Expanding The Uses of Split-Intein Through Protein Engineering

by

Stanley Siu Cheung Wong

A thesis submitted in conformity with the requirements
for the degree of Doctor of Philosophy

Institute of Biomaterials and Biomedical Engineering
University of Toronto

Expanding the Uses of Split-Intein Through Protein Engineering

Stanley Siu Cheung Wong

Doctor of Philosophy

Institute of Biomaterials and Biomedical Engineering
University of Toronto

2013

Abstract

Split-protein systems are invaluable tools used for the discovery and investigations of the complexities of protein functions and interactions. Split-protein systems rely on the non-covalent interactions of two fragments of a split protein to restore protein function. Because of this, they have the ability to restore protein functions post-translationally, thus allowing for quick and efficient responses to a milieu of cellular mechanisms. Despite this, split-protein systems have been largely limited as a reporting tool for protein-protein interactions. The recent discovery of inteins has the potential of broadening the scope of split-protein systems. Inteins are protein elements that possess the unique ability of post-translationally ligating protein fragments together with a native peptide bond, a process termed protein splicing. This allows split-proteins to reassemble in a more natural state. Exploiting this property and utilizing protein engineering techniques and methodologies, several approaches are described here for restoring and controlling split-protein functions using inteins.

First, the protein splicing behaviour was demonstrated with the development of a simple *in vitro* visual fluorescence assay that relies on examining the subcellular localization of different fluorescent proteins. Inteins were then used to reassemble and restore function to artificially split genetically encoded Ca^{2+} indicators.

Second, inteins were shown to be able to simultaneously restore protein function to two target proteins. The first target protein was restored through the normal protein splicing pathway while the second was restored through non-covalent interactions of the split-protein fragments. This is a previous unknown property of inteins.

Lastly, an intein was engineered to respond to an external light-stimulus that triggered protein splicing to restore split-protein function. The photoactivatable intein, coupled with the versatility of light, allows exquisite control in both space and time for the restoration of protein function within cells. The modularity of the photoactivatable intein can be simply attached to a variety of split-proteins. This was demonstrated with the restoration of various split-protein functions.

Acknowledgements

My supervisor, Dr. Kevin Truong,

for his constant encouragement, guidance, and invaluable ideas.

My past and present lab friends and colleagues,

for their friendship, technical assistance, and moral support.

My supervisory committee,

for their mentoring, insights, and challenging me.

My friends and family

for their support, patience and encouragement.

Table of Contents

List of Publications

- **Wong SS.**, Mills E., Truong K. Simultaneous assembly of two target proteins using split inteins. (Submitted).

- **Wong SS**., Pham E., Mills E., Truong K. An engineered naturally split intein for photoactivatable protein *trans*-splicing. (Submitted).

- Nagaraj S., Mills E, **Wong SS.**, Truong K. Programming membrane fusion and subsequent apoptosis into mammalian cells. *ACS Synthetic Biology*. (In Press)

- Nagaraj S., **Wong SS.**, Truong K. Parts-based assembly of synthetic transmembrane proteins in mammalian cells. *ACS Synthetic Biology*. 2012. 4(1):111-117

- Pham E., **Wong SS.**, Nagaraj S., Truong K. Effects of rapamycin-induced oligomerization of Parvalbumin, Stim 1 and Orai 1 in puncta formation. *Cell Calcium*. 2012. 51:57-64

- Mills E., Chen X., Pham E., **Wong SS.**, Truong K. Engineering a photoactivated caspase-7 for rapid induction of apoptosis. *ACS Synthetic Biology*. 2012. 3(1):75-82

- **Wong SS.**, Kotera I., Mills E., Suzuki H., Truong K., Split-intein mediated reassembly of genetically encoded Ca^{2+} indicators. *Cell Calcium*. 2012. 51:57-64

- **Wong SS.**, Truong K. Fluorescent protein-based methods for on-plate screening of gene insertion. *PLoS ONE*. 2010. 5(12):e14274

List of Abbreviations

ATP	Adenosine-5'-triphosphate
BiFC	Bimolecular Fluorescence Complementation
CaM	Calmodulin
Ceru	Cerulean, variant of cyan fluorescent protein
CFP	Cyan Fluorescent Protein
DMEM	Dulbecco's Modified Eagle Medium
EPL	Expressed Protein Ligation
ER	Endoplasmic Reticulum
FBS	Fetal Bovine Serum
FMN	Flavin mononucleotide
FP	Fluorescent Protein
FRET	Fluorescence Resonance Energy Transfer
GECI	Genetically Encoded Ca^{2+} Indicators
GFP	Green Fluorescent Protein
IP3	inositol 1,4,5-triphosphate
IP3R	inositol 1,4,5-triphosphate receptor
Jα	C-terminal helical peptide of LOV2
LOV	Light-Oxygen-Voltage domain
LOVInC	Photoactivatable C-terminal *Npu*DnaE intein
MCS	Multiple Cloning Site
mRFP	Monomeric Red Fluorescent Protein
NMR	Nuclear Magnetic Resonance
*Npu*DnaE	*dnaE* gene from *Nostoc punctiforme*
*Npu*DnaE$_C$	C-terminal intein half of *Npu*DnaE
*Npu*DnaE$_N$	N-terminal intein half of *Npu*DnaE
PC	Pearson's Coefficient
PCR	Polymerase Chain Reaction
PDB	Protein DataBank
PLCβ	Phospholipase-Cβ
PhyB	Phytochrome B
PIF3	Phytochrome Interacting Factor 3
PIP3	Phosphatidylinositol 4,5-bisphosphate
PM	Plasma Membrane
PTS	Protein *Trans*-Splicing
RPM	Revolutions Per Minute
S-DS	Shine-Dalgarno Sequence
SDS-PAGE	Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis
SOCE	Store-Operated Ca^{2+} Entry
t#*Npu*DnaE$_C$	Truncated variants of *Npu*DnaE$_C$
TnC	Troponin C
UTP	Uridine-5'-triphosphate
YFP	Yellow Fluorescent Protein

List of Figures

List of Movies

Movies can be downloaded at:

http://apel.ibbme.utoronto.ca/apel/gallery/sw

Movie 4.1 *C. Elegans* co-expressing split GCaMP2 precursors in pharyngeal muscles

Movie 5.1 HeLa cells undergoing dynamic blebbing

Movie 5.2 HEK293 cells undergoing dynamic blebbing

Movie 6.1 HeLa cells undergoing dynamic blebbing after photo-stimulation

1 Introduction

This chapter provides general motivation for this work and introduces specific research objectives.

1.1 Motivation

Protein engineering tools merge the vast library of functional proteins with genetic methodologies to create original proteins with novel functions. This has resulted in a number of protein-based approaches for studying and perturbing biological functions such as split-protein reassembly [1-3], artificial gene synthesis [4, 5], and rewiring of cellular networks via synthetic biology [6, 7]. Nonetheless, as new proteins and protein functions come to light, innovative techniques utilizing the new proteins are developed to expand on existing methodologies.

Split-protein reassembly is an important protein engineering tool used in deciphering the complexities of protein functions and interactions [1, 8, 9]. The post-translational reassembly of split proteins is generally fast acting allowing temporal resolutions comparable to biological processes. This has been exploited to monitor protein-protein interactions using split reporter proteins such as green fluorescent proteins (GFPs) [3, 10] and luciferases [11, 12]. However, protein fragment association relies on suitable protein interacting partners found within the cell or the use of dimerizable domains inducible by small-chemical molecules, both of which may interfere with cellular functions and other processes. Thus, applications of split-protein reassembly have largely been limited to monitoring protein-protein interactions. The widespread use of post-translational reassembly of split proteins will depend on a robust system that is highly specific, generalizable with respect to the split-proteins used, can be conditionally controlled, and do not interfere with endogenous cell function. The development of such system can form the basis of a protein engineering platform from which therapeutic applications can be developed.

Inteins have emerged as a promising tool to bridge this gap. Inteins are protein elements that spontaneously catalyzes the formation of a peptide bond between any two peptide fragments, a

process termed protein splicing. It subsequently removes itself leaving a single ligated polypeptide chain and the separated intein. The generality of protein splicing to ligate a wide variety of foreign polypeptides has been harnessed for the development of several protein-engineering methods such protein semisynthesis [13-15] and isotopic labelling of proteins for NMR studies [16-19]. As such, this virtually traceless catalytic process can be exploited to reassemble a wide array of split-proteins via a native peptide bond to restore protein function in cells.

1.2 Research Objectives

The goal of this research is to engineer an intein-based approach wherein a foreign input signal can be used to control the reassembly of split-proteins. This thesis achieves its goals in an objective-oriented manner. Three specific research objectives have been identified in order to achieve its overall goal:

1. Determine if naturally occurring split inteins can be used to reassemble artificially split Ca^{2+}-sensitive biosensors. Validate protein splicing by creating fusion proteins using naturally occurring split inteins to reassemble artificially split Ca^{2+} biosensors.

2. Demonstrate the dual functionality of naturally split inteins by showing a single pair of split inteins can simultaneously reassemble two separate proteins.

3. Develop and engineer a control mechanism such that protein splicing can be modulated by an exogenous signal.

1.3 Organization

This thesis is organized into nine chapters that demonstrate the ideas and design strategies employ in the development of a photo-activated protein *trans*-splicing by *NpuDnaE* intein. Chapter one provides an overall motivation for the work and sets forth specific research objectives. Chapter two provides general background knowledge on split-protein systems and how they are used; an introduction to inteins with an emphasis on naturally split inteins and how

inteins can be controlled; and finally a brief description of recurring proteins throughout this thesis. Chapter three describes the experimental methodologies and materials used throughout this work in five sections: Gene construction; *In vitro* protein analysis; Cell culture and Live cell imaging; Data analysis; and Animal model culturing and imaging.

Chapters four, five and six detail the progressive development and expansion on the uses of naturally split inteins. Chapter four describes the development of visual fluorescence assay for protein *trans*-splicing (PTS) activity by observing for distinct localization of fluorescent proteins. Protein splicing is also applied to reassemble split Ca^{2+} biosensors to validate PTS activity. Chapter five builds on chapter four and shows the dual functionality of naturally split inteins. Chapter six describes the development, characterization, and applications of a photo-activated protein *trans*-splicing intein in reassembling split proteins.

Chapter seven provides a general overview and summary of the work. Chapter eight provides some future project directions and expansion of the work. Finally chapter nine is a list of the references.

2. Background

This chapter provides general background information related to the development of a photo-activated protein *trans*-splicing intein for the reassembly of split proteins. This chapter is divided into three sections. The first section reviews split-protein systems and how they have been used to study biological systems and interactions. Selected examples will be given to highlight how split-proteins have been used. The second section introduces and describes inteins. An emphasis will be placed on the naturally split inteins, *Npu*DnaE. This section will also describe current ways inteins are controlled with light and some of the limitations. Finally, chapter three will give a brief overview of the recurring proteins used throughout this study.

2.1 Split-Protein Systems

The restoring of split-protein function within the cells has been extensively cultivated as a protein engineering tool to discover and study new macromolecular interactions, functions and potential therapeutic applications. This approach relies on the appropriate division of proteins into two inactive fragments that are fused to possible interacting partners. The subsequent association of the interacting partners brings the inactive fragments together reassembling and returning function to the split proteins. A major advantage of this approach is that the restoration of split-protein function occurs post-translationally, enabling quicker formation of desired proteins compared to conditional gene expression systems. This property has found far-reaching applications as a reporting technique. As a result, the split-proteins are commonly reporter type proteins such as GFPs (method better known as bimolecular fluorescence complementation [3]), firefly luciferases [11, 12, 20], and thymidine kinase [21] that are able to qualitatively and quantitatively report on the interplay between macromolecules and cellular localizations. Several studies are described here that use split-proteins as a reporter of protein interactions in *in vivo* and *in vitro* environments.

Split-GFP as a reporter for improved thermodynamic stability: By visually inspecting the fluorescence intensity emitted by the reconstituted split-GFPs, a variant of the B1 domain of protein G (PGB1) was found to have improved thermodynamic stability [22]. The 56-residue

PBG1 was divided into two fragments comprised of residues 1 to 40 and 41-56, which has the ability to be reconstituted back together. A library of mutant variants of fragment 1 to 40 was fused to one half of a split GFP while the wild-type fragment 41-56 was fused to the remaining half of the split GFP. Mutant variants with improved affinities and enhanced stability promoted the increased association and chromophore maturation of the GFP fragments leading to brighter fluorescence for *in vivo* screening. This study illustrates the clever use of split fluorescent protein reporters as an efficient method to not only screen for the presence of protein-protein associations but also discern between fragment affinities and stability.

Split-luciferases used to detect protein-DNA interactions: Target DNA and DNA modifications can be detected using a rapid and convenient cell-free method utilizing split-luciferases [23]. The detection of a target DNA was accomplished using two sequence-specific zinc fingers (Zif268/PBSII) that are attached to complementary halves of split-luciferases. The zinc finger domains recognize and bind target DNA harbouring specific adjacent DNA sequences, reconstituting the split-luciferases and emitting an easily measureable luminescence. In a similar manner, modifications to DNA such as DNA-methylation can be detected by replacing one of the zinc finger domains with a methyl-binding domain. This study has a number of distinctions from the study above: split-reporter proteins were reconstituted *in vitro* rather than *in vivo*; split-reporter proteins was used to detect protein-DNA interactions rather than protein-protein interactions; and finally unmodified target DNA were detected using two recognition protein domains resulting in a ternary complex rather than the coupling of protein-protein in a binary complex. However in both cases, the formation of a binary or ternary complex reassembled and returned function to a split-reporter protein.

Detection of protein interactions via split-enzyme and PET-based imaging: An ultra-sensitive *in vivo* approach for detecting protein-protein interactions was developed by combining a split-protein enzyme with positron emission tomography (PET) based imaging [21]. The herpes simplex virus type 1 thymidine kinase (HSV1-TK) enzyme was divided to yield fragments 1-265 (nTK) and 266-376 (cTK). Reassembly of the two fragments restores enzymatic activity that is known to phosphorylate a number of radiolabeled substrates. These radiolabeled substrate probes are transported into cells and trapped once the reassembled split-HSV1-TK phosphorylates them.

The general applicability of this method was used to monitor the complexation of hypoxia-inducible factor 1-α (HIF-1α) with the von Hippel-Lindau tumour suppressor protein (VHL). In addition, as a proof of principle HIF-1α and VHL were replaced with rapamycin-inducible FKBP12 and FRB domains, respectively. Subcutaneous xenografts of 293T cells stably expressing nTK-FKBP12 and FRB-cTK were injected into live mice. Rapamycin-induced reassembly of split-HSV1-TK accumulated injected radiolabeled substrates within localized xenografted regions that were able to be image with PET. This study introduces a split-reporter method that goes beyond using fluorescence or luminescence as a readout, but instead makes use of a phosphorylating enzyme to trap radiolabeled substrates. This was coupled with a highly sensitive and non-invasive PET-based imaging modality to study protein-protein interactions in live animals.

