Modular Design and Directed Evolution for the Development of Genetically Encodable Fluorescent Sensors

By

Landon Cameron Ashlie Zarowny

A thesis submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Chemistry
University of Alberta

Abstract

The development of genetically encodable sensors for optical imaging has revolutionized the way researchers investigate cellular phenomena. Illuminating the molecular dynamics of the cell has become nearly routine with countless improvements in sensor design and directed evolution efforts to enhance their reporting ability. These optical technologies owe their impressive versatility to their genetically encodable nature. As new luminescent and fluorescent reporter domains are discovered, new sensors will inevitably be developed from them. In this thesis, I cover the basics of sensor design while commenting on methods for construction of prototypes, rational structure guided design, directed evolution methods, and briefly touch on new methods for developing split sensors. The goal of this thesis is to explore the currently available technologies for developing genetically encodable optical sensors and provide a methodology for general sensor design with examples provided within each chapter.

In Chapter 2 I describe the development of a series of spectrally distinct fluorescent proteins from an engineered variant of green fluorescent protein derived from Lancelets. Starting with the site-directed mutagenesis of a key residue of the fluorophore, I used a combination of directed evolution and structure guided rational design to improve the optical properties of this protein. The X-ray structure of NeonCyan helped to identify a residue critical for modulating the spectral properties resulting in the shifting of excitation and emission spectra. Further investigation of this new protein revealed interesting physical properties. I conclude this work by demonstrating the ability of NeonCyan variants to label subcellular organelle in cultured cell lines.

In Chapter 3 I describe the development of a Ca^{2+} sensor from the mNeonGreen fluorescent protein. Ca^{2+} is secondary messenger within all eukaryotic cells and the importance of fluorescent Ca^{2+} sensors have been demonstrated with several variants covering the visible spectrum. However, the most optimized Ca^{2+} sensors, e.g.) the GCaMP series, are still relatively dim compared to the brightest fluorescent proteins available. By engineering a Ca^{2+} sensor from the brightest green fluorescent protein available at the time of writing, I have demonstrated that starting from brighter scaffold proteins decreases the engineering effort required to develop bright and functional fluorescent protein-based sensors.

In Chapter 4 I describe the development of a split protein-protein interaction reporter that utilizes and endogenous substrate. The CobA enzyme is a *S*-adenosyl-L-methionine uroporphyrinogen III methyltransferase and when overexpressed in bacteria produces red fluorescence. Using directed evolution, I improved the transcriptional reporting activity of CobA to develop CobA1. I then used CobA1 to develop a split CobA1 sensor. However, the activity of the split CobA1 sensor was not suitable for studying protein-protein interactions in cells. To solve this issue, I took advantage of recently developed split inteins to engineer an intein and protein-protein interaction reporter capable of tracking protein interactions in cultured cell lines.

In Chapter 5 I comment on the outlook for the technologies described within this thesis.

Preface

I was responsible for the design, directed evolution, *in vitro* characterization, and cell-based characterization of NeonCyan and NeonCyan variants. Heather Baker and Abhi Aggarwal contributed to the screening, directed evolution, and *in vitro* characterization of NeonCyan and NeonCyan variants. Antoine Royant and Damien Clavel performed the crystallization and characterization of the NeonCyan structure. Antoine Royant and Damien Clavel wrote section “2.2.2 Crystal structure of NeonCyan” in Chapter 2. Robert Campbell conceived and supervised the experiments and edited Chapter 2.

I was responsible for the design, directed evolution, *in vitro* characterization, and *in vitro* characterization (cultured mammalian cells only) of mNG-GECO and mNG-GECO variants. Tiffany Phan developed the first prototypes of mNG-GECO referred to as mNG-GECO0.1 and mNG-GECO0.2. Abhi Aggarwal contributed to screening, directed evolution, and *in vitro* characterization of mNG-GECO and mNG-GECO variants and the culture and injection of zebrafish embryos. Dissociated neuronal tissues were a gift from Vladimir Rancic in accordance with the University of Alberta Animal Care and Use Committee and Canadian Council on Animal Care standards. Imaging of dissociated neuronal tissues was performed by Yong Qian. Imaging of zebrafish embryos was performed by Richard Kanyo. Robert Campbell conceived and oversaw the project and edited Chapter 3. Ted Allison oversaw culturing and experiments of zebrafish and zebrafish embryos.

I was responsible for the design, directed evolution, *in vitro* characterization of CobA and CobA variants. I performed all the imaging, sensor development, and prototyping

within Chapter 4. Robert Campbell conceived the project, supervised experiments, and edited the chapter.

This is an original work by Landon Zarowny.

No part of this thesis has been previously published.

Acknowledgments

Foremost, I would like to express my sincere gratitude to my supervisor Robert E. Campbell for his support and guidance throughout my Ph.D. studies. Your patience and understanding through this academic journey have helped me develop as a student of chemical biology and as a professional scientist. I will never forget the first time we met when I was an undergrad working construction in Edmonton. For our first meeting I was still dirty from work and I can hardly imagine what you must have thought about having a student with such rough appearance come to your clean and orderly lab. Thank you for giving me the opportunity to follow my dreams of becoming a protein engineer and giving me the freedom to investigate my research interests.

I would like to thank my supervisory committee, Professors John Vederas and Julianne Gibbs, for their guidance through my Ph.D. work.

I would also like to thank all the current and previous group members from Campbell Lab. Several students have helped me throughout my Ph.D. studies both professionally and personally. Working with this group of peers has made the last few years a wonderful experience. Specifically, I would like to thank the undergrads who helped with the work in Chapters 2 and 3: Abhi Aggarwal, Heather Baker, and Tiffany Phan. A special thank you to some of my mentors who taught me the basics when I was still an undergrad: Eason Shen, Yidan Ding, Jiahui Wu, and Matt Wiens.

Without my family, some of whom were sadly not able to see me finish this accomplishment, I don’t believe I could have come this far. Thank you to my mother and

father for supporting me through school and life in general. I would also like to thank my grandparents for everything they were able to pass down from their vast wisdom.

Finally, I would like to thank my girlfriend Claude. Your mastery of chemistry has inspired me to be a better chemist, while your patience and love has inspired me to be a better person. Thank you for your support through the final years of my Ph.D.

“Redemption is a mountain you climb starting in a hole you dug yourself”

Table of Contents

List of Tables

List of Figures

List of Abbreviations

Ac	CH_2COOH
ATP	Adenosine triphosphate
5-ALA	5-aminolevulinic acid
aTaq	after *Taq* mutagenic polymerase
avGFP	*Aequorea victoria* green fluorescent protein
BFP	Blue fluorescent protein
BON	Bacterial OsmY and nodulation domain
BPER	Bacterial protein extraction reagent
BRET	Bioluminescence resonance energy transfer
CaM	Calmodulin
CaMKIV	Calcium/calmodulin-dependent protein kinase type IV
CaMKK	Calcium/calmodulin-dependent protein kinase kinase
CFP	Cyan fluorescent protein
CobA	Cobalamin biosynthetic pathway protein A
cp	Circularly permuted
CPred	Circular permutation predictor
DHFR	Dihydrofolate reductase

DMEM	Dulbucco's modified Eagle medium
E. coli	Escherichia coli
EDTA	Ethylenediaminetetraacetic acid
EGTA	ethylene glycol-bis(2-aminoethylether)-N,N,N′,N′-tetraacetic acid
EP-PCR	Error prone polymerase chain reaction
exo^-	exonuclease deficient
FACS	Fluorescence assisted cell sorting
FKBP	12 kDa FK506 binding protein
FP	Fluorescent Protein
FPLC	Fast protein liquid chromatography
FRB	FKBP-Rapamycin binding domain
FRET	Förster resonance energy transfer
GECI	Genetically encodable Ca^{2+} indicator
GECO	Genetically encodable Ca^{2+} indicator for optical imaging
GFP	Green fluorescent protein
GPCR	G protein-coupled receptors
HBSS	Hank's balanced last solution

IDT	Integrated DNA technologies
kb	kilobase, nucleotides
Kbp	Potassium binding protein
LanRFP	Lancelet red fluorescent protein
LanYFP	Lancelet yellow fluorescent protein
mbUMT	mutant barley Uroporphyrinogen III methyltransferase
mNG	Monomeric NeonGreen fluorescent protein
MOPS	3-(N-morpholino)propanesulfonic acid
mRFP1	Monomeric red fluorescent protein number 1
N.A.	Numerical aperture
NAD+	Nicotinamide adenine dinucleotide
ncp	non-circularly permuted
NEB	New England Biolabs
NirE	Nitrate reductase biosynthetic pathway protein E (SUMT)
P2A	Ribosomal skipping sequence from porcine teschovirus-1 2A
PBG	porphobilinogen
PCR	Polymerase chain reaction

PEG	Polyethylene glycol
PDB	Protein data bank
PDB ID	Protein data bank identification
Pfu	*Pyrococcus furiosus* (species), and polymerase
POI	Protein of interest
Pr	CH_2CH_2COOH
PTM	Post-translational modification
QY	Quantum yield
SAH	*S*-adenosyl-L-homocysteine
SAM	*S*-adenosyl-L-methionine
SirC	Sirohydrochlorin biosynthetic pathway protein C
smURFP	small ultra red fluorescent protein
SPELL	Split protein reassembly by ligands or light
Sso7d	*Sulfolobus solfataricus* 7 kDa DNA binding protein
StEP	Staggered extension polymerase chain reaction
SUMT	*S*-adenosyl-L-methionine uroporphyrinogen III methyltransferases
Taq	*Thermus aquaticus* (species), and polymerase

TEV	Tobacco Etch Virus, refers to protease from TEV
YFP	Yellow fluorescent protein

Chapter 1: Introduction

1.1 Genetically encodable sensors for optical imaging

There are very few technologies that have accelerated our understanding of biology as much as the development and engineering of genetically encodable sensors for optical imaging. Illuminating the molecular dynamics of the cell has now become nearly routine due to countless improvements in genetic engineering, molecular modeling, and microscopy. Two of the most important genes for stimulating and enabling this entire field of research, the luciferase aequorin and the green fluorescent protein (GFP), were discovered in a single organism: *Aequorea victoria*, a bioluminescent jellyfish (1). These two proteins work in a concerted mechanism to convert chemical energy into green light (2). In the first step aequorin oxidizes the luciferin (general term for the substrate of a luciferase) coelenterazine using adenosine triphosphate (ATP) as a co-substrate. The product of this reaction is initially in an excited state that relaxes to the ground state, in isolation, this relaxation to the ground state would be accompanied by emission of a blue photon. However, in the jellyfish (where aequorin and GFP are located in close proximity to each other), this light energy is transferred to GFP through a non-radiative energy transfer mechanism known as bioluminescence resonance energy transfer (BRET) (3). This energy excites the GFP chromophore in the excited state, which then relaxes back to the ground state with the emission of aa green photon.

It was Osama Shimomura's early investigations of the green bioluminescent emission in jellyfish ultimately led to the vast array of genetically encoded bioluminescent and fluorescent proteins that are indispensable to modern biological research (4-6). Shimomura's laborious investigation of GFP isolated from the 'squeezate' of jellyfish led

to his characterization and prediction of the fluorophore structure (7). However, many years passed between Dr. Shimomura's groundbreaking work and the cloning of the GFP gene by Douglas Prasher (8). Cloning of the GFP expression enabled the expression of the GFP transgene in *Escherichia coli* and *Caenorhabditis elegans* (9). Famously, Martin Chalfie genetically fused the gene encoding GFP to a neuron specific protein to selectively illuminate neurons in *C. elegans*. The utility of GFP for this sort of application, was rapidly improved by the work of Roger Tsien and others. Mutagenesis of the gene encoding GFP gave rise to new spectrally diverse genetically encoded fluorophores capable of Förster (or fluorescence) resonance energy transfer (10, 11). These three researchers, Osamu Shimomura, Martin Chalfie, and Roger Tsien, were awarded the 2008 Nobel Prize in Chemistry "for the discovery and development of the green fluorescent protein, GFP".

Soon after the first report of expression of the GFP transgene came the first report of genetically encodable fluorescent protein sensors for Ca^{2+}: the cameleons (12). These Ca^{2+} sensors were engineered by taking a pair of fluorescent proteins that exhibit strong FRET when close together and linking them with a Ca^{2+} sensing domain composed of calmodulin and RS20. In the presence of Ca^{2+}, calmodulin and RS20 interact and bring the fluorescent proteins closer together and thereby increasing the FRET efficiency. As with fluorescent proteins themselves, this fluorescent sensor protein can be genetically targeted to different organelles using known genetically fused subcellular localization tags (13). Furthermore, because the Ca^{2+} sensing mechanism of calmodulin/RS20 has been extensively studied, cameleons can be tuned to have a K_d for Ca^{2+} that is appropriate for the organelle it is being targeted too (14). The advent of the cameleons was an impressive

demonstration of the versatility of fluorescent proteins and inspired further rapid expansion of fluorescent protein diversity and sensor development.

What makes fluorescent proteins so versatile as a reporters is their ability to auto-catalytically form a chromophore without the addition of exogenous co-factors except molecular oxygen (11, 15, 16). Canonical fluorescent proteins, derived from jellyfish, anemones, corals, copepods, and cephalochordates (17-20), consist of eleven β-strands that wrap around an α-helix to form a cylinder called a β-barrel. This β-barrel structure is critical for providing the environment necessary for fluorophore maturation and stabilization. This β-barrel structure is essential for fluorescence because it stabilizes the fluorophore microenvironment, as demonstrated by the fact that synthetic analogues of the GFP fluorophore are weakly fluorescent in aqueous solution (7, 21). The formation of the GFP fluorophore is widely accepted to progress through four stages: 1) cyclization of the peptide backbone, 2) loss of the of the peptide carbonyl oxygen by dehydration, 3) oxidation of the imidazolone conjugates the two ring systems, 4) deprotonation of the phenolate delocalizes the lone pair of electrons facilitating green fluorescence (Figure 1.1) (22, 23). However, there is still some debate regarding the order of the dehydration and oxidation steps (24).

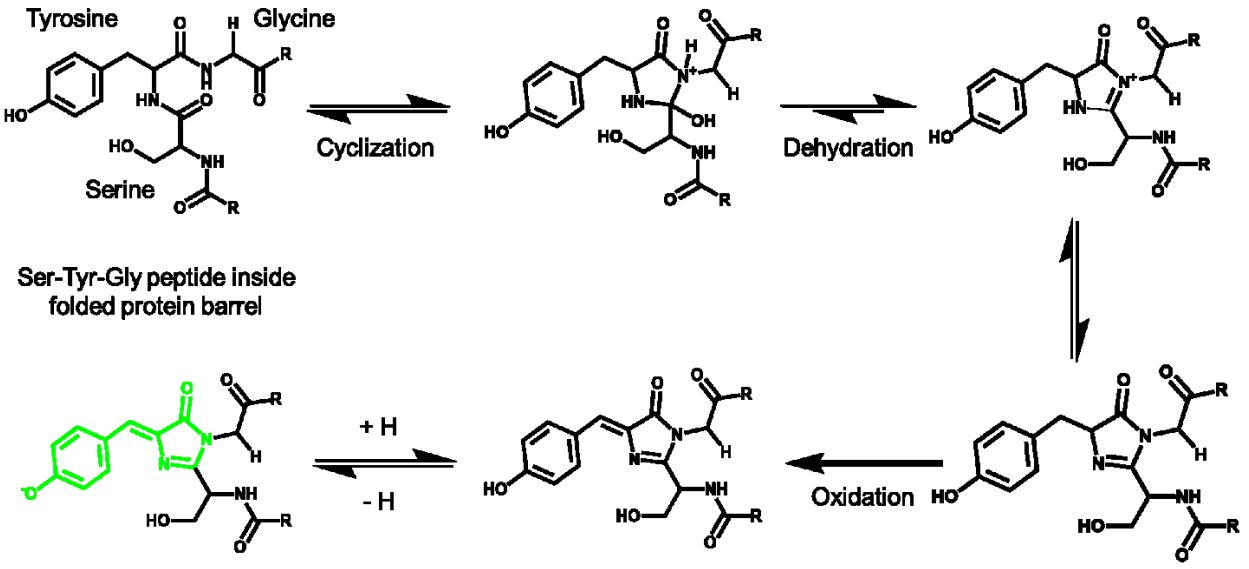


Figure 1.1 Mechanism of fluorescent protein fluorophore formation

The proposed mechanism of the autocatalytic maturation of GFP fluorophore. An α-helix that resides in the center of the β-barrel contains the Ser-Tyr-Gly peptide that folds into a conformation promoting the cyclization the of peptide backbone. Following cyclization of the peptide backbone, the five-membered ring intermediate undergoes a spontaneous dehydration followed by oxidation to extend a conjugated system that connects the phenol of the tyrosine side chain with the new imidazolidinone ring. Deprotonation of the tyrosine frees the lone pair of electrons from the phenolate oxygen to delocalize over the conjugated system.

Since the introduction of genetically encodable sensors in the mid 1990's, there has been a dramatic expansion of techniques for the expression, recording, and detection of optical reporters. For example, as alternatives to FRET based Ca^{2+} sensors such as cameleons, single fluorescent protein-based sensors have been engineered and have

been subject to a massive directed evolution effort towards optimizing them for imaging of neural activity (25, 26). In addition, the excitation and emission spectrum of fluorescent proteins have been expanded into the near-infrared range where tissue absorption and scattering are at their minimum (27). This spectral expansion into the near-infrared has facilitated the development of fluorescent sensors able to utilize excitation light that penetrates deeper into biological tissues (28). An alternative approach for imaging deep into tissue, which continues to increase in popularity, is the two-photon excitation of visible light absorbing fluorescent sensors with pulsed near-infrared excitation light (29-31). New neuro imaging techniques including 'BrainBow' and the 'Crystal Skull' have demonstrated *in vivo* optical recording in mice that can record potentially millions of neurons (32, 33). However, the simultaneous expression of spectrally distinct optical reporters has limited the utility of these techniques. This limitation has driven the expansion of orthogonal optical reporters. In addition to fluorescent proteins, luciferases have also been developed into sensors of Ca^{2+} and other analytes, and have been engineered to efficiently utilize synthetic substrates that emit in the near-infrared (34-37). As a demonstration, light from the AkaLuc/AkaLumine-HCl near-infrared luciferase system was observable within the brain of a live and awake marmoset after injection with the luciferin (34).

However impressive these new optical tools and methods may be, the extent of prototype sensor development has generally out-paced the optimization, implementation, and *in vivo* testing of these new imaging tools. This is acutely demonstrated by the lack of examples of multiplexed imaging of multiple analytes in model organisms. If genetically encodable optical sensors are to be fully utilized for the investigation of disease models

and cellular biology, an approachable, high-throughput, and modular design platform for sensor development and optimization will be required.

In this introductory chapter I intend to describe the basics of genetically encodable sensor engineering, based on generalized examples of sensors that are popular in the community. To this end, I will describe practical strategies for the modular design, efficient sensor construction, optimization of sensor prototypes, iterative directed evolution, methods of mutagenesis, and final optimization techniques. Because of the rapidly expanding diversity of genetically encodable biosensors, this work is not intended to be comprehensive (38). Where applicable, I will also expand on critical aspects of sensor design that may be particularly relevant for assisting the efforts of less experienced researchers.

1.2 Modular design of genetically encodable sensors for optical imaging

Aequorin was the first genetically encodable biosensor for optical imaging (39). The aequorin luciferase requires binding to calcium ions (Ca^{2+}) to enhance the enzymatic activity, making it a natural Ca^{2+} sensor. Observing Ca^{2+} transients in cell lines expressing the aequorin gene was one of the first demonstrations of a genetically encodable sensor that could optically report cellular phenomena. However, the necessity to supply the luciferin substrate for the light emitting reaction is a notable disadvantage of aequorin (40). In its native state, the *Aequorea victoria* jellyfish uses BRET to convert chemical energy into green luminescence. Considering the Ca^{2+} dependence on the BRET

mechanism, this system could be interpreted as a modular sensor. Through dissection and understanding of natural sensor systems we can begin to develop a sense of the modular nature of genetically encoded sensor design.

The engineering of genetically encodable sensors for optical imaging is best described in a modular framework. Fundamentally, there are two components to a sensor: the reporter domain and the sensing domain. Reporter domains can be generalized into three categories: fluorescent proteins, luciferases (enzymes that produce luminescence), and enzymes. Fluorescent proteins report changes in their sensing domain through modulation of their fluorescent intensity (41, 42). Luciferases report activity by increasing or decreasing their light output usually through changes in their catalytic efficiency, although it is conceivable that a reduction in quantum yield of the reaction could be an alternative mechanism. Generally speaking, enzymes are less desirable than fluorescent proteins or luciferases for sensor design because they require addition of substrate and are generally more difficult to engineer. However, enzymes have the distinct advantage of being catalysts and can therefore be expressed at very low levels yet generate very high signals that accumulate over time (43).

Sensor domains are more diverse than reporters. The critical aspect of a sensor component is a conformational change. Sensing domains must have a conformational change associated with the detection of the intended analyte. It is the job of the protein engineer to connect the sensing domain and the reporter domain such that the conformational changes modulates the fluorescence or enzymatic activity of the reporter. For fluorescent proteins this modulation usually involves changes in the protonation state of the fluorophore (44). For luciferases and enzymes this modulation involves

reconstitution of the active site either directly by restoring missing portions of the reporter or indirectly through an allosteric or inhibitory mechanism (45, 46). The development of prototypes with new sensor domains has been a major driver of sensor diversity. For example, switching the calmodulin binding peptide (RS20 from myosin light chain kinase to ckkap from CaMKK-CaMKIV) of a fluorescent protein based Ca^{2+} sensor dramatically increased the affinity for Ca^{2+} and improved the kinetics (47).

Selecting the appropriate reporter and sensor domains is a decision made at the behest of the researcher. A review of popular designs ought to provide initial direction for researcher's intent on developing sensors tailored to their specific experiments. For this work I will use some generalized illustrations to describe each component (Figure 1.2).

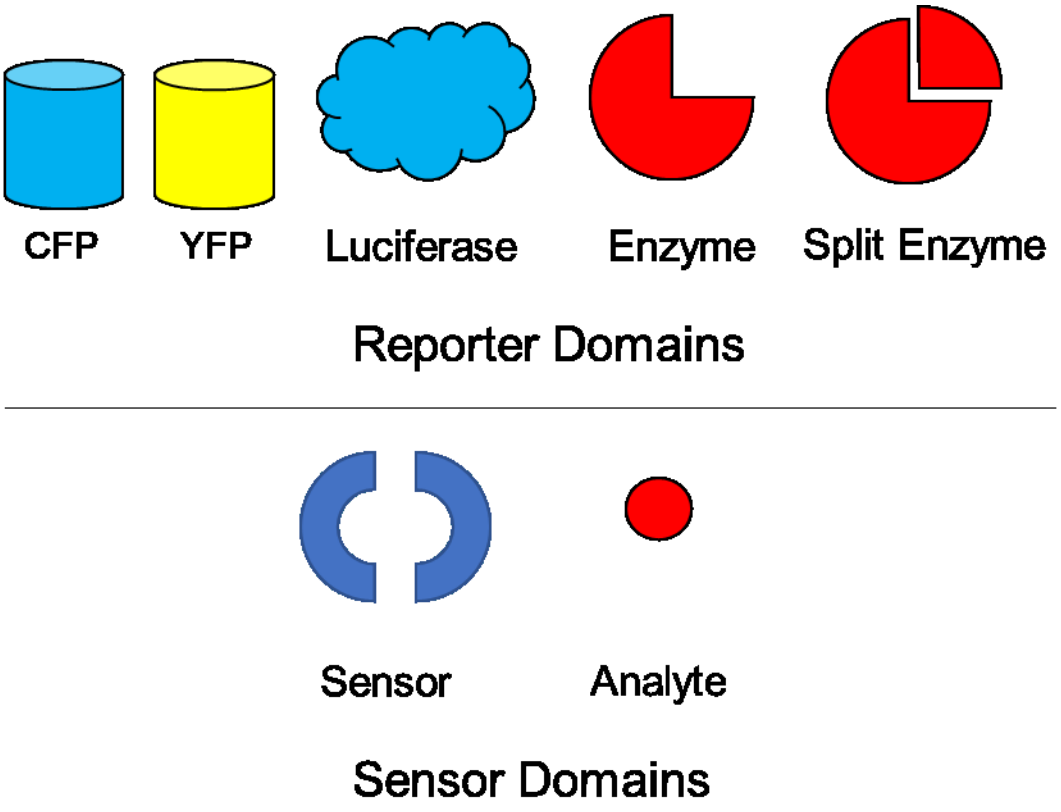


Figure 1.2 General illustrations of sensor components

Reporter domains are drawn as cylinders for fluorescent proteins, clouds for luciferases, ¾ pie for enzymes, and split ¾ and ¼ pie pieces for split enzymes. Sensor domains are generalized as blue half circles for the sensor portion of the protein with the analyte represented as a red circle. Cyan fluorescent protein (CFP) and yellow fluorescent protein (YFP) are used as FRET pairs in the following figures.

1.2.1 Previous designs

Inspiration for any genetically encodable sensor should be guided by the ever-expanding list of previously reported sensors (38). In the immortal words of chemist Frank Henry Westheimer: “A couple of months in the laboratory can frequently save a couple of hours

in the library" (48). Published sensors provide guidance to insertion sites and mechanisms that might otherwise take months or years to rediscover independently. It is advisable to identify the class of reporter or sensor domain that is best suited for a new application. For example, it is more efficient to use a Ca^{2+} sensing domain with the appropriate affinity for the organelle micro-environment than to try and modify the properties of a novel sensing domain. For example, several published sensors in the GCaMP series would represent appropriate starting points for selecting a low affinity Ca^{2+} sensing domain for observing Ca^{2+} dynamics in the endoplasmic reticulum (49-51). To aid in the dissection of sensing and reporter domains, it is best to categorize sensors into three classes: FRET and BRET sensors, complementation/reconstitution (split) sensors, and single fluorophore sensors.

FRET and BRET sensors require that the sensing domains interact in a manner that brings a donor fluorophore and an acceptor fluorophore (or chromophore) for FRET, or a donor luciferase and acceptor fluorophore (or chromophore) for BRET, into close proximity (less than 10 nm) to increase their energy transfer efficiency (Figure 1.3 A) (52). This can be accomplished with a single polypeptide sensor such as the cameleon Ca^{2+} sensor or as protein-protein interaction domains such as the Rapamycin inducible dimerization of FKBP/FRB (described in section 1.2.3) (12, 53). Protease RET sensors work in reverse: an active protease decreases energy transfer efficiency by cleaving a specific sequence within a single polypeptide that contains both the donor and acceptor, allowing the reporter domains to separate (Figure 1.3 B) (54).

A novel sensor design that uses BRET to a single fluorophore reporter has been recently developed (Figure 1.3 D) (55). LUCI-GECO1 is a fusion of NanoLuc luciferase

and non-circularly permuted GCaMP6 (56, 57). The Ca^{2+} sensor component, GCaMP6, modulates BRET efficiency by increasing or decreasing the extinction coefficient. Accordingly, BRET from the luciferase is converted to dynamic Ca^{2+}-dependent signal by GCaMP6. This sensor combines two powerful optical imaging techniques into one: the luciferase activity of NanoLuc with the highly optimized Ca^{2+} sensing domain of GCaMP6.

FRET sensors are recommended for sensor domain testing and prototype development. Small changes in distance or orientation of the fluorophores can be detected as changes in FRET efficiency that are manifested as changes in the fluorescence emission ratio. This makes FRET an attractive first option for testing new sensing domains. The simplest design is to fuse the sensing domain between a CFP and a YFP. This design, CFP-sensor domain-YFP, has been used extensively to test new sensor domains. The response of FRET sensors can be quantified, detected ratiometrically, and corrected for photobleaching with routine image acquisition and processing techniques (58). A combination of simple design and high sensitivity makes FRET sensor designs a first-choice for testing new sensing domains. As an example, Ca^{2+} and K^{+} ion sensors were developed as FRET sensor before porting the sensing domains to single fluorophore reporter scaffolds (59, 60).

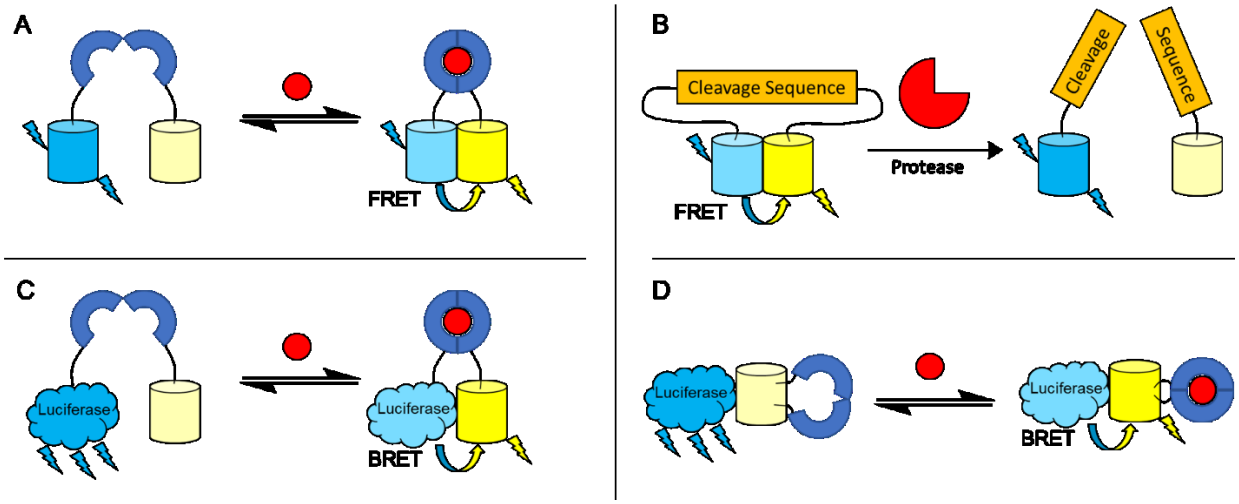


Figure 1.3 FRET and BRET sensor designs

Several FRET and BRET designs have been used to construct sensors. (A) FRET between two fluorescent proteins. Interaction or conformational change in the sensing domain brings the fluorophores closer together, increasing FRET efficiency and resulting in increased emission from the acceptor. (B) An active protease cleaves a peptide that maintains FRET between two fluorescent proteins. (C) BRET sensor utilizing a fluorescent protein acceptor. The luciferase constantly emits light, but the BRET efficiency increases as the sensor domains bring the reporter domains in close proximity. (D) BRET to a fluorescent protein with a sensing domain. BRET efficiency is constant, but acceptor efficiency is regulated by the sensing domain.

Split reporters can be derived from whole reporters by splitting the gene and fusing each portion to sensing domains. The simplest of these designs are the split fluorescent proteins (Figure 1.4) (61, 62). Split versions of canonical fluorescent proteins, composed

of the eleven stranded β-barrel, offer high contrast but are not reversible. In practice, this means that once the proteins interact, and form a fluorophore, the sensing domains are irreversibly trapped in the complexed state. This is the result of the strong intermolecular forces that help to reorganize and reform the reporter domain. Another name for this reporter strategy is bimolecular fluorescence complementation (63). Recently, a split reporter based on the eel fluorescent protein UnaG has been demonstrated to be reversible (64).

In contrast to split fluorescent proteins, split luciferase reporters have been demonstrated to be reversible (45, 65). When the sensing domains interact, it promotes the complementation of the luciferase fragments which then become active. However, the affinity between the split luciferase domains is low enough for them to separate and lose activity in the absence of the analyte. The most developed and widely used split luciferase reporter is the split NanoLuc luciferase, although split version of the luciferases from firefly and *Renilla reniformis* are also routinely used.

Split enzymes are an expanding class of reporters (66). Split β-galactosidase (67), β-lactamase (68), Tobacco Etch virus (TEV) protease (69), dihydrofolate reductase (DHFR) (70), and Horseradish peroxidase (71) have all been demonstrated to be effective reporter domains when combined with a fluorescent substrate. The catalytic turnover of enzymes allows for the low expression of reporter while maintaining high fluorescence signals. Similar to luciferases, the utility of genetically encodable split enzyme sensors is hindered by the necessity to supply an exogenous substrate. Though, protease sensors can be paired with FRET fluorescent proteins to make a fully genetically encodable sensor (54).

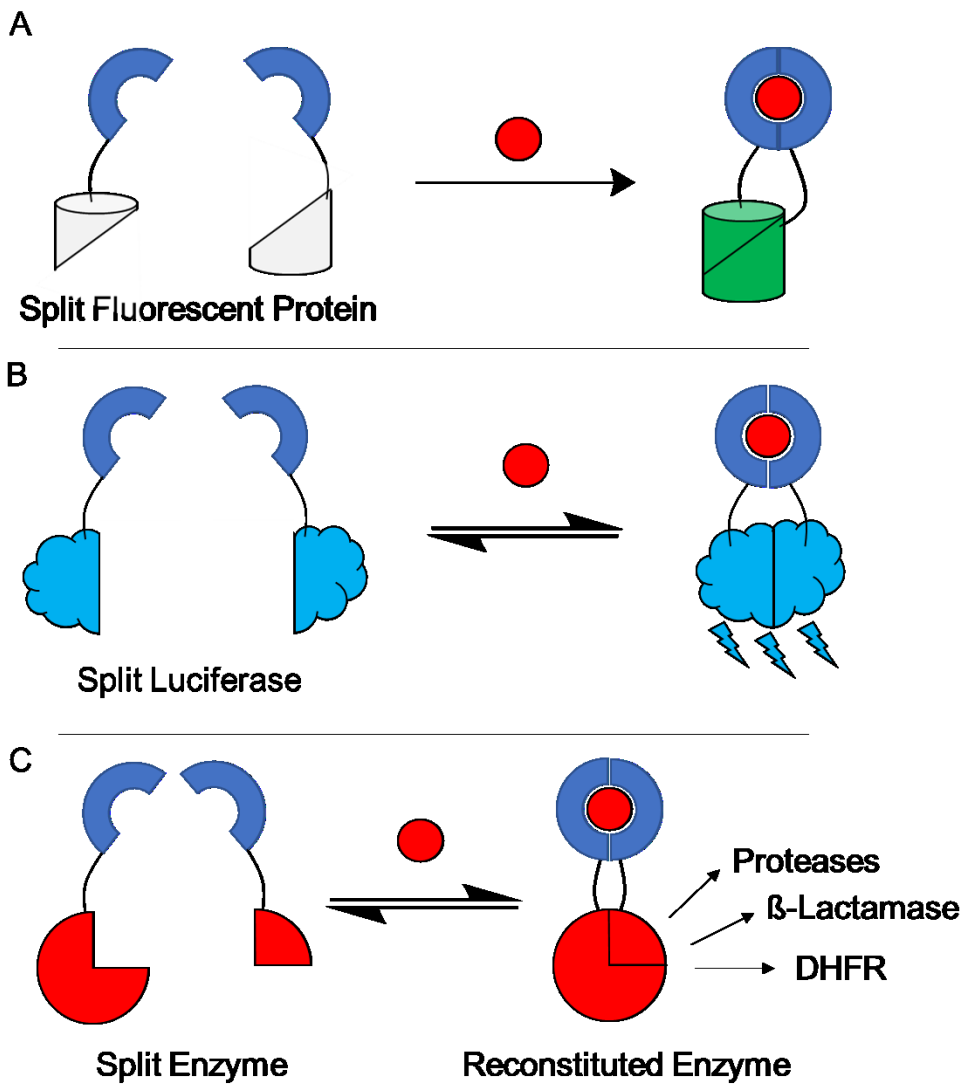


Figure 1.4 Split sensor designs

Split sensors are designed to complement and restore reporter activity when the sensing domain is activated. (A) Split fluorescent proteins lack a fully formed chromophore until the domains interact. Fluorescent protein complementation is generally irreversible. (B) Split luciferase is unable to catalyze the light emitting reaction until interaction of the sensing domains restores the full luciferase. Some luciferase reactions are reversible. (C) Several split enzymes have been developed. The restoration of enzymatic activity is dependent on the interaction of the sensing domains.

Single fluorophore designs are highly desired in order to make multiplex imaging of several optical reporters more practical and accessible. Since this class of sensors uses only one fluorophore (one color), more sensors can be optically imaged in a single experiment. In contrast, FRET sensors require two channels of optical recording to interpret a single sensor and thus require more “spectral bandwidth” in the visual to near-infrared spectrum. The constraint of fluorescent and bioluminescent reporters to work within the visible to near-infrared spectrum is a driving factor towards the development of single fluorophore of sensors.

Two topologies are popular with fluorescent protein sensors: circular permutation and non-circular permutation (Figure 1.5) (72, 73). Each topology has advantages and disadvantages for designing sensor prototypes. For single-domain analyte sensing domains that have been tested in FRET constructs, conversion to the non-circularly permuted fluorescent protein topology has a higher chance of success, since the topology of the sensing domain is preserved. In addition, the non-circularly permuted topology ought to be more compatible with fusions to other proteins or subcellular localization signal sequences because the original N- and C-termini of GFP are retained.

Circular permutation could be advantageous in certain cases because the fluorescent protein is less constrained which may lead to high contrast between the apo and saturated states of the sensing domain. Circular permutation is useful for domains that aren’t natively linked together in their natural protein or natural state. As an example, the evolution of the GCaMP series was built on fusing interacting peptides that sense Ca^{2+} (calmodulin and RS20) to the N- and C-termini. Because the calmodulin and RS20 domains have little affinity for each other in the Ca^{2+} free state, they are able to move far

from the fluorophore and produce very low apo state fluorescence. The result of this conformational freedom is a very high contrast upon binding Ca^{2+}. However, researchers investigating Ca^{2+} dynamics in organelles may find that localization tags and fusion to endogenous proteins can negatively affect the function of the sensor, since the fusion is linked via the functional portion.

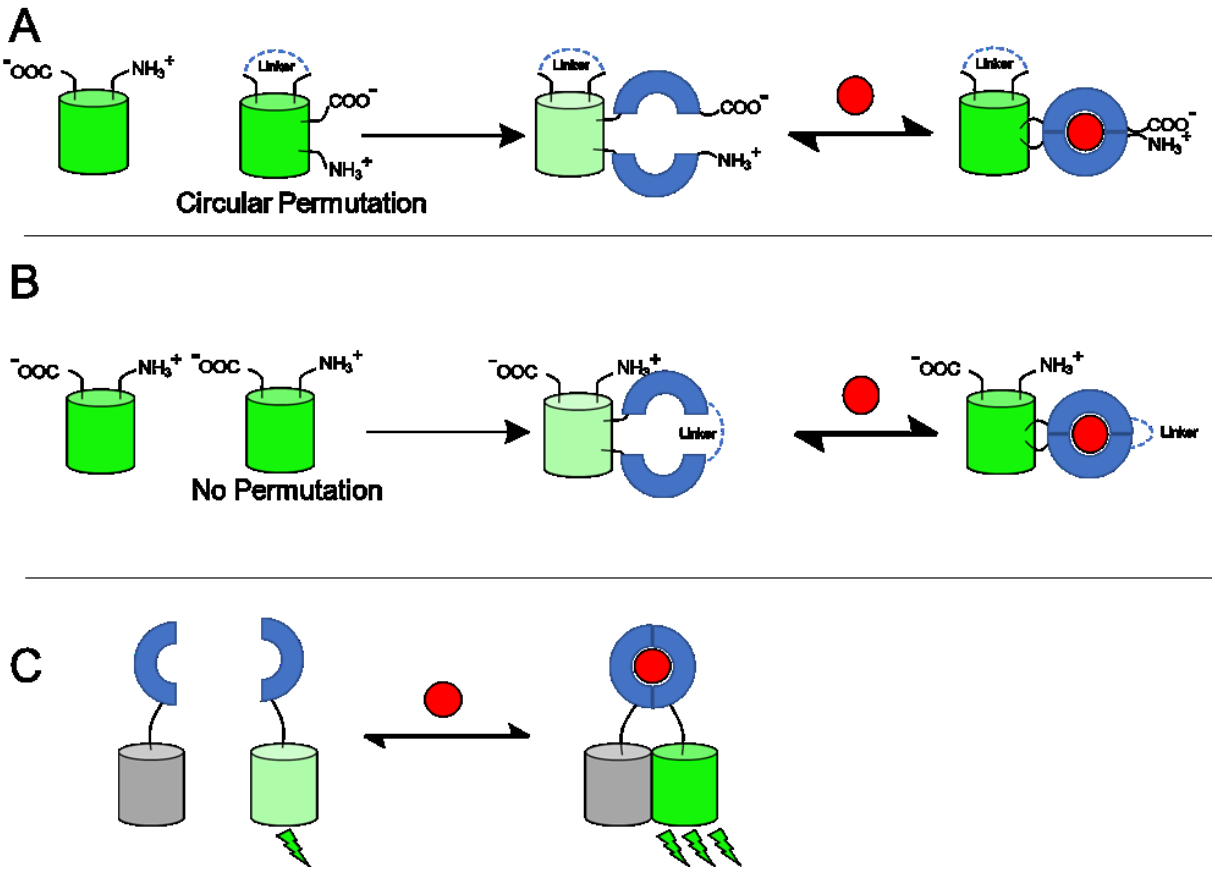


Figure 1.5 Single fluorophore designs

Single fluorophore designs utilizing fluorescent proteins. The general mechanism involves the restructuring of the immediate environment around the chromophore which modulates the fluorescent intensity of the sensor. (A) Circular permutation of the fluorescent protein involves genetically joining the existing termini with a short linker and introducing new termini elsewhere in the protein. The new termini are fused to sensing domains that modulate the fluorophore environment in response to the analyte. (B) Non-circularly permuted designs involve insertion of the sensing domains connected by a linker. The original termini are not modified which allows for the use of localization tags that do not directly affect the sensing domains. (C) Dimerization-dependent fluorescent

proteins have a single fluorophore but are dim without the stabilization of second non-fluorescent monomer.

1.2.2 Selecting a reporter domain

The development of canonical fluorescent proteins, those with an eleven stranded β-barrel, have been the cornerstone of fluorescent reporter engineering (74). Most β-barrel fluorescent proteins found in corals are tetrameric (75), which is undesirable for using fluorescent proteins as fusion partners. Fortunately, extensive protein engineering has been used to convert tetramer fluorescent proteins to dimeric and monomeric derivatives (Figure 1.6) (4, 76). The propensity to form dimers and tetramers varies between fluorescent proteins and this characteristic should be carefully considered when choosing a reporter domain (77). In contrast, the green fluorescent protein from the jellyfish *Aequorea victoria* (avGFP) is naturally a weak dimer. This is likely the result of avGFP being evolved for BRET as opposed to the fluorescent tetrameric proteins found in corals which have evolved for photostability, presumably to protect the host organism from photodamage (78). The versatility of the avGFP of fluorescent proteins for sensor design is evident from the numerous FRET and single fluorophore sensors available.

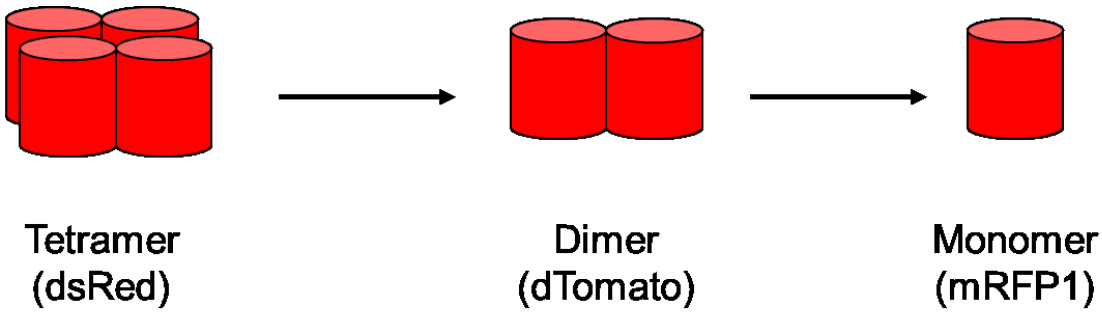


Figure 1.6 Engineering monomeric fluorescent proteins

The majority of fluorescent proteins are found in their native state as a single gene product that assembles to form obligate tetramers. dsRed, named after the organism from which it was discovered (*Discosoma* sp.), is an obligate tetramer in its wild type state. Extensive engineering to disrupt the dimerization face produced the dTomato series of proteins (d is and identifier for 'dimer'). Further engineering to disrupt the final dimer face yielded mRFP1 (m for monomer), a truly monomeric red fluorescent protein suitable for use in sensor development.

Biological tissues have a minimal absorbance of light within the range of 650-800 nm. This makes near-infrared fluorescent proteins highly valued as the next generation of *in vivo* reporters for whole animal studies. Furthermore, near-infrared fluorescence is compatible with canonical fluorescent proteins derived from jellyfish and corals for multiplex imaging. Near-infrared fluorescent proteins are a relatively new class of reporter domain (79-81). These proteins become fluorescent due to covalent binding of an endogenous chromophore, such as biliverdin or a phycocyanobilin, formed from the breakdown of heme. In mammals and higher eukaryotes, the catabolism of heme yields biliverdin and bilirubin (82, 83). In cyanobacteria, biliverdin is converted to a variety of

phycocyanobilins enzymatically (84). Application of reporter domains that use phycocyanobilin in mammalian cells typically requires supplementation with the chromophore, although researchers have also managed to produce it *in situ* (85). Recently, near-infrared reporters have been developed into protease activity sensors that increase in fluorescence after cleaving their substrate (86).

Luciferase reporter domains are becoming increasingly popular due to the discovery and engineering of brighter luciferase enzymes, and the development of improved substrates. Luciferases are unique as reporters because bioluminescence is associated with effectively no background signal. This is a major advantage over fluorescent reporter domains which can have high background in biological tissues due to autofluorescence. Furthermore, luminescent assays only require a dark environment and a sensitive camera. Conversely, fluorescent assays require a camera, light source, excitation, and emission filters, making the experimental setup more complex. To date, the only genetically encodable luciferase with a full biosynthetic pathway that can be recombinantly expressed, has been the bacterial Lux Operon (87-89). Having the entire biosynthetic pathway for the luciferin recombinantly expressible, negates the need to supply the luciferase substrate. However, the relative intensity of the transgenically expressed Lux Operon is quite low (90-92). The fungal luciferase biosynthetic pathway has recently been described and demonstrated for the development of autoluminescent yeast and mammalian cell lines (93). The pathway for firefly and marine organism luciferins, D-luciferin and coelenterazine respectively, have yet to be fully elucidated. NanoLuc is the smallest luciferase at 19 kDa and has been demonstrated as an effective split protein-protein interaction reporter (94). NanoLuc is a commercial product that has

been engineered to use the synthetic substrate furimazine (56), however, it can still efficiently utilize its native substrate coelenterazine.

Fewer directed evolution studies have been performed on luciferases compared to the numerous engineering efforts to create better fluorescent proteins. The majority of luciferase engineering has been performed on firefly luciferase. This can be attributed to the promiscuous active site of firefly luciferase which accepts a range of synthetic substrates (95, 96). Furthermore, the firefly luciferase emission wavelength can be modified through mutation of the active site (96-98). Three independent engineering efforts have produced firefly luciferases variants that can efficiently utilize synthetic substrates with near-infrared emission (34, 35, 99). New bioluminescent creatures with novel luciferase/luciferin pathways are still being discovered (100).

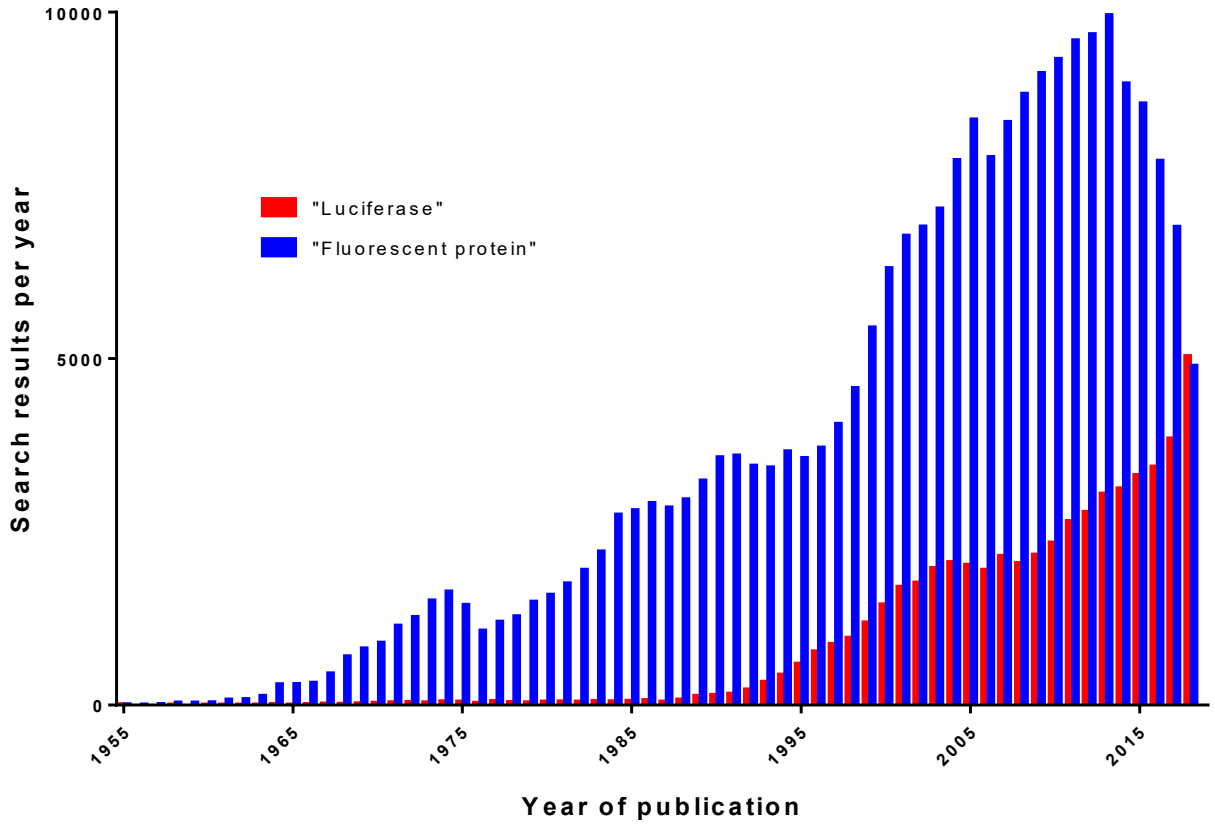


Figure 1.7 PubMed search results for common reporter domains

2018 is the first year that the search term “Luciferase” has more results than “Fluorescent protein”.

