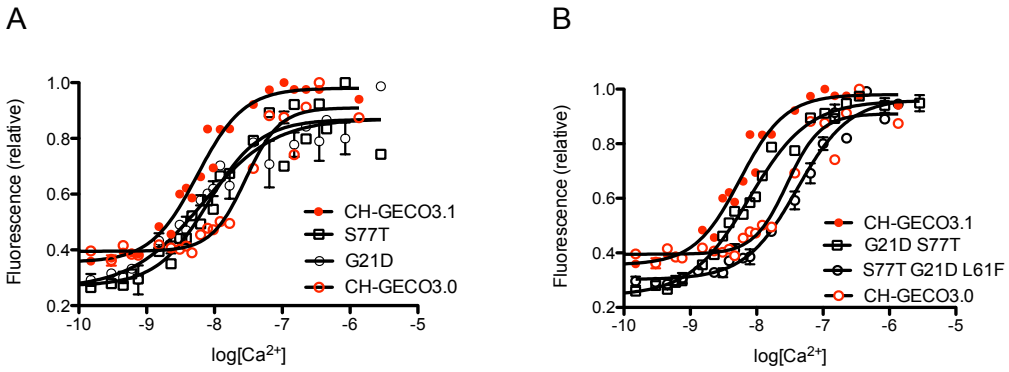


Figure 4.4 Ca^{2+} titration curves for variants with mutations that revert CH-GECO3.1 CaM to CH-GECO3.0 CaM. A) Ser77Thr and Gly21Asp individual mutations compared to CH-GECO3.0 and CH-GECO3.1. B) The combination of Ser77Thr and Gly21Asp and Leu61Phe compared to CH-GECO3.0 and CH-GECO3.1.

In Chapter 3 the potential structural impact of the mutations found in CaM were discussed so the results of the mutagenesis will be discussed in relation to the structural changes. The first mutations introduced were those to revert the CaM sequence of CH-GECO3.1 back to CH-GECO3.0. Ser77 is located between α-helix 4 and α-helix 5 away from EF hands. We might expect that mutation back

to Thr would not affect the Ca^{2+} binding, which is what we see from the Ca^{2+} titration results summarized in Figure 4.4A and Table 4.2. The K_d remains relatively constant at 8 nM and the fluorescence response is similar at 220%. Gly21 as mentioned previously is in the first EF hand in the 'X' position of the Ca^{2+} binding loop. Surprisingly, mutation back to Asp, which should restore a Ca^{2+} chelating site, does not alter the K_d or the fluorescence response. The Hill coefficient decreases to 0.7 indicating negative cooperativity. The response remains the same, but perhaps the single Asp mutation is not sufficient to restore Ca^{2+} binding in that hand as there is still the Asp23Gly mutation. Mutation of Leu61Phe, which is located near the second EF hand, appears to disrupt the proper folding of the CaM domain. This in turn disrupts the proper maturation of the RFP since I was unable to purify a significant amount of red protein. In other mutagenic studies of CaM it has been shown that mutations in the second EF hand are more detrimental to the Ca^{2+} affinity than mutations in the first EF hand (270). The combination of Gly21Asp and Ser77Thr results in a K_d of 13 nM with a response of 280% while the addition of the Leu61Phe mutation yields a K_d of 44 nM, but a similar response of 225%. Interestingly converting these three amino acids back to those in CH-GECO3.0 gave a sensor with a higher response and larger K_d. This difference must mean that there is another mutation found within the RFP that is interacting with CaM and influencing the K_d and response. Mutation of Gly191Asp kept the K_d low at 8 nM while the response remained constant at 210%.

Next, mutations were introduced to match the amino acid sequence of R-GECO1 CaM since the K_d differs by two orders of magnitude. The single mutation Ala23Asp shifts the Hill coefficient to 1.0 and the response increases to 300%, but the K_d remains unchanged (Figure 4.5A, Table 4.2). This amino acid position is also another key Ca^{2+} chelator at position '-Z' in the EF hand loop. In the absence of Asp21 at position 'Z' it may prevent the loop from forming the appropriate orientation to allow Ca^{2+} binding to take place. Asn109Asp is located at the end of α6 away from the EF hands, but the mutation increases the K_d to 13 nM and the response to 313%. Both Asp and Asn are approximately the same size, but Asn is uncharged, though still able to hydrogen bond. In the RCaMP crystal structure (PDB ID 3U0K) Asp109 forms a hydrogen bond network with a

water molecule and the carbonyl backbone of Ser37. When mutated to Asn, the amino group of the side chain could directly hydrogen bond to the backbone, which could confer more stability to both EF hand 2 and 3.

Table 4.2 Summary of the mutations in CH-GECO3.1 CaM

Mutation	Response (%)	K_d (nM) ±1nM	Hill Coefficient
CH-GECO3.1 Ser77Thr	218	8	1.1 ±0.1
CH-GECO3.1 Gly21Asp	260	7	0.7 ±0.1
CH-GECO3.1 Leu61Phe	N/A	N/A	N/A
CH-GECO3.1 Ser77Thr Gly21Asp	280	13	1.18 ±0.05
CH-GECO3.1 Ser77Thr Gly21Asp Leu61Phe	230	44	1.1 ±0.1
CH-GECO3.1 Asn109Asp	313	13	1.5 ±0.2
CH-GECO3.1 Ala23Asp	307	7	1.0±0.1
CH-GECO3.1 Asp191Gly	211	8	1.0 ±0.2
CH-GECO3.1 Ser77Thr Gly21Asp Ala23Asp	270	5	1.7 ±0.1
CH-GECO3.1 Ser77Thr Gly21Asp Leu61Phe Ala23Asp	302	23	1.3 ±0.1
CH-GECO3.1 Ser77Thr Gly21Asp Leu61Phe Ala23Asp Asp191Gly	356	21	1.2 ±0.1
CH-GECO3.1 Ser77Thr Gly21Asp Leu61Phe Ala23Asp Asp191Gly And109Asp	372	35	1.5 ±0.1

To replicate R-GECO1 CaM the Asn109Asp and Ala23Asp mutations needed to be added to the CH-GECO3.0 mutations to determine if the K_d of CH-GECO3.1 could match that of R-GECO1. The addition of Ala23Asp to Ser77Thr, Gly21Asp, and Leu61Phe lowers the K_d from 44 nM to 23 nM; the response also decreases to 300%, and the Hill coefficient shifts to 1.3 (Figure 4.5B). It appears that restoration of the first EF hand increases the affinity for Ca^{2+}, which is likely due to a rearrangement of the loop structure.

In combination, Gly21Asp, Ala23Asp, Leu61Phe, Ser77Thr, Asn109Asp, and Asp191Gly result in a measured K_d of 35 nM with a Hill coefficient of 1.5 and

response of 372% (Figure 4.5B). It is apparent that there must be other mutations in the FP that are interacting with CaM and modulating the K_d since it is still 10-fold smaller than that of R-GECO1. Without a crystal structure of CH-GECO3.1, it is unclear which residues at the interface of mCherry and CaM must be interacting to confer such high affinity for Ca^{2+}.

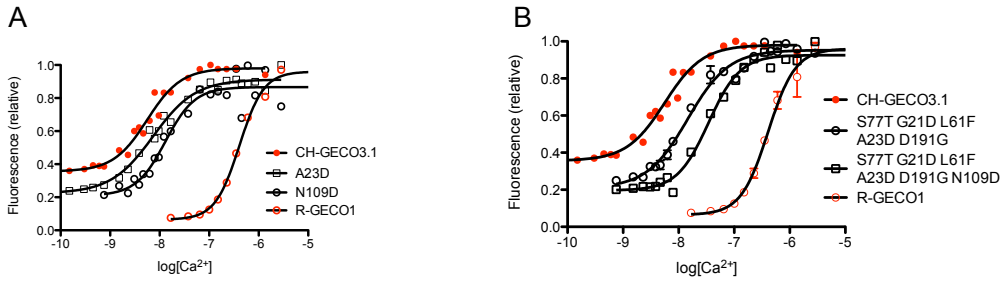


Figure 4.5 Ca^{2+} titration of CH-GECO3.1/R-GECO1 CaM hybrid mutants. A) Titration plots with Ala23Asp (CaM) mutant and Asn109Asp (CaM) mutant. B) Titration plots with Ser77Thr, Gly21Asp, Leu61Phe, Ala23Asp (CaM), and Asp191Gly mutations; Also Ser77Thr, Gly21Asp, Leu61Phe, Ala23Asp, Asn109Asp (CaM), and Asp191Gly mutations.

4.3.3 Investigating the Mechanism

To understand the differences between the CH-GECO3.1 and R-GECO1 fluorescent response mechanisms, I explored which residues were responsible for the mechanistic differences.

4.2.3.1 Differences between CH-GECO3.0 and CH-GECO3.1

There were several mutations between the two generations of sensors, though the major differences were found in CaM rather than the FP and were discussed in the previous section. Just two mutations, Ile147Thr and Gly191Asp, were located in the FP domain. The Ile147Thr mutation is likely responsible for increasing the dynamic range between the two generations as it is adjacent to the split point of the FP. When reverted to Ile in the context of CH-GECO3.1, the response decreases to 115% and the Hill coefficient increases to 1.7, which matches the properties of CH-GECO3.0 (Table 4.3). Since Thr147 is close to the phenolate of the chromophore the Thr could be either pointed inwards towards the phenolate, stabilizing the negative charge or could be enhancing the

interaction between the FP and CaM interface, which would also stabilize the Ca^{2+}-bound fluorescent state. The reversion of Asp191 to Gly was performed during the investigation of the effect of various mutations on K_d. The Asp191Gly reversion CH-GECO3.1 variant retains a similar response and K_d value as indicated in Table 4.2.

4.2.3.2 Hydrogen-Bond Network

From the CH-GECO3.1 pH titration curve in Figure 4.1B there is a pK_a around pH 6. We hypothesized that there is an amino acid residue being deprotonated around physiological pH. In the absence of Ca^{2+} this deprotonation serves to decrease the brightness of the protein. If we mutated the amino acid in question we should see the disappearance of this extra ionization in the pH titration curve.

Since the pK_a in question is around 6.5 our first guess at a likely amino acid was a histidine residue, which has a pKa of 6.0. There are no histidines in direct contact with the chromophore, but the residue could be situated further away participating in a hydrogen bond network with residues linked closer to the phenol group of the chromophore. One likely candidate was His75, which appears to H-bond in a network with Tyr193, Lys70, Glu148 and water molecules (Figure 4.6). I mutated each residue individually and characterized the pH profile and K_d to determine if this network was responsible for the signal increase. The individual pH profiles and Ca^{2+} titrations are shown in Figure 4.7A-D. Mutation of His75Gln gave a pH profile that was very similar to CH-GECO3.1, albeit with a reduced response of 158% (Figure 4.7A, Table 4.3). The Tyr193Phe mutation yielded similar results with a pH curve and Ca^{2+} titration curve that are similar to CH-GECO3.1 (Figure 4.7C, D).

For the mutants Lys70Gln and Glu148Gln I was unable to purify any red protein. As such there are no pH profiles or Ca^{2+} titrations. The absorbance scans of the purified proteins are shown in Figure 4.7E and 4.7F. Absorbance scans of both mutants indicate that the complete maturation of the protein chromophore is significantly disrupted as both mutations result in the formation of the blue intermediate absorbing at 382 nm (58).

From these four mutations I learned that they are not responsible for the pH behaviour of the sensor, but do play an important role in orienting the protein environment during chromophore maturation to promote the formation of the fully conjugated system.

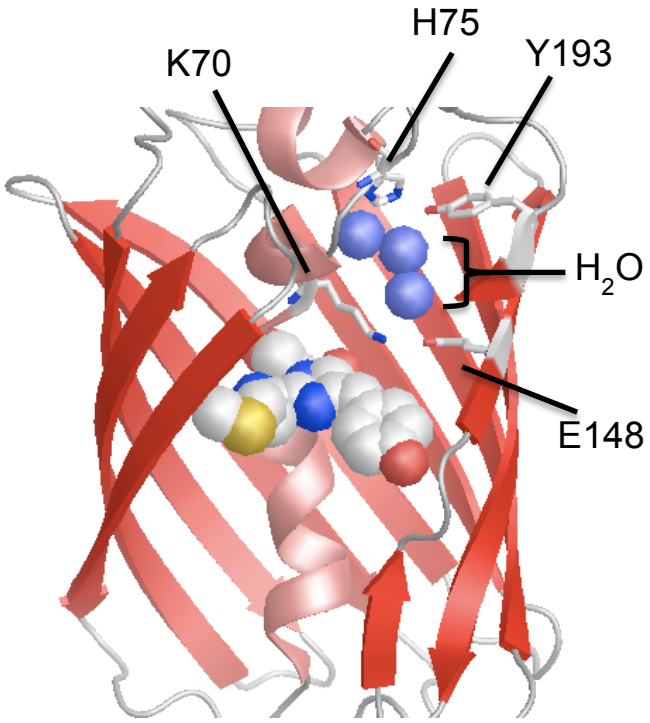


Figure 4.6 Hydrogen-bond network within mCherry. His75 and Tyr193 form hydrogen bonds, along with Lys70 and Glu148. There are three water molecules (blue spheres) between the residues that could also be interacting in hydrogen bonds.

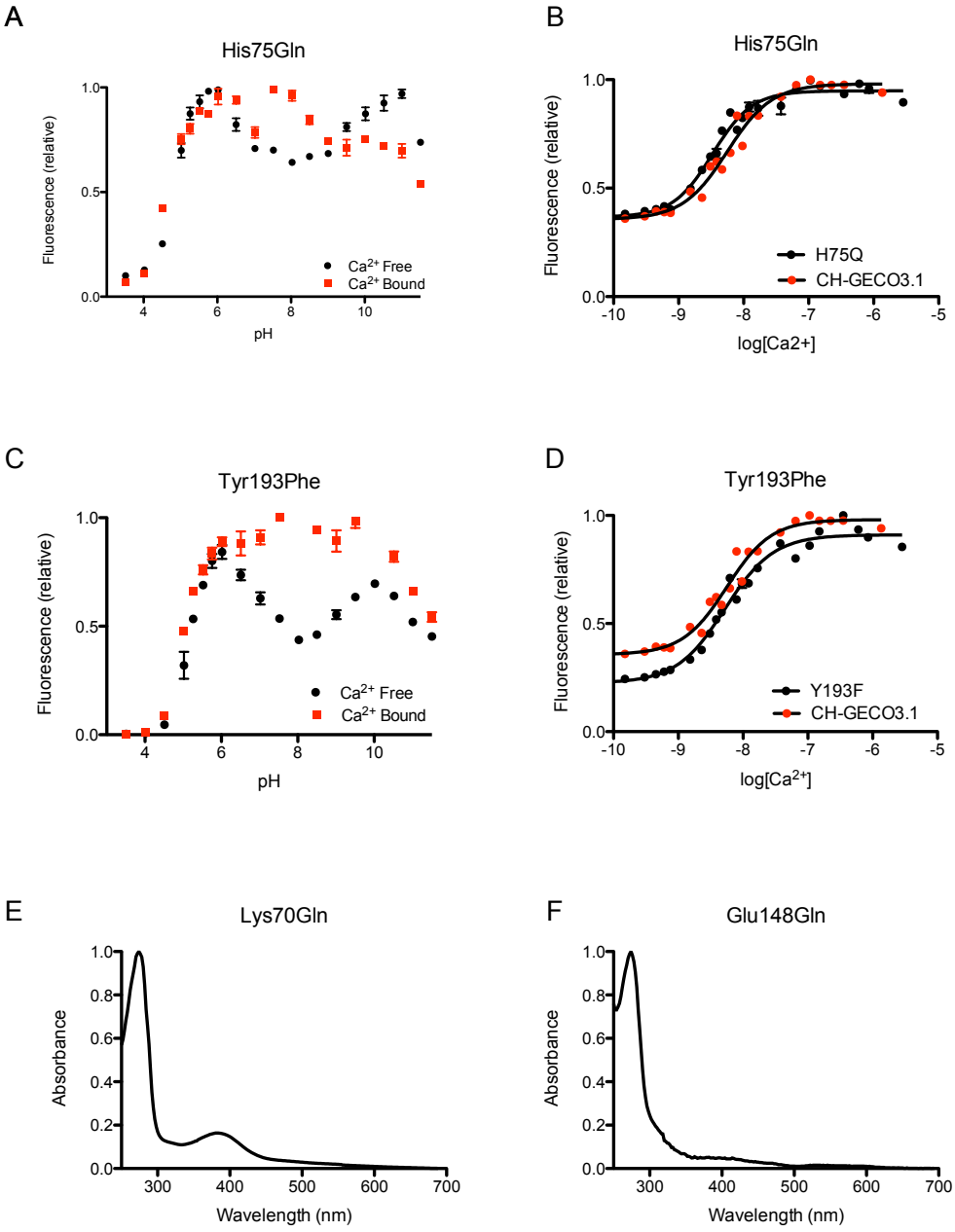


Figure 4.7 pH profiles and Ca^{2+} titration curves of His75Gln, Tyr193Phe and absorbance scans of Lys70Gln and Glu148Gln. A) pH titration curve of His75Gln with EGTA (black circles) or Ca^{2+} (red squares) where the Ca^{2+}-induced fluorescence increase remains B) Ca^{2+} titration curve for His75Gln remains similar to CH-GECO3.1 curve. C) pH titration curve for Tyr193Phe mutation with EGTA (black circles) or Ca^{2+} (red squares). D) Ca^{2+} titration curve (black circles) for Tyr913Phe compared with CH-GECO3.1 (red circles). E) Absorbance scan of purified Lys70Gln mutant. F) Absorbance scan of purified Glu148Gln mutant.

4.2.3.3 Alternate Amino Acids

After discounting our original H-bond network hypothesis the structure was re-examined to find possible alternative residues that contained a titratable functional group near the chromophore. A few other candidates were Tyr214, His204, His172, and Asp1 of CaM (Figure 4.8A). Though these residues are further removed from the chromophore environment, they are near the interface between the RFP and CaM. Once Ca^{2+} is bound the protonation state of one of these residues could feasibly be altered. The mutations were introduced individually in CH-GECO3.1, and the pH profiles were determined (Figure 4.8B-E). The results from the Ca^{2+} titrations are summarized in Table 4.3. Unfortunately the pH profiles of all four mutants still retained the characteristic pH increase at 6.5, indicating that none of these residues played a part in the sensing mechanism. The signal change for His204Gln and His172Gln is around 300%, while Tyr214Phe has a similar response to CH-GECO3.1 of 230%. Asp1 is the first amino acid in CaM and mutation to Gln did not alter the pH behaviour, but it did influence the K_d, which is increased to 53 nM as well as the fluorescence response, which is increased to 330%. From Figure 4.8A Asp1 is located directly at the split point and may likely be forming a hydrogen bond with a basic residue elsewhere in CaM or in the RFP. Mutation to Gln could influence the stability of the interface.

Table 4.3 Summary of Mutation Data

Variant	Response (%)	K_d (nM) ±1	Hill Coefficient
CH-GECO3.0	150	28	1.7 ±0.2
CH-GECO3.1	250	6	1.3 ±0.1
CH-GECO3.1 Thr147Ile	115	5	1.9 ±0.2
CH-GECO3.1 His75Gln	158	3	1.4 ±0.1
CH-GECO3.1 Tyr193Phe	305	5	1.2 ±0.1
CH-GECO3.1 His172Gln	304	3	1.1 ±0.1
CH-GECO3.1 His204Gln	285	7	1.3 ±0.1
CH-GECO3.1 Tyr214Phe	232	8	1.2 ±0.1
CH-GECO3.1 Asp1Gln (CaM)	307	53	1.7 ±0.2

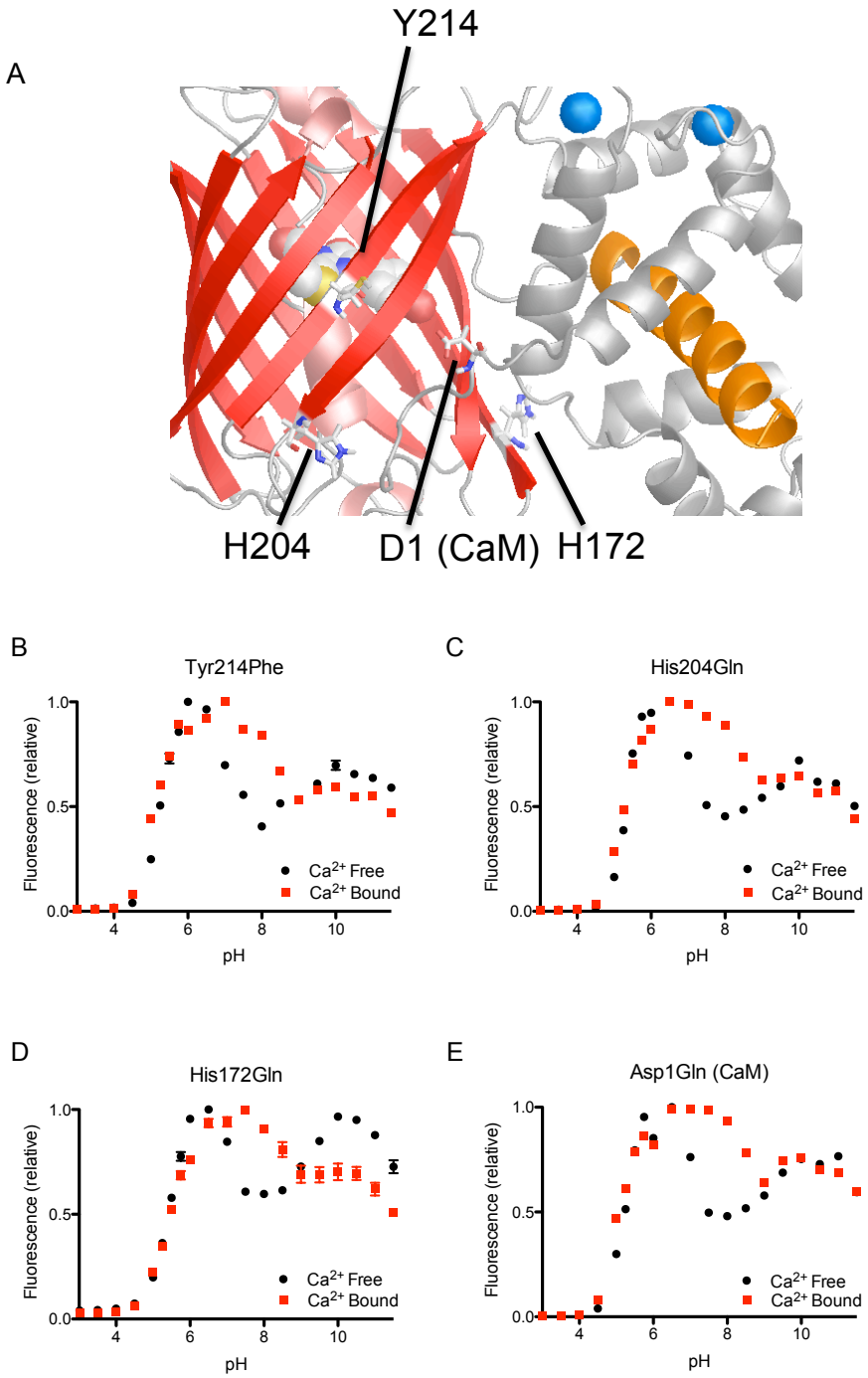


Figure 4.8 Location of H-network mutations in CH-GECO3.1 and the corresponding pH titration curves. A) Four mutation positions depicted in R-GECO1 structure Tyr214, His172, His204, and Asp1(CaM). B) pH titration curved of Tyr214Phe C) pH titration curve of His204Gln D) pH titration curve of His172Gln E) pH titration curve of Asp1Gln (CaM).

From these results it is apparent that neighbouring histidine and tyrosine residues do not play a role in the mechanism of action. Since the sequence of CH-GECO3.1 and R-GECO1 are so similar, yet behave completely different in the presence of Ca^{2+}, I decided to next investigate the regions where the two proteins differed the most, the linker region and the residues in the β8 strand.

4.2.3.4 Linker Region

Previous work in our lab with both single FP and FRET sensors showed that the length and composition of the linkers between the protein domains can have a substantial affect on the response of the sensor to the analyte of interest (153, 154). Accordingly, a lot of work goes into creating linker libraries and screening for improved response. The sequence alignment between R-GECO1 and CH-GECO3.1 (Figure 3.8) shows that the interface between the three domains is significantly different. R-GECO1 has a shorter linker between M13 and the RFP and does not contain the Gly-Gly linker between the RFP and CaM. The linker between M13 and R-GECO1 is Ser-Ser-Pro-Val while in CH-GECO3.1 it is Leu-Glu-Ser-Leu-Leu. At the other end the translated *MluI* restriction site serves as the linker between R-GECO1 and CaM, while in CH-GECO3.1 there is an additional Gly-Gly preceding the restriction site.

If the linkers were influencing the behaviour of the protein, replication of the R-GECO1 linker regions in the context of CH-GECO3.1 may result in a protein with properties that more closely resemble R-GECO1. The N- and C-terminal linker regions were introduced separately and then combined together in CH-GECO3.1 and the CH-GECO3.1-Gln163Lys mutant. CH-GECO3.1-Gln163Lys was also chosen as a template because the unpublished crystal structure of R-GECO1 reveals that Lys163 may play a key role in the Ca^{2+} sensing mechanism. The combination of this mutation, along with the linkers, may be sufficient to alter the function to match R-GECO1.

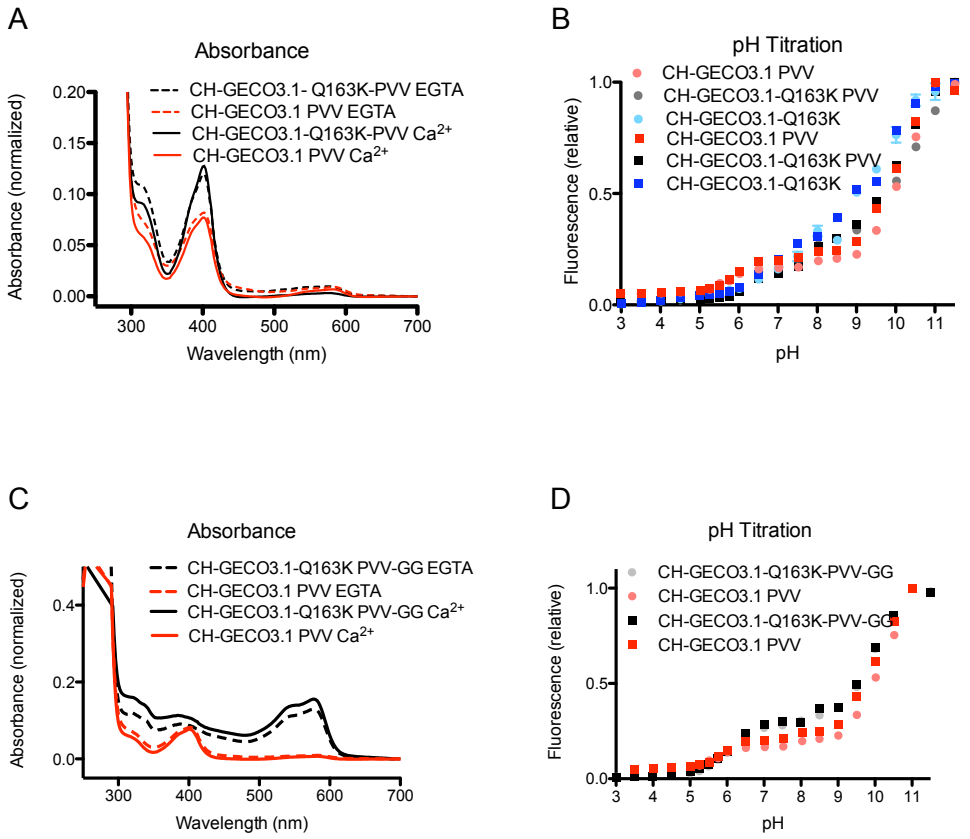


Figure 4.9 Absorbance scans and pH titration data for R-GECO1 linker mutations in CH-GECO3.1 and CH-GECO3.1-Gln163K. A) Absorbance spectra for CH-GECO3.1-PVV and CH-GECO3.1-Q163K-PVV mutants in the presence of EGTA (red dashed line, black dashed line, respectively) or Ca^{2+}(red line, black line, respectively). In both instances there is a much larger portion of protein in the blue form and neither mutant shows any absorbance change in the presence of Ca^{2+}. B) pH titration data of CH-GECO3.1-Q163K with EGTA (light blue circles) or Ca^{2+} (bright blue squares) along with CH-GECO3.1-Q163K-PVV mutant with EGTA (grey circles) and Ca^{2+} (black squares) and CH-GECO3.1-PVV with EGTA (pink circles) or Ca^{2+} (red squares). C) Absorbance spectra of CH-GECO3.1-Q163K-PVV-GG mutant with EGTA (dashed black line) and Ca^{2+} (solid black line) along with CH-GECO3.1-PVV with EGTA (dashed red line) or Ca^{2+} (solid red line). D) pH titration data of the CH-GECO3.1-Q163K-PVV-GG mutant with EGTA (grey circles) or Ca^{2+} (black squares) in comparison with the CH-GECO3.1-PVV mutant (same symbols as 4.9B).

Replacing the Leu-Glu-Ser-Leu-Leu-Ser sequence with Ser-Ser-Pro-Val-Val in both proteins, designated CH-GECO3.1-PVV and CH-GECO3.1-Q163K-PVV, respectively resulted in dimly red proteins. There was also a large portion of the protein in the intermediate blue state, similar to the results from the Lys70Gln and Glu148Gln mutations (Figure 4.9A). The most recent publications regarding the

maturation of the red chromophore indicate that the mature chromophore comes from a blue, rather than green, intermediate (58, 62, 271). It appears that the replacement of the longer linker with a shorter hydrophobic linker disrupts the efficient maturation of the chromophore. The pK_a of the chromophore in both mutants increases slightly, but is difficult to determine because the fluorescence increase associated with deprotonation of the chromophore is reduced. As shown in Figure 4.9B, there is a small inflection point around pH 5.5. This result is possible such that the new linkers along with the introduction of the positively charged Lys residue has generated an environment that distorts the chromophore away from planarity or increases its exposure to fluorescence quenchers, which would explain the decreased fluorescence. In CH-GECO3.1-Q163K-PVV the absorbance curve looks similar to the CH-GECO3.1-PVV (Figure 4.9A), although the pH titration curve in Figure 4.9B still resembles that of CH-GECO3.1-Gln163Lys.

The second step to generate complete R-GECO1 linkers was to delete the C-terminal Gly-Gly linker and mutate the terminal Leu to Asp-Ala in both CH-GECO3.1 (CH-GECO3.1-PVV-GG) and CH-GECO3.1-Gln163Lys (CH-GECO3.1-Q163K-PVV-GG). After several purification attempts no red protein was recovered from the CH-GECO3.1-PVV-GG. Surprisingly, the combination of both linkers in CH-GECO3.1-Q163K-PVV-GG was less disruptive than the N-terminal linker alone. In the absorbance spectrum shown in Figure 4.9C, there is a greater proportion of protein absorbing in the red region, while the pH titration profile in Figure 4.9D has better defined titration regions that more closely matches that of CH-GECO3.1-PVV. Despite the improvement in chromophore maturation CH-GECO3.1-Q163K-PVV-GG did not show any fluorescence response to Ca^{2+}. From these results, like most other fluorescent protein engineering endeavours, the adage “you get what you screen for” applies. Because the initial CH-GECO2.0 construct had the longer linkers all subsequent engineering results favoured this connection. Undoubtedly if the shorter linkers had been introduced at an earlier stage in the evolution, the fluorescence could likely be rescued after several rounds of mutagenesis, and larger responses to Ca^{2+} might have been realized.

4.2.3.5 Differences between CH-GECO3.1 and R-GECO1

Though the linkers in the protein play an important role for the chromophore maturation all of the mutagenesis until now has still not been sufficient to explain the disparity between R-GECO1 and CH-GECO3.1. The final area where amino acids could be directing the behaviour are residues located at the interface between the RFP and CaM in the β8 strand along with Trp83, which plays a role in the pronounced red-shift of mCherry (245). From the sequence alignment in Figure 3.8 there are a few amino acids in the β8 strand, which are near the split point that could potentially play a role in the proteins behaviour: Gln163, Lys166, and Gly159, and Trp83 in between the β3 and β4 strand. The Gln163Lys mutation was discussed briefly in Section 4.2.3.4, but will be discussed again in further detail.

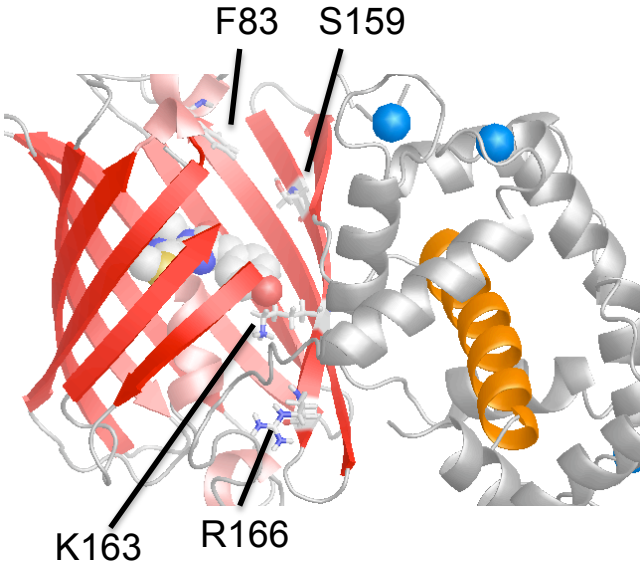


Figure 4.10 Predicted structure of R-GECO1 with Ser159, Lys163, Arg166, and Phe83 shown. The corresponding Gly159Ser, Gln163Lys, Lys166Arg, and Trp83Phe mutations were introduced into CH-GECO3.1 to match the residues of R-GECO1.

Based on the predicted structure of R-GECO1 (14), the four residues that were expected to be pointed inwards towards the chromophore or interacting with CaM are depicted in Figure 4.10. These four positions in CH-GECO3.1 contained

alternate amino acids (Figure 3.8) and they could play a role in the mechanism due to their proximity to the chromophore. Each mutation was introduced separately and a few were combined to determine if the mutations had additive affects on the K_d, pH behaviour, or Ca^{2+} response.

Residue 83 was originally a Lys in DsRed, but mutagenesis to a hydrophobic residue such as Met or Leu red-shifted the fluorescence emission (59). Based on the crystal structures of DsRed and mCherry a reason for this wavelength shift was offered (66). The original Lys83 was found to interact with Lys70, which in turn hydrogen bonded to Glu215, but mutation to a bulkier Arg prevented the formation of the red chromophore, while replacement with a Met caused a red-shift in emission to 602 nm. These changes were thought to occur because of the reorganization of the Lys70 residue as the added bulk of an Arg residue would disrupt the hydrogen bonds between Lys70 and Glu215, while the hydrophobic Met would influence the electron density of the chromophore (66). In mCherry there is the Lys83Leu mutation, which also reorganizes Lys70 and moves Lys away from the methylene bridge of the chromophore and instead forms hydrogen bonds with a water molecule and Glu148. Glu215 in turn becomes protonated and hydrogen bonds with the imidazolinone moiety of the chromophore (245). The loss of electron density around the phenolate of the chromophore is also thought to contribute to the red-shift in excitation/emission. When the Trp83Phe was introduced into CH-GECO3.1, no noticeable changes were observed in the fluorescent response, K_d, or excitation/emission characteristics. The introduction of a conservative Phe mutation could be expected to reorganize the amino acids nearby due to size differences, but the electrostatic environment must not be perturbed as evidenced by the unchanged excitation/emission wavelengths.

Though the Gly159Ser mutation is admittedly a relatively conservative alteration, the presence of a hydroxyl side chain could potentially introduce a new hydrogen bond pair, which may influence the packing of the interior. From the pH and Ca^{2+}titration in Figure 4.11A & B, respectively, the behaviour of the sensor is not changed, but the response is lowered to only 184% and has a smaller Hill coefficient of 0.9 (Table 4.4).