One of the drawbacks of the current reassembly of split-proteins is that it relies on the non-covalent interacting protein pairs. These pairs are commonly endogenous protein species such as zinc finger transcription factors [21, 24] or are protein binding domains that are inducible by small-chemical molecules such as FKBP12 and FRB [25]. Overexpression of endogenous proteins and the use of small-chemical molecules could have adverse effects on native cellular function. In addition, the interacting protein pairs are constitutively linked to the split-protein fragments thus the interacting pairs could potentially interfere with split-protein function through steric hinderance, complicating the design process. To broaden the scope of using split-proteins beyond studying protein-protein interactions, a fundamental change in the methods by which the two protein fragments are brought back together needs to be made.

This study examines the reassembly and restoration of split-protein functions using inteins as the mode of protein coupling and restoration. Rather than bring two split-protein fragments together through non-covalent associations, inteins ligate the protein fragments with a native peptide bond. Thus, a significant improvement and extension of reassembling split-proteins over existing methodologies can be accomplished with the use of inteins.

2.2 Inteins

Intein (**int**ernal prot**eins**) are protein elements that facilitate a unique post-translational process termed protein splicing wherein an intervening peptide sequence (the intein) auto-catalytically removes itself from a precursor protein while concomitantly ligating the flanking ends (extein, **ext**ernal prot**ein**) with a peptide bond to restore host gene function [26-28]. This remarkable process requires no exogenous energy, co-factors, or auxiliary enzymes, making inteins a self-contained modular entity. There are currently more than 480 inteins (http://www.neb.com/neb/inteins.html [29]) found in unicellular organisms from all three domains of life – archaea, bacteria, and eukarya – each sharing conserved sequence motifs that suggests an ancient evolutionary origin. Although inteins are absent from multicellular organisms, similar protein auto-processing do occur and involve intein-like domains; the most notable being the hedgehog-like proteins essential for embryonic development [26]. Inteins generally consist of two distinct domains: the self-splicing intein domain of ~135 amino acids and a homing endonuclease that is situated within the intein (Figure 2.1) [27, 28]. The function of the homing endonuclease domain is to facilitate intein gene mobility by allowing for lateral gene transfers [30-32]. Nonetheless, the homing endonuclease is non-essential to the functioning of protein splicing as is evident by the existence of mini-inteins that lack the endonuclease domain [33]. A biological role for intein and protein splicing has thus far eluded scientists. Inteins' insertion into unrelated genes and locations, many coding essential proteins, adds to its mysterious nature. For this and other reasons, inteins are often viewed as parasitic genetic elements. In addition, inteins are highly promiscuous with respect to its extein sequences; protein splicing proceeds in a foreign context wherein native exteins sequences are replaced with foreign sequences of interest. Because of this, inteins have been cleverly exploited in a number of protein engineering tools such as protein semisynthesis [13, 14] and cell-based biosensors [34, 35].

2.2.1 Protein Splicing Pathway

The protein splicing pathway consists of a series of four intramolecular chemical reactions that often rely on particular chemical side groups for efficient protein splicing to proceed [28, 36]. By this very nature, the splicing junctions between the intein and flanking exteins are often

populated with specific amino acids. For example, most inteins begin with either Ser or Cys residue and end with an Asn, while the first residue of the C-extein (+1 amino acid) is commonly a Ser, Cys, or Thr (Figure 2.1). Recent research have relaxed this requirement with the development of auxiliary ligations and desulfurization methods [37, 38].

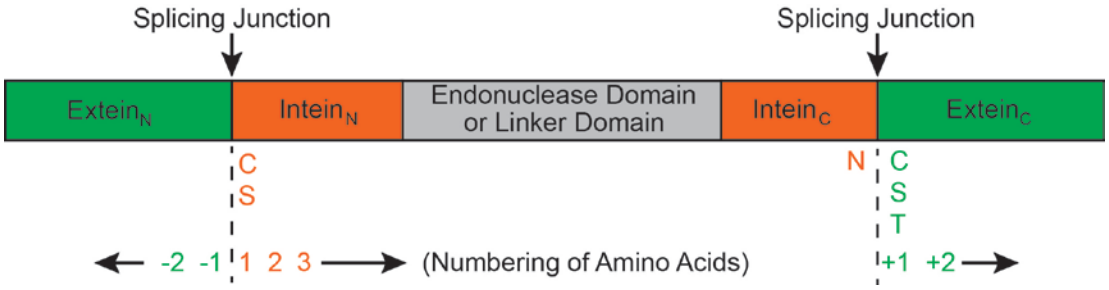

Figure 2.1 Schematic representation of general splicing intein precursor with endonuclease domain

Typical inteins (orange), shown flanking the endonuclease domain, associate to form the functional splicing domain. Mini-inteins are joined with a linker region rather than the endonuclease domain. Standard conserved amino acids near the splice junctions are shown in upper case lettering. The most common amino acids numbering system is shown at the bottom. Negative and positive signs denote N-extein and C-extein amino acids, respectively.

Regardless, the general chemical mechanism of protein splicing involves the following four steps (Figure 2.2) [28, 39-41]:

1. A linear (thio)ester intermediate is formed when Cys1 (or Ser1) undergoes an N-S (or N-O) acyl rearrangement turning the peptide bond at the N-terminal splice junction into a (thio)ester bond. As a result, the N-extein binds to the sulfur of the Cys1 (or oxygen of Ser1).

2. Cys+1 (or Ser+1, Thr+1) undergoes nucleophilic attack on the (thio)ester bond produced in step 1, forming a branched (thio)ester intermediate at the C-terminal splice junction (trans(thio)esterification).

3. The splicing then proceeds through cyclization of the Asn residue adjacent to the C-terminal splice junction, coupled with cleavage of the branched (thio)ester intermediate yielding the

excised inteins with a C-terminal aminosuccinimide residue and the two exteins joined by a (thio)ester bond.

4. A spontaneous hydrolysis of the intein aminosuccinimide residue generates the Asn amino acid while another acyl shift S-N (or O-N) of the (thio)ester linkage of the exteins forms the more stable amide bond.

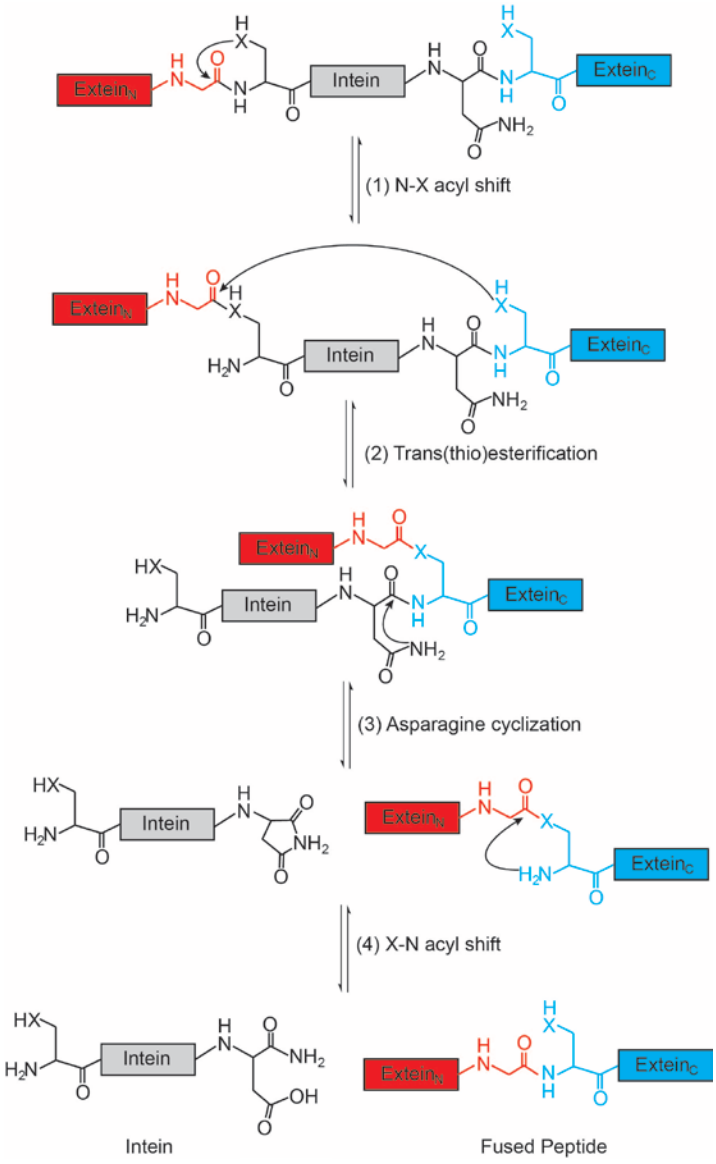

Figure 2.2 General protein splicing mechanism.

The standard chemical mechanism of protein splicing is depicted with X representing either an Oxygen or Sulphur atom of Ser, Cys, or Thr. Splicing is initiated with an acyl shift and catalytically proceeds forward with the intein acting as the catalyst.

The last reaction step is a favourable process and proceeds naturally forward. On the other hand, the first three reaction processes are highly favourable in physiological environments and do not occur at any appreciable rates. Thus, inteins not only initiate protein splicing, but exhibits catalytic properties.

2.2.2 Split Intein

Inteins are found in two splicing modes: *cis*- and *trans*- splicing mode. The vast majority of inteins are *cis*-splicing mode where the inteins either exist as a single self-splicing unit such as the GyrA mini-intein from *Mycobacterium xenopi* [42] or are linked together with a homing endonuclease such as the VMA intein from *Saccharomyces cerevisiae* (*Sce*, also the first intein to be discovered) [43]. On the other hand, a few rare inteins are naturally split and are capable of preforming protein splicing in *trans* [27, 44]. Unlike *cis*-splicing inteins, each precursor half of the naturally split *trans*-splicing inteins are located on separate genes on opposing strands of DNA, with some separated by more than 700 kbp. Most remarkable about these naturally split inteins is the intrinsic nature of the two separate precursor fragments to spontaneously self-associate and fold into its active protein splicing state. Through evolutionary optimizations, the two separate precursor fragments are thought to have better solubility and splice with higher efficiencies, particularly when fused to heterologous protein fragments [15]. *Cis*-splicing inteins could be engineered to yield artificially split inteins by removing the endonuclease domain and separating the intein halves at the endonuclease insertion positions [45-47]. However in contrast to naturally occurring split inteins, they often show marginal splicing activity due to reduced affinity for complementary intein halves and reduced solubility leading to the precursors forming aggregates with each other [17, 18]. To restore protein splicing activity, artificially split inteins are typically fused to association pairs to overcome the low binding affinity and bring the two intein halves together [48]. In addition, the low solubility and high tendency of the intein halves to aggregate requires the laborious process of denaturation and renaturation in order for the intein halves to fold into its proper three-dimensional structure [49]. These processes add complexity to the preparation of artificially split inteins. Thus, naturally split inteins are especially powerful for live cell or animal studies as they do not require interacting partners, do not require specific

chaperones to prevent aggregations, and have no cross-reactivity with any endogenous proteins in eukaryotic cells.

2.2.3 Naturally Split Intein

The naturally split intein of the gene *dnaE* from *Nostoc punctiforme* (*Npu*DnaE) is a highly efficient intein exhibiting an extraordinarily high rate of protein splicing compared to other artificial or naturally split inteins [50, 51]. It is a relatively small protein with the N-terminal precursor half (*Npu*DnaE$_N$) composed of 102 amino acids while the C-terminal precursor half (*Npu*DnaE$_C$) is composed of just 36 amino acids. The genes encoding each of these fragments are found on opposite DNA strands and are located more than 700kpbs away. The associated inteins together comprised of 14 β-strands and 2 short α-helixes with the 2 α-helixes and 10 β-strands found in the N-terminal portion while the remaining 4 β-strands are found in the C-terminal portion. In its associated form, the overall intein structure resembles a disk shape with a diameter ranging from 19-22 Å and a thickness of 14 Å arranged into a horseshoe-like fold [52] (Figure 2.3). The majority of the β-strands are relatively short ranging from 2-5 amino acids. However, there are 2 long β-strands of 8 and 10 amino acids long that are located on the N- and C-terminal intein portion, respectively. Together, these two long β-strands form an anti-parallel β-sheet situated at the bottom of the horseshoe-like fold. These long β-strands play an important role in intein association as the N-terminal β-strand contains clusters of acidic amino acids while the C-terminal β-strand contains opposing clusters of basic amino acids [53]. These highly acidic and basic amino acids participate in the electrostatic interactions that are thought to give *Npu*DnaE inteins its unique property of high affinity self-association with a K_d ~ 3 mM and an association rate of k_{on} ~ 2.8 x10^7 $M^{-1}s^{-1}$ [54]. The splicing ends are located at the centre of the hydrophobic horseshoe-like fold. This region positions the ends in optimal orientations for catalyzing one of the fastest known protein splicing known to date at a first order rate of ~1.1x10^{-2} s^{-1} with a half-time $t_{1/2}$ ~ 60 seconds [51]. In addition, one study found that the *in vivo* PTS reaction efficiency to be almost complete with yields of >98% [50]. To date, this is one of the highest splicing efficient found in inteins.

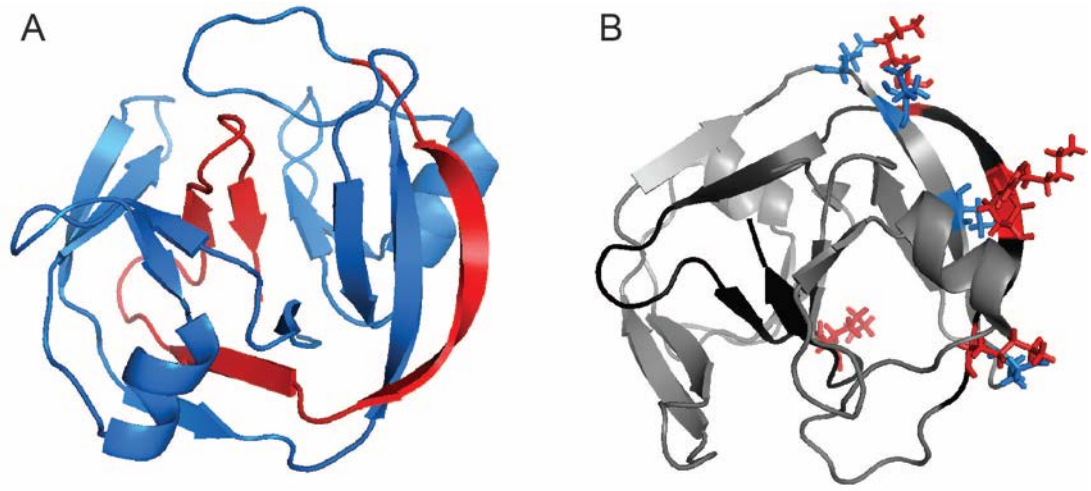

Figure 2.3 Structural modeling of *Npu*DnaE intein.