Enzymes are another class of reporter domain. Often, the enzyme has been genetically split into two non-functioning fragments. When the split enzyme fragments are brought into close proximity, they recombine to restore enzyme activity. This makes split enzymes particularly suited for studying protein-protein interactions. Protein complementation assays are often engineered from split enzymes to produce a

fluorogenic or chromogenic output. When developing new split enzymes, the most common interaction domain for testing prototypes is the Rapamycin inducible dimerization of the 12 kDa FK506 binding protein (FKBP) and the 100-residue section of mTOR (101) known as FKBP-Rapamycin binding domain (FRB) (102, 103). The Rapamycin inducible dimerization is titratable and can be used at very low concentrations of the inducer (103). The relatively small domains and well documented interaction make FKBP/FRB a valuable experimental control when testing new split enzyme reporter domains. Furthermore, this interaction has been extensively characterized and tested in a wide range of cultured cell lines. Using the FKBP/FRB system to prototype a new sensor is highly effective while applying the prototype reporter domains to different proteins of interest is straightforward.

Table 1.1 Common reporter domains

Reporter Domain	Fluorophore/Luciferin	References	PDB ID
Fluorescent protein (11 stand β-barrel)	Auto-catalytic formation of fluorophore from amino acid sequence	Review (77) (92)	1GFL
UnaG Fluorescent protein	Bilirubin	UnaG (104) Split UnaG (64)	4I3B
Near-infrared fluorescent protein	Biliverdin, Bilirubin phycocyanobilin, phycochromobilin	Review (105)	5AKP 4XTQ 4O0P
Bacterial Luciferase	Flavin mononucleotide / long chain aldehyde	Review (88)	3FCG
Firefly luciferase	D-Luciferin and synthetic analogues	With substrate (106)	2D1R
Fungal luciferase	3-Hydroxyhispidin	(93)	N/A
Renilla luciferase	Coelenterazine	(2)	2PSD
Aequorin luciferase	Coelenterazine	(107)	1EJ3
NanoLuc	Coelenterazine (natural) or furimazine (synthetic)	NanoLuc (56, 108, 109) Split NanoLuc (46)	5B0U
Enzymes	Proteases, Dihydrofolate reductase, β-lactamase, Cas9 (several more)	Review (66)	1Q31 7DFR 5F9R

1.2.3 Selecting a sensing domain

A variety of sensing domains have been previously reported in the development of genetically encodable optical sensors (Table 1.2). Sensing domains for Ca^{2+} (and Zn^{2+} to a lesser extent) have been at the forefront of sensor design and a new sensing domain for K^+ has been described recently (110). To develop new sensors for multiplex imaging, starting with domains from previously published ion sensors and modifying the reporter domain is likely to have better success.

Table 1.2 Common sensing domains

Reporter Domain	Analyte	References	PDB ID
Calmodulin and RS20	Ca^{2+}	Review (111)	1QS7
Human centrin 3	Mg^{2+}	(112) (113)	2GGS
Zap1	Zn^{2+}	(114) (115)	1ZW8
Kbp and BON	K^+	(110) (116) (60)	5FIM
FKBP/FRB (Testing protein-protein interaction sensors)	Rapamycin	(103) (117) (118)	3FAP

1.2.4 Insertion of sensing domain into reporter domain

Once a prototype design with diligently chosen reporter and sensing domains has been selected, the arduous search for an insertion site can begin. This is generally the make-or-break step in sensor design. Often, the insertion of an entire protein into another protein is detrimental to the activity of both. However, if a specific site within the sequence of a

reporter domain has been previously characterized to accept the insertion of a sensor domain, this site ought to work for other similarly sized sensor domains. For example, numerous single fluorophore sensors based on the canonical β-barrel fluorescent proteins, and a variety of sensing domains, have been reported (38). This is largely a result of the designs using the same, well characterized, insertion site (in the vicinity of residue 145) close to the chromophore (119).

The development of novel sensors with reporter domains other than the canonical β-barrel fluorescent proteins has been more challenging. The mechanisms and methods by which the fluorescence of canonical fluorescent proteins can be modulated are relatively well understood. Conversely, near-infrared fluorescent proteins have a biliverdin fluorophore with entirely different chemistry and structure from that of canonical fluorescent proteins (120). Overall, the relatively less-studied near-infrared fluorescent proteins have had considerably less investigation of their structures and fewer mechanisms describing the modulation of fluorescence have been reported (121, 122). Because of this, the development of near-infrared fluorescent protein sensors has seen relatively little progress to date.

A common approach for identifying potential insertion sites in new reporter domains is to consider the B-factor (123). B-factor is a mathematical score of local temperature (to describe motion and disorder) in a protein modeled using X-ray diffraction data. Flexible regions (higher B-factor) are generally more accepting of insertions and more likely to retain the native function or folding of the protein. Therefore, identifying these flexible regions is the first step to developing a novel sensor. Rendering and visualization of B-factor is routine using molecular visualization software (124). There are other utilities

available to do this automatically such as the circular permutation predictor webpage CPred (125-127) (http://sarst.life.nthu.edu.tw/CPred/) and the split protein predictor SPELL (128) (https://dokhlab.med.psu.edu/spell/login.php).

As one example, locating sites in the near-infrared fluorescent reporter smURFP (small ultra-red fluorescent protein) by B-factor is straightforward, as illustrated in Figure 1.8. The protein is first rendered and colored by B-factor, and then the structural data is aligned and correlated with the gene sequence. Residues with high B-factor, and are solvent accessible when compared to the molecular model, are selected for designing insertion primers. Usually, several sites are selected within the reporter domain for insertion. There are two structural categories for insertion/fusion of the sensing domain: circularly permutated and non-circularly permuted.

Circular permutation (cp) creates new N- and C-termini in the reporter domain (Figure 1.8). This is possible by genetically linking the normal N- and C-termini with a linker of appropriate length. Next, the new termini are fused to the sensing domains. Non-circularly permuted proteins retain their native N- and C-termini by insertion of the sensing domain within the contiguous reporter sequence (Figure 1.9). The non-circularly permutated topology is likely to facilitate better subcellular localization when fused to a targeting domain, because it retains the native termini of the reporter protein. However, not all sensing domains will function properly if both termini are constrained. Some sensing domains require their termini to be relatively close together, others will require long linker regions. Careful examination of the expected conformational change of the sensing domain will provide insight as to which topology is likely to work best. Overall, the question

of “to cp or not to cp” and careful consideration of the particular reporter and sensing domains being used is best left to the discretion of the researcher.

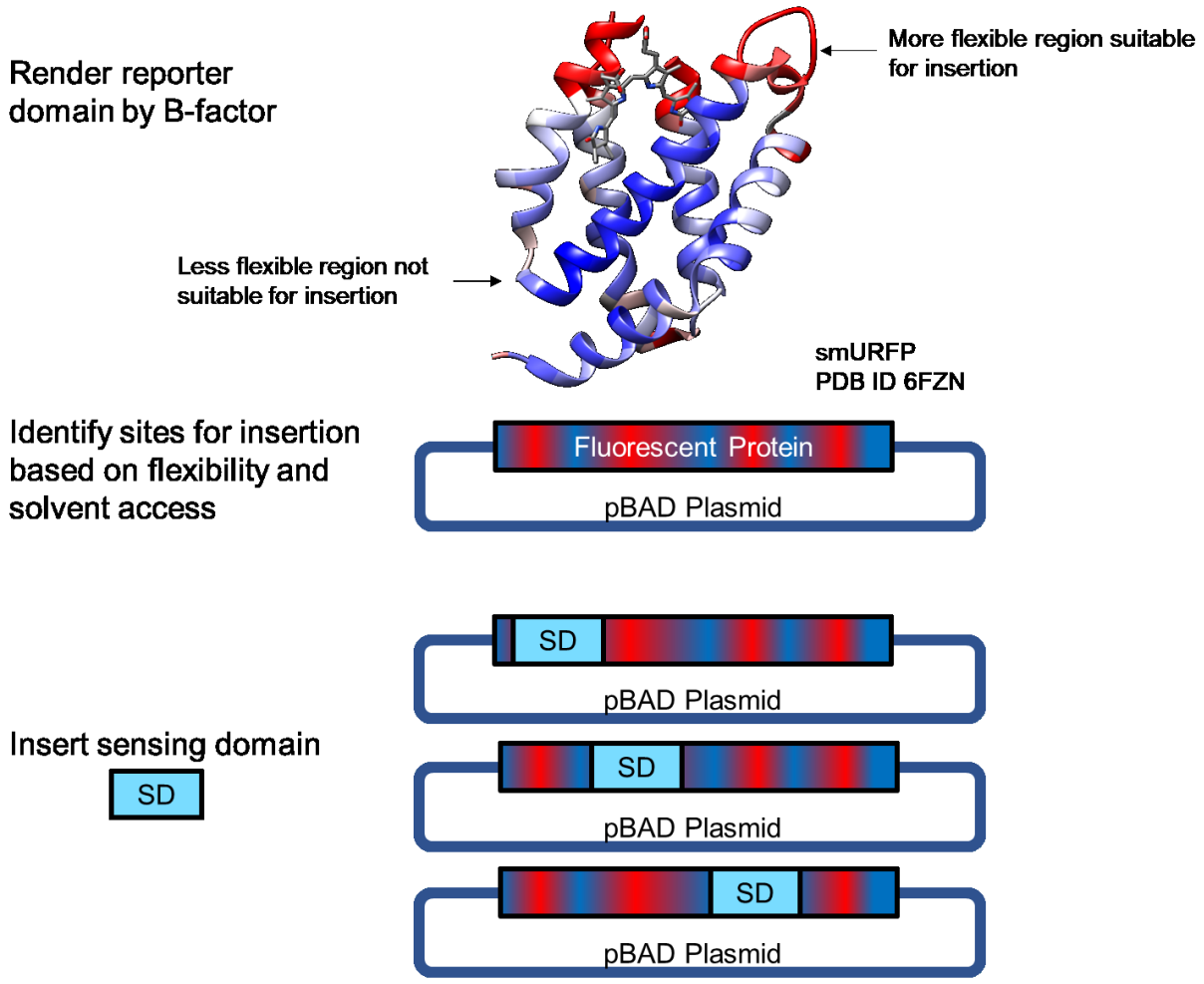


Figure 1.8 Rational design of sensor domain insertion sites

Using the X-ray structure information to determine sites for insertion of a sensing domain. Rendering the structure with residue coloring by B-factor (i.e., red for higher values, blue for lower values) indicates regions of the protein amenable to insertion of new sequences. Blue regions are rigid and red regions are flexible with white and pink regions in-between. Rationally selecting sites that are flexible and solvent accessible will increase the chances of finding insertions sites that fold correctly and retain the key properties (i.e., fluorescence or enzymatic activity) of the reporter domain.

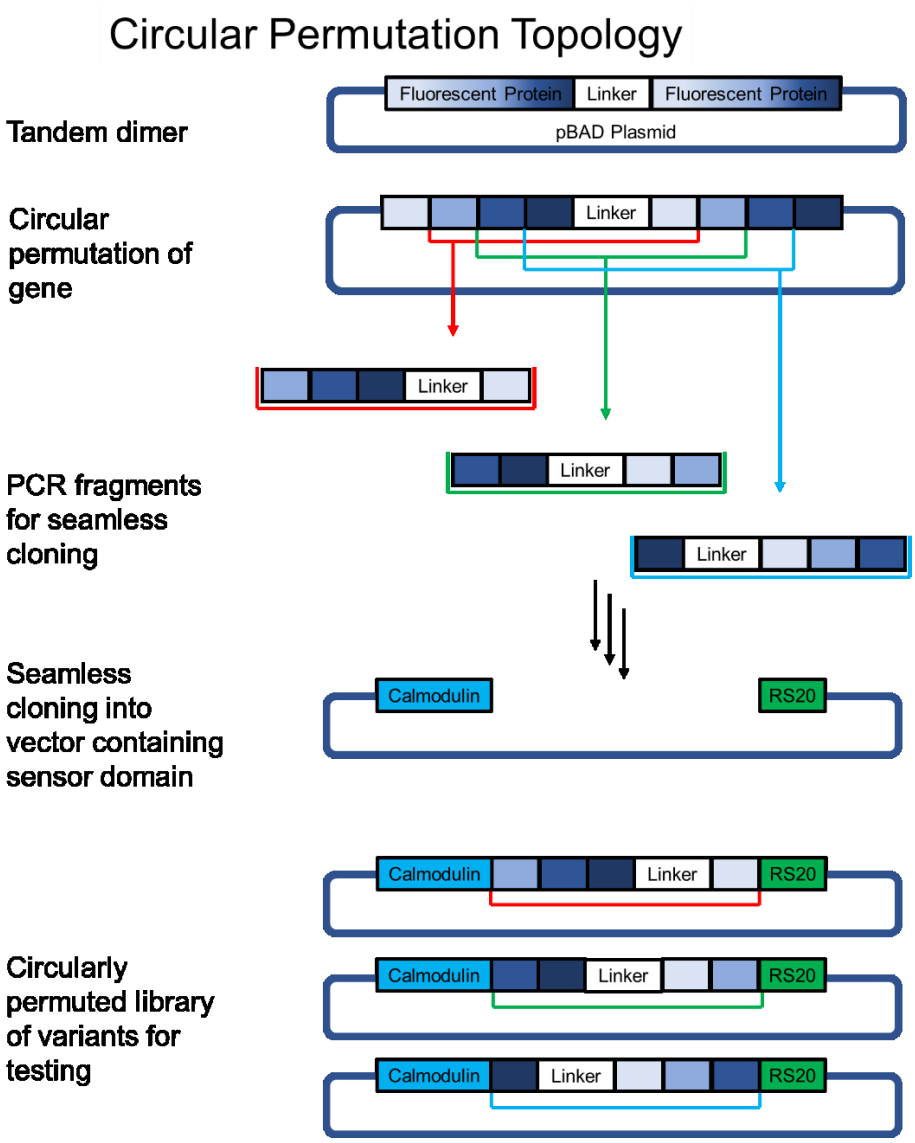


Figure 1.9 Design of circular permutation libraries

Circular permutation library design starts with constructing a tandem dimer of the reporter gene with a flexible linker in-between. A fragment of the gene that represents one whole protein with new N- and C-termini is then created using PCR. The circularly permuted gene fragment can then be cloned into a vector containing the sensing domains. In this illustration, the fluorescent protein is circularly permuted in three sites and cloned into a vector flanked with a Ca^{2+} sensing domain composed of Calmodulin and RS20.

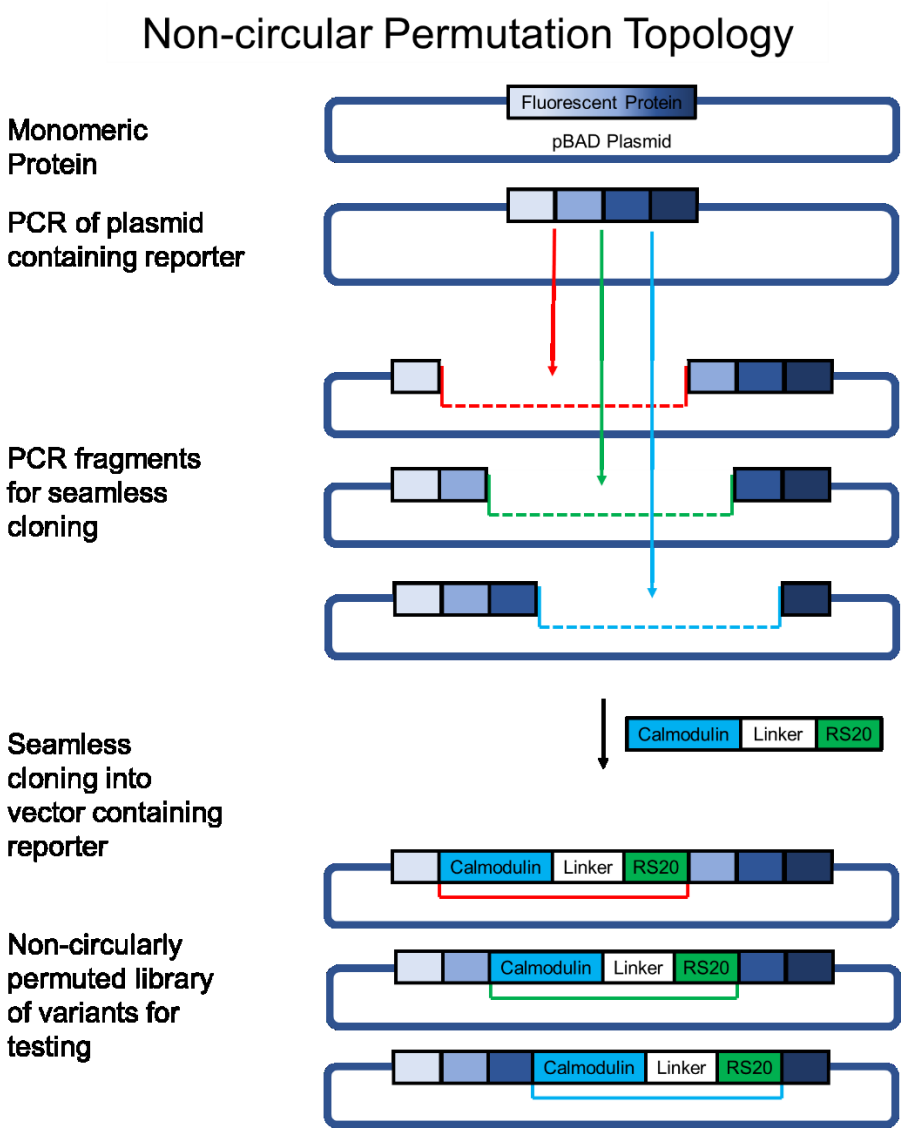


Figure 1.10 Design of non-circularly permuted gene libraries

Non-circularly permuted genes retain their N- and C-termini by inserting the sensing domain within the gene. PCR of the vector creates insertion sites suitable for cloning a sensor domain into. In this illustration, Calmodulin and RS20 (joined by a short linker) are inserted in to multiple sites within the fluorescent protein gene.

1.2.5 Optimization of the linker regions

The regions joining the sensor and reporter domains are referred to as linkers. Because these linkers will communicate the conformational change of the sensing domain to the reporter domain, they can be the key to successful prototype development. Designing insertion libraries with randomized linker regions is recommended when investigating new sensor topologies (Figure 1.10). It is often necessary to add or remove residues to optimize the insertion site and linker regions. An effective strategy for developing libraries is to remove two residues from the reporter domain at the insertion site and 'replace' them with residues that have been randomized to the 20 common amino acids using an NNK codon (where N = A, G, C, or T and K = G or T). In practice, inserting 2 or 3 randomized residues creates very large libraries that can be difficult to screen comprehensively. If possible, a primary screen on-plate for fluorescence or luminescence can effectively identify insertion sites that fold correctly and retain the function of the reporter domain.

To thoroughly investigate the insertion site, a set of PCR primers with randomized linker residues should be created as shown in Figure 1.11. To this end, create 3 primers at the N-terminal junction of the reporter and sensing domain while incorporating randomized residues (NNK, $(NNK)_2$, $(NNK)_3$) this will be primers 1 to 3. Then construct 3 primers in a similar fashion for the C-terminal junction of reporter and sensing domain, this will be primers 4 to 6. Combining the primers will create 9 different possible libraries e.g.) Primer 1 $\times$ Primer 4, Primer 1 $\times$ Primer 5, Primer 1 $\times$ Primer 6, and so on. A Punnett square can be used to visualize the possible libraries (Figure 12 B). Evaluation of variants from each library can then be used to develop a heat map of insertion libraries with favourable response. This is an efficient method to investigate insertion sites with minimal

primers. When these libraries have been exhausted and a sensors with a reproducible have been response characterized, the next step in sensor development is directed evolution.

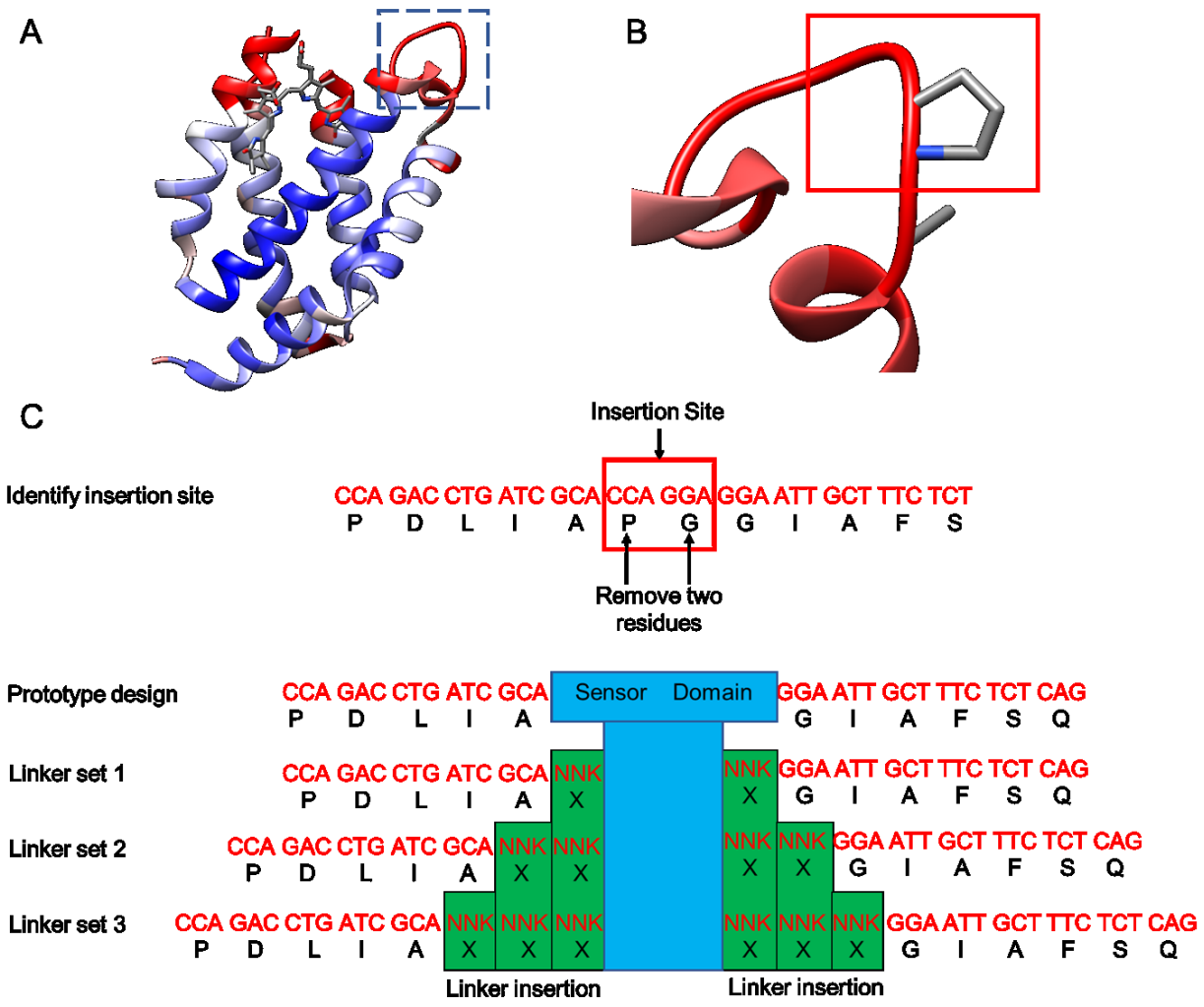
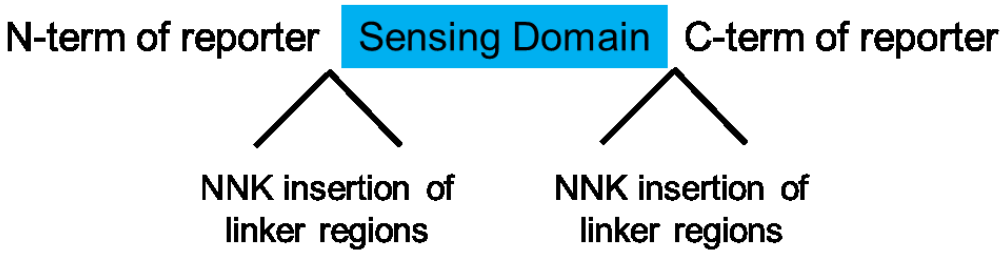


Figure 1.11 Insertion library design

(A) Identification of potential insertion sites based on B-factor. (B) Close up of insertion region, the red box are the two residues of the insertion site. (C) DNA sequences in red with protein sequence underneath in black. The two residues of the insertion site, proline and glycine, are removed and a random codon “NNK” is inserted in their place. Linker set one has no extra residues, linker set two has one randomized amino acid on each side of the sensor domain. Linker set three has two randomized amino acids on each side of the sensor domain.

A



B

Response in percent change	N-term-NNK-SD (Primer 1)	N-term-$(NNK)_2$-SD (Primer 2)	N-term-$(NNK)_3$-SD (Primer 3)
SD-NNK-C-term (Primer 4)	+	++++	+++
SD-$(NNK)_2$-C-term (Primer 5)	++	+++++ (best spacing)	++++
SD-$(NNK)_3$-C-term (Primer 6)	+	++	++

Figure 1.12 Optimization of insertion site

The optimization of an insertion site can be investigated using 6 primers to construct 9 libraries. (A) Visual representation of the N-terminal to sensing domain insertion site and the sensor domain to C-terminal insertion site. (B) A Punnett square of all possible combinations of the 6 primers. Using 3 N-term primers and 3 C-term primers 9 libraries can be created. In this arbitrary example, variants from each library were selected and scored using a scale of 1 to 5 "+" symbols. The library with "+++++" has the best spacing and it was created using Primer 2 and Primer 5.

1.3 Testing the sensor

The method for testing the prototype sensor is dependent on what analyte is being measured. For ion indicators, pelleting the cells and extracting soluble protein with commercially available protein extraction reagents is the first method to try. For example, Bacterial Protein Extraction Reagent (B-PER, ThermoFisher) is routinely used to extract soluble protein. The indicator in the clarified lysate can then be tested with dissolved salts of Ca^{2+} and the Ca^{2+} chelator EGTA (ethylene glycol-bis(β-aminoethyl ether)-N,N,N′,N′-tetraacetic acid). Testing equal amounts of lysate containing the sensor in Ca^{2+} and EGTA provides a quantification of sensor response. This saturation/chelation methodology can be applied to other ion indicators. For instance, EGTA is more selective for Ca^{2+} over Mg^{2+}, however EDTA (ethylenediaminetetraacetic acid) is more selective for Mg^{2+}.

Testing for response to small molecules analytes is more challenging as there are no options to specifically chelate and deplete the molecule as is done with EGTA to remove Ca^{2+}. However, identification of enzymes that can modify the analyte may offer an alternative approach. As an example, sensors for the neurotransmitter glutamate could be tested by saturating the protein with exogenous glutamate then removing the glutamate enzymatically with glutamate dehydrogenase.

1.4 Directed evolution

1.4.1 Brief history of directed evolution

Natural selection is the process where heritable characteristics in a genetic population are selected for over successive generations of reproduction and recombination. This

process was first described by Charles Darwin in 1859 (129), outlined by Gregor Mendel in 1865 (130), and rediscovered and confirmed by multiple researchers in the early 1900s. The chemical basis of natural selection was first proposed by Nobel laureates Francis Crick, James Watson, and Maurice Wilkins in 1953 after determining the structure of DNA (131, 132). This prediction was confirmed with the elucidation of the genetic code for which Marshall W. Nirenberg, Har Gobind Khorana, and Robert W. Holley shared the Nobel Prize in Physiology or Medicine 1968 (133). In the 1980s the development of polymerase chain reaction (PCR) by Kary Mullis transformed molecular biology and genetics by enabling researchers to amplify gene fragments from complex mixtures of DNA quickly and easily (134). PCR is the workhorse of modern life sciences research. The 1993 Nobel Prize in Chemistry was awarded jointly to Kary Mullis for the development of PCR and to Michael Smith for his development of site-directed mutagenesis (135) (136). These landmark achievements, plus the contributions of countless other research researchers, helped defined and reinforce the ‘Central Dogma of Biology’: the genetic mechanism of heredity and the flow of information from DNA to RNA to protein (137).

In the early 1990s an enzyme researcher with training in engineering, Frances Arnold, began experimenting with randomly mutating proteins and selecting the best variants for further mutation. This method led to what is now known as directed evolution. Directed evolution is powerful technique that forms the basis of modern protein engineering (138). By random mutation of a gene or site-directed mutation of specific portions of a gene, new proteins with beneficial properties can be discovered. Using the best variants in iterative rounds of selection can produce proteins with properties that are vastly improved over the natural protein. This technique was recognized with the Nobel Prize in Chemistry

for 2018, which was shared between Frances Arnold (50% of prize) for directed evolution of enzymes, and the other half jointly awarded to George Smith and Gregory Winter for the development of phage display of peptides and antibodies (136).

Directed evolution techniques for protein engineering and their application to genetically encodable sensors for optical imaging have produced an incredible diversity of functional fluorescent probes for microscopy.

1.4.2 Methods of mutagenesis

Mutagenesis in directed evolution is centered around two techniques: error-prone PCR (EP-PCR) and site-directed mutagenesis (135, 139). EP-PCR is a method of amplifying DNA and producing nucleotide mutations at random positions. Central to EP-PCR are thermostable DNA polymerases that lack a functional exonuclease domain. In nature, the exonuclease domain has evolved to lower the mutation rate of DNA polymerases by identifying nucleotide mismatches and excising them from the growing DNA strand. Using DNA polymerases lacking an exonuclease domain, or that have an exonuclease domain with reduced nuclease activity, results in a higher rate of nucleotide mismatches incorporated into the PCR product.

Site-directed mutagenesis is a technique to make mutations at precise positions in a gene (135). Using large PCR primers with either specific mutations, or degenerate codons that code for 2 to 20 different amino acids, a gene can be modified at the nucleotide level. This technique allows researchers to mutate a gene at specific positions and then evaluate the translated protein (or proteins in the case of degenerate codons) for activity. Furthermore, critical residues in an enzyme or protein can be investigate by site-directed mutagenesis. For example, a technique known as alanine scanning involves

systematically mutating amino acid residues to alanine and tests the resulting protein variant for activity (140).

Staggered Extension PCR (StEP) is a technique for combining random mutations from multiple variants of the same gene (141). In a StEP reaction, templates of several variants are combined in a single reaction and the extension time is lowered so that the extension of the growing DNA chain is relatively short in each round. By using a short extension time and many thermal cycles, short fragments of genes containing mutations act as large primers that can anneal to different genes. In practice, the results are a shuffling of mutations from different variants resulting in a library of new variants with recombined mutations. This technique is particularly effective when a round of directed evolution produces multiple promising variants.

1.4.3 Mutagenic DNA polymerases

The options for mutagenic DNA polymerases for directed evolution are limited. *Taq* polymerase has long been the standard because it is cheap, effective, and the rate of mutation can be controlled rather easily (139). The *Taq* mutation rate is increased in two ways: 1) using a skewed dNTP ratio in the PCR reaction; and 2) enhancing the mutagenic rate with the addition of Mn^{2+}. The mechanism by which having a skewed ratio of dNTP's increases the error rate is straightforward. Having a higher concentration of an individual nucleotide present will increase its chances of entering the active site of the polymerase and mistakenly being incorporated in the synthesized PCR product. Adding Mn^{2+} to the reaction modifies the active site by substituting for a catalytic Mg^{2+} ion. This replacement causes an increase in the rate of mismatched nucleotides (142). Between the dNTP ratio and Mn^{2+}/Mg^{2+} ratio, the mutation rate of the PCR reaction can be controlled.

However, an often-overlooked issue with *Taq* mutagenesis is the mutational bias. Due to specific interactions within the polymerase that are not fully understood, certain polymerases show a propensity to misincorporate certain bases over others (143). In *Taq* error-prone PCR, this bias results in more adenine to guanine (A → G) and thymine to cytidine (T → C) substitutions, relative to other possible substitutions (Table 1.3). Over multiple rounds of evolution this can decrease the chances of finding certain mutations with some mutations being substantially less likely to be observed in the library.

A relatively new method of creating random mutations in a gene is the commercially available Genemorph II mutagenesis kit from Agilent Technologies (144). This kit has a modified *Taq* polymerase and another patented error prone polymerase. Interestingly, the two polymerases have complementary nucleotide biases, and so combining them in the same reaction results in a more even distribution of mutations. Furthermore, this kit does not require the addition of Mn^{2+}, uses an even dNTP ratio, and produces blunt ended PCR products suitable for seamless cloning procedures such as Gibson Assembly. This contrasts with *Taq* polymerase which has terminal transferase activity resulting in the non-template addition of residues at the terminal 3' end of the gene making the PCR product less suitable for seamless cloning (145).

A search of the Genemorph II kit-related patents indicates that the second polymerase is a derivative of *Pyrococcus furiosus* (*Pfu*) polymerase. The native *Pfu* polymerase has an exceptionally low rate of mutation in standard PCR reactions and has become a workhorse of routine PCR in molecular biology labs. To increase the rate of mutation in *Pfu* polymerase, the researchers at Stratagene (later acquired by Agilent) mutated a catalytic residue in the exonuclease domain. This mutation reduced the polymerases

ability to remove mismatched nucleotides and therefore increased the rate of mutation. To further increase the rate of mutation, the exonuclease deficient Pfu polymerase was modified in the region that acts as a proofreading mechanism by detecting mismatch pairs. This exonuclease and proofreading deficient polymerase became the Mutazyme™ polymerase used in the Genemorph II kit.

Table 1.3 Mutation rates of mutagenic polymerases

		Transitions			
Polymerase	**dNTP ratio (A:T:C:G)**	**A→G**	**T→C**	**G→A**	**C→T**
Taq/Mn^{2+}	1:1:1:1	32.7	31.5	4.5	5.2
Taq/Mn^{2+}	1:5:5:1	14.5	14.5	7.1	7.1
Mutazyme I	1:1:1:1	5.3	5.3	22.3	22.3
Mutazyme II	1:1:1:1	9.3	9.3	13.5	13.5
Pfu (exo$^-$, D473G)	1:1:1:1	10.2	10.2	11.7	11.7

		Transversions							
Polymerase	**dNTP ratio (A:T:C:G)**	**A→T**	**T→A**	**A→C**	**T→G**	**G→C**	**C→G**	**G→T**	**C→A**
Taq/Mn^{2+}	1:1:1:1	7.4	8.9	1.1	2.9	0.7	0.7	1.5	1.8
Taq/Mn^{2+}	1:5:5:1	21.4	21.4	3.8	3.8	0.7	0.7	2.4	2.4
Mutazyme I	1:1:1:1	5.7	5.7	2.1	2.1	4.5	4.5	10.2	10.2
Mutazyme II	1:1:1:1	15.1	15.1	2.5	2.5	2.2	2.2	7.5	7.5
Pfu(exo$^-$, D473G)	1:1:1:1	14.9	14.9	3.9	3.9	3.9	3.9	5.5	5.5

From Ref. (146)

In our lab we have developed our own mutagenic *Pfu* polymerase by mutating residues that have a similar effect as those in the Mutazyme™ polymerase. An Asp215Ala mutation knocks out exonuclease activity and an Asp473Gly mutation reduces proofreading. We have named this Pfu variant aTaq (*a*fter *Taq*, pronounced “attack”) and a detailed protocol for purification and making error-prone libraries can be found in the

accompanying Technical Note: “Purification and use of aTaq mutagenic polymerase”. aTaq polymerase has a more even mutation bias compared to *Taq*/Mn^{2+}, creates blunt ended PCR products suitable for seamless cloning, uses an even dNTP ratio, does not require Mn^{2+}, and the rate of mutation is easily controlled by the number if thermal cycles. We further modified aTaq for enhanced processivity by fusing the non-sequence-specific DNA binding domain Sso7d to the C-termini with a short flexible linker (147). Fusion of Sso7d domain to *Pfu* has been demonstrated to increase performance during PCR. The Sso7d domain was previously used to develop the non-mutagenic high-performance polymerase Phusion (148).

A novel mutagenesis technique based on CRISPR/Cas9, known as EvolvR, uses a mutagenic polymerase fused to a nicking Cas9 protein (149, 150). When a guide RNA is present, the Cas9-polymerase fusion protein locates to a specific region of the gene to be mutated. The nickase activity of Cas9 opens the DNA for the mutagenic polymerase to displace a small, generally 20-200 nucleotide, gap with new DNA containing mutations in the nucleotide sequence. This technique allows a researcher to target many regions within a gene in a single reaction. The libraries resulting from this technique can be larger than the screening ability of most directed evolution researchers. Paired with fluorescence assisted cell sorting (FACS) this technique may become quite popular (151-153). However, due to the complexity and lack of supporting literature to date, this technique has few published examples of its utility or effectiveness.

1.4.4 Protein expression systems

The host organism for the recombinant production and testing of genetically encodable sensors is determined by the requirements of the sensor. For many small molecule and

ion indicators, bacteria are a suitable host. For sensors based on transmembrane proteins such as voltage indicators and G-protein coupled receptors (GPCR), higher order eukaryotes are required (154-156) There are several advantages and disadvantages to different hosts during sensor development. I will discuss the two most common host systems, and briefly compare their attributes as they relate to the development of genetically encoded sensors.

The bacterium *Escherichia coli* is the workhorse of molecular biology and directed evolution. The genetics of *E. coli* are well understood and the methods to express recombinant protein well described. Some notable attributes of *E. coli* are its fast doubling time (20 minutes), inexpensive growth media, high transformation efficiency, easy storage of competent cells, and ease of screening and colony picking by-hand from agar plates. However, bacteria generally lack the post-translational modifications of higher eukaryotes, have a different cell cycle, have vastly different cellular architecture, have different chaperones, and have different membranes and secretion mechanisms. For these reasons, protein evolved in *E. coli* may ultimately not perform as expected when expressed in another system such as mammalian cells. Despite this, for most sensors and screening techniques, *E. coli* is the first host organism to try.

Yeasts are increasingly becoming a valuable host organism to screen genetically encodable sensors (157). The most commonly used yeast species are *Saccharomyces cerevisiae* and *Pichia pastoris* (158). Both species are eukaryotes and have similar cell architecture and post-translational modifications as mammalian cells. Furthermore, yeasts can be engineered to have similar glycosylation patterns as mammals (159, 160). Because of this, the use of yeast as a host has facilitated the expression and testing of

several membrane bound proteins from humans and the development of GPCR based optical sensors (161, 162). Yeasts are relatively inexpensive to grow and require modest growth conditions that are generally available to most researchers. However, yeasts grow slower than bacteria and yeast competent cells must be made fresh every time a library is to be transformed. The transformation efficiency for yeast is generally lower than bacteria.

Several other protein expression systems have been developed based on cultured cell lines from mammals and insects (163, 164). Complicated transfection protocols and difficult selection procedures prevents these methods from becoming high-throughput (165). However, the major advantage of screening in mammalian cell lines is that the resulting genetically encoded sensors should have the most optimized performance when expressed in mammalian cells. This is because the cellular architecture, post-translational modifications, cell cycle, and interactome is as close to an ultimate target cell as possible. Screening in mammalian cells has been limited to the most advanced research labs with specialized and custom-made equipment (166).

One notable new expression system is the commercially available Vmax cell line. Vmax is a genetically modified and extremely fast growing (doubling time ~10 minutes) variant of the salt water bacterium *Vibrio natriegens* (167, 168). Vmax cells are compatible with most *E. coli* expression plasmids and antibiotics and the genome has been modified to include the T7 RNAP gene under a lac promoter similar to BL21 (DE3) *E. coli*. Vmax bacteria require a higher concentration of salts than *E. coli* making their growth conditions semi-orthogonal which limits contamination of bacterial cultures. Although Vmax has the

characteristics of a faster-growing bacterial screening host, it's adoption in the molecular biology community has been slow.

Table 1.4 Comparison of cell types for sensor screening

Organism	Positive attributes	Negative attributes
Escherichia coli	Fast (overnight), inexpensive, routine techniques, high-throughput	Lacks PTM, membranes and interactome different than mammalian cells
Yeast	Fast (two days), inexpensive, routine techniques, medium-throughput, some PTM	Competent cells cannot be stored, slower growth than bacteria,
Mammalian and insect	Full PTM, similar interactome, higher order metabolites, most accurate for sensor development	Low throughput, slow growth, difficult to screen, expensive growth media
Vmax	Fastest doubling time (10 minutes), similar expression and culture techniques as *E. coli*	Lacks PTM, competent cell preparation is difficult, low transformation efficiency
PTM, Post-translational modifications		

1.4.5 Iterative directed evolution

Engineering efforts toward the development of genetically encodable sensors for optical imaging are generally achieved using some amount of iterative directed evolution. In an iterative directed evolution experimental scheme, a prototype sensor is subjected to a combination of mutagenesis techniques to develop libraries of new variants. These libraries are first screened for optical activity to confirm expression and folding of the sensor. Variants with interesting optical properties are selected and cultured to produce enough protein for further evaluation. The second screen uses a crude protein

preparation for a rapid evaluation of sensor response. DNA from variants selected as the ‘winners’ from the secondary screen are then isolated and used as the template for the next round of iterative directed evolution. This methodology to improving sensor response is highly effective. If sequencing information is collected each round, the impact of certain mutations can be investigated further. Ideally, all mutations that arise during random mutagenesis are investigated in more detail using site-directed mutagenesis. A mutation found in random mutagenesis might confer better fitness of the sensor, but site-directed mutagenesis libraries can evaluate all potential amino acid residues at this critical position. When a mutation is found by screening of a randomly mutated gene, the residue can be considered a ‘hot-spot’ to be investigated in more detail to identify the optimal residue.

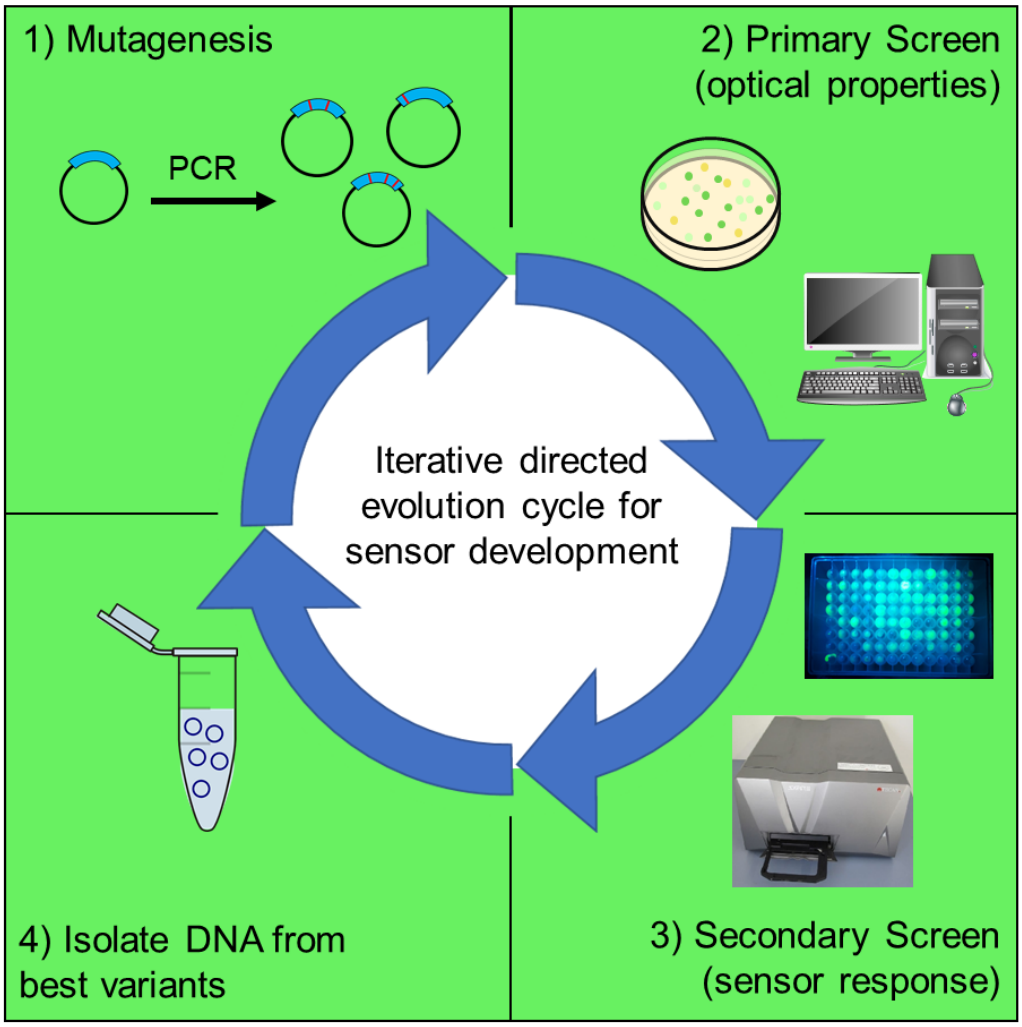


Figure 1.13 Iterative directed evolution scheme for optical sensor development

A simplified scheme for iterative directed evolution of genetically encodable sensors for optical imaging. (1) The gene encoding a sensor is mutated to produce a library of variants. (2) The library is transformed into a host cell. Each colony represents a single genetic clone. The primary screen selects for gene variants that have favourable optical properties (e.g. fluorescence, bioluminescence). (3) The variants selected in the primary screen are cultured so that enough sensor protein can be collected for response screening. (4) DNA from variants that are selected as 'winners' is collected and used in the next round of directed evolution.

1.5 Final modifications

Final modifications of the sensor are usually dependent on the end target cell type and subcellular localization. Often it is necessary to fine tune the affinity of the sensing domain in order for the sensor to be responsive to physiological-relevant concentrations, minimize the pH sensitivity, maximize expression level, minimize oligomerization, remove possible sites of post-translational modification, and add subcellular localization signal sequences. The biophysical properties of the sensor can be investigated with site-directed mutagenesis with a focus on residues that were mutated during evolution and residues in the linker regions between reporter and sensor domains. For example, a bright Ca^{2+} sensor with strong contrast between apo and saturated states might have a K_d that is outside the range required for the application that was initially conceived. Modification of the residues involved in Ca^{2+} binding and linker regions would initially be chosen for investigation.

Correcting for codon bias is another technique that can greatly improve expression and trafficking when moving sensors into higher eukaryotes. Highly evolved sensors that have been subjected to multiple rounds of directed evolution are often codon-biased for their particular evolution host. In effect, directed evolution selects for codons that express better in the host used for directed evolution, which is often different than the target organism for biological applications. This is because the available pool of tRNAs in a bacterium is different than yeast or mammalian cell. Mutations that change the codon to a more common tRNA, yet don't change the protein sequence, can increase expression which is often mistaken for increased activity. These mutations are often selected for when a full characterization is not performed on individual variants, which is a common

practice during directed evolution. Fortunately, codon optimization for a particular host organism is routine, and several DNA synthesis companies now offer this service.