Table 4.4 Summary of β8 strand mutations

Mutation	Response (%)	K_d (nM) ±1	Hill Coefficient
CH-GECO3.1	250	6	1.7 ±0.2
CH-GECO3.1 Trp83Phe	280	7	1.1 ±0.1
CH-GECO3.1 Gly159Ser	184	8	0.9 ±0.1
CH-GECO3.1 Gln163Lys	N/A	N/A	N/A
CH-GECO3.1 Lys166Arg	200	3	1.2 ±0.1
CH-GECO3.1 Gly159Ser Gln163Lys	N/A	N/A	N/A
CH-GECO3.1 Gln163Lys Lys166Arg	57	14	1.8 ±0.5
CH-GECO3.1 Gln163Met	54	2	1.2 ±0.2
CH-GECO3.1 Gln163Asp	N/A	N/A	N/A

As mentioned in the previous section the Gln163Lys mutation completely abolished the response of CH-GECO3.1 and appears to affect the protonation state of the chromophore as evidenced by the pH titration curve in Figure 4.11C. However, the absorbance, excitation, and emission wavelengths remain unchanged. During the evolution of mCherry from DsRed, position 163 remained a hotspot for shifting the fluorescence hue of the chromophore. In the crystal structure of DsRed, position 163 is a Lys residue, which forms a direct salt bridge with the phenolate of the chromophore, stabilizing the phenolate group in the anionic state (66). The engineering of mRFP1 from DsRed found mutations Lys163Gln (dimer2) or Lys163Met (mRFP1), which are no longer able to form a salt bridge with the phenolate chromophore and as a result likely shift the polarity of the chromophore (43). This shift in polarity may shift the electron density towards the imidazolinone ring and promote the extended conjugation. The progenitor to mRFP1, dimer2 had slightly blue-shifted excitation and emission wavelengths (552/579 nm), while mRFP1 is significantly red-shifted to 582/604 nm. The final steps in the evolution of mCherry involved screens for red-shifted emission, tolerance of N-terminal fusions, and improved folding. During those efforts, the Met163Gln mutation was rediscovered and retained since this mutation abolished the 510 nm absorbance peak and provided increased photostability (25, 109). These results reinforce the importance of this residue in mCherry, but given the alternate conformation brought about by permutation and the Ca^{2+}-binding domains this residue must be playing additional roles in R-

GECO1 and CH-GECO3.1. In the unpublished R-GECO1 crystal structure (Eric Schreiter, personal communication), Lys163 appears to be interacting directly with the phenolate of the chromophore in the Ca^{2+} bound state, which suggests that this residue could be responsible for lowering the pK_a of the chromophore by stabilizing the negative charge on the phenolate.

When the two mutations, Gly159Ser and Gln163Lys, were combined together there was still no response to Ca^{2+} (Figure 4.11F) and the pH titration curves looked similar to the Gln163Lys titration curve (Figure 4.11C,E). Introduction of Lys166Arg still yielded a pH curve that had the inflection point at pH 6.5 and a K_d around 3 nM (Figure 4.12A,B). As shown in Figure 4.10, Lys166 is not pointed inside the protein barrel, but is instead pointed outwards and from the predicted structure it could be hydrogen bonding with a backbone amino group of the N-terminal portion of the β7 strand to stabilize it, thereby restricting solvent access to the chromophore. If this is the role of Lys166, mutation to Arg should not significantly disrupt this arrangement, as it is also a basic residue capable of forming a hydrogen bond. More interesting was that the combination of Gln163Lys Lys166Arg seemed to partially compensate for the abolishment of response seen in the Gln163Lys mutant. The pH curve of the double mutant closely resembles the Gln163Lys curve, but there is a slight increase in fluorescence in the presence of Ca^{2+} as seen in Figure 4.12C. The Ca^{2+} titration curve indicates the protein has a K_d of 14 nM and a Hill coefficient of 1.8 (Table 4.4), albeit with greatly diminished response of 57%. It is unclear why the combination of the two mutations would restore some Ca^{2+} responsiveness, but if Lys163 was contacting the phenol of the chromophore in an undesirable conformation the introduction of a larger basic residue at position 166 may reorient the β8 strand such that the positive charge is now engaged in a more useful hydrogen bond or salt-bridge away from the chromophore.

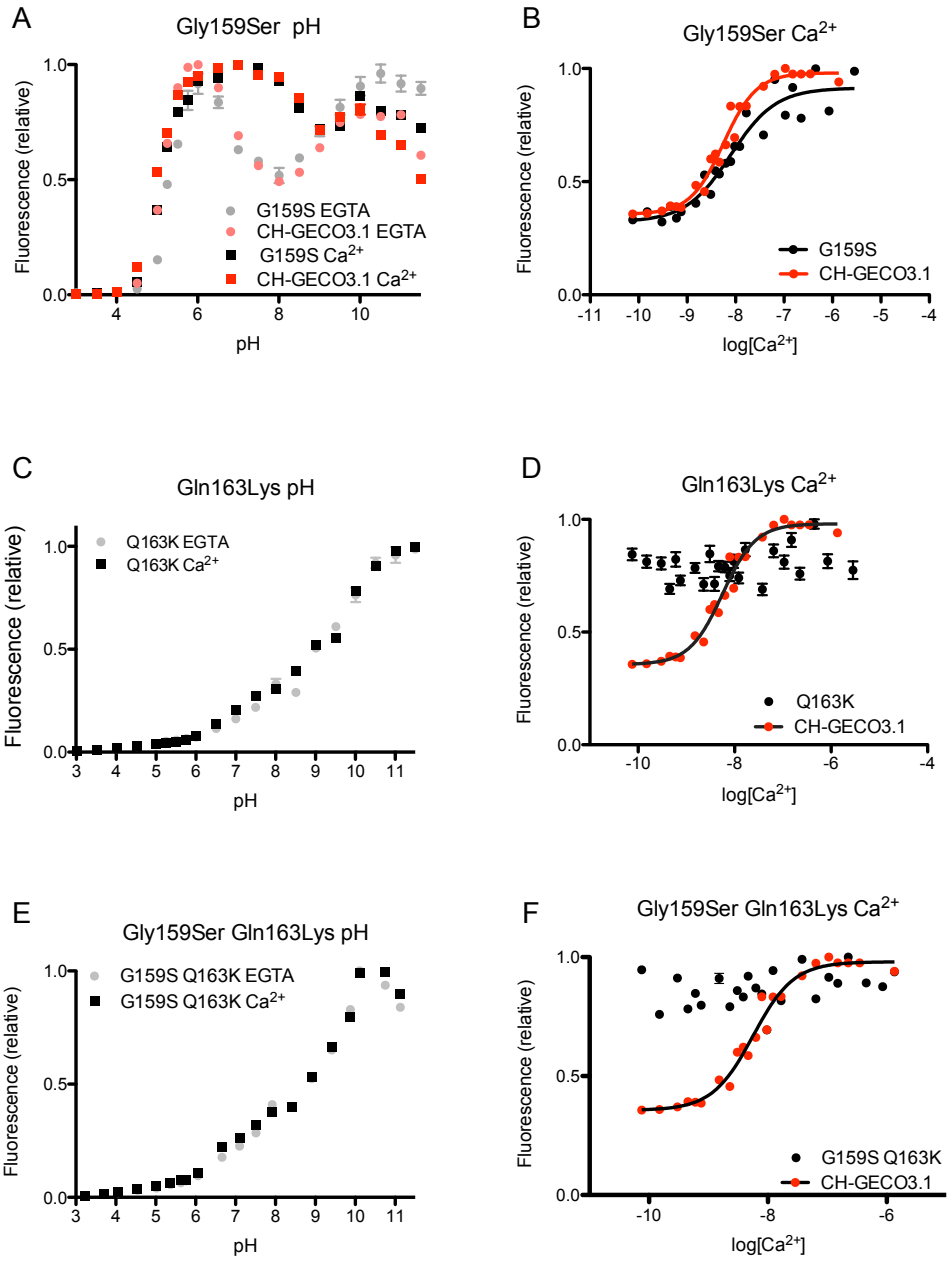


Figure 4.11 pH and Ca^{2+} titration curves for the CH-GECO3.1 mutants Gly159Ser, Gln163Lys, and Gly159Ser/Gln163Lys. A) Gly159Ser pH curve with EGTA (grey circles) or Ca^{2+} (black squares) compared with CH-GECO3.1 (pink circles, red squares). B) Ca^{2+} titration curve of Gly159Ser (black circles). C) pH titration curve of Gln163Lys mutant with EGTA or Ca^{2+}. D) Ca^{2+} titration of Gln163Lys (black circles) compared with CH-GECO3.1 (red circles). E) Gly159Ser/Gln163Lys mutant pH curve. F) Gly159Ser/Gln163Lys mutant Ca^{2+} titration.

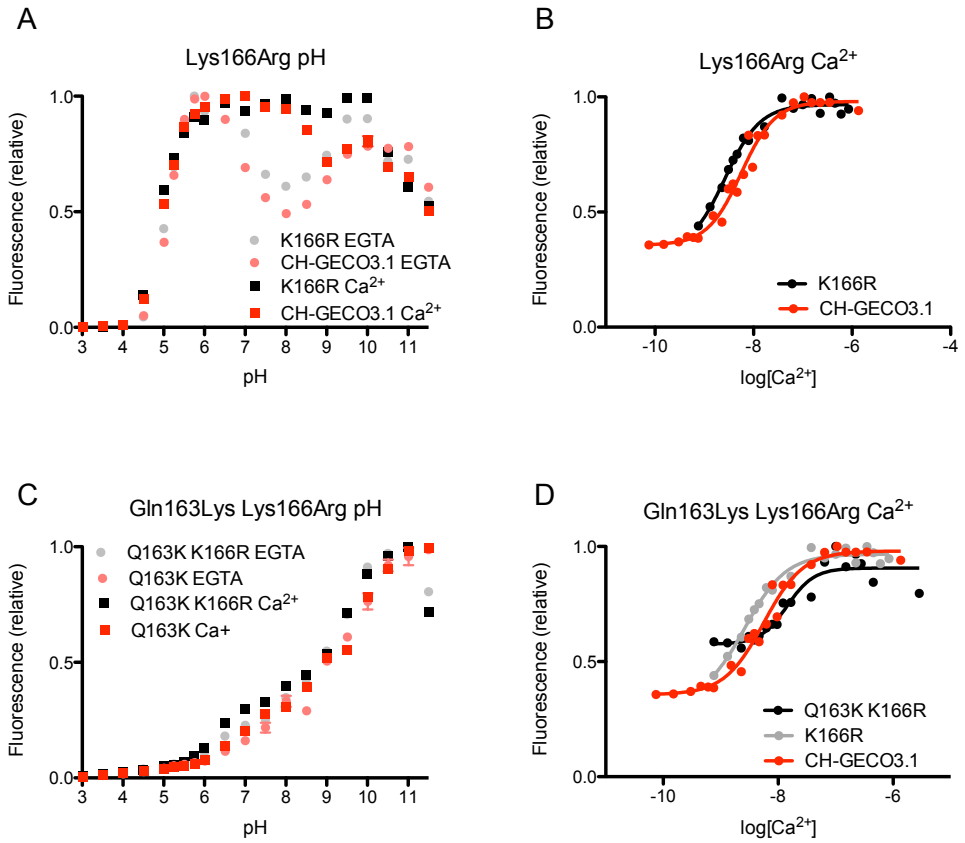


Figure 4.12 pH and Ca^{2+} titration curves for Lys166Arg and Gln163Lys/Lys166Arg mutations in comparison with CH-GECO3.1. A) pH titration of Lys166Arg with EGTA (grey circles) or Ca^{2+} (black squares), overlaps with CH-GECO3.1 with EGTA (pink circles) or Ca^{2+} (red squares). B) Ca^{2+} titration of Lys166Arg (black circles) compared with CH-GECO3.1 (red circles). C) Gly163Lys/Lys166Arg pH titration with EGTA (grey circles) or Ca^{2+} (black squares), in comparison with Gln163Lys with EGTA (pink circles) or Ca^{2+} (red squares). D) Ca^{2+} titration of Gln163Lys/Lys166Arg (black circles) in comparison with Lys166Arg (grey circles) and CH-GECO3.1 (red circles).

4.2.3.6 Conclusions for mechanistic determination

The characterization of CH-GECO3.1's pK_a, K_d, and pH behaviour has led to a more in depth understanding of its mechanism. All previously designed single-FP Ca^{2+} biosensors operate via pK_a modulation, but CH-GECO3.1 did not follow this trend. I set out to understand the structural differences between CH-GECO3.1 and the closely related R-GECO1 that could account for the mechanistic differences. From the pH titration curve in Figure 4.1 there is an extra titratable group present around pH 6 for all of the CH-GECO variants.

Without a crystal structure of the protein it is difficult to predict how specific amino acids are oriented near the chromophore, especially given how distorted we expect the β7 and β8 strand residues to be from the mCherry crystal structure. However, the Pro-Val-Val and Gln163Lys mutations did have the most striking affect on the protein and some potential mechanistic conclusions can be drawn from those results. From the unpublished crystal structure of Ca^{2+}-bound R-GECO1 it appears that Lys163 and Ser147 are both pointed inwards towards the chromophore. For the function of this sensor it may be that the positive charge on the Lys and the hydroxyl group of Ser are stabilizing the phenolate in the Ca^{2+}-bound state, causing the drop in the pK_a. For CH-GECO3.1, the pK_a remains around 5.1 and it has a Gln at position 163 and Thr at 147. Relative to Ser147, Thr147 has an additional methyl group, but it still could be helping to stabilize the negative charge of the phenolate or coordinating with another residue. In going from CH-GECO3.0 to CH-GECO3.1, the Ile147Thr mutation accounted for the increase in response, which indicates that the hydroxyl group must be forming a beneficial hydrogen bond with either the chromophore or a neighbouring residue to increase the fluorescence response. Alternatively, Ser146 in CH-GECO3.1 is pointed towards the chromophore phenolate, hydrogen bonding to stabilize the negative charge. The introduction of the Gln163Lys mutation had the most drastic effect on the pH profile (Figure 4.11C). The protein no longer responds to Ca^{2+} and the pH curve has no extra inflection point at pH 6 and a depressed inflection point corresponding to the titration of the chromophore phenol group. The addition of the Lys166Arg mutation partially restores the Ca^{2+} response (Figure 4.12C and D). Based on these results as well as those from mutagenesis of the linker we have proposed a structural reason for the fluorescence increase in CH-GECO3.1.

We propose that, in the Ca^{2+} free state, Gln163 is pointed outside the barrel, giving the chromophore more rotational and vibrational freedom, which is seen from the low quantum yield of 0.06. When Ca^{2+} is present, Gln163 moves inwards and could be interacting with the chromophore phenolate. In the crystal structure of mCherry, Gln163 is stacked below the chromophore sandwiched between Trp143 and Ile161. In Ca^{2+}-bound CH-GECO3.1 Gln163 is likely in the same orientation as in mCherry, which may both orient the chromophore in a

conformation that increases the quantum yield and extinction coefficient. When a Lys is present the positive charge may create too many repulsive interactions in the hydrophobic interior.

Though Gln163 may play the largest role in the fluorescence increase there must be other residues that are involved. The introduction of the Pro-Val-Val linker also abolished the response of the sensor and the pH curve resembled the mCherry pH titration, which means there must be residues at this junction that are involved in the mechanism. Between M13 and the RFP the *XhoI* restriction site codes for Leu-Glu, which is the same linker that is present in the GCaMP series (16). In the crystal structure of Ca^{2+}-free GCaMP2, Glu61 is in close proximity to the phenol of the chromophore, which helps to stabilize the neutral phenol (19). In the presence of Ca^{2+}, Glu61 in the linker forms a hydrogen bond with Arg81 (GFP numbering 168), which helps to stabilize the interface between CaM and GFP, and may restrict chromophore movement (18). Since the pK_a of CH-GECO3.1 is much lower than GCaMP2 at physiological conditions, the chromophore is already deprotonated so it is unlikely that the Glu is stabilizing the neutral form. An alternate explanation is that in the Ca^{2+}-free state the Glu is still pointed towards the chromophore, which may introduce electrostatic clash with the phenolate, resulting in the decrease of fluorescence in the apo state around pH 6 as seen in Figure 4.1B-D. When Glu61 is no longer near the chromophore, the chromophore can adopt a more favourable orientation, increasing the quantum yield and/or extinction coefficient. This hypothesis is not discounted by the results from the linker mutation study. When the sequence between M13 and RFP is mutated to Pro-Val-Val the Glu residue is removed, along with Ser146. From the pH titration curve of this mutant (Figure 4.9 B and D), the apo state no longer has the decreased emission from pH 6 to pH 9. The Glu and Ser146 could act cooperatively such that in the apo state Glu clashes with the phenolate of the chromophore. In the Ca^{2+} bound state Glu is pointed away from the chromophore and Ser146 is oriented towards the chromophore, stabilizing the negative charge. Ser146 also plays in important role in the proper maturation of the chromophore as a large portion of the protein becomes trapped as the blue intermediate when this residue is mutated.

The results of the K_d mutagenesis were curious such that when trying to manipulate the K_d of the sensor, mutation of residues in CaM were not sufficient to restore the K_d to a value that more closely matched the K_d of R-GECO1 (482 nM). The sum of the differences in the linkers and the protein interface are also playing a critical role in determining the affinity for Ca^{2+}. These results should offer food for thought for scientists who want to manipulate the K_d of Ca^{2+} biosensors for specific applications. Engineering a Ca^{2+} binding pair may not be sufficient for predicting the K_d. Rather, the interface between the Ca^{2+} binding domain and the FP must be considered as well.

4.2.5 Alternate Mutations and Colours

Since residue 163 is important for the function of the CH-GECO3.1 sensor a few other mutations at this position were introduced to determine if the function would be retained after changing the spectroscopic behaviour. The two amino acids Met and Asp were introduced at position 163 because Met is present in the original mRFP1 (43) and Asp is found in an RFP with long Stokes-shift (113).

The Gln163Met mutation introduces a hydrophobic residue next to the chromophore. The absorbance and emission spectra closely resembles mCherry, with the exception of the appearance of a small shoulder around 500 nm in the absorbance spectrum, which is a non-emitting green species (Figure 4.13A). This result could be expected as it was mentioned that the introduction of Gln163 during the evolution of mCherry eliminated this shoulder peak (25). The Ca^{2+} response of this protein drops to 54% while the K_d decreases to 1.5 nM (Figure 4.13B, C), however the absorption and emission is red-shifted to 586/610 nm, which matches that of mCherry. This red-shift result agrees with the general trend observed during mCherry evolution that increasing the hydrophobicity of the chromophore environment caused a red-shift in the fluorescence spectra (245). In the presence of Ca^{2+}, the absorbance at 582 nm increases 11%, which means that the remaining signal increase must come from a 43% increase in quantum yield. The exact quantum yield and extinction coefficient were not determined for this protein.

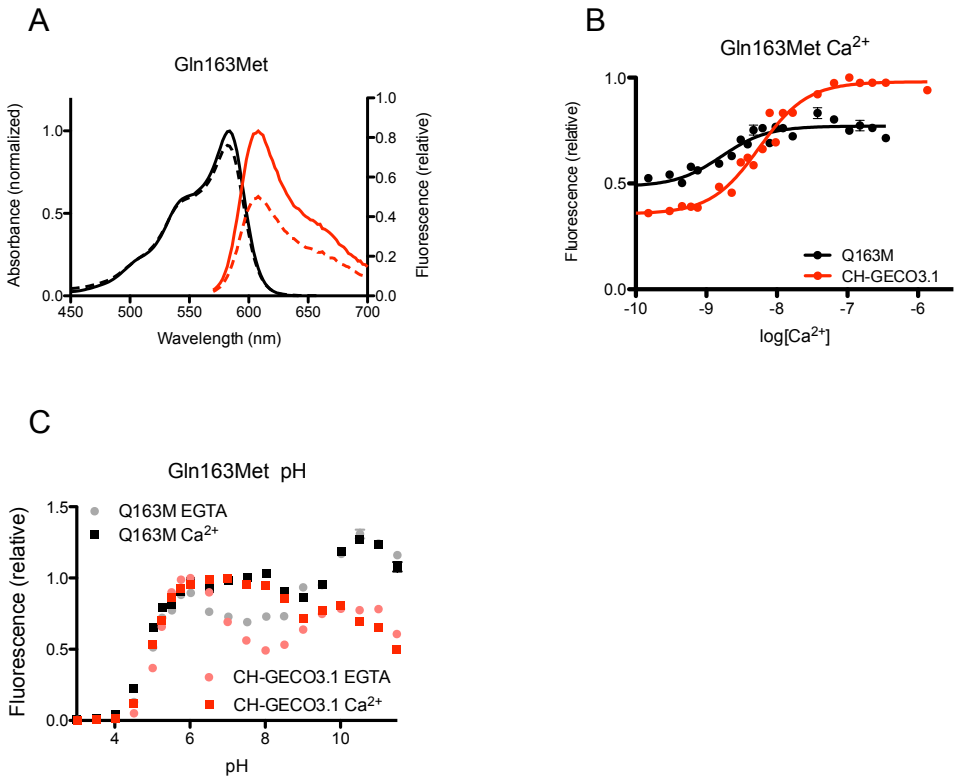


Figure 4.13 Spectral characterization of Gln163Met, including absorbance, fluorescence emission, Ca^{2+} titration, and pH titration. A) Absorbance scan with EGTA (dashed black line) or Ca^{2+} (solid black line) and fluorescence emission of Gln163Met in the presence of EGTA (dashed red line) or Ca^{2+} (solid red line). B) Ca^{2+} titration of Gln163Met (black line) in comparison with CH-GECO3.1 (red line). C) pH titration of Gln163Met with EGTA (grey circles) or Ca^{2+} (black squares) compared with CH-GECO3.1 in EGTA (pink circles) or Ca^{2+} (red squares).

The typical Stokes-shift between excitation and emission wavelengths is 20-30 nm. Piatkevitch and coworkers reported the generation of two long stokes-shift (LSS) RFPs from the far-red emitting mKate (21, 113). The final two variants LSS-mKate1 and LSS-mKate2, had excitation and emission wavelengths of 463/624 nm and 460/605 nm, respectively. They were able to stabilize the ground state neutral chromophore, which undergoes an excited state proton transfer (ESPT) to emit from the deprotonated chromophore. This behaviour is the same for the original GFP, whereby the neutral chromophore absorbs light at 450 nm, but fluorescence emission occurs from the deprotonated chromophore

(48). After excitation the phenol proton is transferred through a proton network. Through their mutagenesis they found that either Ser145Asp, Ser163Asp, or Met165Lys/Glu/Asp were sufficient to introduce the Stokes-shift phenotype, though further optimization was required to generate the final LSS-variants.

An LSS-Ca^{2+} sensor would have the same benefits as an LSS-RFP in that two-photon microscopy could be used allowing for longer, less damaging excitation wavelengths with less autofluorescence (272). The use of longer wavelengths would also allow for multi-parameter imaging with less cross-talk between chromophores. Given these potential benefits, the Gln163Asp mutation was introduced into CH-GECO3.1 to see if a LSS phenotype could be engineered.

Unfortunately the protein showed no responsiveness to Ca^{2+}, but the spectroscopic behaviour was interesting enough to justify further characterization. In Figure 4.14A the absorbance of the Gln163Asp mutation is shown at three time points; black dashed and solid lines for the protein characterized the day of purification, dark grey lines for the protein characterized after being left overnight at 4°C, and light grey lines for the protein after several weeks at 4°C. The initially purified protein had a mixture of deprotonated and protonated red chromophore as evidenced by the absorbance peaks at 582 nm and 445 nm in Figure 4.14A (black solid and dashed line). When left overnight the 582 nm absorbance decreases, leaving a strong absorbance peak centered around 406 nm with a shoulder remaining at 445 nm (Figure 4.14A, dark grey dashed and solid lines). Excitation of the fresh protein at 450 nm resulted in emission around 520 nm and 610 nm (Figure 4.14B, grey line), but the red emission could be from excitation of the tail of the 550 nm absorption peak, while excitation at 550 nm gave the expected 610 nm emission (Figure 4.14B, black line). Monitoring 610 nm emission from 450 nm and 550 nm excitation with increasing pH in Figure 4.14C there is an increase in emission from 550 nm excitation at the expense of the 450 nm excitation. The pH curve for red fluorescence with excitation at 550 nm (Figure 4.14C) shows an increase above pH 9.5 due to the blue shift in emission wavelength, which has been described in the literature (245) and elsewhere in this thesis. In Figure 4.14D the excitation wavelengths were scanned at various pH values while monitoring 610 nm emission. From these curves we see that as

pH increases the proportion of protonated chromophore decreases (450 nm), while the deprotonated portion increases (550 nm), which means that the red peak seen in Figure 4.14B from exciting the tail of the 550 nm excitation peak, rather than an LSS emission from the neutral chromophore.

The emission and excitation spectrum in Figure 4.14E was collected after the protein had been left at 4°C for several weeks. The absorbance spectra of the same sample (Figure 4.14A, light grey line) reveals that the red component is not detectable and only a small amount of absorption remains at 406 nm. The decrease in the absorption peaks is likely due to protein degradation as the sample was not stored with any protease inhibitor. Nonetheless it was possible to collect fluorescence data from a diluted protein solution as evidenced by the spectra in 4.14E. It appears that the main component is a blue species that absorbs at 406 nm and emits at 466 nm, while the 430 nm shoulder emits at 510 nm. The literature contains several descriptions of blue-to-red fluorescent timers generated from mCherry (63). In timer proteins, the FP initially forms a blue chromophore structure and will slowly undergo further maturation to form the fully mature red chromophore. However, based on the chromophore pathway suggested by Strack et al. there is at least one non-reversible dehydroxylation and dehydration step (Figure 1.7). If this is the case it would seem unlikely that the red chromophore can revert to the blue-form. However, there may be a large portion of immature protein that slowly forms the blue structure, which is why it doesn't appear until 6-12 hours after purification. A more likely reason for the loss of the red fluorescent species could be due to the addition of water to the C=N bond in the imidazolinone ring or the slow hydrolysis of the mainchain acylimine moiety. A variety of hydrolytic degradation products of the blue-to-red timers were characterized (64) and a very likely reason for the disappearance of the red peak is due to a similar process. Alernatively, the resulting species could also be the non-absorbing species designated 'X' by Strack et al (58). Regardless of the strange behaviour of this protein, the complex chromophore chemistry and photophysics suggest that the introduction of a LSS phenotype could be possible with further mutagenesis.

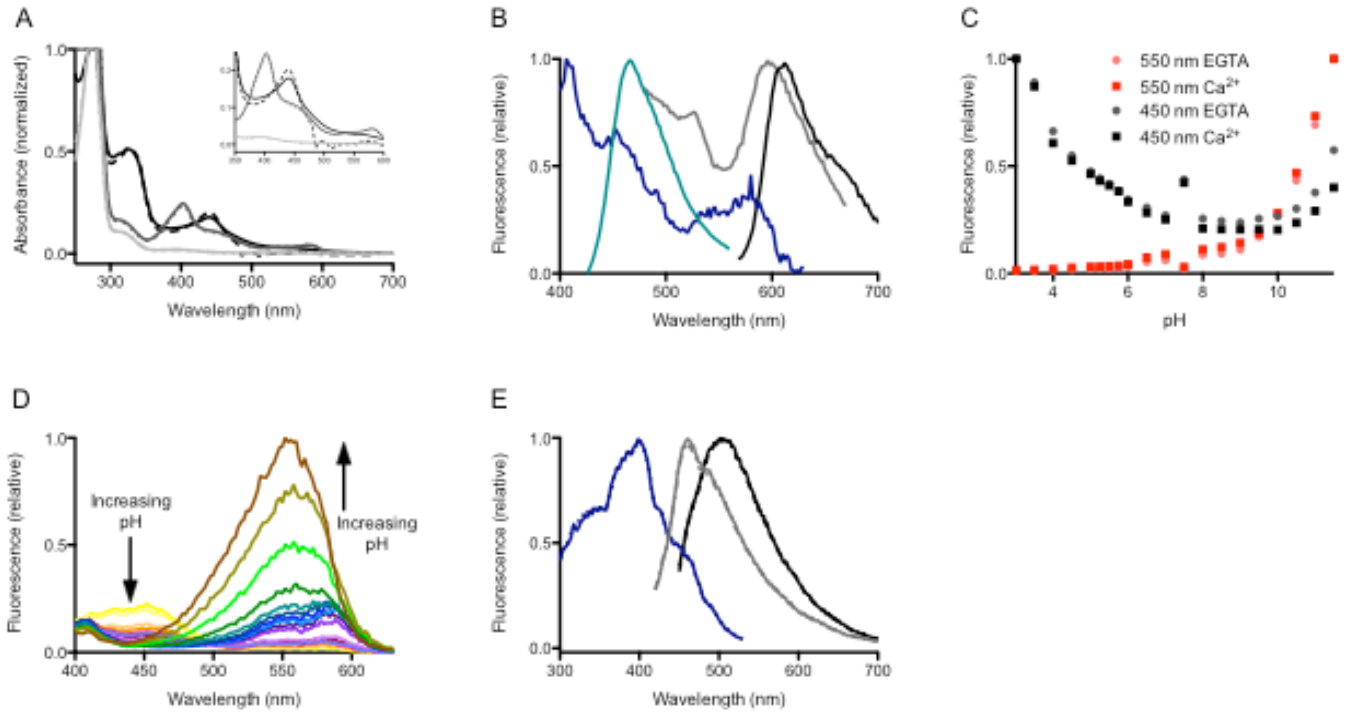


Figure 4.14 Spectral characterization of CH-GECO3.1 Gln163Asp. A) Absorbance scans of Gln163Asp mutant with EGTA (dashed lines) or Ca^{2+} (solid lines). Same day purification (black lines), overnight at 4°C (dark grey), several weeks at 4°C (light grey). The inset is a zoom in between 350 to 600 nm. B) Same-day protein characterization. 550 nm excitation (black line); 450 nm excitation (grey line); 385 nm excitation (teal line); excitation scan from 650 nm emission (dark blue line). C) pH titration of Gln163Asp monitoring the 610 nm emission peak from 450 nm and 550 nm excitation D) Fluorescence excitation spectra measured at 650 nm emission. E) Gln163Asp fluorescence after several weeks of storage. Excitation from 550 nm emission (dark blue line); emission from 400 nm excitation (grey line); emission from 430 nm excitation (black line).

An orange GECO (O-GECO) was engineered from R-GECO1 by introducing several mutations known to cause blue shifted fluorescence hues (Jiahui Wu, unpublished results). The mutation Met65Thr, which is found in mOrange, is the main amino acid change that results in a protein that is blue shifted about 35 nm, with excitation/emission of 557/569 nm (25). The reason for this blue-shift is a cyclization reaction that occurs when the Thr65 hydroxyl group reacts with the carbonyl carbon of Phe64 to form a 2-hydroxy-dihydrooxazole ring structure. This structure is less conjugated than the red chromophore and therefore emits orange light (Figure 1.14, mOrange).

When the Met65Thr mutation was introduced into CH-GECO3.1 the freshly purified protein, designated as CHO-GECO1, had absorption peaks at 410 nm and 550 nm (Figure 4.15A, red solid and dashed line). The 410 nm peak could be either an immature neutral green chromophore or a neutral orange chromophore. In the original description of mOrange there was approximately 20% immature green chromophore while there has also been publication of an LSS-mOrange, in which the neutral mOrange chromophore is stabilized (110, 245). Comparing the absorbance peaks with EGTA or Ca^{2+} in Figure 4.15A, the 410 nm peak decreases and the 550 nm peak increases in the presence of Ca^{2+}. This result indicates that this shorter wavelength peak is the neutral chromophore even though the wavelength does not exactly match the reported absorbance of the neutral mOrange chromophore of 437 nm (110). Also possible is that the 410 nm absorbance is an immature blue chromophore as excitation at 410 nm results in emission at 480 nm. From this result it is not entirely clear why the addition of Ca^{2+} would promote the complete maturation of the orange chromophore, converting the blue chromophore into the fully mature orange chromophore. Since the colour change of the CHO-GECO1 solution can be observed as soon as mixing with Ca^{2+} buffer, it is more likely that a deprotonation is occurring, rather than a chromophore maturation step (i.e. dehydroxylation, dehydration, and/or cyclization reaction, which would be expected to be relatively slow). The pH titration indicates the pK_a of the chromphore has increased to around 6.0, which is consistent with mOrange, which has an acid sensitive pK_a of 6.5 (109). Despite the increased pK_a the response of CHO-GECO1 is improved to 400%.

As with several of the mutations made to CH-GECO3.1, a peculiar change in the absorbance spectrum was observed after the protein was stored several weeks at 4°C. The black traces in Figure 4.15A indicate an extra absorbance peak at 615 nm and in the presence of Ca^{2+} both the 550 nm and 615 nm peaks increase at the expense of the 415 nm peak. Excitation at 415 nm results in emission at 475 nm and excitation at the other two peaks resulted in fluorescence emission around 570nm and 630nm, respectively. The Ca^{2+} titration curves resemble the titration curve of newly purified CHO-GECO1 with excitation at 515 nm and 600 nm in Figure 4.15B (115). The 615 nm peak could be a further conjugated chromophore that forms from mOrange (Figure 1.16A). In 2011, a PSmOrange was decribed, which, when irradiated with green light extended the conjugation of the chromophore producing a chromophore with absorption at 615 nm (115). As far as I am aware nothing untoward happened with this protein sample, though it was stored at a relatively high concentration. Further work is necessary with this variant to determine whether irradiation with green light will irreversibly convert CHO-GECO1 into a far-red emitting species.

As with CH-GECO3.1, the same Gln163Asp LSS mutation was introduced into CHO-GECO1 with different results. From the spectra in Figure 4.15D the absorbance has completely shifted to 430 nm, while excitation at this peak produces orange emission of 570 nm, indicative of a LSS emission. Unfortunately this protein did not show any responsiveness to Ca^{2+}. The negatively charged Asp residue at position 163 may be interfering with mechanism of Ca^{2+} response, similar to the affect seen with Gln163Lys mutation in CH-GECO3.1. Though there is fluorescence emission from the deprotonated chromophore the close proximity of Asp163 may reorganize the interface such that beneficial interactions between the chromophore and Asp163 is removed due to electrostatic clashes in the protein interior. As with all of these mutations without a crystal structure it is difficult to say which amino acids are involved, though the replacement of Gln163 with a negatively charged amino acid is detrimental to the Ca^{2+} response of the CH-GECO series sensors.

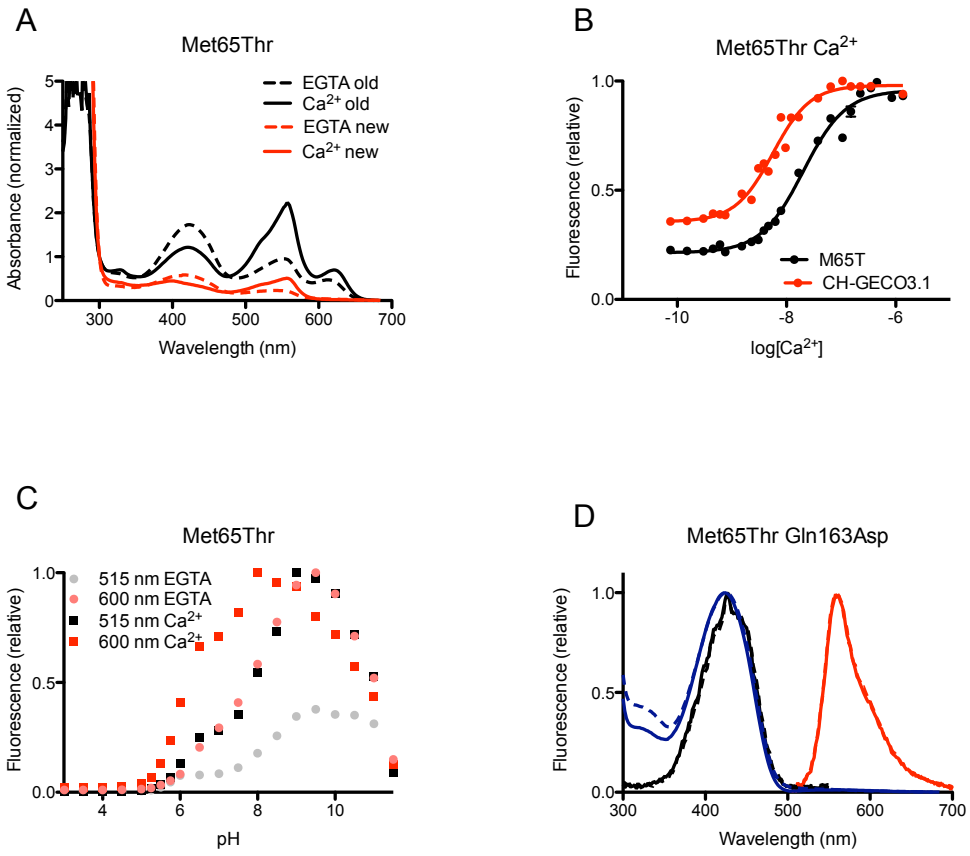


Figure 4.15 Characterization of orange fluorescent CHO-GECO1 and CHO-GECO1-Gln163Asp. A) Absorbance scans of CHO-GECO1 that was freshly purified (EGTA, dashed red; Ca^{2+}, solid red) and longterm storage (EGTA, dashed black line; Ca^{2+}, solid black line). B) Ca^{2+} titration of CHO-GECO1, monitoring orange emission. C) pH titration curves of CHO-GECO1 monitoring 570 nm emission with EGTA (grey circles) or Ca^{2+} (black squares), and 630 nm emission with EGTA (pink circles) or Ca^{2+} (red squares). D) Absorption and fluorescence excitation/emission of the CHO-GECO1-Gln163Asp mutant. Absorption with EGTA (dashed blue line) and Ca^{2+} (solid blue line) overlap. Excitation is at 430 nm (black line) and emission is at 570 nm (red line).