(A) Structural modeling of *Npu*DnaE intein showing the N-terminal portion (*Npu*DnaE$_N$) in blue and the C-terminal portion (*Npu*DnaE$_C$) in red. Note the two long anti-parallel β-strands forming the β-sheet as a result of intermolecular association of the intein halves. **(B)** Structure of *Npu*DnaE highlighting the intermolecular static interactions with the acidic residues shown in blue and the basic residues in red. *Npu*DnaE$_N$ is shown in gray and *Npu*DnaE$_C$ in black.

For efficient protein *trans*-splicing (PTS) to occur, it has been found that certain native extein amino acids are required near the splice junctions, particularly near the C-terminal splice junction [50, 51, 55]. For example, at the C-terminal splice junction, the three native C-terminal extein amino acids 'CFN' and restriction enzyme site 'GT' was incorporated downstream of *Npu*DnaE$_C$. At the N-terminal splice junction, the amino acids 'GS' representing a restriction enzyme site used in cloning, was included upstream of *Npu*DnaE$_N$ (Figure 2.3). Of note is that after PTS activity, amino acids 'GSCFNGT' remains with the newly formed protein.

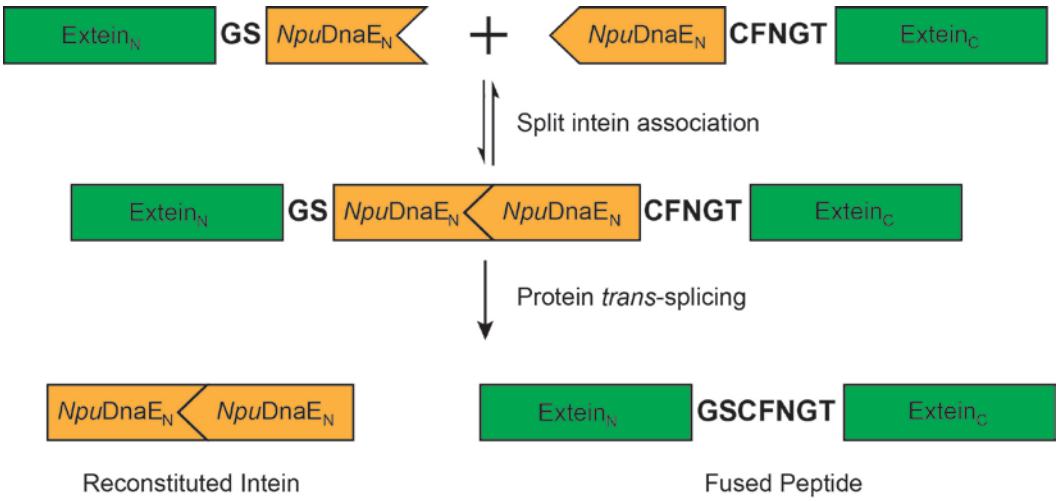

Figure 2.4 Protein *trans*-splicing of *Npu*DnaE intein.

The two precursor fragments are initially separated before they self-associate to form the functional intein for splicing.

2.2.4 Controlling Protein Splicing – Optogenetics

There are currently no known natural regulators of protein splicing - splicing occurs spontaneously after translation [37]. Nevertheless, the ability to control splicing activity is valuable as it can provide a means to post-translationally synthesize target proteins. This can potentially yield a split-protein reassembly platform that is much faster than inducible genes, can be tunable so that protein levels can be adjusted, and targeted such that protein formation can be localized to specific cellular regions or groups of cells. Towards this goal, protein splicing by intein has been regulated by temperature [56, 57], pH [58], or by the introduction of small molecule [48, 59, 60]. However, temperature- and pH-sensitive inteins are limited to biological systems that can tolerate temperature and pH shifts while small-chemical molecules often mimic 'drug-like' compounds that can potentially cause adverse effects. A particularly promising approach relies on the use of light as a regulating input signal. The use of light as a stimulus is easily controlled in both time and space offering a potentially powerful and convenient means of an input stimulus. To date, only a handful of studies have addressed the photo-control of inteins [61-64]. Three studies will be discussed that utilize photo-cleavable protection group or photodimerization domains to control protein splicing activity.

Photo-cleavable protein group disrupts protein splicing by local perturbations: The PTS activity of a naturally split intein was disrupted by the introduction of a photo-cleavable amino-protecting group [61]. This strategy is based on the incorporation of an *O*-acyl isomer at a Ser or Thr amino acid that introduces of a kink in the main peptide backbone, disrupting the normal function of the protein. A photocleavable Nvoc group (*O*-acyl isomer analogue) was synthesized on the penultimate amino acid near the splicing junction of the C-terminal DnaE intein from the cyanobacterium *Synechocystis* sp. Strain PCC6803 (*Ssp*). The proximity of the kink to the splicing junction disrupts the active site, preventing PTS activity from proceeding. UV-light exposure removes the protecting group, triggering spontaneous O to N acyl migration. This removes the kink and re-establishes the protein splicing active site.

Photo-cleavable protein group modulates protein splicing by regulating binding affinity: The PTS activity of the naturally split *Ssp* DnaE was efficiently blocked by the introduction of two photolabile groups that significantly reduced the native binding affinity of the two precursors [62]. Two photolabile Nvl groups were introduced on specific points on the peptide backbone of *Ssp*DnaE$_C$ (amino acids 19 and 31). The introduced Nvl groups prevent β-sheet formation by interfering with the required hydrogen bonding. Consequently, the binding affinities of the intein moieties are adversely affected, resulting in the arrest of PTS activity.

Activation of protein splicing by photodimerizable domains: PTS by an artificially split intein was controlled by a pair of photodimerizable domains [64]. The VMA intein from *Sce* was artificially divided as described in [48] yielding two inactive and low-binding affinity precursors. To facilitate the coupling of the inactive precursors, the plant phytochrome B (PhyB) from *Arabidopsis thaliana* was used together with the transcription factor phytochrome interacting factor 3 (PIF3) that are known to interact in a light-dependent manner. Attachment of PhyB and PIF3 to its respective VMA intein fragments yielded a photo-inducible PTS intein. In contrast to the previous two studies, an artificially split intein was used which required the use of a pair of coupling mechanism, in this case a genetically encoded phytochrome-PIF system. However in the present study, the expression of active PhyB requires the addition of a tetrapyrrole chromophore during growth, which may not be amenable to all organisms.

The use of a genetically encoded photo-dimerization domain represents a departure from previous chemical synthetic means, and may be more attractive for biological applications. This technique is commonly known as optogenetics. Optogenetics has emerged as a new class of tools that combine optical imaging technologies with genetic engineering to alter and control cellular and biological systems [65-67]. Light of varying wavelengths acting on native photo-activatable proteins have been used in a number of novel ways to interrogate biological and neural networks [68, 69]. In particular, the photo-receptive protein domains have been functionally engineered to act on effector proteins such that the photo-induced conformation of the photo-receptive protein domains regulates effector activity. Typically used photo-receptive protein domains include the phytochrome-PIF system [64, 70], channelrhodopsins [68, 71, 72], and LOV domains [73-76].

LOV domains are a class of light-activated serine/threonine kinases first discovered in phototropins that are responsible for plant phototropism [76-78]. The second LOV domain (LOV2) of phototropin-1 from *Avena sativa* is particularly well-studied and characterized [79-83] and has been used as a building block for a number of optogenetic tools [73, 74, 84]. The LOV2 domain exists in two conformational states. In the absence of light (dark state), the LOV2 domain non-covalently associates with a flavin mononucleotide (FMN) chromophore, promoting the carboxyl-terminal helical extension (Jα) to dock against the LOV2 domain in a tight conformation [80, 81]. Blue-light photoexcitation of LOV2 leads to the formation of a covalent adduct between the FMN chromophore and a conserved Cys residue. This triggers a large conformational change that unwinds and undocks the Jα helix. This FMN-Cys covalent adduct persists for several tens of seconds before decaying back to its non-covalent dark state [80-82].

2.3 Recurring Model Proteins

One of the features of a robust system is the generality of the system – the ability to be used in a number of non-native sequences. Here, a selection of diverse proteins is presented to examine its efficacy in reconstituting split-proteins. These proteins include genetically encoded Ca^{2+} indicators, cell morphology and migration regulator, and cell death protein.

2.3.1 Genetically Encoded Ca^{2+} Indicators

Ca^{2+} is a ubiquitous second messenger involved in the regulation of diverse cellular processes such as cell signalling events, immune response activation, cellular migration and cell death [85-89]. To reveal the relationship between Ca^{2+} signalling events and other cellular processes, specific Ca^{2+} indictors are used to monitor such changes. Genetically encoded Ca^{2+} indicators (GECIs) represent an exquisite set of protein-based sensors that are categorized into two main groups based on the use of dual or single fluorophores. Dual fluorophore GECIs such as cameleons [90-92] and TN-XL [93] rely on changes in fluorescence resonance energy transfer (FRET) between two complementary donor (i.e. cyan fluorescent protein) and acceptor (i.e. yellow fluorescent protein) fluorophores to produce a ratiometric output (Figure 2.4). These ratiometric measurements help mitigate artefacts related to cell movements, morphology changes, or fluorescence intensity changes. Single fluorophore GECIs such as pericam [94], R-GECO [95], and GCaMP [96] operate by emitting changing fluorescent intensity in response to differing Ca^{2+} concentrations (Figure 2.4). These GECIs typically have larger dynamic ranges and use simpler optics than dual fluorophore GECIs but are more sensitive to the aforementioned artefacts.

In this study, changes in endogenous Ca^{2+} concentration are used to test the efficacy of the reconstituted split-GECIs. These changes are elicited with either exogenous adenosine-5'-triphosphate (ATP) or uridine-5'-triphosphate (UTP), or ionomycin (Figure 2.4). Depending on the cell and receptors present, ATP or UTP acts on either P2X or P2Y purinergic receptors, or a combination of both. P2X is a dual plasma membrane spanning ligand gated ion channel. Binding of ATP or UTP to the extracellular domain of P2X brings forth a conformational change, opening the ion-permeable pore for Na^{+} and Ca^{2+} entry [97, 98]. The Ca^{2+} release mechanisms for P2Y receptors are slightly more complex. P2Y receptors are seven-transmembrane G protein-coupled receptors [99]. Binding of ATP or UTP triggers the G_q subunit to activate phospholipase-Cβ (PLCβ) that is responsible for the generation of inositol 1,4,5-triphosphate (IP3). IP3 then binds and activates IP3-receptors (IP3R) causing Ca^{2+} mobilization from endoplasmic reticulum (ER) internal stores. Replenishment of the Ca^{2+} stores within the ER lumen can occur through store-operated calcium entry (SOCE) via STIM1 and ORAI1 mechanism [100, 101]. Ionomycin enhances Ca^{2+} influx through SOCE as well [102].

2.3.2 Rho GTPases

The Rho family of GTPases is a 20 member sub-group of the much larger superfamily of Ras-like proteins. These small Rho GTPases are highly conserved molecules that primarily regulate intracellular actin dynamics such as cell migration and morphogenesis [103, 104]. Of all the members, RhoA is one of the most studied of the Rho GTPase family. RhoA is a molecular switch that cycles between the active GTP-bond and inactive and GDP-bond state to regulate the actin cytoskeleton [105, 106]. Interestingly, expression of the dominant positive RhoA (Q63L) in certain epithelial cells (e.g. HeLa cells) induces cells to acquire a distinctive dynamic non-apoptotic blebbing phenotype most likely through constitutive activation of its effector kinase Rho-associated kinase (ROCK) and phosphorylation of myosin light chain [107-109]. In cells undergoing dynamic blebbing, small blebs initially protrude out from the PM and subsequently shrink back to the PM. This is repeated throughout the whole cells with varying size, shape, and duration of the blebs that are formed. This distinctive blebbing phenotype was used to differentiate cells that have restored activity of the split-RhoA protein.

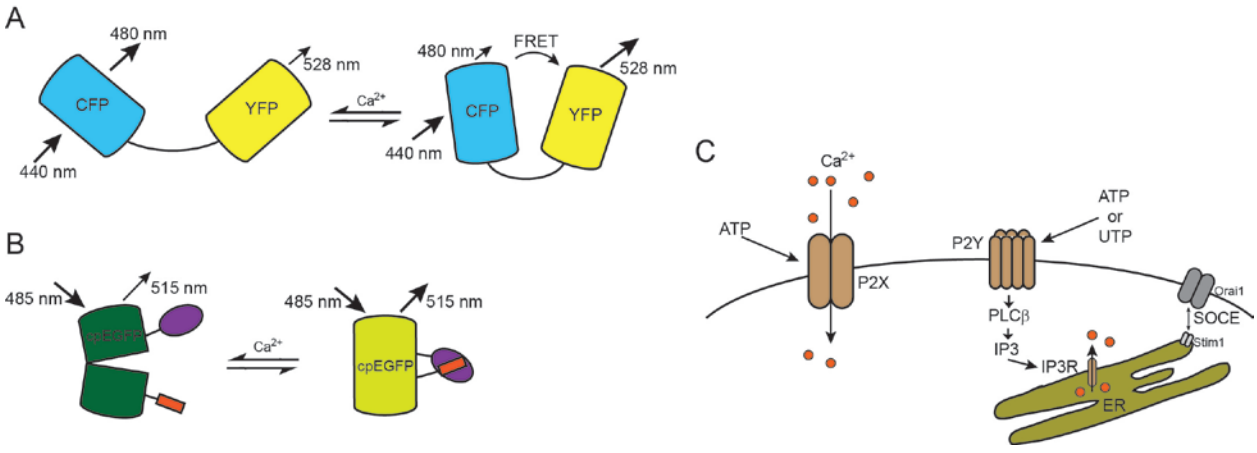

Figure 2.5 Two representative GECIs to measure induced Ca^{2+} transients in this work.