Table 1.5 Codon usage table of four common protein expression systems

Residue	Codon	*E. coli*	*S. cerevisiae*	*P. pastoris*	*H. sapiens*
Phenylalanine	TTT	0.58	0.59	0.56	0.45
	TTC	0.42	0.41	0.44	0.55
Leucine	TTA	0.14	0.28	0.15	0.07
	TTG	0.13	0.29	0.33	0.13
	CTT	0.12	0.13	0.17	0.13
	CTC	0.1	0.06	0.08	0.2
	CTA	0.04	0.14	0.12	0.07
	CTG	0.47	0.11	0.16	0.41
Isoleucine	ATT	0.49	0.46	0.51	0.36
	ATC	0.39	0.26	0.31	0.48
	ATA	0.11	0.27	0.18	0.16
Valine	GTT	0.28	0.39	0.42	0.18
	GTC	0.2	0.21	0.23	0.24
	GTA	0.17	0.21	0.16	0.11
	GTG	0.35	0.19	0.2	0.47
Serine	TCT	0.17	0.26	0.29	0.18
	TCC	0.15	0.16	0.2	0.22
	TCA	0.14	0.21	0.19	0.15
	TCG	0.14	0.1	0.09	0.06
	AGT	0.16	0.16	0.15	0.15
	AGC	0.25	0.11	0.09	0.24
Proline	CCT	0.18	0.31	0.35	0.28
	CCC	0.13	0.15	0.16	0.33
	CCA	0.2	0.42	0.4	0.27
	CCG	0.49	0.12	0.1	0.11
Threonine	ACT	0.19	0.35	0.4	0.24
	ACC	0.4	0.22	0.24	0.36
	ACA	0.17	0.3	0.25	0.28
	ACG	0.25	0.14	0.11	0.12
Alanine	GCT	0.18	0.38	0.45	0.26
	GCC	0.26	0.22	0.25	0.4
	GCA	0.23	0.29	0.24	0.23
	GCG	0.33	0.11	0.06	0.11

Residue	Codon	*E. coli*	*S. cerevisiae*	*P. pastoris*	*H. sapiens*
Tyrosine	TAT	0.59	0.56	0.45	0.43
	TAC	0.41	0.44	0.55	0.57
Arginine	CGT	0.36	0.14	0.16	0.08
	CGC	0.36	0.06	0.05	0.19
	CGA	0.07	0.07	0.11	0.11
	CGG	0.11	0.04	0.05	0.21
	AGA	0.07	0.48	0.47	0.2
	AGG	0.04	0.21	0.16	0.2
Glycine	GGT	0.35	0.47	0.43	0.16
	GGC	0.37	0.19	0.14	0.34
	GGA	0.13	0.22	0.32	0.25
	GGG	0.15	0.12	0.1	0.25
Histidine	CAT	0.57	0.64	0.54	0.41
	CAC	0.43	0.36	0.46	0.59
Glutamine	CAA	0.34	0.69	0.62	0.25
	CAG	0.66	0.31	0.38	0.75
Asparagine	AAT	0.49	0.59	0.48	0.46
	AAC	0.51	0.41	0.52	0.54
Lysine	AAA	0.74	0.58	0.47	0.42
	AAG	0.26	0.42	0.53	0.58
Apartic acid	GAT	0.63	0.65	0.59	0.46
	GAC	0.37	0.35	0.41	0.54
Glutamic acid	GAA	0.68	0.7	0.58	0.42
	GAG	0.32	0.3	0.42	0.58
Cysteine	TGT	0.46	0.63	0.65	0.45
	TGC	0.54	0.37	0.35	0.55
Stop	TAA	0.61	0.48	0.53	0.28
	TAG	0.09	0.22	0.29	0.2
	TGA	0.3	0.3	0.18	0.52
Methionine	ATG	1	1	1	1
Tryptophan	TGG	1	1	1	1

https://www.genscript.com/tools/codon-frequency-table

1.7 Conclusion

Herein I have described a modular approach to genetically encodable sensor design and practical development for optical imaging. This modular design concept should provide future researchers a starting point for developing or modifying sensors. The rational approach presented in this work illustrates the engineering component of the sensor design. However, in most examples of sensor development, the random and seemingly

irrational directed evolution component teaches us the most about our creation. Using a proper balance of both methods will be required to master this discipline.

1.8 Scope of my thesis

In this introductory chapter I have provided a practical guideline, with supporting methodology, for the modular construction of genetically encodable sensors for optical imaging. My thesis introduction serves three purposes for interpreting the following chapters. First, this introduction covers the basic elements of optical sensors with the discovery, mechanism, and application of green fluorescent protein used as an example. In Chapter 2 I use techniques and rational engineering, reminiscent of those used in the early years of GFP development, to evolve a set of CFPs from mNeonGreen. Secondly, this introduction covers a brief history and methodology of directed evolution of proteins. Directed evolution is used extensively in my doctoral work and its application to future sensor development cannot be understated. Thirdly, the modular dissection of previously published sensor designs provides a framework for understanding the rationale behind the development of mNG-GECO (Chapter 3) and iCobA (Chapter 4). The final chapter of my thesis is a summary of the work contained herein with a brief discussion of the future directions.

Chapter 2: Structure guided engineering of a mNeonGreen-derived cyan fluorescent protein

2.1 Introduction

As we approach the 25th anniversary of the first transgenic expression of *Aequorea victoria* green fluorescent protein (avGFP) (169), the transformative influence of fluorescent proteins (FPs) for the investigation of cellular biology is astounding. The discovery and cloning of FPs from various clades, the majority from Anthozoa (marine invertebrates including coral and anemone), has blossomed into an impressive palette of practically useful genetically encoded fluorophores that span the visible spectrum from ultra-violet into the near infra-red (20, 77, 81).

The first generation of FPs demonstrated their versatility by illuminating protein expression and trafficking through their use in chimeric fusions to various proteins of interest (POI). By fusing the POI to spectrally separated FPs and using suitable excitation and emission filters, multiple cellular phenomena could be investigated simultaneously in a single cell. Further development of these imaging tools has focused on decreasing their oligomeric state (i.e., from tetramer to dimer to monomer) and improving their photophysical properties using directed evolution. Later generations of evolved FPs are generally brighter, more photostable, effectively monomeric, and have new properties including long Stokes-shift, photoswitching, photoconversion, and improved Förster resonance energy transfer (FRET) efficiency (170-172).

While the Anthozoan clade has been a gold mine for FP discovery, new structural classes and chromophores have been sought by genome miners to further expand the suite of imaging tools. Recently, the discovery of FPs from cephalochordates (i.e.,

lancelet) has yielded the brightest monomeric protein: mNeonGreen (173). This discovery has sparked a renewed interest in the development of brighter canonical FPs from a scaffold outside the Anthozoan clade.

Fluorescent proteins were first reported in cephalochordates in 2007 (19, 174). In 2013, Shaner et al. reported the development of the brightest green fluorescent protein, mNeonGreen (173). An extensive engineering effort, guided by the development of the first monomeric red FP (76), was executed in pursuit of a bright and versatile reporter capable of surpassing the performance of avGFP as a fusion tag for both traditional imaging and single-molecule superresolution imaging. In the five years since the initial report of mNeonGreen, there have been no reports of altered color variants of cephalochordate FPs, despite the structures of *Lancelet* yellow FP (LanYFP) and *Lancelet* red FP (LanRFP) being previously described (175, 176).

The pioneering efforts of the late Nobel laureate Dr. Roger Y. Tsien inspired us to attempt to engineer a series of color variants derived from mNeonGreen. Dr. Tsien's work on developing avGFP into blue and cyan shifted variants was the first demonstration that the spectral properties of FPs could be manipulated by modifying the protein coding sequence (10, 11). Initially, site-directed mutagenesis of the key spectrally determining residue in avGFP, Tyr66, was used to convert this residue into either phenylalanine, histidine, or tryptophan. This work led to the variants blue FP (BFP) (with Tyr66His), cyan FP (CFP) (with Tyr66Trp), and the Tyr66Phe variant (11). Subsequent site directed mutagenesis and directed evolution improved the photophysical properties of these avGFP variants and ultimately produced the BFP and GFP FRET pair (10).

The development of the BFP and CFP derivatives from avGFP guided our attempt to blue shift the cephalochordate repertoire of FPs. Using the previously published structures of mNeonGreen, LanYFP, and LanRFP, we introduced mutations that we reasoned were likely to result in altered fluorescent color. We decided to use site-directed mutagenesis to mutate the tyrosine in the mNeonGreen chromophore to another aromatic acid, while also randomizing the first residue (Gly68) of the chromophore to create a library. Color-shifted variants discovered in these initial libraries which could then be improved with directed evolution. Our initial attempt involved randomizing the first residue of the chromophore (Gly68) while simultaneously mutating the chromophore tyrosine (Tyr69) to phenylalanine (Try69Phe), histidine (Tyr69His), and tryptophan (Tyr69Trp). One very dimly fluorescent colony was found in the Y69W library that retained the glycine of mNeonGreen at position 68. This variant, designated NeonCyan0.1 served as the starting point for an extensive process of directed evolution that ultimately led to three blue shifted variants that we have named NeonCyan1, NeonCyan1-Thr207Asp, and NeonCyan1-Thr207Met.

2.2 Results and discussion

2.2.1 Iterative directed evolution of NeonCyan

Directed evolution of NeonCyan was implemented in an iterative fashion (Figure 2.1). Briefly, starting with NeonCyan0.1, we subjected the gene to a combination of error prone polymerase chain reaction (EP-PCR), site directed mutagenesis, and staggered extension PCR (177, 178). Beneficial mutations that arise from EP-PCR were typically subjected to further investigation, along with other residues in close proximity to the site

of mutation. To further investigate these residues, the position would be fully randomized to all 20 possible amino acids using site directed mutagenesis. When multiple improved variants were discovered in a round of screening from a library produced by EP-PCR, the corresponding plasmids would be mixed and mutations shuffled using staggered extension PCR. After 10 rounds of evolution it was apparent the protein started to plateau in terms of brightness and so we decided to halt the process of directed evolution.

The final variant, designated as NeonCyan1, has 9 mutations relative to mNeonGreen: Asp42Glu/Val67Met/Tyr69Trp/Ala146Val/Trp148Ser/Arg150Trp/Lys175Glu/Met201Val/ Lys213Glu. A very late stage variant, designated as NeonCyan0.95 (lacking only the Val67Met mutation relative to NeonCyan1), was sent to collaborators Dr. Antoine Royant and Damien Clavel (Institut de Biologie Structurale, Grenoble, France) for structure determination by X-ray crystallography.

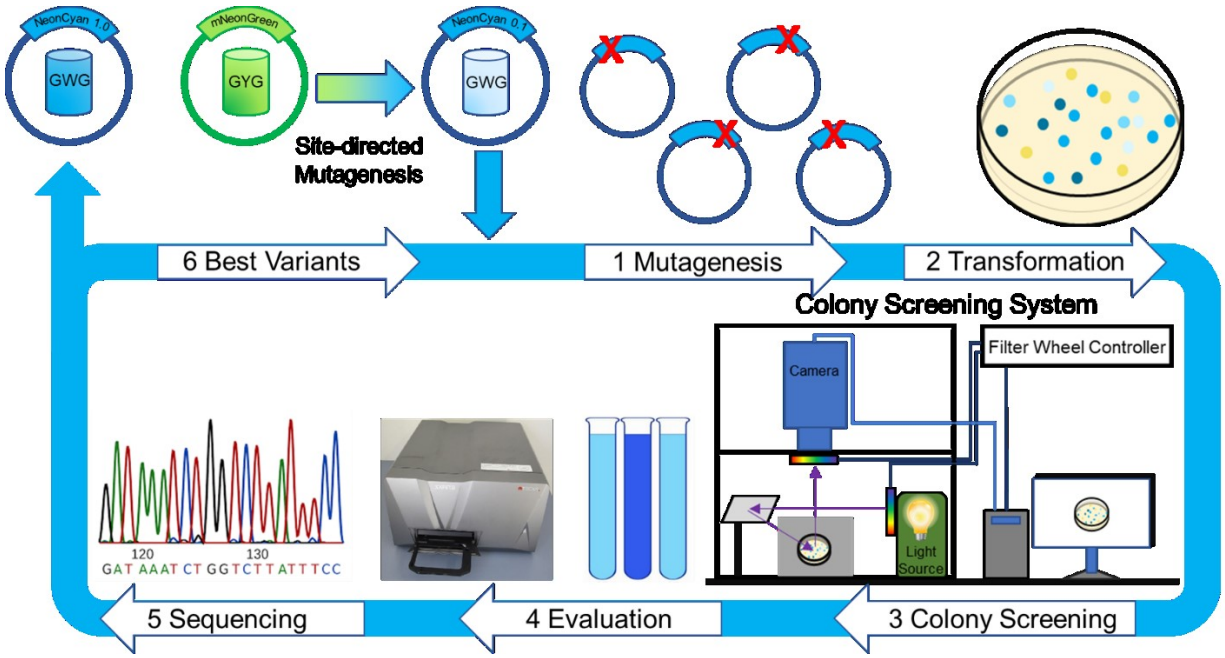


Figure 2.1 Iterative directed evolution scheme for NeonCyan

Directed evolution of NeonCyan fluorescent protein was performed in an iterative fashion. In a typical round, the brightest variants from the previous round are used as a template for EP-PCR. Mutant libraries are transformed into *Escherichia coli* and plated onto LB agar plates. A fluorescent colony screening system is used to identify the brightest colonies which are picked and cultured for further evaluation. The most promising variants are sent for Sanger sequencing and used as the template for the next round of iterative directed evolution (179).

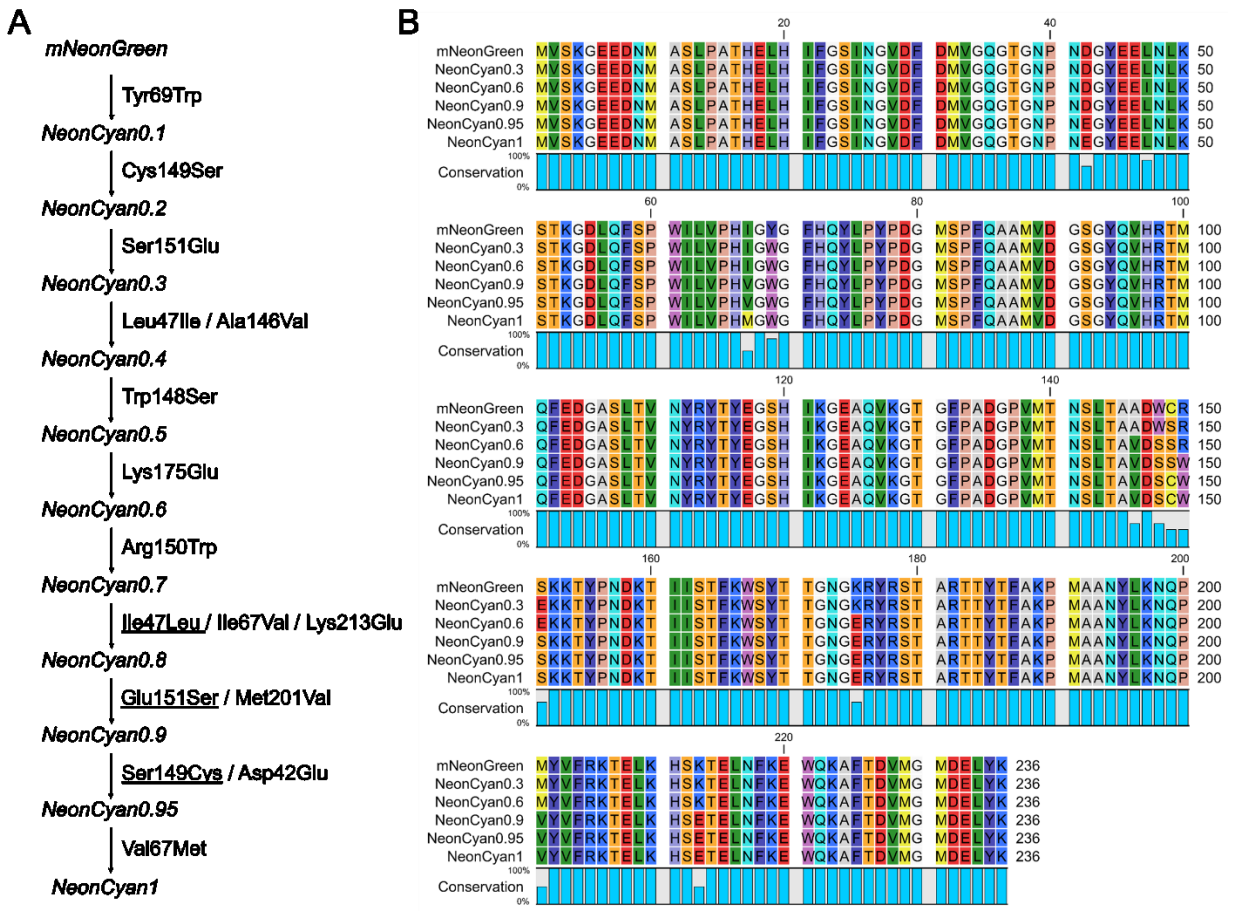
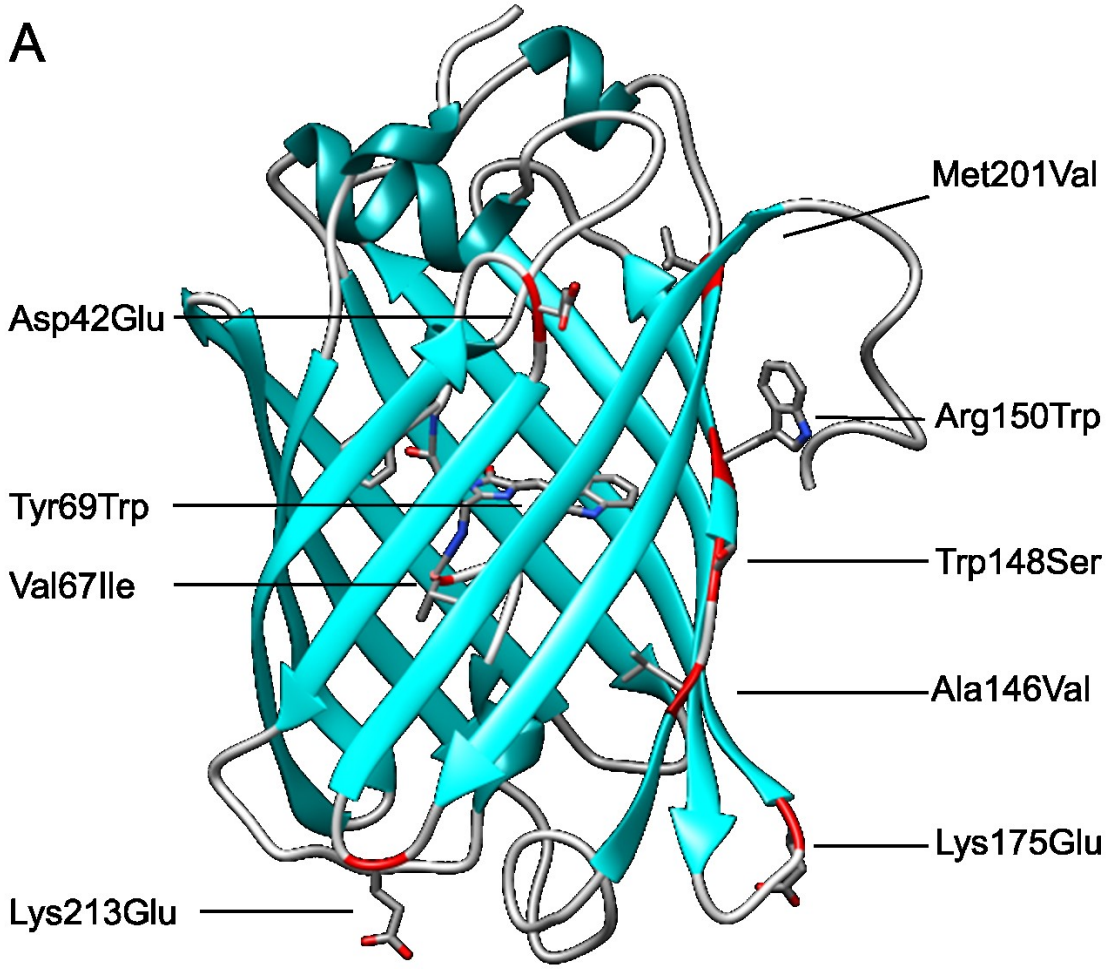


Figure 2.2 Lineage and alignment of NeonCyan variants

A) Lineage of mNeonGreen to NeonCyan1. NeonCyan1 has a total of 9 mutations and 3 reversions (underlined text) from mNeonGreen. B) Sequence alignment of NeonCyan1 to mNeonGreen with intermediates 0.3, 0.6, 0.9, and 0.95. A bar chart representing the percentage conservation of the most frequently observed residue, across the 6 variants displayed here, is shown underneath the alignment.


A
Met201Val
Asp42Glu
Arg150Trp
Tyr69Trp
Trp148Ser
Val67Ile
Ala146Val
Lys175Glu
Lys213Glu

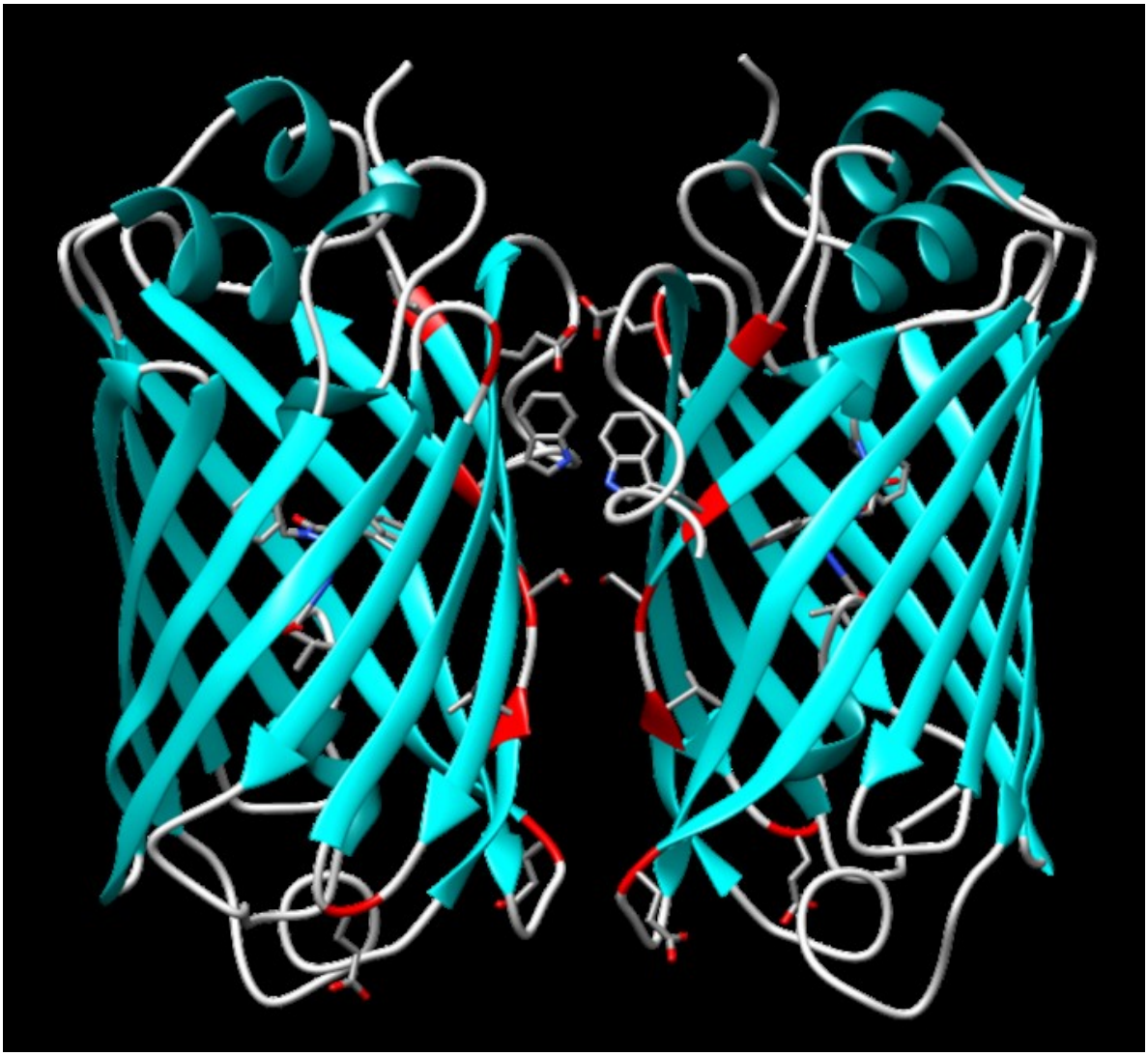


Figure 2.3 Crystal structure of NeonCyan0.95

A) X-ray structure of the monomeric unit of NeonCyan0.95 represented as a ribbon structure. The locations of mutations in NeonCyan0.95, relative to mNeonGreen, are highlighted red and sidechains are shown. B) The two structured loops located in the dimerization interface of NeonCyan0.95 are the C-terminal ‘MDELYK’ sequence.

2.2.2 Crystal structure of NeonCyan

Crystallogenesis

Crystallization screening was performed at the HTXlab (EMBL Grenoble, Fr) on 288 conditions (commercial screens: JCSG+, Wizard I&II and Classics Suite from Molecular Dimension and Rigaku, respectively). The initial conditions were reproduced and refined using the hanging drop vapour diffusion method with a drop size of 2 µL (ratio 1:1) at 293 K. The protein concentration used for crystallization was 20 mg/mL with a crystallization condition that contains 0.1 M HEPES pH 7.5 and 18% w/v PEG 8,000. Needle-shaped crystal grew within a week (Figure 2.4).

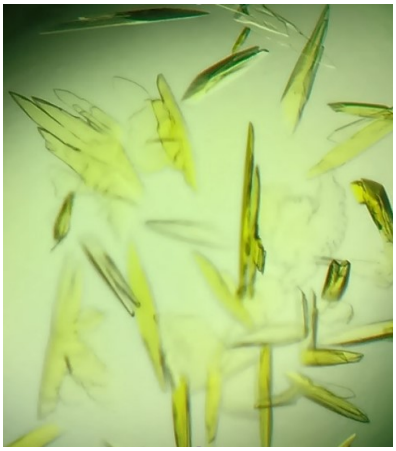


Figure 2.4 Crystals of NeonCyan0.95 obtained in 0.1 M HEPES pH 7.5 and 18% w/v PEG8000

Data collection, integration and reduction

Prior to automated data collection at the ESRF beamline MASSIF-1 (180), a NeonCyan0.95 crystal was progressively cryoprotected with a solution of mother liquor supplemented with 20% (v/v) glycerol. Data were collected at 100 K at 1.64 Å resolution,

integrated and reduced in space group P1 with the XDS program suite (181). Data reduction statistics can be found in Table 2.1.

Table 2.1 Data reduction and structure refinement statistics

Data integration and reduction	
Beamline	ID30A-1 (ESRF)
Wavelength (Å)	0.9660
Temperature (K)	100
Space group	P1
a, b, c (Å)	49,91, 60.68, 67.14
α, β, γ (°)	89.95, 71.75, 90.02
Resolution range (Å)	47.4 – 1.64 (1.70 – 1.64)
Number of measured reflections	138,647 (13,734)
Unique reflections	83,725 (8222)
Multiplicity	1.7 (1.7)
Completeness (%)	91.34 (90.43)
<I/σ(I)>	13.2 (2.0)
$CC_{1/2}$	0.999 (0.708)
R_{meas} (%)	5.3 (59.0)

Structure refinement	
Resolution range	
Molecules in the asymmetric unit	4
R_{work}	0.2364 (0.3528)
R_{free}	0.2788 (0.4028)
Number of atoms	
Protein	7056
Chromophore	84
Water	314
B-factors (Å^2)	
Protein	32.0
Chromophore	27.4
Water	30.3
R.m.s. deviations	
On bonds (Å)	0.014
On angles (°)	1.99

Structure determination and refinement

The structure of NeonCyan0.95 was solved by the molecular replacement method with the program Phaser (182) using the model of mNeonGreen at physiological pH as a search model (175) (PDB entry code: 5LTR). Model rebuilding was carried out in the visualisation software Coot (183) and the structure was refined with Refmac5 (184). Structure refinement statistics can be found in Table 2.1.

Structural analysis

The asymmetric unit of the crystal lattice contains four molecules of NeonCyan0.95 (Figure 2.5). The calculation of interaction surfaces between the various monomers (779 $Å^2$ between monomers A and C, 439 $Å^2$ between A and B, 312 $Å^2$ between A and D, 272 $Å^2$ between B and D, 272 $Å^2$ between B and D, 0 $Å^2$ between B and C) strongly suggests the existence of a physiological dimer formed by monomers A and C (Figure 2.6), while other contacts must only be due to crystal packing. The C-terminal DELYK motif folds as a helix (while it is mostly agitated in engineered FP variants containing this very motif) and contributes to the interaction interface.

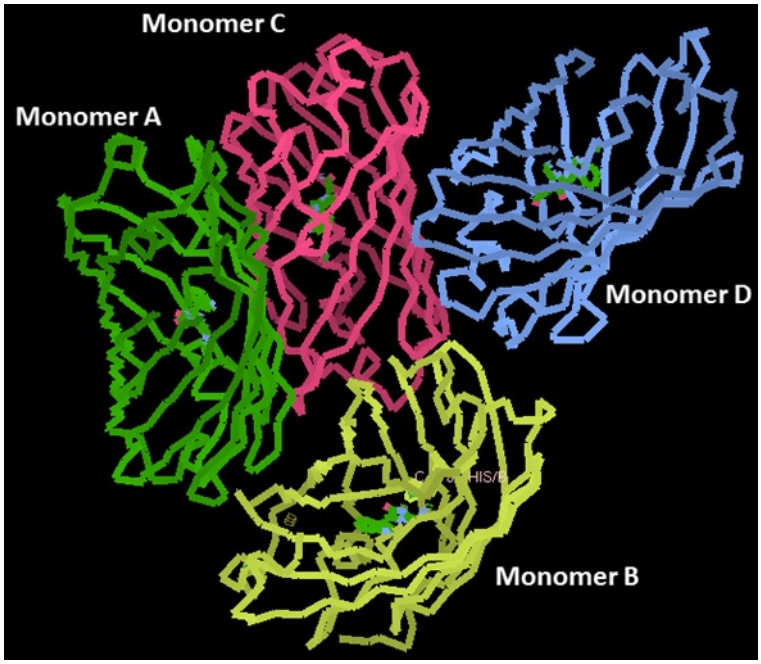


Figure 2.5 Content of the asymmetric unit of the NeonCyan0.95 crystal lattice

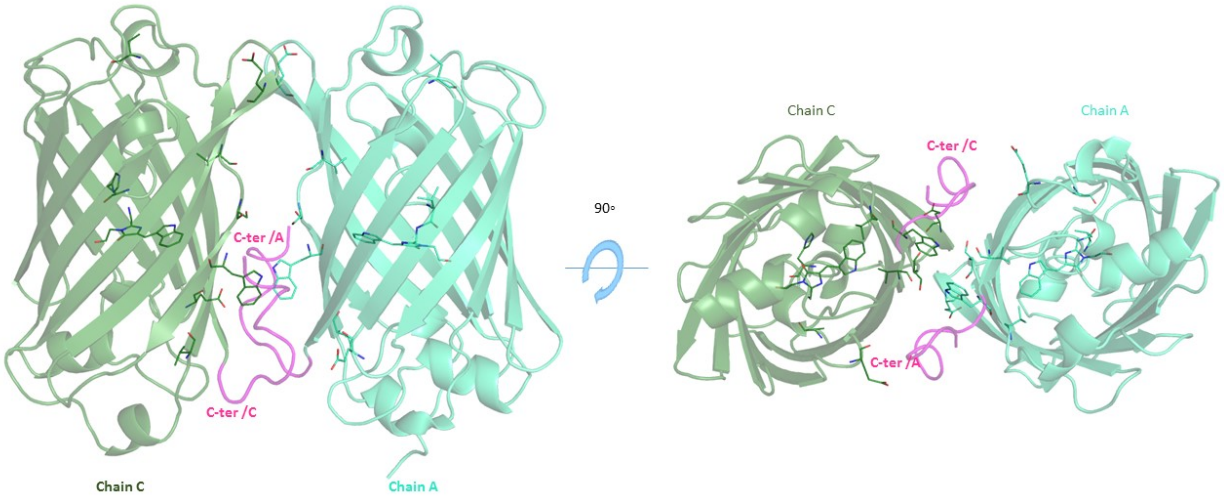


Figure 2.6 Putative physiological dimer.

The chromophore is essentially stabilized by two hydrogen bonds: one between one nitrogen atom of the imidazolinone ring and residue Glu220 (which is crucial for chromophore maturation) and one between one nitrogen atom of the indole ring and the

first water molecule of a three water molecule chain (Figure 2.7). One shall note that Thr207 is hydrogen-bonded to the second water molecule of this chain.

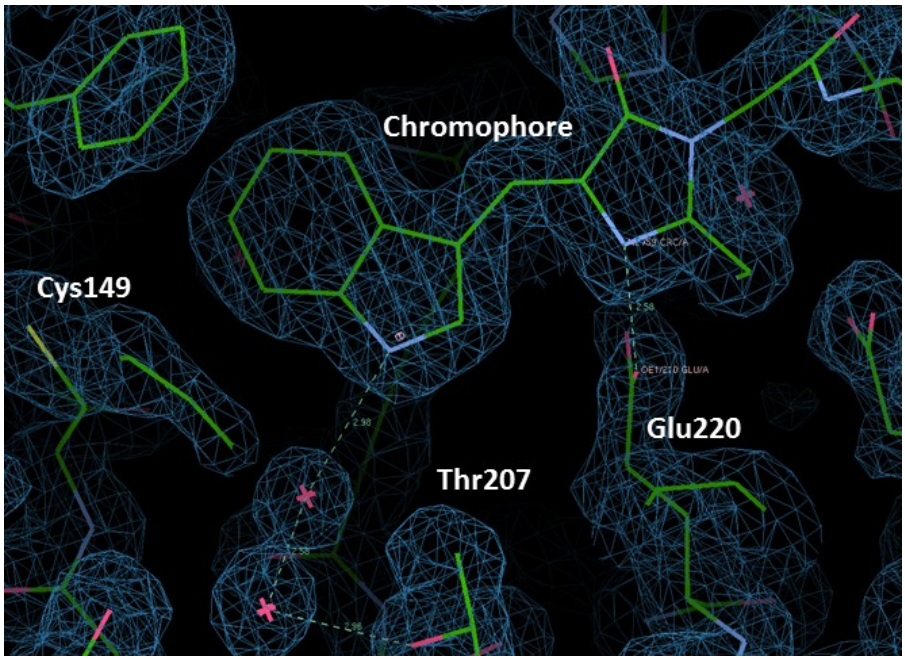


Figure 2.7 Chromophore stabilization.

Electron density superimposed on the protein structure at the chromophore location (contour at a 1.5 σ level). Two hydrogen bonds between the chromophore and a water molecule and Glu220 are represented, respectively.

One side of the chromophore is packed with hydrophilic residues with side chains potentially bearing positive charges (His72, Lys 153, Arg205) and negative charges (Glu45 and Glu220), whose positioning may affect the spectroscopic properties of the chromophore (Figure 2.8).

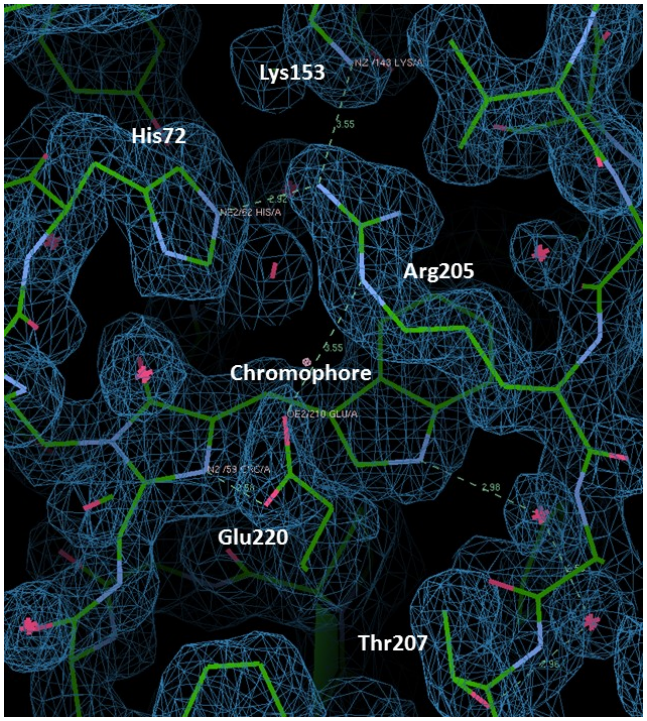


Figure 2.8 Charged residues positioned next to the chromophore.

Electron density superimposed on the protein structure at the chromophore location (contour at a 1.5 σ level). Charged residues in the vicinity are highlighted.

2.2.3 Notable aspects of the NeonCyan0.95 crystal structure

It is evident from the crystal structure that NeonCyan0.95 has a propensity to form a dimer at high concentrations. During the evolution of mNeonGreen from LanYFP the dimer interface was purposely disrupted with the Val140Arg mutation (175). Generally speaking, a dimeric structure that buries this hydrophobic interface is associated with increased FP stability and brightness. Accordingly, it follows that the directed evolution of NeonCyan would have a tendency to lead to the introduction of mutations that reverted the interface-disrupting Val150Arg mutation and would thereby allow the dimer structure to form. Indeed, 4 out of 9 the mutations introduced in going from mNeonGreen to NeonCyan1

are within this interface. The Arg150Trp mutation contributes to this dimeric character by restoring a hydrophobic pocket with Phe204 and Phe225. The localization of the mutations that accumulated during the evolution of NeonCyan1 highlights the importance of the dimer interface for stabilization of the new chromophore structure.

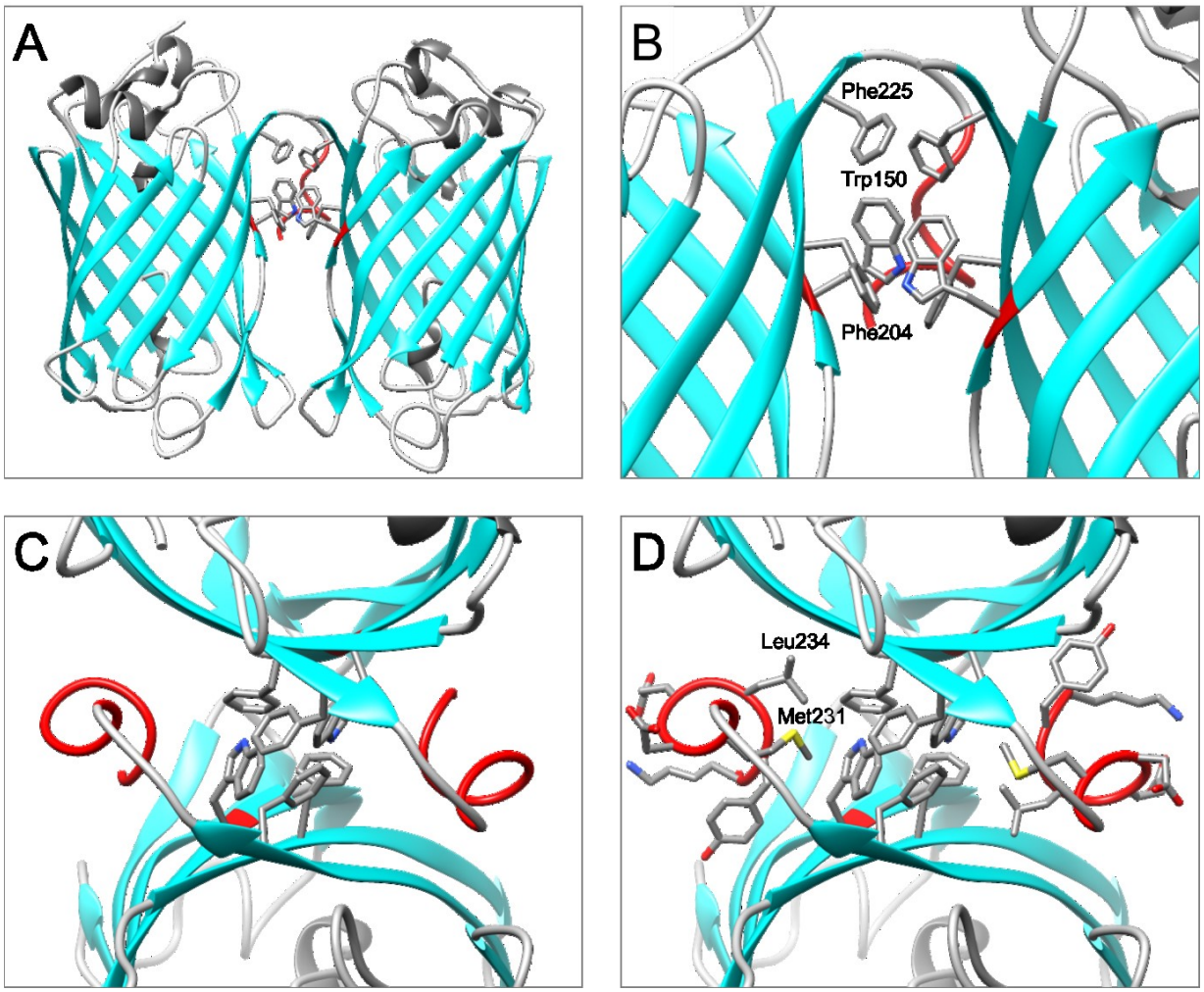


Figure 2.9 Hydrophobic pocket of NeonCyan0.95 dimer interface

The Arg150Trp mutation introduced in NeonCyan0.7 resulted in the formation of a hydrophobic pocket within the dimer interface, as revealed in the crystal structure. A) Hydrophobic pocket of the dimer interface of NeonCyan0.95. B) zoomed in view of pocket residues: Trp140, Phe204, Phe225, symmetrically interacting with the same 3 residues across the dimer interface. C) Top-view of pocket with C-terminal 'MDELYK' ribbon backbone colored red. D) Top-view of pocket with both C-terminal 'MDELYK' residues shown. Two hydrophobic residues of the structured C-terminal 'MDELYK', Met231 and Leu234, are pointed towards the hydrophobic pocket.

The mutation of a key dimer interface residue during the evolution of NeonCyan1, Arg150Trp, contributed to the formation of a hydrophobic pocket within the dimer interface (Figure 2.9). We speculate that this hydrophobic pocket is an important factor in stabilizing the dimeric quaternary structure observed in the crystal structure. To determine if the dimeric structure observed in the crystal structure was retained in the fully soluble state, we subjected NeonCyan1 to size-exclusion fast protein liquid chromatography (FPLC). The elution profile of NeonCyan1 is remarkable because it reveals that the protein does not behave as a true monomer, or a true dimer, compared to known standards. Rather, it elutes between the peaks for monomer and dimer standards, suggesting that it exists in a monomer-dimer equilibrium. Based on the X-ray crystal structure, we speculated that interactions of the structured C-terminal tail were contributing to the stability of the dimeric species. In the mNeonGreen crystal structure, the C-terminal tail was not observed, presumably due to it being unstructured and in different conformations in different copies of the protein in the crystal. To test whether the C-terminal “MDELYK” tail of NeonCyan1 was contributing to the dimeric character, we genetically truncated this sequence. Analysis of the resulting protein by size exclusion chromatography revealed a substantial shift in the FPLC trace towards a more monomeric character. This indicates that the hydrophobic pocket caused by Arg140Trp mutation works in concert with the C-terminal portion of NeonCyan1 to stabilize the dimeric quaternary structure of the protein *in vitro*.

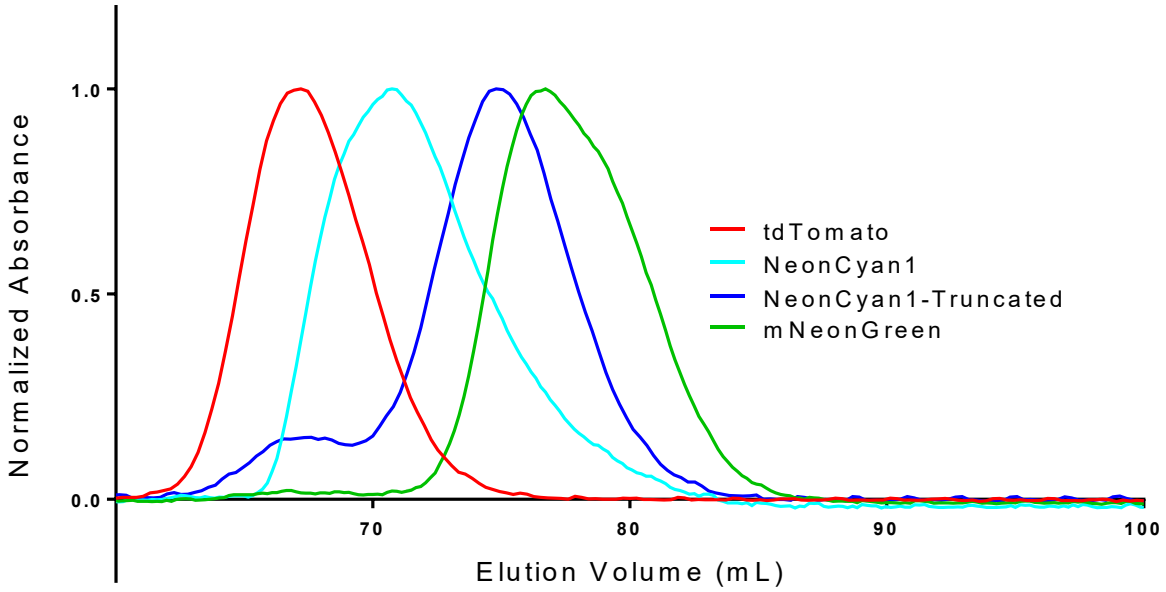


Figure 2.10 FPLC trace of NeonCyan1 and NeonCyan1-Truncated

Size exclusion chromatography of tdTomato (dimer standard), NeonCyan1 (dimer/monomer), NeonCyan1-Truncated (dimer/monomer), and mNeonGreen (monomer standard). Each sample was run individually under identical conditions with absorption detection of 555 nm, 430 nm, 430 nm, and 505 nm respectively. Removal of the structured C-terminal 'MDELYK' peptide sequence in NeonCyan1-Truncated induced a substantial shift from dimer to monomeric character.

We anticipated that the crystal structure of NeonCyan0.95 would facilitate further directed evolution and rational mutation to improve its biophysical properties. A notable insight drawn from the structure was the presence of 3 structured water molecules that form a hydrogen bond network that extends from the indole moiety of the chromophore to the hydroxyl groups of Thr207 (on the interior of the β-barrel) and Ser148 (on the

outside of the β-barrel) (Figure 2.11). We anticipated that site-directed mutagenesis of Thr207 of NeonCyan1 could potentially create new interactions with the chromophore that may serve to stabilize it in a more highly planar conformation and thereby increase the quantum yield (QY). We speculated that a suitably bulky residue (e.g., leucine, asparagine, glutamine, cystine, methionine), or a residue capable of directly forming a hydrogen bond to the indole nitrogen of the chromophore (e.g., aspartate, glutamate, asparagine, or glutamine), would be effective in displacing the water molecule(s) and forming a new interaction with the indole moiety of the chromophore. These new interactions could lead to improved or otherwise altered spectral properties. To test this idea, we screened a library of NeonCyan1 variants in which position Thr207 was randomized to all other possible amino acids. From this library we identified several new variants with spectral shifts and increased brightness (Table 2.1).

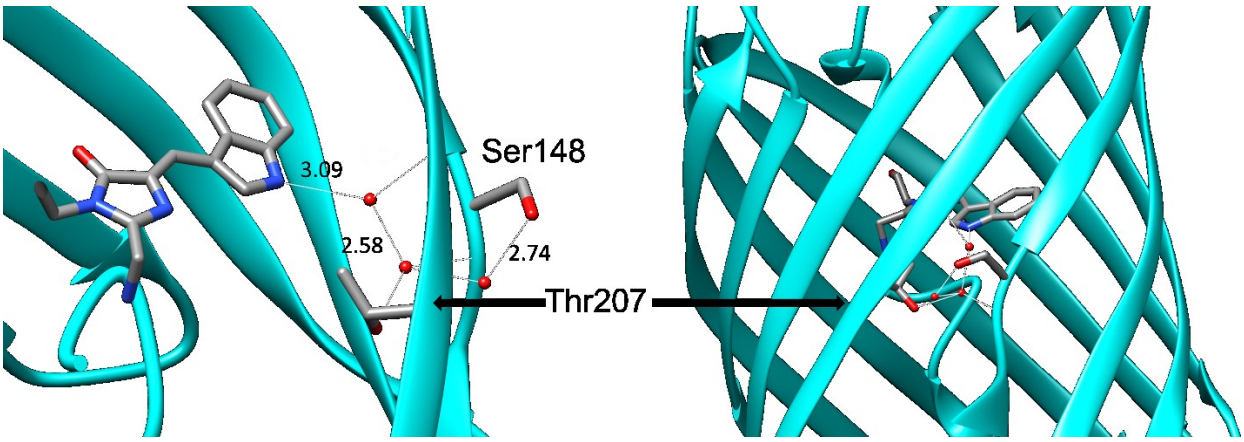


Figure 2.11 Hydrogen bond network of NeonCyan0.95 chromophore with Thr207

The crystal structure of NeonCyan0.95 revealed that the residue Thr207 was a candidate for site-directed mutagenesis. The indole moiety of the chromophore, hydroxyl side chain of Ser148, and the hydroxyl side chain of Thr207 all participate in a hydrogen bond network that stabilizes 3 water molecules. Hydrogen bond lengths are labelled in Angstroms.