Since the mechanism of fluorescence increase does not depend on the pK_a of the chromophore the introduction of a non-titratable residue at position 66 could in theory still function since the quantum yield or extinction coefficient could still increase. With this idea in mind, the introduction of the Tyr66Phe mutation to generate a blue fluorescent protein (BFP) could yield a functioning protein sensor. Introduction of this mutation into CH-GECO3.1 yielded a protein with absorption

maxima of 402 nm and emission at 466 nm, but no response to Ca^{2+}(Figure 4.16A). However, these results are consistent with the published report on BFPs with alternate chromophore structures, in particular this mutation is consistent with the generation of mBlueberry1 (106). In Figure 4.16B the pH titration indicates there is a titratable group with a pK_a of 6.2, which cannot be the phenyl group of the chromophore since it has no available proton. Instead the same amino acid group that is being titrated in the pH curve of CH-GECO3.1 could be responsible for this titration curve. From this result it appears that a titratable group needs to be present on the ring system of the chromophore in order for the extinction coefficient and quantum yields to be increased. This may suggest that a Tyr66His mutation could result in a 'blueberry'-GECO that responds to Ca^{2+}. Despite the pK_a of CH-GECO3.1 remaining unaffected by Ca^{2+} binding, the presence of the titratable hydroxyl group plays an important role in the electron density of the chromophore such that removal of the hydroxyl group may disrupt H-bonds or salt bridges between the chromophore and surrounding residues.

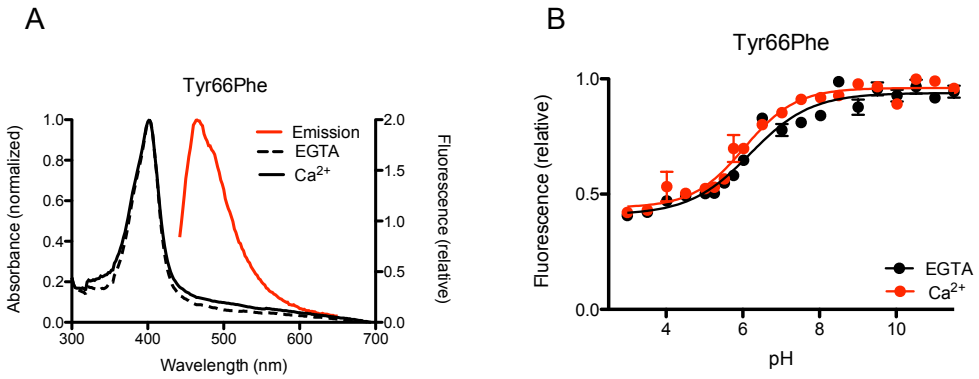


Figure 4.16 Spectroscopic characterization of Tyr66Phe mutation. A) Absorbance scans with EGTA (dashed line) or Ca^{2+} (solid line) B) pH titration in the presence of EGTA (black circles) or Ca^{2+} (red squares).

From the results of the additional mutations the adaptability of the chromophore and its surrounding environment is demonstrated. With a few mutations the CH-GECO3.1 was converted into a BFP, OFP (CHO-GECO1), LSS-OFP, and LSS-RFP, though unfortunately the Ca^{2+} response was not nearly as robust. Only the CHO-GECO1 variant retained Ca^{2+} response despite the protein retaining fluorescence for almost all of the single/multi-site mutations, with

the exception of some key chromophore maturation positions such as Lys70 and Asp59 along with Leu61 in CaM. It may be possible to re-engineer Ca^{2+} response with some of these variants. Alternatively, the LSS-mutation could be introduced into the R-GECO1 or O-GECO proteins as they operate under a different mechanism that may not be adversely affected by the introduction of a negatively charged residue at position 163.

4.2.6 Kinetic Tests

In addition to the steady-state Ca^{2+} affinities, another characteristic of Ca^{2+} sensors that must be taken into account before being used for imaging or *in* vivo applications is the kinetics of Ca^{2+} binding. If the Ca^{2+} binding kinetics are slower than the transient Ca^{2+} signal the measured signal will be time delayed or in the case of a rapid succession of transient signals, several may be missed. A Ca^{2+} imaging protocol has been published in the literature, which touches upon the importance of the kinetics (273, 274) and a modified version of this method was used for the mammalian cell imaging described in Section 4.2.7.

The Ca^{2+}-association kinetics were determined using an SX20 stopped-flow spectrometer (Applied Photophysics). The Ca^{2+}-sensor was diluted in 1 mM EGTA-MOPS buffer and mixed 1:1 with a buffered solution containing various amounts of Ca^{2+}. Injections at each Ca^{2+} concentration were repeated several times in order to obtain averaged K_{obs} measurements. The kinetic curves were fit using a one-phase association curve equation where the k_{obs} can be extracted. Plots of k_{obs} and $[Ca^{2+}]$ were fit using the equation:

$$\boldsymbol{k_{obs} = k_{on}[Ca^{2+}]^{n} + k_{off}}$$

Equation 4.1 Determination of kinetic constants

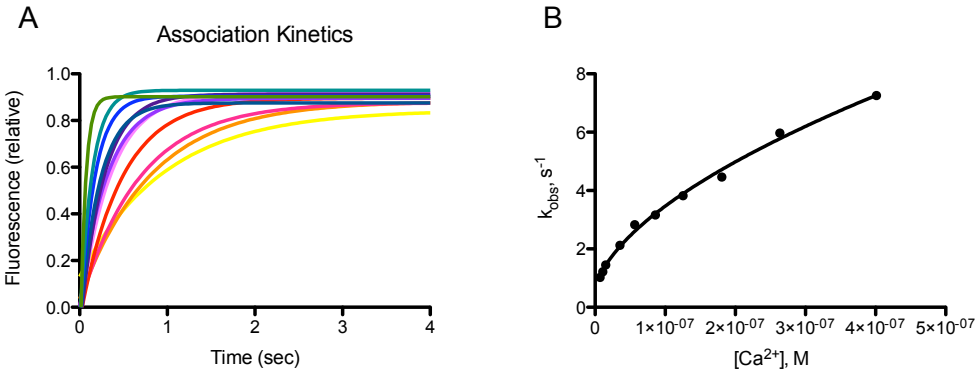


Figure 4.17 Stopped flow kinetic measurements of CH-GECO3.1. A) Raw association kinetic data fit to a single exponential curve. k_{obs} are determined for each Ca^{2+} concentration. B) Observed relaxation rate constants (k_{obs}) are plotted vs. Ca^{2+} concentration. The association and dissociation constants (k_{on} and k_{off}, respectively) were determine by fitting the data to the equations $k_{obs}=k_{on}[Ca^{2+}]^n + k_{off}$.

From the fitted data in Figure 4.17B the relevant kinetic parameters were extracted and are summarized below in Table 4.5. One of the striking differences between R-GECO1 and CH-GECO3.1 is value of *n*, which indicates the degree of cooperativity between Ca^{2+} binding sites in CaM. In CH-GECO3.1 *n* is less than 1, indicating that there is negative cooperativity between EF-hands, which is contrary to most reports on Ca^{2+} binding kinetics of CaM (254). This result is also at odds with the steady-state Ca^{2+} binding curves, which have a hill coefficient of 1.3, indicating a positive cooperativity. From the fit data the k_{on} value is several orders of magnitude smaller than the rest of the GECO series and could be explained by the large number of mutations found in CaM as compared to the rest of the GECO series (14). As mentioned in Section 4.2, there are mutations at the critical residues Asp21 and Asp23 in the first CaM EF-hand as well as the mutation Leu61Phe in the second EF-hand. These mutations could be sufficient to disrupt Ca^{2+} binding at one or both EF-hands. The disruption or weakened Ca^{2+} binding at the N-terminal EF-hands could be playing a role in the negative cooperativity as the mutations may make successful binding events in these two EF-hands less favourable. For the steady-state measurements the protein is given enough time to reach equilibrium between the Ca^{2+}-free and Ca^{2+}-bound states so the signal being measured is not reflective of initial binding, where the negative cooperativity takes place.

Table 4.5 Kinetic Characterization of CH-GECO3.1

Protein	k_{on} ($M^{-n}s^{-1}$)	k_{off} (s^{-1})	n	K_d',kinetic (nM)[1]	K_d',static (nM)[2]
CH-GECO3.1	3.70 x 10^5	0.423	0.58	3.4	6.0
R-GECO1[3]	9.52 X 10^9	0.752	1.6	484	482

[1] $K_d'_{,kinetic} = (k_{off}/k_{on})^{1/n}$, [2] From Figure 4.3B
[3] Values taken from (14).

Not shown in Figure 4.17A are the kinetic curves for Ca^{2+} concentrations between 270 pM and 4 nM. At these low concentrations instead of fluorescence association curves they looked like dissociation curves. It is possible that upon rapid mixing there is a dissociation of fluorescent protein dimers. There is no precedent for this in the literature at extremely low Ca^{2+} concentrations. However, the publication of the crystal structures of the RFP-based Ca^{2+} sensors Case12 and Case16 at "zero" Ca^{2+} gave a crystal with only the C-terminal lobes of CaM bound to Ca^{2+} (275). It is possible that at these low concentrations Ca^{2+} is bound in a fraction of the EF hands of CaM and these proteins form a domain-swapped structure, similar to the ones observed when GCaMP2 was crystallized (18, 19).

To explore the possibility that CH-GECO3.1 was forming dimers at high concentration, I examined the concentration dependence of the Ca^{2+}-dependent change in fluorescence. The results of the characterization are shown in Table 4.6. The data reveals that, at concentrations above 1-2 μM, there appears to be a suppression of fluorescence response. At these higher protein concentrations it is also possible that the solutions were experiencing fluorescence quenching, whereby the emission of one chromophore could be partially re-absorbed by a neighboring protein due to very close proximity.

Table 4.6 Ca^{2+} response tests at decreasing protein concentration

Dilution	Concentration (μM)	% Response
25	20.16	144
50	10.08	170
100	5.04	193
200	2.52	218
300	1.68	229
400	1.26	205
600	0.84	246
800	0.63	240

The most convenient way to test for the presence of protein dimers is to run a pseudo-native SDS-PAGE at several different concentrations in the presence of EGTA or Ca^{2+}. The SDS-PAGE is pseudo-native because the proteins are not boiled with SDS, but the prepared gels are the same composition as normal SDS gels, as described in Section 4.4.9. The fluorescent and white light images of both gels are shown in Figure 4.18 and contains two control proteins, dimer-Tomato (dTomato) and vivid Verde fluorescent protein (VFP). VFP was reported to be a dimer, but runs as a tetramer during size exclusion chromatography (276). From the white light images in Figure 4.18 there are several bands in the lanes containing the sensor. Faint bands at MW 100 kDa, which decrease in intensity as the protein concentration decreases, correspond to the size of sensor dimers. The large band at near 50 kDa is the intact sensor, while the bands located between MW 27-20kDa are partially proteolyzed proteins. The large broad bands around 55-60 kDa are not large enough to be sensor dimers, but could be a combination of cleaved CaM binding to M13 of the intact sensor. An alternate explanation is that the 55-60 kDa band is the dimerized CH-GECO3.1 migrating faster than what might be expected on the basis of size alone. From the fluorescent images the emission from the intact protein sensor is clearly visible, along with dTomato. Overall, there are no substantial differences between the EGTA or Ca^{2+} incubated protein gels. The most important conclusion from these gels is that there exists a higher molecular weight species that is non-fluorescent. While the exact nature of this species is unknown, the non-fluorescent band does appear to become more pronounced at higher protein concentrations. This could

partly explain the decrease in fluorescence response observed at high protein concentrations. The presence of non-fluorescent protein dimers could potentially be a confounding factor for imaging in mammalian cells since transient transfection often can cause cells to accumulate protein at concentration up to 10s of μM.

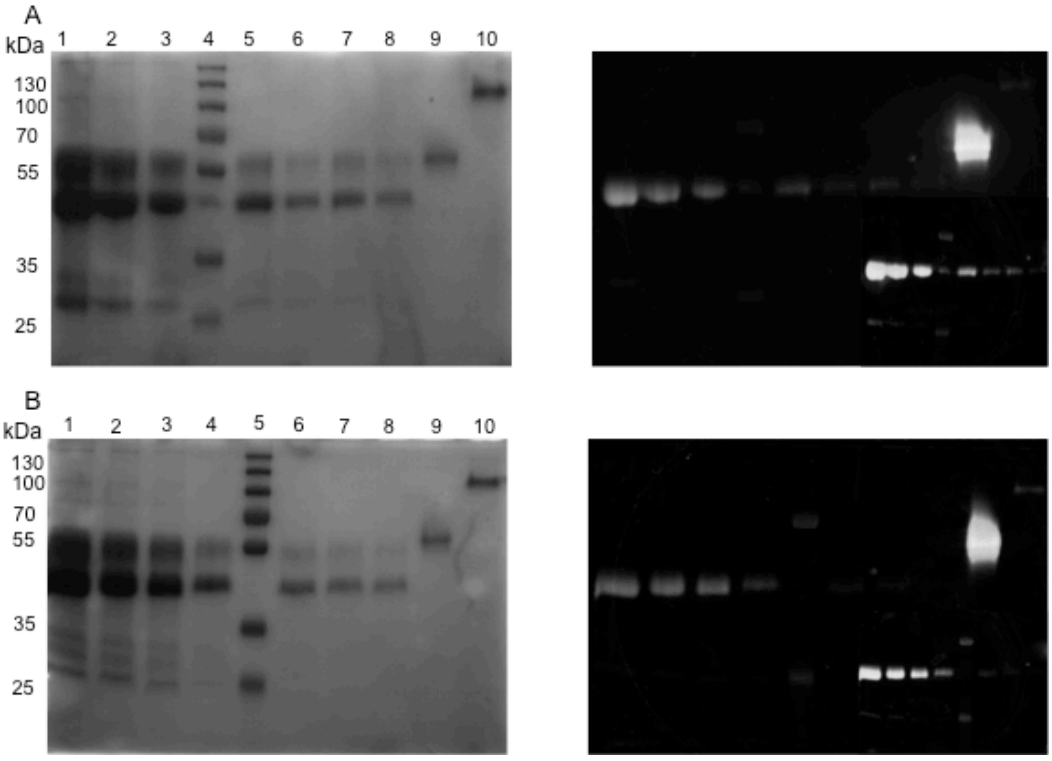


Figure 4.18 White light and fluorescent images of pseudo-native SDS-PAGE gels of CH-GECO3.1 incubated with EGTA or Ca^{2+}. A) Left panel is white light image of SDS-PAGE with EGTA. Large band around 49 kDa is the intact protein. Smaller bands around 25 kDa are partial cleavage products containing the RFP. Lane 9 is control protein dTomato and lane 10 contains the tetrameric VFP. Broad bands around 60 kDa could be partial dimers. Right panel is fluorescent images of the gels before coomassie staining. The inset image is higher exposure times clearly showing the fluorescent sensor, and fluorescent cleavage products, but no fluorescence from the dimers around 60 kDa. B) White light and fluorescent images of Ca^{2+} saturated protein gel showing very similar results as the EGTA gel.

4.2.7 Mammalian Cell Imaging

Imaging of Ca^{2+} indicators can be problematic for several reasons including Ca^{2+} buffering, interference from endogenous CaM, and protein dimerization or

aggregation (273, 274). Considering that CH-GECO3.1 has a much lower K_d than other commonly used indicators, there was a concern that Ca^{2+} signals in mammalian cells would not be detected. Both CH-GECO3.0 and CH-GECO3.1 were transfected in mammalian HeLa cells and a modified Ca^{2+} imaging protocol was used (273). Some of the time traces are shown below in Figure 4.19 and evident in both cases are the histamine-induced Ca^{2+} oscillations, as well as the influx of Ca^{2+} when ionomycin is added with EGTA or Ca^{2+}. Analysis of signal change for both proteins gave smaller fluorescence changes than that recorded *in vitro*. This was not unexpected as conditions in mammalian cells are much different than in a buffered solution. The lower magnitude of fluorescence change could also be due to the lower K_d value, which means that the protein will always largely be in the bound state even at resting Ca^{2+} concentrations. The cells picked for all imaging experiments were dimly fluorescent because cells that are very bright are likely over expressing the protein, which could pose several problems: the signal increases could saturate the detector, causing a lower signal change: the signal change could be diminished due to the formation of protein dimers; or high protein expression levels could buffer intracellular Ca^{2+} and disrupt the proper function of the cell.

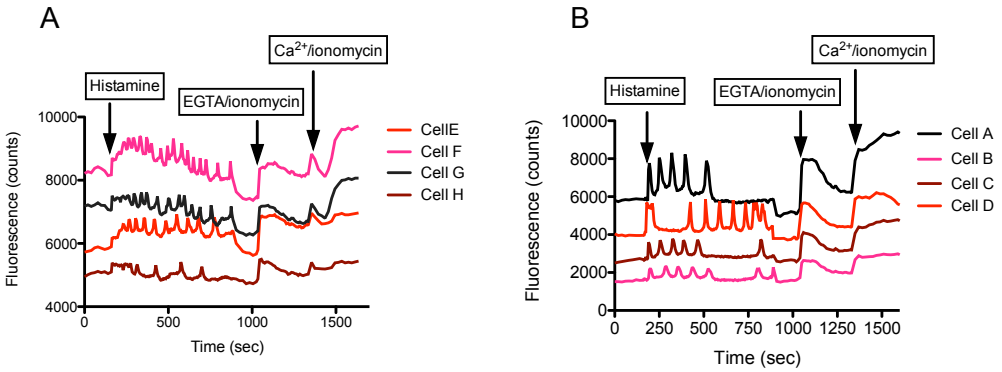


Figure 4.19 Hela cell traces of CH-GECO3.1 and CH-GECO3.0. A) Time trace of CH-GECO 3.1 indicating addition of histamine, EGTA/ionomycin, and Ca^{2+}/ionomycin. B) Time trace of CH-GECO3.0 indicating addition of histamine, EGTA/ionomycin, and Ca^{2+}/ionomycin.

Table 4.7 Characterization of Ca^{2+}-dependent fluorescence of CH-GECOs in HeLa cells. Cells were treated with histamine (His), then EGTA/ionomycin (EGTA), and finally Ca^{2+}/ionomycin (Ca^{2+})

Protein	n[1]	Maximum Ca^{2+} to Minimum EGTA ratio	Maximum His to Minimum His ratio	Maximum His to Maximum Ca^{2+}
CH-GECO3.1	9	1.13 ±0.02	1.07 ±0.02	0.89 ±0.02
CH-GECO3.0	13	1.40 ±0.05	1.37 ±0.05	0.92 ±0.07

[1] n is the number of unique HeLa cells imaged

In Table 4.7 the summarized results from the HeLa cells indicate that CH-GECO3.0 had better performance than CH-GECO3.1 in terms of signal change. This could be for a couple of reasons. As mentioned previously, CH-GECO3.1 has a lower K_d of 6 nM and despite the addition of EGTA/ionomycin there may still be some amount of protein bound to Ca^{2+}. Alternatively, the 5 mutations in CaM between CH-GECO3.0 and CH-GECO3.1 could cause CaM to interact with more endogenous binding partners in mammalian cells. Either way the results from this screening help to reinforce the fickleness of biosensor engineering. Characteristics screened for by *in vitro* assays may not always translate into better performance when used in a different system. As well, performance in one type of cell culture line does not always correlate with other systems. Extensive amounts of work have gone into improving the GCaMP series for use in neuronal experiments, while their performance in other systems is typically more than sufficient (20, 269, 274).

4.3 Conclusions

Several important properties of CH-GECO3.1 were discovered and characterized in this Chapter. The most interesting was the mechanism of action, which is completely distinct from all previously published single-FP Ca^{2+} sensors. Analysis of the pH titration curve indicated that the pK_a remains effectively constant, while both extinction coefficient and quantum yield are increased in the presence of Ca^{2+}. With this mechanism it should be possible to predict a theoretical maximum signal increase as we would expect the ideal sensor to match the quantum yield and extinction coefficient of the parent RFP, mCherry.

Whether these improvements can be realized remains to be seen. Nonetheless CH-GECO3.1 will be useful for scientists not only for its use in fluorescence imaging, but perhaps even more so for its insights into RFP chromophore maturation studies. Attempts to discern the amino acid responsible for the increased quantum yield and extinction coefficient led to a likely candidate, Gln163. Mutation of this residue to Lys (as found in R-GECO1), abolished any Ca^{2+} response, while mutation to other amino acids such as Met or Asp produced RFPs with unique spectral properties.

An investigation of the linker regions reinforced the importance of optimizing linkers when designing FP biosensors. The initial design of the CH-GECO contained Gly-Gly linkers in between the Ca^{2+} binding domains and the N-terminal residues were mutated to Ser-Leu. Interestingly when the N-terminal linker was converted to the analogous region from R-GECO1, there was a significant portion of protein trapped in the blue intermediate. In contrast, adding the C-terminal R-GECO1 linker abolished any expression of red fluorescent protein. The choice to include linkers and their composition should be very carefully considered and continually optimized, as this portion of the construct could become a crucial portion of the protein.

Characterization of the kinetics and behaviour in mammalian cells gave somewhat unexpected results. Despite the low K_d, CH-GECO3.1 was able to detect histamine-induced Ca^{2+} oscillations in HeLa cells, though the dynamic range was decreased several fold as compared with the *in vitro* characterization. CH-GECO3.0 actually out performed CH-GECO3.1 in mammalian cells with a higher dynamic range, which was better matched to the *in vitro* experiments. This result serves as a reminder that the screening methods used when evolving FP-based biosensors is very important and positive results from one screening system may not necessarily equate to improvements when using it for another application.

Finally, the kinetics of CH-GECO3.1 were measured using a stop-flow instrument. These studies revealed a kinetic K_d of 3 nM which closely matched the measured steady-state K_d of 6 nM. However, these studies also revealed that the protein appears to bind Ca^{2+} with negative cooperativity. This unprecedented

behaviour of CaM/M13 behaviour could be due to the several mutations found in CaM as compared with R-GECO1 and previous CH-GECO versions.

Future work with CH-GECO3.1 could go down several avenues. A crystal structure would offer more conclusive results as to the key amino acids that are interacting with the chromophore. This protein is far from having optimal performance *in vitro* and in cell imaging experiments, but more mutagenesis could be done to see if the Ca^{2+} bound quantum yield and extinction coefficient can be engineered to more closely match that of mCherry. Expression of the protein in *E. coli* is not efficient at 37 °C so an added bonus of continuing engineering would be to improve protein folding and maturation at physiological relevant temperatures. Ultimately, CH-GECO3.1 is only the third reported RFP-based Ca^{2+} sensor, and is distinguished by the unique mechanism of its Ca^{2+}-dependent fluorescence increase.

4.4 Materials and Methods

4.4.1 General Materials and Methods

The plasmid DNA from the bacteria was purified using a chloroform extraction protocol. Briefly, 150 μL of solution I (50 mM Tris, 10 mM EDTA, 100 μg/mL RNaseA, pH 8.0) is used to resuspend the bacterial pellet. 150 μL of solution II (1% SDS, 0.2 M NaOH) is added and the mixture is gently inverted several times. Solution III (2 M acetic acid, 3 M KOAc, pH 5.5) was added to a total volume of 450 μL and mixed to pellet the non-soluble cell debris. Finally, 150 μL of chloroform is added and mixed several times before being centrifuged at 14000 rpm, 4°C for 5 minutes. The top aqueous layer is aliquotted into a separate 1.5 mL Eppendorf tube and 800 μL of 100% ethanol is added. The solution is then centrifuged again at 4°C for 5 minutes. The DNA should form a small visible pellet at the bottom of the tube and the residual ethanol is removed. Finally 500 μL of 70% ethanol is added to the tube and the contents are spun at 14000 rpm for 1 minute. Afterwards the ethanol is removed and the contents are allowed to air-dry.

Sequencing reactions were run with either oligo 2.1 or 2.2 to generate the forward or reverse sequence. All sequencing reactions were cleaned up using the sodium acetate/ethanol protocol and analyzed by MBSU.

4.4.2 Quikchange Mutagenesis

Oligos 4.1 through 4.42 were designed using the Agilent Technologies Quikchange primer design software. Single-site and multi-site Quikchange mutagenesis were run with CH-GECO3.1 as the template. The protocol was the same as described above in section 3.2.9.

4.4.3 Protein Expression and Purifications

A single colony of the desired mutant was picked and grown overnight in a 5 mL LB/amp culture tube (37 °C). The 5 mL culture tube was used to inoculate 500 mL of LB/amp. After 4 hours of shaking at 37 °C arabinose was added to a final concentration of 0.02% and continued shaking overnight. For poorly folding mutants the flasks were transferred to a shaker at 30 °C for two nights after induction.

To increase the amount of expressed protein a modified TB (Terrific broth) growth media was used. The recipe for 1L is as follows: 20 g LB; 14 g tryptone; 7 g yeast; 9.4 g K_2HPO_4; 2.2 g $KHPO_4$; 0.8% w/v glycerol. A 5mL LB/amp culture tube is used to inoculate 250 mL of TB. After allowing the OD to reach 0.6 protein expression was induced with 0.004% arabinose and grown over night at 37 °C or for two nights at 30 °C, depending on the brightness of the protein construct. Once ready for purification the bacteria were pelleted at 10000 rpm, 4 °C for 10 minutes. The pellet is resuspended in 4°C 10 mM Tris-Cl, 150 mM NaCl pH 7.4 and lysed using a cell disruptor (Constant Systems Ltd.). The lysed product was pelleted at 14000 rpm and the supernatant was shaken with Ni-NTA beads on ice. The Ni-NTA beads were collected on a column with a vacuum manifold and washed twice with 10 mM Tris-Cl, 30 mM imidazole, 150 mM NaCl pH 8.0. The beads were gravity washed once and then eluted with 300 mM imidazole, 10 mM Tris-Cl pH 8.0. The excess imidazole was removed via buffer exchange with Amicon columns (MWCO 10000) and 10 mM Tris-Cl, 150 mM NaCl pH 7.3.

4.4.4 pH Titrations

To determine the pK_a of the proteins I prepared a series of pH buffers as follows. A solution of 30 mM trisodium citrate and 30 mM sodium borate was adjusted to pH 11.5. The pH of the solute on was then adjusted with HCl (12 M and 1 M) in 0.5 pH units and 10-15 mL was collected at each desired pH value. A few extra pH values were collected at pH 5.25 and 5.75. 5 μL of protein was mixed with 100 μL of either a Ca^{2+}-free buffer (30 mM MOPS, 100 mM KCl, 10 mM EGTA, pH 7.2) or Ca^{2+}-saturated buffer (30 mM MOPS, 100 mM KCl, 10 mM CaEGTA, pH 7.2). 5 μL of the protein solution was mixed with 50 μL of the desired pH buffer and aliquotted in triplicate into a 396-well plate. Fluorescence emission was recorded using the Tecan microplate reader.

4.4.5 Ca^{2+} Titrations

The apparent K_d was determined by mixing the appropriate protein with buffers containing various amount of Ca^{2+}. The buffers were prepared following the recipe from the Invitrogen Ca^{2+} calibration buffer kits. The Ca^{2+}-free buffer (30 mM MOPS, 100 mM KCl, 10 mM EGTA, pH 7.2) and Ca^{2+}-saturated buffer (30 mM MOPS, 100 mM KCl, 10 mM CaEGTA, pH 7.2) were mixed in different ratios to generate buffers with Ca^{2+} concentrations ranging from zero Ca^{2+} to 39 μM Ca^{2+}. Similar to the pH titrations 5-10 μL of the protein was mixed with 150-200 μL of each Ca^{2+} buffer. 50 μL of each solution was aliquotted in triplicate into a 396-well plate and the fluorescence emission was recorded using the plate reader. Each emission peak was integrated and plotted against the log of the calculated free Ca^{2+} concentration. The concentration of free Ca^{2+} was calculated using the following equation:

$$[Ca^{2+}]_{free} = K_d^{EGTA} \frac{[CaEGTA]}{[K_2EGTA]}$$

Equation 4.2 Determination of free-Ca^{2+} where K_d is the K_d of EGTA for Ca^{2+} and the last term is the ratio of CaEGTA to K_2EGTA. The K_d of EGTA at pH 7.2 and 20°C is 150.5 nM.

Using the software in Prism each Ca^{2+} titration curve was fit with a sigmoidal curve and from the data the K_d and the Hill coefficient could be obtained.

4.4.6 Determining the quantum yield and extinction coefficient

Similar to the determination of the extinction coefficients of the cpRFPs, mCherry was used as the control protein for quantum yield and extinction coefficient measurements. A BCA assay (Pierce) was used to determine the concentration of both proteins according the manufacturers protocol. Solutions of mCherry and CH-GECO3.1 were diluted to absorbance values of approximately 0.25, 0.4, and 0.60 and prepared in quadruplet. Extinction coefficients were determined in two ways using Beer's Law:

$$A = \varepsilon bc$$

Equation 4.3 Beer's Law where A is the absorbance; ε is extinction coefficient ($M^{-1}cm^{-1}$); b is path length (cm); c is concentration (M)

Using the concentrations obtained from the BCA assay the extinction coefficients were calculated. The other method of determination is the same one used in section 2.2.7 and 3.2.6 and was used to determine the extinction coefficient of CH-GECO3.0. The ensemble ε of the proteins was determined by comparing the intensity of the 587 nm absorbance peak for each protein to that of a solution of mCherry with matched absorbance at 280 nm. The ε was calculated by multiplying the ratio of absorbance intensities at 587 nm by the ε of mCherry (72 000 M cm^{-1}).

The quantum yield of CH-GECO3.1 in the presence and absence of Ca^{2+} was determined in a similar fashion described elsewhere in the literature. Briefly, the concentration of protein in a buffered solution (30 mM MOPS, pH 7.2, with either 10 mM EGTA or 10 mM Ca-EGTA) was adjusted such that the absorbance at the excitation wavelength was between 0.2 and 0.6. A series of dilutions of each protein solution and standard, with absorbance values ranging from 0.01 to 0.05, was prepared. The fluorescence spectra of each dilution of each standard and protein solution were recorded and the total fluorescence intensities obtained by integration. Integrated fluorescence intensity vs. absorbance was plotted for each protein and each standard. Quantum yields were determined from the slopes (S) of each line using the equation:

$$\Phi_{protein} = \Phi_{standard}\left(\frac{S_{protein}}{S_{standard}}\right)$$

Equation 4.4 Determination of quantum yield of the protein

Using the same protocol as described in sections 2.2.7 and 3.2.6 the quantum yield of CH-GECO3.0 was determined.

4.4.7 Kinetic Measurements

The Ca^{2+} association kinetics of CH-GECO3.1 were determined using a SX20 stopped-flow spectrometer (Applied Photophysics). The Ca^{2+} sensor diluted 400X in buffer (30 mM MOPS, 100 mM KCl, 1 mM EGTA, pH 7.2) was mixed (1:1) with a series of Ca^{2+} buffers that were prepared by mixing a buffered solution (30mM MOPS, 100 mM KCl) with different ratios of 10 mM EGTA and 10 mM CaEGTA. The change in fluorescence signal as a result of the rapid mixing was used to determine the relaxation coefficient (k_{obs}) for the Ca^{2+} association reactions at various Ca^{2+} concentrations (700 pM to 1300 nM) by fitting the curve to a one-phase decay. The k_{obs} was then fit to the equation:

$$k_{obs} = k_{on}[Ca^{2+}]^n + k_{off}$$

Equation 4.5 Determination of k_{on} and k_{off}

A plot of $[Ca^{2+}]$ versus k_{obs} would yield a slope equal to k_{on} and y-intercept of k_{off}.

4.4.8 Mammalian Cell Imaging

Both CH-GECO3.0 and CH-GECO3.1 were cloned into the mammalian expression plasmid PCDNA3.1 (+). Using oligos 4.44 and 4.45 *BamHI* and *EcoRI* sites were introduced at 5' and 3' ends, respectively. The DNA was doubly digested with Fermentas fast digest enzymes and ligated into the PCDNA3.1 plasmid, which was also digested with the same enzymes. The ligation products were electroporated into *E. coli* DH10B cells and plated onto LB/agar/amp dishes. Colonies from the plate were picked up and grown overnight in 5 mL LB/amp culture tubes. Plasmid minipreps were done using the GeneJet miniprep kit (Fermentas) according to the manufacturers protocol. All of the plasmid DNA was

sequenced with oligos 4.45 (forward) and 4.46 (reverse) and the products were run by MBSU.

HeLa cells were maintained in Dulbecco's Modified Eagle Medium (DMEM) supplemented with 10% fetal bovine serum (FBS) and glutamax (Invitrogen) at 37 °C and 5% CO_2. To prepare cells for imaging a small amount of cell culture is seeded onto homemade glass-bottom petri dishes and incubated overnight at 37 °C, 5% CO_2. The dishes were checked under a light microscope to ensure that the cells were 70-90% confluent before transfection. Transfection solution was prepared with Turbofect (Invitrogen) according to the manufacturer's protocol. Briefly, 100 μL of serum-free media was mixed with 2 μL of Turbofect and 750 ng of plasmid DNA. The mixture was allowed to sit for 20 minutes before being added to the petri dishes that contained 1 mL of serum-free media. After dropwise addition of the transfection reagent the dishes were left to incubate at 37 °C for 2 hours. After two hours the media was removed and replaced with complete medium (FBS, glutamax, pen/strep) overnight. On the day of imaging the complete medium was removed and the dishes were washed twice with PBS and left with 1 mL of hepes-buffered hanks balanced saline solution (HHBSS).

An inverted Nikon Eclipse Ti microscope equipped with a 200W metal halide lamp (PRIOR Lumen), 60x oil objective, and a QuantEM; 512SC 16-bit cooled CCD camera (Photometrics) was used for fluorescence microscopy. The protocol for Ca^{2+} imaging was an adapted version from the literature (273). Briefly, the rinsed dishes containing the transfected cells are imaged using 300 ms exposure every 30 seconds for 2.5 minutes to measure baseline fluorescence, after which 1 mL of 5 mM histamine is added and images are captured every 5 seconds for 12 minutes. After the 12 minutes the dish is rinsed three times with HHBSS and images are taken every 30 seconds for 1.5 minutes to reestablish a baseline signal, following baseline measurements the cells are washed three times with Ca^{2+}-free HHBSS and 1 ml of 1 mM EGTA/5 μM ionomycin is added and images are captured every 10 seconds for 5 minutes. Afterwards the cells are rinsed three times again with HHBSS before 1 ml of 2 mM Ca^{2+}/5 μM ionomycin is added and images taken every 30 seconds for 2.5 minutes.

4.4.9 Dimerization

A pseudo-native sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) (10%) was prepared as described elsewhere (277). Several dilutions of the protein were prepared with either 10 mM EGTA or 10 mM CaEGTA. Dimer tomato and VFP, a green tetrameric protein, were run as controls. The gels were imaged using our colony screener for red and green fluorescence before being stained with Coomassie.