(**A**) General mechanism of FRET-based Ca^{2+} biosensor. In the presence of high Ca^{2+} environments, the two fluorophores are brought into close proximity of each other such that there

is a non-radiative transfer of energy between the excited donor (CFP) and the acceptor (YFP) resulting in strong emitted acceptor fluorescence and low donor fluorescence. In low Ca^{2+} environments, the two fluorophores are separated such that FRET is not as efficient and the donor (CFP) emits strong fluorescence while the acceptor emits lower fluorescence. (**B**) General mechanism of single fluorophore based Ca^{2+} biosensor. In low Ca^{2+} environments, a binding peptide (orange) is released from a Ca^{2+}-sensitive domain (purple). This action causes the destabilization of the cpEGFP fluorophore resulting in reduced emitted fluorescence. In the presence of high Ca^{2+} environments, the Ca^{2+}-sensitive domain binds the binding peptide, restabilising the fluorophore, and emitting strong fluorescence. **C.** Calcium transients were induced in by stimulation with ATP/UTP or with ionomycin. ATP binds P2X channel receptor eliciting direct Ca^{2+} entry from the external environment. ATP or UTP can also bind P2Y receptors which initiate a cascade of events leading to Ca^{2+} release from the ER. Ionomycin utilizes the SOCE mechanism for Ca^{2+} entry.

2.3.3 Caspase-7

Casp7 is a member of a conserved family of tightly-regulated proteases whose activation results in the dismantling of the cell machinery and the irreversible commitment to apoptosis, or programmed cell death [110, 111]. Casp7 is an effector caspase which is situated downstream of initiator caspases in the apoptosis cascade. The catalytically inactive pro-Casp7 is activated when initiator caspases proteolytically cleave a specific internal sequence to yield p20 and p11 subunits followed by removal of a short prodomain cap [112]. The p20 and p11 subunits associate to form the active Casp7. The active Casp7 cleaves a series of apoptotic substrates such as PARP and DNA-PK involved in DNA repair mechanisms, and directs cells to proceed with apoptosis [112-115]. Some specific hallmarks of cells undergoing apoptosis include shrinking and rounding of cell and loss of nuclear membrane integrity [116, 117].

3 Methods and Materials

Portions of the contents from this chapter were modified from the peer-reviewed journal article [118]:

Stanley SC Wong, Kevin Truong. Fluorescent Protein-Based Methods for On-Plate Screening of Gene Insertion. *PLOS One*. 2010. 5(12):e14274

3.1 Gene Construction – Subcloning

The creation of gene constructs utilizes a cassette-based system for the insertion of new and existing genes [119]. The cassette-based system builds upon a standard vector cassette structure such that generation of new fusion genes maintains the vector cassette structure, enabling efficient and rapid generation of gene libraries with minimal restriction enzyme management. Bacteria (T7) and mammalian (CMV) promoters are also conveniently included in the vector cassette structure providing simple expression in mammalian cells and bacteria colonies. In addition, an ampicillin resistance allows for simple bacteria selection. Here, an overview of gene construction including the various materials and methods used will be giving.

3.1.1 Materials, Reagents, and Equipment

The subcloning process, from gene construction to plasmid propagation involves a number of materials, reagents, and equipment. These will be described here.

Cassette-based system: The cassette-based system revolves around the vector pCfvtx [119], which is based on the pTriEx3 – Hygro vector (Novagen) backbone. The innate pTriEx3 MCS is replaced with the standard vector structure (Figure 3.1). This standard vector structure has three pairs of restriction enzyme sites producing compatible cohesive ends: SpeI/NheI, BamHI/BglII, and StuI/SmaI (although this pair is rarely used). These restriction sites are divided between site 2a and site 2b which flank a stop codon. Collectively, they are inserted between a start codon and a fluorescent protein gene. The fluorescent protein gene is important for screening purposes. Fluorescence does not occur from the cassette itself or in the event that gene is unsuccessfully

inserted. When a new gene (usually a PCR product) is successfully inserted using non-compatible restriction sites on either side of the stop codon, the stop codon is removed and replaced with the new gene (Section 3.1.2). This allows full transcription of both the inserted new gene and the fluorescent protein (FP). Once the new gene is inserted, the new cassette can be combined with existing cassette libraries to generate new fusion proteins (Section 3.1.3). The removal of the fluorescent protein gene is accomplished by PmeI enzyme digestion and ligation (Section 3.1.4). Other strategies for improving fluorescence screening and gene insertions can also be made (Section 3.1.5).

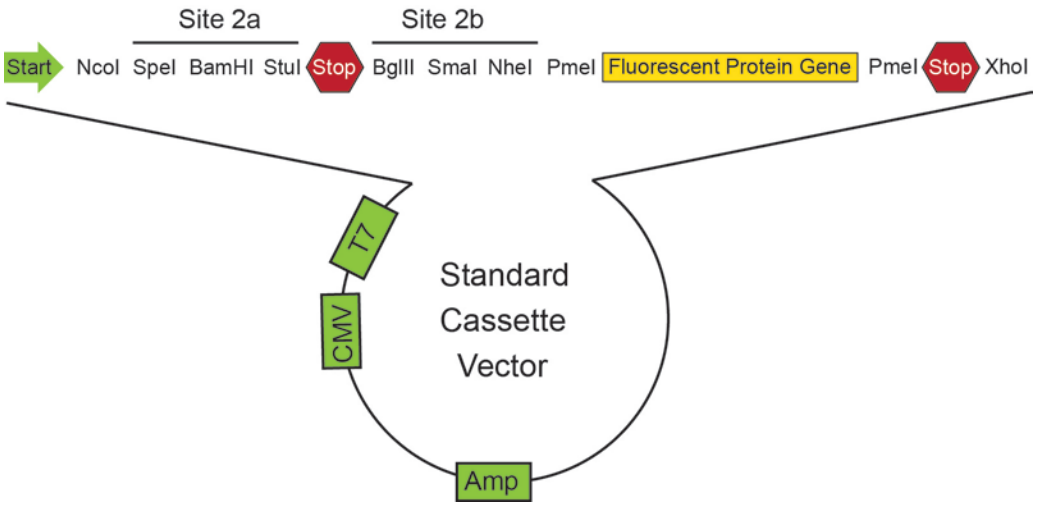

Figure 3.1 Schematic layout of the standard cassette vector and structure.

The MCS contains three pairs of compatible restriction enzyme sites divided between Sites 2a and 2b (SpeI/NheI, BamHI/BglII, and StuI/SmaI). The restriction sites on either end are unique restriction sites. The two PmeI restriction sites facilitate removal of the fluorescent protein. The CMV and T7 promoters allow bacterial and mammalian expression.

PCR reaction and purification: PCR reactions were performed in a 50 µL reaction mix prepared by combining 5 µL of 10x PCR buffer with $MgSO_4$, 5 µL of 5 mM dNTP, 0.5 µL of *Pfu* DNA polymerase (all above from Fermentas), 5 µL each of forward and reverse primers at a concentration of 0.5 µM (custom ordered from Invitrogen), 28.5 µL of RNAse/DNAse-free water (Invitrogen), and 1 µL of cDNA or plasmid source. PCR reactions took place using the

Eppendorf Mastercycler Personal. Separation of the PCR product from the reaction buffer was performed using the PureLink PCR Purification kit (Invitrogen).

Gel electrophoresis and purification: Agarose gel for electrophoresis was prepared by dissolving 1% agarose gel in 0.5x TAE buffer (both from Invitrogen) and allowed to set in a well-forming apparatus. Size reference on gel separation was provided with O'Gene DNA Ladder (range of 100bp to 10kbp; Fermentas). Gene samples were mixed with 6x Orange Loading Dye (Fermentas) before loading onto gel. To view bands after gene separation, the gel was soaked in ethidium bromide bath. Gene bands were viewed on UV transilluminator. Purification of gene samples from gel, if needed, was performed using PureLink Quick Gel Extraction Kit (Invitrogen).

Enzyme digestion and ligation: Enzyme digestion reactions were performed in a 30 µL reaction mix prepared by combining 3 µL of pre-mixed 10x buffer (provided by manufacturer), 1 µL 30x BSA buffer (all buffers from Fermentas), 1 µL each of the restriction enzyme (commonly any two of: NcoI, SpeI, NheI, BamHI, BglII, PmeI, XhoI) (NEB), and 24 µL RNAse/DNAse-free water. Digestions were incubated in the Eppendorf Mastercycler Personal. Ligation reactions were performed in 20 µL reaction mix consisting of 2 µL of 10x ligase buffer, 1 µL of T4 DNA ligate (both from NEB) and 17 µL mixture of plasmid host and insert. Ligations were incubated in the Eppendorf Mastercycler Personal.

Bacteria transformation, culturing, purification, plating, and fluorescence screening: DH5α Subcloning Efficiency chemically competent *E. coli* cells were used for transformation. Transformed *E. coli* cells were cultured in LB broth with 100 µg/mL ampicillin or 80 µg/mL kanamycin. LB broth was prepared by dissolving Luria Broth powder (Sigma-Aldrich) in distilled water and autoclaved. Extraction and purification of plasmids from bacteria culture was performed with the use of PureLink Quick Plasmid Miniprep Kit (Invitrogen). Bacteria colonies were screened on LB Agar (powder, Sigma-Aldrich) plates using a fluorescent plate reader (Illumatool Tunable Lighting System; Montreal Biotech Inc). Excitation filters with peak wavelength emissions centered at 440 nm, 488 nm, and 540 nm were used to excite the

fluorescent proteins (CFP, YFP, and mRFP, respectively) while long pass emission filters at 480 nm, 525 nm, and 560 nm, respectively were used to view fluorescence.

3.1.2 Generating a Cassette from a New Gene of Interest

The procedures for constructing cassettes from new genes of interest are presented. Here, the new genes are fused to the FP gene creating a FP-tagged construct. A summary of the process wil be given (Figure 3.2). Occasionally, FP may interfere with proper folding of the new gene. An alternative method that separates the new gene from the FP gene can be used (Section 3.1.4).

1. Forward and reverse primers flanking the gene of interest are designed to include overhangs containing restriction sites matching those on the standard vector structure. Restriction sites are chosen to include non-compatible pairs from Site 2a and 2b. The design of the primers depend on a number of factors such as annealing temperatures, length, G-C and A-T content relative to the gene of interest being amplified.

2. PCR was performed with 50 µL of reaction mixture using primers from above. Typically, PCR conditions are as follows: an initial denaturation step for 10 minutes at 95^{o}C; followed by 32 cycles of 3 minutes of denaturation at 95^{o}C, annealing for 2 minutes at 60^{o}C, and amplification at 72^{o}C for 2 minutes for every 1 kpb. After the 32 cycles, a final extension step for 10 minutes at 72^{o}C was performed. Slight variations may be needed for optimal performance.

3. Each PCR samples were mixed with loading dye and loaded into individual wells in the agarose gel. DNA ladder was also loaded into a separate well. Gene separation was performed in electrophoresis apparatus at 100V for 25 mins. Bands on gel were then prepared for UV viewing in ethidium bromide bath. Bands of the gene of interest should appear was a well-defined clear single band of the appropriate size. In the event of smeared results, wrong band size, or no visible bands, the PCR conditions should be adjusted for better performance. Continued failure may require redesign of primers.

4. Bands of interest were cut from gel using a clean stainless steel knife, making sure knife does not cut into band or touch nearby bands to reduce cross-contamination. Successfully extracted gel bands were purified with PureLink Quick Gel Extraction Kit.

5. The standard vectors (host) and PCR fragments (insert) were enzyme digested by incubating in Eppendorf Mastercycler Personal for 3 hours at 37^{o}C. Digested host vector and PCR fragments were purified with PureLink PCR Purification kit.

6. Ligation reaction mixture was prepared with digested host vector and PCR fragments from above. Typical amounts added for each were 4 µL host vector and 13 µL PCR fragments. Ligation proceeded in Eppendorf Mastercycler Personal set at 16^{o}C for 3 hours.

7. Ligation products were transformed into 20 µL aliquots of *E. coli* cells. 1-2 µL of ligation products from above was mixture with the *E. coli* cells and incubated at 42^{o}C for 1 minute to heat-shock the cells and then quickly placed back on ice for 5 minutes.

8. Transformed *E. coli* cells were transferred to 1 mL of LB broth with ampicillin and incubated for 3 hours at 37^{o}C in a shaking incubator at 200 rpm. After 3 hours, 60 µL of the LB broth was transferred onto a warm LB agar plate and evenly spread across the agar surface with a sterile glass spreader. Plates were labelled and left overnight in incubator at 37^{o}C.

9. Plates are examined for colonies and screened with the plate reader the following day. Successfully ligated and transformed colonies expressing fluorescence were picked off with a sterile pipette tip and cultured in 4 mL of LB broth with ampicillin overnight in shaker incubator at 37^{O}C. Sometimes, colonies may be non-fluorescent and require additional incubation at room temperature or at 4^{o}C for 24-48 hours, presumable due to low folding efficiencies. Persistent non-fluorescent colonies or absence of colonies may require backtracking to step 6 or further.

10. Plasmids were extracted and purified from the cultured *E. coli* cells with the PureLink Quick Plasmid Miniprep kit. Purified plasmids were size checked by performed electrophoresis with digestion enzymes flanking the inserted gene of interest (typically NcoI and XhoI). Further analysis was obtained by sequencing (TCAG , Toronto).