2.2.4 Spectral shift of NeonCyan1 variants

The X-ray crystal structure of NeonCyan0.95 revealed that the side chain of Thr207 was involved in a hydrogen bond network through 3 structured water molecules. Site-directed mutagenesis of Thr207 was performed to investigate the possible stabilization of the indole moiety of the NeonCyan1 chromophore. Screening of a library in which Thr207 was mutated to all 19 other amino acids led to the identification of two variants with drastically different spectral properties relative to NeonCyan1: Thr207Met (NeonCyan1-Thr207Met) with blue shifted excitation and emission, and Thr207Asp (NeonCyan1-Thr207Asp) with red shifted excitation and emission (Figure 2.12). The Thr207Asp mutation, as evident from the spectral profile, led to formation of the anionic form of the

indole nitrogen due to stabilization by the carboxylic acid side chain (185, 186). It is likely that the Thr207Asp mutation induces a second negative charge resulting in the reorganization of Arg205 and Lys153, two positive side-chains inside the β-barrel and in close proximity to the indole nitrogen, which stabilizes the anionic form of the fluorophore. The Thr207Met mutation, however, led to a blue shift. This shift to higher energy absorbance and emission may indicate that the Thr207Met mutation has forced the chromophore into a *Z*,*E* conformation (187). This rotation would be possible through the displacement of the structured water molecules within the barrel by introduction of the relatively bulky methionine residue.

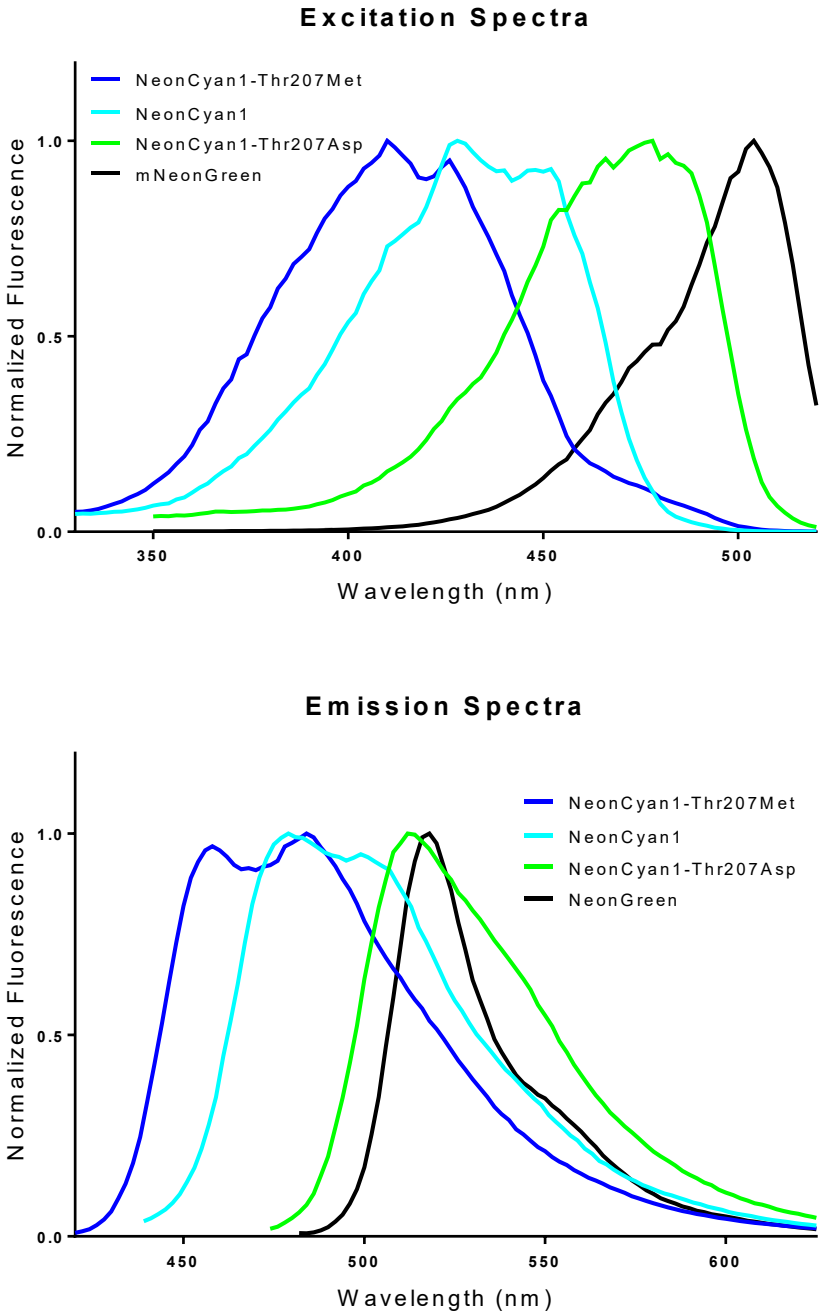


Figure 2.12 Excitation and emission of NeonCyan1 variants

Site directed mutagenesis of Thr207 resulted in the discovery of the NeonCyan1-Thr207Met with a blue shifted excitation and emission profile, and NeonCyan1-Thr207Asp with a red shifted excitation and emission spectra.

Table 2.2 Biophysical properties of NeonCyan1 variants

Protein	λ_{Exc} (nm)	λ_{Em} (nm)	ε^a	Φ	Brightness[b]	pK_a
NeonCyan1	430	480	44	0.29	13	6.4
NeonCyan1-Thr207Met	410/424	460/485	44	0.18	8	6.5
NeonCyan1-Thr207Asp	478	504	65	0.27	18	5.4/8.0
NeonCyan1-Truncated	430	480	45	0.27	12	6.7
mNeonGreen[c]	506	518	113	0.80	90	5.4

[a]Units of $mM^{-1}cm^{-1}$

[b]Product of ε in $mM^{-1}cm^{-1}$ and Φ

[c]From Ref. (77)

2.2.5 Biophysical characterization of NeonCyan1 variants

NeonCyan1, NeonCyan1-Thr207Met, and NeonCyan1-Thr207Asp were characterized in terms of their fundamental spectral and biophysical properties (Table 2.1). Relative to NeonCyan1, NeonCyan1-Thr207Met has an ~30% lower quantum yield (0.18 versus 0.29), but its pK_a and extinction coefficient remain essentially unchanged. Relative to NeonCyan1, NeonCyan1-Thr207Asp has a substantially increased extinction coefficient (65 $mM^{-1}cm^{-1}$ versus 44 $mM^{-1}cm^{-1}$) and a small decrease in QY (0.27 versus 0.29). Interestingly, while NeonCyan1 and NeonCyan1-Thr207Met both exhibit a sigmoidal dependency on pH that is characterized with a single pK_a, NeonCyan1-Thr2017Asp's pH dependency is best fit as a biphasic curve with pK_a values of 5.4 and 8.0 (Figure 2.13). based on this *in vitro* pH dependency, the fluorescence of NeonCyan1 and NeonCyan1-Thr207Met are sensitive to pH changes within the physiologically relevant range roughly

pH 5-8. In contrast, NeonCyan1-Thr207Asp is substantially less pH sensitive in this range. However, NeonCyan1-Thr207Asp has only approximately half of its maximum brightness (i.e., half the brightness at pH 9) within the physiologically relevant pH range.

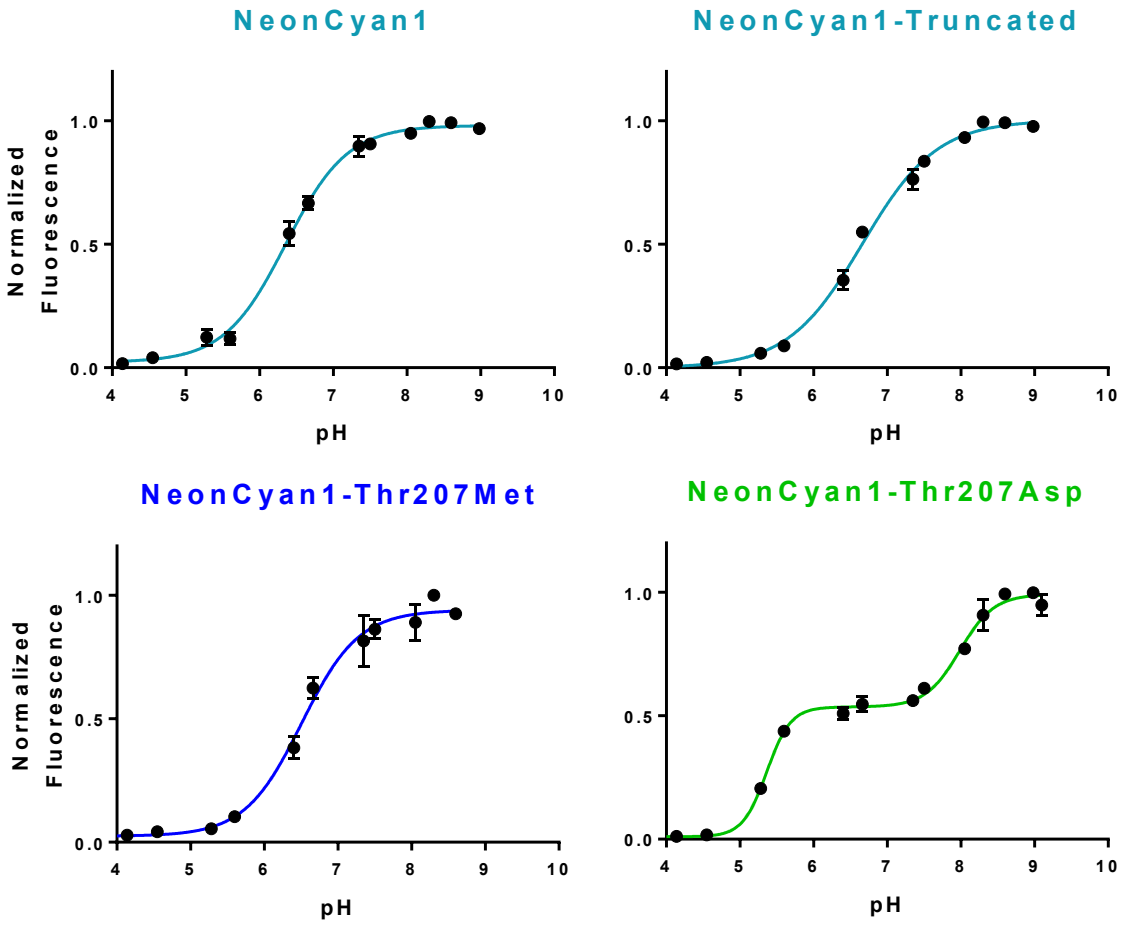


Figure 2.13 pH sensitivity of NeonCyan1 variants

Four variants of NeonCyan1 were tested for their fluorescent brightness in buffered solutions adjusted to a variety of different pH values. pK_as for these variants are listed in Table 2.1.

2.2.6 Imaging of F-actin labelled with NeonCyan1 variants

To assess the suitability of NeonCyan1 variants for live cell imaging we created a modified pcDNA vector with XhoI/HindIII restriction sites downstream of the DNA sequence coding for the F-Tractin peptide. The F-Tractin peptide is derived from the N-terminal portion of rat neuronal inositol-3-phosphate kinase A and has been demonstrated as an excellent filamentous-actin (F-actin) labelling peptide (188). We human codon optimized the 51 residues of the gene encoding peptide, added restriction sites for easy cloning, and DNA encoding a short flexible linker to accommodate the comparatively bulky FP chimeras. Into the Xho1/HindIII sites of this vector, we inserted the genes encoding NeonCyan1, NeonCyan1-Thr207Met, and NeonCyan1-Thr207Asp. We then transfected HeLa cells with our pcDNA-F-Tractin-NeonCyan1 variant constructs and imaged them using a wide-field fluorescence microscope with appropriate excitation and emission filters. As shown in Figure 2.14, our results demonstrate that F-Tractin-NeonCyan1 variant fusions localize to F-actin and illuminate the cytoskeleton with no substantial perturbation of natural F-actin structure (189).

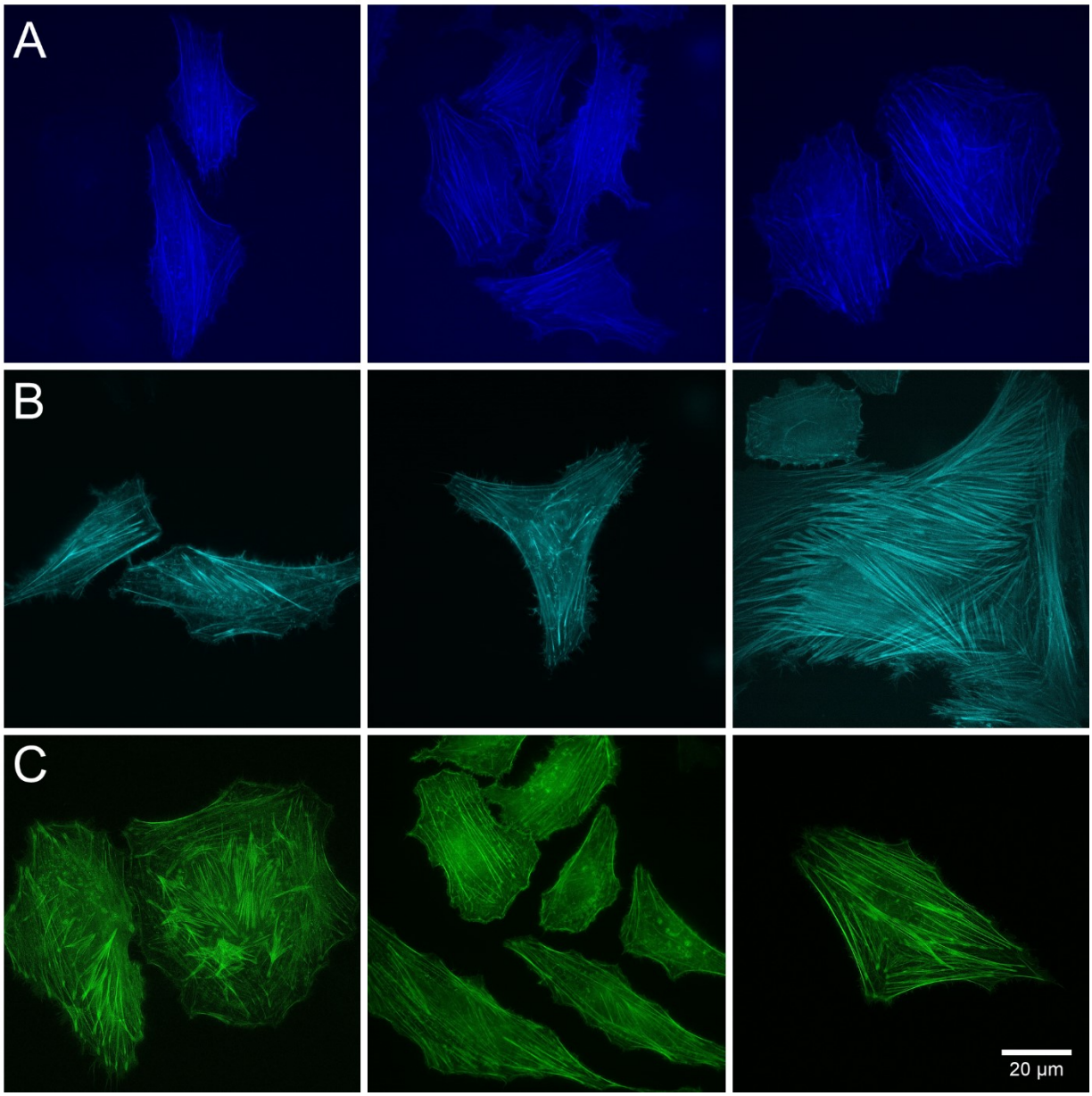


Figure 2.14 Imaging of actin labelled HeLa cells with NeonCyan1 variants

HeLa cells transfected with NeonCyan1 variants fused to the F-Tractin localization peptide. Each row (A-C) contains representative images of cells transfected with one of the NeonCyan1 variants. (A) NeonCyan1-Thr207Met with 100 ms exposure. (B) NeonCyan1 with 25 ms exposure. (C) NeonCyan-Thr207Asp with 25 ms exposure. All images are pseudo-colored.

2.3 Conclusion

In summary, we have engineered the first color-shifted tryptophan chromophore variants based on the mNeonGreen cephalochordate FP scaffold. These new FPs have interesting photophysical and biophysical characteristics as evident from the X-ray crystal structure, their spectral diversity, their oligomeric character, and their pH sensitivity. The NeonCyan1 variants express well in cultured HeLa cells and exhibit excellent F-actin labelling using the F-Tractin peptide. Further work to improve these new variants should focus on improving their monomeric character and increasing their brightness.

2.4 Experimental procedures

2.4.1 General

Synthetic DNA oligonucleotides and gBlocks were purchased from Integrated DNA Technologies (IDT). Plastic consumables, restriction endonucleases, *Taq* polymerase, Phusion polymerase, T4 DNA ligase, deoxynucleotides, DH10B *E. coli.*, pBAD/His B plasmid, pcDNA3.1(+) plasmid, Bacterial Protein Extraction Reagent (BPER), Penicillin-Streptomycin, Fetal Bovine Serum (FBS), TurboFect, and GeneJet gel or plasmid purification kits were purchased from Thermo Scientific. Agarose, $MnCl_2 \cdot 4H_2O$, D-glucose, ampicillin, L-arabinose, Hank's balanced salt solution (HBSS), DMEM, TryplE Express, and LB Lennox media were purchased from Fisher Scientific. Nickel NTA immobilized metal affinity chromatography protein purification beads were purchased from G-BioSciences. Ethidium bromide and PCR machines (T100 Thermal Cycler) were purchased from BioRad. Gibson Assembly reagent was purchased from New England Biolabs. QuikChange mutagenesis kits were purchased from Agilent Technologies. 96-

well Nunc MicroWell 96-Well Optical-Bottom Plates (cat. 265301) were purchased from Thermo Scientific. Molecular weight cut off filters were purchased from Millipore-Sigma. Sequencing was completed by the Molecular Biology Services Unit at the University of Alberta. mNeonGreen gene was a kind gift from Jiwu Wang at Allele Biotech.

2.4.2 Directed evolution

Following ligation into pBAD/His B vector, the NeonCyan library was transformed into DH10B *E. coli* bacteria and plated onto 100 μg/mL ampicillin/1.5% agar plates with 0.02% L-arabinose and grown overnight (12-18 hours) at 37 °C. Images acquired using the fluorescent colony screening system were analyzed with Image Pro 6 software (Media Cybernetics) to identify the brightest fluorescent colonies which are then picked and grown in 4 mL of LB Lennox growth medium with 0.02% L-arabinose and 100 μg/mL ampicillin. Culturing overnight (approximately 12-18 hours), yielded sufficient bacterial mass to perform BPER protein extraction. The fluorescent protein containing lysate was then dispensed into 384 well plates and evaluated on a Tecan Safire^2 for brightness.

2.4.2 PCR and EP-PCR

All PCR reactions were performed according to the manufacture's recommendations. EP-PCR was performed using *Taq* polymerase with 0.1 mM $MnCl_2$ and a skewed dNTP ratio (0.2 mM ATP, 1 mM CTP, 0.2 mM GTP, 1 mM TTP) resulting in approximately 1 nucleotide mutation per 800 base pairs. Cycling conditions were 35 rounds of 95 °C for 20 seconds, 55 °C for 30 seconds, 72 °C for 60 seconds, and 5 minutes at 72 °C for final extension. PCRs utilizing Phusion polymerase were performed with the supplied buffer with the concentrations of DNA and oligos suggested by the manufacturer. Conditions for

PCR were 98 °C for 10 seconds, 63 °C for 20 seconds, 72 °C for 30 seconds, and 72 °C for 60 seconds for final extension.

2.4.3 Purification of PCR products and plasmid DNA

PCR products were run on 1% agarose gel in TAE buffer, excised after viewing on a UV transilluminator, and recovered with the Thermo Scientific gel preparation kit according to the manufacturer's recommendations. Plasmid DNA was recovered from 4 mL of overnight *E. coli* culture using Thermo Scientific plasmid mini prep kit according to the manufacturer's recommendations.

2.4.4 Construction of F-Tractin-NeonCyan1 variant pcDNA vectors

Construction of vectors for mammalian expression that label endogenous filamentous actin chains were completed in 4 steps. First, a modified pcDNA plasmid with Xho1/HindIII restriction sites was digested to be used as vector. Second, PCR amplified NeonCyan1 variants were digested with Xho1 and HindIII and ligated into the digested pcDNA to create pcDNA-NeonCyan1. Third, pcDNA-NeonCyan1 variants were amplified with primers 'pcDNA MVSK Vec Fwd' and 'pcDNA FT GA Rv' resulting in a linear product suitable for Gibson assembly, this required an increased extension time of 5 minutes. Finally, a human codon optimized gBlock fragment from IDT (F-Tractin gBlock) was combined with the linear vector from step 3 using Gibson Assembly. The resulting product was transformed, picked, and sequenced to identify proper insertions with no frameshifts or mutations. It should be noted that this method, although potentially seamless, retains the original ORF and restriction sites as pBAD vector thus allowing for quick transfer of candidate variants to a mammalian expression plasmid with filamentous actin localization.

2.4.5 Protein purification and characterization

To purify NeonCyan variants for spectroscopic and physical characterization, the pBAD plasmid containing the desired variant was transformed into electrocompetent DH10B *E. coli* and plated on LB/ampicillin (400 µg/mL) agar plates. Following overnight incubation at 37 °C, a single colony was picked and cultured in 500 mL of LB liquid media with ampicillin (200 µg/mL) and 0.02% L-arabinose. Bacterial cultures were grown for 30 hours at 37 °C and 250 rpm shaking followed by 24 hours at 30 °C. The cells were harvested by cooling on ice to 4 °C then centrifugation (6,000 × g for 6 minutes), re-suspended in 30 mL of ice-cold TBS buffer pH 7.2, and lysed by sonication (QSonica Q700, amplitude 50, 1 second on, 2 seconds off, 2 minutes sonication time). All subsequent purification procedures were performed on ice. The resulting lysate was clarified of cell debris by centrifugation for 1 hour at 21,000 × g, filtered through a Kim-wipe into a 50 mL conical tube, and incubated for 3 hours with Ni-NTA resin. Resin containing NeonCyan variant protein was washed with 100 mL of 20 mM imidazole TBS wash buffer and eluted with 250 mM imidazole TBS elution buffer. Purified protein was buffer exchanged into TBS using a 10,000 Da molecular weight cut-off filter (Millipore-Sigma) through 3 successive washes. Absorption spectra were recorded on a Beckman-Coulter DU-800 UV-visible spectrophotometer and fluorescence spectra recorded on a Tecan Safire^2 plate reader.

Determination of the extinction coefficient of NeonCyan variants was achieved using the alkaline denaturation method (77) with mCerulean3 as a standard. Briefly, the concentration of protein was adjusted by dilution in TBS to reach an absorbance of 0.6 to 1.0. A dilution series was then prepared with absorbances of 0.01, 0.02, 0.03, 0.04, and 0.05 for NeonCyan variants and mCerulean3. Integration of the fluorescent peaks

provides a total fluorescent emission value which was plotted against the absorbance to provide a slope. The quantum yields of NeonCyan variants were determined using the published (77) QY value in a ratiometric manner:

$$\Phi_{protein} = \Phi_{standard} \times (S_{protein}/S_{standard})$$

Extinction coefficients were determined by measuring the absorption spectrum in TBS buffer and 2 M NaOH. Absorbance of the denatured fluorescent protein at 460 nm was divided by the previously determined extinction coefficient of 46,000 $M^{-1}cm^{-1}$ to give the concentration of protein (186). The extinction coefficient is then determined by dividing the TBS sample absorbance maximum by the protein concentration.

2.4.6 Determination of pK_a

The pK_as of NeonCyan variants were determined using a buffer series created following Carmody's method (190) and the protocol of Cranfill *et al* (77). Briefly, 2 µg of purified protein was added to 100 µL of each respective pH buffer in a black walled 96 well microplate. Each variant was measured in triplicate and background corrected before normalization of the data. The pK_a value was determined with GraphPad Prism 7 software using a four-parameter variable-slope fit.

2.4.7 Expression in HeLa cells for *in vitro* imaging

HeLa cells cultured in DMEM with 10% fetal bovine serum supplemented with penicllin-G potassium salt (50 units/mL) and streptomycin sulphate (50 µg/mL) were plated on collagen coated 35 mm glass bottom dishes (Matsumami) and transfected with 1 µg of plasmid DNA using 2 µL of TurboFect when cells reached roughly 60% confluency. After overnight incubation at 37 °C with 5% CO_2, adherent cells were washed twice with warm Hank's balanced salt solution immediately before imaging.

Cell imaging was performed on a Zeiss 200M wide-field microscope with Semrock filters loaded into filter cubes specific for each NeonCyan1 variant:

Table 2.3 Filters for NeonCyan1 imaging

Variant	Excitation filter (nm)	Emission filter (nm)
NeonCyan1-Thr207Met	387/11	460/40
NeonCyan1	436/20	483/32
NeonCyan1-Thr207Asp	470/40	525/50

The Zeiss 200M was fitted with an OrcaFlash 4.0 – C13440 (Hamamatsu) camera and images were acquired using MetaMorph 7.8.0.0 software and an MS-2000 automated stage (Applied Scientific Instrumentation). Images were acquired at room temperature through a 60× oiled objective lens (N.A. 1.4). Images were analyzed and pseudo-colored with ImageJ2 (191, 192). UCSF Chimera was used for modelling and representations of NeonCyan0.95 (124).

Table 2.4 Oligos and gBlocks used in this work

Primer Name	Sequence (5’ to 3’)
MVSK Xho1 Fwd	TATATCTCGAGGATGGTGAGCAAGGGCGAGGAG
MDELYK H3 Rv	ATATAAAGCTTTTACTTGTACAGCTCGTCCATGCC
NC XWG QC	CCCTGGATTCTGGTCCCTCATGTCNNKTGGGGCTTCCATCAGTACCTG
NC XFG QC	CCCTGGATTCTGGTCCCTCATGTCNNKTTTGGCTTCCATCAGTACCTG
NC XHG QC	CCCTGGATTCTGGTCCCTCATGTCNNKCATGGCTTCCATCAGTACCTG
NC XXWG QC	CCCCTGGATTCTGGTCCCTCATNNKNNKGGGTGGGGCTTCCATCAGTA CCTGCCC
NC 0.7 E151X QC	GCTGTGGACTCGTCCTGGNNKAAGAAGACTTACCCCAACGACAAAACC

NC 0.8 S149X_S151X QC	CTGACCGCTGTGGACTCGNNKTGGNNKAAGAAGACTTACCCCAACGACAAAAC
NC Arg150X QC	CCGCTGTGGACTCGTGCNNKAGTAAGAAGACTTACCCCAACGACAAAACC
pcDNA MVSK Vec Fwd	ATGGTGAGCAAGGGCTCGG
pcDNA FT GA Rv	GGTGGCGGCACGCTAGCC
F-Tractin gBlock	GCTAGCGTGCCGCCACCATGGCTAGACCCCGAGGTGCCGGCCCTTGTAGTCCTGGTCTGGAGCGCGCTCCTCGACGCTCAGTAGGTGAACTGAGGTTGTTGTTCGAGGCCAGATGTGCTGCTGTGGCCGCTGCAGCAGCTGCCGGAGGGTTGGCTCTGCCGGTAGCCACCGGCGGGAGCGGGGGCAGCTCGAGGATGGTGAGCAAGGGCTCGG
NC Thr207X QC	GCCGGTGTACGTGTTCCGTAAGNNKGAGCTCAAGCACTCCGAGACC

2.4.8 Expression and purification of NeonCyan0.95 for crystallization purposes

The overexpression and purification of NeonCyan was performed according to a previously described protocol (193). The pBAD-His B plasmid of NeonCyan was dissolved in distilled water to 100 µg/mL. Bacterial transformation was performed with competent *E. coli* DH10B cells (MAX Efficiency® DH10B, ThermoFischer) and plated on LB/ampicillin (400 µg/mL) agar plates. Cells grew in 1 L of LB medium at 37 °C until OD_{600} reached 1.25. Protein expression was then induced overnight at 17 °C with the addition of 10 mL of L-arabinose 20% w/v. Cells were centrifuged at 4,000 × g during 20 min (Avanti J-26 XP, Beckman Coulter) and pellets were resuspended, prior to flash-freezing at -80 °C, with 20 mL of lysis buffer made of 50 mM Tris pH 8.0, 300 mM NaCl, 10 mM imidazole, 0.25 mg/mL lysozyme, 400 µg/mL DNAse I and anti-protease cocktail (Complete EDTA-free, Roche). Sonication was carried out on an ice bath with 5 cycles of

10 s at 50 % maximum power. Cell debris underwent centrifugation at 17,500 × g during 30 min at 4 °C. The tagged protein was separated from the clarified lysate using a nickel affinity column (HisTrap QFF 5 mL, GE Healthcare) followed by a size exclusion chromatography (Superdex 75 10/300 GL, GE Healthcare). Final concentration of the protein was 20 mg/mL leading to a production yield of 15 mg/L of medium. Prior to crystallization, the purified proteins were submitted to tryptic digestion for 1 h at room temperature at a 1:10 trypsin:protein ratio.

Chapter 3: Development mNeonGreen-GECO

3.1 Introduction

Genetically encodable Ca^{2+} indicators (GECIs) have become powerful tools for the investigation of Ca^{2+} dynamics both *in vitro* and *in vivo* (12, 194-197). The popularity of these sensors is derived from their ability to mitigate numerous concerns that arise from the inherent nature of optical recording in live organisms. Because they are fully genetically encodable, they can be targeted for expression only in specific cell types, or subcellularly localized to specific organelles. Their low cellular toxicity, minimal perturbation of endogenous cellular functions, and biological turnover make them ideal for long-term imaging experiments (57). The Ca^{2+}-dependent fluorescent response of GECI's is routinely used as a proxy for neural activity due to the transient changes in Ca^{2+} concentration that accompany action potentials and subthreshold membrane depolarization (198-201). Recently, GECIs have facilitated the optical recording of thousands of neurons simultaneously in the surgically exposed brains of mice (33). Despite their wide acceptance by the scientific community, there are some properties of GECIs that could be further improved. This property includes the kinetics of the response, their fluorescent brightness, and their role in Ca^{2+} buffering. Some GECIS under aggregation in neurons, and recently, some of the most highly optimized GECIs have been demonstrated to cause aberrant cortical activity in murine models (202-204).

The most prominent members of the GECI family are the GCaMP-style sensors. GCaMP sensors are a three-part fusion of circularly permuted fluorescent protein (FP), calmodulin (CaM), and the RS20 peptide from avian smooth muscle myosin light chain kinase (205, 206). Since the first report of green fluorescent G-CaMP (as the name was

originally written) by Junichi Nakai in 2001 (59), this lineage of indicators has been the focus of extensive engineering efforts by multiple research groups, most notably the Howard Hughes Medical Institute funded GENIE project at Janelia Research Campus. The modest Ca^{2+} 4.3-fold increase in fluorescent intensity for the original G-CaMP has been improved through many generations to the 50-fold *in vitro* response of GCaMP6 series of sensors (57). In addition, extensive optimization of kinetics, brightness, and Ca^{2+} affinity are all major improvements that have facilitated the acceptance of GCaMPs as the standard GECI for optical reporting of neuronal activity.

Expansion of the range of colors of GCaMP type sensors was has been achieved by changing the fluorescent protein (to both blue-shifted and red-shifted variants) and retaining the CaM/RS20 sensing domain (207). Red shifted GECIs have a number of advantages, including the fact that tissue is more transparent to more red-shifted wavelengths of light. That is, red light is less absorbed by heme (as in hemoglobin) and other biomolecules (27). Accordingly, there is great interest in using red-shifted reporters that allow for deeper optical recording. Red GECI's are also compatible with blue light excited Channelrhodopsins (208), allowing for the simultaneous optical stimulation and monitoring of neural activity (209-211). However, some red fluorescent protein-based GCaMP class sensors accumulate in punctate structures when expressed in neurons, which hinders their wide acceptance (212). On the other end of the visible spectrum, development of blue shifted GECI's has facilitated multiplex imaging of subcellular localized Ca^{2+} dynamics (207). Overall, the spectral expansion of GCaMP sensors has broadened the research capabilities of GECI imaging, but no derivative has yet dethroned the highly evolved green GCaMP sensor as the best-in-class tool for neuronal imaging.

An important issue that is common to all GECIs is their intrinsic Ca^{2+} buffering capacity. The Ca^{2+} binding domains of GECIs act as Ca^{2+} buffers within the cell and must necessarily compete with endogenous proteins for binding to Ca^{2+} (Refs. (213, 214)). Comprehensive investigations of this phenomenon are limited, but a few reports have indicated abnormal morphology and behaviour of neurons after long term or high expression of GCaMPs, and Ca^{2+} buffering is a possible cause (215). One solution to the Ca^{2+} buffering phenomenon is to reduce the reporter protein expression, leading to a lower concentration of GECIs, and reduced buffering capacity. However, reduced expression requires increased intensity of excitation light to achieve an equivalent fluorescent signal, this can lead to increased phototoxicity and photobleaching. One solution to this problem is to develop GECIs with increased brightness such that they could be expressed at lower concentration while retaining a similar fluorescent intensity with similar intensity of excitation light.

GCaMP6s, the brightest GCaMP class sensor available at the time of writing, is brighter than the enhanced green fluorescent protein (EGFP) from which it is derived, but recent generations have plateaued in brightness (GCaMP6s +Ca^{2+} brightness = 39.6; EGFP brightness = 37.5) (57, 77). The GCaMP series has characteristic trade-offs with brightness, affinity, kinetics, and sensitivity. That is, the brightest sensors tend to have decreased sensitivity or slower kinetics (26, 216).

We propose that increasing the brightness of GCaMP class sensors would lead to better tools for optical imaging of neuronal activity and decrease the occurrence of biological artifacts resulting from Ca^{2+} buffering (215). This proposal is inspired, in part, by the advent of a bright monomeric fluorescent protein from Lancelet sp., mNeonGreen

(173) (mNG). Due to its high brightness, mNeonGreen is an excellent starting point from which to develop a brighter GCaMP. mNeonGreen is a fluorescent protein derived from a cephalochordate (lancelet). In contrast, GCaMP was developed from an Anthozoan (jellyfish) fluorescent protein. While mNeonGreen is a very bright fluorescent protein from a distinct and relatively unexplored evolutionary clade, it has also been demonstrated to be an excellent subcellular localization tag. To develop a series of Ca^{2+} sensors that are brighter than GCaMP6s with a new topology that should facilitate efficient subcellular localization, we undertook an effort to rationally engineer mNeonGreen to be a Ca^{2+} indicator. Here we introduce the mNeonGreen genetically encodable Ca^{2+} sensor for optical imaging (mNG-GECO1), a non-circularly permuted Ca^{2+} sensor that is brighter than GCaMP6s.

3.2 Results and Discussion

3.2.1 Rational engineering and iterative directed evolution of mNG-GECO1

We undertook a rational approach to the engineering of mNG-GECO1. We started from an unpublished variant of REX-GECO1 that was developed by a former Ph.D. student in the Campbell lab, Dr. Jiahui Wu (217). This was a topological variant of REX-GECO1 in which the original termini of the fluorescent protein scaffold had been reintroduced and a 5 amino acid linker (Gly_4Ser) had been introduced between CaM and RS20. Based on this sensor topology i.e.) insertion at residue 145, we used PCR to produce a linear fragment of the plasmid containing the mNG gene that was suitable for Gibson Assembly cloning (218). Next, we used PCR to amplify the DNA encoding 'CaM-GlyGlyGlyGlySer-RS20' from the unpublished REX-GECO1 plasmid and clone it into the linear fragment

containing the mNG gene to restore the circular plasmid (218, 219). We named this new Ca^{2+} sensor prototype non-circularly permuted mNG-GECO0.1 (Figure 3.1). This construct had minimal response, roughly 0.3-fold, to Ca^{2+} but we anticipated that optimization of the sequence around the insertion site would yield a suitable template for directed evolution. Indeed, we found that removal of Ala146 (mNG-GECO0.2) improved the sensitivity to a 2-fold increase in fluorescence upon binding to Ca^{2+}. We then began a process of iterative directed evolution to increase the brightness and sensitivity of mNG-GECO0.2 (Figure 3.2). In our primary screen, we used a fluorescent colony screening system to screen for colonies harbouring fluorescent mNG-GECO variants. These colonies were picked and cultured overnight in liquid media. A secondary screen for Ca^{2+} sensitivity was performed the next day from bacterial lysate containing variants of mNG-GECO protein. The genes encoding promising candidate variants are sent for Sanger sequencing and used as templates for the next round of directed evolution (179).

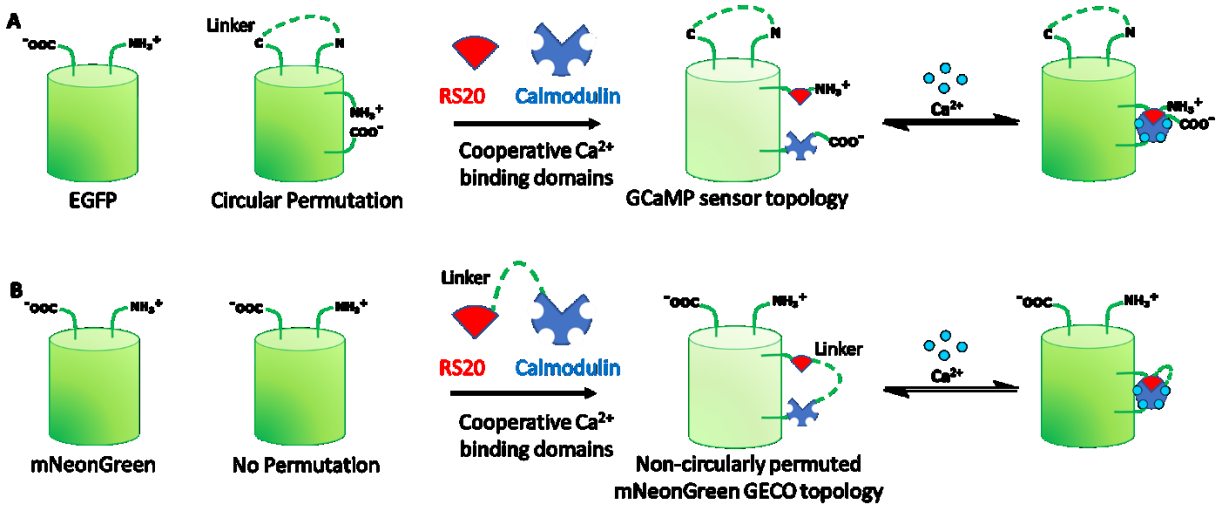


Figure 3.1 Topology of mNG-GECO1

The topology of GCaMP and mNG-GECO1 Ca^{2+} sensors. In the presence of Ca^{2+}, the RS20 and CaM domains interact and modulate chromophore environment resulting in increased fluorescent intensity. (A) Circular permutation of enhanced green fluorescent protein (EGFP) with a linker between the original termini and fusion of the RS20 peptide to the new N-terminus and CaM to the new C-terminus results in a GCaMP topology. (B) Design of mNG-GECO1 by retaining the original termini and insertion of RS20 and CaM fused together with a short peptide linker.

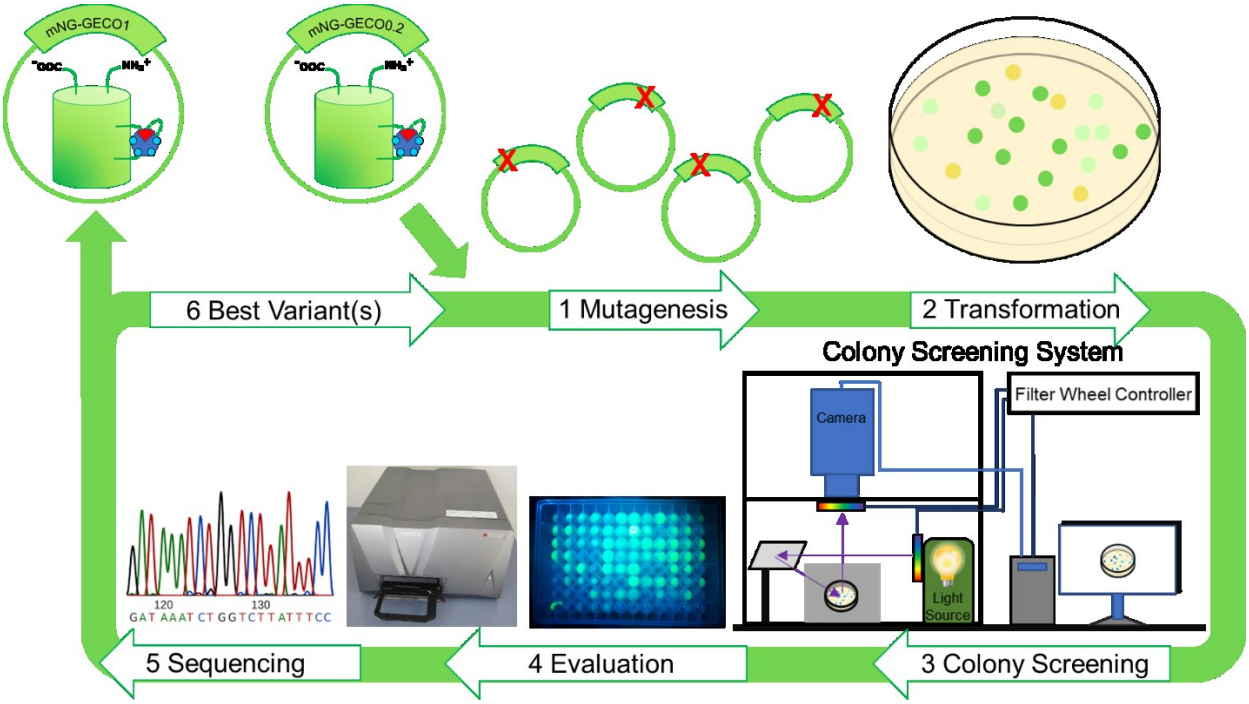


Figure 3.2 Iterative directed evolution screening of mNG-GECO1

Directed evolution of mNG-GECO1 was performed in an iterative fashion. A non-circularly permuted variant was subjected to 7 rounds of iterative directed evolution. Libraries constructed with error-prone PCR or site-directed mutagenesis were used to transform electrocompetent *E. coli* which was then plated on agar plates. A colony screening system was used to image the plates and identify fluorescent colonies harbouring mNG-GECO variants suitable for further investigation. Candidate variants selected for brightness and sensitivity were sent for sequencing and used as the template for the subsequent round of directed evolution.

After seven rounds of iterative directed evolution, *E. coli* colonies harbouring mNG-GECO0.7 were brightly fluorescent after overnight incubation. However, the Ca^{2+} response of mNG-GECO0.7 was low, roughly 5-fold, in comparison to other GCaMP class

sensors (generally greater than 20-fold). We anticipated that optimization of the linkers using site-directed mutagenesis to develop libraries of random mutations within these regions of the sensor would improve the response. Indeed, we identified two mutations of the linker region between mNeonGreen barrel and CaM: Ala145Gly and Leu143Ile. This variant, mNG-GECO0.9, had a 12-fold response to Ca^{2+} *in vitro*. Further investigation of the linker region between RS20 and CaM yielded a particularly bright variant with the Cys325Asn mutation (Figure 3.3). This variant, mNG-GECO1, is brighter than mNG-GECO0.9 but has a decreased response to Ca^{2+} of 3.5-fold (Table 3.1). mNG-GECO0.9 and mNG-GECO1 were selected for further characterization in cultured cells.

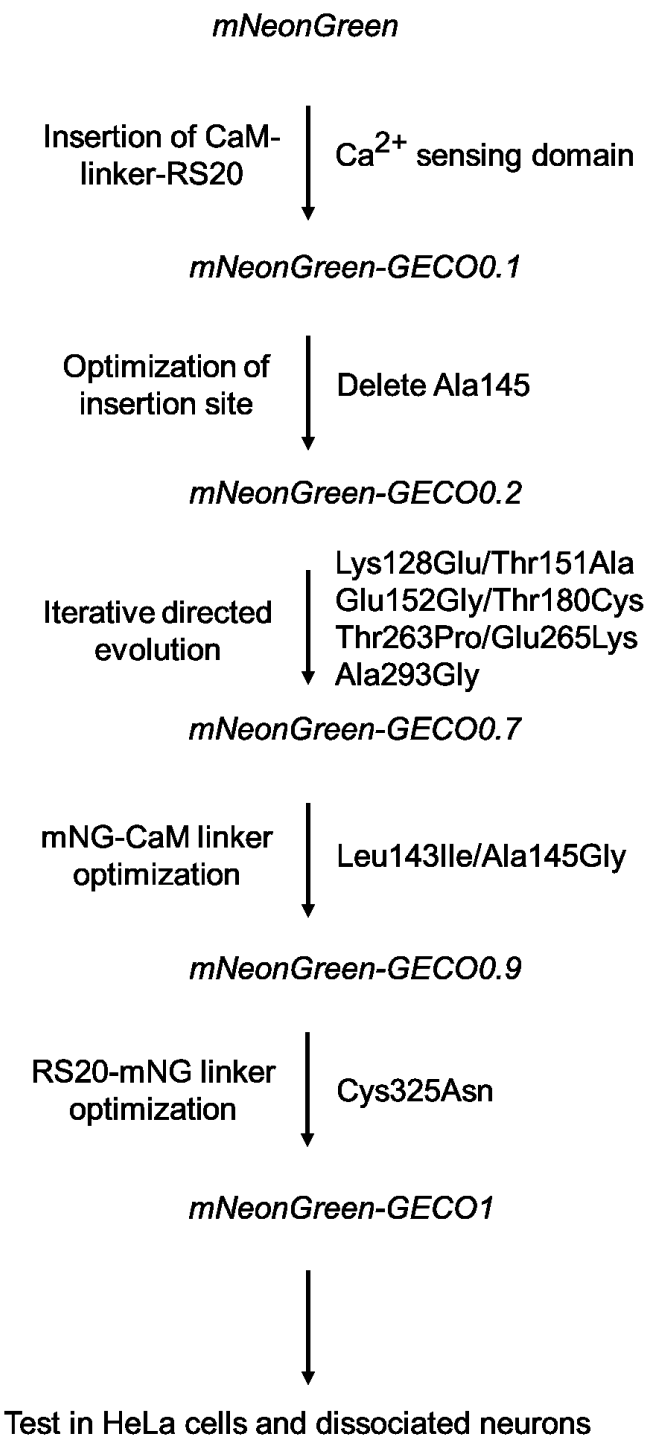


Figure 3.3 Lineage of mNG-GECO1

Lineage of mNG-GECO variants from initial insertion of Ca^{2+} sensing domain to mNG-GECO1 for testing in mammalian cells.

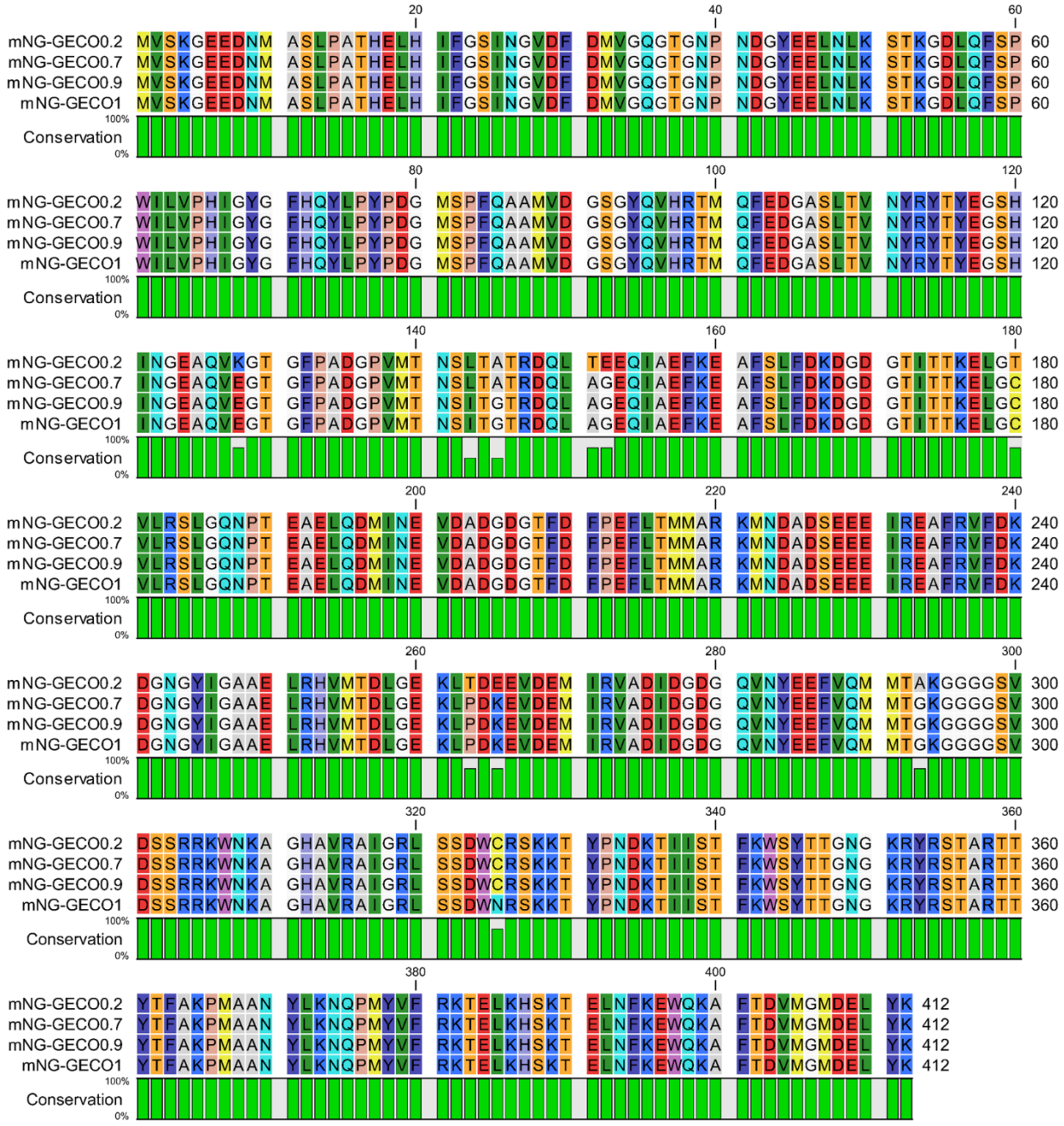


Figure 3.4 Alignment of mNG-GECO variants

Alignment of mNG-GECO variants 0.2, 0.7, 0.9, and 1. A green bar chart representing the percentage of the most frequently observed residue is displayed below the alignment.