The fluorescence emission of the protein in the presence of EGTA or Ca^{2+} was measured at several dilutions (25x, 50x, 100x, 200x, 300x, 400x, 600x, 800x) using the microplate reader. The percent fluorescence increase was determined by comparing the integrated fluorescence emission with Ca^{2+} to EGTA.

Chapter 5 Splicing-based reconstitution of red fluorescent Ca^{2+} indicators

5.1 Introduction

As mentioned in Section 1.3.6.2, inteins are proteins that excise themselves from in between flanking exteins by re-forming a peptide bond between the N-terminal extein and C-terminal extein (Figure 5.1). The first known intein was discovered over 20 years ago in *Saccharomyces cerevisae* (162), and was initially referred to as a 'spacer domain' until the terms intein and extein were coined by Perler et al. (163). The first naturally split intein, wherein separate N-terminal and C-terminal intein fragments associate and splice together the N-extein and C-extein (175, 176) was identified in the cyanobacterium *Synechtosis sp*. Currently there are over 400 reported inteins from all three branches of the tree of life (164). Inteins are divided into three classes depending on their mechanism and conserved amino acid sequence. Only the Class 1 inteins will be discussed here in greater detail as the intein used for the work in this Chapter is from Class 1.

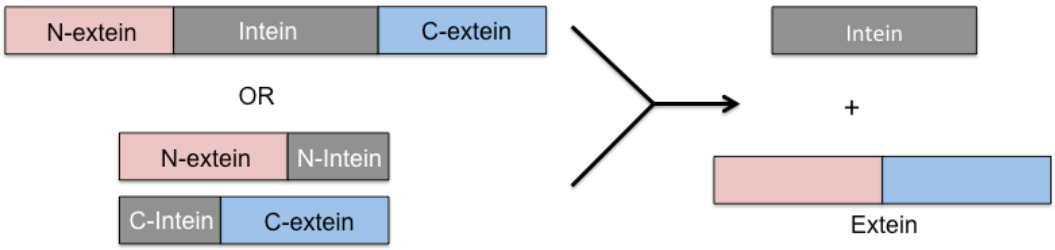


Figure 5.1 General schematic of the splicing process from either an intact intein or split intein.

The protein splicing of Class 1 inteins occurs via a four-step mechanism, which was described in detail in Section 1.3.6.2. To briefly reiterate, the first step is an N-O or N-S acyl shift if the N-terminal initiating amino acid is Ser or Cys, respectively. The (thio)ester bond is then attacked by the OH or SH group located at the first residue in the C-extein. The third step involves the cyclization of a conserved Asn residue, which releases the intein and connects the two

exteins by a (thio)ester bond. Finally, rearrangement of the (thio)ester bond forms the new peptide bond.

The intein used in this work comes from the gene *dnaE* of *Nostoc punctiforme* (*Npu*), a cyanobacteria. Protein splicing by this naturally split intein has been shown to be very robust and efficient (176, 278) and makes a good choice for studying or exploiting spliced proteins in a variety of cellular settings. Dr. Kevin Truong (University of Toronto) generously provided several *dnaE*-based intein constructs for us to work with. He has reported using this intein for the reconstitution of TN-XL and GCaMP2 (190), and we wanted to undertake a similar effort with GECO-series Ca^{2+} sensors.

Our initial plan was to generate split R-GECO1 and CH-GECO3.1 to determine if these proteins could be spliced together and regenerate the intact and fully functional Ca^{2+} indicator. Our goal was to demonstrate that such reconstitution could be performed in mammalian cell culture. If this goal could be achieved, we would send the constructs to the lab of Hiroshi Suzuki (University of Toronto), for application in *C. elegans*. In this Chapter the results of the *in vitro* splicing and mammalian cell imaging experiments will be summarized.

5.2 Results and Discussion

5.2.1 *In vitro* splicing

We decided to clone each intein construct into pBAD/His B (Life Technologies) for ease of DNA manipulation and protein expression. The exact oligos and cloning strategy for each intein construct is summarized in Table 5.2. There are several different plasmids discussed so abbreviations were generated for each series (Figure 5.2). Each Ca^{2+} sensor contains its name followed by and 'N' or 'C' to dictate if it is attached to the N-intein or C-intein, respectively, then the plasmid name is abbreviated for pTriex3Hygro (pT), pBAD/His B (pB), or pCDNA3 (pC). For example, CH-GECO3.1 cloned into pBAD attached to the C-intein is abbreviated CH-GECO3.1CpB. For the FPs excised from the Ca^{2+} biosensors, the R-GECO1 RFP is designated as mApple, CH-GECO3.1 is designated as cp145-3.1, and the RFP from 196V1.2-GECO in Chapter 3 is abbreviated cp196V1.2.

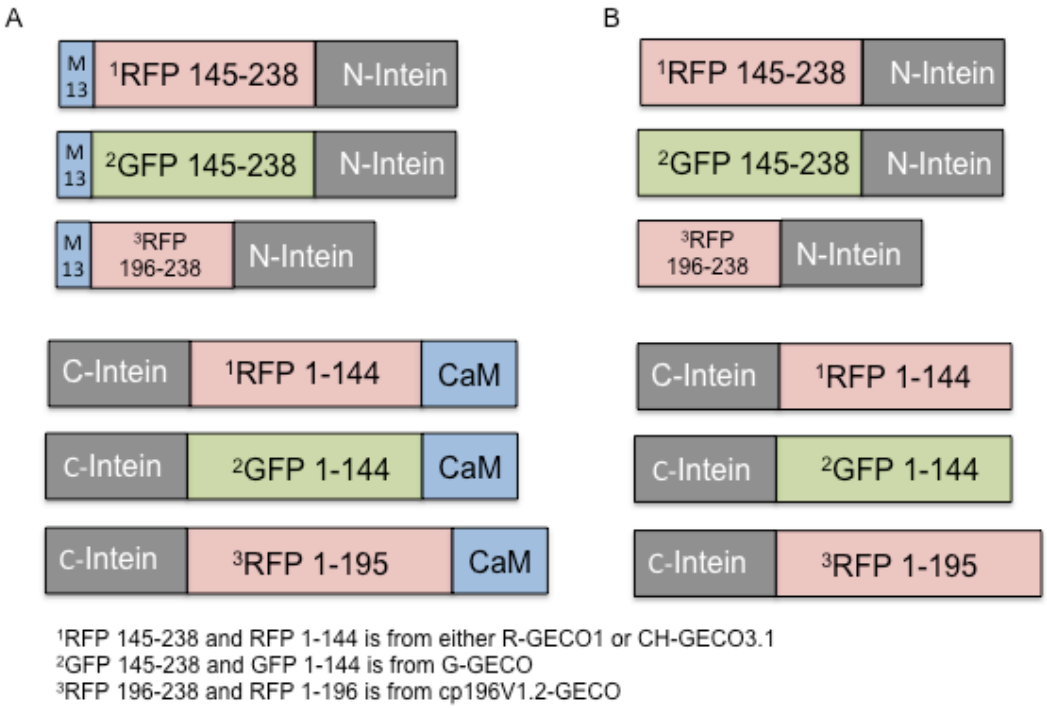


Figure 5.2 General schematic of intein constructs for *in vitro* splicing tests. A) Ca^{2+} sensor constructs. RFP 1-144 and RFP 145-238 fragments come from R-GECO1 and CH-GECO3.1. GFP 1-144 and GFP 145-238 are from G-GECO. RFP 1-195 and RFP 196-238 are from cp196V1.2-GECO. B) RFP and GFP protein intein fragments.

Once the protein fragments were successfully cloned into pBAD/His B and their identity confirmed by sequencing, attempts were made to express and purify each protein fragment from *E. coli*. Below are the images of the SDS-PAGE gels run after purification (two different protocols) as described in Methods and Materials (Figure 5.2A and B). For mAppleCpB, cp145-3.1CpB, mApple-NpB, cp145-3.1NpB, and cp145-3.1CpB the expected band size should be 26 kDa. The cp196V2.1NpB and cp196V1.2CpB should be 20 kDa and 31 kDa, respectively. R-GECO1NpB and CH-GECO3.1-NpB proteins should be 29 kDa, while R-GECO1-CpB and CH-GECO3.1-CpB should be 42 kDa. Finally, the 196V1.2-GECO-NpB and 196V1.2-GECO-CpB would be expected to show up at 23 kDa and 48 kDa, respectively.

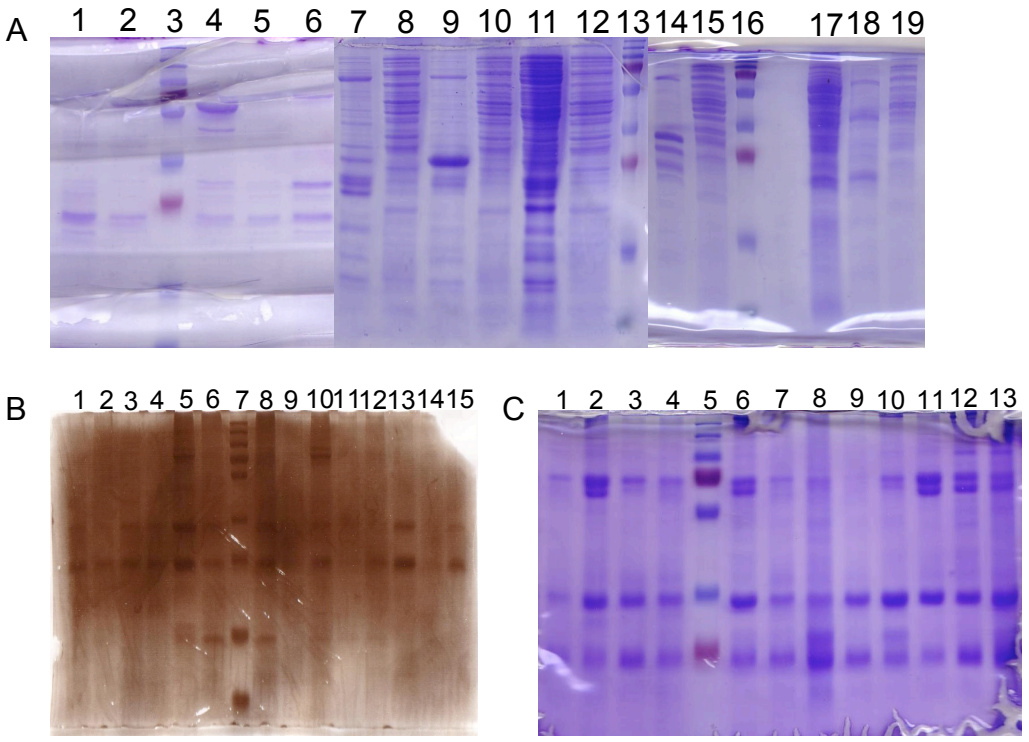


Figure 5.3 SDS-PAGE gels of intein constructs. Details of samples loaded in each lane are provided in Table 5.1. A) Three SDS-PAGE gels of purified intein proteins expressed overnight at 37°C. B) and C) Two SDS-PAGE gels of intein proteins expressed for 4-5 hours at 37°C. The left panel is a grey-scale image of the gel.

Table 5.1 Summary Table of SDS-PAGE Protein Bands

Lane	Panel	Protein	Expected Size (kDa)	Band Size(s) (kDa)
1	A	mAppleC	25.8	22, 25
2	A	R-GECO1C	42.5	25
3	A	Ladder	-	-
4	A	mAppleN	25.6	25, 31, 47, 63
5	A	R-GECO1N	29	25, 28, 67
6	A	CH-GECO3.1N	29	22, 25.3, 28, 31, 67
7	A	196V1.2-GECON	23	22, 25, 64
8	A	196V1.2-GECON (whole cell lysate)	23	-
9	A	cp145-3.1N	26	24.6, 27
10	A	cp145-3.1N (whole cell lysate)	26	30

11	A	cp196V1.2N	20	21
12	A	cp196V1.2N (whole cell lysate)	20	21
13	A	Ladder	-	-
14	A	cp196V1.2C	31	21, 26, 29, 32
15	A	cp196V1.2C (whole cell lysate)	31	29, 32
16	A	Ladder	-	-
17	A	mAppleC	25.8	23, 26, 29
18	A	CH-GECO3.1C	42.5	21, 42, 65
19	A	CH-GECO3.1C (whole cell lysate)	42.5	-
1	B	CH-GECO3.1N	29	26, 34
2	B	CH-GECO3.1C	42.5	26, 34
3	B	R-GECO1N	29	26, 34
4	B	R-GECO1C	46	26, 34
5	B	G-GECON	29	26, 34, ~85
6	B	G-GECOC	42.5	26, 34
7	B	Ladder	-	-
8	B	196V1.2-GECON	23	26, 34
9	B	196V1.2-GECOC	48	26, 34
10	B	cp145-3.1N	26	26, 34, ~85
11	B	cp145-3.1C	26	34
12	B	mAppleN	25.6	26, 34
13	B	mAppleC	25.8	26, 34
14	B	cp196V1.2N	20	-
15	B	cp196V1.2C	31	26, 34
1	C	R-GECO1C	42.5	26, 34, 67
2	C	mAppleN	25.6	26, 34, 62, 67
3	C	mAppleC	25.8	26, 34, 62, 67
4	C	CH-GECO3.1N	29	26, 34, 67
5	C	Ladder	-	-
6	C	CH-GECO3.1C	42.5	26, 34, 62, 67
7	C	cp145-3.1N	26	26, 34, 67
8	C	cp145-3.1C	26	26, 29, 34, 62, 67
9	C	G-GECON	29	26, 34, 67
10	C	G-GECOC	42.5	26, 29, 34, 62, 67
11	C	196V1.2-GECON	23	26, 34, 62, 67
12	C	196V1.2-GECOC	48	26, 34, 62, 67
13	C	cp196V1.2N	20	26, 34, 62, 67

The summary of the prominent bands in each lane of the SDS-PAGE gels shown in Figure 5.3 are found in Table 5.1. Overall, it appears that overnight expression of the inteins resulted in more proteins at the correct size (Figure 5.3A), as compared with the proteins that were purified after only 4 to 5 hours of expression (Figure 5.3B and C). In Figure 5.3B and C there are several prominent bands for all protein solutions at 26, 34, 62, and 67 kDa. It is unlikely that these bands are non-specifically bound proteins as all Ni-NTA purification beads were rinsed multiple times with wash buffer. The expected band size for the N-intein fragments as mentioned above were at 20 kDa, 23 kDa, 26 kDa, and 29 kDa. The C-intein fragments should occur at 26 kDa, 31 kDa, 29 kDa, 42 kDa, and 48 kDa. It is possible that the bands at 26 kDa are the mAppleNpB, mAppleCpB, cp145-3.1NpB, and cp145-3.1CpB proteins, and the presence of these bands in other lanes may be degradation products that form during cell lysis. Degradation by proteolysis of C-intein fragments has been described previously in the literature (278). Intein fragments were successfully purified, but intact extein proteins were used, rather than split-FPs. The remainder of the bands are either smaller than expected (34 kDa) or larger than expected (62 and 67 kDa), which suggest that the expressed proteins are folding improperly and aggregating or perhaps degrading due to proteolysis. It is also likely that improperly folded protein is being sent to inclusion bodies and purification would necessitate lengthy denaturation and renaturation protocols.

Very few FPs will fold properly after being split and the few that do are the result of extensive engineering to improve folding (104, 219). Ottman et al. described the extensive efforts they underwent to generate a superfolder-YFP that would tolerate splitting at position 158 for *in vitro* bi-molecular fluorescence complementation (BiFC) (219). Over 15 mutations were introduced to obtain protein halves that were still soluble. A split mCherry was successfully generated at position 159, which is located in the beginning of the β8 strand, but all experiments were completed in Vero cells where both halves were co-transfected into the same cells to monitor complementation of p53 (human tumor protein) and LTag (large T-antigen) (279). In comparison with the published split mCherry, CH-GECO3.1 and R-GECO1 are split in β7 near some crucial chromophore forming residues (245, 262). Generally speaking, publications describing split

fluorescent proteins do not report the purification of the fragments and *in vitro* reconstitution, because of poor folding efficiency and solubility when expressed in *E. coli* (203, 207). Most publications discussing the purification of split inteins use dual expression plasmids where protein expression of both fragments is induced at the same time so that upon expression the properly spliced protein is present (177, 280).

Protein splicing was tested by combining the individually purified proteins *in vitro*. Despite the lack of the large C-intein bands identifying each fragment, especially in lanes 2A, 2B, 4B, 6B, 9B, 1C, 6C, 10C, and 12C of Figure 5.3, the BCA assay was used to determine the concentration of each purified protein solution. Each N-intein protein was then mixed with each C-intein protein at equal concentrations of 14 μM in a 396-well microplate reader and red fluorescence was measured over several hours. Included as controls was each individual intein protein. We did consider cloning all potential N-intein and C-intein pairs into a single dual expression vectors. However, this strategy was not pursued due to the sheer volume of cloning that would have been required. With 6 different N-intein fragments and 6 different C-intein fragments a total of 36 unique plasmids would need to be constructed, which was not considered to be an efficient use of time.

The proteins were incubated at room temperature in wells of a 96-well plate in a plate reader, though the recorded temperature was around 28-29 °C. Figure 5.4 depicts the splicing data as time vs. fluorescence counts. Included as controls were the individual N- and C-inteins. For most cases no increase in fluorescence was detected with the exception of a few combinations. For CH-GECO3.1, the fluorescence from the C-intein fragment increases over time, though this result was not seen for any other combinations of CH-GECO3.1C (Figure 5.4A). A few other combinations that had a fluorescence increase were 196V1.2-GECON/196V1.2-GECOC (Figure 5.4D, black triangle), 196V1.2-GECON/cp196V1.2C (Figure 5.4D, open circle), cp145-3.1N/196V1.2-GECOC (Figure 5.4E, black triangle), cp145-3.1N/cp196V1.2C (Figure 5.4E, open circle), and mAppleN/196V1.2-GECOC (Figure 5.4F, black triangle). The background signal of 196V1.2-GECOC was consistently higher than the rest, but for splicing

mixtures where fluorescence increased the recorded signal was larger than 196V1.2-GECOC alone, indicating it was not just background signal.

Overall, the results of the in vitro protein splicing assay were relatively disappointing. Specifically, for those combinations that did undergo splicing, the observed fluorescence changes were not as large as expected. The reason for the very poor splicing efficiency can be attributed to a few things, namely poor protein fragment solubility resulting in very low levels of purified protein; disrupted/partially folded proteins such that even after splicing the original FP is not reconstituted; or interference of the C-intein His tag. Zhao et al. found that expression levels of intact GFP*mut3** cloned between inteins was variable depending on temperature and induction time (188). It is also possible that the His-tag downstream of the C-intein could interfere with the correct splicing. These constructs are cloned into the pBAD/His B plasmid, which all contain a 5' 6-His-tag attached to the N-terminus for protein purification purposes. Although work has been done to test the robustness of the extein sequences, most *in vitro* splicing tests contain a C-terminal His-Tag on the C-intein fragment (179, 180). It has also been reported that more efficient splicing occurs when the purified proteins are denatured then renatured in the presence of each other in order to promote correct protein folding and splicing (278, 281, 282). The results of the splicing tests were not very encouraging, but, considering that most of the literature on inteins describes their use in live mammalian cells, we decided to test all the same combinations in mammalian cell culture. Also, the published results of Wong et al. were encouraging, prompting us to test RFP and GFP-based Ca^{2+} sensors in a variety of combinations (190).

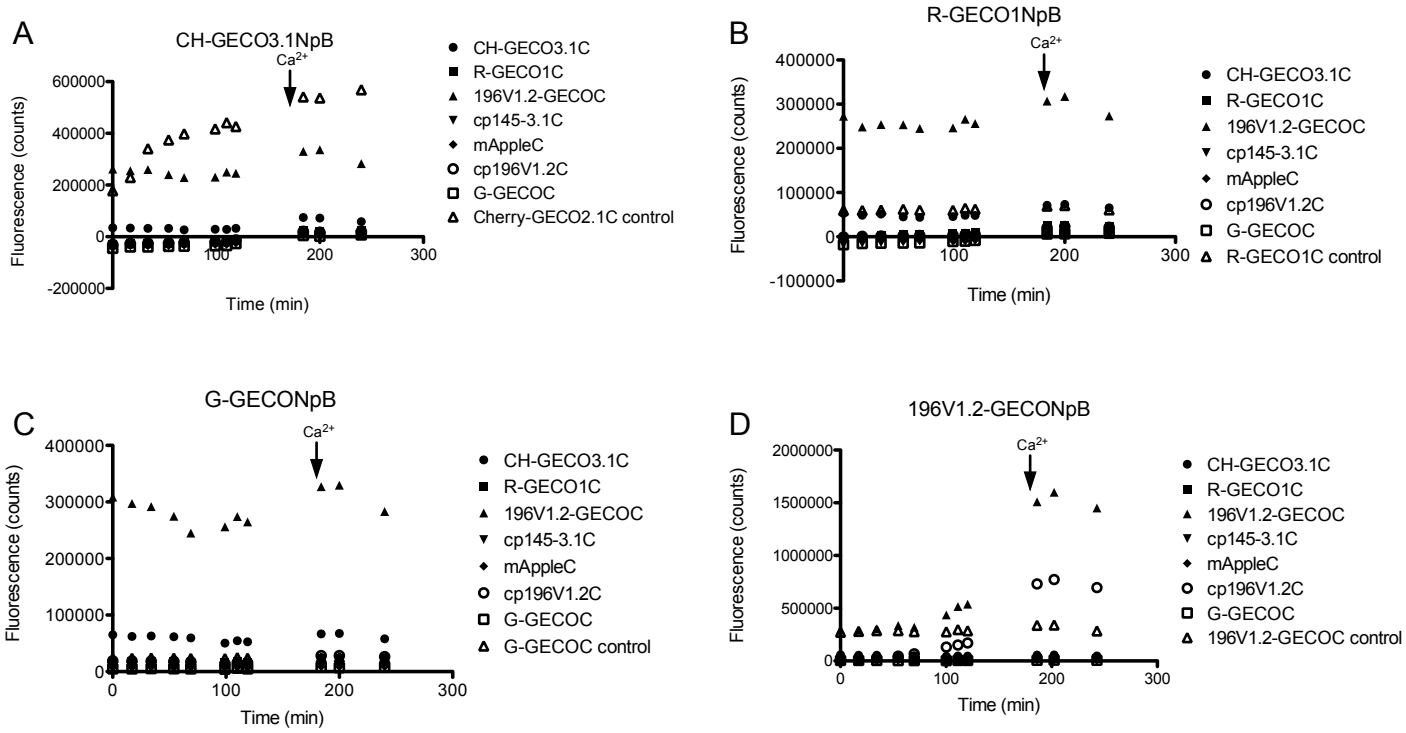
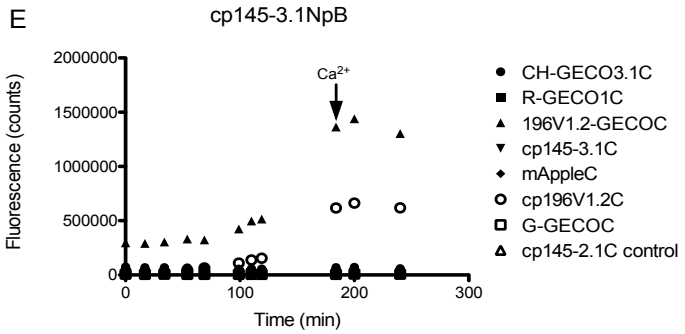
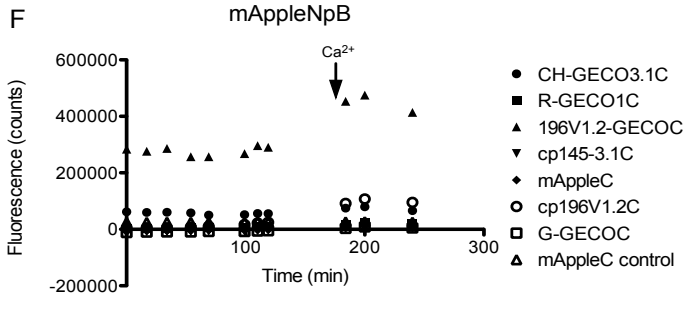
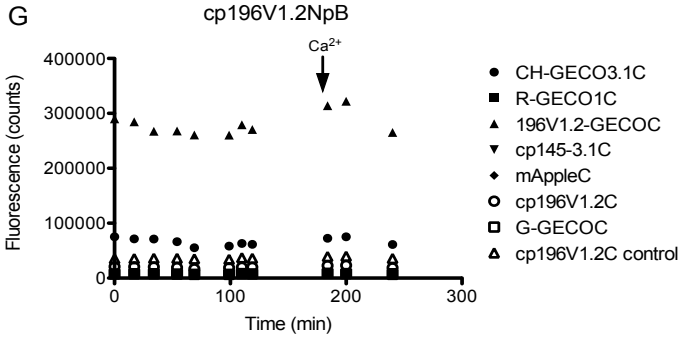


Figure 5.4 *In vitro* splicing results of Ca^{2+} biosensor proteins. Each graph is the summarized results of the mixing of designated N-intein fragment. Frames E through G are shown on the following page.


E
cp145-3.1NpB
Fluorescence (counts)
2000000
1500000
1000000
500000
0
Ca2+
0
100
200
300
Time (min)
CH-GECO3.1C
R-GECO1C
196V1.2-GECOC
cp145-3.1C
mAppleC
cp196V1.2C
G-GECOC
cp145-2.1C control


F
mAppleNpB
Fluorescence (counts)
600000
400000
200000
0
-200000
Ca2+
100
200
300
Time (min)
CH-GECO3.1C
R-GECO1C
196V1.2-GECOC
cp145-3.1C
mAppleC
cp196V1.2C
G-GECOC
mAppleC control


G
cp196V1.2NpB
Fluorescence (counts)
400000
300000
200000
100000
0
Ca2+
0
100
200
300
Time (min)
CH-GECO3.1C
R-GECO1C
196V1.2-GECOC
cp145-3.1C
mAppleC
cp196V1.2C
G-GECOC
cp196V1.2C control

5.2.2 Mammalian Cell Imaging

As the splicing data was obtained from SDS-PAGE and *in vitro* incubation was not definitive, rather than optimizing protein expression conditions we decided that the intein products would be co-transfected in mammalian cells and imaged for fluorescence and Ca^{2+} response as described in Methods and Materials. In addition to the RFP-based sensors, split G-GECO (14) was also cloned into pcDNA3. Part of the inspiration for combining several different FP fragments came from a publication by Hu et al, who were able to demonstrate using BiFC, that halves of different FPs could associate to form functional FPs (206). In their paper they generated split YFP, CFP, and GFP at positions 155 and 173, as well as BFP split at 173. Each of these fragments was fused to either bFos or bJun, proteins from the basic region leucine zipper (bZIP) family (283, 284). Out of a total 24 combinations they found 12 combinations that complemented to generate a fluorescent signal. In these experiments the two halves were not physically connected, as with inteins, but rather brought together in close proximity by the dimerization of the bFos or bJun domains such that the two halves could associate. As with inteins this complementation is essentially irreversible. The authors speculate that the differing amino acid sequences influence the efficiency of fluorescent complex formation, but do not interfere with dimerization of the binding partners.

In the case of the GECO-based intein constructs, the FP partners have a much larger sequence variation with R-GECO1 and CH-GECO3.1 (both DsRed-derived) differing by 26 amino acids. In addition, G-GECO is evolved from *Aequorea* GFP and 196V1.2-GECO is split at an entirely different position relative to all of the other constructs. Any difficulties arising from the difference in protein sequence could potentially be over come by the splicing process. This will link the two halves together, increasing the likelihood that they will fold up into a complete barrel structure to minimize exposed hydrophobic residues that would normally be internal to the barrel structure. In addition, the previous paper only monitored re-constitution of fluorescence, while this work looked at splicing of different Ca^{2+} sensors, which could lead to a unique fluorescence response depending on which halves combine. For completeness each N-intein half was

co-transfected with each C-intein, along with individual intein halves as controls. The results of all of imaging experiments have been summarized below in Table 5.2.

Overall, the constructs containing the G-GECO C-intein were the most robust, with cells displaying green fluorescence as well as responding to Ca^{2+} for all splice products. This result is interesting because the sequence identity of G-GECO compared with R-GECO1 or CH-GECO3.1 is only 20-30% for the FP portion, yet the chromophore is still able to form and Ca^{2+} response can be measured, albeit with a lower dynamic range. None of constructs containing the 196V1.2-GECO C-intein fragment showed a response to Ca^{2+}, which would be expected since original tests with the intact sensor did not show any response. However, splicing of other FPs with the 196V1.2-GECO N-intein fragment did provide some interesting results, as the protein would be missing the β8 and β9 strands. R-GECO1 was the poorest performing protein when spliced with alternate proteins, where no fluorescence response could be detected when spliced with G-GECONpC. Very weak Ca^{2+} response, relative to intact R-GECO1 was measured when spliced with either CH-GECO3.1NpC or 196V1.2-GECONpC. Finally, red fluorescence and a Ca^{2+} response was detected in all CH-GECO3.1CpC splice products, though in some cases histamine-oscillations were absent. Taken as a whole, intein splicing in mammalian cells seems to be a very robust process, as well as chromophore maturation in FPs composed of halves from relatively diverse variants.

5.2.2.1 CH-GECO3.1 Chromophores

During the evolution of a cpmCherry multiple rounds of mutagenesis were done as every change in the circular permutation site significantly reduced the proteins folding and maturation efficiency. Also, considering the mechanism of fluorescence response was determined to be different from the GECO series, it was unexpected that splicing G-GECO and R-GECO1 N-intein fragments with CH-GECO3.1CpC would result in an RFP-sensor that responded to Ca^{2+}. However, despite the obstacles when manipulating the protein in *E. coli,* mammalian cell expression was more successful. Next to the G-GECO constructs, CH-GECO3.1CpC spliced proteins were the most resilient. The cells

containing the CH-GECO3.1CpC fragment were only monitored for red fluorescence. The response ratios in Table 5.2 are close to the original values calculated from CH-GECO3.1 in Chapter 4 (Table 4.7). In all four intein constructs the ratio of Ca^{2+} responses are larger than the original sensor. This difference in performance could be due to several reasons such as cell behaviour, DNA quality, or differences in protein structure. For both the R-GECO1N and 196V1.2-GECON combinations fewer Ca^{2+} oscillations were detected despite a higher initial His peak and Ca^{2+}/ionomycin peak. Representative cells traces for both constructs are shown below in Figure 5.5.

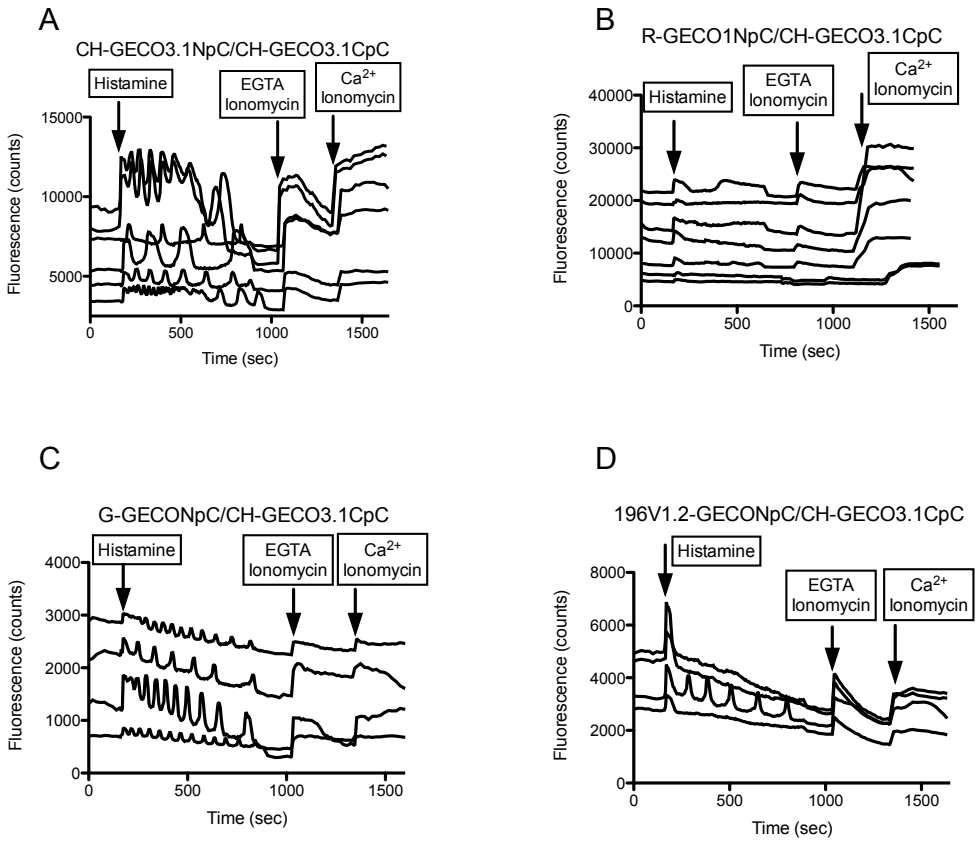


Figure 5.5 Cell traces for HeLa cell expression of CH-GECO3.1CpC spliced proteins. A) CH-GECO3.1NpC and CH-GECO3.1CpC. B) R-GECO1NpC and CH-GECO3.1CpC. C) G-GECONpC and CH-GECO3.1CpC. D) 196V1.2-GECONpC and CH-GECO3.1CpC. All cell traces were recorded monitoring red fluorescence.

His-induced oscillations were detected in CH-GECO3.1NpC/CH-GECO3.1CpC (Figure 5.5A), as well as the G-GECONpC/CH-GECOCpC combination (Figure 5.5C). The lack of oscillations in the R-GECO1NpC/CH-GECO3.1CpC combination is surprising because there are the same number of mutations in G-GECO M13 and CaM as compared with R-GECO1, so you might expect similar Ca^{2+} binding behaviour. These combinations indicate that the chromophore environment is very resilient and can still promote complete maturation with a different N-terminal sequence. Despite the differences in amino acid composition the overall environment formed by the G-GECONpC fragment must be similar. Interestingly, the residue Gln163, which seems to be responsible for the function of CH-GECO3.1, is absent in the G-GECONpC sequence. Rather, an Ile and His residue occupy positions 163 and 165, respectively. Either of these residues could be replacing the Gln163 in original CH-GECO3.1. Some red fluorescent images of HeLa taken during Ca^{2+} imaging are displayed in Figure 5.6 to highlight some of the fluorescent signal increases observed.

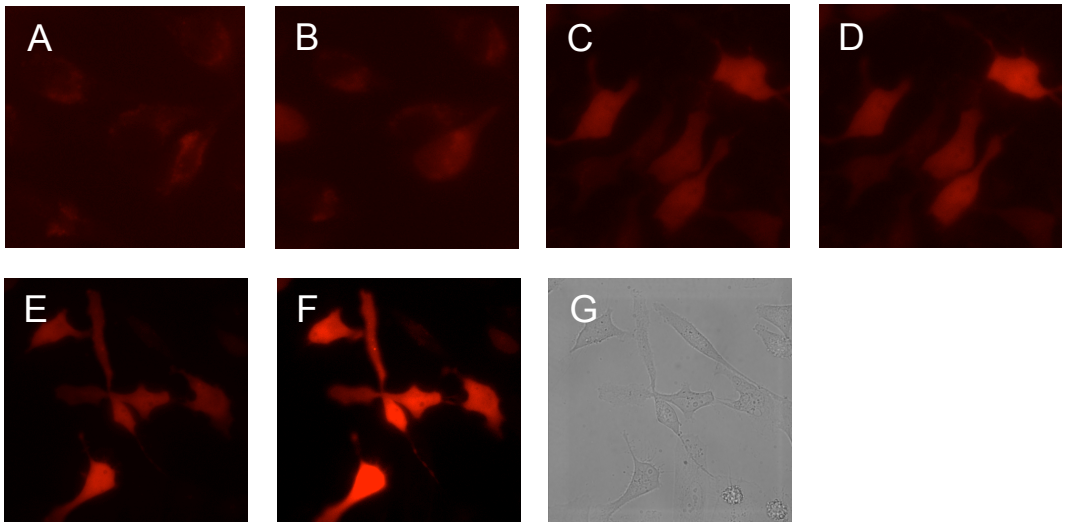


Figure 5.6 Red fluorescent images of HeLa cells expressing selected CH-GECOCpC intein combinations. A) G-GECONpC combination before addition of Histamine. B) G-GECONpC combination after addition of His/Ionomycin. C) CH-GECONpC combination baseline image. D) CH-GECONpC after addition of His/Ionomycin. E) R-GECO1NpC combination baseline fluorescence. F) Fluorescent image after addition of Ca^{2+}/ionomycin. G) DIC image of R-GECO1NpC/CH-GECOCpC expressing cells.