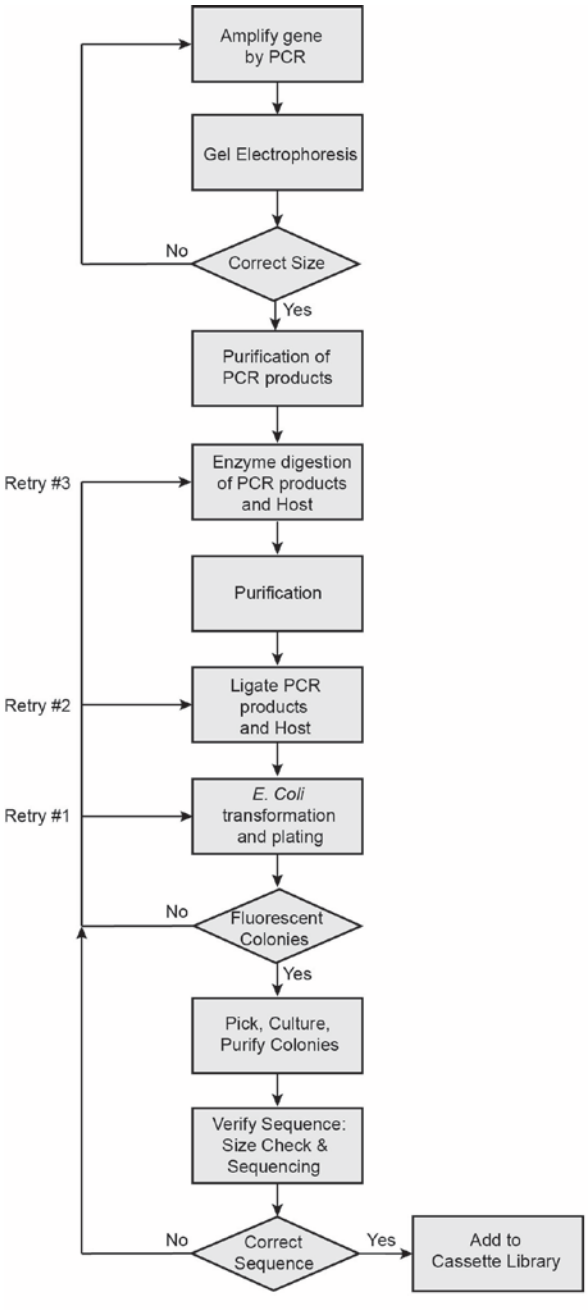

Figure 3.2 Flowchart for the insertion of a new gene into the standard cassette vector

3.1.3 Generating Fusion Proteins from Two Cassettes

The standard vector structure enables simple combination of two cassettes-systems, cassette A and cassette B, to generate fusion proteins (Figure 3.3). There are two possible paths to generate the fusion protein, where either cassettes can serve as the host vector or insert fragment. The determination of which is the host vector and insert fragment is largely governed by the fluorescent proteins present and/or further downstream genetic manipulations. Generally, to ease fluorescence screening, the cassette with the FP gene removed is designated the host vector while the cassette with the intact FP gene is the insert fragment. If the two cassettes have different FP genes present (i.e. one cassette has CFP and the other has mRFP), removal of the FP gene from one cassette is not necessary as screening can be performed using the two different fluorescence. A summary of the process will be given (Figure 3.4).

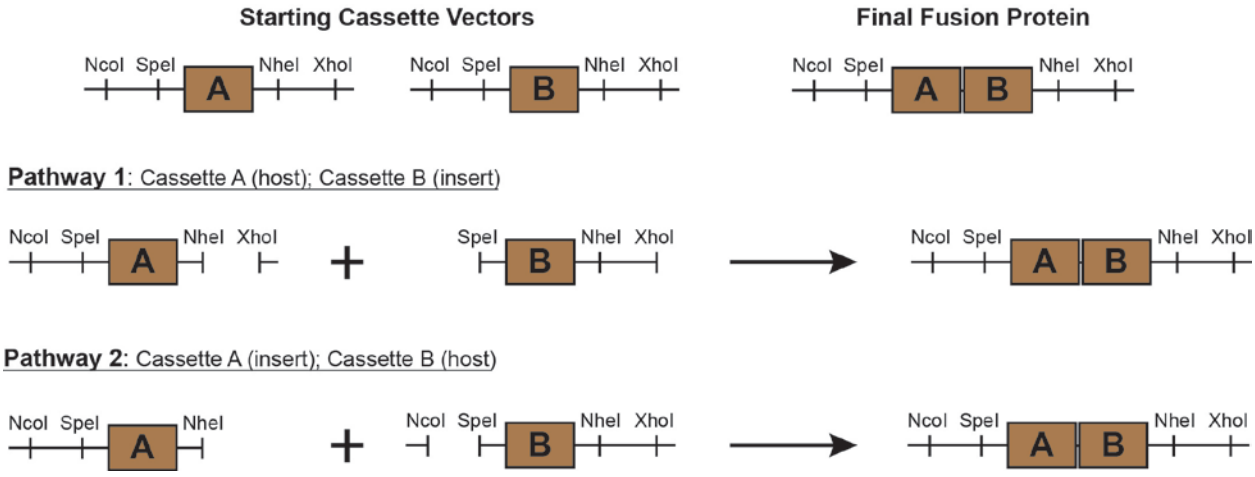

Figure 3.3 Two methods of creating fusion proteins from cassette.

There are two pathways to create fusion protein A-B from cassette A and cassette B. In Pathway 1, cassette A is the host vector while cassette B is the insert. In Pathway 2, cassette B is now the host while cassette A is the insert. For simplicity, representative restriction sites are shown. Restriction sites that have been enzyme cut are shown as ending vertical bars. Note, in both cases, the fusion cassette A-B maintains the standard vector structure.

The procedures to generate fusion protein A-B are presented below using representative restriction enzyme sites NcoI, SpeI, NheI, and XhoI, although other sites can be used. Fusion protein B-A can be created in the reverse order.

1. Pathway 1: Cassette A serving as the host vector is digested with restriction enzymes NheI and XhoI while the insert fragment is obtained from cassette B by enzyme digestion with SpeI and XhoI. Pathway 2: Cassette B serves as the host vector and is enzyme digested with NcoI and SpeI while the insert fragment is obtained from cassette A by enzyme digestion with NcoI and NheI.

2. Gel electrophoresis was used to separate insert fragments from cassette vectors, or vice versa. The much larger cassette vectors is located as the top-most band while the insert fragment band is below it. Required bands are extracted and purified as previously discuss (Section 3.1.2).

3. Cassette vectors and insert fragments are ligated, transformed, and screened similar to steps 5-10 of Section 3.1.2.

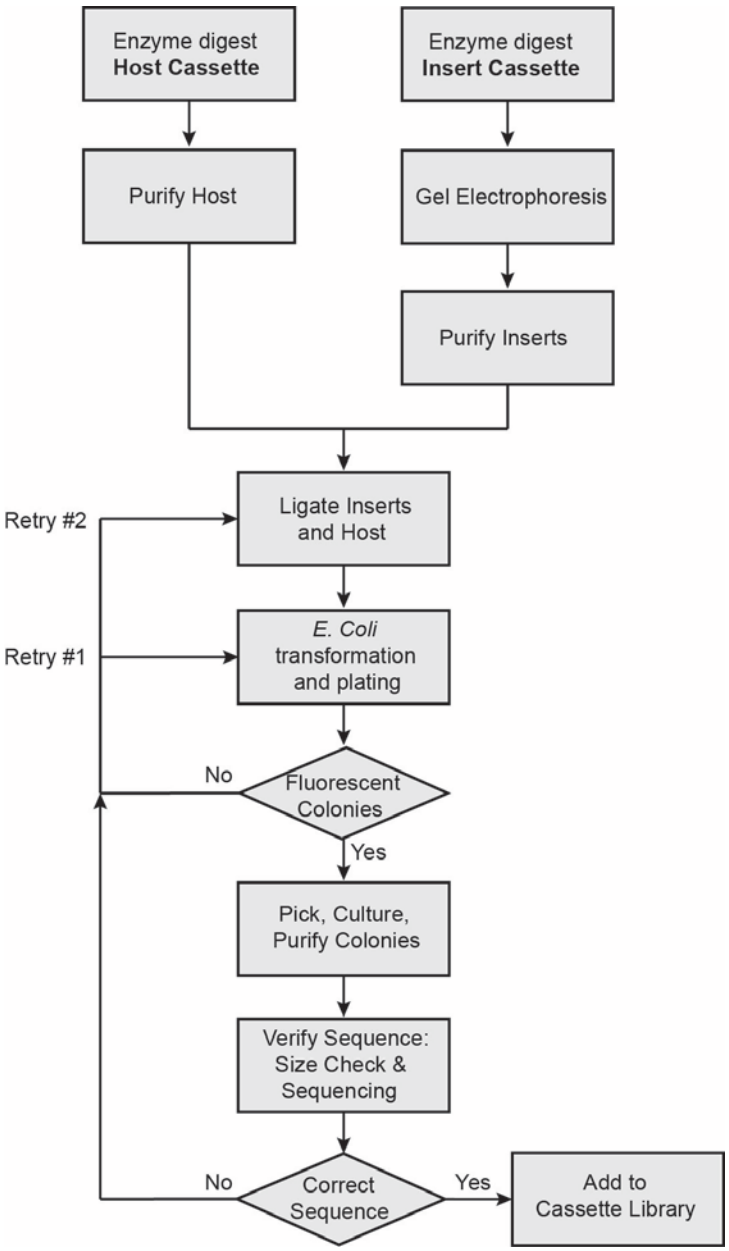

Figure 3.4 Flowchart for generating a fusion protein from two cassettes (previous page).

The generation of a fusion protein begins with two cassettes: one non-fluorescent cassette and one fluorescent cassette.

3.1.4 Removing Fluorescent Protein Gene

Often times, the FP gene is removed to facilitate fusion protein construction or for experimental studies. The PmeI enzyme restriction sites flanking the FP gene greatly eases removal of the FP gene without altering the standard cassette structure. The procedures are presented below (Figure 3.5).

1. The cassette vector is enzyme digested with PmeI and purified with PureLink PCR Purification Kit.

2. Self-ligation and transformation procedures are similar to steps 6-8 of Section 3.1.2. For self-ligation, 17 µL of the digested and purified vectors was used.

3. Plates were then screened for non-fluorescent colonies designating cassette vectors with FP gene removed. A possible issue here is that a FP gene may be re-inserted into the cassette vector in the reverse orientation, resulting in non-fluorescent colonies. However, the possibility of this occurring relative to the self-ligation is miniscule and normally do not present a problem. Nevertheless, size check can be performed to dispute suspicions.

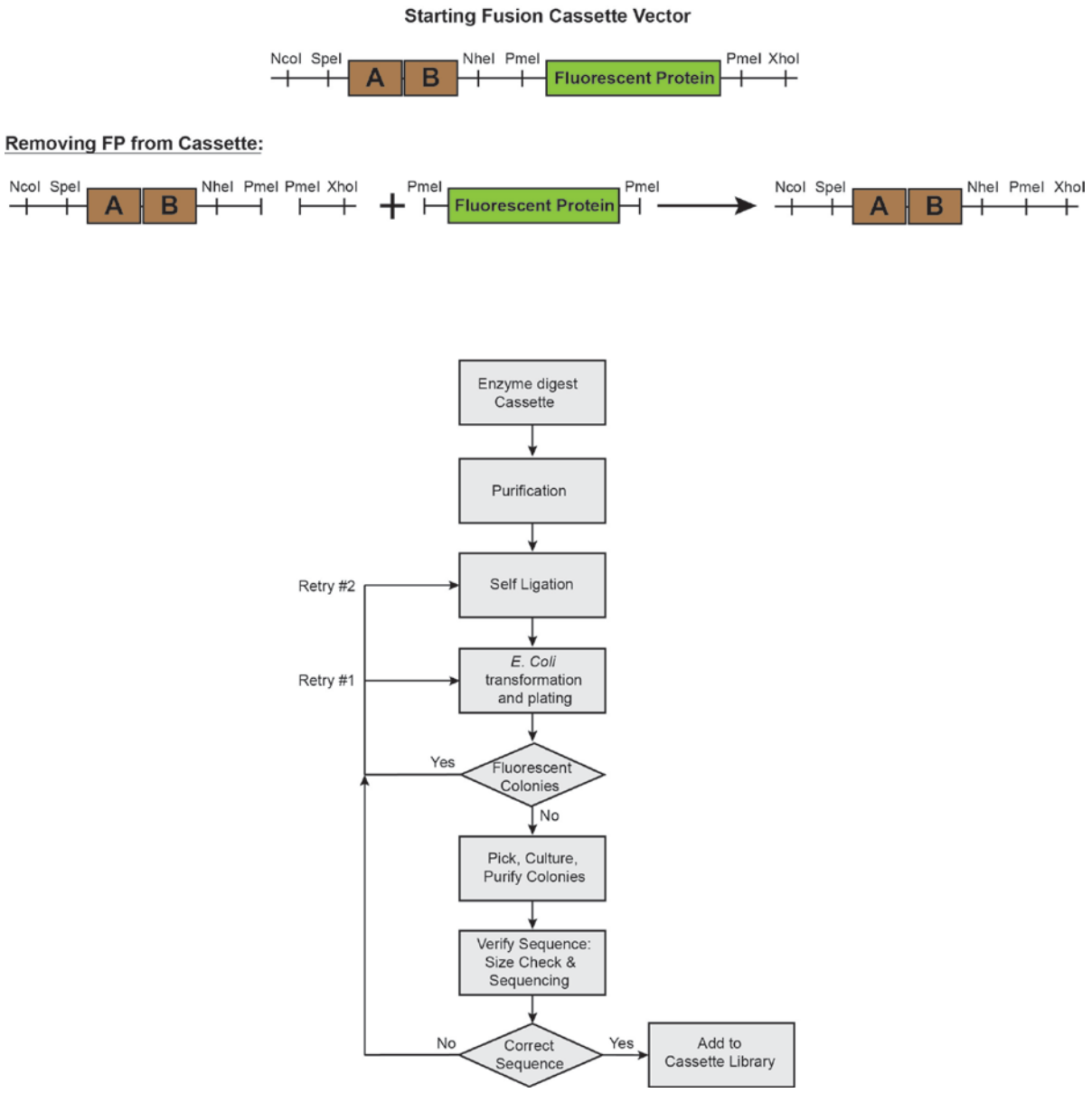

Figure 3.5 Removing fluorescent protein from cassette and flowchart.

To remove the fluorescent protein, PmeI restriction sites are enzyme cut and self-ligated. Flowchart of the procedure is also shown.

3.1.5 Other Strategies for Improving Fluorescent Screening and Gene Insertion

While fluorescence screening vastly assist in generating cassette libraries and fusion proteins, it may be limited by the folding stability of the inserted gene as proper folding of the downstream

fluorescent protein is directly related to the folding robustness of fused upstream proteins [120, 121]. Also, since gene cloning is performed in *E. coli* which may lack required chaperon proteins, the translated protein may mis-fold or aggregate in the host organism [122]. These factors adversely affect the folding of the fluorescent protein reporter can arise especially when working with new genes with unknown properties. To address these issues, three alternative methods can be utilized to improve fluorescence screening while maintaining the standard cassette structure (Figure 3.6 and Appendix 2).