3.2.2 *In vitro* characterization of mNG-GECO

We characterized mNG-GECO0.9 and mNG-GECO1 alongside GCaMP6s (as a reference) for direct comparison of biophysical properties (Table 3.1). We found that the excitation and emission maximum for Ca^{2+} saturated GCaMP6s, mNG-GECO0.9, and mNG-GECO1 to be 497/512, 497/513, and 509/517 nm respectively (Figure 3.5). There is a shift in excitation and emission spectra from mNG-GECO0.9 and mNG-GECO1 resulting from the Cys325Asn mutation, the mechanism of this spectral shift has not been determined.

mNG-GECO0.9 shows a relatively low fluorescent brightness in the Ca^{2+} bound state, making it is approximately half as bright as GCaMP6s. The extinction coefficient of 57,000 $M^{-1}cm^{-1}$ and quantum yield of 0.78 are both lower than that of GCaMP6s. However, the *in vitro* Ca^{2+}-dependent fluorescent change of mNG-GECO0.9 is about 4 times higher than mNG-GECO1. mNG-GECO0.9 has a K_d of 124 nM, which is similar to GCaMP6s' K_d of 168 nM. GCaMP6s' K_d has been highly optimized for neuronal activity imaging, indicating that mNG-GECO0.9 may also have a K_d that is well suited for imaging of neuronal activity (Figure 3.6). The Ca^{2+} dissociation constant (K_{off}) ($t_{1/2}$ = 0.87 s^{-1}) is lower than that of GCaMP6s ($t_{1/2}$ = 1.34 s^{-1}).

mNG-GECO1 has a very high fluorescent brightness in the Ca^{2+} bound state. Its extinction coefficient of 97,000 $M^{-1}cm^{-1}$ and quantum yield of 0.93 make mNG-GECO1 the brightest green GECI available today. Although not quite as bright as mNeonGreen, the combined effect of a decreased extinction coefficient and increased quantum yield make mNG-GECO1 98% as bright as its fluorescent protein scaffold. The Ca^{2+}

dissociation constant for mNG-GECO1 was measured to be 69 nM, and the $t_{1/2}$ for Ca^{2+} dissociation is 1.90 s^{-1}.

With two promising variants having interesting properties, we sought to further evaluate the Ca^{2+} dynamics reporting capability of mNG-GECO0.9 and mNG-GECO1 in HeLa cells and dissociated rat cortical neurons.

Table 3.1 Biophysical characterization of Ca^{2+} saturated mNG-GECO variants

Sensor	Excitation Max (nm) (Ca^{2+} sat.)	Emission Max (nm) (Ca^{2+} sat.)	Quantum Yield (Ca^{2+} sat.)	Extinction Coefficient ($mM^{-1}cm^{-1}$; Ca^{2+} sat.)	Brightness[a]	Brightness Relative to GCaMP6s	Brightness Relative to mNG	ΔF/F	K_d (nM)	k_{off} (s^{-1})
mNeonGreen[b]	506	517	0.8	115,000	92	143%	100%	NA	NA	NA
mNG-GECO0.9	497	513	0.78	57,000	36	56%	39%	12	124 +/- 15	0.87 +/- 0.01
mNG-GECO1	509	517	0.93	97,000	90	140%	98%	3.5	69 +/- 15	1.90 +/- 0.02
GCaMP6s	497	512	0.84	76,000	64	100%	70%	39	168 +/- 14	1.34 +/- 0.02

[a]Product of ε in $mM^{-1}cm^{-1}$ and Φ (no units).

[b]From Ref. 30

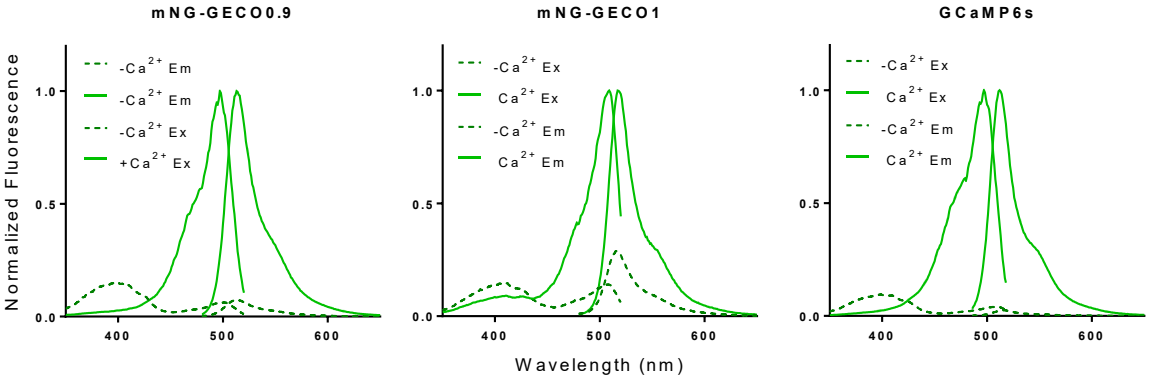


Figure 3.5 Excitation and emission spectra of mNG-GECO and GCaMP6s

Normalized excitation and emission spectra for mNG-GECO0.9, mNG-GECO1, and GCaMP6s. The Cys325Asn mutation in mNG-GECO1 increases the fluorescent brightness of the Ca^{2+}-free state which reduces the Ca^{2+}-dependent fluorescent response relative to mNG-GECO0.9.

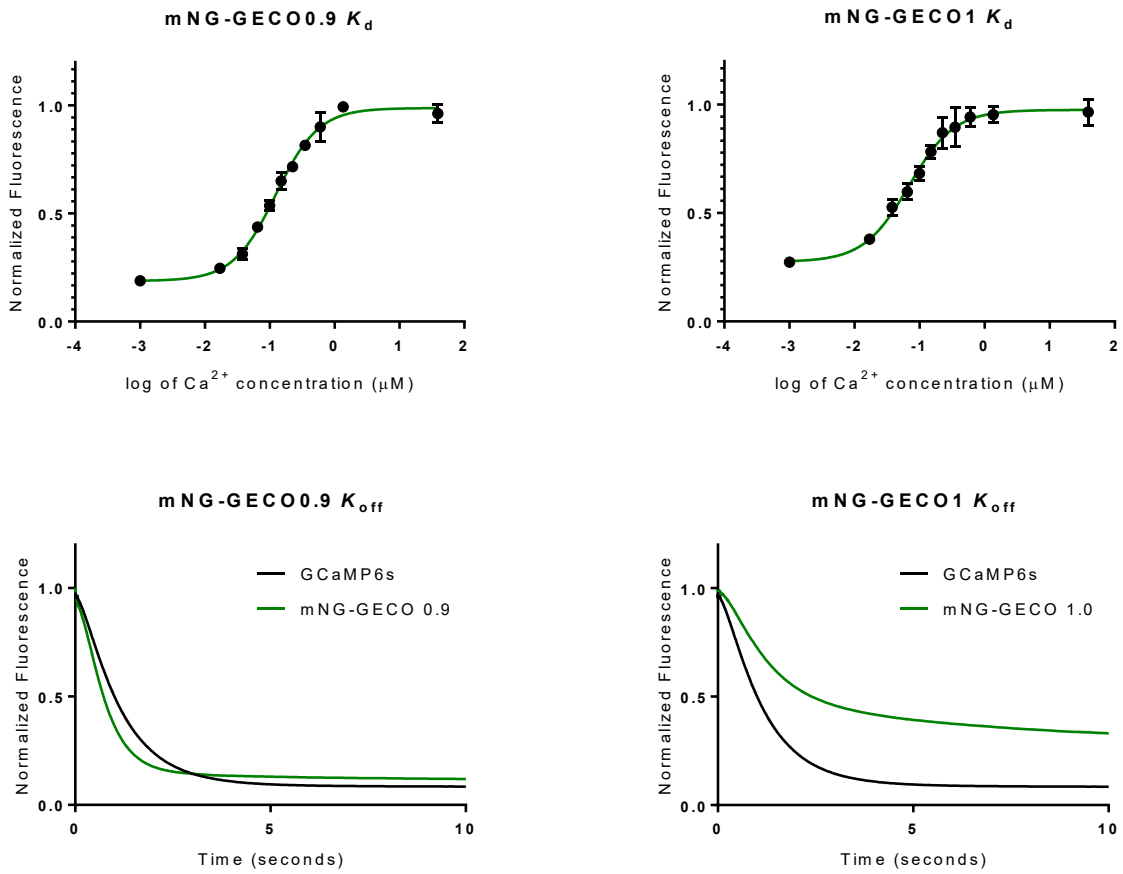


Figure 3.6 Kd and kinetics of mNG-GECO variants

We characterized our mNG-GECO variants for dissociation constant (K_d) and Ca^{2+} off kinetics (K_{off}). The K_d for mNG-GECO0.9 and mNG-GECO1 were 124 nM and 69 nM respectively. The Ca^{2+} dissociation constant for mNG-GECO0.9 was slightly faster than GCaMP6s while mNG-GECO1 was relatively slow in comparison.

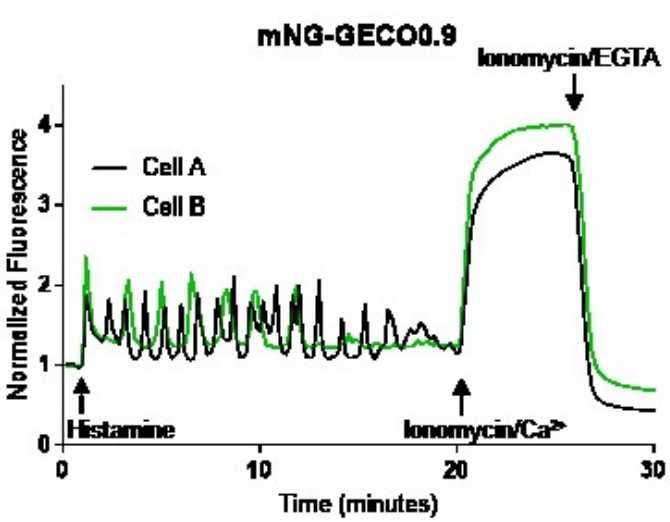
3.2.3 Characterization of mNG-GECO0.9 and mNG-GECO1 in cultured cells

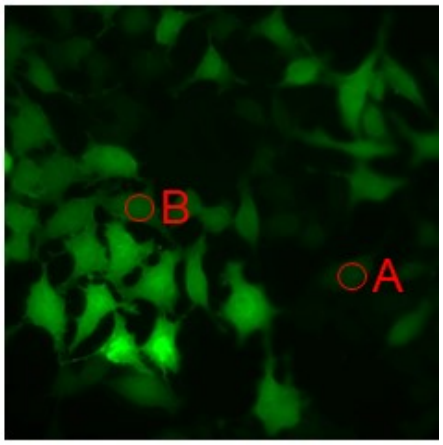
To assess the effectiveness of mNG-GECO variants to report Ca^{2+} concentration changes in mammalian cell lines, we created plasmids to transiently express mNG-

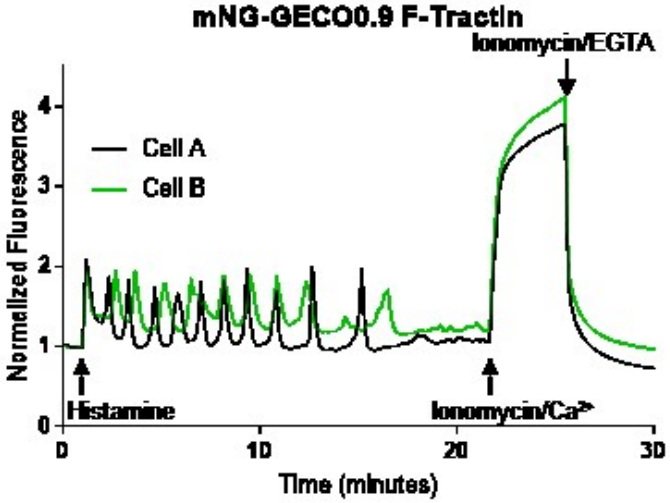
GECO0.9 and mNG-GECO1 in cultured HeLa cells. Using established protocols (220), we exposed treated mNG-GECO transfected HeLa cells with histamine for the pharmacological induction of Ca^{2+} oscillations (Figure 3.7).

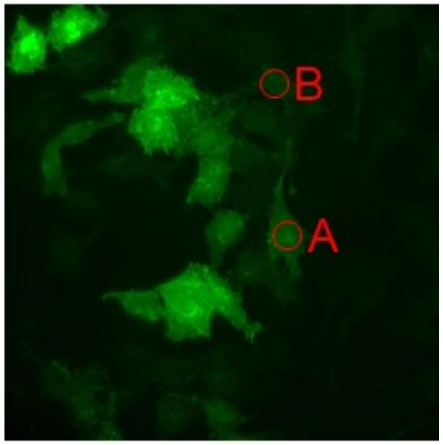
mNG-GECO0.9 shows a robust 100% increase in fluorescence during histamine induced Ca^{2+} oscillations. The ionomycin Ca^{2+}/EGTA transient, a method of inducing maximum and minimum fluorescence, revealed a maximum 8-fold fluorescence change in HeLa cells for mNG-GECO0.9 but only 3-fold fluorescence change for mNG-GECO1.

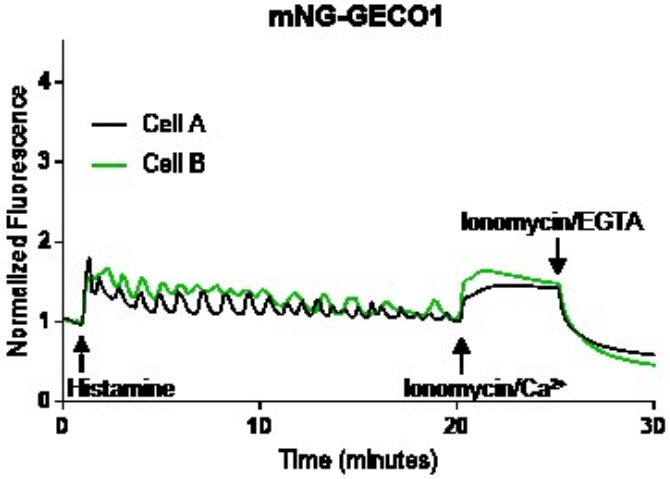
To determine if mNG-GECO0.9 could retain its function when fused to a localization tag, we genetically fused the gene to the DNA encoding an N-terminal F-Tractin localization tag (188, 189). Filamentous actin labelled with mNG-GECO0.9 harbouring this N-terminal F-Tractin localization tag has an equivalent response to mNG-GECO0.9 with no localization tag. This result indicates that localization tags do not affect the response of non-circularly permuted mNG-GECO0.9.


mNG-GECO0.9
Ionomycin/EGTA
Cell A
Cell B
Normalized Fluorescence
Histamine
Ionomycin/Ca2+
Time (minutes)


B
A


mNG-GECO0.9 F-Tractin
Ionomycin/EGTA
Cell A
Cell B
Normalized Fluorescence
Histamine
Ionomycin/Ca2+
Time (minutes)


B
A


mNG-GECO1
Cell A
Cell B
Ionomycin/EGTA
Normalized Fluorescence
Histamine
Ionomycin/Ca2+
Time (minutes)

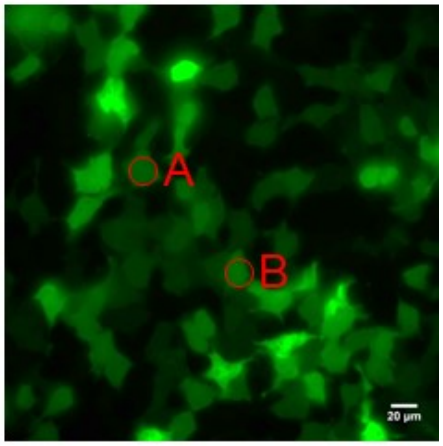

A
B
20 µm

Figure 3.7 Trace of mNG-GECO variants in HeLa cells

Fluorescent intensity time-course trace for HeLa cells expressing mNG-GECO variants. The trace is normalized to the average fluorescence before addition of histamine (60 seconds of baseline fluorescence). After 1 minute, histamine is added to a final concentration of 20 μM. After approximately 20 minutes, ionomycin/Ca^{2+} is added to produce a fluorescence maximum. Approximately 5 minutes after addition of ionomycin/Ca^{2+}, addition of ionomycin/EGTA chelates the available Ca^{2+} to produce a fluorescence minimum. Pseudo-colored image of HeLa cells represented by the trace with their region of interest marked with a red circle.

mNG-GECO1, having a lower K_d and lower maximum fluorescence intensity change, shows a relatively low response to histamine induced Ca^{2+} oscillations in HeLa cells. The mNG-GECO1 change in fluorescent intensity was 0.3-fold change during Ca^{2+} oscillations and roughly 3-fold for Ca^{2+}/EGTA treatment. The low K_d of mNG-GECO1 and limited sensitivity suggest that the majority of HeLa cells have a basal intracellular Ca^{2+} concentration higher than 69 nM, which is consistent with previous reports of the basal Ca^{2+} in HeLa cells being 100 nM (221). As both sensors were functional in cultured cells we proceeded with testing in dissociated neural tissue.

With the successful expression and Ca^{2+} dynamics reporting of mNG-GECO variants in HeLa cells, we next tested our sensors in dissociated neural tissue. Rat cortical tissue was prepared from Sprague Dawley rats, postnatal day P0-P2, enzymatically dissociated and allowed sufficient time to adhere to optical imaging dishes (55). After eight days we transfected the neural tissue with pcDNA3.1(+) plasmids containing mNG-GECO0.9 and

mNG-GECO1 and allowed the tissues to recover for two to three days before imaging. Upon observation of transfected neuronal tissues, we were able to detect spontaneous activity in a neuron harbouring mNG-GECO0.9 (Figure 3.8). This mNG-GECO0.9 transfected neuron showed spontaneous and robust ~60% increases in fluorescent intensity. Despite our best efforts, we were unable to find neurons expressing mNG-GECO1 that were able to report neural activity. This may be due to the low K_d of the mNG-GECO or that our efforts to locate active neurons were unsuccessful.

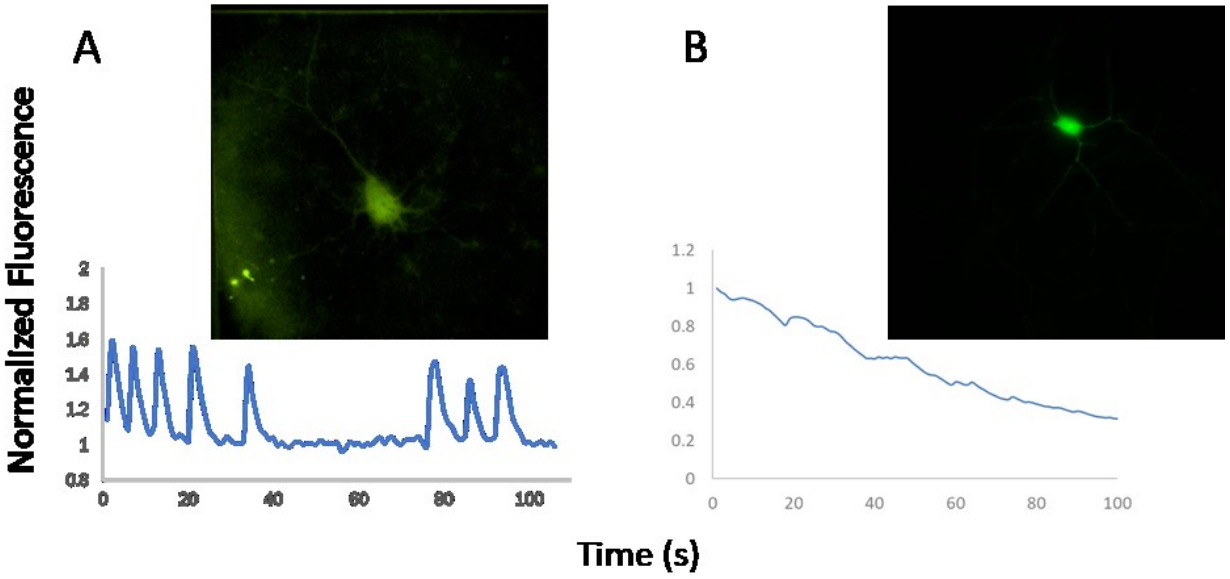


Figure 3.8 Expression in dissociated rat cortical neurons

mNG-GECO variants, mNG-GECO0.9 (A) and mNG-GECO1 (B), were expressed in dissociated rat cortical neurons and imaged 48-72 hours post transfection. Fluorescence intensity was normalized to the baseline fluorescence and showed spontaneous increases of ~60% for mNG-GECO0.9. Identical experiments with mNG-GECO1 did not lead to the identification of neurons with spontaneous changes in fluorescence intensity. An image of each neuron from these recordings is featured in the inset.

Having observed ~60% increases in fluorescent intensity when imaging of mNG-GECO0.9 in dissociated rat cortical neurons, we anticipated that we could also use this indicator to visualize spontaneous activity in zebrafish larvae. We constructed an expression vector under the zebrafish neuron specific Elavl3 promoter using the Gateway cloning protocol to place mNG-GECO0.9 into the Tol2 transposon transgenesis vector (222). We then co-transfected zebrafish embryos with our mNG-GECO0.9 Tol2 plasmid

and transposase mRNA. Upon observation of zebrafish larvae at post-fertilization day three, we observed neurons expressing mNG-GECO0.9 (Figure 3.9). However, we were unable to capture fluorescent intensity changes associated with spontaneous neural activity. The inability to observe neural activity in zebrafish neurons, despite having neuron specific localization and fluorescent signal, was unexpected. We were unable to determine the reason for the absence of response.

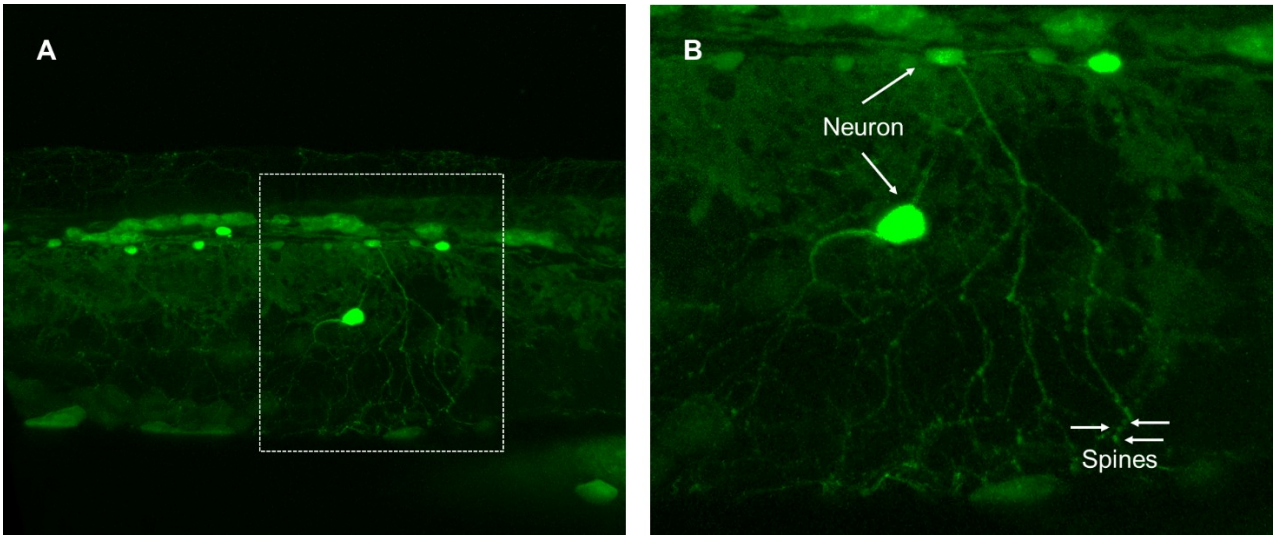


Figure 3.9 Transient expression of mNG-GECO0.9 in zebrafish larvae

zebrafish larvae transfected with mNG-GECO0.9 under control of the Elavl3 promoter for neuron specific expression. (A) Fluorescent image of zebrafish larval tail spinal cord zoomed in to show neurons expressing mNG-GECO0.9 three days post-fertilization. (B) Zoom-in on the area in a white box in image A. Neuron body and apparent neuronal spines are visible with mNG-GECO0.9 expression.

3.3 Recent improvements of mNG-GECO1

We have continued the iterative directed evolution of mNG-GECO1 and, after many additional rounds of screening, identified a more promising variant. This variant, mNG-GECO1a, has a 35-fold Ca^{2+}-dependent fluorescence change *in vitro*. However, the K_d has been substantially increased from 69 nM to 1000 nM. The higher K_d of mNG-GECO1a indicated that it might be suitable for imaging Ca^{2+} transients in HeLa cells. Indeed, mNG-GECO1a was able to report histamine induced Ca^{2+} oscillations with up to 10-fold response (Figure 3.10).

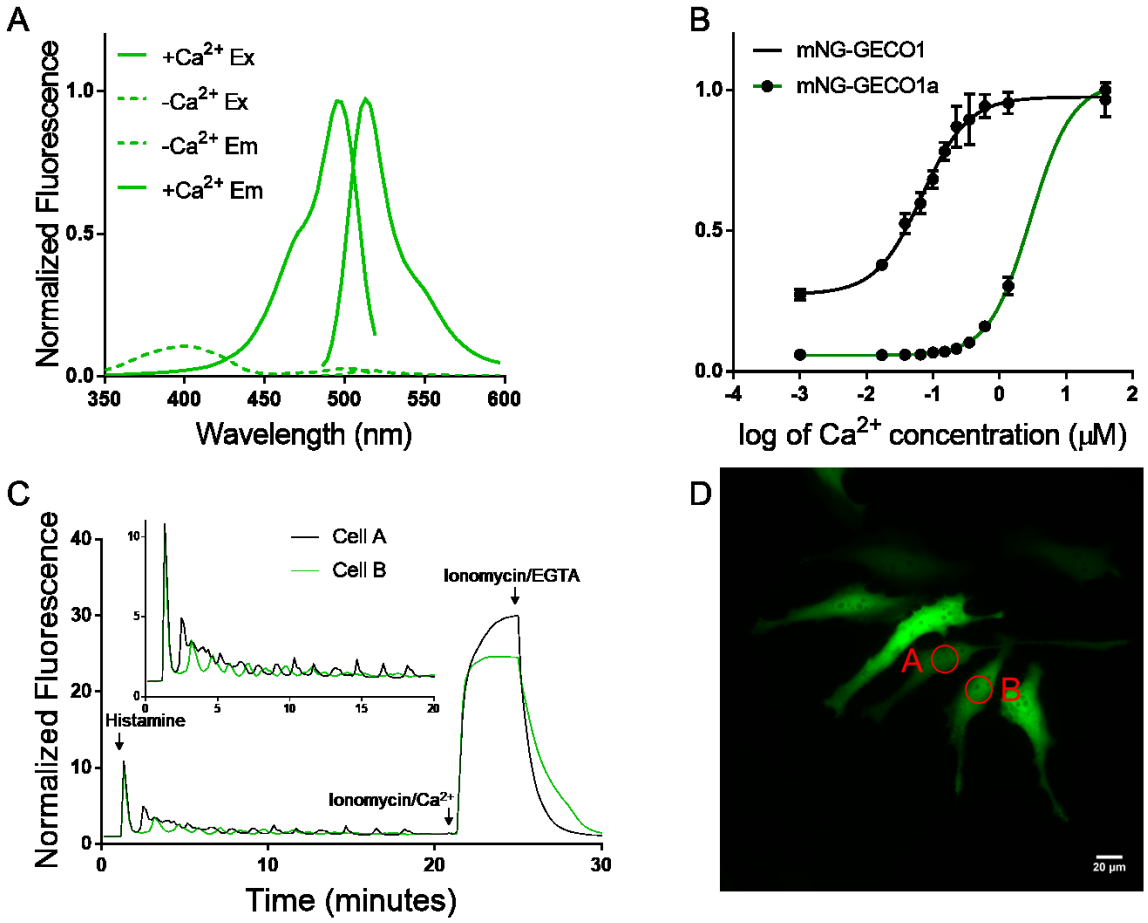


Figure 3.10 Recent developments: mNG-GECO1a

Characterization of recently developed Ca^{2+} reporter mNG-GECO1a. (A) Spectrum of mNG-GECO1a in Ca^{2+} saturated and Ca^{2+} free buffers. (B) Comparison of Ca^{2+} affinity for mNG-GECO1 and mNG-GECO1a. The same buffer system was used for this titration to compare the K_d directly for these two variants. (C) Trace of 2 representative HeLa cells induced with histamine to stimulate Ca^{2+} oscillations. (D) Image of cells from trace in (C) with the two regions of interest in red circles.

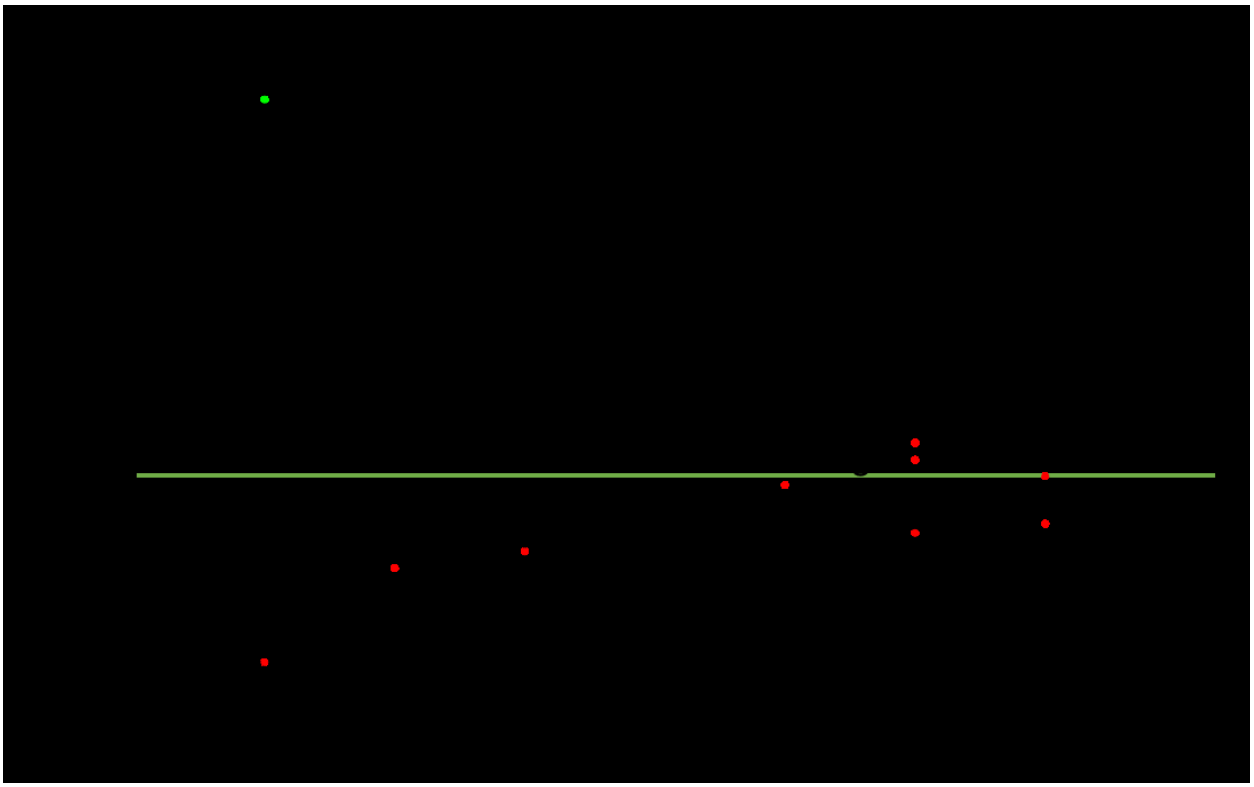


Figure 3.11 Brightness of mNG-GECO1a

The brightness of mNG-GECO1a compared to the GCaMP series of green fluorescent Ca2+ indicators. Over seven generations, the brightness of the GCaMP series plateaus near the brightness of its scaffold protein EGFP (brightness of EGFP is 37). mNeonGreen-GECO1a is brighter than its scaffold protein in the first generation, more than twice the brightness of the brightest GCaMP variants. GCaMP brightness data from Ref. (26, 59, 77).

3.4 Conclusion

We undertook the development of a genetically encodable Ca^{2+} sensor to be the template for brighter and more versatile fluorescent GCaMPs. Currently, mNG-GECO0.9, mNG-GECO1, and mNG-GECO1a represent a promising new class of Ca^{2+} sensors derived from a fluorescent protein homologue outside the Anthozoan clade. We have demonstrated with mNG-GECO0.9 that the non-circularly permuted topology is amenable to subcellular localization without aggregation or loss of sensitivity. Although we describe a green fluorescent sensor brighter than the GCaMP lineage, our efforts are still hampered through the biophysical trade-offs that often must be made during development of genetically encodable Ca^{2+} sensors. Our mNG-GECO1 sensor is brighter than GCaMP6s but suffers from decreased sensitivity and kinetics. However, as a first-in-class GCaMP-type sensor, mNG-GECO1 proves that a brighter fluorescent protein scaffold can yield a brighter fluorescent sensor. Moreover, if mNG-GECO1 is to be practically useful and competitive with current state-of-the-art neural imaging indicators, substantial additional directed evolution for improved Ca^{2+} sensing properties (i.e., decreased K_d and faster kinetics) will be required.

3.5 Experimental procedures

3.5.1 General procedures

Synthetic DNA oligonucleotides and gBlocks were purchased from Integrated DNA Technologies. Plastic consumables, restriction endonucleases, Taq polymerase, Phusion polymerase, T4 DNA ligase, deoxynucleotides, DH10B *E. coli*, pBAD/His B plasmid,

pcDNA3.1(+) plasmid, Bacterial Protein Extraction Reagent (B-PER), Penicillin-Streptomycin, Fetal Bovine Serum (FBS), TurboFect, Lipofectamine 2000, Gateway BP Clonase II kits, and GeneJet gel or plasmid purification kits were purchased from Thermo Scientific. Agarose, $MnCl_2 \cdot 4H_2O$, tryptone, D-glucose, ampicillin, L-arabinose, Hank's balanced salt solution (HBSS), DMEM, TryplE Express, and LB Lennox media were purchased from Fisher Scientific. NbActiv4 and neuron transfection media were purchased from Brain Bits.

3-(N-morpholino)propanesulfonic acid (MOPS), ethylene glycol-bis(2-aminoethylether)-N,N,N′,N′-tetraacetic acid (EGTA), and nitrilotriacetic acid (NTA), were purchased from VWR. Nickel NTA immobilized metal affinity chromatography protein purification beads were purchased from G-BioSciences. Ionomycin and tricaine methanesulfonate were purchased from Millipore-Sigma. Ethidium bromide and PCR machines (T100 Thermal Cycler) were purchased from BioRad. Gibson Assembly reagent was purchased from New England Biolabs (NEB). Genemorph II Random Mutagenesis kits and QuikChange mutagenesis kits were purchased from Agilent Technologies. Nunc 96-Well Polypropylene DeepWell Storage Plates (cat. 278743) and 96-well Nunc MicroWell 96-Well Optical-Bottom Plates (cat. 265301) were purchased from Thermo Scientific. Molecular weight cut off filters were purchased from Millipore-Sigma. Sequencing was completed by the Molecular Biology Services Unit at the University of Alberta. mNeonGreen was a kind gift from Jiwu Wang at Allele Biotech. pGP-CMV-GCaMP6s was a gift from Douglas Kim (Addgene plasmid # 40753; http://n2t.net/addgene:40753; RRID:Addgene_40753) (57). Plasmids for the generation of zebrafish transgenesis were provided by Ted Allison's lab at the University of Alberta.

3.5.2 Molecular biology and protein engineering

Libraries for iterative directed evolution were created using Genemorph II Random Mutagenesis kits and NEB's Gibson Assembly reagent. Blunt ended linear DNA fragments with random mutations are created using the Genemorph II kit according to the manufacturer's recommendations. Genemorph II PCR product was ligated into a pBAD vector cut with XhoI/HindIII using NEB Gibson Assembly reagent. Site-directed mutagenesis libraries were created using single and multi QuikChange mutagenesis kits according the manufacturers recommendations.

Libraries are then transformed into DH10B *E. coli* and plated on 100 μg/mL ampicillin/1.5% agar plates with 0.02% L-arabinose and grown overnight (12-18 hours) at 37 °C. Colonies are selected on the basis of fluorescence intensity, picked, and placed into 96 DeepWell blocks containing 1.3 mL of LB Lennox media supplemented with 100 μg/mL ampicillin and 0.02% L-arabinose. Deepwell blocks were shaken overnight at 37 °C. The next day, blocks are centrifuged at 6000 $\times$ g for 5 minutes to pellet cells. Media was discarded and 30 µL of B-PER was added to each well. After shaking for 15 minutes, 200 µL of 10 mM EGTA in 30 mM MOPS/100 mM KCl pH 7.2 (MOPS/KCl buffer) is added to each well of the blocks before mixing briefly and being centrifuged again for 5 minutes at 6000 $\times$ g. 90 µL of the resulting lysate is loaded in each well of 96-well optical bottom plates. Fluorescence intensity for each well of the plate is then read with a Tecan Safire^2 microplate reader to determine the low Ca^{2+} intensity for each variant. High Ca^{2+} intensity is acquired by adding 15 µL of 100 mM Ca^{2+} in 30 mM MOPS pH 7.2 with a 60 second shake before reading. Taking the value of the high Ca^{2+} intensity divided by the low Ca^{2+} intensity gives a relative sensitivity value. Promising candidates, usually 10% of each 96-

well block, are retested from the lysate in 10 mM low (EGTA chelated) and 10 mM high Ca^{2+} solution diluted in MOPS/KCl buffer. The plasmids associated with the promising variants are sent for sequencing and used as template for the next round of directed evolution.

To create constructs for subcellular localization, we used PCR of pcDNA plasmid using Phusion polymerase to create linear fragments suitable for Gibson Assembly cloning. We constructed an F-Tractin (filamentous actin) labelling vector in a similar manner except the template plasmid for inverse PCR had mNG-GECO ligated into it previously using XhoI/HindIII and T4 ligase according to the manufacturer's recommendations. IDT gBlock for F-Tractin was synthesized and Gibson Assembly was used to assemble the vector. XhoI/HindIII restriction digest sites were used to create Gibson Assembly ready vectors for testing new variants in each localization construct.

Constructs for zebrafish transfection were created using the Gateway BP Clonase II kit and the Tol2kit system as previously described(222). Briefly, PCR of mNG-GECO variants using 'attB1 MVSK Fwd' and 'attB1 MDELYK Rv' was used to produce a linear fragment suitable for BP Clonase recombination into pDONOR-221. The resulting donor plasmid harbouring the mNG-GECO variant is then used for recombination into the pDEST-221 plasmid. The final plasmid harbouring mNG-GECO under Elavl3 promoter is then purified for injection into fertilized zebrafish embryos.

3.5.3 Protein purification and *in vitro* characterization

To purify mNG, mNG-GECO variants, and GCaMP6s for *in vitro* characterization, pBAD/His B plasmid containing the gene of interest was used to transform electrocompetent DH10B *E. coli*, which were then streaked on 100 μg/mL ampicillin/1.5%

agar plates. After overnight incubation at 37 °C, a single colony was picked and transferred to a 2 L flask containing 500 mL of 100 μg/mL ampicillin/0.02% L-arabinose liquid media and cultured for 24-30 hours at 37 °C. The culture is then centrifuged at 6000 $\times$ g for 6 minutes to collect the cells. Cells are resuspended in 30 mL of cold Tris buffered saline (TBS, 150 mM NaCl, 50 mM Tris-HCl) pH 8.0 and lysed by sonication (QSonica Q700, amplitude 50, 1 second on, 2 seconds off, 3 minutes sonication time). All subsequent purification procedures were performed on ice. The resulting lysate was clarified of cell debris by centrifugation for 1 hour at 21,000 $\times$ g, filtered through a Kim-wipe into a 50 mL conical bottom tube, and incubated for 3 hours with Ni-NTA resin. Resin containing NTA bound protein was washed with 100 mL of 20 mM imidazole TBS wash buffer and eluted with 250 mM imidazole TBS elution buffer. Purified protein was buffer exchanged into TBS using a 10,000 Da molecular weight cut-off filter (Millipore-Sigma) through 3 successive washes. Absorption spectra were recorded on a Beckman-Coulter DU-800 UV-visible spectrophotometer and fluorescence spectra recorded on a Tecan Safire^2 plate reader.

Extinction coefficient determination for mNG-GECO variants was performed using the alkaline denaturation method with mNG as a standard(77). Briefly, the concentration of protein was adjusted by dilution in MOPS/KCl pH 7.2 to reach an absorbance of 0.6 to 1.0. A dilution series with MOPS/KCl and 10 mM Ca^{2+} was then prepared with absorbances of 0.01, 0.02, 0.03, 0.04, and 0.05 for mNG, mNG-GECO variants, and GCaMP6s. Integration of the fluorescent peaks provides a total fluorescent emission value which was plotted against the absorbance to provide a slope. The quantum yields

of mNG-GECO variants were determined using the published(77) QY value of mNeonGreen in a ratiometric manner:

$$\Phi_{protein} = \Phi_{standard} \times (S_{protein}/S_{standard})$$

Extinction coefficients were determined by measuring the absorption spectrum in MOPS/KCl pH 7.2 and 2 M NaOH. The absorbance value for the denatured fluorescent protein at 440 nm was divided by the previously determined extinction coefficient of 44,000 $M^{-1}cm^{-1}$ to give the concentration of protein[ref]. The extinction coefficient was then determined by dividing the TBS sample absorbance maximum by the calculated protein concentration.

Determination of K_d was performed as previously described(223, 224). Briefly, a reciprocal dilution series was created with either 10 mM EGTA/10 mM Ca^{2+} EGTA ranging in free Ca^{2+} concentration of 0 to 0.039 mM or 10 mM NTA/10 mM Ca^{2+} NTA ranging in free Ca^{2+} concentration from 0 to 1.13 mM (225). An equal amount of purified mNG-GECO was diluted 100× into 100 µL of buffer and the intensity plotted against free Ca^{2+} in triplicate. The data are then fit to a four-parameter variable-slope in GraphPad Prism 7 software to determine the K_d.

3.5.4 Fluorescence live cell imaging

Imaging in HeLa cells. We followed previously established protocols for our Ca^{2+} imaging experiments (220). Briefly, HeLa cells cultured in DMEM with 10% fetal bovine serum supplemented with penicillin-G potassium salt (50 units/mL) and streptomycin sulphate (50 µg/mL) were plated on collagen coated 35 mm glass bottom dishes. HeLa cells are transfected at 60% confluency with 1 µg of pcDNA3.1(+) harbouring the variant

of interest using 2 µL of TurboFect according to the manufacturer's recommendation. After overnight incubation at 37 °C with 5% CO_2, cells were washed twice with prewarmed Hank's balanced salt solution immediately before imaging.

Imaging of transfected HeLa cells was performed on an inverted Zeiss 200M microscope with Semrock filters (excitation 470/40, emission 525/50) and captured with an OrcaFlash 4.0 – C13440 (Hamamatsu). Images were acquired through a 40× (N.A. 1.3) oil immersion lens using MetaMorph 7.8.0.0 software and an MS-2000 automated stage (Applied Scientific Instrumentation).

Imaging in dissociated rat cortical neurons. Cortical neurons were prepared from Sprague Dawley rats, postnatal day P0-P2, using TryplE Express dissociation reagent and plated on poly-D-lysine coated 24-well glass bottom imaging plates. Neurons were maintained in NbActiv4 medium supplemented with 2% FBS, penicillin-G potassium salt (50 units/mL) and streptomycin sulphate (50 µg/mL) at 37 °C and 5% CO_2 with half of the media replaced every four days. Dissociated neuronal tissue was transfected on day 8 using 1 µg of pcDNA3.1(+) harbouring the variant of interest and 2 µL of Lipofectamine 2000. Both the DNA and transfection reagent are mixed with 50 µL NbActive4 media before being combined and set aside for 15 minutes. Before the 100 µL of transfection solution is applied to the dissociated tissue, half of the media (500 µL) was removed and mixed with an equal part NbActive4 and set aside in the incubator at 37 °C with 5% CO_2. Three hours after addition of the transfection solution, the media is exchanged with the media set aside and the tissue allowed to recover for two to three days in the incubator before imaging. Before imaging, the tissue samples are washed three times with

prewarmed and equilibrated HBSS. Images were acquired using the same microscope, camera, software, and filters as described for HeLa cell imaging experiments.

Imaging of zebrafish larvae. zebrafish were maintained and bred in according to approved protocols of the University of Alberta's Animal Care and Use Committee. Briefly, fertilized zebrafish embryos were prepared for injection by separating males and females the evening before and allowing them to mate in the morning for a maximum of four hours. The fertilized eggs are collected and loaded onto a custom egg injection apparatus. Embryos are injected with an osmotically balanced solution containing transposase mRNA and Tol2 transgenesis vector containing the mNG-GECO variant under Elavl3 neuron specific promoter (222). Injected embryos are kept at 30 °C in petri plates filled with E3 media and dead embryos removed twice daily. Larvae were screened for green fluorescence in the central nervous system and upon confirmation of positive transgene expression they are separated and allowed to develop. At 3 days post-fertilization, the larvae are anesthetized in 0.24 mg/mL tricaine methanesulfonate and embedded in 2% low gelling agarose. Maximum intensity projections were acquired from Z-stacks (4 µm steps) using a 20×/0.8 objective and a laser-point scanning confocal microscope (Zeiss 700). Images were further processed for proper orientation using Imaris 7.6 (Bitman, Zuerich).