5.2.2.2 R-GECO1 Chromophores

When R-GECO1NpC and R-GECO1CpC were spliced the reconstituted R-GECO1 intein had a similar Ca^{2+}/EGTA ratio of 4.59 ±0.66, as compared with the reported value of 4.9 ±1.9. However, the His ratio was lower at 5.01 ±1.17 compared with 9.2 ±1.3 (Table 5.2). Somewhat unexpectedly the R-GECO1CpC constructs performed very poorly as compared with CH-GECO3.1CpC. When spliced with G-GECONpC or 196V1.2-GECONpC both ratios were significantly lowered to levels matching CH-GECO3.1, while no Ca^{2+} oscillations were detected when spliced with CH-GECO3.1NpC. A few possibilities for the poor performance are poor transfection efficiency, poor splicing efficiency, or simply that R-GECO1 is less tolerant of an alternate amino acid sequence. Some representative cells traces are shown below in Figure 5.7. Since the R-GECO1NpC/R-GECO1CpC splicing resulted in a protein with very similar performance to the original R-GECO1, the poor performance is not likely due to incomplete splicing or co-transfection. To better evaluate the splicing efficiency of the intein proteins both fragments would need to be in a dual-expression plasmid where transfection efficiency would not influence the amount of each fragment expressed. It was also reported the *Npu dnaE* intein splicing is extraordinarily resilient to a variety of extein sequences as well as temperature (278), so given the similarity of various extein sequences in this work we shouldn't expect R-GECO1CpC to hinder splicing efficiency.

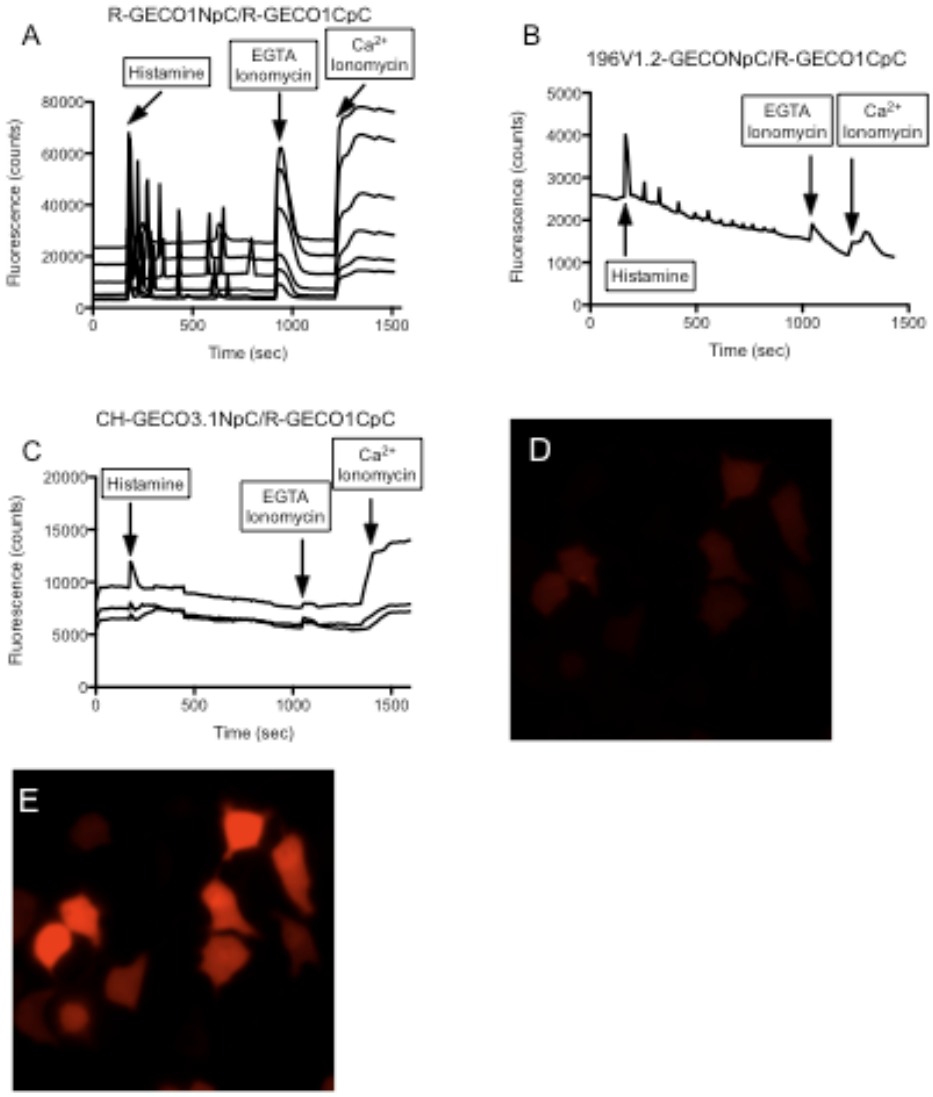


Figure 5.7 HeLa cell traces for R-GECO1CpC intein combinations. D) Baseline fluorescent image of R-GECO1NpC/R-GECO1CpC E) Fluorescent image of R-GECO1NpC/R-GECO1CpC after addition of histamine. Cell traces and fluorescent images were taken monitoring red fluorescence.

5.2.2.3 G-GECO Chromophores

For the G-GECOCpC splice products green fluorescence was monitored instead of red. After sequencing the G-GECO construct two mutations were found, Lys66Met, and Gly88Ser in comparison to G-GECO1.1 (G-GECO1.1 numbering). Rather than having a deleterious effect the performance in cells was improved over G-GECO1.1. The Lys66Met is right near the linker between M13

and EGFP in an analogous position to Gln163 of CH-GECO3.1, and could be affecting the behaviour of the protein. As seen with CH-GECO3.1, a more hydrophobic Met or Gln resulted in a larger fluorescence response when compared with the Lys163 mutation. The role of the Gly88Ser mutation is less clear, but it is the interface between CaM and the RFP and could be introducing a beneficial hydrogen bond. For the G-GECONpC/G-GECOCpC construct the Ca^{2+}/EGTA ratio improved to 27.3 ± 1.66 from 11.6 ± 2.9. The His ratio was relatively unchanged at 8.53 ± 3.04 compared with 11.7 ± 3.4 (Table 5.2).

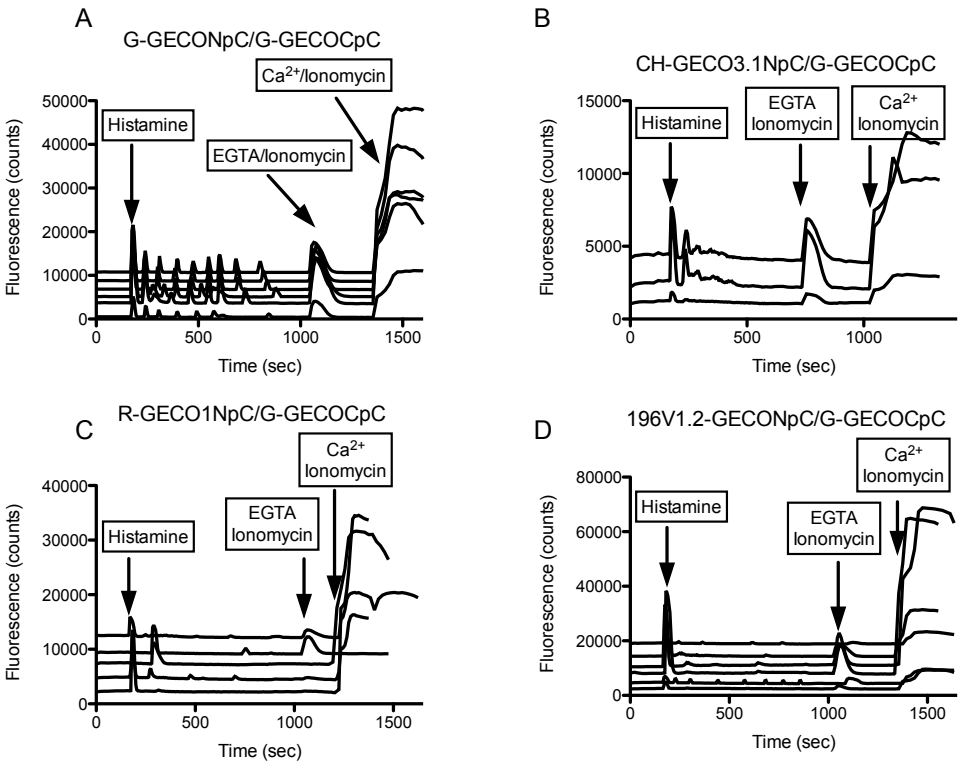


Figure 5.8 HeLa cell traces for G-GECOC intein combinations.

When spliced with R-GECO1Npc or CH-GECONpC, G-GECO still recorded some Ca^{2+} oscillations along with reasonably large Ca^{2+}/EGTA ratios (Figure 5.8B and C, respectively). When imaging, red fluorescence was also monitored, but no detectable signal was measured (data not shown). Representative cells traces are shown in Figure 5.8 along with some fluorescent images taken of the cells in Figure 5.9. Because the formation of the green chromophore requires one

less step, the maturation should be less interrupted by an RFP-based N-terminus sequence.

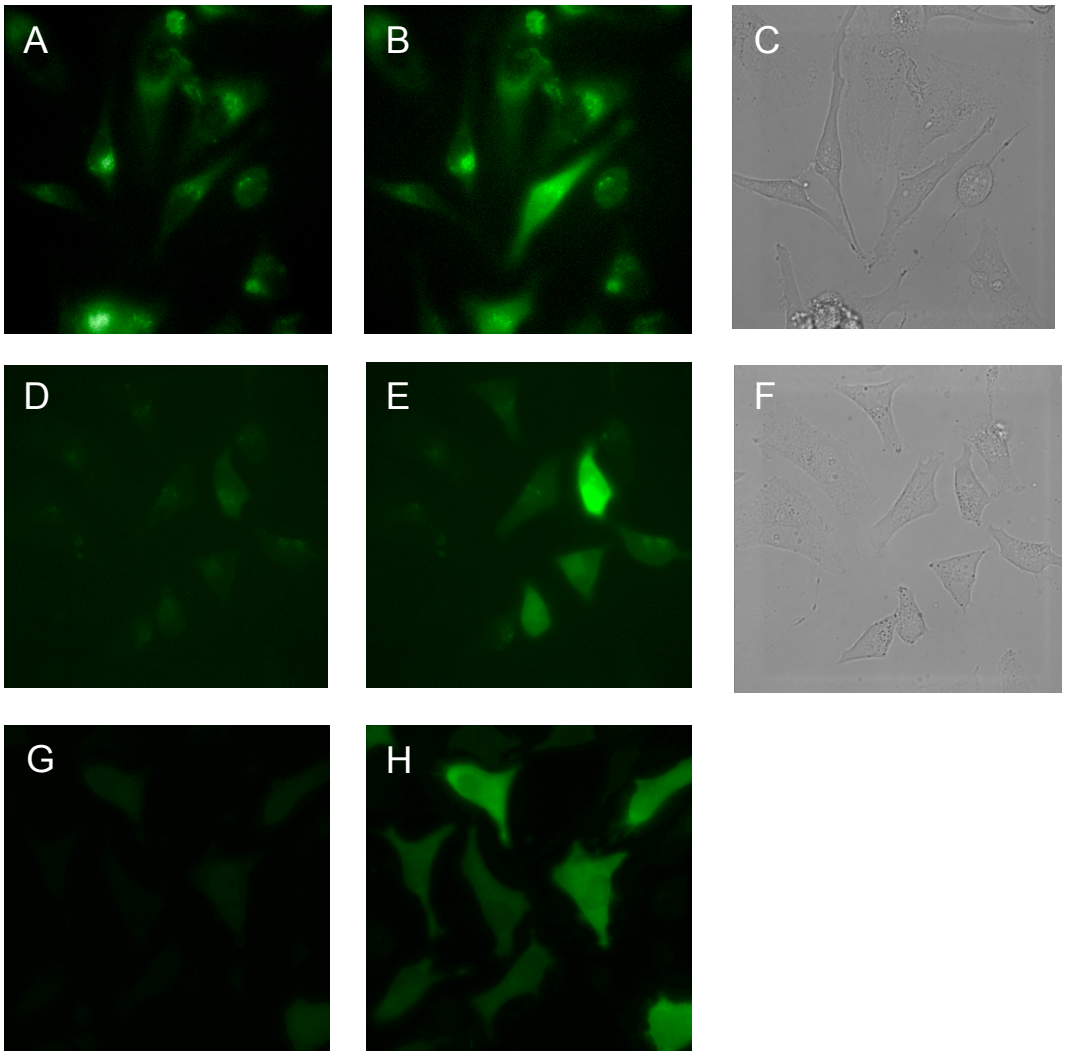


Figure 5.9 Green fluorescent images of HeLa expressing G-GECO intein constructs. A-C) R-GECO1NpC combination baseline fluorescence, fluorescent image after addition of Ca^{2+}/ionomycin, and DIC image of HeLa cells. D-F) 196V1.2-GECONpC combination baseline fluorescence, fluorescence after addition of Ca^{2+}/ionomycin, and DIC image of cells. G-H) Baseline fluorescence of G-GECONpC/CpC combination and fluorescence after addition of Ca^{2+}/ionomycin.

5.2.2.4 196V1.2-GECO Chromophores

From Chapter 3, initial tests with a 196V1.2-GECO construct did not result in a protein that responded to Ca^{2+}. The choice to include this construct was made primarily to determine if the other proteins would still function if there was a

duplication or deletion in sequence in the case of 196V1.2-GECONpC and 196V1.2-GECOCpC, respectively. Where 196V1.2-GECOCpC has been spliced together there will be a duplication of the sequence from position 144-196, while with 196V1.2-GECONpC, amino acids 145-195 will be missing (Figure 5.10).

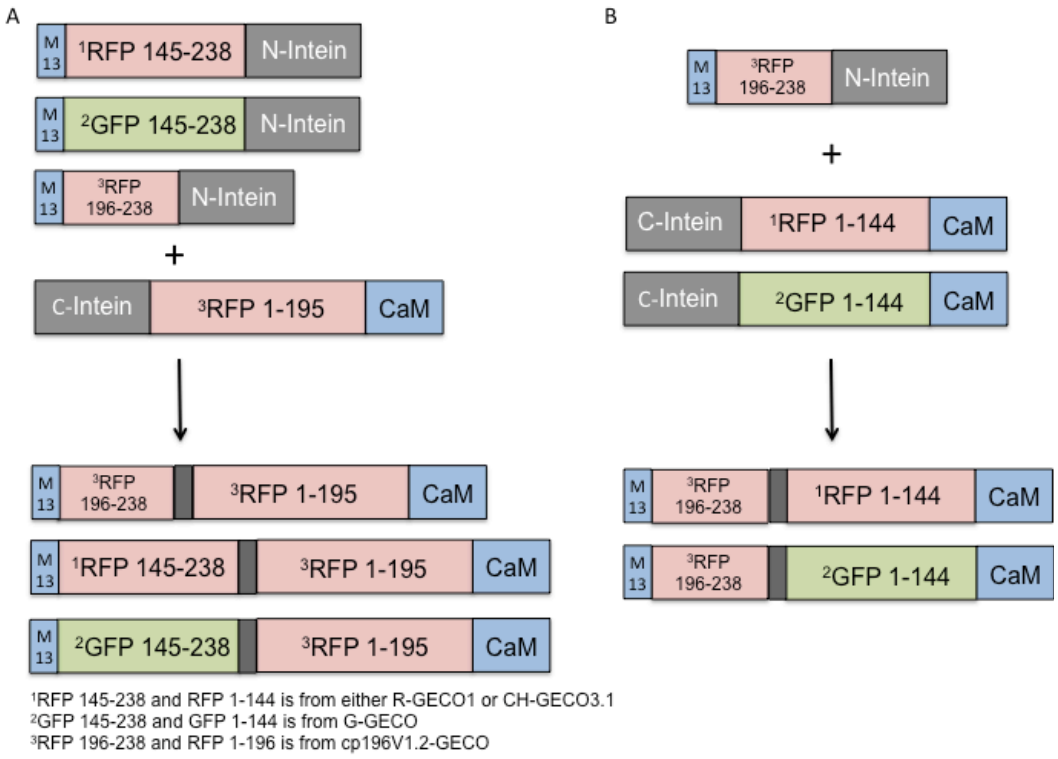


Figure 5.10 Schematic of Ca^{2+} biosensors when spliced with 196V1.2-GECOCpC and 196-GECONpC. A) Expected spliced proteins for R-GECO1NpC, CH-GECO3.1NpC, and G-GECONpC with 196V1.2-GECOCpC. These proteins will have a repetition of residues 145-195. B) Expected spliced proteins for R-GECO1CpC, CH-GECOCpC, and G-GECOCpC with 196V1.2-GECONpC. These proteins will be missing residues 145-195.

Red fluorescence was detected in cell expressing the R-GECO1NpC/196V1.2-GECOCpC, CH-GECO3.1NpC/196V1.2-GECOCpC, and G-GECONpC/196V1.2-GECOCpC spliced proteins, but no fluorescence response was recorded in the presence of Ca^{2+}. A few representative cell traces as well as fluorescent images of some of the HeLa cell are shown below in Figure 5.11. This result must mean that the split point still remains at position 196 and the addition N-terminal β-strands are not part of the barrel structure. These additional residues could form an extended loop region or a partially folded β-

sheet on the periphery of the protein. As mentioned in the previous sections the combination of the 196V1.2-GECONpC fragment still resulted in RFPs that responded to Ca^{2+}. This result seems unlikely since the β8 and β9 strands are missing, along with the purported residue at position 163 that plays an important role in interacting with the chromophore. Work done with *sf*GFP, which is missing the β11 strand reported that without the 11th strand the protein was non-fluorescent (285, 286). The added bulk of M13 and CaM must prevent solvent access allowing the chromophore to fully mature and Ca^{2+} binding must occlude the chromophore from the surrounding cellular environment, causing an increase in fluorescence.

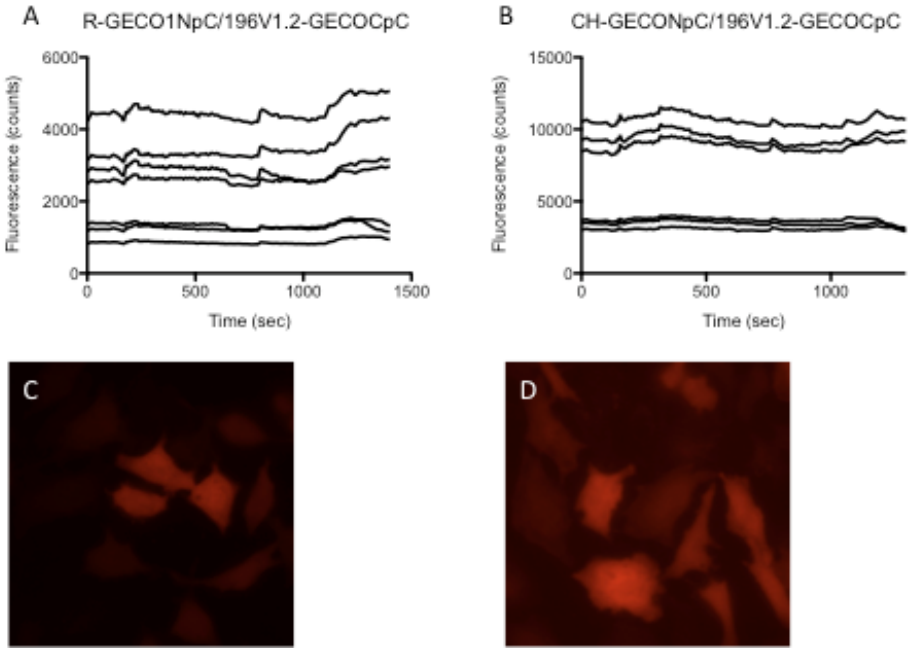


Figure 5.11 Cell traces and red fluorescent images of R-GECO1NpC/196V1.2-GECOCpC and CH-GECO3.1NpC/196V1.2-GECOCpC cells. A) Cell traces during normal Ca^{2+} monitoring. Small fluctuations around 750 sec are likely due to loss of focus or drift during addition of EGTA/ionomycin. The gradual fluorescence increase at 1100 sec is from cell death, rather than true fluorescence increase. B) Cell traces of CH-GECO3.1NpC splice products. C) Red fluorescent images of R-GECO1NpC/196V1.2-GECOCpC cells. D) Red fluorescent images of CH-GECO3.1CpC cells.

5.2.2.5 Control Inteins

Imaging the controls for the inteins was challenging since the cells were expected to be non-fluorescent. The N-intein fragments do not contain the chromophore so when imaging, regions were selected with at least half a dozen cells were visible. G-GECONpC, G-GECOCpC, 196V1.2-GECONpC, 196V1.2-GECOCpC, R-GECO1NpC, and CH-GECO3.1NpC imaging gave cell traces with no response. In some cells there is some signal fluctuation, which is more likely due to background signal or a slight change in the focus from addition or removal of solution, which can cause the cells to appear dimmer or brighter than they are. When there was a noticeable change in focus the cells were brought back into focus and any anomalous image frames were removed from the trace.

Table 5.2 HeLa cell characterization of split intein Ca^{2+} sensors

Intact protein or Intein fragments		n[1]	Maximum Ca^{2+} to Minimum EGTA ratio	Maximum His to Minimum His ratio	Maximum His to Maximum Ca^{2+} ratio
N-intein	C-intein				
[2]G-GECO1.1		31	11.6 ±2.9	11.7 ±3.4	0.69 ±0.22
[2]R-GECO1		22	4.9 ±1.9	9.2 ±1.3	0.98 ±0.15
CH-GECO3.1		9	1.13 ±0.02	1.07 ±0.02	0.89 ±0.02
CH-GECO3.1	CH-GECO3.1	17	1.36 ±0.10	1.28 ±0.10	1.06 ±0.05
R-GECO1	R-GECO1	17	4.59 ±0.66	5.01 ±1.17	0.94 ±0.16
G-GECO	G-GECO	23	27.3 ±1.66	8.53 ±3.04	3.72 ±1.72
196V1.2-GECO	196V1.2-GECO	-	nd[3]	nd	nd
R-GECO1	CH-GECO3.1	25	1.77 ±0.24	1.09 ±0.08	1.63 ±0.22
R-GECO1	G-GECO	23	4.10 ±4.40	1.84 ±1.34 (n=21)[4]	2.06 ±1.03
R-GECO1	196V1.2-GECO	10	1.16 ±0.20	1.17 ±0.01 (n=2)	0.97 ±0.003
CH-GECO3.1	R-GECO1	18	1.29 ±0.12	nd	N/A
CH-GECO3.1	G-GECO	6	6.10 ±6.30	2.54 ±1.53	2.06 ±0.69
CH-GECO3.1	196V1.2-GECO	-	nd	nd	nd
G-GECO	R-GECO1	2	1.59 ±0.11	1.23 ±0.04	1.29 ±0.05
G-GECO	CH-GECO3.1	13	1.20 ±0.06	1.35 ±0.08	1.13 ±0.06
G-GECO	196V1.2-GECO	-	nd	nd	nd
196V1.2-GECO	G-GECO	17	7.94 ±6.78	4.34 ±3.29 (n=14)	2.10 ±0.43
196V1.2-GECO	CH-GECO3.1	4	1.44 ±0.12	1.29 ±0.10	1.12 ±0.15
196V1.2-GECO	R-GECO1	2	1.39 ±0.01	1.51 ±0.16	0.92 ±0.03

[1]Number of transfected cells on which Ca^{2+} measurements were made

[2]Data for G-GECO1.1 and R-GECO1 are From Zhao et al., *Science*. 2011, **333**, 1888- 1891. Reprinted with permission from AAAS.

[3]nd indicates that no Ca^{2+} response was detected therefore no measurements could be made

[4]When the His cell count does not match n, it means that no fluorescence response was detected in some cells when His was added, so the new cell count reflects only those with measured responses to His.

5.3 Conclusions

Fluorescent protein splicing using split inteins offers a new tool for scientists to use when trying to probe specific protein-protein interactions, cell-specific gene expression, or simply as a protein engineering tools to test different split combinations of fluorescent proteins. It is apparent from the *in vitro* screening that dual-expression plasmids could potentially be a more convenient avenue to pursue when purifying spit proteins in *E. coli* due to the reduced solubility of the protein halves. If the proteins are expressed separately then an additional denaturation/renaturation step must be included in protein purification to promote proper protein folding.

Mammalian cell imaging was more successful in that co-transfected inteins were expressed and able to splice together the FP halves to regenerate functional sensors. Interestingly, FP-based sensors with β-strand deletions not only retained fluorescence, but were also functional. No new combinations of different FP-sensors surpassed the behaviour of the original constructs, but it could serve a purpose for screening a variety of FP combinations for new responses. Of the non-native pairs the CH-GECO3.1N/G-GECOC and R-GECO1N/G-GECOC intein constructs were the most promising, perhaps the chromophore maturation was more robust and more tolerant of alternate protein halves. It would be interesting to try splicing a wider variety of Ca^{2+} biosensors together, perhaps from the GECO series or other colors and characterize their behaviours. Based on the results here it would seem more efficient to do preliminary screening in mammalian cells and the most promising variants could then be cloned into a dual expression plasmid to ensure efficient *in vitro* splicing.

In order to confirm if *in vitro* splicing takes place as opposed to dimerization the C-intein fragment should be cloned with a 3' His-tag to eliminate the possibility of His-tag interference during splicing. Additionally, SDS-PAGE could be used before and after splicing reactions to monitor the extent of splicing.

Very little work has been done with inteins and fluorescent proteins as compared with other complementation applications. This could be due to either a lack of awareness of inteins or perhaps a lack of cross-talk between scientists

who specialize in using intein/FP technology with individuals studying unique biological systems where traditional FP applications fall short. The preliminary work discussed in this Chapter has not only demonstrated the utility and robustness of intein splicing, but more importantly has further illustrated the ability of FPs and FP-based biosensors to tolerate splicing and reconstitute fluorescence-based response.

5.4 Methods and Materials

5.4.1 General Methods and Materials

The standard reagents for PCR reactions were from Fermentas and a 1:1 mixture of Pfu/Taq was used for amplifying DNA. PCR products were separated using agarose gel electrophoresis (10%) and purified using the GeneJet gel extraction kit (Fermentas). Fast digest enzymes were purchased from Fermentas and used according to the manufacturers protocol. Again the DNA was purified using the same methods. For ligations T4 DNA ligase (Invitrogen) was used and the ligation products were electroporated into *E. coli* DH10B or JM109 cells. Oligos 4.40 and 4.41 were used to sequence the ligation products in the pTriEx-3 plasmid, oligos 2.1 and 2.2 were used to sequence the products in the pBAD/His B B plasmid, and oligos 4.40 and 4.41 were used to sequence the products in pcDNA3.1.

5.4.2 Cloning RFP-based Ca^{2+} sensors and RFPs

The plasmids containing the N-intein and C-intein were provided by Dr. Kevin Truong from University of Toronto. Both constructs were contained in the pTriEx^{TM}-3Hygro DNA vector and the plasmid design is shown in Figure 5.2. Initially oligos for CH-GECO3.1, R-GECO1, 196V1.2-GECO, and the individual fluorescent proteins cp145-3.1, mApple, and cp196V1.2, respectively, were designed to insert each portion into the intein plasmids. Seven out of the twelve possible constructs were successfully cloned into pTriEx-3, but difficulties with duplicated restriction sites and cloning efficiency led me to decide to insert the DNA into pBAD/His B. The C-intein was cloned into pBAD/His B using oligos 5.45/5.25 and the N-intein was cloned using oligos 5.21/5.29. Each distinct DNA fragment was amplified with the oligos indicated in Table 5.1 and digested with

the corresponding restriction enzymes. For the FP constructs that could not be cloned initially into the pTriEx-3 plasmid the desired DNA fragments were cloned into either Cintein/pB or N-intein/pB. Due to some restriction site compatibility the two extra BamHI sites in pBAD/His B were removed using oligos 5.36 and 5.37 before mAppleNpB could be ligated into the digested plasmid. Similarly for cp196V1.2GECOCpB, the MluI site between EGFP and CaM had to be replaced with SalI using oligo 5.42 because there was an internal MluI site in the pBAD plasmid.

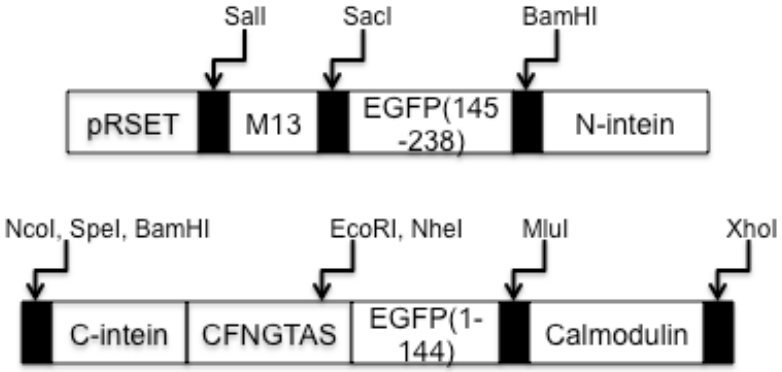


Figure 5.12 Layout of initial N-intein and C-intein constructs in the pTriEx-3 Hygro plasmid.

Table 5.3 Cloning Strategy for each construct described in Chapter 5

		Amplification		Cloning
Construct	**Plasmid**	**Template**	**Oligos**	**RE sites (5'/3')**
CH-GECO3.1 NpT	pTriEx-3	CH-GECO3.1	5.1/5.2	SacI/BamHI
CH-GECO3.1 CpT	pTriEx-3	CH-GECO3.1	5.3/5.4	EcoRI/XhoI
cp145-3.1 CpT	pTriEx-3	CH-GECO3.1	5.3/5.6	EcoRI/XhoI
R-GECO1 CpT	pTriEx-3	R-GECO1	5.4/5.5	EcoRI/XhoI
mApple CpT	pTriEx-3	R-GECO1	5.3/5.7	EcoRI/XhoI

196-GECO1.2 NpT	pTriEx-3	cp196V1.2	5.2/5.8	SacI/BamHI
cp196V1.2 CpT	pTriEx-3	cp196V1.2	5.3/5.46	EcoRI/XhoI
C-intein pB	pBAD/His B	C-intein pT	5.45/5.25	XhoI/HindIII
N-intein pB	pBAD/His B	N-intein pT	5.21/5.29	XhoI/HindIII
CH-GECO3.1N pB	pBAD/His B	CH-GECO3.1NpT	5.11/5.29	XhoI/HindIII
CH-GECO3.1C pB	pBAD/His B	CH-GECO3.1CpT	5.25/5.45	XhoI/HindIII
cp145-3.1 NpB	pBAD/His B	CH-GECO3.1	5.12/5.29	XhoI/HindIII
cp145-3.1 CpB	pBAD/His B	cp145-3.1 CpT	5.16/5.45	XhoI/HindIII
R-GECO1 NpB	pBAD/His B	R-GECO1	5.11/5.29	XhoI/HindIII
R-GECO1 CpB	pBAD/His B	R-GECO1 CpT	5.21/5.25	XhoI/HindIII
mApple NpB	pBAD/His B	R-GECO1	5.13/5.29	XhoI/BamHI
mApple CpB	pBAD/His B	R-GECO1 CpT	5.17/5.45	XhoI/HindIII
196V1.2-GECO NpB	pBAD/His B	cp196V1.2-GECO NpT	5.10/5.25	SacI/HindIII
196V1.2-GECO CpB	pBAD/His B	cp196V1.2 CpT	5.44/5.45	XhoI/SalI
cp196V1.2 NpB	pBAD/His B	cp196V1.2 NpT	5.24/5.29	XhoI/HindIII
cp196V1.2 CpB	pBAD/His B	cp196V1.2 CpT	5.3/5.28	XhoI/HindIII
CH-GECO3.1N pC	PCDNA3.1	CH-GECO3.1NpB	5.38/5.40	HindIII/EcoRI
CH-GECO3.1C pC	PCDNA3.1	CH-GECO3.1CpB	5.39/5.41	HindIII/EcoRI
R-GECO1 NpC	PCDNA3.1	R-GECO1 NpB	5.38/5.40	HindIII/EcoRI
R-GECO1 CpC	PCDNA3.1	R-GECO1 CpB	5.39/5.41	HindIII/EcoRI

196V1.2-GECO NpC	PCDNA3.1	196V1.2-GECO NpB	5.38/5.40	HindIII/EcoRI
196V1.2-GECO CpC	PCDNA3.1	196V1.2-GECO CpB	5.39/5.41	HindIII/EcoRI
G-GECO NpC	PCDNA3.1	G-GECO	5.38/5.40	HindIII/EcoRI
G-GECO CpC	PCDNA3.1	G-GECO	5.39/5.41	HindIII/EcoRI

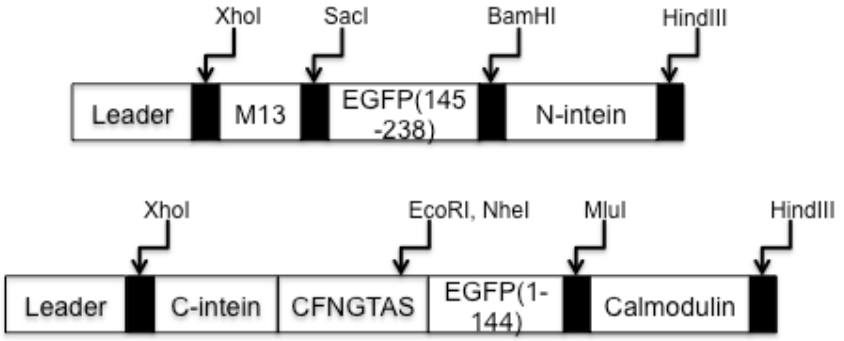


Figure 5.13 General layout for the intein constructs in the pBAD/His B plasmid. The 'leader' sequence is the purification epitope (MGGSHHHHHHGMASMTGGQQMGRDLYDDDDKDPSSR).

5.4.3 Protein Expression and Purification

Once cloned into pBAD attempts were made to express and purify each intein protein. An individual colony was grown overnight at 37°C in a 5 mL LB/amp culture tube. This 5 mL culture tube was used to inoculate a 250 mL flask of modified TB. The flask was grown at 37 °C until the OD was 0.6 and protein expression was induced by addition of arabinose to 0.2 %. The flasks shook overnight and the proteins were purified from *E. coli* using the same method as described in Section 2.4.6 and 3.4.5. A second round of purification was done using the modified LB media described in Section 4.4.3. After protein expression was induced the flasks continued to shake for another 5 hours before isolating the bacterial pellet. Protein purification was carried out using the same methods described in Section 2.4.6 and 3.4.5.

5.4.4 SDS-PAGE

Each purified protein was run on a 10% SDS-PAGE gel prepared following the protocol described in literature (277) Each protein solution was boiled with SDS-loading buffer for 5 minutes before loading 20 μL onto each lane.

5.4.5 BCA Assay

The concentrations of each protein solution were determined using the BCA assay (Pierce) according to the manufacturers protocol. Proteins were measured in triplicate at three different dilutions.

5.4.6 *In vitro* protein splicing

To test the *in vitro* splicing of the intein fragments the proteins were mixed together in 50 μL to a final concentration of 14 μM each. Control wells containing each individual intein at the same final concentration were also measured. The fluorescence emission at 460 nm and 550 nm excitation was collected every 15 minutes for two and a half hours. The microwell plate was left in the fridge overnight and 1 μL of 10 mM CaEGTA was added to each well the next day. After warming to room temperature the fluorescence emission was collected again.