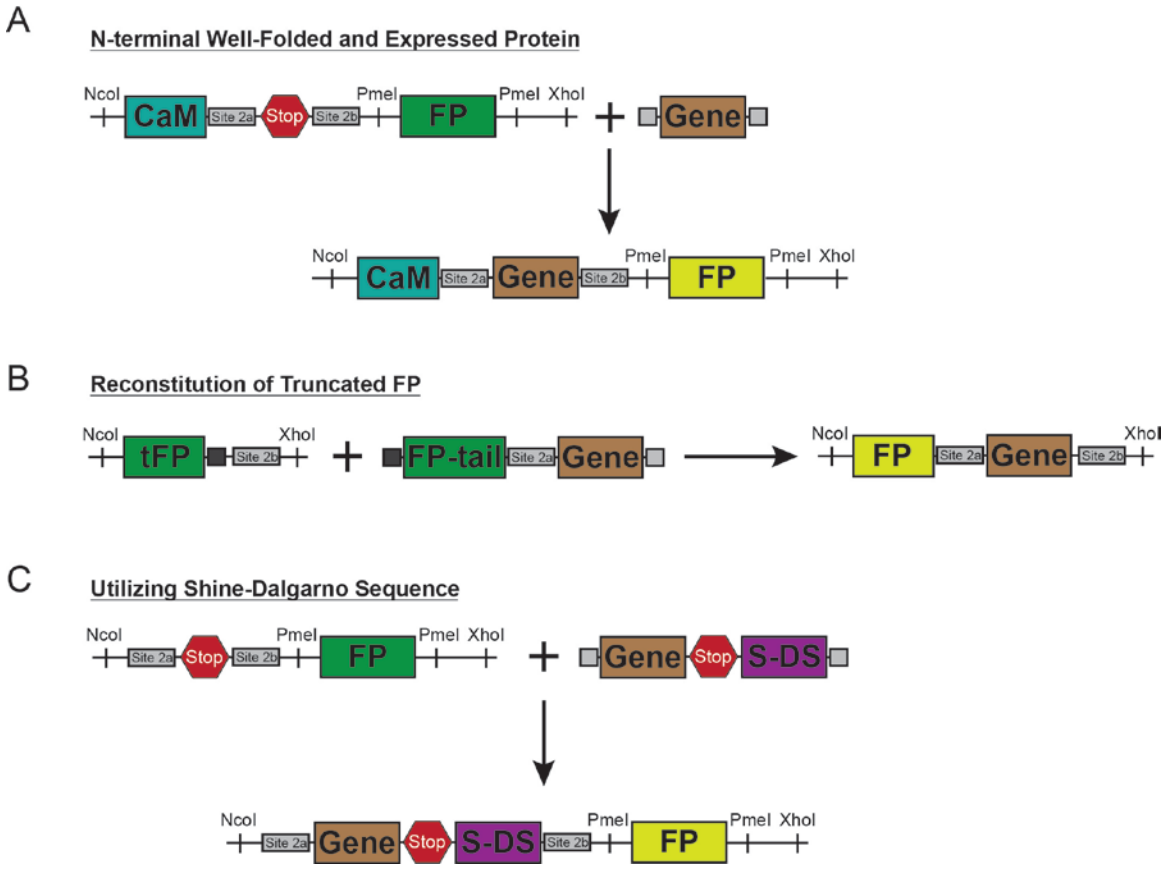

Figure 3.6 Strategies for improving fluorescent screening

(**A**) Emitted fluorescence from fluorescent protein can be improved with an N-terminally situated well-folded and expressed protein such as CaM. The insertion of a gene removes the Stop codon. (**B**) Abolished fluorescence from a truncated fluorescent protein can be restored by inserting a gene with the truncated FP portion returned. (**C**) The use of a Shine-Dalgarno sequence separates the FP expression from the inserted gene. Compatible restriction sites are shown in grey square boxes.

Use of N-terminally situated well-folded and expressed protein: To improve fluorescence screening, an N-terminally situated protein that is well-folded and expressed, such as calmodulin (CaM), can be used to promote folding and enhanced expressing of the gene insert. Since a poorly folded protein negatively affects downstream proteins, a well-folded protein might facilitate increased protein expression levels by overcoming translation initiation. The result is increased production of downstream fluorescent proteins allowing for improved fluorescence detection. In previous studies, fusion of CaM to the FP produced noticeably brighter fluorescence than is typically seen. Drawing from these observations, a new cassette vector was created with CaM situated between enzyme restriction site NcoI and Site 2a (Figure 3.6). The standard vector structure is maintained so that new gene insertion (Section 3.1.2), cassette combinations (Section 3.1.3), and FP removal (Section 3.1.4) can be performed as described above.

Reconstitution of N-terminally situated truncated FP: Since N-terminally situated proteins has the best opportunity for proper protein-folding of the reporter FP, another strategy to enhance on-plate fluorescence screening is creating gene inserts such that its integration will reconstitute a truncated N-terminally situated FP to the inserted gene. Since the fluorescence of a fluorescent protein is dependent on the proper formation of the β-barrel structure [123-125], fluorescence can be abolished by truncating the FP at the C-terminus (tFP) to disrupt the β-barrel structure. By designing the insert gene fragments to include the truncated portion of the FP, the inserted gene fragments complete the β-barrel structure of the truncated FP allowing restoration of fluorescence (Figure 3.6).

Separating the dependence of fluorescence screens on protein-folding efficiency of inserted protein: To eliminate the dependence of fluorescence screens on the protein-folding efficiency of the fusion protein, protein translation of the reporter FP can be separated from the protein translation of the inserted gene. This can be accomplished with the use of a stop codon, Shine-Dalgarno sequence (S-DS) [126], and a start codon that is designed into the reverse primer sequence of the inserted gene in the respective order (Figure 3.6). Upon insertion of the gene fragment, the stop codon halts translation after the inserted gene is translated while the Shine-Dalgarno sequence and start codon together initiate translation of the reporter FP. This

effectively separates the expression of the reporter FP from any folding influences of the inserted gene.

3.2 In Vitro Protein Analysis – SDS-PAGE

SDS-PAGE was performed on some protein constructs for protein analysis. *E. coli* cells were the organism used in the production of proteins. Proteins of interest were extracted as crude cell lysates by mechanical cytolysis of the cell membrane. No purification of proteins of interest from the crude cell lysate was required as proteins of interest are fluorescent and thus SDS-PAGE gel can be viewed by fluorescence [127]. Procedures for the obtaining, extracting, and performing SDS-PAGE are described.

1. Plasmids containing gene of interest were transformed into competent *E. coli* cells and cultured overnight in 4 mL of LB broth containing ampicillin in a shaker incubator set at 200 rpm at 37^{o}C.

2. Cells were retrieved by transferring LB broth and cells into a 1.5 mL Eppendorf tube and centrifuging at 13,200 rpm for 2 minutes and discarding the broth. This was repeated until all cultured LB broth has been used and cells are in a single pellet.

3. The cells were then washed three times by re-suspending the pellet in 1 mL buffer solution of 50 mM Tris and NaCl and centrifuging and discarding buffer solution. After the final wash, 300 µL of Tris/NaCl buffer was used to re-suspend cells.

4. Proteins were harvested from cells by sonication with a Branson Sonifier Model 250 (Thermo-Fisher Scientific) for 1 minute at an output of 2 and 30% duty cycle.

5. After sonication, cell suspensions were centrifuged at 13,200 rpm for 3 minutes to separate the insoluble cell components from the supernatant containing the protein of interest. The fluorescence brightness of the supernatant can be compared to the pellet at this point to

determine if sufficient proteins have been released from the cells. Otherwise, further sonication can be done.

6. The fluorescent supernatant was transferred to a clean 1.5 mL Eppendorf tube.

7. Supernatant concentration was determined by qualitative comparison of fluorescence intensity with a set of fluorescent protein solution of known concentrations. Supernatant concentrations were adjusted for optimal SDS-PAGE analysis by diluting with Tris/NaCl buffer solution.

8. Preparation of SDS-PAGE was done according to manufacturer's protocols (Invitrogen). Samples were prepared with the addition of NpuPage 1x LDS sample buffer and 1x reducing agent. Typically, 20 µL of samples were prepared and heated at 70^{o}C for 10 minutes before loading into gels. Gels were pre-cast 1.5 mm thick 4%-12% Bis-Tris 10-well gels and used with 1x SDS-MOPS buffer for electrophoresis. Electrophoresis apparatus was powered by Owl EC105 power supply (Thermo-Fisher) at 120V for 90 minutes or until desired separation. PageRuler protein ladder (Fermentas) was used for size reference.

9. After SDS-PAGE electrophoresis, gels were removed from plastic casing and gently washed with distilled water and set on plastic wrap. Fluorescent proteins were viewed with fluorescence plate reader (Illumatool system) with appropriates filters. Images were taken with Canon Powershot A350 digital camera.

10. Gels were stained overnight by soaking in PageBlue Coomassie blue-based stain (Fermentas) to visualize protein ladder. Images were taken with Canon Powershot A350 digital camera and compared with fluorescent proteins from above.

3.3 Cell Culture and Live-Cell Imaging

Live-cell imaging provided the primary means to test and characterize proteins while closely monitoring cellular events. Cell culture reagents, equipment, and procedures will be described, followed by imaging equipment and procedures.

3.3.1 Cell Culture Reagents, Materials, and Equipment

In this work, several standard cell lines were primary used for live-cell imaging. They were COS-7, CHO, HeLa, and HEK293.

Cell Culture Reagents: Cells were cultured and maintained using Dulbecco's Modified Eagle Medium (DMEM) supplemented with 25 mM D-glucose, 1 mM sodium pyruvate, 4mM L-glutamine and 10% FBS (Invitrogen). To combat contamination, cell culture medium was also supplemented with 10 mL/L penicillin-streptomycin solution (Sigma-Aldrich). Cells were passaged by first washing with Phosphate Buffer Saline (PBS) (Invitrogen) and using 0.05% trypsin-EDTA (Sigma-Alrich). Cell transfections were performed with Lipofectamine 2000 reagent (Invitrogen). Cells were stored with pre-mixed cell freezing dimethyl-sulfoxide (DMSO) and methylcellulose medium (Sigma-Aldrich). PBS supplemented with 1 mM $CaCl_2$ and 1 mM $MgCl_2$ (Invitrogen) was also used for imaging in some cases. To stimulate cells for experimentation, the following stimulatory reagents were used: ionomycin (1 µM; Sigma-Aldrich), ATP and UTP (10 µM; Fermentas), $CaCl_2$ (1 mM; Sigma-Alrich), and EDTA (2 mM; Sigma-Alrich). Stimulants were prepared as 10X stock solution and added to cells to the final concentration as given.

Cell Culture Material and Equipment: Cells were maintained in uncoated screw-top T-25 flasks (Sarstedt) in a HeraCell incubator (Thermo-Fisher) maintained at 37^{o}C and 5% CO_2. A Class IIA biosafety cabinet (Thermo-Fisher) was used for all live-cell preparations. For imaging, cells were transferred to uncoated 35 mm glass bottom wells with 14 mm glass coverslips at thickness No. 0 (MatTek).

3.3.2 Procedures for Cell Culture, Transfection, and Storage

Cells were cultured from frozen stocks and continuously maintained in flasks. Passaged cells can either be used to continue propagation of cell line, used for storage, or prepared for transfection.

1. Complete DMEM culture medium was pre-heated to 37^{o}C in incubator prior to culturing cells from frozen stock. Frozen cell stocks stored at -80^{o}C are then slowly thawed in incubator at 37^{o}C. Once complete DMEM culture medium has warmed up, 5 mL and thawed cells are added to T-25 flask and gently rocked back and forth to evenly disperse cells on the bottom flask surface. Cells are then set back in incubator and allowed to grow. Depending on the cell line, cells should be 80-100% confluent in the next 1-3 days.

2. Once cells are >90% confluent, cells are passage to maintain health and growth of cells. Cell culture medium is removed from cells and 5 mL of PBS is gently added to cells with pipette and rocked back and forth to wash cells. Remove PBS from cells.

3. 1 mL of trypsin was added to the washed cells and incubated for 3 minutes at 37^{o}C. After 3 minutes, flask was gently tapped to detach cells from flask surface and transferred to a clean 1.5 mL Eppendorf tube.

4. Cells were centrifuged at 2,000 rpm for 3 minutes to pellet cells. Left over trypsin solution was discarded and cells re-suspended in 100 µL of complete DMEM.

5. To maintain cells, a clean T-25 flask was loaded with 5 mL of pre-heated complete DMEM culture medium. Cells were seeded on the flask at a dilution ratio of 1:10 (i.e. 10 µL of the 100 µL re-suspended cells). This allows cells to grow for 3-5 days before the next passage.

Re-suspended cells can also be used to seed glass-bottom wells for transfection.

1. Glass bottom wells are filled with 2 mL of pre-heated complete DMEM and cells are seeded at a dilution ratio of 1:8 and grown overnight in incubator. Cells should be ~90% confluent the next day.

2. For transfection the next day, transfection solution was prepared according to manufacturer's protocols. Briefly, 1 µg of purified plasmids were mixed with 100 µL of DMEM (without FBS) in an Eppendorf tube. In another Eppendorf tube, 100 µL of DMEM (without FBS) was mixed with 7 µL of Lipofectamine 2000. Both tubes were incubated at room temperature for 5 minutes.

3. After 5 minutes of incubation, the contents of one tube was transferred to the other tube and gently mixed. The combined solution was incubated at room temperature for another 20 minutes.

4. After the second incubation, 800 mL of pre-heated DMEM (without FBS) was added to the combined solution. At the same time, cell seeded glass bottom wells were removed from incubator and culture medium was removed from cell and washed with 1 mL of PBS.

5. Transfection mixture was added to cells and incubated in the incubator for 3 hours. After 3 hours, transfection mixture was removed and 2 mL of pre-heated complete DMEM was added to cells and incubated overnight in incubator set at $37^{o}C$ and 5% CO_2. For experiments requiring overnight photostimulation, the transfected cell wells were place on top of an iPod for periodic illumination. For experiments requiring complete darkness, transfected cell wells were covered with aluminum foil before placing into the incubator. Transgenes were usually expressed 18-24 hours after transfection.

Generally, any remaining re-suspended cells not used for transfection are prepared for storage. Before cells are stored in -80^{o}C freezer, they are first frozen in Styrofoam boxes at -20^{o}C overnight and then transferred to -80^{o}C freezer for long term storage. This allows the cells to freeze slowly and prevent ice crystal formation within the cell.

1. Remaining re-suspended cells are divided between two screw-top cryovials (Diamed) and 0.5 mL of pre-warmed cell freezing medium are drop-wise added to each cryovial.

2. Cells are gently mixed to distribute cell freezing medium across cells.

3. Cryovials are labelled and set in Styrofoam boxes in -20^{o}C freezer overnight and then moved to -80^{o}C freezer for permanent storage.