Table 3.2 Filters for mNG-GECO variant imaging

Sensor	Excitation (nm)	Emission (nm)
mNG-GECO variants	470/40	525/50

Table 3.3 Oligos and gBlocks used in this work

Primer Name	Sequence (5' to 3')
Construction and Directed Evolution	
pBAD Vec Fwd	AAGCTTGGCTGTTTTGGCGGATG
pBAD Vec Rv	CCTCGAGCTCGGATCCTTATCGTC
FP GA MVSK Fwd	GGATCCGAGCTCGAGGATGGTGAGCAAGGGCGAG
FP GA MDELYK Rv	CCAAAACAGCCAAGCTTTTATTTGTACAAACCGCCGGCC
mNG CaM/RS20 vec fwd	GACTGGTGCAGGTCGAAGAAGAC
mNG CaM/RS20 vec Rv	CGCAGCGGTCAGCGAG
CaM GA Fwd	CTCGCTGACCGCTGCGACGCGTGACCAACTGACTGAAG
RS20 Rv	CCTGCACCAGTCTGAGCTCAGCCGACCTATAGCTC
mNG-QC-del_A146	CTCGCTGACCGCTACGCGTGACCAACTGACTGAAG
mNG-QC1-P263X	GATGACAGACCTTGGAGAGAAGTTANNKGATNNKGAGGTTGATGAAATGATCAGGGTAGC
mNG-QC2-A180X	CGATAACAACCAAGGAGCTGGGGNNKGTGTTGCGGTCTCTGGGGC
mNG-QC3-G293X	CAGGTAAACTACGAAGAGTTTGTACAAATGATGACANNKAAGGGTGGCGGAGGTTCTG
mNG-QC4-W324/R326X	GCTATAGGTCGGCTGAGCTCAGACNNKTGCNNKTCGAAGAAGACTTACCCCAACGAC
mNG-QC5-F380/K382X	CTATCTGAAGAACCAGCCGATGTACGTGNNKCGTNNKACGGAGCTCAAGCACTCCAA
mNG-QC6-W324G/R326G	GCTATAGGTCGGCTGAGCTCAGACGGGTGCGGGTCGAAGAAGACTTACCCCAACGAC
mNG-QC7-N243X	CCGCGTGTTTGATAAGGACGGCNNKGGCTACATCGGCGCAGCAGAG
mNG-QC8-L250X	GCTACATCGGCGCAGCAGAGNNKCGCCACGTGATGACAGACCTTG
mNG-QC9- L143X	CGGTCCTGTGATGACCAACTCGNNKACCGCTACGCGTGACCAAC
mNG-QC10-T144X	GTCCTGTGATGACCAACTCGCTGNNKGCTACGCGTGACCAACTGGC
mNG-QC11-A145X	CTGTGATGACCAACTCGCTGACCNNKACGCGTGACCAACTGGC
mNG-QC12-D323X	GAGCTATAGGTCGGCTGAGCTCANNKTGGTGCAGGTCGAAGAAGACTTACC
mNG-QC13-W324X	GCTATAGGTCGGCTGAGCTCAGACNNKTGCAGGTCGAAGAAGACTTACCCC
mNG-QC14-C325X	GTCGGCTGAGCTCAGACTGGNNKAGGTCGAAGAAGACTTACCCCAACG
mNG-QC15-R326X	GGCTGAGCTCAGACTGGTGCNNKTCGAAGAAGACTTACCCCAACGACAAAAC
mNG-QC16-323/324X	GAGCTATAGGTCGGCTGAGCTCANNKNNKTGCAGGTCGAAGAAGACTTACCCC
mNG-QC17-324/325X	GCTATAGGTCGGCTGAGCTCAGACNNKNNKAGGTCGAAGAAGACTTACCCCAACG
mNG-QC18-325/326X	GTCGGCTGAGCTCAGACTGGNNKNNKTCGAAGAAGACTTACCCCAACGACAAAACC
mNG-QC19-R326/S327X	CGGCTGAGCTCAGACTGGAATNNKNKKAAGAAGACTTACCCCAACGACAAAACC
mNG-QC20-S327/K328X	GCTGAGCTCAGACTGGAATAGGNNKNNKAAGACTTACCCCAACGACAAAACCATCATCAG
mNG-QC21-K238/K239X	GCTGAGCTCAGACTGGAATAGGTCGNNKNNKACTTACCCCAACGACAAAACCATCATCAG
mNG-QC22-A151/G152X	GATTACCGGTACGCGTGACCAACTGNNKNNKGAGCAGATCGCAGAGTTTAAAGAGGC
mNG-QC23-P263/K265X	GATGACAGACCTTGGAGAGAAGTTANNKGATNNKGAGGTTGATGAAATGATCAGGGTAGC
mNG-QC24-T292/G293X	GTCAGGTAAACTACGAAGAGTTTGTACAAATGATGNNKNNKAAGGGTGGCGGAGGTTCTG
mNG-QC25-G293/K294X	CAGGTAAACTACGAAGAGTTTGTACAAATGATGACANNKNNKGGTGGCGGAGGTTCTGTC
mNG-QC26-L209X	CGACGGTGACGGCACANNKGACTTCCCTGAGTTCCTGACGATGATG
mNG-QC27-I346X	CAAAACCATCATCAGTACCTTTAAGTGGAGTTACNNKACTGGAAATGGCAAGCGCTAC
mNG-QC28-G209X	GATCAATGAAGTAGATGCCGACGGTNNKGGCACACTCGACTTCCCTGAG
mNG-QC29-D277X	GGTTGATGAAATGATCAGGGTAGCAGACATCNNKGGGGATGGTCAGGTAAACTACGAAG
ER Targetting	
pcER Fwd MVSK	CCTGTGTTGTTGCTGGGACTGCTGGGGGCGGCCGCGGACATGGTGAGCAAGGGCGAG
pcER Fwd 2	AGCGTGCCGCCACCATGCTCTTGCCTGTGTTGTTGCTGGGACTGCTG
ER MDELYK Rv	GGCATGGACGAGCTGTACAAGAAAGACGAGCTTTGAGGTAAACCGCTGATCAGCCTC
pcDNAlocal Vec Fwd	GGTAAACCGCTGATCAGCCTC
pcDNAlocal Vec Rv	GGTGGCGGCACGCTG
Nesprin 3	
Nes3 GA Fwd	AAGCTTTCCAGTGGATATTCGTCTCGGTCGG
Nes3 GA Rv	GTCGAGGCTGATCAGCGGTTTACCTTAGGTGGGTGGTGGGCCATTG
MVSKGE pcDNA GA Fwd	AGCGTGCCGCCACCATGGGCTCGAGGATGGTGAGCAAGGGCGAG
GMDELYK Nes3 GA Rv	CCGACCGAGACGAATATCCACTGGAAAGCTTCTTGTACAGCTCGTCCATGCC
pcDNA Nes3 Vec Rv	CGAGCCCATGGTGGCGGCACGCT
pcDNA Nes3 Vec Fwd	GGTAAACCGCTGATCAGCCTCGAC
Nes3 L44A	GCCCTGGAGGCCAGGGCGTGGGAGACCGAGAAAATATGCCAG
F-Tractin	
pcDNA MVSK Vec Fwd	ATGGTGAGCAAGGGCTCGG
pcDNA FT GA Rv	GGTGGCGGCACGCTAGCC
F-Tractin gBlock	GCTAGCGTGCCGCCACCATGGCTAGACCCCGAGGTGCCGGCCCTTGTAGTCCTGGTCTGGAGCGCGCTCCTC GACGCTCAGTAGGTGAACTGAGGTTGTTGTTCGAGGCCAGATGTGCTGCTGTGGCCGCTGCAGCAGCTGCCG GAGGGTTGGCTCTGCCGGTAGCCACCGGCGGGAGCGGGGGCAGCTCGAGGATGGTGAGCAAGGGCTCGG
Zebrafish	
attB1 MVSK Fwd	GGG GACAAG TTT GTA CAA AAA AGC AGG CTT AGC CGC CAC CAT GGT GAG CAA GGG C
attB1 MDELYK Rv	GGG GAC CAC TTT GTA CAA GAA AGC TGG GTG TTA CTT GTA CAG CTC GTC CAT GCC CAT

Chapter 4: iCobA as an enzymatic reporter of protein-protein interactions that uses a ubiquitous endogenous substrate

4.1 Introduction

Genetically encodable fluorescent reporters of transcription and protein-protein interactions have helped to illuminate the inner biological world of the cell. Prior to the development of genetically encoded fluorescent reporters, researchers interested in studying living cells and organisms relied on chemically-synthesized reporters and probes to illuminate subcellular structures and investigate cellular metabolism. Synthetic fluorescent reporters are less desirable because they can be challenging to deliver inside the cell, are not cell type specific, and must be injected prior to performing imaging experiments (226). In contrast, genetically encodable fluorescent reporters are routinely expressed in specific cell types, can be easily targeted subcellularly, and do not require multiple rounds of administration. Furthermore, cell lines and model organisms that have been genetically modified to stably express a reporter are highly valuable to the research community because they facilitate sharing and experimental reproducibility between labs and enable biological questions to be addressed in the context of a live animal.

Genetically encoded fluorophores are now routinely used as transcriptional reporters and subcellular localization tags. A more challenging, and ultimately more informative, type of application is to use versions of them to illuminate dynamic changes in cellular metabolism and the underlying protein-protein interactions. The importance of this work was made evident with the completion of the human genome project, which delivered a plethora of data for biochemists and biologists to annotate on a biomolecular level (227).

There are estimated to be 650,000 protein-protein interactions in the human interactome (228). To investigate these interactions, researchers require non-invasive and genetically encodable high-throughput assays that can be genetically integrated into model organisms and cell lines. Furthermore, genetically encodable optical reporters are well suited for high-throughput assays because they can be spectrally separated and recorded simultaneously without any chemical perturbation of the sample.

A number of optical protein-protein interaction reporters have been developed using split proteins (229-231). Split protein reporters are composed of two genetically separated coding sequences, each encoding for part of a reporter gene. These two separate coding sequences can be separately genetically fused to the open reading frames of the two genes of interest. When the chimeric genes (that is, two genes of interest, each fused to a gene encoding part of the reporter gene) are expressed and their corresponding proteins interact, the two fragments associate and the full active reporter gene is reconstituted. For split fluorescent proteins as an example, successful interactions would result in an increased fluorescent signal. For a split luciferase enzyme, a positive interaction would lead to an increase in the emission of light. There are numerous split proteins to choose from, each with its pros and cons, when designing an assay for protein-protein interactions. For example, split fluorescent protein reporters are not reversible and can therefore suffer from a high level of background fluorescence due to transient or non-specific interaction. Many enzymatic reporters do undergo reversible association and avoid this problem, but they require the addition of fluorogenic or luminescent substrate. This need for an exogenous substrate is associated with some disadvantages. The exogenous substrate must be membrane permeable in order to be used in a live cell

assay, it could potentially be toxic, and it some substrates are quite expensive. In addition, the need for an external substrate greatly limits the utility of enzymatic reporter assays in animal models where substrate delivery would be impractical.

We sought to expand the toolbox of protein-protein interaction reporters and overcome some specific limitation associated with currently available systems. In particular, the development of a genetically encodable fluorescent reporter that uses an endogenous substrate was intriguing to us. Using an endogenous substrate ought to give the reporter a natural advantage relative to other systems because it would not require addition of any exogenous cofactors or reagents. A particularly promising class of enzymes for developing such a system are the *S*-adenosyl-L-methionine uroporphyrinogen III methyltransferases (SUMT) (232, 233). SUMT enzymes are involved in the synthesis of heme in higher eukaryotes and the synthesis of siroheme and cobalamin in bacteria. The first SUMT gene to be used as a transcriptional reporter was the *CobA* gene (cobalamin biosynthetic pathway protein A) from *Pseudomona freudenrechii* (234-236). The CobA enzyme catalyzes the transfer of two methyl groups to uroporphyrinogen III, a ubiquitous early intermediate in the pathway to heme, siroheme, and cobalamin (Figure 4.1). High expression of CobA in *Escherichia coli* bacteria results in the production and accumulation of red fluorescent products, trimethylpyrrocorphin and sirohydrochlorin (Figure 4.2).

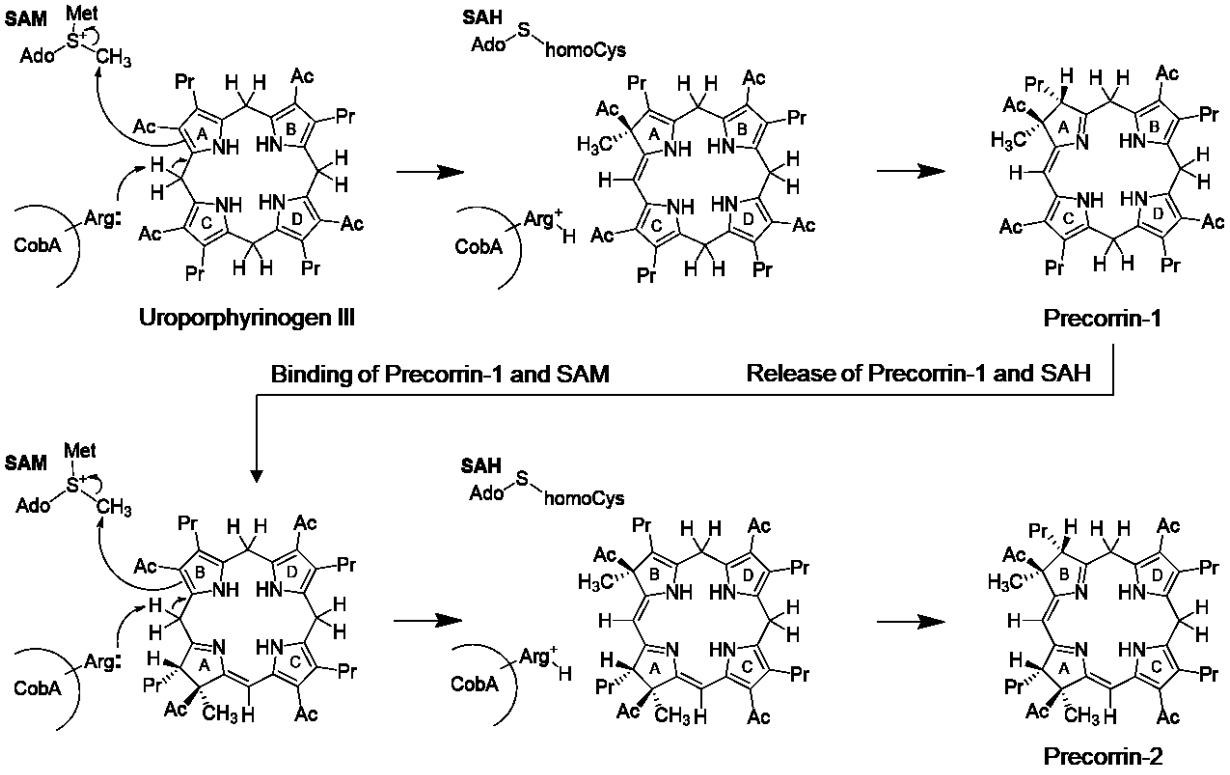


Figure 4.1 Proposed mechanism of CobA methyl transfer to Uroporphyrinogen III

Proposed mechanism of CobA methyltransferase activity. SAM, *S*-adenosyl-L-methionine, SAH, *S*-adenosyl-L-homocysteine Ac = CH_2COOH, Pr = CH_2CH_2COOH. This figure adapted from reference (237).

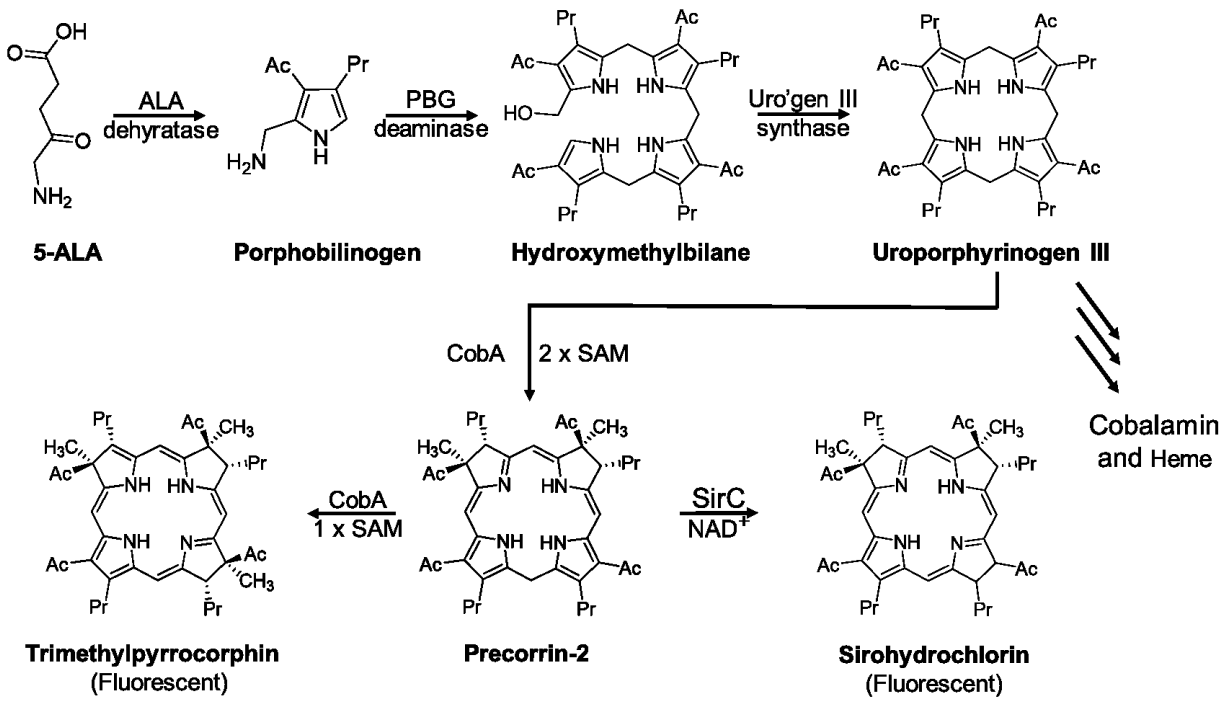


Figure 4.2 Biosynthesis of red fluorescent products from CobA enzyme

Biosynthetic pathway resulting in red fluorescence in *E. coli*. Overexpression of the *CobA* gene from *P. freudenreichii* results in the production of red fluorescent porphynoid compounds. Abbreviations: 5-ALA, 5-Aminolevulinic acid; PBG, porphobilinogen; Uro'gen, Uroporphyrinogen; SAM, *S*-adenosyl-L-methionine; SirC, Precorrin-2 dehydrogenase; NAD^+, Nicotinamide adenine dinucleotide, Ac = CH_2COOH, Pr = CH_2CH_2COOH. This figure is adapted from reference (238).

The attractive properties of CobA as a potential protein-protein interaction reporter inspired us to develop iCobA as a split enzyme reporter of protein-protein interactions that uses a ubiquitous endogenous substrate. Herein we report the directed evolution of CobA for increased fluorescent emission to facilitate low expression transcriptional

reporting and the combination of split CobA with a split intein for the development of the iCobA reporter.

4.2 Results and Discussion

4.2.1 Rational engineering and iterative directed evolution of CobA

We used a combination of rational structure-guided engineering and iterative directed evolution to develop an improved CobA reporter enzyme (Figure 4.3). Libraries generated by random mutagenesis and site-directed mutagenesis were used to transform *E. coli* and plated onto agar plates. If multiple promising variants were selected during a single round of screening, the corresponding plasmids were mixed and used as the template for a staggered extension polymerase chain reaction resulting in the recombination of multiple mutations (178). When a library of CobA variants did not provide a distinguishable range of fluorescent brightness after overnight incubation (i.e., the majority of colonies were bright), the expression of CobA was decreased by reducing the amount of transcriptional inducer L-arabinose.

The directed evolution strategy was continued for ten rounds until the expression of CobA was clearly evident at the lowest titratable concentration of L-arabinose recommended by the manufacturer (0.00002% w/v L-arabinose). By the 10^{th} round of directed evolution, the red fluorescence of the most highly evolved CobA variants could not be distinguished from each other. At this point we stopped the directed evolution of CobA. The variant with the most red fluorescence at the lowest expression level was named CobA1. CobA1 has twelve mutations compared to wild-type CobA (CobA-WT): Asp54Asn/Gln76Arg/Val98Ile/Gln107Leu/Gln143Arg/Ser149Thr/His151Gln/His181L/Ser 204Arg/Ala221Val/His229Gln/Leu235Val (Figure 4.4). Having maximized the

performance of the enzyme using our directed evolution scheme, its performance was compared to a previously published SUMT homologue: mutant barley Uroporphyrinogen III methyltransferase (mbUMT) (239).

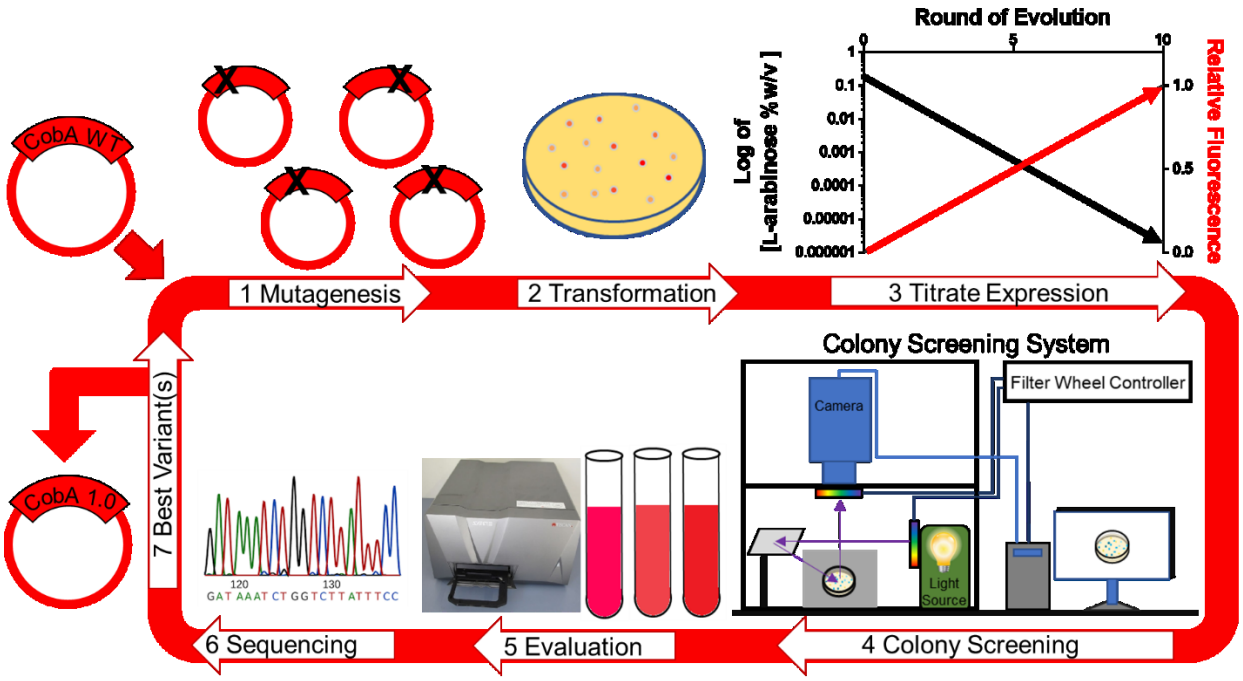


Figure 4.3 Directed evolution of CobA

Directed evolution scheme for CobA. The gene encoding the CobA enzyme was subjected to random and site-directed mutagenesis over ten rounds of iterative directed evolution. Colonies were screened for brightness and the best variants cultured for further evaluation. As the fluorescent output of the colonies increased the expression of CobA was decreased as a method to screen more active variants.

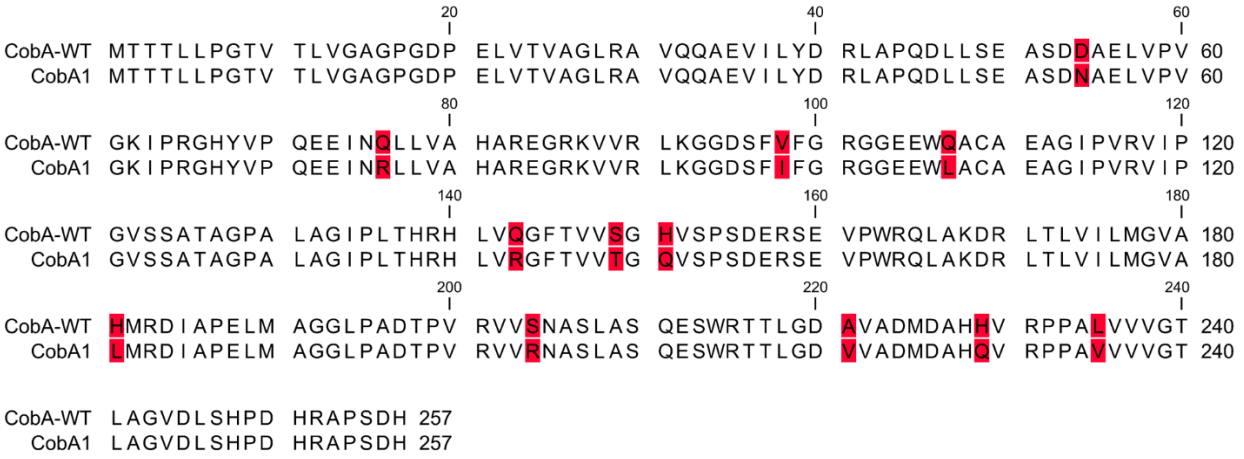


Figure 4.4 Alignment of CobA-WT and CobA1

Protein sequence alignment of CobA-WT and CobA1. The twelve mutations introduced during the directed evolution of CobA1 are highlighted in red.

We compared our evolved CobA1, and other selected variants, to a previously published SUMT that had also been subjected to directed evolution for improved red fluorescent molecule production. This previously published SUMT was a mutant barley uroporphyrinogen methyltransferase (mbUMT) (239). All variants were cloned into the same expression vector, pBAD/His-B, and streaked on high (0.2%) and low (0.00002%) expression L-arabinose/agar plates (Figure 4.5). Fluorescence imaging of the plates revealed that, at high expression levels, CobA-WT, CobA0.5 (the variant from round 5), CobA1, and mbUMT, all produce a strong red fluorescent emission. In contrast, under the low expression level conditions, only CobA1 gave a strong red fluorescent emission. Using the production of red fluorescence as a proxy for enzyme activity, this result indicates that CobA1 has superior enzyme activity relative to CobA-WT and mbUMT at low expression levels.

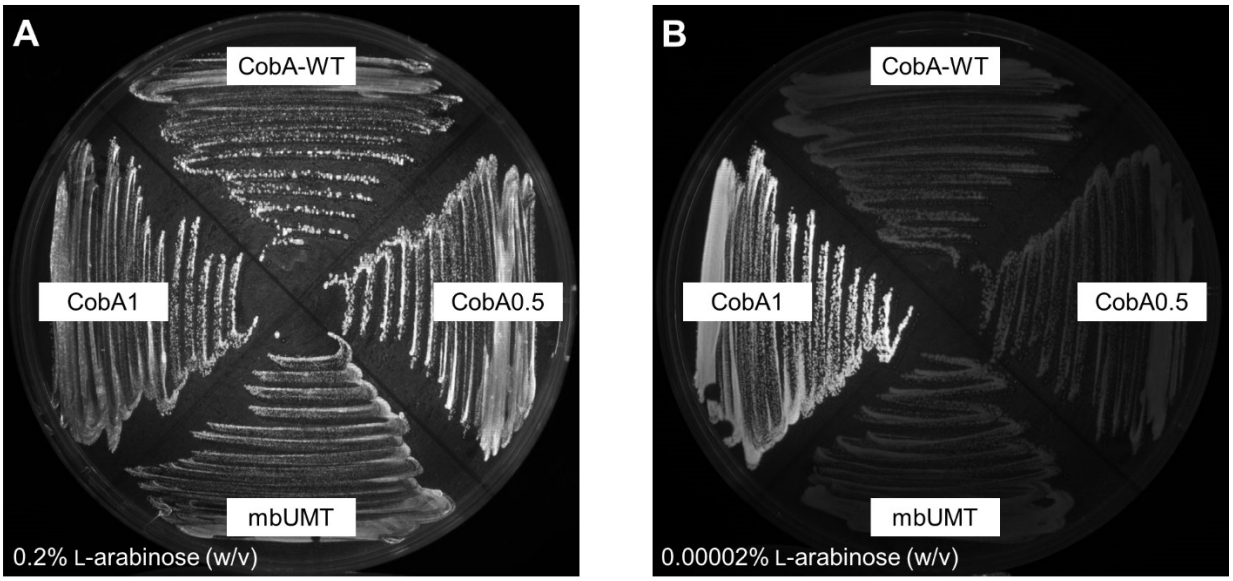


Figure 4.5 CobA1 expressing *E. coli* colony fluorescence

Expression of CobA-WT, CobA0.5 (an intermediate in the evolution of CobA1), mbUMT, and CobA1. At high expression (A) all bacteria expressing SUMT genes are red fluorescent, at low expression (B) only colonies expressing CobA1 exhibit substantial fluorescence.

Based on the strong fluorescence observed for CobA1 at low level expression, we concluded that CobA1 is a superior enzyme for reporting low levels of gene transcription on agar plates. We further evaluated the ability CobA1 to report transcription in liquid culture by growing *E. coli* harbouring plasmids containing CobA-WT, CobA1, and mbUMT. Overnight expression (approximately 18 hours) at different concentrations of transcriptional inducer L-arabinose indicated that CobA1 provides stronger red fluorescence at low concentrations and similar red fluorescence at high concentrations. To mitigate the potential metabolic burden diverted SAM consumption, we normalized the

raw fluorescent intensity of the cultures to their optical density at 600 nm (Figure 4.6). We also observed that, over a range of expression levels, CobA-WT is an equal transcriptional reporter to mbUMT.

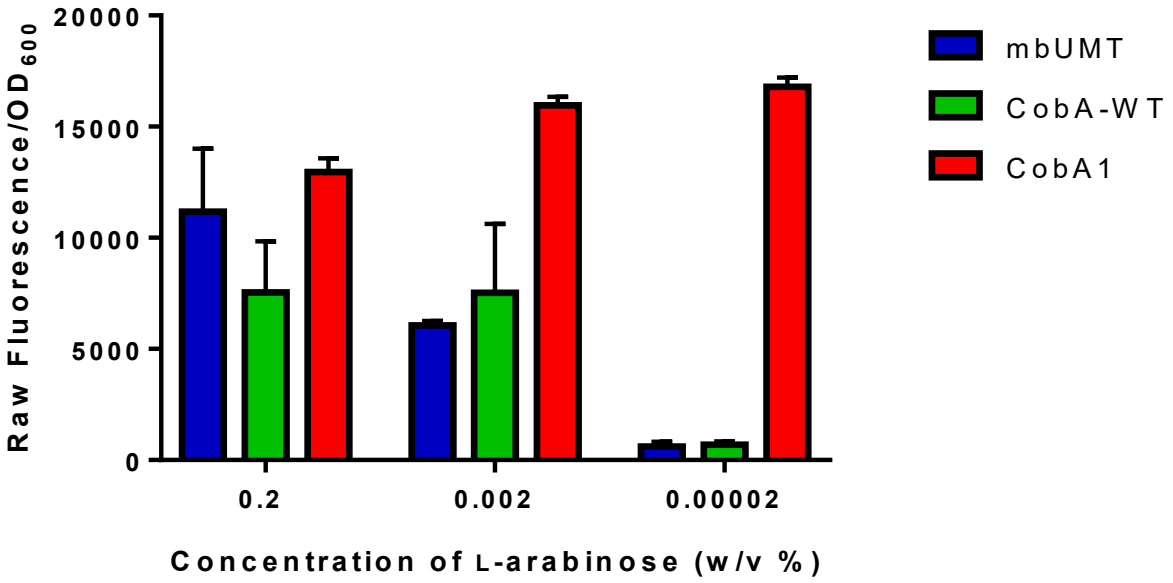


Figure 4.6 Fluorescent emission of SUMT variants at different expression levels

E. coli expressing CobA-WT, CobA1, and mbUMT were cultured overnight, and the fluorescence emission was evaluated by normalizing to the optical density at 600 nm absorption. Cultures expressing CobA1 are significantly brighter than CobA-WT or mbUMT at very low L-arabinose induced expression.

Characterization of the excitation and emission spectra of CobA-WT and CobA1 products yielded interesting results. It is known that the overexpression of CobA and homologous SUMT genes produces a mixture of red fluorescent products (233, 238). However, when comparing our evolved variant to the wild type variant we noticed

substantial shifts in the excitation and emission spectra (Figure 4.7). The excitation maximum of CobA1 is 18 nm red-shifted relative to CobA-WT (360 nm to 378 nm). The excitation spectrum of CobA1 also has a substantial increase in the 600 nm range and a noticeably different spectral profile overall. The emission spectrum also has some substantial differences relative to CobA-WT. Previous characterization of the product overexpression of the CobA gene has suggested that it is a mixture of mono to tetra methylated porphynoid compounds (234). During the directed evolution of CobA1, it appears that the enzyme has begun to favour one or more of these products as evident from the change in emission profile.

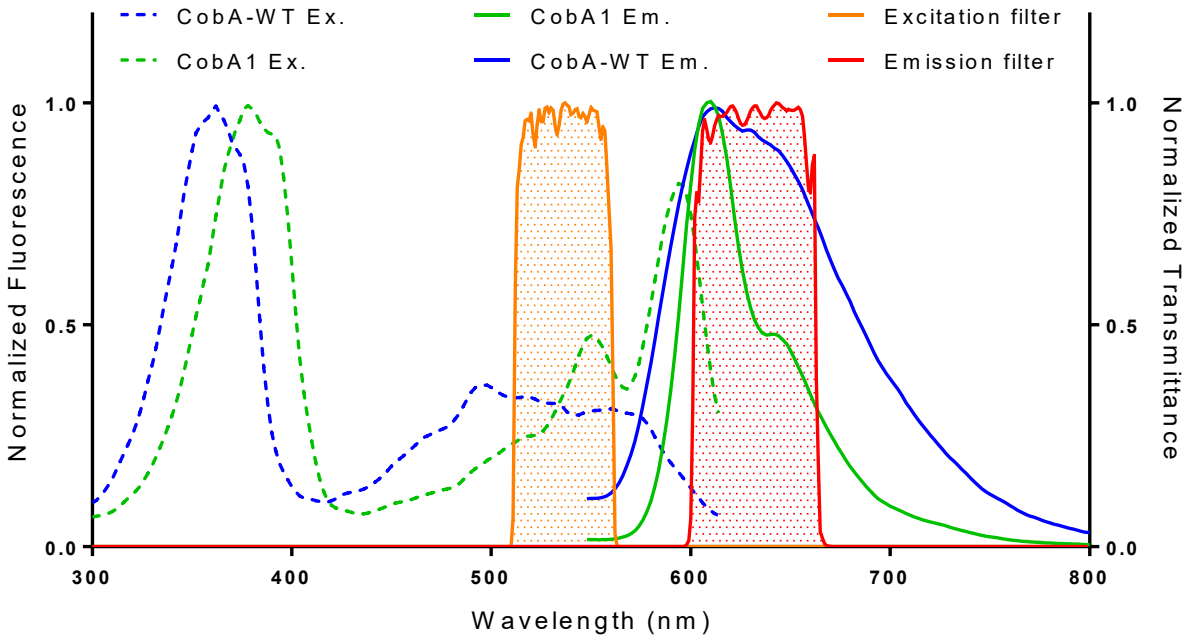


Figure 4.7 Excitation and emission spectrum of CobA-WT and CobA1

Excitation and emission spectrum of CobA-WT and CobA1 from *E. coli* lysate. There is a shift in the fluorescent emission and excitation peaks from CobA-WT to CobA1 that likely reflects differences in the product distribution produced by the enzymes. The shift in peak shape suggests CobA1 is producing more of one fluorescent molecule than CobA-WT. Excitation filter (orange pattern filled) 510-560 nm, emission filter (red pattern filled) 600-665 nm.

The change in fluorescent excitation and emission spectral profile from CobA-WT to CobA1 was unexpected. However, when considering the evolution scheme utilized in the development of CobA1, we can speculate as to why an enzyme that produces one fluorescent product over the other may have been selected for. The colony screening system used in this work uses excitation (510-560 nm) and emission (600-665 nm)

bandpass filters. Observation of the spectral profile suggests that fluorescent molecules with a higher absorption of light closer to the red emission should appear brighter. Furthermore, we selected variants based on the fluorescent intensity of the peak at 610 nm using a plate reader which neglects the broader total fluorescent emission from CobA-WT. Therefore, during the screening of CobA, colonies expressing variants with a higher proportion of fluorescent molecules with increased absorption at 535 nm excitation light would be selected for due to higher fluorescent emission. However, the photophysical properties of the porphynoid compounds could also affect the selection process i.e.) CobA variants that favour the creation of brighter fluorescent molecules might also be selected. Whether the change in spectral profile is the result of the particular protocol used for screening, or if this result would have been obtained regardless of the details of the screening, is as yet unknown.

4.2.2 Rational design of split CobA

Structural information for the CobA enzyme from *P. freudenrechii* has not been described, however, detailed structural information for a handful SUMT homologues has been reported (233, 240).. For this work, we used the recently published structure of NirE from *Pseudomonas aeruginosa* (PDB ID 2BYQ) as a model of the CobA enzyme (237). We chose this homologue for our model because the researchers were able to crystalize the structure with both the substrate, uroporphyrinogen III, and *S*-adenosyl-L-homocysteine (SAH; the demethylated cofactor formed from *S*-adenosyl-L-methionine (SAM)) (Figure 4.8) bound to the protein. The structure of NirE and other SUMTs reveals that this class of enzymes has dumbbell-like structure with two lobes connected by a short stretch of alpha-helix. The large open active site cleft sits between the two lobes. Two copies of the

protein associate to form an obligate homodimeric structure with the two subunits of the protein sitting ‘back-to-back’ and an extended hydrophobic interface between them. SAH is buried deep within the active site cleft and the uroporphyrinogen III is positioned higher in the active site cleft. The relatively open active site cleft likely results from the SUMT having to bind both natural substrates uroporphyrinogen III and Precorrin-1 while also binding Precorrin-2 to synthesize trimethylpyrrocorphin. Two highly flexible loops that overhang the active site, not shown due to lack of structural data, are less conserved than the rest of the structure and are purportedly involved in substrate recognition and stabilization (233, 237). The residues involved in substrate binding and putative catalytic residues for hydrogen abstraction have been previously described (237).

Using the crystal structure of the CobA homologue NirE as a guide, we rationally designed constructs in which the CobA1 gene was split into two fragments. We used molecular modelling software to evaluate the CobA1 structure for potential split sites. Rendering the structure by B-factor (an atom-by-atom measure of “temperature” in the crystal structure) facilitated the visualization of highly flexible regions that were selected for evaluation (Figure 4.8 A).

To evaluate split CobA1 designs, we developed a bicistronic expression plasmid based on pBAD/His-B. This *E. coli* expression plasmid, which we have named pBiC, facilitates the expression of two separate gene fragments from one mRNA by utilizing two ribosomal binding sites. pBiC allows for the efficient construction and expression of libraries consisting of split CobA gene fragments. We used this expression plasmid to express several pairs of CobA gene fragments and select for active reconstituted CobA enzyme by observing red fluorescence.

Ten rationally designed constructs were evaluated for fluorescence (split sites indicated with red arrows in Figure 4.8 C). In rationally designing these constructs, we opted to pick highly flexible regions of the protein that were distant from the active site and solvent accessible. Ultimately, we would want to genetically fuse the new termini to other proteins-of-interest to test for protein-protein interactions. Unfortunately, we were unable to visualize fluorescence from any of these ten split CobA designs. We then turned our investigation to a highly flexible region between the α-helices ‘C’ and ‘D’. This region showed a high B-factor despite having a structured beta-sheet (numbered 3 in Figure 4.8 C) within it. We systematically constructed a series of split CobA1 fragments, in which the protein was split at each residue within this section (residues colored cyan in Figure 4.8 C). Upon testing of twelve potential split sites within this region, we found that the combination of fragments sCobA-N (residues 1-60) and sCobA-C (residues 61-257) was visibly fluorescent after overnight incubation (Figure 4.9).

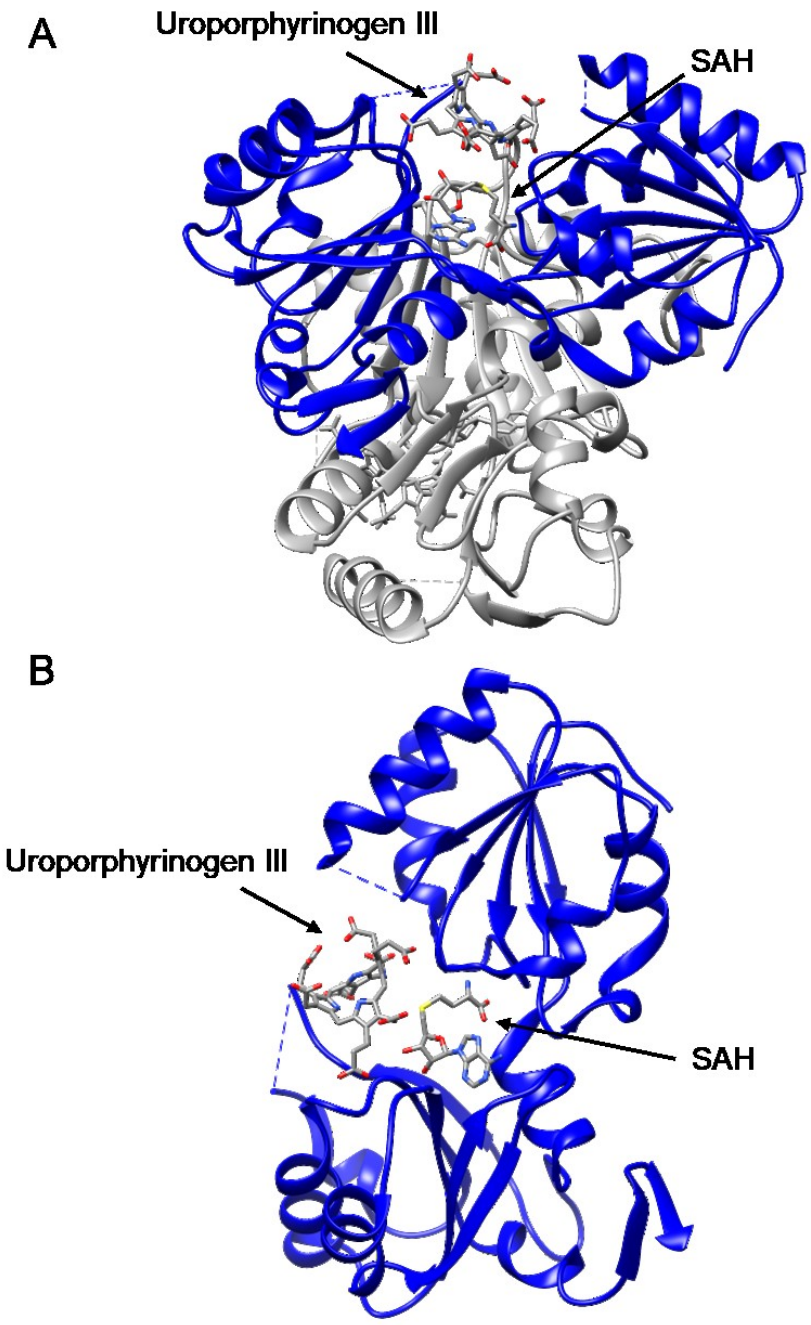


Figure 4.8 Structure of SUMT with substrate and cofactor

We used the crystal structure of NirE, a SUMT from *P. aeruginosa*, to guide the development of a split CobA (PDB ID 2YBQ) (237). (A) Dimer of NirE with Uroporphyrinogen III and SAH bound to the enzyme. (B) Monomer of NirE with SAH and Uroporphyrinogen III shown (PDB ID 2YBQ).

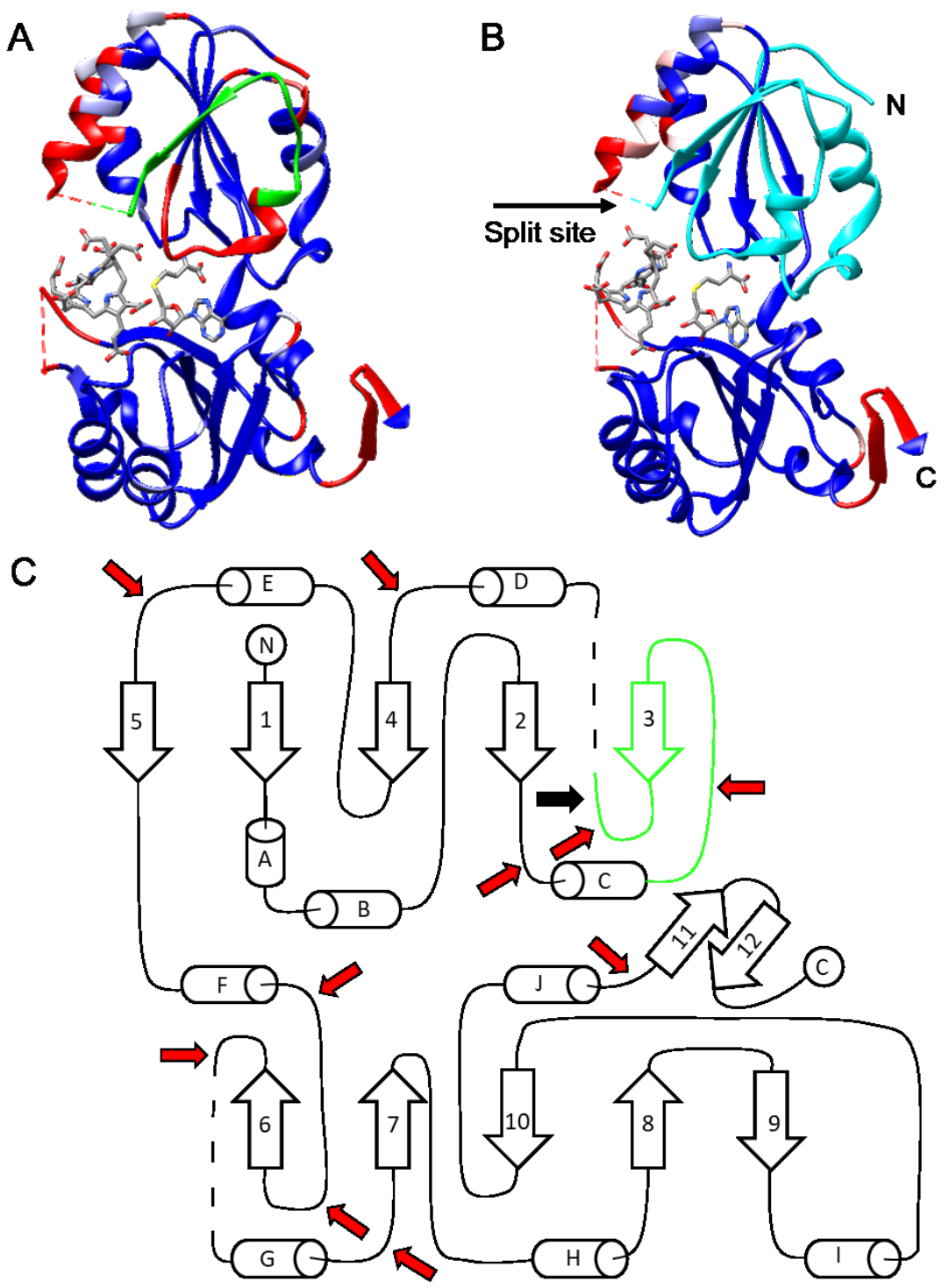


Figure 4.9 Structure guided design of split CobA

We used molecular modelling software to evaluate the NirE crystal structure for potential split sites. (A) A ribbon backbone rendering of the protein structure in which each residue

is color coded according to B-factor. Blue colored regions are more rigid (lower B-factor) and red regions are less rigid (higher B-factor). The contiguous region selected for investigation is colored in green. (B) Successful split site for CobA variant with the N-terminal 60 residues colored in cyan. (C) Schematic representation of topology of NirE with unsuccessful split site interrogations marked with red arrows. The region colored in green is the same as in (A) with a black arrow marking the split site.

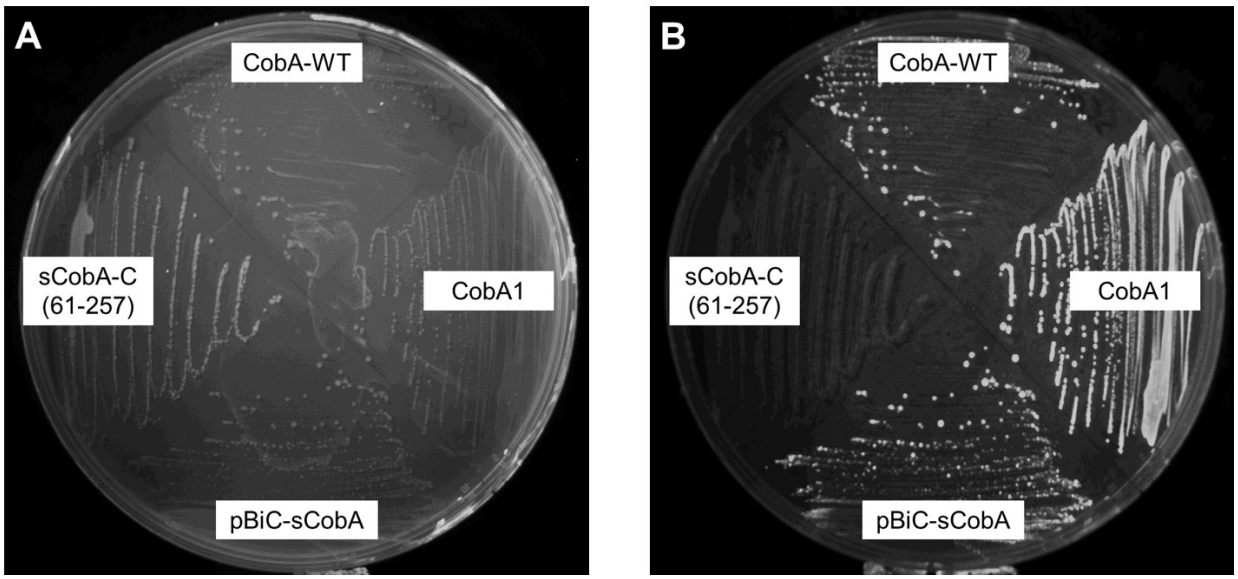


Figure 4.10 Expression of split CobA

Coexpression of split CobA1 (sCobA) constructs under white light (A) and with imaging of red fluorescence (B). pBiC-sCobA1 plasmid contains two ribosomal binding sites for expression of two separate peptides that interact in the cytosol of *E. coli* to produce one CobA1 monomer (sCobA1-N residues 1-60 and sCobA1-C residues 61-257). When sCobA1-C (the large fragment of the split protein) is expressed alone, no fluorescence

was observed. The on-plate brightness of split CobA was lower than CobA1 but similar to CobA-WT.