5.4.7 Mammalian Cell Imaging

A total of eight different constructs were cloned into the PCDNA3.1 plasmid as indicated in Table 5.1. Hela cells were used for all of the imaging experiments. The same microscope was used as indicated in section 4.2.7 as well as the same Ca^{2+} imaging protocol. To test the robustness of the splicing each N-intein portion was co-transfected with each C-intein construct. For example RGECONpC was imaged with CH-GECO3.1CpC, RGECOCpC, 196V1.2GECOCpC, and GGECOCpC. For co-transfection 500 ng of each plasmid DNA were mixed together with 2 μL Turbofect and 100 μL of serum-free medium. Controls were also run with only the N-intein or C-intein present. For these transfections 1 μg of DNA was introduced into the cells. The cells were imaged for red or green fluorescence where appropriate.

Chapter 6 Conclusions and Future Directions

Fluorescent protein technology has made tremendous progress since *av*GFP was cloned, sequenced, and first expressed in the worm *C.elegans* (46, 287). Red-shifted FPs discovered in other sea creatures such as corals and copepods further expanded the colour palette (24, 78). In addition, protein engineering has enabled the development of FPs with altered wavelengths, reduced oligomeric tendencies, improved protein expression, and many other favourable characteristics (25, 43, 48, 74). Still the most common use for FPs is in live cell microscopy as fluorescent tags to allow scientists to image the localization of cellular proteins. A wider array of colours with spectrally distinct excitation and emission peaks has allowed for better multiparameter imaging without the need for complex post-imaging processing (101).

A more challenging and therefore less developed application of FPs is their use as FP-based biosensors. Using the phenomena of FRET or single FPs engineered to be environmentally sensitive, biosensors have been developed to detect a subset of small molecules such as Ca^{2+}, H_2O_2, cGMP, and Cl^- along with kinase or protease activity (7, 15, 16, 37, 39, 288, 289). Despite the numerous advances in the area of FP-based biosensors, this field is still growing and there is much room for improvement to increase their utility for *in vivo* studies. One of the main areas for improvement is the sensitivity of Ca^{2+} indicators. The Ca^{2+} biosensor GCaMP3 has been used to monitor neuronal activity (290-292), but is still unable to detect single action potentials (32). The sensitivity of Ca^{2+} indicators is limited by lower affinity and smaller fluorescent signal changes in the range of physiologically relevant Ca^{2+} concentrations. Another area of improvement is in the development of voltage sensitive biosensors (41, 293). The kinetics of most sensors do not allow for baseline resolution of action potentials and the signal change is small (294, 295). Despite these current limitations, scientist's eagerness to better understand neurological connections and functions will undoubtedly continue to motivate research groups to design biosensors with improved function.

Crucial to the success development of single FP-based biosensors is the location of the external binding domains in relation to the chromophore. FPs are formed of an eleven stranded β-barrel with the chromophore located in the centre, protected from the external environment. In order to link an external binding event with the fluorescence output a FP must be circularly permutated, such that the new termini are in close proximity to the chromophore. Applying this approach top the development of a single RFP-based Ca^{2+} biosensor was the primary goal of my thesis work.

As Ca^{2+} is an ubiquitious second messenger in intracellular signalling (239, 240), much work has been invested in engineering FRET-based and single FP-based Ca^{2+} sensors. Despite the discovery of RFPs, up until 2011 all of the Ca^{2+} biosensors were generated from GFPs, YFPs, and CFPs (15, 16, 296), which made multiparameter fluorescence imaging with FPs impractical. We hoped to remedy this deficiency by engineering an RFP-based single-FP Ca^{2+} biosensor.

In this chapter I hope to succinctly summarize the body of work described in this thesis, namely the development of a cpRFP-based Ca^{2+} biosensor along with the use of split inteins to reconstitute split Ca^{2+} biosensors in mammalian cells. Alongside this summary I will discuss potential future work in the context of my research results. Potential future work with inteins and FPs will also be discussed in the context of new screening strategies and imaging applications. Finally, I will also focus on the broader application of cpRFP engineering and their use in biosensor designs.

6.1 Summary of thesis work

In this thesis I describe the development of circular permutated RFPs that served as the FP scaffold for attaching the Ca^{2+} binding domains, M13 and CaM. The RFP chosen for this work was mCherry, an RFP engineered from mRFP1.0 (25). Previous work in the Campbell lab identified positions in mCherry that would tolerate circular permutation and one of these constructs was chosen as the starting point for my research (23). The initial cp193V1.0 mCherry was plagued with poor folding and maturation efficiency so random mutagenesis was used to screen for improved variants. Starting with an improved cp193 mCherry (cp193g7) a library of circularly permutated variants was generated at all

positions in the β10 strand. To increase the likelihood of finding a variant that retained fluorescence the N and C-terminal residues were randomized and after screening 15 positions, red fluorescent circularly permutated variants were located at positions 193, 195, 196, 201, 202, 203, 204, 205, 206, and 207.

The two most likely candidates for a Ca^{2+} biosensor were cp196V1.0 and cp201V1.0, however they were also the dimmest variants with quantum yields of 0.15 and 0.22, and extinction coefficients of 4600 $M^{-1}cm^{-1}$ and 2600 $M^{-1}cm^{-1}$, respectively. In Chapter 3 the mutagenesis of these variants along with the construction and screening of potential Ca^{2+} biosensors was discussed. After mutagenesis of cp196V1.0, the final improved version, cp196V1.2 had an improved quantum yield of 0.21 and extinction coefficient of 49,500 $M^{-1}cm^{-1}$. From cp196V1.2, potential GCaMP-type Ca^{2+} biosensors were constructed at positions 197-200 and 144-147. The most promising variant was empirically found to be the one permutated at position 145. Several rounds of mutagenesis to improve the fluorescence response produced a variant named CH-GECO3.1, whose fluorescence increased 2.5x upon addition of Ca^{2+}.

In Chapter 4 I took a more in depth look at the mechanism of fluorescence increase and other properties of CH-GECO3.1. CH-GECO3.1 has a K_d of only 6 nM and a pK_a of 5.1 that remains constant in the presence or absence of Ca^{2+}. The pH titration curves indicated the mechanism was not due to pK_a modulation, but rather via increased quantum yield and extinction coefficient. Without Ca^{2+} the quantum yield and extinction coefficient are 0.05 and 37,000 $M^{-1}cm^{-1}$, which increase to 0.17 and 48,000 $M^{-1}cm^{-1}$ with Ca^{2+}, respectively. Both CH-GECO3.1 and the previous generation CH-GECO3.0 were successfully used to image Ca^{2+} dynamics in mammalian HeLa cells, though CH-GECO3.0 performed better, likely due to the differences in the K_d of the sensors. Site-directed mutagenesis pointed to a few key residues that seemed to play an important role in the proteins function and proper chromophore maturation. Gln163 is thought to sandwich the chromophore with Ile161 and Trp143 in the Ca^{2+}-bound state, stabilizing the chromophore, which could increase the quantum yield. A few other residues such as the Glu in the M13-RFP linker along with Ser146 may be interacting with the phenolate group of the chromophore in either the apo state or Ca^{2+}-bound state, respectively, also influencing the quantum yield and extinction

coefficient. Without a crystal structure no definite structural conclusions can be made, but based on the results of the mutagenesis and structures of related FPs such as mCherry (245) and R-GECO1 (Eric R. Schreiter, personal communication) I believe Gln163, Ser146, and the Glu in the linker play key roles in modulating the fluorescence response. I was also able to demonstrate that CH-GECO3.1 could be turned into a functioning orange GECO (CHO-GECO). A blue emitting fluorescent protein and a long Stokes-shift variant were also engineered, but unfortunately neither retained their response to Ca^{2+}. Further work with theses variants could lead to more Ca^{2+} biosensors with alternate spectral properties. A family tree connecting all of the cpmCherry variants and mutations described in this thesis can be found in Appendix C.

The final chapter in my thesis was more application based and involved the use of the *Npu dnaE* split inteins, kindly provided to me by Dr. Kevin Truong, from the University of Toronto. Inteins are proteins that catalyze protein splicing of exteins, which are protein sequences that flank the N- and C-terminal sides of the intein. The first intein was discovered over 30 years ago (162) and to date over 400 inteins have been reported in all three branches of the tree of life (164). The combined use of FPs and inteins has been rather limited and only one recent paper reported the successful co-expression and splicing of FP-based biosensors (190). The first report of the successful reconstitution of the Ca^{2+} biosensors TN-XL (297) and GCaMP2 (31) in mammalian cells inspired us to expand the repertoire of inteins to other split Ca^{2+} biosensors such as R-GECO1, CH-GECO3.1, G-GECO, along with a cp196 construct, cp196V1.2-GECO. Splicing was tested with all combinations of protein and it was shown that reconstituted R-GECO1, CH-GECO3.1, and G-GECO retained function. In addition, some interesting recombinations of R-GECO1, CH-GECO3.1, cp196V1.2-GECO, and G-GECO also resulted in functional proteins. No combination had superior performance to the original Ca^{2+} biosensors, but the results present a unique opportunity for screening possible new biosensor constructs.

6.2 Future Directions

Despite the amount of work described in this thesis, there remain many avenues that could be further explored with both the intein work as well as with CH-GECO3.1. For CH-GECO3.1 there is room for improvement concerning both the quantum yield and extinction coefficient in the Ca^{2+} bound state. As mentioned, the mechanism for fluorescence is increased quantum yield and extinction coefficient, and the theoretical limit would be a protein whose quantum yield and extinction coefficient matched that of mCherry (0.22, 72,000 $M^{-1}cm^{-1}$) when Ca^{2+} bound. As further improvements are made, new colony screening techniques may need to be employed to detect smaller, incremental changes. Such techniques might include spraying plates with EGTA to modulate the fluorescence response, or perhaps even the paper-based screening protocols developed by a fellow lab member (Yi Shen, unpublished results). Based on the mammalian cell imaging results it may also be necessary to alter K_d, by reverting some of the mutations in CaM back to the original amino acids to increase its utility in environments with higher Ca^{2+} concentrations. More immediate work could be focused on obtaining a crystal structure of both the apo- and Ca^{2+}-bound states to help confirm if the amino acids I proposed were playing the suggested structural roles.

Beyond the original CH-GECO3.1, another future directions will be work that expands upon some of the results I discussed in Section 4.2.5. Preliminary mutagenesis results suggested that CHO-GECO was a functional orange-emitting biosensor that also has the potential for photoconversion to a far-red emitting species. The protein could be irradiated with green laser light to confirm if photoconversion occurs and to what degree. If the photoconversion is reproducible it would be worth screening for both improved orange fluorescence response along with red fluorescence response. Currently there is a green to red photoconvertible Ca^{2+} indicator (298), but the addition of an orange to far-red sensor would be a welcomed addition and could be use in conjuction with green emitting protein. Further exploration of the long Stokes-shift variant would also be useful. Although my preliminary results did not suggest that Ca^{2+} responsiveness was retained, there are other mutation sites that could be introduced to generate a long Stokes-shift that may retain the desired Ca^{2+} sensing function (110, 114).

Finally, the use of split inteins with FPs represent a potential new stream of screening or protein-protein interaction detection system similar to BiFC (203). Similar to BiFC, intein based protein splicing would not be reversible, but splicing effiecient would be expected to be higher compared with BiFC fragment association, which is reported to be less than ideal (203). There would still be a time delay depending on the size of the two fragments and if the chromophore maturation is only initiated upon splicing. The ability to reconstitute functional biosensors also presents an interesting opportunity to place the two intein fragments under the control of two different promoters such that expression of both fragments only occurs under specific cellular triggers and/or localization. As an added bonus, once reconstituted, the protein could then be used to monitor cellular dynamics such as Ca^{2+}. As mentioned in Section 1.3.6.2, split Ca^{2+}-biosensors could be used to locate interneurons and then used to monitor neuron activity indirectly through Ca^{2+} fluctuations. Another potential use for inteins is at the front end of FP biosensor design during the screening phase. As discussed in Section 5.2.1 *in vitro* screening was not very successful, but optimization of protein expression and purification may allow testing of a variety of spliced proteins composed of two different FP halves. BPER purification could be used and splicing of proteins could be monitored using a microplate reader. Libraries of FP fragments could also be cloned into the appropriate dual expression plasmids and co-expression would allow the testing of the spliced products. Alternately, combinations of FPs or FP biosensor fragments co-transfected into mammalian cells would offer a new screening strategy for potentially unique biosensors.

6.3 Concluding Remarks

Beyond the work in this thesis, the broader topic of cpRFP development still remains a relatively unexplored area of investigation. Including the work described in this thesis, there are only a handful of successful single cpRFP-based biosensors (14, 22, 298). It is likely that disruption of the RFP structure with circular permutation has more deleterious effects on chromophore maturation (relative to GFP) since it requires an additional post-translational modification step compared with GFPs (58). However, despite this engineering hurdle, the benefits of having additional red-shifted biosensors sensitive to a

multitude of biomolecules justifies the investment of effort. A review of RFP engineering has been published that touches upon several different considerations including starting template, screening method and mutagenesis strategies (263). Of course we must look beyond Ca^{2+} and towards other interesting physiological functions and processes that would benefit from FP technology including many in the field of neuroscience. Ca^{2+} biosensors have traditionally been used to monitor neurological activity indirectly through Ca^{2+} fluctuations, but more recently there has been a push to develop voltage-sensitive FPs (293, 299). A cpmKate variant has been engineered as a VSFP (22), but there are a wide variety of voltage sensitive domains that could be used as the sensing domain. Another related field is that of optogenetics that looks to monitor or control specific cell populations using light (294, 295). Both these emerging and important fields will benefit from engineered RFPs along with more traditional biosensors where a lack of red-shifted emissions limits the ability to do multiparameter experiments.

References

1. Campbell, R. E. Fluorescent Proteins. *Scholarpedia* **2008,** *3*, 5410.

2. Yang, F., Moss, L. G. and Phillips, G. N. The molecular structure of green fluorescent protein. *Nat Biotechnol* **1996**, *14*, 1246-1251.

3. VanEngelenburg, S. B. and Palmer, A. E. Fluorescent biosensors of protein function. *Curr Opin Chem Biol* **2008**, *12*, 60-65.

4. Campbell, R. E. Fluorescent-Protein-Based Biosensors: Modulation of Energy Transfer as a Design Principle. *Anal Chem* **2009**, *81*, 5972-5979.

5. Zhang, J., Campbell, R. E., Ting, A. Y. and Tsien, R. Y. Creating new fluorescent probes for cell biology. *Nat Rev Mol Cell Bio* **2002**, *3*, 906-918.

6. Johnson, D. E., Ai, H. W., Wong, P., Young, J. D., Campbell, R. E. and Casey, J. R. Red fluorescent protein pH biosensor to detect concentrative nucleoside transport. *J Biol Chem* **2009**, *284*, 20499-20511.

7. Jayaraman, S., Haggie, P., Wachter, R. M., Remington, S. J. and Verkman, A. S. Mechanism and cellular applications of a green fluorescent protein-based halide sensor. *J Biol Chem* **2000**, *275*, 6047-6050.

8. Baird, G. S., Zacharias, D. A. and Tsien, R. Y. Circular permutation and receptor insertion within green fluorescent proteins. *Proc Natl Acad Sci U S A* **1999**, *96*, 11241-11246.

9. Zou, J., Hofer, A. M., Lurtz, M. M., Gadda, G., Ellis, A. L., Chen, N., Huang, Y., Holder, A., Ye, Y., Louis, C. F., Welshhans, K., Rehder, V. and Yang, J. J. Developing sensors for real-time measurement of high Ca2+ concentrations. *Biochemistry* **2007**, *46*, 12275-12288.

10. Chen, N., Zou, J., Wang, S. M., Ye, Y. M., Huang, Y., Gadda, G. and Yang, J. J. Designing Protease Sensors for Real-Time Imaging of Trypsin Activation in Pancreatic Cancer Cells. *Biochemistry* **2009**, *48*, 3519-3526.

11. Carlson, H. J. and Campbell, R. E. Genetically encoded FRET-based biosensors for multiparameter fluorescence imaging. *Curr Opin Biotechnol* **2009**, *20*, 19-27.

12. Topell, S., Hennecke, J. and Glockshuber, R. Circularly permuted variants of the green fluorescent protein. *FEBS Lett* **1999**, *457*, 283-289.

13. Carlson, H. J., Cotton, D. W. and Campbell, R. E. Circularly permuted monomeric red fluorescent proteins with new termini in the beta-sheet. *Protein Science* **2010**, *19*, 1490-1499.

14. Zhao, Y., Araki, S., Wu, J., Teramoto, T., Chang, Y. F., Nakano, M., Abdelfattah, A. S., Fujiwara, M., Ishihara, T., Nagai, T. and Campbell, R. E. An expanded palette of genetically encoded Ca2+ indicators. *Science* **2011**, *333*, 1888-1891.

15. Nagai, T., Sawano, A., Park, E. S. and Miyawaki, A. Circularly permuted green fluorescent proteins engineered to sense Ca2+. *Proc Natl Acad Sci U S A* **2001**, *98*, 3197-3202.

16. Nakai, J., Ohkura, M. and Imoto, K. A high signal-to-noise Ca2+ probe composed of a single green fluorescent protein. *Nat Biotechnol* **2001**, *19*, 137-141.

17. Souslova, E. A., Belousov, V. V., Lock, J. G., Stromblad, S., Kasparov, S., Bolshakov, A. P., Pinelis, V. G., Labas, Y. A., Lukyanov, S., Mayr, L. M. and Chudakov, D. M. Single fluorescent protein-based Ca2+ sensors with increased dynamic range. *BMC Biotechnol* **2007**, *7*, 37.

18. Wang, Q., Shui, B., Kotlikoff, M. I. and Sondermann, H. Structural basis for calcium sensing by GCaMP2. *Structure* **2008**, *16*, 1817-1827.

19. Akerboom, J., Rivera, J. D., Guilbe, M. M., Malave, E. C., Hernandez, H. H., Tian, L., Hires, S. A., Marvin, J. S., Looger, L. L. and Schreiter, E. R. Crystal structures of the GCaMP calcium sensor reveal the mechanism of fluorescence signal change and aid rational design. *J Biol Chem* **2009**, *284*, 6455-6464.

20. Akerboom, J., Chen, T. W., Wardill, T. J., Tian, L., Marvin, J. S., Mutlu, S., Calderon, N. C., Esposti, F., Borghuis, B. G., Sun, X. R., Gordus, A., Orger, M. B., Portugues, R., Engert, F., Macklin, J. J., Filosa, A., Aggarwal, A., Kerr, R. A., Takagi, R., Kracun, S., Shigetomi, E., Khakh, B. S., Baier, H., Lagnado, L., Wang, S. S., Bargmann, C. I., Kimmel, B. E., Jayaraman, V., Svoboda, K., Kim, D. S., Schreiter, E. R. and Looger, L. L. Optimization of a GCaMP Calcium Indicator for Neural Activity Imaging. *J Neurosci* **2012**, *32*, 13819-13840.

21. Shcherbo, D., Merzlyak, E. M., Chepurnykh, T. V., Fradkov, A. F., Ermakova, G. V., Solovieva, E. A., Lukyanov, K. A., Bogdanova, E. A., Zaraisky, A. G., Lukyanov, S. and Chudakov, D. M. Bright far-red fluorescent protein for whole-body imaging. *Nature methods* **2007**, *4*, 741-746.

22. Gautam, S. G., Perron, A., Mutoh, H. and Knopfel, T. Exploration of fluorescent protein voltage probes based on circularly permuted fluorescent proteins. *Front Neuroeng* **2009**, *2*, 14.

23. Li, Y., Sierra, A. M., Ai, H. W. and Campbell, R. E. Identification of sites within a monomeric red fluorescent protein that tolerate peptide insertion and testing of corresponding circular permutations. *Photochem Photobiol* **2008**, *84*, 111-119.

24. Matz, M. V., Fradkov, A. F., Labas, Y. A., Savitsky, A. P., Zaraisky, A. G., Markelov, M. L. and Lukyanov, S. A. Fluorescent proteins from nonbioluminescent Anthozoa species. *Nat Biotechnol* **1999**, *17*, 969-973.

25. Shaner, N. C., Campbell, R. E., Steinbach, P. A., Giepmans, B. N., Palmer, A. E. and Tsien, R. Y. Improved monomeric red, orange and yellow fluorescent proteins derived from Discosoma sp. red fluorescent protein. *Nat Biotechnol* **2004**, *22*, 1567-1572.

26. Niwa, H., Inouye, S., Hirano, T., Matsuno, T., Kojima, S., Kubota, M., Ohashi, M. and Tsuji, F. I. Chemical nature of the light emitter of the Aequorea green fluorescent protein. *Proc Natl Acad Sci U S A* **1996**, *93*, 13617-13622.

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

28. Shimomura, O. Structure of the chromophore of Aequorea green fluorescent protein. *FEBS Lett* **1979**, *104*, 220-222.

29. Cody, C. W., Prasher, D. C., Westler, W. M., Prendergast, F. G. and Ward, W. W. Chemical structure of the hexapeptide chromophore of the Aequorea green-fluorescent protein. *Biochemistry* **1993**, *32*, 1212-1218.

30. Nagai, T., Yamada, S., Tominaga, T., Ichikawa, M. and Miyawaki, A. Expanded dynamic range of fluorescent indicators for Ca(2+) by circularly permuted yellow fluorescent proteins. *Proc Natl Acad Sci U S A* **2004**, *101*, 10554-10559.

31. Tallini, Y. N., Ohkura, M., Choi, B. R., Ji, G. J., Imoto, K., Doran, R., Lee, J., Plan, P., Wilson, J., Xin, H. B., Sanbe, A., Gulick, J., Mathai, J., Robbins, J., Salama, G., Nakai, J. and Kotlikoff, M. I. Imaging cellular signals in the heart in vivo: Cardiac expression of the high-signal Ca2+ indicator GCaMP2. *Proc Natl Acad Sci U S A* **2006**, *103*, 4753-4758.

32. Tian, L., Hires, S. A., Mao, T., Huber, D., Chiappe, M. E., Chalasani, S. H., Petreanu, L., Akerboom, J., McKinney, S. A., Schreiter, E. R., Bargmann, C. I., Jayaraman, V., Svoboda, K. and Looger, L. L. Imaging neural activity in worms, flies and mice with improved GCaMP calcium indicators. *Nat Methods* **2009**, *6*, 875-881.

33. Siegel, M. S. and Isacoff, E. Y. A genetically encoded optical probe of membrane voltage. *Neuron* **1997**, *19*, 735-741.

34. Ataka, K. and Pieribone, V. A. A genetically targetable fluorescent probe of channel gating with rapid kinetics. *Biophys J* **2002**, *82*, 509-516.

35. Lundby, A., Mutoh, H., Dimitrov, D., Akemann, W. and Knopfel, T. Engineering of a genetically encodable fluorescent voltage sensor exploiting fast Ci-VSP voltage-sensing movements. *Plos One* **2008**, *3*,

36. Hung, Y. P., Albeck, J. G., Tantama, M. and Yellen, G. Imaging Cytosolic NADH-NAD(+) Redox State with a Genetically Encoded Fluorescent Biosensor. *Cell Metab* **2011**, *14*, 545-554.

37. Belousov, V. V., Fradkov, A. F., Lukyanov, K. A., Staroverov, D. B., Shakhbazov, K. S., Terskikh, A. V. and Lukyanov, S. Genetically encoded fluorescent indicator for intracellular hydrogen peroxide. *Nat Methods* **2006**, *3*, 281-286.

38. Berg, J., Hung, Y. P. and Yellen, G. A genetically encoded fluorescent reporter of ATP:ADP ratio. *Nat Methods* **2009**, *6*, 161-166.

39. Kawai, Y., Sato, M. and Umezawa, Y. Single color fluorescent indicators of protein phosphorylation for multicolor imaging of intracellular signal flow dynamics. *Anal Chem* **2004**, *76*, 6144-6149.

40. Nausch, L. W. M., Lecloux, J., Bonev, A. D., Nelson, M. T. and Dostmann, W. R. Differential patterning of cGMP in vascular smooth muscle cells revealed by single GFP-linked biosensors. *Proc Natl Acad Sci U S A* **2008**, *105*, 365-370.

41. Barnett, L., Platisa, J., Popovic, M., Pieribone, V. A. and Hughes, T. A fluorescent, genetically-encoded voltage probe capable of resolving action potentials. *Plos One* **2012**, *7*, e43454.

42. Shui, B., Wang, Q., Lee, F., Byrnes, L. J., Chudakov, D. M., Lukyanov, S. A., Sondermann, H. and Kotlikoff, M. I. Circular permutation of red fluorescent proteins. *Plos One* **2011**, *6*,

43. Campbell, R. E., Tour, O., Palmer, A. E., Steinbach, P. A., Baird, G. S., Zacharias, D. A. and Tsien, R. Y. A monomeric red fluorescent protein. *Proc Natl Acad Sci U S A* **2002**, *99*, 7877-7882.

44. Shcherbakova, D. M., Subach, O. M. and Verkhusha, V. V. Red fluorescent proteins: advanced imaging applications and future design. *Angew Chem Int Ed Engl* **2012**, *51*, 10724-10738.

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

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

47. Heim, R., Prasher, D. C. and Tsien, R. Y. Wavelength mutations and posttranslational autoxidation of green fluorescent protein. *Proc Natl Acad Sci U S A* **1994**, *91*, 12501-12504.

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

49. Day, R. N. and Davidson, M. W. The fluorescent protein palette: tools for cellular imaging. *Chem Soc Rev* **2009**, *38*, 2887-2921.

50. Delagrave, S., Hawtin, R. E., Silva, C. M., Yang, M. M. and Youvan, D. C. Red-shifted excitation mutants of the green fluorescent protein. *Biotechnology (N Y)* **1995**, *13*, 151-154.

51. Tsien, R. Y. The green fluorescent protein. *Annu Rev Biochem* **1998**, *67*, 509-544.

52. Cubitt, A. B., Heim, R., Adams, S. R., Boyd, A. E., Gross, L. A. and Tsien, R. Y. Understanding, improving and using green fluorescent proteins. *Trends Biochem Sci* **1995**, *20*, 448-455.

53. Reid, B. G. and Flynn, G. C. Chromophore formation in green fluorescent protein. *Biochemistry* **1997**, *36*, 6786-6791.

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

55. Rosenow, M. A., Huffman, H. A., Phail, M. E. and Wachter, R. M. The crystal structure of the Y66L variant of green fluorescent protein supports a cyclization-oxidation-dehydration mechanism for chromophore maturation. *Biochemistry* **2004**, *43*, 4464-4472.

56. Rosenow, M. A., Patel, H. N. and Wachter, R. M. Oxidative chemistry in the GFP active site leads to covalent cross-linking of a modified leucine side chain with a histidine imidazole: Implications for the mechanism of chromophore formation. *Biochemistry* **2005**, *44*, 8303-8311.

57. Barondeau, D. P., Kassmann, C. J., Tainer, J. A. and Getzoff, E. D. Understanding GFP posttranslational chemistry: structures of designed variants that achieve backbone fragmentation, hydrolysis, and decarboxylation. *J Am Chem Soc* **2006**, *128*, 4685-4693.

58. Strack, R. L., Strongin, D. E., Mets, L., Glick, B. S. and Keenan, R. J. Chromophore formation in DsRed occurs by a branched pathway. *J Am Chem Soc* **2010**, *132*, 8496-8505.

59. Baird, G. S., Zacharias, D. A. and Tsien, R. Y. Biochemistry, mutagenesis, and oligomerization of DsRed, a red fluorescent protein from coral. *Proc Natl Acad Sci U S A* **2000**, *97*, 11984-11989.

60. Terskikh, A., Fradkov, A., Ermakova, G., Zaraisky, A., Tan, P., Kajava, A. V., Zhao, X., Lukyanov, S., Matz, M., Kim, S., Weissman, I. and Siebert, P. "Fluorescent timer": protein that changes color with time. *Science* **2000**, *290*, 1585-1588.

61. Remington, S. J. Fluorescent proteins: maturation, photochemistry and photophysics. *Curr Opin Struct Biol* **2006**, *16*, 714-721.

62. Verkhusha, V. V., Chudakov, D. M., Gurskaya, N. G., Lukyanov, S. and Lukyanov, K. A. Common pathway for the red chromophore formation in fluorescent proteins and chromoproteins. *Chem Biol* **2004**, *11*, 845-854.

63. Subach, F. V., Subach, O. M., Gundorov, I. S., Morozova, K. S., Piatkevich, K. D., Cuervo, A. M. and Verkhusha, V. V. Monomeric fluorescent timers that change color from blue to red report on cellular trafficking. *Nat Chem Biol* **2009**, *5*, 118-126.

64. Pletnev, S., Subach, F. V., Dauter, Z., Wlodawer, A. and Verkhusha, V. V. Understanding Blue-to-Red Conversion in Monomeric Fluorescent Timers and Hydrolytic Degradation of Their Chromophores. *J Am Chem Soc* **2010**, *132*, 2243-2253.

65. Ormo, M., Cubitt, A. B., Kallio, K., Gross, L. A., Tsien, R. Y. and Remington, S. J. Crystal structure of the Aequorea victoria green fluorescent protein. *Science* **1996**, *273*, 1392-1395.

66. Yarbrough, D., Wachter, R. M., Kallio, K., Matz, M. V. and Remington, S. J. Refined crystal structure of DsRed, a red fluorescent protein from coral, at 2.0-angstrom resolution. *Proc Natl Acad Sci U S A* **2001**, *98*, 462-467.

67. Wall, M. A., Socolich, M. and Ranganathan, R. The structural basis for red fluorescence in the tetrameric GFP homolog DsRed. *Nat Struct Biol* **2000**, *7*, 1133-1138.

68. gross, L. A. b., g.s.; hoffman, r.c.; baldridge, k.k & tsien, r.y. The structure of the chromophore within DsRed, a red fluorescent protein from coral. *Proc Natl Acad Sci U S A* **2000**, *97*, 119900-111995.

69. Vrzheshch, P. V., Akovbian, N. A., Varfolomeyev, S. D. and Verkhusha, V. V. Denaturation and partial renaturation of a tightly tetramerized DsRed protein under mildly acidic conditions. *FEBS Lett* **2000**, *487*, 203-208.

70. Shagin, D. A., Barsova, E. V., Yanushevich, Y. G., Fradkov, A. F., Lukyanov, K. A., Labas, Y. A., Semenova, T. N., Ugalde, J. A., Meyers, A., Nunez, J. M., Widder, E. A., Lukyanov, S. A. and Matz, M. V. GFP-like proteins as ubiquitous metazoan superfamily: evolution of functional features and structural complexity. *Mol Biol Evol* **2004**, *21*, 841-850.

71. Lauf, U., Lopez, P. and Falk, M. M. Expression of fluorescently tagged connexins: a novel approach to rescue function of oligomeric DsRed-tagged proteins. *FEBS Lett* **2001**, *498*, 11-15.

72. Gavin, P., Devenish, R. J. and Prescott, M. An approach for reducing unwanted oligomerisation of DsRed fusion proteins. *Biochem Biophys Res Commun* **2002**, *298*, 707-713.

73. Soling, A., Simm, A. and Rainov, N. Intracellular localization of Herpes simplex virus type 1 thymidine kinase fused to different fluorescent proteins depends on choice of fluorescent tag. *FEBS Lett* **2002**, *527*, 153-158.

74. Ai, H. W., Henderson, J. N., Remington, S. J. and Campbell, R. E. Directed evolution of a monomeric, bright and photostable version of Clavularia cyan fluorescent protein: structural characterization and applications in fluorescence imaging. *Biochem J* **2006**, *400*, 531-540.

75. Karasawa, S., Araki, T., Yamamoto-Hino, M. and Miyawaki, A. A green-emitting fluorescent protein from Galaxeidae coral and its monomeric version for use in fluorescent labeling. *J Biol Chem* **2003**, *278*, 34167-34171.

76. Karasawa, S., Araki, T., Nagai, T., Mizuno, H. and Miyawaki, A. Cyan-emitting and orange-emitting fluorescent proteins as a donor/acceptor pair for fluorescence resonance energy transfer. *Biochem J* **2004**, *381*, 307-312.

77. Henderson, J. N., Osborn, M. F., Koon, N., Gepshtein, R., Huppert, D. and Remington, S. J. Excited State Proton Transfer in the Red Fluorescent Protein mKeima. *J Am Chem Soc* **2009**, *131*, 13212-+.

78. Merzlyak, E. M., Goedhart, J., Shcherbo, D., Bulina, M. E., Shcheglov, A. S., Fradkov, A. F., Gaintzeva, A., Lukyanov, K. A., Lukyanov, S., Gadella, T. W. and Chudakov, D. M. Bright monomeric red fluorescent protein with an extended fluorescence lifetime. *Nat Methods* **2007**, *4*, 555-557.

79. Kredel, S., Oswald, F., Nienhaus, K., Deuschle, K., Rocker, C., Wolff, M., Heilker, R., Nienhaus, G. U. and Wiedenmann, J. mRuby, a Bright Monomeric Red Fluorescent Protein for Labeling of Subcellular Structures. *Plos One* **2009**, *4*,

80. Gurskaya, N. G., Fradkov, A. F., Terskikh, A., Matz, M. V., Labas, Y. A., Martynov, V. I., Yanushevich, Y. G., Lukyanov, K. A. and Lukyanov, S. A. GFP-like chromoproteins as a source of far-red fluorescent proteins. *FEBS Lett* **2001**, *507*, 16-20.

81. Wang, L., Jackson, W. C., Steinbach, P. A. and Tsien, R. Y. Evolution of new nonantibody proteins via iterative somatic hypermutation. *Proc Natl Acad Sci U S A* **2004**, *101*, 16745-16749.

82. Ward, W. W. C., C; Hart, R; Cormier, M. Spectrophotometric identity of the energy transfer chromophores in Renilla and Aequorea green-fluorescent proteins. *Photochem Photobiol* **1980**, *31*, 611-615.

83. Brejc, K., Sixma, T. K., Kitts, P. A., Kain, S. R., Tsien, R. Y., Ormo, M. and Remington, S. J. Structural basis for dual excitation and photoisomerization of the Aequorea victoria green fluorescent protein. *Proc Natl Acad Sci U S A* **1997**, *94*, 2306-2311.

84. Llopis, J., McCaffery, J. M., Miyawaki, A., Farquhar, M. G. and Tsien, R. Y. Measurement of cytosolic, mitochondrial, and Golgi pH in single living cells with green fluorescent proteins. *Proc Natl Acad Sci U S A* **1998**, *95*, 6803-6808.

85. Kneen, M., Farinas, J., Li, Y. X. and Verkman, A. S. Green fluorescent protein as a non-invasive intracellular pH indicator. *Biophys J* **1998**, *74*, A181-A181.

86. Elsliger, M. A., Wachter, R. M., Hanson, G. T., Kallio, K. and Remington, S. J. Structural and spectral response of green fluorescent protein variants to changes in pH. *Biochemistry* **1999**, *38*, 5296-5301.

87. Griesbeck, O., Baird, G. S., Campbell, R. E., Zacharias, D. A. and Tsien, R. Y. Reducing the environmental sensitivity of yellow fluorescent protein. Mechanism and applications. *J Biol Chem* **2001**, *276*, 29188-29194.

88. Rekas, A., Alattia, J. R., Nagai, T., Miyawaki, A. and Ikura, M. Crystal structure of venus, a yellow fluorescent protein with improved maturation and reduced environmental sensitivity. *J Biol Chem* **2002**, *277*, 50573-50578.

89. Zapata-Hommer, O. and Griesbeck, O. Efficiently folding and circularly permuted variants of the Sapphire mutant of GFP. *BMC Biotechnol* **2003**, *3*,

90. Whitaker, J. E., Haugland, R. P. and Prendergast, F. G. Spectral and photophysical studies of benzo[c]xanthene dyes: dual emission pH sensors. *Anal Biochem* **1991**, *194*, 330-344.

91. Hanson, G. T., McAnaney, T. B., Park, E. S., Rendell, M. E., Yarbrough, D. K., Chu, S., Xi, L., Boxer, S. G., Montrose, M. H. and Remington, S. J. Green fluorescent protein variants as ratiometric dual emission pH sensors. 1. Structural characterization and preliminary application. *Biochemistry* **2002**, *41*, 15477-15488.

92. McAnaney, T. B., Park, E. S., Hanson, G. T., Remington, S. J. and Boxer, S. G. Green fluorescent protein variants as ratiometric dual emission pH sensors. 2. Excited-state dynamics. *Biochemistry* **2002**, *41*, 15489-15494.

93. Galietta, L. J. V., Haggie, P. M. and Verkman, A. S. Green fluorescent protein-based halide indicators with improved chloride and iodide affinities. *FEBS Lett* **2001**, *499*, 220-224.