3.3.3 Live-Cell Imaging Equipment

Cell imaging was conducted on an inverted IX-81 microscope (Olympus) with a Lambda DG4 xenon lamp source (light intensity at 25 mW/mm^2) and QuantEM 512C CCD camera. The microscope was fitted with 10x (air) and 20x, 40x, 60x, and 100x (oil immersion) objective lens. Excitation filters (Semrock) for CFP, YFP, and mRFP were 438/24 nm, 500/24 nm, and 630/60 nm, respectively. Emission filters were 482/32 nm, 542/27 nm, 630/60 nm, respectively. Bleed through for the filter sets are described in Appendix D. FRET imaging was administered with a dual-band filter for CFP and YFP (480/30 nm and 535/30 nm, respectively) with a Photometrics DV2 beamsplitter to monitor both fluorescence channels simultaneously. Microscope data acquisition was provided by MetaMorph Advanced (Olympus), ImagePro-Plus (MediaCybernetics) and NIH ImageJ software. Long term imaging experiments were done in microscope stage incubator set at 37^{o}C (Solent Scientific) with CO_2 provided by a pound of dry ice. Overnight illumination was provided by an iPod 3 programmed to flash blue light (intensity of approximately 1 mW/cm^2) for 1 second at intervals of 30 seconds.

3.3.4 Imaging Procedures

Depending on the experiment, culture media may be removed from transfected cells and replaced with PBS supplemented with $CaCl_2$ and $MgCl_2$ to prevent cells from drying out. In other cases, the culture media was left with the cell. Different imaging procedures were employed for different experiments. They are outlined as follows.

1. In experiments were stimulants are added, the desired amounts are added dropwise to the glass-bottom wells away from the imaging region. This minimizes perturbation to the cells being imaged. In addition, a 5 minute initial control period was obtained.

2. To obtain the dynamic range, the ratio between the highest $[Ca^{2+}]$ measured and the lowest $[Ca^{2+}]$, cells were grown in culture media supplemented with 1 mM Ca^{2+} and cells were allowed to equilibrate at baseline $[Ca^{2+}]$. After UTP- (or ATP) induced Ca^{2+} transient experiments, cells were stimulated with ionomycin to induced enhanced Ca^{2+} influx (i.e. elevated cytoplasmic $[Ca^{2+}]$). This was followed by the addition of EDTA to chelate the Ca^{2+}.

3. In FRET experiments, images were captured at 1 frame per second. Regions of interest were defined and fluorescence intensity measured.

4. For light-activated experiments, cells were wrapped in aluminum foil to prevent after transfection to prevent premature photostimulation.

3.4 Data Analysis

In this work, two methods were used to analyze results: statistical comparison analysis and fluorescence signal co-localizations.

3.4.1 Statistical Comparison Analysis

Two data sets were compared using unpaired Student's t-test to determine significant differences. Two-tailed *P* values less than 0.05 were considered significant. Data is usually presented as mean ± standard deviation unless otherwise noted.

3.4.2 Quantification of Co-Localization

The light-activated translocation kinetics of cytoplasmic YFP to PM-bound CFP was quantified by Pearson's Coefficient (PC). PC is a measurement of linear dependence between two variables and ranges in value from -1 to +1, where -1 represents perfect exclusion, 0 means no relationship, and +1 denotes perfect correlation. To determine PC, JACoP plug-in for Image J [128] was used to calculate the time-lapsed relationship between YFP and CFP. To minimize background noise and improve accuracy, images were cropped to cell membrane edge before analysis was performed. Co-localizations for overnight experiments were qualitatively and quantitatively determined. For quantitative measurements two other methods methods, PC with Costes' automatic thresholding and Van Steensel's cross-corelation function (CFF), were used in addition to PC.

3.4.3 Generation, Imaging and Tracking of Transgenic *C. Elegans*

In this study, transgenic *C. Elegans* was used as an animal model for some experiments. All culturing procedures and experimentations were conducted by the collaborators (Ippei Kotera and Hiroshi Suzuki). Thus, only a brief description of the generation, tracking and imaging of *C. Elegans* will be provided.

Generation of transgenic *C. Elegans*: *C. Elegans* were cultured and propagated on Nematode Growth Medium agar plates and fed with a strain of *E. Coli* that grows slowly and provides

nutrients without overgrowing them. *C. Elegans* were kept constant at ~20^{o}C for optimal growth. Transgenic *C. Elegans* were created by injecting plasmids at a concentration of 20 ng/µL into the distal gonads of *C. Elegans* to induce extra chromosomal expression of the gene of interest. All introduced genes were under the control of the myo-2 promoter for expression in pharyngeal muscle cells.

Imaging and tracking of *C. Elegans*: Fluorescence imaging of *C. Elegans* was obtained by a fluorescence filter cube with an excitation filter (470/40 nm, Chroma), a long pass dichroic mirror (495 nm, Chroma), and an emission filter (525/50 nm, Chroma). An objective lens (MVPLAPL 2XC NA=0.5, Olympus) and a zoom lens were combined to give 12.5x magnification power. Exposure time was 10 ms for Ca^{2+} imaging and 200 ms for still images (MetaMorph v7.14, Molecular Devices). In order to simultaneously track and image moving *C. Elegans*, a custom-made tracking system scripted in MATLAB (v7.10, MathWorks) was arranged to control a fluorescent microscope (MVX10, Olympus), a motozies stage (99S106, Ludl), a stage controller (MAC5000, Ludl), a CCD camer (CoolSNAP HQ2, Photometrics), and a light source (exacte, X-cite). Fluorescent animals with uniform expression pattern among the pharyngeal muscles were selected for imaging. Transgenic *C. Elegans* were submerged in 1 mg/mL serotonin solution (H7752, Sigma) for 10 minutes to induce feeding behaviour and then transferred to a 2% agarose (15517 – 022, Invitrogen)pad with a buffer containing 50 mM NaCl, 10 mM HEPES-NaOH, 1 mM $CaCl_2$ and 1 mM $MgSO_4$ at a pH of 7.1 for imaging.

4 Reassembly of Split GECIs

The content of this chapter was modified from the following peer-reviewed journal article [129].

Stanley SC Wong, Ippei Kotera Evan Mills, Hiroshi Suzuki, Kevin Truong. Split-intein mediated re-assembly of genetically encoded Ca^{2+} indicators. *Cell Calcium*. 2012. 51:57-64

4.1 Chapter Introduction and Aims

Two representative GECIs, one each from FRET-based dual fluorophore and intensity-based single fluorophore, were chosen to test the reassembling behaviour of *Npu*DnaE intein. At the time of this study, *Npu*DnaE inteins was generally used as a protein ligation tool for fusing simple discrete protein domains [50, 51] or for directing site specific chemical modifications to proteins [52], but has not been applied or shown to be effectively in more complex bioengineered proteins such as GECIs. The FRET-based biosensor, TN-XL, was chosen as the first experimental subject to test for the effectiveness of *Npu*DnaE intein in restoring function to more complex split proteins, specifically split-GECIs. Some design strategies pertaining to TN-XL were laid out and those insights were extended to the designing of the single fluorophore intensity-based biosensor, GCaMP2, and verified in an animal model. In addition, a simple and efficient live-cell based *in vitro* assay for PTS activity was developed which will be used throughout this study to test for the presence of PTS activity.

For this chapter, the following three specific aims were established to achieve the overall research goal:

1. Develop an *in vitro* visual assay to test for the presence of protein *trans*-splicing by observing co-localization and translocation.

2. Reveal the capability of *Npu*DnaE intein to splice with both the splicing and non-splicing ends fused to exogenous polypeptides.

3. Confirm functionality of *Npu*DnaE intein in reassembling and restoring function to GECIs.

In this chapter study, several engineered fusion proteins were created to address the specific aims and study PTS activity by *Npu*DnaE intein (Figure 4.1). All engineered proteins were created in accordance with the gene construction methodologies as presented in Chapter 3.

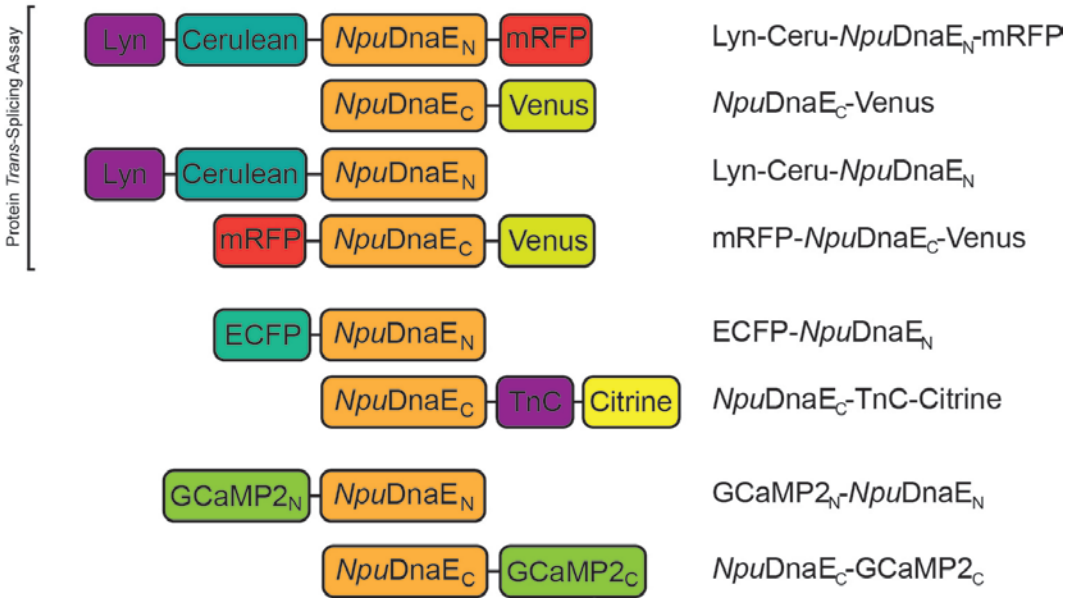

Figure 4.1 Outline of the fusion proteins used in this chapter.

Block diagrams of the fusion protein are shown. The names of the fusion proteins used in this chapter are also given.

4.2 Results

This section will describe the experimental results. First, the visual fluorescence assay will be developed beginning with a brief description of the strategy (section 4.2.1) followed by the validation and characterization in mammalian cells (section 4.2.2). PTS activity is then applied to TN-XL (section 4.2.3) and GCaMP2 (section 4.2.4) Ca^{2+} biosensors. Finally, the split GCaMP2 biosensor is tested in the animal model *C. Elegans* (section 4.2.5).

4.2.1 General Strategy of the Visual Assay

Protein *trans*-splicing by *Npu*DnaE inteins can be visually assayed by observing distinct subcellular localizations of fluorescent proteins. A fundamental property of PTS activity (or protein splicing in general) is that after the active intein catalyzes the ligation of the flanking extein sequences (host protein), the intein subsequently frees itself from the newly ligated extein. By exploiting this property, it was speculated that by tracking the individual components (i.e. the flanking exteins and intein), the occurrence of PTS activity can be assayed by examining how the components are rearranged before and after PTS. To facilitate this, PTS can be made more apparent by initially localizing un-spliced precursors to defined subcellular compartments. A convenient and easily distinguishable cellular localization is the PM which can be targeted by utilizing a 10 amino acid PM-localization sequence from Lyn kinase [130]. In addition, several fluorescent proteins with separate emission spectrums were used to track the components of PTS such that PTS activity can be easily assayed by fluorescence microscopy.

4.2.2 Establishing Protein *Trans*-Splicing Assay in Mammalian Cells

As a proof of concept, the N-terminal precursor Lyn-Ceru-*Npu*DnaE$_N$-mRFP was created as a tandem fusion of Lyn, the CFP mutant Cerulean [131], *Npu*DnaE$_N$, and monomeric red fluorescent protein (mRFP) [132] while the C-terminal precursor, *Npu*DnaE$_C$-Venus, was constructed as a tandem fusion of *Npu*DnaE$_C$ and the YFP mutant Venus [133]. It was anticipated that PTS would form a new PM-bound fusion protein Lyn-Ceru-Venus while the reconstituted *Npu*DnaE-mRFP would be removed from the PM (Figure 4.2).

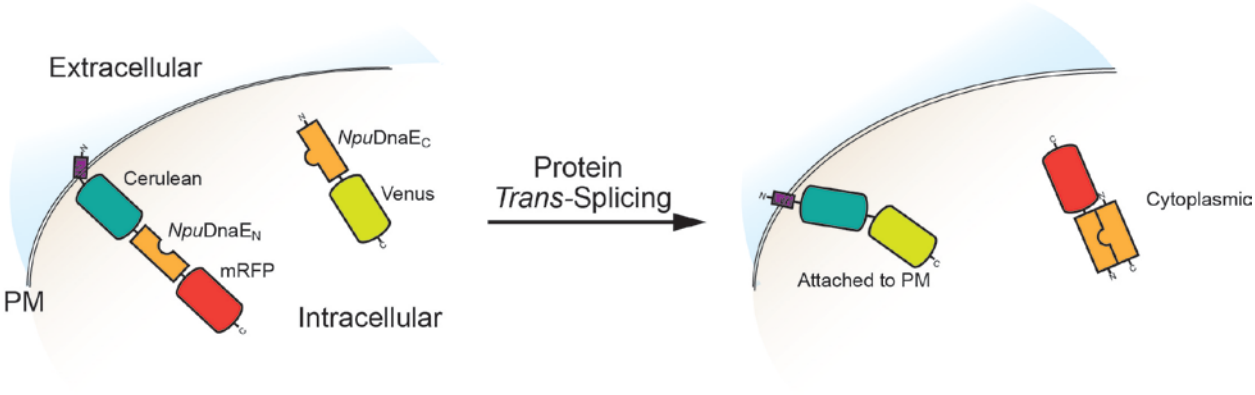

Figure 4.2 Overview of protein *trans*-splicing assay.

The general strategy for visual assay is depicted. Before PTS, Cerulean and mRFP are initially localized to the PM while Venus is localized to the cytoplasm. After PTS, Venus is ligated to the PM-bound Cerulean while the mRFP is excised out along with the intein to the cytoplasm. It is this translocation of mRFP and Venus that forms the basis of the assay.

The high binding affinity of the *Npu*DnaE halves (K_d ~ 3 nM) is postulated to be largely due to the global and local intermolecular charge distributions that contribute to the electrostatic interactions between the intein halves [31, 53, 54]. As a result, the mere presence of the precursor halves induces spontaneous PTS activity. To address the absence of PTS activity, each precursor halves were expressed separately. As expected, HeLa cells expressing Lyn-Ceru-*Npu*DnaE$_N$-mRFP had the PM labelled with cyan and red fluorescence while HeLa cells expressing only *Npu*DnaE$_C$-Venus had the cytoplasm labelled with yellow fluorescence (Figure 4.3). When HeLa cells were co-expressed with Lyn-Ceru-*Npu*DnaE$_N$-mRFP and *Npu*DnaE$_C$-Venus, spontaneous PTS resulted in the cytoplasmic Venus being PM-bound labelling the PM with cyan and yellow fluorescence, while mRFP translocated from the PM to the cytoplasm labelling it with red fluorescence (Figure 4.3).