4.2.3 Design of a split CobA1 protein-protein interaction reporter

With the development of a split CobA1 protein capable of re-associating to restore SUMT activity in *E. coli*, we attempted to use these fragments for the development a protein-protein interaction reporter. For the purposes of development and validation, the well characterized Rapamycin inducible dimerization of FKBP (12 kDa FK506 binding protein) and FRB (FKBP-Rapamycin binding domain) suited our needs for testing split CobA1 (102, 103). FKBP and FRB are soluble monomeric proteins that dimerize in the presence of Rapamycin. This system has been characterized extensively and is routinely used for the validation of protein-protein interaction reporters (53, 64, 68).

To test the effectiveness of our Rapamycin-inducible split CobA1 reporter we developed a single mRNA that codes for two protein fragments (Figure 4.11). The sCobA1-C fragment was fused to the FKBP domain using a short flexible linker. Downstream of the sCobA1-C-linker-FKBP fusion is ribosomal cleavage peptide known as P2A (241). The P2A peptide is well characterized as a sequence capable of causing protein cleavage. However, this moniker is inaccurate since no peptide bond is actually formed and so the protein is not actually cleaved. The P2A sequence is better described as a ribosomal skipping site, which causes the ribosome to fail to join the growing polypeptide chain to the next amino acid, thereby resulting in a break in the polypeptide chain. The result of incorporating a P2A sequence is that two protein fragments are produced from a single mRNA. We used this P2A sequence to separate the two

fragments of our split CobA1 protein-protein interaction reporter. Downstream of the P2A is a cyan fluorescent protein (CFP) for marking successful transformation and expression. The cyan fluorescent protein is fused to sCobA1-N and FRB, with short flexible linkers connecting all of the domains. The sequence encoding our reporter was ligated into pcDNA3.1(+) vector which is suitable for expression in cultured mammalian cell lines.

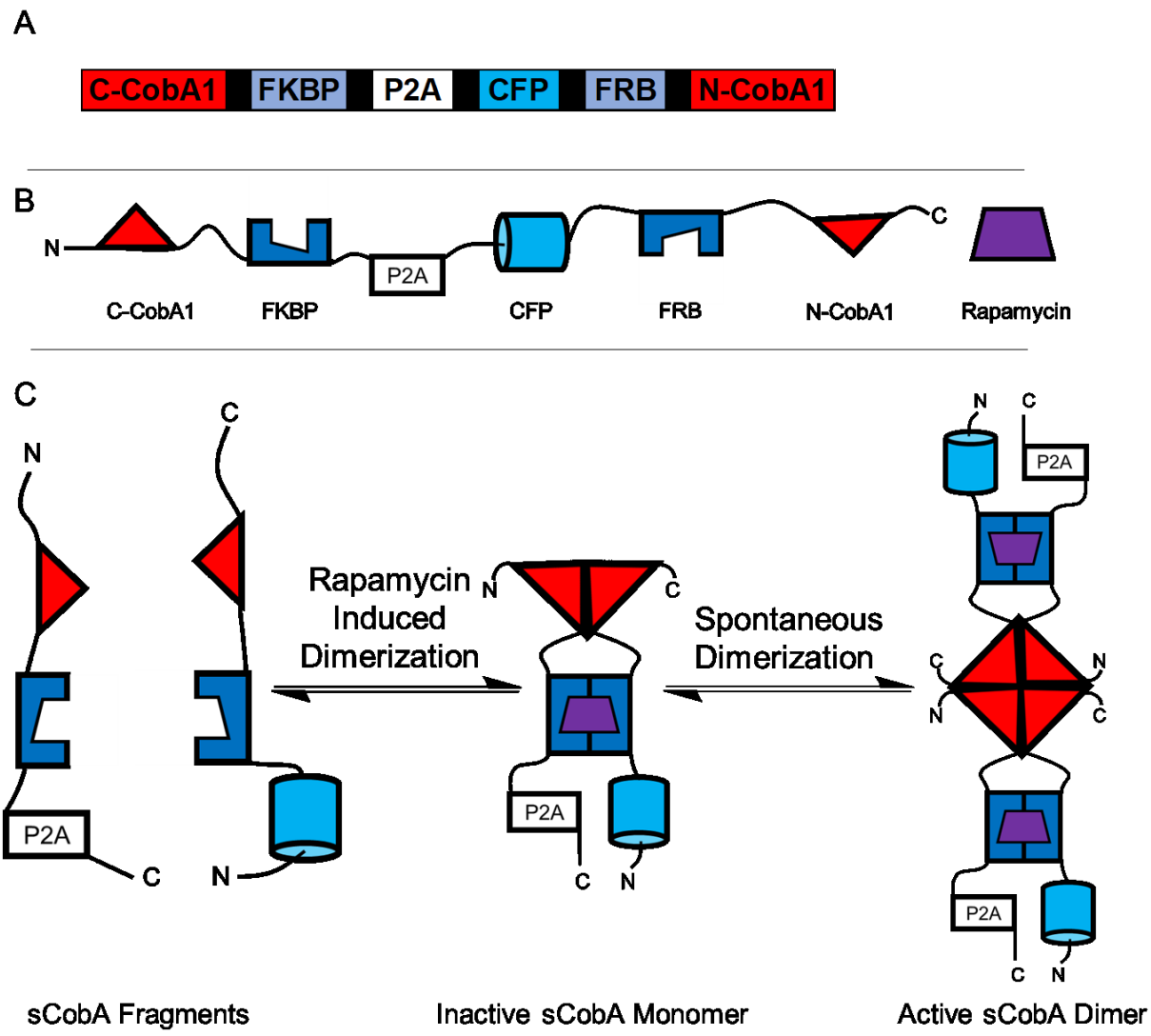


Figure 4.11 Design of split CobA1 protein-protein interaction reporter

The design of split CobA1 protein-protein interaction reporter using FKBP and FRB for Rapamycin-inducible complementation. (A) Representation of the mRNA transcript that codes for split CobA1, linker regions are shaded in black (B) Illustrated representation of the domains in the mRNA transcript that codes for the split CobA1 reporter. CFP is incorporated as an expression marker. (C) The P2A ribosomal cleavage site produces two protein fragments from translation of a single mRNA depicted in (A). Dimerization of FKBP and FRB is Rapamycin-dependent. When Rapamycin is present, the two fragments

come together to form the complete CobA1 monomer. Two CobA1 monomers can then spontaneously dimerize to form an active CobA enzyme.

We transfected HeLa cells with the plasmid encoding the Rapamycin-inducible split CobA1 reporter. Although HeLa cells were expressing the gene, as evident from the fluorescent signal in the CFP channel, we were unable to detect red fluorescence upon incubation with Rapamycin. That is, incubation with cells with Rapamycin (100 nM) for two to three days failed to produce a detectable red fluorescent signal. Cells appeared healthy although they expressed a lower than anticipated amount of CFP. We suspect that the reconstitution of the split CobA1 was optimal under the conditions of high expression (and corresponding high concentration) in *E. coli*, and in the *E. coli* cytosolic environment. This *E. coli* intracellular environment is substantially different than that found in mammalian cell lines. We speculate that these differences contributed to our lack of success in demonstrating the Rapamycin-inducible split CobA1 protein-protein interaction reporter in mammalian cells.

4.2.4 Engineering a split CobA1 intein mediated protein-protein interaction reporter

Undeterred by the failure of split CobA1 to report Rapamycin-inducible dimerization, we searched the relevant literature for methods to further improve the system. We determined that a promising avenue for potentially improving the system was to use protein splicing. Protein splicing is a post-translational autoprocessing event that results in the removal of residues in a protein, called the ‘intein’, to form a new protein with a peptide bond where the intein was excised (a newly formed protein comprised of the ‘extein’ fragments). Most inteins occur as contiguous sequences and are processed

during the translation of the protein (242). Less common are protein that undergo trans-splicing between two separate proteins to excise an intein and form a new protein with a peptide bond joining the two original proteins. However, such trans-splicing proteins have been identified and have found widespread use in chemical biology for their ability to restore the activity of split proteins (243).

One caveat of the split CobA1 is that the split site is in relatively close proximity to the enzyme active site. Accordingly, in the functional protein that results from association of sCobA1-N and sCobA1-C, there are new termini near the active site and the highly flexible substrate recognition loop of the N-terminal lobe. We suspected that this split site was problematic for fully reconstituted tetrameric form of the split CobA1 protein-protein interaction reporter. To alleviate this issue, and to restore activity to split CobA1, we utilized previously characterized split inteins to restore the normal peptide backbone at the split site (Figure 4.12).

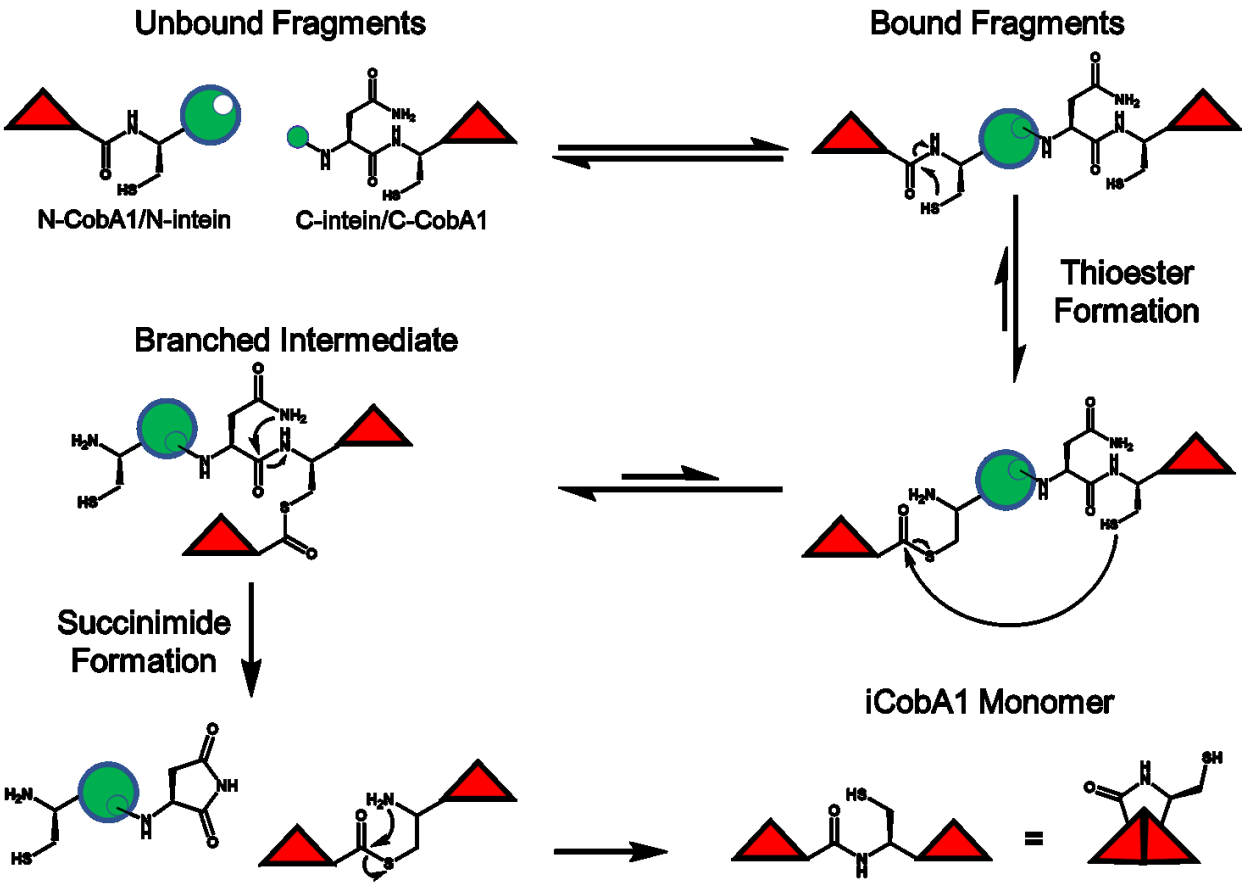


Figure 4.12 Mechanism of split intein formation of iCobA fragments

The mechanism of a protein trans-splicing split intein used to construct iCobA from sCobA fragments. N-CobA1 and C-CobA1 are fused to the split intein peptides. When the intein fragments interact they induce a N to S acyl shift followed by a thioester formation. A branched intermediate decomposes to form a succinimide at the intein termini. The newly formed extein intermediate transitions to a peptide bond through a S to N acyl shift. The product of the reaction is an iCobA monomer with a cysteine residue joining the previously open protein termini.

We used our split CobA1 design to make an intein and Rapamycin-dependent protein-protein interaction reporter we have named iCobA1. With complementary trans-splicing inteins fragments to the amino and carboxylic acid termini of the split CobA1 fragments, Rapamycin-induced dimerization should cause ligation of our split CobA fragments into a fully formed CobA1 monomer with a single cysteine insertion (Figure 4.13). A drawback of this approach is that the formation of intact iCobA monomers would not be reversible. Because the enzyme is an obligate homodimer, once the iCobA monomers are formed the enzyme would dimerize with another spliced iCobA and become active.

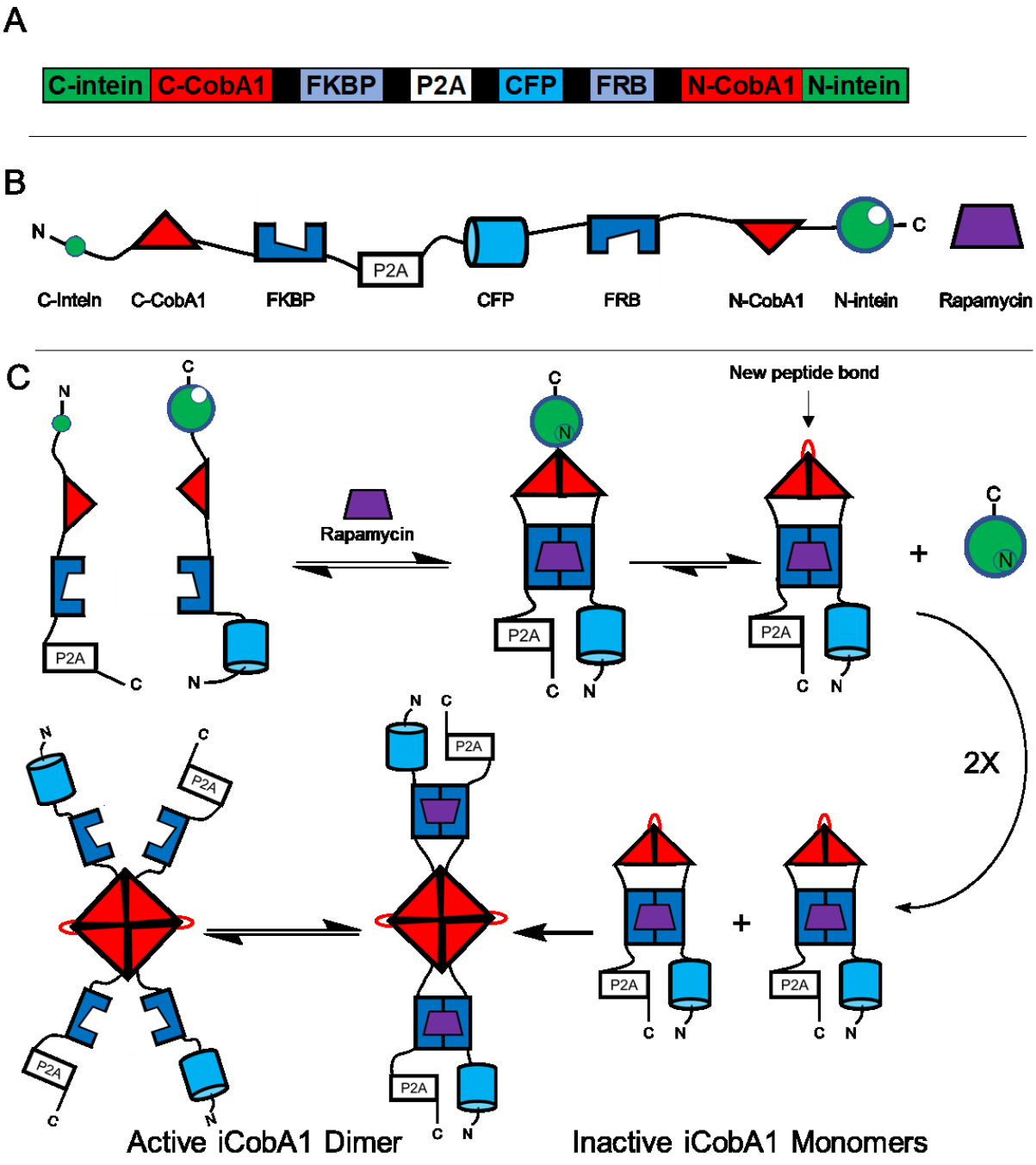


Figure 4.13 Design of a split intein-based split CobA1 protein-protein interaction sensor

(A) Representation of the mRNA transcript that codes for the iCobA sensor (B) Illustration of the mRNA sequence that codes for the iCobA1 protein-protein interaction sensor. (B)

Illustrated representation of the domains in the mRNA transcript that codes for the iCobA1 sensor. The P2A ribosomal cleavage site ensures that two protein fragments are produced from the mRNA. When Rapamycin is present, the FKBP and FRB fragments dimerize and promote interaction of split CobA and split intein. When the split intein is converted to the intein and extein, the split CobA1 fragments are fused with a peptide bond incorporating a single cysteine residue to form an iCobA1 monomer. Two iCobA1 monomers can now interact to form an active iCobA1 dimer. Once formed, the iCobA1 dimer would be active with or without Rapamycin.

4.2.5 Testing iCobA1 protein-protein interaction reporter in cultured cell lines

To evaluate the ability of the iCobA1 reporter to report protein-protein interactions in cultured cell lines, we cloned the gene from iCobA1 into pcDNA3.1(+) plasmid. We then transfected HeLa cells with this plasmid and allowed them to recover for one day. Approximately 24 hours post transfection, we added Rapamycin to a final concentration of 100 nM, or an equivalent volume of dimethylsulphoxide as a control, to individual wells containing transfected HeLa cells. One to two days later we imaged the dishes to determine if iCobA1 could report protein-protein interactions.

Upon imaging of the iCobA transfected HeLa cells, we observed that cells contained red fluorescence puncta. These puncta have been described previously for SUMT expression in cultured mammalian cells (236). The observation of red fluorescent puncta indicated the iCobA reporter was indeed functioning in HeLa cells. However, when we compared the Rapamycin-treated cells to the control cells, we found that both the control cells and treated cells had similar levels of red fluorescent puncta. This suggested to us

that the intein reaction was more favourable than the Rapamycin-induced protein-protein interaction.

It has been previously reported that the identity of the residue at the +2 position of the C-intein can have a dramatic effect on the kinetics of the intein formation reaction (244). We had used a set of previously reported, and highly optimized, inteins for the development of iCobA1. However, during the engineering of this particular set of inteins, Stevens *et al* characterized slower and otherwise less efficient intein pairs (244). These slower and less efficient pairs were developed by mutating the +2 position. We anticipated that modification of this residue in the iCobA design would decrease the intein formation efficiency and restore the Rapamycin-dependent dimerization. Indeed, when we mutated the important +2 residue from a phenylalanine (iCobA1-Phe) to a glycine (iCobA-Gly), which has been shown to reduce the activity of this split intein system (244), we observed an increased number of red fluorescent puncta in the Rapamycin treated HeLa cells.

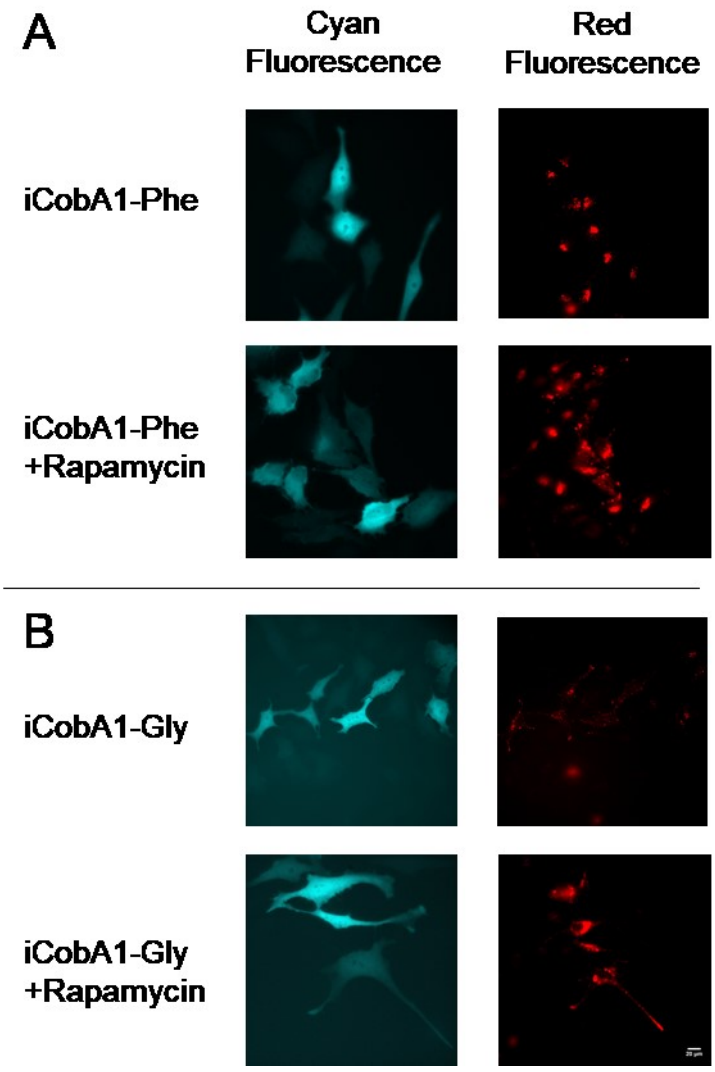


Figure 4.14 Testing of iCobA1 variants in cultured cell lines

Transfected HeLa cells expressing iCobA1 variants. (A) The relatively high intein activity of iCobA1-Phe allows for the formation of the iCobA monomer and results in red fluorescent puncta within the transfected cells. No increase in red fluorescence was observed for iCobA1-Phe treated with Rapamycin. (B) Reducing the intein activity with a Phe->Gly mutation (iCobA1-Phe to iCobA1-Gly) reduced the production of red fluorescence. Addition of Rapamycin restores the red fluorescence of iCobA1-Gly in HeLa cells.

4.5 Conclusion

Through directed evolution and rational structure-guided engineering of the CobA gene we have developed a highly sensitive reporter of gene expression, as well as a fluorescent, intein-modulated, protein-protein interaction reporter. The sensitive gene expression reporter, CobA1, will be useful for observing transcriptional activity of weak promoters that are associated with low abundance proteins. The enzymatic activity ensures that a small amount of expression can yield a large fluorescent signal. Because CobA1 enzyme utilizes an endogenous substrate, it could potentially find applications in a wide range of model organisms. One note of caution about this tool is that high expression of CobA1 can be taxing on cellular metabolism and potentially be detrimental to cell growth.

The development of iCobA1 into a functional protein-protein interaction reporter was a very challenging endeavour. We experimented with a number of designs before arriving at a working prototype. The first prototype was derived from a split CobA1. However, the activity of the split CobA1 was insufficient to report protein-protein interactions in cells. We then modified split CobA1 to be reconstituted with split inteins and developed split iCobA1. Using previously reported mutations, we modified the activity of the inteins in split iCobA1 to enhance the contrast reported by the protein-protein interaction system of Rapamycin-inducible dimerization. We anticipate that iCobA1 could be improved further using directed evolution techniques to optimize the dependence on the interaction of the proteins of interest and increase the contrast between a positive interaction and a negative control.

4.6 Experimental Procedures

4.6.1 General Procedures

Synthetic DNA oligonucleotides and gBlocks were purchased from Integrated DNA Technologies (IDT). Plastic consumables, restriction endonucleases, *Taq* polymerase, Phusion polymerase, T4 DNA ligase, deoxynucleotides, DH10B *E. coli*, pBAD/His-B plasmid, pcDNA3.1(+) plasmid, Bacterial Protein Extraction Reagent (B-PER), Penicillin-Streptomycin, Fetal Bovine Serum (FBS), TurboFect, and GeneJet gel or plasmid purification kits were purchased from Thermo Scientific. Agarose, $MnCl_2 \cdot 4H_2O$, D-glucose, ampicillin, L-arabinose, Hank's balanced salt solution (HBSS), DMEM, TryplE Express, and LB Lennox media were purchased from Fisher Scientific. Ethidium bromide and PCR machines (T100 Thermal Cycler) were purchased from BioRad. Gibson Assembly reagent was purchased from New England Biolabs. QuikChange mutagenesis kits were purchased from Agilent Technologies. 96-well Nunc MicroWell 96-Well Optical-Bottom Plates (cat. 265301) were purchased from Thermo Scientific. Molecular weight cut off filters were purchased from Millipore-Sigma. Sequencing was completed by the Molecular Biology Services Unit at the University of Alberta. The CobA gene in the pISA417 plasmid was a generous gift from the lab of Charles Roessner.

4.6.2 Directed evolution

Following ligation into pBAD/His-B vector, the CobA library was transformed into DH10B *E. coli* bacteria and plated onto 100 μg/mL ampicillin/1.5% agar plates with 0.02% L-arabinose and grown overnight (12-18 hours) at 37 °C. The concentration of L-arabinose was reduced by a factor of ten when the majority of colonies in a mutagenesis library

were brightly fluorescent after overnight incubation. Images acquired using the fluorescent colony screening system were analyzed with Image Pro 6 software (Media Cybernetics) to identify the brightest fluorescent colonies which are then picked and grown in 4 mL of LB Lennox growth medium with 0.02% L-arabinose and 100 μg/mL ampicillin. Culturing overnight (approximately 12-18 hours), yielded sufficient bacterial mass to perform B-PER protein extraction. When the expression of advanced CobA intermediates was detrimental to culture growth, as observed from lack of cell density after overnight incubation, the concentration of L-arabinose was reduced to 0.002% w/v. The red fluorescent porphyrinoid containing lysate was then dispensed into 384 well plates and evaluated on a Tecan Safire^2 for brightness.

4.6.3 PCR and EP-PCR

All PCR reactions were performed according to the manufacture's recommendations. EP-PCR was performed using Taq polymerase with 0.1 mM $MnCl_2$ and a skewed dNTP ratio (0.2 mM ATP, 1 mM CTP, 0.2 mM GTP, 1 mM TTP) resulting in approximately 1 nucleotide mutation per 800 base pairs. Cycling conditions were 35 rounds of 95 °C for 20 seconds, 55 °C for 30 seconds, 72 °C for 60 seconds, and 5 minutes at 72 °C for final extension. PCRs utilizing Phusion polymerase were performed with the supplied buffer with the concentrations of DNA and oligos suggested by the manufacturer. Conditions for PCR were 98 °C for 10 seconds, 63 °C for 20 seconds, 72 °C for 30 seconds, and 72 °C for 60 seconds for final extension.

4.6.4 Construction of pcDNA vectors containing sCobA and iCobA variants

Construction of plasmids for the transfection and expression in cultured HeLa was completed using Gibson Assembly seamless cloning. We amplified the genes required to

make sCobA from our pBAD/His-B vector containing CobA1. Genes for FKBP, FRB, and mCerulean (CFP) were sourced from plasmids available in the Robert Campbell Group plasmid inventory. When all the genetic elements were amplified by PCR and purified, a single Gibson Assembly reaction was used to fuse all the genes together with a linear pcDNA3.1(+) vector. The pcDNA vector containing sCobA construct was sequenced to confirm the correct sequence with no insertions, deletions, or frameshifts.

iCobA was constructed using PCR to amplify the sCobA1 fragment with overlaps matching the N and C intein genes described previously. The N and C inteins were codon optimized for expression in HeLa cells and supplied by IDT as gBlock. Construction was completed with a Gibson Assembly cloning reaction into a linear pcDNA3.1(+) vector. The completed pcDNA plasmid was sequenced to confirm the insert had no insertions, deletions, or frameshifts. We used site-directed mutagenesis with the QuikChange kit from Agilent to convert iCobA1-CFN to iCobA1-CGN.

4.6.5 Expression in HeLa cells for *in vitro* imaging

HeLa cells cultured in DMEM with 10% fetal bovine serum supplemented with penicllin-G potassium salt (50 units/mL) and streptomycin sulphate (50 µg/mL) were plated on collagen coated 35 mm glass bottom dishes (Matsumami) and transfected with 1 µg of plasmid DNA using 2 µL of TurboFect when cells reached roughly 60% confluency. After overnight incubation at 37 °C with 5% CO_2, adherent cells were washed twice with warm Hank's balanced salt solution immediately before imaging.

Cell imaging was performed on a Zeiss 200M wide-field microscope with Semrock or Chroma filters loaded into filter cubes specific for cyan and red fluorescent channels:

Table 4.1 Filters for sCobA1 and iCobA1 imaging

Channel	Excitation Filter (nm)	Emission Filter (nm)
mCerulean (CFP) (Cyan Fluorescence)	D436/20	D480/40
sCobA and iCobA (Red Fluorescence)	HQ535/50x	FF01-585/29

The Zeiss 200M was equipped with an OrcaFlash 4.0 – C13440 (Hamamatsu) camera and images were acquired using MetaMorph 7.8.0.0 software and an MS-2000 automated stage (Applied Scientific Instrumentation). Images were acquired at room temperature through a 40× oiled objective lens (N.A. 1.3). Images were analyzed and pseudo-colored with ImageJ2 (191). UCSF Chimera was used for modelling and representations of CobA (124).

Table 4.2 Oligos and gBlocks used in this work

Primer Name	Sequence (5' to 3')
CobA Fwd	TATATCTCGAGGATGACCACCACACTGTTG
CobA Rv	ATATAAAGCTTTCAGTGGTCGCTGGGC
sCobA V1 Fwd	TATATGAATTCGATATAAGATCTTCACAGGTGGCGATGGGTC
sCobA V1 Rv	ATATAAGATCTTCACAGGTGGCGATGGGTC
sCobA V2 Fwd	TATATGAATTCGGACGAGCGCTCCGAG
sCobA V2 Rv	ATATAAGATCTTCAGCTGGGCGATACCTGC
sCobA V3 Fwd	TATATGAATTCGCGGCTCACGCTGG
sCobA V3 Rv	ATATAAGATCTTCAGTCCTTGGCGAGTTGTC
sCobA V4 Fwd	TATATGAATTCGCCGGTGCGCGTG
sCobA V4 Rv	ATATAAGATCTTCAGATGCCGGCCTCGG
sCobA V5 Fwd	TATATGAATTCGACCCTGGCCGGC
sCobA V5 Rv	ATATAAGATCTTCAACCCACCACCACCAC
sCobA V6 Fwd	TATATGAATTCGGGGTTTACCGTCGTGAC
sCobA V6 Rv	ATATAAGATCTTCACCGCACCAGGTGG
sCobA V7 Fwd	TATATGAATTCGGGCCGTAAGGTGGTGC
sCobA V7 Rv	ATATAAGATCTTCACTCACGGGCGTGCG
sCobA V8 Fwd	TATATGAATTCGATATAAGATCTTCAGGCCTCGGCGCAG
sCobA V8 Rv	ATATAAGATCTTCAGGCCTCGGCGCAG
sCobA V9 Fwd	TATATGAATTCGATATAAGATCTTCACCTGTCGTAGAGGATCACCTC
SCobA V9 Rv	ATATAAGATCTTCACCTGTCGTAGAGGATCACCTC
sCobA V10 Fwd	TATATGAATTCGAACGCCGAACTCGTG
sCobA V10 Rv	ATATAAGATCTTCAGTCGGACGCCTCCGAC
Nterm S49 Rv	ATATAAGATCTTCACAGCAGGTCCTGCGGG
Cterm S49 Fwd	TATATGAATTCGTCGGAGGCGTCCGACAAC
Nterm E50 Rv	ATATAAGATCTTCACGACAGCAGGTCCTGCG
CTerm E50 Fwd	TATATGAATTCGGAGGCGTCCGACAACGC
Nterm A51 Rv	ATATAAGATCTTCACTCCGACAGCAGGTCCTGC
Cterm A51 Fwd	TATATGAATTCGGCGTCCGACAACGCCG
Nterm S52 Rv	ATATAAGATCTTCACGCCTCCGACAGCAGG
Cterm S52 Fwd	TATATGAATTCGTCCGACAACGCCGAACTCG
Nterm N54 Rv	ATATAAGATCTTCAGTCGGACGCCTCCGAC
Cterm N54 Fwd	TATATGAATTCGAACGCCGAACTCGTGCC
Nterm A55 Rv	ATATAAGATCTTCAGTTGTCGGACGCCTCCGAC
Cterm A55 Fwd	TATATGAATTCGGCCGAACTCGTGCCGG
Nterm E56 Rv	ATATAAGATCTTCAGGCGTTGTCGGACGCC
Cterm E56 Fwd	TATATGAATTCGGAACTCGTGCCGGTCGG
Nterm L57 Rv	ATATAAGATCTTCATTCGGCGTTGTCGGACG
Cterm L57 Fwd	TATATGAATTCGCTCGTGCCGGTCGGCAAG
Nterm V58 Rv	ATATAAGATCTTCAGAGTTCGGCGTTGTCGGAC
Cterm V58 Fwd	TATATGAATTCGGTGCCGGTCGGCAAGATC
Nterm P59 Rv	ATATAAGATCTTCACACGAGTTCGGCGTTGTCG
Cterm P59 Fwd	TATATGAATTCGCCGGTCGGCAAGATCCC
Nterm V60 Rv	ATATAAGATCTTCACGGCACGAGTTCGGCG
Cterm V60 Fwd	TATATGAATTCGGTCGGCAAGATCCCGCG
Nterm G61 Rv	ATATAAGATCTTCAGACCGGCACGAGTTCGG
Cterm G61 Fwd	TATATGAATTCGGGCAAGATCCCGCGCG

(continued from Table 4.2)

Primer Name	Sequence (5' to 3')
Split CobA1 Primers	
C-CobA1 Fwd	CACCATGGCCTCGAGGGGCAAGATCCCGCGC
C-CobA1 Rv	CCGCTCCCTCCGCCGCTCCCTCCGCCGTGGTCGCTGGGCGC
FKBP Fwd	CGGAGGGAGCGGCGGAGGGAGCGGAGTGCAGGTGGAAACCATC
FKBP Rv	CAGCAGGCTGAAGTTAGTAGCTCCGCTTCCTTCCAGTTTTAGAAGCTCCACATCGAA
P2A Rv	AGGTCCAGGGTTCTCCTCCACGTCTCCAGCCTGCTTCAGCAGGCTGAAGTTAGTAC
MVSKG P2A Fwd	GAACCCTGGACCTATGGTGAGCAAGGGCGAG
GMDELYK Rv	CTAGAAGCCATTCCTCCGCCGCTACCGCCCTTGTACAGCTCGTCCATGCC
FRB Fwd	GGAGGAATGGCTTCTAGAATCCTCTGGCATGAGATGTG
FRB Rv	CCCGCTTCCTGACCCGGCTCCTGACCCGGCTCCCTTTGAGATTCGTCGGAACAC
N-CobA1 Fwd	CAGGAGCCGGGTCAGGAAGCGGGGGCACGGGCACAATGACCACCACACTGTTGC
N-CobA1 Rv	GGGTTTACCAAGCTTTCAGACCGGCACGAGTTCGG
iCobA1 Primers	
pcDNA Rv	GCGTGCCGCCACCATG
pcDNA Fwd	AAGCTTGGTAAACCCGCTG
sCobA Fwd	GGCAAGATCCCGCGC
sCobA Rv	GACCGGCACGAGTTCGG
N-Intein Fwd	CGAACTCGTGCCGGTCTGCCTCTCTTACGATACGGAAATATTG
N-Intein pcDNA Rv	GTTTACCAAGCTTTCAAGGAAGGCCATCAACCTGC
C-Intein pcDNA Fwd	GTGCCGCCACCATGGGTGTCAAGATCATCAGTCGCAAATC
C-Intein Rv	GCGGGATCTTGCCGTTGAAGCAGTTAGAAGCCACC
iCob CGN QC	CTTAAGAATGGTTTGGTGGCTTCTAACTGCGGGAACGGCAAGATCCCGCGC
iCob CEN QC	CTTAAGAATGGTTTGGTGGCTTCTAACTGCGAAAACGGCAAGATCCCGCGC
IDT gBlocks	
C-Intein	GGTGTCAAGATCATCAGTCGCAAATCTCTGGGCACGCAGAATGTATACGACATCGGGGTCGAGAAA GATCACAACTTCCTTCTTAAGAATGGTTTGGTGGCTTCTAACTGCTTCAAC
N-Intein	TGCCTCTCTTACGATACGGAAATATTGACAGTTGAATACGGATTTCTTCCAATAGGGAAGATAGTGG AAGAGCGGATCGAGTGTACTGTCTATACGGTGGATAAGAACGGTTTCGTTTATACTCAACCCATAGC ACAGTGGCATAATCGCGGAGAACAGGAGGTCTTCGAATACTGTTTGGAAGACGGCTCCATAATACG CGCGACAAAGGATCACAAGTTTATGACCACGGATGGGCAAATGCTCCCAATAGACGAAATTTTTGA GCGCGGTCTTGATCTGAAGCAGGTTGATGGCCTTCCT

Chapter 5: Conclusion, future directions and perspective

5.1 Conclusions and future directions

Cellular biology has entered a new era – an era illuminated by tools for the optical investigation of cellular phenomena. Of these new optical tools, the most promising are the genetically encodable sensors. The ability to genetically target optical sensors to organelle or specific cells types has provided researchers a window into the biological world that no other technology has provided. The efforts to understand biology and model disease will continue to require new and improved reporters.

The NeonCyan series of fluorescent proteins are a new set of reporters with interesting optical and physical attributes. In this work I have demonstrated their usefulness for labeling of subcellular structures in HeLa cells. With the help of collaborators Antoine Royant and Damien Clavel, we have produced a molecular model of a NeonCyan variant from X-ray crystallization efforts. I then used this structural information to probe critical residues of NeonCyan leading to the serendipitous discovery of new spectral properties of this protein. NeonCyan is still far from being the brightest fluorescent protein with a tryptophan chromophore (245). Directed evolution efforts guided by the structural data presented in this thesis should aid in the development of brighter variants with improved physical properties. The green shifted variant, NeonCyan-Thr207Asp is only the second example of a tryptophan chromophore in an anionic state. This interesting behaviour might be suitable for the development of a ratiometric NeonCyan Ca^{2+} sensor. Furthermore, the residue involved in modulating the spectral properties of the protein (Thr207) should be investigated as an alternative insertion site for the construction of sensors based on single fluorescent proteins.

In Chapter 3 I detailed the development of a green fluorescent Ca^{2+} sensor. This sensor is based on the scaffold of the brightest green fluorescent protein reported to date, mNeonGreen. The resulting sensors, mNG-GECO1 and its variants, were demonstrated to be capable of reporting Ca^{2+} histamine induced Ca^{2+} oscillations in cultured HeLa cells. mNG-GECO1 has a low sensitivity to Ca^{2+} but is nearly twice as bright as the brightest engineered GCaMP series sensor. Our latest variant, mNG-GECO1a, is substantially improved in sensitivity but this improvement came at the expense of a higher K_d. Taken together, the mNG-GECO series of Ca^{2+} sensors will provide an excellent starting point for future directed evolution efforts for the development of brighter sensors.

In Chapter 4 I described efforts to develop a protein-protein interaction reporter that utilizes a ubiquitous endogenous substrate, Uroporphyrinogen III. Starting with the wild type CobA gene from *Pseudomonas freudenreichii*, I used directed evolution to improve the activity of the CobA enzyme. Using this improved CobA, CobA1, I split CobA into two polypeptide fragments that could reconstitute the active enzyme in *E. coli*. However, this split CobA variant does not produce a fluorescent signal in cultured HeLa cells. By fusing a complementary split intein pair to each portion of the split CobA1, I was able to rescue the enzymatic activity. Modulation of the split intein activity enhanced the contrast of Rapamycin-induced dimerization in HeLa cells. This sensor is the first split *S*-adenosyl-L-methionine uroporphyrinogen III methyltransferase (SUMT) protein-protein interaction reporter. Although the contrast is limited, this prototype could be further improved using directed evolution. iCobA1 should be useful across a wide range of cell types due to the ubiquitous presence of the Uroporphyrinogen III substrate.

5.2 Perspective

Genetically encodable optical sensors will continue to be at the forefront of biology research for many years to come. As the use of these powerful tools to study biology becomes routine, the diversity of metabolic targets and protein-protein interactions will increase. To facilitate the development of new sensors, a modular design framework has been presented here which should serve as a starting point for novel sensor design. This framework, combined with the recent advances in molecular biology techniques such as Gibson Assembly, mutagenic polymerases, and EvolvR, will facilitate fast and efficient sensor development.

Directed evolution has been recognized with the 2018 Nobel Prize in chemistry and the power of this technology cannot be understated for the improvement of genetically encodable sensors. To harness the full power of directed evolution, the ability to screen very large libraries of sensors in a high throughput manner remains a top priority. Sensors developed in this work relied heavily on 96-well format screening which allows for testing 1000's of potential variants for activity. However, technologies for screening millions of cells, such as microfluidics or FACS, are becoming more readily available. One critical aspect of directed evolution is the matrix effect of the cell type being used to screen variants. This has previously led to less than optimal performance of sensors when they have been transitioned to *in vivo* applications. Recently, screening directly in cultured mammalian cell lines has been demonstrated to be effective, but remains practical for only several specialized research labs. As robotics become more available to the average researcher, it is easy to imagine that medium throughput screening systems (such as 96-well format) aided by robots will become incredibly powerful. The best equipped labs may

be able to combine robotics and FACS: very high throughput FACS screening for brightness followed by a robotic (or semi-automated) high throughput secondary screening.

Researchers interested in continuing to improve the technologies presented here should start with focus their efforts on increasing the brightness of NeonCyan and its variants. NeonCyan may also serve as a template for the development of blue fluorescent proteins by genetic manipulation of the tryptophan residue, possibly to phenylalanine or histidine. mNG-GEOC1a has recently shown promising activity when applied for Ca^{2+} imaging in cultured cells. This variant would benefit from lowering the K_d to a level appropriate for visualizing Ca^{2+} transients in neurons. The K_d of mNG-GECO1a could also be increased for studying Ca^{2+} transients in the endoplasmic reticulum or mitochondria, two organelles that have been somewhat neglected by sensor engineers to date. iCobA is an excellent prototype for tracking protein-protein interactions in cells and the fact that it uses a ubiquitous substrate makes it highly advantageous relative to other technologies. Further development of iCobA should focus on the development of an high throughput on-plate screen for complementation.

Throughout my Ph.D. I have been able to master the techniques of molecular biology and apply them to the development of genetically encodable sensors. The knowledge I have gained from the development of these skills, from reverse engineering reagents to the development of my own mutagenic polymerase, have helped me to evolve from a naïve but eager undergraduate student into a well-rounded and informed scientist. I am now an accomplished protein engineer with a strong background in the leading edge of modern biology research.

Technical Note: Purification and use of aTaq mutagenic polymerase

Abstract: Error prone polymerase chain reaction (EP-PCR) is an indispensable tool for creating libraries of random mutations in a gene. The typical EP-PCR reaction utilizes the thermostable polymerase from *Thermus aquaticus*, known as *Taq* polymerase. *Taq* has a known bias for A->T mutations in the presence of Mn^{2+} and skewed dNTP ratio. Recently, a new polymerase for developing EP-PCR libraries has become preferred over *Taq* because it has less mutational bias, does not require skewed dNTP's or the addition of Mn^{2+}, and generates blunt end PCR products suitable for Gibson Assembly cloning. This polymerase is based on the widely used thermostable polymerase from *Pyrococcus furiosus* known as Pfu polymerase. aTaq (after Taq, pronounced 'attack') mutagenic polymerase has two mutations relative to wild-type Pfu polymerase: D215A (exonuclease domain knockout) and D473G (loss of mismatch fidelity) (246). aTaq also contains the sso7d DNA binding domain fused to the polymerase for increased speed and processivity (247).

Controlling the rate of mutation: The mutational rate of aTaq is controlled by the amount of template in the reaction and the number of thermal cycles. Low mutation rates, 1-5 nucleotide mutations per kilobase (kb), result from high DNA concentrations in the initial reaction (200-500 ng per 50 µL reaction). High mutation rates, 5-15 nucleotide mutations per kb, result from low concentrations of DNA in the initial reaction (10-100 ng per 50 µL reaction). These numbers are only a starting point and should serve as a guide to adjusting the mutation rate as needed.

Recommended reaction conditions:

Reagent *add to PCR tube in the following order	Volume (µL)
Water	40
10× Buffer	5
dNTP's (10 mM stock)	1
Primers (100 µM stock)	0.5 / 0.5
Template DNA	1-2
aTaq	1

Mutation Rate	DNA (ng/µL)	Cycles
Low	5	25
Medium	3	30
High	1	35

Mix by slowly pipetting up and down, transfer to a thermal cycler.

Amplification of target library:

Step	Temperature (°C)	Time
1	95	2 minutes
2	95	20s
3	55	20s
4	72	1 minute per kb
5	Cycle steps 2-4 30×	
6	72	5 minutes

10× PCR Buffer: 200 mM Tris-HCl (pH 8.5), 100 mM KCl, 100 mM $(NH_4)_2SO_4$, 20 mM $MgSO_4$, 1.0% Triton X-100, 1 mg/mL BSA.

For convenience, aTaq is active in commercially available Pfu buffer.

Purification of aTaq polymerase:

This modified protocol makes enough polymerase for several thousand 50 µL reactions (147).

Day 1 - Streak out a fresh plate of BL21 (DE3) *E. coli.* harboring the aTaq plasmid on kanamycin selection plates. Prepare a 50 mL flask of LB.

Day 2 - Pick a single colony for a 4 mL liquid subculture, incubate overnight at 37 °C.

Day 3 - Subculture 50:1 and grow cells to an OD_{600} of 1.0, then induce with 0.5 mM IPTG. Culture for 3 hours at 30 °C and collect the cells in a 50 mL centrifuge bottle.

- Resuspend the cells in 12 mL of cold TBS in the 50 mL centrifuge tube and transfer to an icebox. Sonicate the cells for 3 minutes using the microtip (amplitude 50, 1 s on, 2 s off).
- Spin the cells down at max RPM for 5 minutes to collect cell debris.
- Transfer the supernatant into 2 mL tubes and incubate at 65-70 °C for 10 minutes
- Centrifuge for 15 minutes at max RPM to remove precipitate
- **the polymerase can be used directly from this step at a 50:1 dilution*
- NTA purify the protein, concentrate the protein with a 30 kDa molecular weight cut off filter
- Buffer exchange into storage buffer, store at -20 °C
- Adjust the activity of the purified protein by diluting with storage buffer, this protocol makes approximately 2 mL of 1 U/µL

Storage Buffer: TBS with 0.2 mM EDTA, 2 mM DTT, 0.2% Tween 20, 50% glycerol.

TBS Buffer: 50 mM Tris-HCl, 150 mM NaCl, pH 8.0

Troubleshooting

1. Low PCR yield: add 1-4% DMSO to the reaction.
2. Smeared PCR product: use an annealing temperature 0-3 °C higher than the Tm of your primers and increase the denaturation temperature to 98 °C.
3. Mutation rate is too low or too high: adjust initial template DNA concentration. Increase number of amplification cycles to increase PCR yield. However, increasing the number of cycles will result in more mutations.

Bibliography

1. Kendall JM, Badminton MN. Aequorea victoria bioluminescence moves into an exciting new era. Trends in Biotechnology. 1998;16(5):216-224.

2. Loening AM, Fenn TD, Gambhir SS. Crystal structures of the luciferase and green fluorescent protein from Renilla reniformis. J Mol Biol. 2007;374(4):1017-1028.

3. Xu Y, Kanauchi A, von Arnim AG, Piston DW, Johnson CH. [12] Bioluminescence resonance energy transfer: Monitoring protein-protein interactions in living cells. Methods Enzymol. 360: Academic Press; 2003. p. 289-301.

4. Shaner NC, Campbell RE, Steinbach PA, Giepmans BN, Palmer AE, Tsien RY. Improved monomeric red, orange and yellow fluorescent proteins derived from Discosoma sp. red fluorescent protein. Nat Biotechnol. 2004;22(12):1567-1572.

5. Ai HW, Shaner NC, Cheng Z, Tsien RY, Campbell RE. Exploration of new chromophore structures leads to the identification of improved blue fluorescent proteins. Biochemistry. 2007;46(20):5904-5910.

6. Labas YA, Gurskaya NG, Yanushevich YG, Fradkov AF, Lukyanov KA, Lukyanov SA, et al. Diversity and evolution of the green fluorescent protein family. Proc Natl Acad Sci U S A. 2002;99(7):4256-4261.

7. Shimomura O. Structure of the chromophore of Aequorea green fluorescent protein. FEBS Lett. 1979;104(2):220-222.

8. Prasher DC, Eckenrode VK, Ward WW, Prendergast FG, Cormier MJ. Primary structure of the Aequorea victoria green-fluorescent protein. Gene. 1992;111(2):229-233.