94. Shaner, N. C., Steinbach, P. A. and Tsien, R. Y. A guide to choosing fluorescent proteins. *Nat Methods* **2005**, *2*, 905-909.

95. Leung, D. W. C., E.Y.; Goeddel, D.V. A method for random mutagenesis of a defined DNA segment using a modified polymerase chain reaction. *Technique 1* **1989**, *1*, 11-15.

96. Cadwell, R. C. and Joyce, G. F. Randomization of genes by PCR mutagenesis. *PCR Methods Appl* **1992**, *2*, 28-33.

97. Taylor, J. W., Ott, J. and Eckstein, F. The rapid generation of oligonucleotide-directed mutations at high frequency using phosphorothioate-modified DNA. *Nucleic Acids Res* **1985**, *13*, 8765-8785.

98. Fromant, M., Blanquet, S. and Plateau, P. Direct Random Mutagenesis of Gene-Sized DNA Fragments Using Polymerase Chain-Reaction. *Anal Biochem* **1995**, *224*, 347-353.

99. Crameri, A., Whitehorn, E. A., Tate, E. and Stemmer, W. P. C. Improved green fluorescent protein by molecular evolution using DNA shuffling. *Nat Biotechnol* **1996**, *14*, 315-319.

100. Zimmermann, T., Rietdorf, J., Girod, A., Georget, V. and Pepperkok, R. Spectral imaging and linear un-mixing enables improved FRET efficiency with a novel GFP2-YFP FRET pair. *FEBS Lett* **2002**, *531*, 245-249.

101. Zimmermann, T., Rietdorf, J. and Pepperkok, R. Spectral imaging and its applications in live cell microscopy. *FEBS Lett* **2003**, *546*, 87-92.

102. Cormack, B. P., Valdivia, R. H. and Falkow, S. FACS-optimized mutants of the green fluorescent protein (GFP). *Gene* **1996**, *173*, 33-38.

103. Patterson, G. H., Knobel, S. M., Sharif, W. D., Kain, S. R. and Piston, D. W. Use of the green fluorescent protein and its mutants in quantitative fluorescence microscopy. *Biophys J* **1997**, *73*, 2782-2790.

104. Pedelacq, J. D., Cabantous, S., Tran, T., Terwilliger, T. C. and Waldo, G. S. Engineering and characterization of a superfolder green fluorescent protein. *Nat Biotechnol* **2006**, *24*, 79-88.

105. Miyawaki, A., Llopis, J., Heim, R., McCaffery, J. M., Adams, J. A., Ikura, M. and Tsien, R. Y. Fluorescent indicators for Ca2+ based on green fluorescent proteins and calmodulin. *Nature* **1997**, *388*, 882-887.

106. Ai, H. W., Shaner, N. C., Cheng, Z. H., Tsien, R. Y. and Campbell, R. E. Exploration of new chromophore structures leads to the identification of improved blue fluorescent proteins. *Biochemistry* **2007**, *46*, 5904-5910.

107. Ai, H. W., Olenych, S. G., Wong, P., Davidson, M. W. and Campbell, R. E. Hue-shifted monomeric variants of Clavularia cyan fluorescent protein: identification of the molecular determinants of color and applications in fluorescence imaging. *BMC Biol* **2008**, *6*, 13.

108. Nagai, T., Ibata, K., Park, E. S., Kubota, M., Mikoshiba, K. and Miyawaki, A. A variant of yellow fluorescent protein with fast and efficient maturation for cell-biological applications. *Nat Biotechnol* **2002**, *20*, 87-90.

109. Shaner, N. C., Lin, M. Z., McKeown, M. R., Steinbach, P. A., Hazelwood, K. L., Davidson, M. W. and Tsien, R. Y. Improving the photostability of bright monomeric orange and red fluorescent proteins. *Nat Methods* **2008**, *5*, 545-551.

110. Shcherbakova, D. M., Hink, M. A., Joosen, L., Gadella, T. W. and Verkhusha, V. V. An orange fluorescent protein with a large Stokes shift for single-excitation multicolor FCCS and FRET imaging. *J Am Chem Soc* **2012**, *134*, 7913-7923.

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

112. Van Thor, J. J. L., C.N.; Kellner, B.; Bourdakos, K.N.; Thompson, L.M.; Bearpark, M.J.; Champion, P.M; Sage, J.T. Ultrafast vibrational dynamics of parallel excited state proton transfer reactions in the Green Fluorescent Protein. *Vib Spectrosc* **2012**, *62*, 1-6.

113. Piatkevich, K. D., Hulit, J., Subach, O. M., Wu, B., Abdulla, A., Segall, J. E. and Verkhusha, V. V. Monomeric red fluorescent proteins with a large Stokes shift. *Proc Natl Acad Sci U S A* **2010**, *107*, 5369-5374.

114. Piatkevich, K. D., Malashkevich, V. N., Almo, S. C. and Verkhusha, V. V. Engineering ESPT Pathways Based on Structural Analysis of LSSmKate Red Fluorescent Proteins with Large Stokes Shift. *J Am Chem Soc* **2010**, *132*, 10762-10770.

115. Subach, O. M., Patterson, G. H., Ting, L. M., Wang, Y., Condeelis, J. S. and Verkhusha, V. V. A photoswitchable orange-to-far-red fluorescent protein, PSmOrange. *Nat Methods* **2011**, *8*, 771-777.

116. Subach, F. V., Patterson, G. H., Manley, S., Gillette, J. M., Lippincott-Schwartz, J. and Verkhusha, V. V. Photoactivatable mCherry for high-resolution two-color fluorescence microscopy. *Nat Methods* **2009**, *6*, 153-159.

117. Lippincott-Schwartz, J. and Patterson, G. H. Photoactivatable fluorescent proteins for diffraction-limited and super-resolution imaging. *Trends Cell Biol* **2009**, *19*, 555-565.

118. Dickson, R. M., Cubitt, A. B., Tsien, R. Y. and Moerner, W. E. On/off blinking and switching behaviour of single molecules of green fluorescent protein. *Nature* **1997**, *388*, 355-358.

119. Ai, H. W. and Campbell, R. E. Teal fluorescent proteins: Characterization of a reversibly photoswitchable variant. *Proc SPIE* **2008**, *6868*,

120. Henderson, J. N., Ai, H. W., Campbell, R. E. and Remington, S. J. Structural basis for reversible photobleaching of a green fluorescent protein homologue. *Proc Natl Acad Sci U S A* **2007**, *104*, 6672-6677.

121. Patterson, G. H. and Lippincott-Schwartz, J. A photoactivatable GFP for selective photolabeling of proteins and cells. *Science* **2002**, *297*, 1873-1877.

122. Subach, F. V., Patterson, G. H., Renz, M., Lippincott-Schwartz, J. and Verkhusha, V. V. Bright Monomeric Photoactivatable Red Fluorescent Protein for Two-Color Super-Resolution sptPALM of Live Cells. *J Am Chem Soc* **2010**, *132*, 6481-6491.

123. Gunewardene, M. S., Subach, F. V., Gould, T. J., Penoncello, G. P., Gudheti, M. V., Verkhusha, V. V. and Hess, S. T. Superresolution imaging of multiple fluorescent proteins with highly overlapping emission spectra in living cells. *Biophys J* **2011**, *101*, 1522-1528.

124. Fuchs, J., Bohme, S., Oswald, F., Hedde, P. N., Krause, M., Wiedenmann, J. and Nienhaus, G. U. A photoactivatable marker protein for pulse-chase imaging with superresolution. *Nat Methods* **2010**, *7*, 627-630.

125. Adam, V., Moeyaert, B., David, C. C., Mizuno, H., Lelimousin, M., Dedecker, P., Ando, R., Miyawaki, A., Michiels, J., Engelborghs, Y. and Hofkens, J. Rational design of photoconvertible and biphotochromic fluorescent proteins for advanced microscopy applications. *Chem Biol* **2011**, *18*, 1241-1251.

126. Chudakov, D. M., Lukyanov, S. and Lukyanov, K. A. Tracking intracellular protein movements using photoswitchable fluorescent proteins PS-CFP2 and Dendra2. *Nat Protoc* **2007**, *2*, 2024-2032.

127. McKinney, S. A., Murphy, C. S., Hazelwood, K. L., Davidson, M. W. and Looger, L. L. A bright and photostable photoconvertible fluorescent protein. *Nat Methods* **2009**, *6*, 131-133.

128. Habuchi, S., Tsutsui, H., Kochaniak, A. B., Miyawaki, A. and van Oijen, A. M. mKikGR, a monomeric photoswitchable fluorescent protein. *Plos One* **2008**, *3*, e3944.

129. McEvoy, A. L., Hoi, H., Bates, M., Platonova, E., Cranfill, P. J., Baird, M. A., Davidson, M. W., Ewers, H., Liphardt, J. and Campbell, R. E. mMaple: A photoconvertible fluorescent protein for use in multiple imaging modalities. *Plos One* **2012**, *7*, e51314.

130. Valentin, G., Verheggen, C., Piolot, T., Neel, H., Coppey-Moisan, M. and Bertrand, E. Photoconversion of YFP into a CFP-like species during acceptor photobleaching FRET experiments. *Nat Methods* **2005**, *2*, 801.

131. Sinnecker, D., Voigt, P., Hellwig, N. and Schaefer, M. Reversible photobleaching of enhanced green fluorescent proteins. *Biochemistry* **2005**, *44*, 7085-7094.

132. Cunningham, B. A., Hemperly, J. J., Hopp, T. P. and Edelman, G. M. Favin versus concanavalin A: Circularly permuted amino acid sequences. *Proc Natl Acad Sci U S A* **1979**, *76*, 3218-3222.

133. Lindqvist, Y. and Schneider, G. Circular permutations of natural protein sequences: structural evidence. *Curr Opin Struct Biol* **1997**, *7*, 422-427.

134. Uliel, S., Fliess, A. and Unger, R. Naturally occurring circular permutations in proteins. *Protein Eng* **2001**, *14*, 533-542.

135. Weiner, J., 3rd, Thomas, G. and Bornberg-Bauer, E. Rapid motif-based prediction of circular permutations in multi-domain proteins. *Bioinformatics* **2005**, *21*, 932-937.

136. Lo, W. C. and Lyu, P. C. CPSARST: an efficient circular permutation search tool applied to the detection of novel protein structural relationships. *Genome Biol* **2008**, *9*, R11.

137. Lo, W. C., Lee, C. C., Lee, C. Y. and Lyu, P. C. CPDB: a database of circular permutation in proteins. *Nucleic Acids Res* **2009**, *37*, D328-D332.

138. Eisenbeis, S. and Hocker, B. Evolutionary mechanism as a template for protein engineering. *J Pept Sci* **2010**, *16*, 538-544.

139. Ponting, C. P. and Russell, R. B. Swaposins: circular permutations within genes encoding saposin homologues. *Trends Biochem Sci* **1995**, *20*, 179-180.

140. Jeltsch, A. Circular permutations in the molecular evolution of DNA methyltransferases. *J Mol Evol* **1999**, *49*, 161-164.

141. Guruprasad, K., Tormakangas, K., Kervinen, J. and Blundell, T. L. Comparative modelling of barley-grain aspartic proteinase: a structural rationale for observed hydrolytic specificity. *FEBS Lett* **1994**, *352*, 131-136.

142. Tsai, L. C., Shyur, L. F., Lee, S. H., Lin, S. S. and Yuan, H. S. Crystal structure of a natural circularly permuted jellyroll protein: 1,3-1,4-beta-D-glucanase from Fibrobacter succinogenes. *J Mol Biol* **2003**, *330*, 607-620.

143. Ribeiro, E. A., Jr. and Ramos, C. H. Circular permutation and deletion studies of myoglobin indicate that the correct position of its N-terminus is required for native stability and solubility but not for native-like heme binding and folding. *Biochemistry* **2005**, *44*, 4699-4709.

144. Yu, Y. and Lutz, S. Circular permutation: a different way to engineer enzyme structure and function. *Trends Biotechnol* **2011**, *29*, 18-25.

145. Graf, R. and Schachman, H. K. Random circular permutation of genes and expressed polypeptide chains: application of the method to the catalytic chains of aspartate transcarbamoylase. *Proc Natl Acad Sci U S A* **1996**, *93*, 11591-11596.

146. Krishna, M. M. and Englander, S. W. The N-terminal to C-terminal motif in protein folding and function. *Proc Natl Acad Sci U S A* **2005**, *102*, 1053-1058.

147. Thornton, J. M. and Sibanda, B. L. Amino and carboxy-terminal regions in globular proteins. *J Mol Biol* **1983**, *167*, 443-460.

148. Nagi, A. D. and Regan, L. An inverse correlation between loop length and stability in a four-helix-bundle protein. *Fold Des* **1997**, *2*, 67-75.

149. Hu, X., Wang, H., Ke, H. and Kuhlman, B. High-resolution design of a protein loop. *Proc Natl Acad Sci U S A* **2007**, *104*, 17668-17673.

150. George, R. A. and Heringa, J. An analysis of protein domain linkers: their classification and role in protein folding. *Protein Eng* **2002**, *15*, 871-879.

151. Wriggers, W., Chakravarty, S. and Jennings, P. A. Control of protein functional dynamics by peptide linkers. *Biopolymers* **2005**, *80*, 736-746.

152. Akemann, W., Raj, C. D. and Knopfel, T. Functional characterization of permuted enhanced green fluorescent proteins comprising varying linker peptides. *Photochem Photobiol* **2001**, *74*, 356-363.

153. Ding, Y., Ai, H. W., Hoi, H. and Campbell, R. E. Forster resonance energy transfer-based biosensors for multiparameter ratiometric imaging of Ca2+ dynamics and caspase-3 activity in single cells. *Anal Chem* **2011**, *83*, 9687-9693.

154. Ibraheem, A., Yap, H., Ding, Y. and Campbell, R. E. A bacteria colony-based screen for optimal linker combinations in genetically encoded biosensors. *BMC Biotechnol* **2011**, *11*, 105.

155. van Dongen, E. M., Evers, T. H., Dekkers, L. M., Meijer, E. W., Klomp, L. W. and Merkx, M. Variation of linker length in ratiometric fluorescent sensor proteins allows rational tuning of Zn(II) affinity in the picomolar to femtomolar range. *J Am Chem Soc* **2007**, *129*, 3494-3495.

156. Waldo, G. S. S., B.M.; Berendzen, J.; Terwillegar, T.C. Rapid protein-folding assay using green fluorescent protein. *Nat Biotechnol* **1999**, *17*, 691-695.

157. Fukuda, H., Arai, M. and Kuwajima, K. Folding of green fluorescent protein and the cycle3 mutant. *Biochemistry* **2000**, *39*, 12025-12032.

158. Huang, J. R., Craggs, T. D., Christodoulou, J. and Jackson, S. E. Stable intermediate states and high energy barriers in the unfolding of GFP. *J Mol Biol* **2007**, *370*, 356-371.

159. Hisatomi, Y., Katagiri, D., Neya, S., Hara, M. and Hoshino, T. Analysis of the unfolding process of green fluorescent protein by molecular dynamics simulation. *J Phys Chem B* **2008**, *112*, 8672-8680.

160. Perez-Jimenez, R., Garcia-Manyes, S., Ainavarapu, S. R. and Fernandez, J. M. Mechanical unfolding pathways of the enhanced yellow fluorescent protein revealed by single molecule force spectroscopy. *J Biol Chem* **2006**, *281*, 40010-40014.

161. Abu-Lail, N. I., Ohashi, T., Clark, R. L., Erickson, H. P. and Zauscher, S. Understanding the elasticity of fibronectin fibrils: unfolding strengths of FN-III and GFP domains measured by single molecule force spectroscopy. *Matrix Biology* **2006**, *25*, 175-184.

162. Kane, P. M., Yamashiro, C. T., Wolczyk, D. F., Neff, N., Goebl, M. and Stevens, T. H. Protein splicing converts the yeast TFP1 gene product to the 69-kD subunit of the vacuolar H(+)-adenosine triphosphatase. *Science* **1990**, *250*, 651-657.

163. Perler, F. B., Davis, E. O., Dean, G. E., Gimble, F. S., Jack, W. E., Neff, N., Noren, C. J., Thorner, J. and Belfort, M. Protein splicing elements: inteins and exteins--a definition of terms and recommended nomenclature. *Nucleic Acids Res* **1994**, *22*, 1125-1127.

164. Perler, F. B. InBase: the Intein Database. *Nucleic Acids Res* **2002**, *30*, 383-384.

165. Liu, X. Q. Protein-splicing intein: Genetic mobility, origin, and evolution. *Annu Rev Genet* **2000**, *34*, 61-76.

166. Gimble, F. S. and Wang, J. Substrate recognition and induced DNA distortion by the PI-SceI endonuclease, an enzyme generated by protein splicing. *J Mol Biol* **1996**, *263*, 163-180.

167. Nishioka, M., Fujiwara, S., Takagi, M. and Imanaka, T. Characterization of two intein homing endonucleases encoded in the DNA polymerase gene of Pyrococcus kodakaraensis strain KOD1. *Nucleic Acids Res* **1998**, *26*, 4409-4412.

168. Saves, I., Ozanne, V., Dietrich, J. and Masson, J. M. Inteins of Thermococcus fumicolans DNA polymerase are endonucleases with distinct enzymatic behaviors. *J Biol Chem* **2000**, *275*, 2335-2341.

169. Pietrokovski, S. Identification of a virus intein and a possible variation in the protein-splicing reaction. *Curr Biol* **1998**, *8*, R634-635.

170. Pietrokovski, S. Conserved sequence features of inteins (protein introns) and their use in identifying new inteins and related proteins. *Protein Science* **1994**, *3*, 2340-2350.

171. Perler, F. B., Olsen, G. J. and Adam, E. Compilation and analysis of intein sequences. *Nucleic Acids Res* **1997**, *25*, 1087-1093.

172. Elleuche, S. and Poggeler, S. Inteins, valuable genetic elements in molecular biology and biotechnology. *Appl Microbiol Biotechnol* **2010**, *87*, 479-489.

173. Xu, M. Q. and Perler, F. B. The mechanism of protein splicing and its modulation by mutation. *EMBO J* **1996**, *15*, 5146-5153.

174. Paulus, H. Protein splicing and related forms of protein autoprocessing. *Annu Rev Biochem* **2000**, *69*, 447-496.

175. Wu, H., Hu, Z. and Liu, X. Q. Protein trans-splicing by a split intein encoded in a split DnaE gene of Synechocystis sp. PCC6803. *Proc Natl Acad Sci U S A* **1998**, *95*, 9226-9231.

176. Iwai, H., Zuger, S., Jin, J. and Tam, P. H. Highly efficient protein trans-splicing by a naturally split DnaE intein from Nostoc punctiforme. *FEBS Lett* **2006**, *580*, 1853-1858.

177. Aranko, A. S., Zuger, S., Buchinger, E. and Iwai, H. In vivo and in vitro proteinlLigation by naturally occurring and engineered split DnaE inteins. *Plos One* **2009**, *4*,

178. Zuger, S. and Iwai, H. Intein-based biosynthetic incorporation of unlabeled protein tags into isotopically labeled proteins for NMR studies. *Nat Biotechnol* **2005**, *23*, 736-740.

179. Southworth, M. W., Adam, E., Panne, D., Byer, R., Kautz, R. and Perler, F. B. Control of protein splicing by intein fragment reassembly. *EMBO J* **1998**, *17*, 918-926.

180. Mills, K. V., Lew, B. M., Jiang, S. and Paulus, H. Protein splicing in trans by purified N- and C-terminal fragments of the Mycobacterium tuberculosis RecA intein. *Proc Natl Acad Sci U S A* **1998**, *95*, 3543-3548.

181. Ozawa, T., Nogami, S., Sato, M., Ohya, Y. and Umezawa, Y. A fluorescent indicator for detecting protein-protein interactions in vivo based on protein splicing. *Anal Chem* **2000**, *72*, 5151-5157.

182. Iwai, H., Lingel, A. and Pluckthun, A. Cyclic green fluorescent protein produced in vivo using an artificially split PI-PfuI intein from Pyrococcus furiosus. *J Biol Chem* **2001**, *276*, 16548-16554.

183. Xu, R., Ayers, B., Cowburn, D. and Muir, T. W. Chemical ligation of folded recombinant proteins: segmental isotopic labeling of domains for NMR studies. *Proc Natl Acad Sci U S A* **1999**, *96*, 388-393.

184. Lee, Y. T., Su, T. H., Lo, W. C., Lyu, P. C. and Sue, S. C. Circular permutation prediction reveals a viable backbone disconnection for split proteins: an approach in identifying a new functional split intein. *Plos One* **2012**, *7*, e43820.

185. Lo, W. C., Wang, L. F., Liu, Y. Y., Dai, T., Hwang, J. K. and Lyu, P. C. CPred: a web server for predicting viable circular permutations in proteins. *Nucleic Acids Res* **2012**, *40*, W232-237.

186. Lo, W. C., Dai, T., Liu, Y. Y., Wang, L. F., Hwang, J. K. and Lyu, P. C. Deciphering the preference and predicting the viability of circular permutations in proteins. *Plos One* **2012**, *7*, e31791.

187. Kanno, A., Ozawa, T. and Umezawa, Y. Detection of protein-protein interactions in bacteria by GFP-fragment reconstitution. *Methods Mol Biol* **2011**, *705*, 251-258.

188. Zhao, Z. L., W; Dun, B; Jin, D; Ping, S; Zhang, W; Chen, M; Xu, M; Lin, M. Purification of green fluorescent protein using a two-intein system. *Appl Microbiol Biotechnol* **2008**, *77*, 1175-1180.

189. Ramsden, R., Arms, L., Davis, T. N. and Muller, E. G. D. An intein with genetically selectable markers provides a new approach to internally label proteins with GFP. *BMC Biotechnol* **2011**, *11*,

190. Wong, S. S., Kotera, I., Mills, E., Suzuki, H. and Truong, K. Split-intein mediated re-assembly of genetically encoded Ca(2+) indicators. *Cell Calcium* **2012**, *51*, 57-64.

191. Camire, O. and Topolnik, L. Functional compartmentalisation and regulation of postsynaptic Ca2+ transients in inhibitory interneurons. *Cell Calcium* **2012**, *52*, 339-346.

192. Krosnowski, K., Ashby, S., Sathyanesan, A., Luo, W., Ogura, T. and Lin, W. Diverse populations of intrinsic cholinergic interneurons in the mouse olfactory bulb. *Neuroscience* **2012**, *213*, 161-178.

193. Alford, S. C., Abdelfattah, A. S., Ding, Y. D. and Campbell, R. E. A Fluorogenic Red Fluorescent Protein Heterodimer. *Chem Biol* **2012**, *19*, 353-360.

194. Alford, S. C. D., Y.; Simmen, T.; Campbell, R.E. dimerization-dependent green and yellow fluorescent proteins. *ACS Synth Biol* **2012**, *1*, 569-575.

195. Johnsson, N. and Varshavsky, A. Split ubiquitin as a sensor of protein interactions in vivo. *Proc Natl Acad Sci U S A* **1994**, *91*, 10340-10344.

196. Pelletier, J. N., Campbell-Valois, F. X. and Michnick, S. W. Oligomerization domain-directed reassembly of active dihydrofolate reductase from rationally designed fragments. *Proc Natl Acad Sci U S A* **1998**, *95*, 12141-12146.

197. Rossi, F., Charlton, C. A. and Blau, H. M. Monitoring protein-protein interactions in intact eukaryotic cells by beta-galactosidase complementation. *Proc Natl Acad Sci U S A* **1997**, *94*, 8405-8410.

198. Hu, C. D., Chinenov, Y. and Kerppola, T. K. Visualization of interactions among bZip and Rel family proteins in living cells using bimolecular fluorescence complementation. *Mol Cell* **2002**, *9*, 789-798.

199. Wehrman, T., Kleaveland, B., Her, J. H., Balint, R. F. and Blau, H. M. Protein-protein interactions monitored in mammalian cells via complementation of beta -lactamase enzyme fragments. *Proc Natl Acad Sci U S A* **2002**, *99*, 3469-3474.

200. Paulmurugan, R., Umezawa, Y. and Gambhir, S. S. Noninvasive imaging of protein-protein interactions in living subjects by using reporter protein complementation and reconstitution strategies. *Proc Natl Acad Sci U S A* **2002**, *99*, 15608-15613.

201. Kerppola, T. K. Visualization of molecular interactions by fluorescence complementation. *Nat Rev Mol Cell Bio* **2006**, *7*, 449-456.

202. Kerppola, T. K. Bimolecular fluorescence complementation (BiFC) analysis as a probe of protein interactions in living cells. *Annu Rev Biophys* **2008**, *37*, 465-487.

203. Kerppola, T. K. Visualization of molecular interactions using bimolecular fluorescence complementation analysis: characteristics of protein fragment complementation. *Chem Soc Rev* **2009**, *38*, 2876-2886.

204. Craggs, T. D. Green fluorescent protein: structure, folding and chromophore maturation. *Chem Soc Rev* **2009**, *38*, 2865-2875.

205. Ghosh, I., Hamilton, A. D. and Regan, L. Antiparallel leucine zipper-directed protein reassembly: Application to the green fluorescent protein. *J Am Chem Soc* **2000**, *122*, 5658-5659.

206. Hu, C. D. and Kerppola, T. K. Simultaneous visualization of multiple protein interactions in living cells using multicolor fluorescence complementation analysis. *Nat Biotechnol* **2003**, *21*, 539-545.

207. Jach, G., Pesch, M., Richter, K., Frings, S. and Uhrig, J. F. An improved mRFP1 adds red to bimolecular fluorescence complementation. *Nat Methods* **2006**, *3*, 597-600.

208. Shyu, Y. J. and Hu, C. D. Fluorescence complementation: an emerging tool for biological research. *Trends Biotechnol* **2008**, *26*, 622-630.

209. Saka, Y., Hagemann, A. I., Piepenburg, O. and Smith, J. C. Nuclear accumulation of Smad complexes occurs only after the midblastula transition in Xenopus. *Development* **2007**, *134*, 4209-4218.

210. Rackham, O. and Brown, C. M. Visualization of RNA-protein interactions in living cells: FMRP and IMP1 interact on mRNAs. *EMBO J* **2004**, *23*, 3346-3355.

211. Robida, A. M. and Kerppola, T. K. Bimolecular fluorescence complementation analysis of inducible protein interactions: effects of factors affecting protein folding on fluorescent protein fragment association. *J Mol Biol* **2009**, *394*, 391-409.

212. Heim, R., Cubitt, A. B. and Tsien, R. Y. Improved green fluorescence. *Nature* **1995**, *373*, 663-664.

213. Zhao, H. M., Giver, L., Shao, Z. X., Affholter, J. A. and Arnold, F. H. Molecular evolution by staggered extension process (StEP) in vitro recombination. *Nat Biotechnol* **1998**, *16*, 258-261.

214. Hutchison, C. A., 3rd, Phillips, S., Edgell, M. H., Gillam, S., Jahnke, P. and Smith, M. Mutagenesis at a specific position in a DNA sequence. *J Biol Chem* **1978**, *253*, 6551-6560.

215. Stemmer, W. P. C. DNA shuffling by random fragmentation and reassembly - in-vitro recombination for molecular evolution. *Proc Natl Acad Sci U S A* **1994**, *91*, 10747-10751.

216. Stemmer, W. P. C. Rapid Evolution of a Protein in-Vitro by DNA Shuffling. *Nature* **1994**, *370*, 389-391.

217. McCullum, E. O., Williams, B. A., Zhang, J. and Chaput, J. C. Random mutagenesis by error-prone PCR. *Methods Mol Biol* **2010**, *634*, 103-109.

218. Drummond, D. A., Iverson, B. L., Georgiou, G. and Arnold, F. H. Why high-error-rate random mutagenesis libraries are enriched in functional and improved proteins. *J Mol Biol* **2005**, *350*, 806-816.

219. Ottmann, C., Weyand, M., Wolf, A. and Kuhlmann, J. Applicability of superfolder YFP bimolecular fluorescence complementation in vitro. *Biol Chem* **2009**, *390*, 81-90.

220. Subach, F. V., Zhang, L. J., Gadella, T. W. J., Gurskaya, N. G., Lukyanov, K. A. and Verkhusha, V. V. Red Fluorescent Protein with Reversibly Photoswitchable Absorbance for Photochromic FRET. *Chem Biol* **2010**, *17*, 745-755.

221. Kirsch, R. D. and Joly, E. An improved PCR-mutagenesis strategy for two-site mutagenesis or sequence swapping between related genes. *Nucleic Acids Res* **1998**, *26*, 1848-1850.

222. Aharoni, A., Griffiths, A. D. and Tawfik, D. S. High-throughput screens and selections of enzyme-encoding genes. *Curr Opin Chem Biol* **2005**, *9*, 210-216.

223. Lin, H. and Cornish, V. W. Screening and selection methods for large-scale analysis of protein function. *Angew Chem Int Ed Engl* **2002**, *41*, 4402-4425.

224. Moore, J. C., Jin, H. M., Kuchner, O. and Arnold, F. H. Strategies for the in vitro evolution of protein function: Enzyme evolution by random recombination of improved sequences. *J Mol Biol* **1997**, *272*, 336-347.

225. Schuster, S., Enzelberger, M., Trauthwein, H., Schmid, R. D. and Urlacher, V. B. pHluorin-based in vivo assay for hydrolase screening. *Anal Chem* **2005**, *77*, 2727-2732.

226. Arnold, F. H. G., G. *Directed Evolution Library Creation Methods and Protocols in Methods in Molecular Biology*, 2003, Humana Press, Inc., Totowa,

227. Labrou, N. E. Random mutagenesis methods for in vitro directed enzyme evolution. *Curr Protein Pept Sci* **2010**, *11*, 91-100.

228. Fisher, A. C. M., T.J.; DeLisa, M.P. Protein folding and solubility: pathways and high-throughput assays S. B. Lutz, U.T, *Protein Engineering Handbook*, 2009, Wiley-VCH, 121-145.

229. Sui, J., Cotard, S., Andersen, J., Zhu, P., Staunton, J., Lee, M. and Lin, S. Optimization of a Yellow fluorescent protein-based iodide influx high-throughput screening assay for cystic fibrosis transmembrane conductance regulator (CFTR) modulators. *Assay Drug Dev Technol* **2010**, *8*, 656-668.

230. Pedemonte, N., Zegarra-Moran, O. and Galietta, L. J. High-throughput screening of libraries of compounds to identify CFTR modulators. *Methods Mol Biol* **2011**, *741*, 13-21.

231. Schmidt-Dannert, C. and Arnold, F. H. Directed evolution of industrial enzymes. *Trends Biotechnol* **1999**, *17*, 135-136.

232. Salih, A., Larkum, A., Cox, G., Kuhl, M. and Hoegh-Guldberg, O. Fluorescent pigments in corals are photoprotective. *Nature* **2000**, *408*, 850-853.

233. Leutenegger, A., Kredel, S., Gundel, S., D'Angelo, C., Salih, A. and Wiedenmann, J. Analysis of fluorescent and non-fluorescent sea anemones from

the Mediterranean Sea during a bleaching event. *J Exp Mar Biol Ecol* **2007**, *353*, 221-234.

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

235. Muller-Taubenberger, A. and Anderson, K. I. Recent advances using green and red fluorescent protein variants. *Appl Microbiol Biotechnol* **2007**, *77*, 1-12.

236. Author; 22 July 1999. Protein derived from Anemonia useful in gene technology and molecular biology D. P.-u. Markenamt Germany

237. Kogure, T., Karasawa, S., Araki, T., Saito, K., Kinjo, M. and Miyawaki, A. A fluorescent variant of a protein from the stony coral Montipora facilitates dual-color single-laser fluorescence cross-correlation spectroscopy. *Nat Biotechnol* **2006**, *24*, 577-581.

238. Fischer, M., Haase, I., Simmeth, E., Gerisch, G. and Muller-Taubenberger, A. A brilliant monomeric red fluorescent protein to visualize cytoskeleton dynamics in Dictyostelium. *FEBS Lett* **2004**, *577*, 227-232.

239. Campbell, A. K. *Intracellular Ca2+: Its universal role as a regulator*, 1983, John Wiley and Sons, New York, NY, 556.

240. Berridge, M. J., Lipp, P. and Bootman, M. D. The versatility and universality of calcium signalling. *Nat Rev Mol Cell Bio* **2000**, *1*, 11-21.

241. Schultz, C., Schleifenbaum, A., Goedhart, J. and Gadella, T. W., Jr. Multiparameter imaging for the analysis of intracellular signaling. *Chembiochem* **2005**, *6*, 1323-1330.

242. Iwakura, M., Nakamura, T., Yamane, C. and Maki, K. Systematic circular permutation of an entire protein reveals essential folding elements. *Nat Struct Biol* **2000**, *7*, 580-585.

243. Hennecke, J., Sebbel, P. and Glockshuber, R. Random circular permutation of DsbA reveals segments that are essential for protein folding and stability. *J Mol Biol* **1999**, *286*, 1197-1215.

244. Chiang, J. J. H., Li, I. and Truong, K. Creation of circularly permutated yellow fluorescent proteins using fluorescence screening and a tandem fusion template. *Biotechnol Lett* **2006**, *28*, 471-475.

245. Shu, X. K., Shaner, N. C., Yarbrough, C. A., Tsien, R. Y. and Remington, S. J. Novel chromophores and buried charges control color in mFruits. *Biochemistry* **2006**, *45*, 9639-9647.

246. Kiss, C., Temirov, J., Chasteen, L., Waldo, G. S. and Bradbury, A. R. M. Directed evolution of an extremely stable fluorescent protein. *Protein Engineering Design & Selection* **2009**, *22*, 313-323.

247. Zapata-Hommer, O. and Griesbeck, O. Efficiently folding and circularly permuted variants of the Sapphire mutant of GFP. *BMC Biotechnol* **2003**, *3*,

248. Cheng, Z. H. and Campbell, R. E. Assessing the structural stability of designed beta-hairpin peptides in the cytoplasm of live cells. *Chembiochem* **2006**, *7*, 1147-1150.

249. Fiser, A. and Sali, A. ModLoop: automated modeling of loops in protein structures. *Bioinformatics* **2003**, *19*, 2500-2501.

250. Rohl, C. A., Strauss, C. E. M., Chivian, D. and Baker, D. Modeling structurally variable regions in homologous proteins with rosetta. *Proteins: Structure Function and Bioinformatics* **2004**, *55*, 656-677.

251. Delano-Wood, L. (2008) *The PyMOL Molecular Graphics System* (1.5) Available @ http://www.pymol.org. (May 10, 2010)

252. Gangola, P. R., B.P. Maintenance of intracellular calcium in Escherichia coli. *J Biol Chem* **1987**, *262*, 12570.

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

254. Gifford, J. L., Walsh, M. P. and Vogel, H. J. Structures and metal-ion-binding properties of the Ca2+-binding helix-loop-helix EF-hand motifs. *Biochem J* **2007**, *405*, 199-221.

255. Shaner, N. C., Patterson, G. H. and Davidson, M. W. Advances in fluorescent protein technology. *J Cell Sci* **2007**, *120*, 4247-4260.

256. Lakowicz, J. R. *Principles of Fluorescence Spectroscopy*, 2006, Springer,

257. Bravaya, K. B., Grigorenko, B. L., Nemukhin, A. V. and Krylov, A. I. Quantum Chemistry Behind Bioimaging: Insights from Ab Initio Studies of Fluorescent Proteins and Their Chromophores. *Acc Chem Res* **2012**, *45*, 265-275.

258. Chica, R. A., Moore, M. M., Allen, B. D. and Mayo, S. L. Generation of longer emission wavelength red fluorescent proteins using computationally designed libraries. *Proc Natl Acad Sci U S A* **2010**, *107*, 20257-20262.

259. Nantasenamat, C., Isarankura-Na-Ayudhya, C., Tansila, N., Naenna, T. and Prachayasittikul, V. Prediction of GFP spectral properties using artificial neural network. *J Comput Chem* **2007**, *28*, 1275-1289.

260. Topol, I., Collins, J., Savitsky, A. and Nemukhin, A. Computational strategy for tuning spectral properties of red fluorescent proteins. *Biophys Chem* **2011**, *158*, 91-95.

261. Treynor, T. P., Vizcarra, C. L., Nedelcu, D. and Mayo, S. L. Computationally designed libraries of fluorescent proteins evaluated by preservation and diversity of function. *Proc Natl Acad Sci U S A* **2007**, *104*, 48-53.