9. Chalfie M, Tu Y, Euskirchen G, Ward W, Prasher D. Green fluorescent protein as a marker for gene expression. Science. 1994;263(5148):802-805.

10. Heim R, Tsien RY. Engineering green fluorescent protein for improved brightness, longer wavelengths and fluorescence resonance energy transfer. Curr Biol. 1996;6(2):178-182.

11. Cubitt AB, Heim R, Adams SR, Boyd AE, Gross LA, Tsien RY. Understanding, improving and using green fluorescent proteins. Trends in Biochemical Sciences. 1995;20(11):448-455.

12. Miyawaki A, Llopis J, Heim R, McCaffery JM, Adams JA, Ikura M, et al. Fluorescent indicators for Ca2+ based on green fluorescent proteins and calmodulin. Nature. 1997;388(6645):882-887.

13. Llopis J, McCaffery JM, Miyawaki A, Farquhar MG, Tsien RY. Measurement of cytosolic, mitochondrial, and Golgi pH in single living cells with green fluorescent proteins. Proc Natl Acad Sci U S A. 1998;95(12):6803-6308.

14. Miyawaki A, Griesbeck O, Heim R, Tsien RY. Dynamic and quantitative Ca2+ measurements using improved cameleons. Proc Natl Acad Sci U S A. 1999;96(5):2135-2140.

15. Reid BG, Flynn GC. Chromophore formation in green fluorescent protein. Biochemistry. 1997;36(22):6786-6791.

16. Heim R, Prasher DC, Tsien RY. Wavelength mutations and posttranslational autoxidation of green fluorescent protein. Proc Natl Acad Sci U S A. 1994;91(26):12501-12504.

17. Yue JX, Holland ND, Holland LZ, Deheyn DD. The evolution of genes encoding for green fluorescent proteins: insights from cephalochordates (amphioxus). Sci Rep. 2016;6:1-12.

18. Hunt ME, Scherrer MP, Ferrari FD, Matz MV. Very bright green fluorescent proteins from the Pontellid copepod Pontella mimocerami. PLoS One. 2010;5(7):1-8.

19. Baumann D, Cook M, Ma L, Mushegian A, Sanders E, Schwartz J, et al. A family of GFP-like proteins with different spectral properties in lancelet Branchiostoma floridae. Biol Direct. 2008;3:1-12.

20. Matz MV, Fradkov AF, Labas YA, Savitsky AP, Zaraisky AG, Markelov ML, et al. Fluorescent proteins from nonbioluminescent Anthozoa species. Nat Biotechnol. 1999;17(10):969-973.

21. Niwa H, Inouye S, Hirano T, Matsuno T, Kojima S, Kubota M, et al. Chemical nature of the light emitter of the Aequorea green fluorescent protein. Proc Natl Acad Sci U S A. 1996;93(24):13617-13622.

22. Barondeau DP, Putnam CD, Kassmann CJ, Tainer JA, Getzoff ED. Mechanism and energetics of green fluorescent protein chromophore synthesis revealed by trapped intermediate structures. Proc Natl Acad Sci U S A. 2003;100(21):12111-12116.

23. Remington SJ. Fluorescent proteins: maturation, photochemistry and photophysics. Curr Opin Struct Biol. 2006;16(6):714-721.

24. Grigorenko BL, Krylov AI, Nemukhin AV. Molecular Modeling Clarifies the Mechanism of Chromophore Maturation in the Green Fluorescent Protein. J Am Chem Soc. 2017;139(30):10239-10249.

25. https://www.janelia.org/project-team/genie.

26. Dana H, Sun Y, Mohar B, Hulse B, Hasseman J, Tsegaye G, et al. High-performance GFP-based calcium indicators for imaging activity in neuronal populations and microcompartments. bioRxiv. 2018:1-40.

27. Jacques SL. Optical properties of biological tissues: a review. Phys Med Biol. 2013;58(11):37-61.

28. Shcherbakova DM, Stepanenko OV, Turoverov KK, Verkhusha VV. Near-Infrared Fluorescent Proteins: Multiplexing and Optogenetics across Scales. Trends Biotechnol. 2018;36(12):1230-1243.

29. Stosiek C, Garaschuk O, Holthoff K, Konnerth A. In vivo two-photon calcium imaging of neuronal networks. Proc Natl Acad Sci U S A. 2003;100(12):7319-7324.

30. Sadakane O, Masamizu Y, Watakabe A, Terada S, Ohtsuka M, Takaji M, et al. Long-Term Two-Photon Calcium Imaging of Neuronal Populations with Subcellular Resolution in Adult Non-human Primates. Cell Rep. 2015;13(9):1989-1999.

31. Svoboda K, Yasuda R. Principles of two-photon excitation microscopy and its applications to neuroscience. Neuron. 2006;50(6):823-839.

32. Livet J, Weissman TA, Kang H, Draft RW, Lu J, Bennis RA, et al. Transgenic strategies for combinatorial expression of fluorescent proteins in the nervous system. Nature. 2007;450(7166):56-62.

33. Kim TH, Zhang Y, Lecoq J, Jung JC, Li J, Zeng H, et al. Long-Term Optical Access to an Estimated One Million Neurons in the Live Mouse Cortex. Cell Rep. 2016;17(12):3385-3394.

34. Iwano S, Sugiyama M, Hama H, Watakabe A, Hasegawa N, Kuchimaru T, et al. Single-cell bioluminescence imaging of deep tissue in freely moving animals. Science. 2018;359(6378):935-939.

35. Hall MP, Woodroofe CC, Wood MG, Que I, Van't Root M, Ridwan Y, et al. Click beetle luciferase mutant and near infrared naphthyl-luciferins for improved bioluminescence imaging. Nat Commun. 2018;9(1):1-12.

36. Evans MS, Chaurette JP, Adams ST, Jr., Reddy GR, Paley MA, Aronin N, et al. A synthetic luciferin improves bioluminescence imaging in live mice. Nat Methods. 2014;11(4):393-395.

37. Jones KA, Porterfield WB, Rathbun CM, McCutcheon DC, Paley MA, Prescher JA. Orthogonal Luciferase-Luciferin Pairs for Bioluminescence Imaging. J Am Chem Soc. 2017;139(6):2351-2358.

38. Greenwald EC, Mehta S, Zhang J. Genetically Encoded Fluorescent Biosensors Illuminate the Spatiotemporal Regulation of Signaling Networks. Chem Rev. 2018:1-88.

39. Tanahashi H, Ito T, Inouye S, Tsuji FI, Sakaki Y. Photoprotein aequorin: use as a reporter enzyme in studying gene expression in mammalian cells. Gene. 1990;96(2):249-255.

40. Close DM, Hahn RE, Patterson SS, Baek SJ, Ripp SA, Sayler GS. Comparison of human optimized bacterial luciferase, firefly luciferase, and green fluorescent protein for continuous imaging of cell culture and animal models. J Biomed Opt. 2011;16(4):1-11.

41. Wang Q, Shui B, Kotlikoff MI, Sondermann H. Structural basis for calcium sensing by GCaMP2. Structure. 2008;16(12):1817-1827.

42. Ding J, Luo AF, Hu L, Wang D, Shao F. Structural basis of the ultrasensitive calcium indicator GCaMP6. Sci China Life Sci. 2014;57(3):269-274.

43. Acker MG, Auld DS. Considerations for the design and reporting of enzyme assays in high-throughput screening applications. Perspectives in Science. 2014;1(1-6):56-73.

44. Kneen M, Farinas J, Li Y, Verkman AS. Green Fluorescent Protein as a Noninvasive Intracellular pH Indicator. Biophys J. 1998;74(3):1591-1599.

45. Dale R, Kato N. Truly quantitative analysis of the firefly luciferase complementation assay. Curr Plant Biol. 2016;5:57-64.

46. Dixon AS, Schwinn MK, Hall MP, Zimmerman K, Otto P, Lubben TH, et al. NanoLuc Complementation Reporter Optimized for Accurate Measurement of Protein Interactions in Cells. ACS Chem Biol. 2016;11(2):400-408.

47. Inoue M, Takeuchi A, Horigane S, Ohkura M, Gengyo-Ando K, Fujii H, et al. Rational design of a high-affinity, fast, red calcium indicator R-CaMP2. Nat Methods. 2015;12(1):64-70.

48. https://news.harvard.edu/gazette/story/2007/04/frank-h-westheimer-major-figure-in-20th-century-chemistry-dies-at-95/.

49. Henderson MJ, Baldwin HA, Werley CA, Boccardo S, Whitaker LR, Yan X, et al. A Low Affinity GCaMP3 Variant (GCaMPer) for Imaging the Endoplasmic Reticulum Calcium Store. PLoS One. 2015;10(10):1-17.

50. Greotti E, Wong A, Pozzan T, Pendin D, Pizzo P. Characterization of the ER-Targeted Low Affinity Ca(2+) Probe D4ER. Sensors (Basel). 2016;16(9):1-13.

51. Wu J, Prole DL, Shen Y, Lin Z, Gnanasekaran A, Liu Y, et al. Red fluorescent genetically encoded Ca2+ indicators for use in mitochondria and endoplasmic reticulum. Biochem J. 2014;464(1):13-22.

52. Clare ER, Carl WB, III, Igor LM, James BD. Intracellular FRET-based probes: a review. Methods and Applications in Fluorescence. 2015;3(4):1-45.

53. Ohmuro-Matsuyama Y, Chung CI, Ueda H. Demonstration of protein-fragment complementation assay using purified firefly luciferase fragments. BMC Biotechnol. 2013;13:1-9.

54. Ding Y, Li J, Enterina JR, Shen Y, Zhang I, Tewson PH, et al. Ratiometric biosensors based on dimerization-dependent fluorescent protein exchange. Nat Methods. 2015;12(3):195-198.

55. Qian Y, Rancic V, Wu J, Ballanyi K, Campbell RE. A Bioluminescent Ca(2+) Indicator Based on a Topological Variant of GCaMP6s. Chembiochem. 2018;0(0):1-6.

56. Hall MP, Unch J, Binkowski BF, Valley MP, Butler BL, Wood MG, et al. Engineered luciferase reporter from a deep sea shrimp utilizing a novel imidazopyrazinone substrate. ACS Chem Biol. 2012;7(11):1848-1857.

57. Chen TW, Wardill TJ, Sun Y, Pulver SR, Renninger SL, Baohan A, et al. Ultrasensitive fluorescent proteins for imaging neuronal activity. Nature. 2013;499(7458):295-300.

58. Spiering D, Bravo-Cordero JJ, Moshfegh Y, Miskolci V, Hodgson L. Chapter 25 - Quantitative Ratiometric Imaging of FRET-Biosensors in Living Cells. In: Sluder G, Wolf DE, editors. Methods Cell Biol. 114: Academic Press; 2013. p. 593-609.

59. Nakai J, Ohkura M, Imoto K. A high signal-to-noise Ca(2+) probe composed of a single green fluorescent protein. Nat Biotechnol. 2001;19(2):137-141.

60. Shen Y, Wu S-Y, Rancic V, Qian Y, Miyashita S-I, Ballanyi K, et al. Genetically encoded ratiometric indicators for potassium ion. bioRxiv. 2018:1-31.

61. Koker T, Fernandez A, Pinaud F. Characterization of Split Fluorescent Protein Variants and Quantitative Analyses of Their Self-Assembly Process. Sci Rep. 2018;8(1):1-15.

62. Feng S, Sekine S, Pessino V, Li H, Leonetti MD, Huang B. Improved split fluorescent proteins for endogenous protein labeling. Nat Commun. 2017;8(1):1-11.

63. Kerppola TK. Bimolecular fluorescence complementation (BiFC) analysis as a probe of protein interactions in living cells. Annu Rev Biophys. 2008;37:465-487.

64. To TL, Zhang Q, Shu X. Structure-guided design of a reversible fluorogenic reporter of protein-protein interactions. Protein Sci. 2016;25(3):748-753.

65. Lund CH, Bromley JR, Stenbaek A, Rasmussen RE, Scheller HV, Sakuragi Y. A reversible Renilla luciferase protein complementation assay for rapid identification of protein-protein interactions reveals the existence of an interaction network involved in xyloglucan biosynthesis in the plant Golgi apparatus. J Exp Bot. 2015;66(1):85-97.

66. Shekhawat SS, Ghosh I. Split-protein systems: beyond binary protein-protein interactions. Curr Opin Chem Biol. 2011;15(6):789-797.

67. Broome AM, Bhavsar N, Ramamurthy G, Newton G, Basilion JP. Expanding the utility of beta-galactosidase complementation: piece by piece. Mol Pharm. 2010;7(1):60-74.

68. Wehrman T, Kleaveland B, Her JH, Balint RF, Blau HM. Protein-protein interactions monitored in mammalian cells via complementation of beta -lactamase enzyme fragments. Proc Natl Acad Sci U S A. 2002;99(6):3469-3474.

69. Wehr MC, Laage R, Bolz U, Fischer TM, Grünewald S, Scheek S, et al. Monitoring regulated protein-protein interactions using split TEV. Nat Methods. 2006;3:985-993.

70. Pelletier JN, Campbell-Valois FX, Michnick SW. Oligomerization domain-directed reassembly of active dihydrofolate reductase from rationally designed fragments. Proc Natl Acad Sci U S A. 1998;95(21):12141-12146.

71. Martell JD, Yamagata M, Deerinck TJ, Phan S, Kwa CG, Ellisman MH, et al. A split horseradish peroxidase for the detection of intercellular protein–protein interactions and sensitive visualization of synapses. Nature Biotechnol. 2016;34:774-780.

72. Nagai T, Sawano A, Park ES, Miyawaki A. Circularly permuted green fluorescent proteins engineered to sense Ca2+. Proc Natl Acad Sci U S A. 2001;98(6):3197-3202.

73. Yu Y, Lutz S. Circular permutation: a different way to engineer enzyme structure and function. Trends Biotechnol. 2011;29(1):18-25.

74. Rodriguez EA, Campbell RE, Lin JY, Lin MZ, Miyawaki A, Palmer AE, et al. The Growing and Glowing Toolbox of Fluorescent and Photoactive Proteins. Trends Biochem Sci. 2017;42(2):111-129.

75. Gross LA, Baird GS, Hoffman RC, Baldridge KK, Tsien RY. The structure of the chromophore within DsRed, a red fluorescent protein from coral. Proc Natl Acad Sci U S A. 2000;97(22):11990-11995.

76. Campbell RE, Tour O, Palmer AE, Steinbach PA, Baird GS, Zacharias DA, et al. A monomeric red fluorescent protein. Proc Natl Acad Sci U S A. 2002;99(12):7877-7882.

77. Cranfill PJ, Sell BR, Baird MA, Allen JR, Lavagnino Z, de Gruiter HM, et al. Quantitative assessment of fluorescent proteins. Nat Methods. 2016;13(7):557-562.

78. Alieva NO, Konzen KA, Field SF, Meleshkevitch EA, Hunt ME, Beltran-Ramirez V, et al. Diversity and evolution of coral fluorescent proteins. PLoS One. 2008;3(7):1-12.

79. Baloban M, Shcherbakova DM, Pletnev S, Pletnev VZ, Lagarias JC, Verkhusha VV. Designing brighter near-infrared fluorescent proteins: insights from structural and biochemical studies. Chem Sci. 2017;8(6):4546-4557.

80. Shcherbakova DM, Baloban M, Emelyanov AV, Brenowitz M, Guo P, Verkhusha VV. Bright monomeric near-infrared fluorescent proteins as tags and biosensors for multiscale imaging. Nature Comm. 2016;7:1-12.

81. Chernov KG, Redchuk TA, Omelina ES, Verkhusha VV. Near-Infrared Fluorescent Proteins, Biosensors, and Optogenetic Tools Engineered from Phytochromes. Chem Rev. 2017;117(9):6423-6446.

82. Gozzelino R, Jeney V, Soares MP. Mechanisms of cell protection by heme oxygenase-1. Annu Rev Pharmacol Toxicol. 2010;50(1):323-354.

83. Waza AA, Hamid Z, Ali S, Bhat SA, Bhat MA. A review on heme oxygenase-1 induction: is it a necessary evil. Inflamm Res. 2018;67(7):579-588.

84. Beale SI. Biosynthesis of phycobilins. Chem Rev. 1993;93(2):785-802.

85. Uda Y, Goto Y, Oda S, Kohchi T, Matsuda M, Aoki K. Efficient synthesis of phycocyanobilin in mammalian cells for optogenetic control of cell signaling. Proc Natl Acad Sci U S A. 2017;114(45):11962-11967.

86. To TL, Piggott BJ, Makhijani K, Yu D, Jan YN, Shu X. Rationally designed fluorogenic protease reporter visualizes spatiotemporal dynamics of apoptosis in vivo. Proc Natl Acad Sci U S A. 2015;112(11):3338-3343.

87. Meighen EA. Enzymes and Genes from the lux Operons of Bioluminescent Bacteria. Ann Rev Microbiol. 1988;42(1):151-176.

88. Meighen EA. Molecular biology of bacterial bioluminescence. Microbiol Rev. 1991;55(1):123-142.

89. Campbell ZT, Weichsel A, Montfort WR, Baldwin TO. Crystal structure of the bacterial luciferase/flavin complex provides insight into the function of the beta subunit. Biochemistry. 2009;48(26):6085-6094.

90. Close DM, Patterson SS, Ripp S, Baek SJ, Sanseverino J, Sayler GS. Autonomous bioluminescent expression of the bacterial luciferase gene cassette (lux) in a mammalian cell line. PLoS One. 2010;5(8):1-12.

91. Gregor C, Gwosch KC, Sahl SJ, Hell SW. Strongly enhanced bacterial bioluminescence with the ilux operon for single-cell imaging. Proc Natl Acad Sci U S A. 2018;115(5):962-967.

92. Waidmann MS, Bleichrodt FS, Laslo T, Riedel CU. Bacterial luciferase reporters: the Swiss army knife of molecular biology. Bioeng Bugs. 2011;2(1):8-16.

93. Kotlobay AA, Sarkisyan KS, Mokrushina YA, Marcet-Houben M, Serebrovskaya EO, Markina NM, et al. Genetically encodable bioluminescent system from fungi. Proc Natl Acad Sci U S A. 2018;115(50):12728-12732.

94. Dixon AS, Schwinn MK, Hall MP, Zimmerman K, Otto P, Lubben TH, et al. NanoLuc Complementation Reporter Optimized for Accurate Measurement of Protein Interactions in Cells. ACS Chemical Biology. 2016;11(2):400-408.

95. Kaskova ZM, Tsarkova AS, Yampolsky IV. 1001 lights: luciferins, luciferases, their mechanisms of action and applications in chemical analysis, biology and medicine. Chem Soc Rev. 2016;45(21):6048-6077.

96. Mofford DM, Reddy GR, Miller SC. Aminoluciferins extend firefly luciferase bioluminescence into the near-infrared and can be preferred substrates over D-luciferin. J Am Chem Soc. 2014;136(38):13277-13282.

97. Mezzanotte L, Que I, Kaijzel E, Branchini B, Roda A, Lowik C. Sensitive dual color in vivo bioluminescence imaging using a new red codon optimized firefly luciferase and a green click beetle luciferase. PLoS One. 2011;6(4):1-9.

98. McCutcheon DC, Porterfield WB, Prescher JA. Rapid and scalable assembly of firefly luciferase substrates. Org Biomol Chem. 2015;13(7):2117-2121.

99. Kuchimaru T, Iwano S, Kiyama M, Mitsumata S, Kadonosono T, Niwa H, et al. A luciferin analogue generating near-infrared bioluminescence achieves highly sensitive deep-tissue imaging. Nat Commun. 2016;7:1-8.

100. Mitani Y, Yasuno R, Isaka M, Mitsuda N, Futahashi R, Kamagata Y, et al. Novel gene encoding a unique luciferase from the fireworm Odontsyllis undecimdonta. Sci Rep. 2018;8(1):1-7.

101. Sabatini DM. Twenty-five years of mTOR: Uncovering the link from nutrients to growth. Proc Natl Acad Sci U S A. 2017;114(45):11818-11825.

102. Chen J, Zheng XF, Brown EJ, Schreiber SL. Identification of an 11-kDa FKBP12-rapamycin-binding domain within the 289-kDa FKBP12-rapamycin-associated protein and characterization of a critical serine residue. Proc Natl Acad Sci U S A. 1995;92(11):4947-4951.

103. Banaszynski LA, Liu CW, Wandless TJ. Characterization of the FKBP.rapamycin.FRB ternary complex. J Am Chem Soc. 2005;127(13):4715-4721.

104. Kumagai A, Ando R, Miyatake H, Greimel P, Kobayashi T, Hirabayashi Y, et al. A bilirubin-inducible fluorescent protein from eel muscle. Cell. 2013;153(7):1602-1611.

105. Chernov KG, Redchuk TA, Omelina ES, Verkhusha VV. Near-Infrared Fluorescent Proteins, Biosensors, and Optogenetic Tools Engineered from Phytochromes. Chemical Reviews. 2017;117(9):6423-6446.

106. Nakatsu T, Ichiyama S, Hiratake J, Saldanha A, Kobashi N, Sakata K, et al. Structural basis for the spectral difference in luciferase bioluminescence. Nature. 2006;440(7082):372-376.

107. Head JF, Inouye S, Teranishi K, Shimomura O. The crystal structure of the photoprotein aequorin at 2.3 A resolution. Nature. 2000;405(6784):372-376.

108. England CG, Ehlerding EB, Cai W. NanoLuc: A Small Luciferase Is Brightening Up the Field of Bioluminescence. Bioconjug Chem. 2016;27(5):1175-1187.

109. Tomabechi Y, Hosoya T, Ehara H, Sekine SI, Shirouzu M, Inouye S. Crystal structure of nanoKAZ: The mutated 19 kDa component of Oplophorus luciferase catalyzing the bioluminescent reaction with coelenterazine. Biochem Biophys Res Commun. 2016;470(1):88-93.

110. Ashraf KU, Josts I, Mosbahi K, Kelly SM, Byron O, Smith BO, et al. The Potassium Binding Protein Kbp Is a Cytoplasmic Potassium Sensor. Structure. 2016;24(5):741-749.

111. Tidow H, Nissen P. Structural diversity of calmodulin binding to its target sites. FEBS J. 2013;280(21):5551-5565.

112. Lindenburg LH, Vinkenborg JL, Oortwijn J, Aper SJ, Merkx M. MagFRET: the first genetically encoded fluorescent Mg2+ sensor. PLoS One. 2013;8(12):1-10.

113. Cox JA, Tirone F, Durussel I, Firanescu C, Blouquit Y, Duchambon P, et al. Calcium and magnesium binding to human centrin 3 and interaction with target peptides. Biochemistry. 2005;44(3):840-850.

114. Qin Y, Sammond DW, Braselmann E, Carpenter MC, Palmer AE. Development of an Optical Zn(2+) Probe Based on a Single Fluorescent Protein. ACS Chem Biol. 2016;11(10):2744-2751.

115. Wang Z, Feng LS, Matskevich V, Venkataraman K, Parasuram P, Laity JH. Solution structure of a Zap1 zinc-responsive domain provides insights into metalloregulatory transcriptional repression in Saccharomyces cerevisiae. J Mol Biol. 2006;357(4):1167-1183.

116. Bischof H, Rehberg M, Stryeck S, Artinger K, Eroglu E, Waldeck-Weiermair M, et al. Novel genetically encoded fluorescent probes enable real-time detection of potassium in vitro and in vivo. Nat Commun. 2017;8(1):1-12.

117. Liang J, Choi J, Clardy J. Refined structure of the FKBP12–rapamycin–FRB ternary complex at 2.2 Å resolution. Acta Crystallogr D Struct Biol. 1999;55(4):736-744.

118. Choi J, Chen J, Schreiber SL, Clardy J. Structure of the FKBP12-rapamycin complex interacting with the binding domain of human FRAP. Science. 1996;273(5272):239-242.

119. Baird GS, Zacharias DA, Tsien RY. Circular permutation and receptor insertion within green fluorescent proteins. Proc Natl Acad Sci U S A. 1999;96(20):11241-11246.

120. Ding WL, Miao D, Hou YN, Jiang SP, Zhao BQ, Zhou M, et al. Small monomeric and highly stable near-infrared fluorescent markers derived from the thermophilic phycobiliprotein, ApcF2. Biochim Biophys Acta Mol Cell Res. 2017;1864(10):1877-1886.

121. Rodriguez EA, Tran GN, Gross LA, Crisp JL, Shu X, Lin JY, et al. A far-red fluorescent protein evolved from a cyanobacterial phycobiliprotein. Nat Methods. 2016;13(9):763-769.

122. Fuenzalida-Werner JP, Janowski R, Mishra K, Weidenfeld I, Niessing D, Ntziachristos V, et al. Crystal structure of a biliverdin-bound phycobiliprotein: Interdependence of oligomerization and chromophorylation. J Struct Biol. 2018;204(3):519-522.

123. Wlodawer A, Minor W, Dauter Z, Jaskolski M. Protein crystallography for non-crystallographers, or how to get the best (but not more) from published macromolecular structures. FEBS J. 2008;275(1):1-21.

124. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, et al. UCSF Chimera--a visualization system for exploratory research and analysis. J Comput Chem. 2004;25(13):1605-1612.

125. Lo WC, Lyu PC. CPSARST: an efficient circular permutation search tool applied to the detection of novel protein structural relationships. Genome Biol. 2008;9(1):1-16.

126. Lo WC, Lee CC, Lee CY, Lyu PC. CPDB: a database of circular permutation in proteins. Nucleic Acids Res. 2009;37(Database issue):328-332.

127. Lo WC, Wang LF, Liu YY, Dai T, Hwang JK, Lyu PC. CPred: a web server for predicting viable circular permutations in proteins. Nucleic Acids Res. 2012;40(Web Server issue):232-237.

128. Dagliyan O, Krokhotin A, Ozkan-Dagliyan I, Deiters A, Der CJ, Hahn KM, et al. Computational design of chemogenetic and optogenetic split proteins. Nat Commun. 2018;9(1):1-8.

129. Darwin C. On the origin of species by means of natural selection, or preservation of favoured races in the struggle for life: London : John Murray, 1859; 1859.

130. <gm-65.pdf>.

131. Watson JD, Crick FH. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. Nature. 1953;171(4356):737-738.

132. <https://www.nobelprize.org/prizes/medicine/1962/summary/> TNPiPoMNoNMASD. [

133. <https://www.nobelprize.org/prizes/medicine/1968/summary/> TNPiPoMNoNMASD. [

134. Saiki RK, Gelfand DH, Stoffel S, Scharf SJ, Higuchi R, Horn GT, et al. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science. 1988;239(4839):487-491.

135. Hutchison CA, Phillips S, Edgell MH, Gillam S, Jahnke P, Smith M. Mutagenesis at a specific position in a DNA sequence. J Biol Chem. 1978;253(18):6551-6560.

136. <https://www.nobelprize.org/prizes/chemistry/1993/press-release/> PrNoNMASD. [

137. Crick F. Central dogma of molecular biology. Nature. 1970;227(5258):561-563.

138. Davis AM, Plowright AT, Valeur E. Directing evolution: the next revolution in drug discovery? Nat Rev Drug Discov. 2017;16(10):681-698.

139. Cadwell RC, Joyce GF. Randomization of genes by PCR mutagenesis. PCR Methods Appl. 1992;2(1):28-33.

140. Cunningham BC, Wells JA. High-resolution epitope mapping of hGH-receptor interactions by alanine-scanning mutagenesis. Science. 1989;244(4908):1081-1085.

141. Zhao H, Giver L, Shao Z, Affholter JA, Arnold FH. Molecular evolution by staggered extension process (StEP) in vitro recombination. Nat Biotechnol. 1998;16(3):258-261.

142. El-Deiry WS, Downey KM, So AG. Molecular mechanisms of manganese mutagenesis. Proc Natl Acad Sci U S A. 1984;81(23):7378-7382.

143. Laos R, Thomson JM, Benner SA. DNA polymerases engineered by directed evolution to incorporate non-standard nucleotides. Front Microbiol. 2014;5:1-14.

144. https://www.agilent.com/en/product/mutagenesis-cloning/mutagenesis-kits/random-mutagenesis-kits/genemorph-ii-233115.

145. Clark JM. Novel non-templated nucleotide addition reactions catalyzed by procaryotic and eucaryotic DNA polymerases. Nucleic Acids Res. 1988;16(20):9677-9686.

146. Wong TS, Roccatano D, Zacharias M, Schwaneberg U. A statistical analysis of random mutagenesis methods used for directed protein evolution. J Mol Biol. 2006;355(4):858-871.

147. Chae Y, Woochun, J, Kyoung, SK. Rapid and Simple Method to Prepare Functional Pfu DNA Polymerase

Expressed in Escherichia coli Periplasm. J Microbiol Biotechnol. 2002;12(5):841-843.

148. Dolgova AS, Stukolova OA. High-fidelity PCR enzyme with DNA-binding domain facilitates de novo gene synthesis. 3 Biotech. 2017;7(2):1-6.

149. Sadanand S. EvolvR-ing to targeted mutagenesis. Nat Biotechnol. 2018;36(9):819.

150. Halperin SO, Tou CJ, Wong EB, Modavi C, Schaffer DV, Dueber JE. CRISPR-guided DNA polymerases enable diversification of all nucleotides in a tunable window. Nature. 2018;560(7717):248-252.

151. Bindels DS, Haarbosch L, van Weeren L, Postma M, Wiese KE, Mastop M, et al. mScarlet: a bright monomeric red fluorescent protein for cellular imaging. Nat Methods. 2017;14(1):53-56.

152. Xiao H, Bao Z, Zhao H. High Throughput Screening and Selection Methods for Directed Enzyme Evolution. Ind Eng Chem Res. 2015;54(16):4011-4020.

153. Wojcik M, Telzerow A, Quax WJ, Boersma YL. High-Throughput Screening in Protein Engineering: Recent Advances and Future Perspectives. Int J Mol Sci. 2015;16(10):24918-24945.

154. Lundby A, Mutoh H, Dimitrov D, Akemann W, Knopfel T. Engineering of a genetically encodable fluorescent voltage sensor exploiting fast Ci-VSP voltage-sensing movements. PLoS One. 2008;3(6):1-5.

155. Dowell SJ, Brown AJ. Yeast Assays for G Protein-Coupled Receptors. In: Leifert WR, editor. G Protein-Coupled Receptors in Drug Discovery. Totowa, NJ: Humana Press; 2009. p. 213-229.

156. Schutz M, Schoppe J, Sedlak E, Hillenbrand M, Nagy-Davidescu G, Ehrenmann J, et al. Directed evolution of G protein-coupled receptors in yeast for higher functional production in eukaryotic expression hosts. Sci Rep. 2016;6:1-16.

157. Gonzalez-Perez D, Garcia-Ruiz E, Alcalde M. Saccharomyces cerevisiae in directed evolution: An efficient tool to improve enzymes. Bioeng Bugs. 2012;3(3):172-177.

158. Pourmir A, Johannes TW. Directed evolution: selection of the host organism. Comput Struct Biotechnol J. 2012;2:1-7.

159. Hamilton SR, Gerngross TU. Glycosylation engineering in yeast: the advent of fully humanized yeast. Curr Opin Biotechnol. 2007;18(5):387-392.

160. De Pourcq K, De Schutter K, Callewaert N. Engineering of glycosylation in yeast and other fungi: current state and perspectives. Appl Microbiol Biotechnol. 2010;87(5):1617-1631.

161. Adeniran A, Sherer M, Tyo KE. Yeast-based biosensors: design and applications. FEMS Yeast Res. 2015;15(1):1-15.

162. Ault AD, Broach JR. Creation of GPCR-based chemical sensors by directed evolution in yeast. Protein Eng Des Sel. 2006;19(1):1-8.

163. van Oers MM, Pijlman GP, Vlak JM. Thirty years of baculovirus-insect cell protein expression: from dark horse to mainstream technology. J Gen Virol. 2015;96(Pt 1):6-23.

164. Bandaranayake AD, Almo SC. Recent advances in mammalian protein production. FEBS Lett. 2014;588(2):253-260.

165. Park J, Werley CA, Venkatachalam V, Kralj JM, Dib-Hajj SD, Waxman SG, et al. Screening fluorescent voltage indicators with spontaneously spiking HEK cells. PLoS One. 2013;8(12):1-10.

166. Piatkevich KD, Jung EE, Straub C, Linghu C, Park D, Suk HJ, et al. A robotic multidimensional directed evolution approach applied to fluorescent voltage reporters. Nat Chem Biol. 2018;14(4):352-360.

167. Lee HH, Ostrov N, Wong BG, Gold MA, Khalil A, Church GM. Vibrio natriegens, a new genomic powerhouse. bioRxiv. 2016:1-30.

168. Weinstock MT, Hesek ED, Wilson CM, Gibson DG. Vibrio natriegens as a fast-growing host for molecular biology. Nat Methods. 2016;13(10):849-851.

169. Chalfie M, Tu Y, Euskirchen G, Ward WW, Prasher DC. Green Fluorescent Protein as a Marker for Gene-Expression. Science. 1994;263(5148):802-805.

170. Lukyanov KA, Chudakov DM, Lukyanov S, Verkhusha VV. Innovation: Photoactivatable fluorescent proteins. Nat Rev Mol Cell Biol. 2005;6(11):885-891.

171. Ai HW, Hazelwood KL, Davidson MW, Campbell RE. Fluorescent protein FRET pairs for ratiometric imaging of dual biosensors. Nat Methods. 2008;5(5):401-403.

172. Zhou XX, Lin MZ. Photoswitchable fluorescent proteins: ten years of colorful chemistry and exciting applications. Curr Opin Chem Biol. 2013;17(4):682-690.

173. Shaner NC, Lambert GG, Chammas A, Ni Y, Cranfill PJ, Baird MA, et al. A bright monomeric green fluorescent protein derived from Branchiostoma lanceolatum. Nat Methods. 2013;10(5):407-409.

174. Deheyn DD, Kubokawa K, McCarthy JK, Murakami A, Porrachia M, Rouse GW, et al. Endogenous green fluorescent protein (GFP) in amphioxus. Biol Bull. 2007;213(2):95-100.

175. Clavel D, Gotthard G, von Stetten D, De Sanctis D, Pasquier H, Lambert GG, et al. Structural analysis of the bright monomeric yellow-green fluorescent protein mNeonGreen obtained by directed evolution. Acta Crystallogr D Struct Biol. 2016;72(Pt 12):1298-1307.

176. Pletnev VZ, Pletneva NV, Lukyanov KA, Souslova EA, Fradkov AF, Chudakov DM, et al. Structure of the red fluorescent protein from a lancelet (Branchiostoma

lanceolatum): a novel GYG chromophore covalently bound to a nearby tyrosine. Acta Crystallogr D Biol Crystallogr. 2013;69(Pt 9):1850-1860.

177. Wilson DS, Keefe AD. Random mutagenesis by PCR. Curr Protoc Mol Biol. 2001;Chapter 8(1):1-9.

178. Zhao H, Giver L, Shao Z, Affholter JA, Arnold FH. Molecular evolution by staggered extension process (StEP) in vitro recombination. Nat Biotechnol. 1998;16:258-261.

179. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci U S A. 1977;74(12):5463-5467.

180. Bowler MW, Svensson O, Nurizzo D. Fully automatic macromolecular crystallography: the impact of MASSIF-1 on the optimum acquisition and quality of data. Crystallography Reviews. 2016;22(4):233-249.

181. Kabsch W. Integration, scaling, space-group assignment and post-refinement. Acta Crystallogr D Struct Biol. 2010;66(2):133-144.

182. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. Phaser crystallographic software. J Appl Crystallogr. 2007;40(Pt 4):658-674.

183. Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of Coot. Acta Crystallogr D Struct Biol. 2010;66(4):486-501.

184. Murshudov GN, Skubak P, Lebedev AA, Pannu NS, Steiner RA, Nicholls RA, et al. REFMAC5 for the refinement of macromolecular crystal structures. Acta Crystallogr D Struct Biol. 2011;67(4):355-367.

185. Pletnev VZ, Pletneva NV, Sarkisyan KS, Mishin AS, Lukyanov KA, Goryacheva EA, et al. Structure of the green fluorescent protein NowGFP with an anionic tryptophan-based chromophore. Acta Crystallogr D Biol Crystallogr. 2015;71(Pt 8):1699-1707.

186. Sarkisyan KS, Yampolsky IV, Solntsev KM, Lukyanov SA, Lukyanov KA, Mishin AS. Tryptophan-based chromophore in fluorescent proteins can be anionic. Sci Rep. 2012;2:1-5.

187. Gotthard G, von Stetten D, Clavel D, Noirclerc-Savoye M, Royant A. Chromophore Isomer Stabilization Is Critical to the Efficient Fluorescence of Cyan Fluorescent Proteins. Biochemistry. 2017;56(49):6418-6422.

188. Schell MJ, Erneux C, Irvine RF. Inositol 1,4,5-trisphosphate 3-kinase A associates with F-actin and dendritic spines via its N terminus. J Biol Chem. 2001;276(40):37537-37546.

189. Belin BJ, Goins LM, Mullins RD. Comparative analysis of tools for live cell imaging of actin network architecture. Bioarchitecture. 2014;4(6):189-202.

190. Carmody WR. Easily prepared wide range buffer series. J Chem Educ. 1961;38(11):559-560.

191. Rueden CT, Schindelin J, Hiner MC, DeZonia BE, Walter AE, Arena ET, et al. ImageJ2: ImageJ for the next generation of scientific image data. BMC Bioinformatics. 2017;18(1):1-26.

192. Schneider CA, Rasband WS, Eliceiri KW. NIH Image to ImageJ: 25 years of image analysis. Nat Methods. 2012;9(7):671-675.

193. Lelimousin M, Noirclerc-Savoye M, Lazareno-Saez C, Paetzold B, Le Vot S, Chazal R, et al. Intrinsic Dynamics in ECFP and Cerulean Control Fluorescence Quantum Yield. Biochemistry. 2009;48(42):10038-10046.

194. Mank M, Griesbeck O. Genetically encoded calcium indicators. Chem Rev. 2008;108(5):1550-1564.

195. Tian L, Hires SA, Looger LL. Imaging neuronal activity with genetically encoded calcium indicators. Cold Spring Harb Protoc. 2012;2012(6):647-656.

196. Hasan MT, Friedrich RW, Euler T, Larkum ME, Giese G, Both M, et al. Functional fluorescent Ca2+ indicator proteins in transgenic mice under TET control. PLoS Biol. 2004;2(6):1-13.

197. Knopfel T. Genetically encoded optical indicators for the analysis of neuronal circuits. Nat Rev Neurosci. 2012;13(10):687-700.

198. Pologruto TA, Yasuda R, Svoboda K. Monitoring neural activity and [Ca2+] with genetically encoded Ca2+ indicators. J Neurosci. 2004;24(43):9572-9579.

199. Tian L, Akerboom J, Schreiter ER, Looger LL. Neural activity imaging with genetically encoded calcium indicators. Prog Brain Res. 2012;196:79-94.

200. Deo C, Lavis LD. Synthetic and genetically encoded fluorescent neural activity indicators. Curr Opin Neurobiol. 2018;50:101-108.

201. Broussard GJ, Liang R, Tian L. Monitoring activity in neural circuits with genetically encoded indicators. Front Mol Neurosci. 2014;7:1-17.

202. Podor B, Hu YL, Ohkura M, Nakai J, Croll R, Fine A. Comparison of genetically encoded calcium indicators for monitoring action potentials in mammalian brain by two-photon excitation fluorescence microscopy. Neurophotonics. 2015;2(2):1-7.

203. Bootman MD, Allman S, Rietdorf K, Bultynck G. Deleterious effects of calcium indicators within cells; an inconvenient truth. Cell Calcium. 2018;73:82-87.

204. Steinmetz NA, Buetfering C, Lecoq J, Lee CR, Peters AJ, Jacobs EAK, et al. Aberrant Cortical Activity in Multiple GCaMP6-Expressing Transgenic Mouse Lines. eNeuro. 2017;4(5):1-15.

205. Buschmeier B, Meyer HE, Mayr GW. Characterization of the calmodulin-binding sites of muscle phosphofructokinase and comparison with known calmodulin-binding domains. J Biol Chem. 1987;262(20):9454-9462.

206. Lukas TJ, Burgess WH, Prendergast FG, Lau W, Watterson DM. Calmodulin binding domains: characterization of a phosphorylation and calmodulin binding site from myosin light chain kinase. Biochemistry. 1986;25(6):1458-1464.

207. Zhao Y, Araki S, Wu J, Teramoto T, Chang YF, Nakano M, et al. An expanded palette of genetically encoded Ca(2)(+) indicators. Science. 2011;333(6051):1888-1891.

208. Boyden ES, Zhang F, Bamberg E, Nagel G, Deisseroth K. Millisecond-timescale, genetically targeted optical control of neural activity. Nat Neurosci. 2005;8(9):1263-1268.

209. Akerboom J, Carreras Calderon N, Tian L, Wabnig S, Prigge M, Tolo J, et al. Genetically encoded calcium indicators for multi-color neural activity imaging and combination with optogenetics. Front Mol Neurosci. 2013;6(2):1-29.

210. Dana H, Mohar B, Sun Y, Narayan S, Gordus A, Hasseman JP, et al. Sensitive red protein calcium indicators for imaging neural activity. Elife. 2016;5:1-24.

211. Forli A, Vecchia D, Binini N, Succol F, Bovetti S, Moretti C, et al. Two-Photon Bidirectional Control and Imaging of Neuronal Excitability with High Spatial Resolution In Vivo. Cell Rep. 2018;22(11):3087-3098.

212. Shen Y, Lai T, Campbell RE. Red fluorescent proteins (RFPs) and RFP-based biosensors for neuronal imaging applications. Neurophotonics. 2015;2(3):1-12.

213. Helmchen F, Imoto K, Sakmann B. Ca2+ buffering and action potential-evoked Ca2+ signaling in dendrites of pyramidal neurons. Biophys J. 1996;70(2):1069-1081.

214. Ashworth R, Zimprich F, Bolsover SR. Buffering intracellular calcium disrupts motoneuron development in intact zebrafish embryos. Brain Res Dev Brain Res. 2001;129(2):169-179.

215. Yang Y, Liu N, He Y, Liu Y, Ge L, Zou L, et al. Improved calcium sensor GCaMP-X overcomes the calcium channel perturbations induced by the calmodulin in GCaMP. Nat Commun. 2018;9(1):1-18.

216. Rose T, Goltstein PM, Portugues R, Griesbeck O. Putting a finishing touch on GECIs. Front Mol Neurosci. 2014;7:1-15.

217. Wu J, Abdelfattah AS, Miraucourt LS, Kutsarova E, Ruangkittisakul A, Zhou H, et al. A long Stokes shift red fluorescent Ca2+ indicator protein for two-photon and ratiometric imaging. Nat Commun. 2014;5:1-26.

218. Gibson DG, Young L, Chuang RY, Venter JC, Hutchison CA, 3rd, Smith HO. Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat Methods. 2009;6(5):343-345.

219. Benoit RM, Ostermeier C, Geiser M, Li JS, Widmer H, Auer M. Seamless Insert-Plasmid Assembly at High Efficiency and Low Cost. PLoS One. 2016;11(4):1-13.

220. Palmer AE, Tsien RY. Measuring calcium signaling using genetically targetable fluorescent indicators. Nat Protoc. 2006;1(3):1057-1065.

221. Badminton MN, Campbell AK, Rembold CM. Differential regulation of nuclear and cytosolic Ca2+ in HeLa cells. J Biol Chem. 1996;271(49):31210-31214.

222. Kwan KM, Fujimoto E, Grabher C, Mangum BD, Hardy ME, Campbell DS, et al. The Tol2kit: a multisite gateway-based construction kit for Tol2 transposon transgenesis constructs. Dev Dyn. 2007;236(11):3088-3099.

223. Tsien R, Pozzan T. [14] Measurement of cytosolic free Ca2+ with quin2. Methods Enzymol. 172: Academic Press; 1989. p. 230-262.

224. Dweck D, Reyes-Alfonso A, Potter JD. Expanding the range of free calcium regulation in biological solutions. Analytical Biochemistry. 2005;347(2):303-315.

225. Dweck D, Reyes-Alfonso A, Jr., Potter JD. Expanding the range of free calcium regulation in biological solutions. Anal Biochem. 2005;347(2):303-315.

226. Arrowsmith CH, Audia JE, Austin C, Baell J, Bennett J, Blagg J, et al. The promise and peril of chemical probes. Nature Chemical Biology. 2015;11:536.

227. International Human Genome Sequencing C, Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, et al. Initial sequencing and analysis of the human genome. Nature. 2001;409:860-921.

228. Shekhawat SS, Ghosh I. Split-protein systems: beyond binary protein–protein interactions. Current Opinion in Chemical Biology. 2011;15(6):789-797.

229. Moustaqil M, Bhumkar A, Gonzalez L, Raoul L, Hunter DJB, Carrive P, et al. A Split-Luciferase Reporter Recognizing GFP and mCherry Tags to Facilitate Studies of Protein-Protein Interactions. Int J Mol Sci. 2017;18(12):1-15.

230. Filonov GS, Verkhusha VV. A near-infrared BiFC reporter for in vivo imaging of protein-protein interactions. Chem Biol. 2013;20(8):1078-1086.

231. Massoud TF, Paulmurugan R, Gambhir SS. A molecularly engineered split reporter for imaging protein-protein interactions with positron emission tomography. Nat Med. 2010;16(8):921-926.

232. Blanche F, Debussche L, Thibaut D, Crouzet J, Cameron B. Purification and characterization of S-adenosyl-L-methionine: uroporphyrinogen III methyltransferase from Pseudomonas denitrificans. J Bacteriol. 1989;171(8):4222-4231.

233. Vevodova J, Graham RM, Raux E, Schubert HL, Roper DI, Brindley AA, et al. Structure/function studies on a S-adenosyl-L-methionine-dependent uroporphyrinogen III C methyltransferase (SUMT), a key regulatory enzyme of tetrapyrrole biosynthesis. J Mol Biol. 2004;344(2):419-433.

234. Sattler I, Roessner CA, Stolowich NJ, Hardin SH, Harris-Haller LW, Yokubaitis NT, et al. Cloning, sequencing, and expression of the uroporphyrinogen III methyltransferase cobA gene of Propionibacterium freudenreichii (shermanii). J Bacteriol. 1995;177(6):1564-1569.

235. Warren MJ, Roessner CA, Santander PJ, Scott AI. The Escherichia coli cysG gene encodes S-adenosylmethionine-dependent uroporphyrinogen III methylase. Biochem J. 1990;265(3):725-729.

236. Wildt S, Deuschle U. cobA, a red fluorescent transcriptional reporter for Escherichia coli, yeast, and mammalian cells. Nat Biotechnol. 1999;17(12):1175-1178.

237. Storbeck S, Saha S, Krausze J, Klink BU, Heinz DW, Layer G. Crystal structure of the heme d1 biosynthesis enzyme NirE in complex with its substrate reveals new insights into the catalytic mechanism of S-adenosyl-L-methionine-dependent uroporphyrinogen III methyltransferases. J Biol Chem. 2011;286(30):26754-26767.

238. Feliciano J, Liu Y, Daunert S. Novel reporter gene in a fluorescent-based whole cell sensing system. Biotechnol Bioeng. 2006;93(5):989-997.

239. Wang Z, Li S, Li J, Li J, Rong L, Cheng B, et al. Engineering uroporphyrinogen III methyltransferase as a red fluorescent reporter in E. coli. Enzyme Microb Technol. 2014;61-62:1-6.

240. Stroupe ME, Leech HK, Daniels DS, Warren MJ, Getzoff ED. CysG structure reveals tetrapyrrole-binding features and novel regulation of siroheme biosynthesis. Nat Struct Biol. 2003;10(12):1064-1073.

241. Kim JH, Lee SR, Li LH, Park HJ, Park JH, Lee KY, et al. High cleavage efficiency of a 2A peptide derived from porcine teschovirus-1 in human cell lines, zebrafish and mice. PLoS One. 2011;6(4):1-8.

242. Shah NH, Muir TW. Inteins: Nature's Gift to Protein Chemists. Chem Sci. 2014;5(1):446-461.

243. Lennon CW, Belfort M. Inteins. Curr Biol. 2017;27(6):204-206.

244. Stevens AJ, Brown ZZ, Shah NH, Sekar G, Cowburn D, Muir TW. Design of a Split Intein with Exceptional Protein Splicing Activity. J Am Chem Soc. 2016;138(7):2162-2165.

245. Goedhart J, von Stetten D, Noirclerc-Savoye M, Lelimousin M, Joosen L, Hink MA, et al. Structure-guided evolution of cyan fluorescent proteins towards a quantum yield of 93%. Nat Commun. 2012;3:1-9.

246. Biles BD, Connolly BA. Low-fidelity Pyrococcus furiosus DNA polymerase mutants useful in error-prone PCR. Nucleic Acids Res. 2004;32(22):1-7.

247. Wang Y, Prosen DE, Mei L, Sullivan JC, Finney M, Vander Horn PB. A novel strategy to engineer DNA polymerases for enhanced processivity and improved performance in vitro. Nucleic Acids Res. 2004;32(3):1197-1207.