262. Subach, O. M., Gundorov, I. S., Yoshimura, M., Subach, F. V., Zhang, J., Gruenwald, D., Souslova, E. A., Chudakov, D. M. and Verkhusha, V. V. Conversion of red fluorescent protein into a bright blue probe. *Chem Biol* **2008**, *15*, 1116-1124.

263. Subach, F. V., Piatkevich, K. D. and Verkhusha, V. V. Directed molecular evolution to design advanced red fluorescent proteins. *Nat Methods* **2011**, *8*, 1019-1026.

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

265. Maddalo, S. L. and Zimmer, M. The role of the protein matrix in green fluorescent protein fluorescence. *Photochem Photobiol* **2006**, *82*, 367-372.

266. Heim, N. and Griesbeck, O. Genetically encoded indicators of cellular calcium dynamics based on troponin C and green fluorescent protein. *J Biol Chem* **2004**, *279*, 14280-14286.

267. Hasan, M. T., Friedrich, R. W., Euler, T., Larkum, M. E., Giese, G., Both, M., Duebel, J., Waters, J., Bujard, H., Griesbeck, O., Tsien, R. Y., Nagai, T., Miyawaki, A. and Denk, W. Functional fluorescent Ca2+ indicator proteins in transgenic mice under TET control. *PLoS Biol* **2004**, *2*, e163.

268. Mori, M. X., Erickson, M. G. and Yue, D. T. Functional stoichiometry and local enrichment of calmodulin interacting with Ca2+ channels. *Science* **2004**, *304*, 432-435.

269. Palmer, A. E., Giacomello, M., Kortemme, T., Hires, S. A., Lev-Ram, V., Baker, D. and Tsien, R. Y. Ca2+ indicators based on computationally redesigned calmodulin-peptide pairs. *Chem Biol* **2006**, *13*, 521-530.

270. Maune, J. F., Klee, C. B. and Beckingham, K. Ca2+ binding and conformational change in two series of point mutations to the individual Ca(2+)-binding sites of calmodulin. *J Biol Chem* **1992**, *267*, 5286-5295.

271. Bravaya, K. B., Subach, O. M., Korovina, N., Verkhusha, V. V. and Krylov, A. I. Insight into the common mechanism of the chromophore formation in the red fluorescent proteins: the elusive blue intermediate revealed. *J Am Chem Soc* **2012**, *134*, 2807-2814.

272. Rubart, M. Two-photon microscopy of cells and tissue. *Circ Res* **2004**, *95*, 1154-1166.

273. Palmer, A. E. and Tsien, R. Y. Measuring calcium signaling using genetically targetable fluorescent indicators. *Nat Protoc* **2006**, *1*, 1057-1065.

274. Palmer, A. E., Qin, Y., Park, J. G. and McCombs, J. E. Design and application of genetically encoded biosensors. *Trends Biotechnol* **2011**, *29*, 144-152.

275. Leder, L., Stark, W., Freuler, F., Marsh, M., Meyerhofer, M., Stettler, T., Mayr, L. M., Britanova, O. V., Strukova, L. A., Chudakov, D. M. and Souslova, E. A. The structure of Ca2+ sensor Case16 reveals the mechanism of reaction to low Ca2+ concentrations. *Sensors (Basel)* **2010**, *10*, 8143-8160.

276. Ilagan, R. P., Rhoades, E., Gruber, D. F., Kao, H. T., Pieribone, V. A. and Regan, L. A new bright green-emitting fluorescent protein--engineered monomeric and dimeric forms. *FEBS Journal* **2010**, *277*, 1967-1978.

277. Gallagher, S. R. One-dimensional SDS gel electrophoresis of proteins, F. M. Ausubel, *Current Protocols in Molecular Biology* 2012, John Wiley & Sons, 10.12A.11- 10.12A.44.

278. Zettler, J., Schutz, V. and Mootz, H. D. The naturally split Npu DnaE intein exhibits an extraordinarily high rate in the protein trans-splicing reaction. *FEBS Lett* **2009**, *583*, 909-914.

279. Fan, J. Y., Cui, Z. Q., Wei, H. P., Zhang, Z. P., Zhou, Y. F., Wang, Y. P. and Zhang, X. E. Split mCherry as a new red bimolecular fluorescence

complementation system for visualizing protein-protein interactions in living cells. *Biochem Biophys Res Commun* **2008**, *367*, 47-53.

280. Ozawa, T., Takeuchi, T. M., Kaihara, A., Sato, M. and Umezawa, Y. Protein splicing-based reconstitution of split green fluorescent protein for monitoring protein-protein interactions in bacteria: improved sensitivity and reduced screening time. *Anal Chem* **2001**, *73*, 5866-5874.

281. Tasayco, M. L. and Chao, K. NMR study of the reconstitution of the beta-sheet of thioredoxin by fragment complementation. *Proteins: Structure Function and Bioinformatics* **1995**, *22*, 41-44.

282. Gross, M., Wyss, M., Furter-Graves, E. M., Wallimann, T. and Furter, R. Reconstitution of active octameric mitochondrial creatine kinase from two genetically engineered fragments. *Protein Science* **1996**, *5*, 320-330.

283. Glover, J. N. and Harrison, S. C. Crystal structure of the heterodimeric bZIP transcription factor c-Fos-c-Jun bound to DNA. *Nature* **1995**, *373*, 257-261.

284. Vinson, C. R., Hai, T. and Boyd, S. M. Dimerization specificity of the leucine zipper-containing bZIP motif on DNA binding: prediction and rational design. *Genes Dev* **1993**, *7*, 1047-1058.

285. Kaddoum, L., Magdeleine, E., Waldo, G. S., Joly, E. and Cabantous, S. One-step split GFP staining for sensitive protein detection and localization in mammalian cells. *Biotechniques* **2010**, *49*, 727-728, 730, 732 passim.

286. Zhou, J., Lin, J., Zhou, C., Deng, X. and Xia, B. An improved bimolecular fluorescence complementation tool based on superfolder green fluorescent protein. *Acta Biochim Biophys Sin (Shanghai)* **2011**, *43*, 239-244.

287. Chalfie, M., Tu, Y., Euskirchen, G., Ward, W. W. and Prasher, D. C. Green fluorescent protein as a marker for gene expression. *Science* **1994**, *263*, 802-805.

288. Niino, Y., Hotta, K. and Oka, K. Blue Fluorescent cGMP Sensor for Multiparameter Fluorescence Imaging. *Plos One* **2010**, *5*,

289. Russwurm, M., Mullershausen, F., Friebe, A., Jager, R., Russwurm, C. and Koesling, D. Design of fluorescence resonance energy transfer (FRET)-based cGMP indicators: a systematic approach. *Biochem J* **2007**, *407*, 69-77.

290. Huber, D., Gutnisky, D. A., Peron, S., O'Connor, D. H., Wiegert, J. S., Tian, L., Oertner, T. G., Looger, L. L. and Svoboda, K. Multiple dynamic representations in the motor cortex during sensorimotor learning. *Nature* **2012**, *484*, 473-478.

291. Dombeck, D. A., Harvey, C. D., Tian, L., Looger, L. L. and Tank, D. W. Functional imaging of hippocampal place cells at cellular resolution during virtual navigation. *Nat Neurosci* **2010**, *13*, 1433-1440.

292. O'Connor, D. H., Clack, N. G., Huber, D., Komiyama, T., Myers, E. W. and Svoboda, K. Vibrissa-based object localization in head-fixed mice. *J Neurosci* **2010**, *30*, 1947-1967.

293. Mutoh, H., Akemann, W. and Knopfel, T. Genetically Engineered Fluorescent Voltage Reporters. *Acs Chem Neurosci* **2012**, *3*, 585-592.

294. Knopfel, T., Lin, M. Z., Levskaya, A., Tian, L., Lin, J. Y. and Boyden, E. S. Toward the second generation of optogenetic tools. *J Neurosci* **2010**, *30*, 14998-15004.

295. Mutoh, H., Perron, A., Akemann, W., Iwamoto, Y. and Knopfel, T. Optogenetic monitoring of membrane potentials. *Exp Physiol* **2011**, *96*, 13-18.

296. Miyawaki, A., Mizuno, H., Nagai, T. and Sawano, A. Development of genetically encoded fluorescent indicators for calcium. *Method Enzymol* **2003**, *360*, 202-225.

297. Mank, M., Reiff, D. F., Heim, N., Friedrich, M. W., Borst, A. and Griesbeck, O. A FRET-based calcium biosensor with fast signal kinetics and high fluorescence change. *Biophys J* **2006**, *90*, 1790-1796.

298. Hoi, H., Matsuda, T., Nagai, T. and Campbell, R. E. Highlightable ca(2+) indicators for live cell imaging. *J Am Chem Soc* **2013**, *135*, 46-49.

299. Baker, B. J., Mutoh, H., Dimitrov, D., Akemann, W., Perron, A., Iwamoto, Y., Jin, L., Cohen, L. B., Isacoff, E. Y., Pieribone, V. A., Hughes, T. and Knopfel, T. Genetically encoded fluorescent sensors of membrane potential. *Brain Cell Biol* **2008**, *36*, 53-67.

Appendices

Appendix A: Oligonucleotide sequence supplement

Table A.1 List of oligonucleotides used in this work

OLIGO #	Lay Name	SEQUENCE (5' to 3')
2.1	REC_A1	ATGCCATAGCATTTTTATCC
2.2	REC_B1	GATTTAATCTGTATCAGG
2.3	T7_seq_F	TAATACGACTCACTATAGGG
2.4	T7_seq_R	TAGAAGGCACAGTCGAGG
2.5	HJC_48_1	GAC CTC GAG CAT GGT GAG CAA GGG C
2.6	HJC_49_2	CCG AAT TCT CAC TTG TAC AGC TCG TCC ATG CC
2.7	cplinker_96_1	CTG TAC AAG GGT GGC ACT GGC GGT TCC ATG GTG AGC AAG GGC
2.8	cplinker_97_2	GCT CAC CAT GGA ACC GCC AGT GCC ACC CTT GTA CAG CTC GTC CAT
2.9	G7_nooverlap_206_1	GTA CGA GCT CGA GGG TCG ACA TCA AGT TGG ACA TCA CCT CA
2.10	G7_nooverlap_207_2	GTA CGA ATT CTC AGC CGG GCA GCT GCA CGG G
2.11	G7_overlap_208_1	GTA CGA GCT CGA GGG CCT ACA ACG TCG ACA TCA CCT CA
2.12	G7_overlap_209_2	GTA CGA ATT CTC AGT TGT AAG CGC CGG GCA G
2.13	192CHE_133_1	GTACCTCGAGCATGTCCAGCTCCGGGGCACTGGGCGGTNNK[1]TACAACGTCGACATCAAGT
2.14	192CHE_134_2	TAC GAA TTC TCA CTC TTC TGT CAG TTG GTC ACC GCC MNN[2] GTA AGC GCC GGG CAG
2.15	193CHE_135_1	GTACCTCGAGCATGTCCAGCTCCGGGGCACTGGGCGGTNNKAACGTCGACATC
2.16	193CHE_136_2	GTA CGA ATT CTC ACT CTT CTG TCA GTT GGT CAC CGC CMN NGT TGT AAG CGC C
2.17	194CHE_137_1	GTACCTCGAGCATGTCCAGCTCCGGGGCACTGGGCGGTNNKGTCGACATCAAG

2.18	194CHE_138_2	GTA CGA ATT CTC ACT CTT CTG TCA GTT GGT CAC CGC CMN NGA CGT TGT AAG C
2.19	195CHE_139_1	GTACCTCGAGCATGTCCAGCTCCGGGGCACTGGGCGGTNNKGACATCAAGTTG
2.20	195CHE_140_2	GTA CGA ATT CTC ACT CTT CTG TCA GTT GGT CAC CGC CMN NGT CGA CGT TGT A
2.21	196CHE_141_1	GTACCTCGAGCATGTCCAGCTCCGGGGCACTGGGCGGTNNKATCAAGTTGGAC
2.22	196CHE_142_2	GTA CGA ATT CTC ACT CTT CTG TCA GTT GGT CAC CGC CMN NGA TGT CGA CGT TGT A
2.23	197CHE_143_1	GTACCTCGAGCATGTCCAGCTCCGGGGCACTGGGCGGTNNKAAGTTGGACATCAC
2.24	197CHE_144_2	GTA CGA ATT CTC ACT CTT CTG TCA GTT GGT CAC CGC CMN NCT TGA TGT CGA CGT TGT A
2.25	198CHE_145_1	GTACCTCGAGCATGTCCAGCTCCGGGGCACTGGGCGGTNNKTTGGACATCACCTCA
2.26	198CHE_146_2	GTA CGA ATT CTC ACT CTT CTG TCA GTT GGT CAC CGC CMN NCA ACT TGA TGT CGA CGT T
2.27	199CHE_147_1	GTACCTCGAGCATGTCCAGCTCCGGGGCACTGGGCGGTNNKGACATCACCTCA
2.28	199CHE_148_2	GTA CGA ATT CTC ACT CTT CTG TCA GTT GGT CAC CGC CMN NGT CCA ACT TGA TGT C
2.29	200CHE_149_1	GTACCTCGAGCATGTCCAGCTCCGGGGCACTGGGCGGTNNKATCACCTCACAC
2.30	200CHE_150_2	GTA CGA ATT CTC ACT CTT CTG TCA GTT GGT CAC CGC CMN NGA TGT CCA ACT T
2.31	201CHE_151_1	GTACCTCGAGCATGTCCAGCTCCGGGGCACTGGGCGGTNNKACCTCACACAAC
2.32	201CHE_152_2	GTA CGA ATT CTC ACT CTT CTG TCA GTT GGT CAC CGC CMN NGG TGA TGT CCA ACT T
2.33	202CHE_153_1	GTACCTCGAGCATGTCCAGCTCCGGGGCACTGGGCGGTNNKTCACACAACGAG
2.34	202CHE_154_2	GTA CGA ATT CTC ACT CTT CTG TCA GTT GGT CAC CGC CMN NTG AGG TGA TGT C
2.35	203CHE_155_1	GTACCTCGAGCATGTCCAGCTCCGGGGCACTGGGCGGTNNKCACAACGAGGACTAC
2.36	203CHE_156_2	GTA CGA ATT CTC ACT CTT CTG TCA GTT GGT CAC CGC CMN NGT GTG AGG TGA TGT C
2.37	204CHE_157_1	GTACCTCGAGCATGTCCAGCTCCGGGGCACTGGGCGGTNNKAACGAGGACTAC
2.38	204CHE_158_2	GTA CGA ATT CTC ACT CTT CTG TCA GTT GGT CAC CGC CMN NGT TGT GTG AGG TGA T
2.39	205CHE_159_1	GTACCTCGAGCATGTCCAGCTCCGGGGCACTGGGCGGTNNKGAGGACTACACCATC
2.40	205CHE_160_2	GTA CGA ATT CTC ACT CTT CTG TCA GTT GGT CAC CGC CMN NCT CGT TGT GTG A
2.41	206CHE_161_1	GTACCTCGAGCATGTCCAGCTCCGGGGCACTGGGCGGTNNKGACTACACCATCGTG
2.42	206CHE_162_2	GTA CGA ATT CTC ACT CTT CTG TCA GTT GGT CAC CGC CMN NGT CCT CGT TGT G
2.43	207CHE_163_1	GTACCTCGAGCATGTCCAGCTCCGGGGCACTGGGCGGTNNKTACACCATCGTG

2.44	207CHE_164_2	GTA CGA ATT CTC ACT CTT CTG TCA GTT GGT CAC CGC CMN NGT AGT CCT CGT T
2.45	208CHE_165_1	GTACCTCGAGCATGTCCAGCTCCGGGGCACTGGGCGGTNNKACCATCGTGGAA
2.46	208CHE_166_2	GTA CGA ATT CTC ACT CTT CTG TCA GTT GGT CAC CGC CMN NGG TGT AGT CCT CGT T
3.1	196_Fwd_235_1	GTA CCT CGA GCA TGT CCA GCT CCG GGG C
3.2	196_Rev_236_2	CGG GCG GTG ACC AAC TGA CAG AAG AGT GAG AAT TCG TAC
3.3	196_L65X_227_1	GCC TGG GAC ATC CTG TCC CCT CAG NDB[3] ATG TAC GGC TCC AAG GCC TAC GTG AAG C
3.4	196_L65X_228_2	GCT TCA CGT AGG CCT TGG AGC CGT ACA TVH N[4]CT GAG GGG ACA GGA TGT CCC AGG C
3.5	196_Y193X_N194X_229_1	GTG CAG CTG CCC GGC GCT NNK NNK GCG GGC GGT GAC CAA CTG ACA GAA GAG TGA GAA TTC
3.6	196_Y193X_N194X_230_2	GAA TTC TCA CTC TTC TGT CAG TTG GTC ACC GCC CGC MNN MNN AGC GCC GGG CAG CTG CAC
3.7	196_S147X_M150X_231_1	ACC ATG GGC TGG GAG GCC TCC NNK GAG CGG NNK TAC CCC GAG GAC GGC GCT CTG
3.8	196_S147X_M150X_232_2	CAG AGC GCC GTC CTC GGG GTA MNN CCG CTC MNN GGA GGC CTC CCA GCC CAT GGT
3.9	196_Q188X_233_1	GTC AAG ACC ACC TAC AAG GCC AAG AAG CCT GTG NNK CTG CCC GGC GCT CAC AAC G
3.10	196_Q188X_234_2	CGT TGT GAG CGC CGG GCA GMN NCA CAG GCT TCT TGG CCT TGT AGG TGG TCT TGA C
3.11	196_H221X_237_1	CTC GAG CAT GTC CAG CTC CGG GGC ACT GGT CGG TTT TAT CAA GTT GGA CAT CAC CTC ACA CAA CGA GGA CTA CAC CAT CGT GGA ACA GTA CGA ACG CGC CGA GGT CCG CNN KTC C
3.12	Q188_Y193_N194_238_2	GTC AAG ACC ACC TAC AAG GCC AAG AAG CCT GTG NNK CTG CCC GGC NNK NNK AAC GCG GGC GGT GAC CAA CTG ACA GAA GAG TGA GAA TTC GTA C
3.13	cp196_GCaMP_239_1	CAT GCT CGA GGT CGG TTT TAT CAA GTT GGA CAT CAC CTC ACA CAA CGA GGA CTA CAC CAT C
3.14	cp197_GCaMP_240_1	CAT GCT CGA GGT CGG TNN KAA GTT GGA CAT CAC CTC ACA CAA CGA GGA CTA CAC CAT C
3.15	cp198_GCaMP_241_1	CAT GCT CGA GGT CGG TNN KTT GGA CAT CAC CTC ACA CAA CGA GGA CTA CAC CAT C
3.16	cp199_GCaMP_242_1	CAT GCT CGA GGT CGG TNN KGA CAT CAC CTC ACA CAA CGA GGA CTA CAC CAT C
3.17	cp200_GCaMP_243_1	CAT GCT CGA GGT CGG TNN KAT CAC CTC ACA CAA CGA GGA CTA CAC CAT C
3.18	cp201_GCaMP_244_1	CAT GCT CGA GGT CGG TNN KAC CTC ACA CAA CGA GGA CTA CAC CAT CGT C
3.19	cp196_GCaMP_245_2	CAT GAC GCG TAC CGC CCG CGT TGT GAG CGC CGG GCA G

3.20	cp197_GCaMP_246_2	CAT GAC GCG TAC CGC CMN NCG CGT TGT GAG CGC CGG GC
3.21	cp198_GCaMP_247_2	CAT GAC GCG TAC CGC CMN NAA ACG CGT TGT GAG CGC CGG GC
3.22	cp199_GCaMP_248_2	CAT GAC GCG TAC CGC CMN NGA TAA ACG CGT TGT GAG CGC CGG GC
3.23	cp200_GCaMP_249_2	CAT GAC GCG TAC CGC CMN NCT TGA TAA ACG CGT TGT GAG CGC CGG GC
3.24	cp201_GCaMP_250_2	CAT GAC GCG TAC CGC CMN NCA ACT TGA TAA ACG CGT TGT GAG CGC CGG GC
3.25	cp144_271_1	GTA CCT CGA GGG CGG TNN KGC CTC CAT TGA GCG GCT TTA CCC C
3.26	cp144_272_2	CAT GAC GCG TAC CGC CMN NGC CCA TGG TCT TCT TCT GCA TGA C
3.27	cp145_273_1	GTA CCT CGA GGG CGG TNN KTC CAT TGA GCG GCT TTA CCC CGA G
3.28	cp145_274_2	CAT GAC GCG TAC CGC CMN NCC AGC CCA TGG TCT TCT TCT GCA TGA C
3.29	cp146_275_1	GTA CCT CGA GGG CGG TNN KAT TGA GCG GCT TTA CCC CGA GGA C
3.30	cp146_276_2	CAT GAC GCG TAC CGC CMN NCT CCC AGC CCA TGG TCT TCT TCT GCA TGA C
3.31	cp147_277_1	GTA CCT CGA GGG CGG TNN KGA GCG GCT TTA CCC CGA GGA CG
3.32	cp147_278_2	CAT GAC GCG TAC CGC CMN NGG CCT CCC AGC CCA TGG TCT TCT TC
3.33	cp196_cp147_279_2	GTG TGA GGT GAT GTC CAA CTT GAT AAA GGC GTT GTG AGC GCC GGG
3.34	FW_XbaI-6His	GCGATGTCTAGAGGTTCTCATCATCATCATCATCATGGTATGGCTAGC
3.35	RV_GCaMP-Stop-HindIII	GCGATGAAGCTTCTACTTCGCTGTCATCATTTGTACAAACTCTTCGTAGTTT
3.36	G144aX_G144bX_291_1	TAGGTCGGCTGAGCTCACTCGAGNYNNYN^{5}CTGTCCATTGAGCGGCTTTACCC
3.37	XhoI_LEXX_294_1	AGAGCTATAGGTCGGCTGAGCTCANYNNYNCTGTCCATTGAGCGGCTTTACCC
4.1	RFP_G159S_341_1	ACGGCGCTCTGAAGAGCGAGATCAAGCAG
4.2	RFP_Q163K_342_1	GAAGGGCGAGATCAAGAAGAGGCTGAAGCTGAA
4.3	RFP_T41F_343_1	CGCCCCTACGAGGGCTTTCAGACCGCCAAGCTG
4.4	RFP_W83F_344_1	GACATCCCCGACTACTTCAAGCTGTCCTTCCCCG
4.5	RFP_G159S_Q163K_349_1	TGAAGAGCGAGATCAAGAAGAGGCTGAAGCTGAAG
4.6	RFP_W83L_350_1	GACATCCCCGACTACTTGAAGCTGTCCTTCCCCG
4.7	RFP_W83K_351_1	CGACATCCCCGACTACAAGAAGCTGTCCTTCCCC
4.8	RFP_K70R_352_1	GTACGGCTCCAAGGCCTACATTCGCCACCCAGCCGAC
4.9	RFP_Q163M_353_1	GAAGGGCGAGATCAAGATGAGGCTGAAGCTGAA

4.10	RFP_Y66F_359_1	CTGTCCCCTCAGTTCATGTTCGGCTCCAAG
4.11	RFP_K70Q_360_1	AGTTCATGTACGGCTCCCAGGCCTACATTAAGCAC
4.12	RFP_E148Q_361_1	GCTGCTGTCCACCCAGCGGCTTTACCC
4.13	RFP_Y193F_362_1	CTGCCCGACGCCTTCATCGTCGACATC
4.14	RFP_H75E_363_1	CTCCAAGGCCTACATTAAGGAGCCAGCCGACATCC
4.15	RFP_H75K_364_1	CTCCAAGGCCTACATTAAGAAGCCAGCCGACATCC
4.16	RFP_H75Q_365_1	TCCAAGGCCTACATTAAGCAGCCAGCCGACAT
4.17	RFP_W83K_366_1	CGACATCCCCGACTACAAGAAGCTGTCCTTCCCC
4.18	RFP_Y193H_367_1	GCTGCCCGACGCCCACATCGTCGACAT
4.19	RFP_M65T_368_1	TGTCCCCTCAGTTCACCTACGGCTCCAAGGCC
4.20	RFP_K70M_369_1	CAGTTCATGTACGGCTCCATGGCCTACATTAAGC
4.21	RFP_M65Q_370_1	CTGTCCCCTCAGTTCCAGTACGGCTCCAAGGC
4.22	CAM_A23D_372_1	TTTGACAAGGACGGGGATGGGACGATAACAACC
4.23	RFP_LLS_PVV_371_1	TGAGCTCACTTGAGTCGCCCGTGGTCACCGAGCGGCTTTACCC
4.24	RFP_LES_XXX_372_1	GCTATAGGTCGGCTGAGCTCACCCGTGGTCAC
4.25	RFP_WLG_WEA_373_1	GAAGACCATGGGCTGGGAGGCTGGTACGCGTGACCAAC
4.26	RFP_GG_XX_374_1	CCATGGGCTGGGAGGCTACGCGTGAC
4.27	RFP_H172Q_377_1	GACGGCGGCCAGTACGCCGCCGA
4.28	RFP_Y214F_378_1	TACACCATCGTGGAACAGTTCGAACGCGCC
4.29	RFP_D59Q_379_1	GCCCTTCGCCTGGCAGATCCTGTCCCCTC
4.30	RFP_H204Q_380_1	GGACATCGTGTCCCAGAACGAGGACTACACC
4.31	RFP_Y208_F_381_1	CCCACAACGAGGACTTCACCATCGTGGAACA
4.32	RFP_D239Q_382_1	CTGGGCGGTACGCGTCAGCAACTGACTGAAGAG
4.33	CaM_N100D_383_1	GAGCTTCGCCACGTGATGACAGACCTTGGAGAG
4.34	EP1_Q163D_393_1	GAAGGGCGAGATCAAGGACAGGCTGAAGCTGAAGG
4.35	EP1_Q163D_2	CCTTCAGCTTCAGCCTGTCCTTGATCTCGCCCTTC
4.36	EP1_K166R_395_1	GCGAGATCAAGCAGAGGCTGAGGCTGAAGGACG

4.37	EP1_Q163K_K166R_396_1	GGCGAGATCAAGAAGAGGCTGAGGCTGAAGGACGG
5.1	Nint_RCaMP_308_1	GTACGAGCTCCTTGAGTCGCTGCTGTCCACC
5.2	Nint_RCaMP_309_2	CGGCATGGACGAGCTGTACAAGGGCGGATCCGTAC
5.3	Cint_RCaMP_311_1	GTACGGTACCGCTAGCATGGTGAGCAAGGGCGTGGAG
5.4	Cint_RCaMP_312_2	GAGTTTGTACAAATGATGACAGCGAAGTGACTCGAGGTAC
5.5	Cint_RGECO_313_1	GTACGGTACCGCTAGCCTGGAGAGCAAGGGCGTGGA
5.6	Cint_RFP_316_2	GAAGAAGACCATGGGCTGGCTGTGACTCGAGGTAC
5.7	Cint_MAC_317_2	AAGACCATGGGCTGGGAGGCTTGACTCGAGGTAC
5.8	196CaMP_Nint_318_1	GTACGAGCTCGTCTGTTTTATCAAGTTGGACATCACCTCAAACGAGGAC
5.9	Nint_RCaMP_322_1	GTACGAGCTCACTTGAGTCGCTGCTGTCCACC
5.10	196CaMP_Nint_323_1	GTACGAGCTCAGTCTGTTTTATCAAGTTGGACATCACCTCAAACGAGGAC
5.11	Nint_RGECO_330_1	GTACCTCGAGTCGTCGTAAGTGGAATAAGTGGAATAAGACAGGTCACGC
5.12	Nint_RFP_331_1	GTACCTCGAGTTCACTTGAGTCGCTGCTGTCCACC
5.13	Nint_MAC_332_1	GTACCTCGAGTCCCGTGGTTTCCGAGCGGAT
5.14	196_Nint_333_1	GTACCTCGAGTGTCTGTTTTATCAAGTTGGACATCACCTCAAACGAGGAC
5.15	Cint_RCaMP_334_2	GTACAAGCTTTCACTTCGCTGTCATCATTTGTACAAACTC
5.16	Cint_RFP_335_2	GTACAAGCTTTCACAGCCAGCCCATGGTCTTCTTC
5.17	Cint_MAC_336_2	GTACAAGCTTTCAAGCCTCCCAGCCCATGGTCTT
5.18	196_Cint_337_2	GTACAAGCTTTCACGCGTTGTGAGCGCCGG
5.19	Nint_RCaMP_322_1	GTACGAGCTCACTTGAGTCGCTGCTGTCCACC
5.20	196CaMP_Nint_323_1	GTACGAGCTCAGTCTGTTTTATCAAGTTGGACATCACCTCAAACGAGGAC
5.21	Nint_RGECO_330_1	GTACCTCGAGTCGTCGTAAGTGGAATAAGTGGAATAAGACAGGTCACGC
5.22	Nint_RFP_331_1	GTACCTCGAGTTCACTTGAGTCGCTGCTGTCCACC
5.23	Nint_MAC_332_1	GTACCTCGAGTCCCGTGGTTTCCGAGCGGAT
5.24	196_Nint_333_1	GTACCTCGAGTGTCTGTTTTATCAAGTTGGACATCACCTCAAACGAGGAC
5.25	Cint_RCaMP_334_2	GTACAAGCTTTCACTTCGCTGTCATCATTTGTACAAACTC
5.26	Cint_RFP_335_2	GTACAAGCTTTCACAGCCAGCCCATGGTCTTCTTC

5.27	Cint_MAC_336_2	GTACAAGCTTTCAAGCCTCCCAGCCCATGGTCTT
5.28	196_Cint_337_2	GTACAAGCTTTCACGCGTTGTGAGCGCCGG

[1] IDT degenerate oligo designation – N: A,G,C,T; K:G,T
[2] Reverse compliment of NNK – M: A,C
[3] N: A,G,C,T; D: A,G,T; B: C,G,T
[4] Reverse compliment of NDB – V: A,C,G; H: A,C,T
[5] Y: C,T

Appendix B: Chapter 3 cp196V1.0 and cp201V1.0 Mutation summary

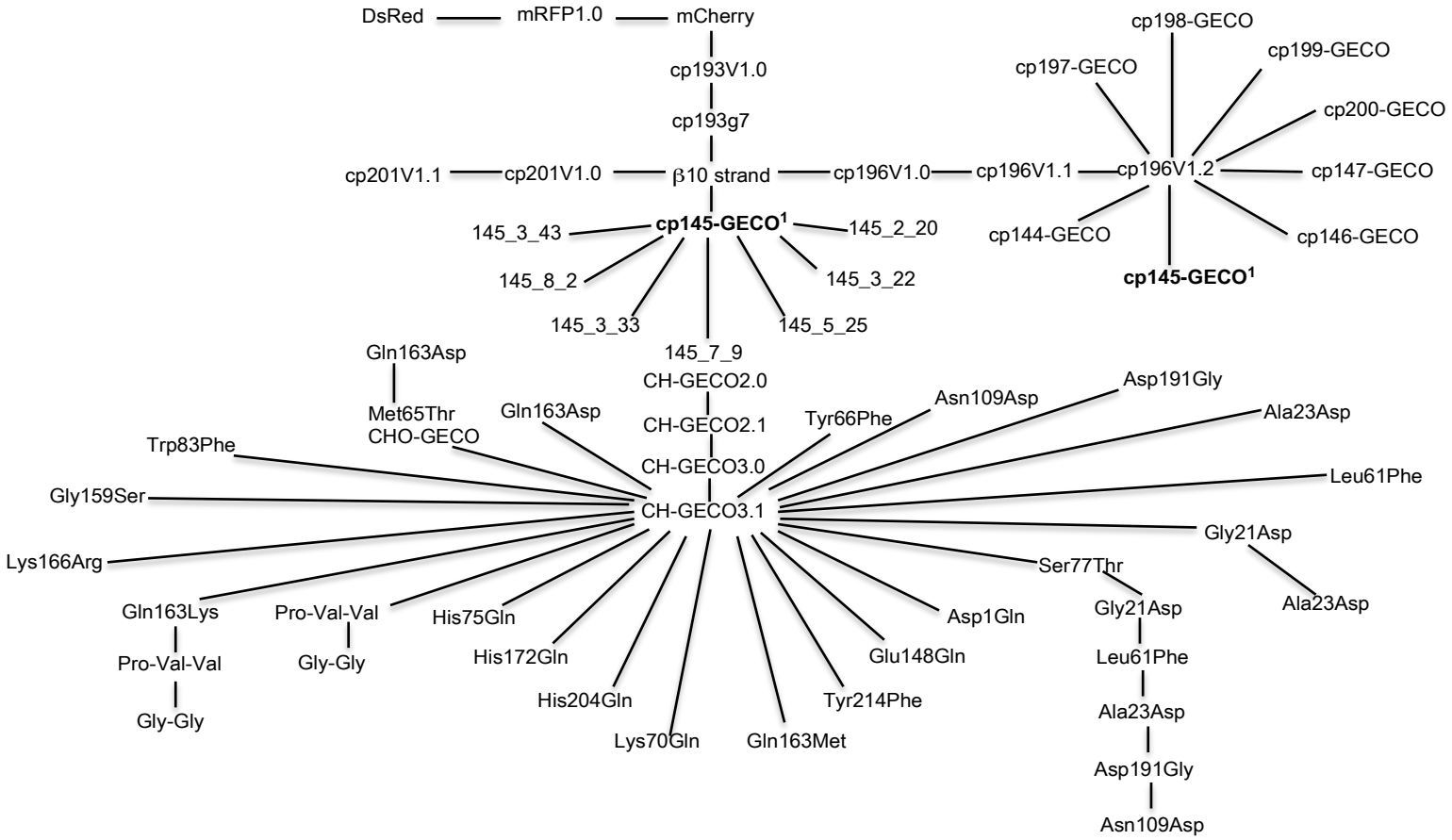
Table B.1 Summary of cp196V1.0 EPPCR/StEP mutagenesis variants

Variant	Mutations
196_3_1	G196aV, S131P, M150L
196_3_3	G196aV, M150L, Q188R
196_3_5	G196aV, D78G, M150L, N194S
196_3_11	G196aV, E19V, M150L, Q188R
196_3_12	G196aV, T43A, M150L, H221R, K231E
196_4_1	G196aV, M150L, E160A, D174V
196_4_2	G196aV, G196bC, M18V, H25Y, T49I, T108I, S147Y, H172Y, Q188R
196_4_4	G196aV, N6cS, T43A, M150L
196_4_5	G196aV, M6dL, M150L
196_4_8	G196aV, M150L
196_4_11	G196aV, M150L, Y193H, T223I
196_4_16	G196aV, M150L, P190T
196_4_18	G196aV, M6dT, M150L
196_B1	G196aS, M6dT, D132G, N194D
196_2_3	S147Y, Y214H

Table B.2 Summary of cp201V1.0 EPPCR/StEP mutagenesis variants

Variant	Mutations
201_A3	E34G, K231R
201_A5	D227V
201_A6	T202A, Y214H
201_A8	K4R, E7G
201_AA	G201aS, R13C
201_AB	L65F, E176D
201_AC	G201aC
201_AD	S147T
201_AE	G201D, M18L, T43A, K139R
201_AF	N6cY, L65F
201_AG	G201aS, R125C
201_AH	L65F, V135I
201_AI	R125C, K198E
201_AK	Q213L
201_AM	A6eV, K139M, Q213L
201_AO	L65F, L229R
201_B2	L65F
201_B4	K168R
201_B5	L65F, Y230C, K231M
201_B7	A6eV, A145V, A156V
201_B8	Y151H
201_B9	E160V
201_BC	G201aS, A77T, E206G
201_BD	G52D, G155D, N194aS, Q213L
201_3_3	L65F, E117V, S147T, G200aS
201_3_5	L65F, S147T, G200aS
201_3_7	L65F, S147T, G200aS, D227G
201_3_9	L65F, K139M, S147T, G200aS
201_3_12	L65F, S147T, K178T, G200aS
201_4_17	L65F, S147T, N194I, G200aC
201_4_19	L65F, A77T, S147T, G200aC, M6dV, I29V
201_4_20	L65F, A77T, S147T, G200aC

Appendix C: Family Tree of mCherry constructs described in this thesis



[1]cp145-GECO is the same construct. The branch point has been re-organized for clarity.