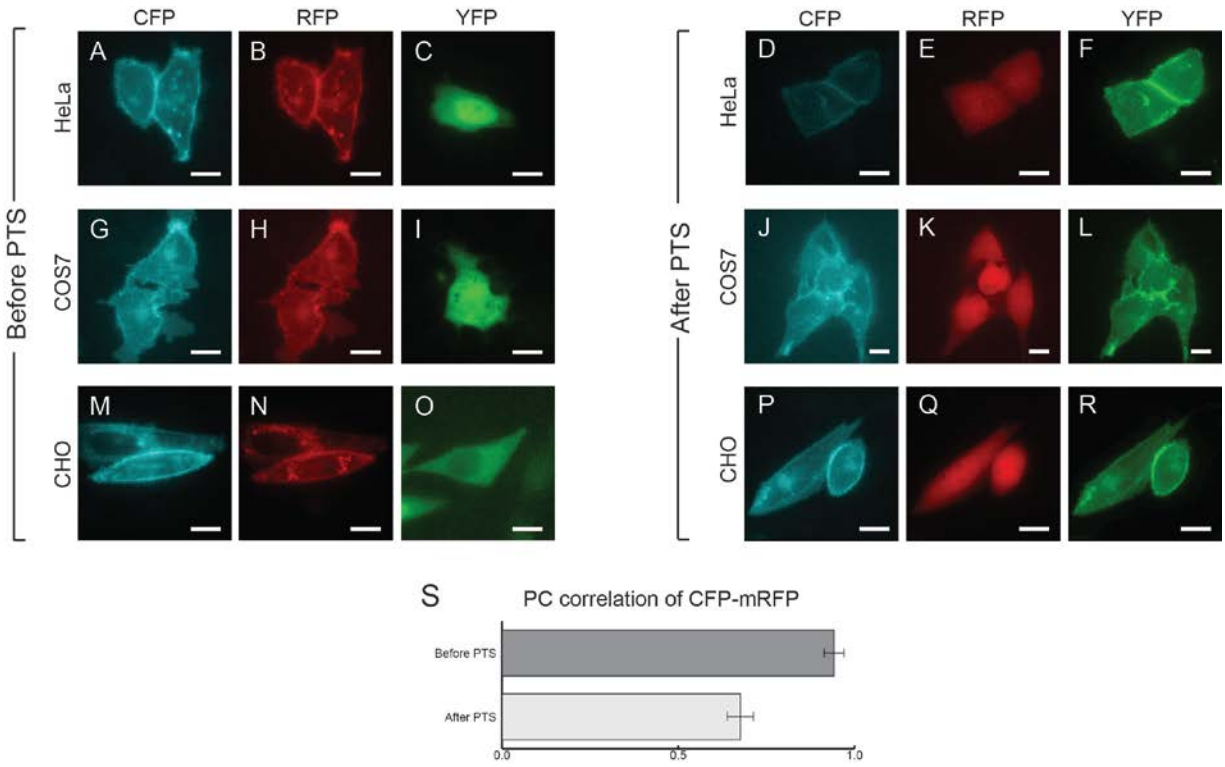

Figure 4.3 Establishing visual protein *trans*-splicing assay in mammalian cells

HeLa cells (**A-C**) before and (**D-F)** after PTS. Similar, in COS7 cells (**G-I**) before and (**J-L**) after PTS and CHO cells (**M-O**) before and (**P-R**) after PTS. In all cases, mRFP that was initially in the PM is translocated to the cytoplasm after PTS activity. (**S**) Pearson coefficient showing the decreased in correlation of CFP-mRFP before and after PTS. Cell images are false coloured. Scale bars represent 10 μm.

Generally, the correlation between CFP and mRFP co-localization as measured by PC, dropped from ~0.942 before PTS to ~0.676 after PTS as a result of the *Npu*DnaE separating from its precursor. Similar results were obtained in COS7 and CHO cell lines (Figure 4.3). The formation of spliced products was also verified by SDS-PAGE.

In a similar fashion, an N-terminal precursor, Lyn-Ceru-*Npu*DnaE$_N$, was created as a tandem fusion of Lyn, Ceru, and *Npu*DnaE$_N$ while the C-terminal precursor was composed of a tandem fusion of mRFP, *Npu*DnaE$_C$, and Venus to yield mRFP-*Npu*DnaE$_C$-Venus (Figure 4.4).

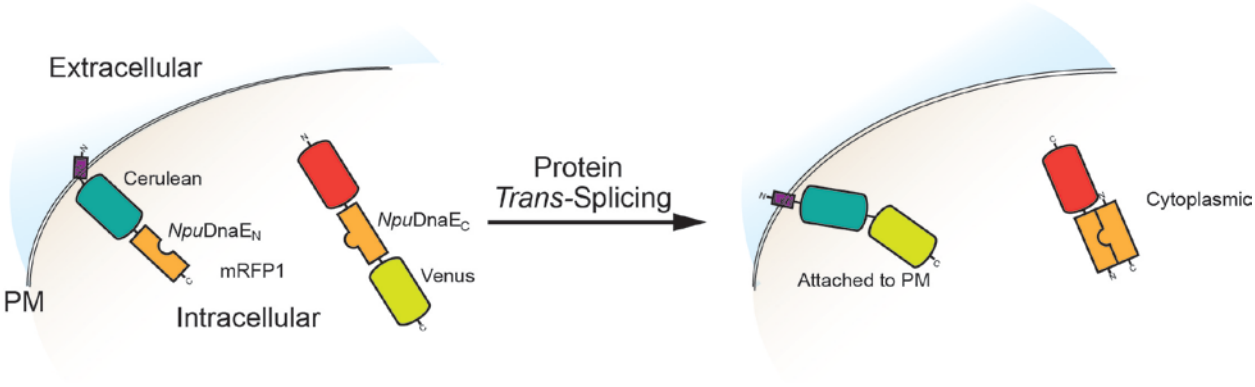

Figure 4.4 Overview of protein *trans*-splicing assay cont'd.

mRFP is now attached to *Npu*DnaE$_C$. Before PTS, Venus and mRFP are initially localized to the cytoplasm while Cerulean is PM-bound. After PTS, Venus is translocated to the PM while mRFP remains in the cytoplasm.

Expression of Lyn-Ceru-*Npu*DnaE$_N$ or mRFP-*Npu*DnaE$_C$-Venus separately in HeLa cells resulted in the labelling of the PM with cyan fluorescence while the cytoplasm was yellow and red fluorescence, respectively (Figure 4.5). Cell expressing both constructs underwent PTS and had the PM labelled with cyan-yellow fluorescence, while the cytoplasm remained red fluorescent (Figure 4.5). The initial co-localization of mRFP and Venus typically dropped in value from ~0.965 to ~0.707, consistent with PTS activity. Again, SDS-PAGE verified the formation of spliced products. To show that the sub-cellular localizations were a result of PTS and not some other mechanism, several mutations were introduced to abolish PTS. To prevent intermediates from forming at the N-terminal splice site and premature splicing, the mutation Cys1Ala was introduced at the first amino acid to yield the non-splicing mutant, ns*Npu*DnaE$_N$.

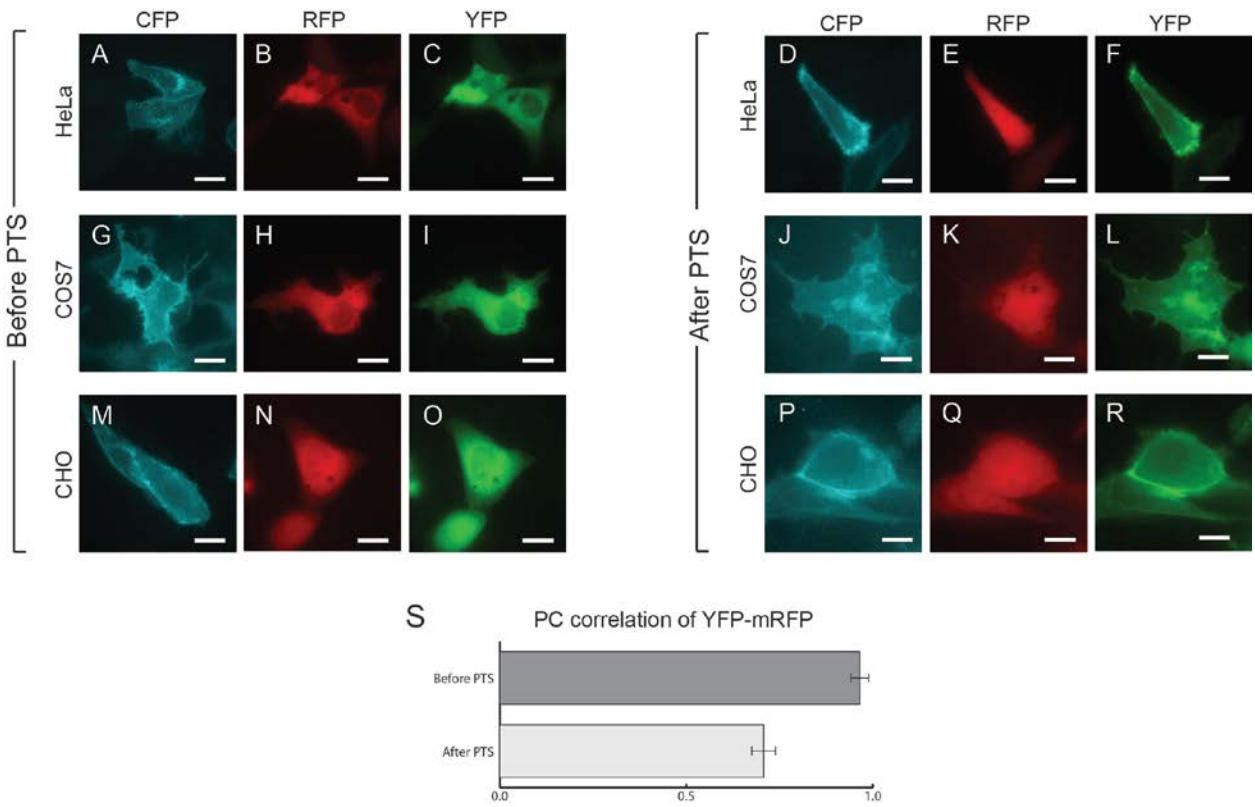

Figure 4.5 Establishing visual protein *trans*-splicing assay in mammalian cells cont'd.

Before PTS, HeLa (**A-C**), COS7 (**G-I**), and CHO (**M-O**) show cytoplasmic expression of mRFP and Venus while Cerulean is PM-bound. After PTS, HeLa (**D-F**), COS7 (**J-L**), and CHO (**P-R**) all show the subcellular localization of Venus to the PM while mRFP remains in the cytoplasm. (**S**) Pearson coefficient showing the decreased in correlation of YFP-mRFP before and after PTS activity. Cell images are false coloured. Scale bars represent 10 μm.

For the same reasons, the non-splicing C-terminal intein precursor, ns*Npu*DnaE$_C$, was created by introducing the following mutations to the last carboxy-terminal amino acid and the first two extein amino acids at the splicing junction: Asn36Ala, Cys+1Ser, Phe+2Pro, respectively [50, 134, 135]. When the precursor constructs, Lyn-Ceru-ns*Npu*DnaE$_N$-mRFP and ns*Npu*DnaE$_C$-Venus, were co-expressed in HeLa cells, PTS was abolished and the PM was labelled with all three fluorescence: red, cyan, and yellow (Figure 4.6).

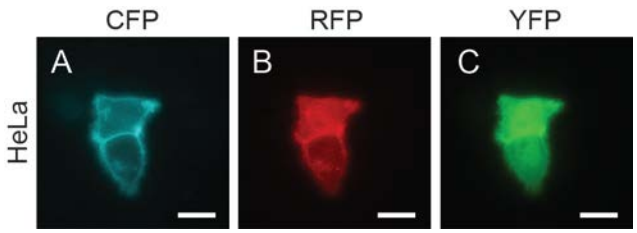

Figure 4.6 Non-splicing intein do not exhibit translocation of fluorescent proteins.

Non-splicing inteins are able to associated by are not able to splice. The colocalization of all three fluorescent proteins show that the translocation of the fluorescent proteins from the previous two scenarios are the result of PTS activity.

4.2.3 Reassembly of TN-XL Ca^{2+} Biosensor by *Npu*DnaE Protein *Trans*-Splicing

An artificially split TN-XL biosensor was reassembled by *Npu*DnaE protein *trans*-splicing. TN-XL is a commonly used FRET-based Ca^{2+} biosensor composed of the chicken skeletal muscle troponin C (TnC) and its binding peptide situated between the FRET donor-acceptor pair, ECFP and citrine [93]. TnC has a dumbbell-shaped structure with two globular calcium binding domains connected by a central linear linker [136, 137]. The TnC used in TN-XL and has been finely tuned to bind tightly to its target to give enhanced conformational changes [93, 138]. To minimally perturb this, TN-XL was divided at the junction separating the N-terminus ECFP and TnC-citrine. ECFP was then tandem fused to *Npu*$DnaE_N$ to obtain the N-terminal fusion protein ECFP-*Npu*$DnaE_N$ while the TnC-citrine was fused to *Npu*$DnaE_C$ to obtain the C-terminal precursor construct *Npu*$DnaE_C$-TnC-citrine (Figure 4.7). Reassembly of the split TN-XL biosensor was the result of PTS activities. Non-splicing precursor pairs, Ceru-*Npu*$DnaE_N$ and TnC-citrine, co-expressed in HeLa cells produced no measureable UTP-induced Ca^{2+} transients. Likewise, co-expression of ECFP and *Npu*$DnaE_C$-TnC-Venus produced no measureable UTP-induced Ca^{2+} transients.

To assess the effectiveness of the reassembled split TN-XL biosensor in measuring Ca^{2+} transients, HeLa cells transfected with the split TN-XL precursors were used to measure the FRET responses from UTP-induced Ca^{2+} transients (Figure 4.7). Dynamic ranges were obtained by loading cells with ionomycin in the presence of $CaCl_2$ to obtained maximum cytosolic Ca^{2+}

followed by the loading of excess EDTA to chelate the Ca^{2+}. As a control, HeLa cells were transfected with unaltered TN-XL biosensors and used to measure similarly induced Ca^{2+} transients and dynamic ranges. What was found was that the reassembled TN-XL biosensor had a significantly reduced FRET-dynamic range when compared to the unaltered TN-XL biosensor (1.1 ± 0.2 and 1. 7 ± 0.4, respectively with $P < 0.001$, $n = 9$) (Figure 4.7). Similar results were obtained when repeated in COS7 and CHO cell lines.

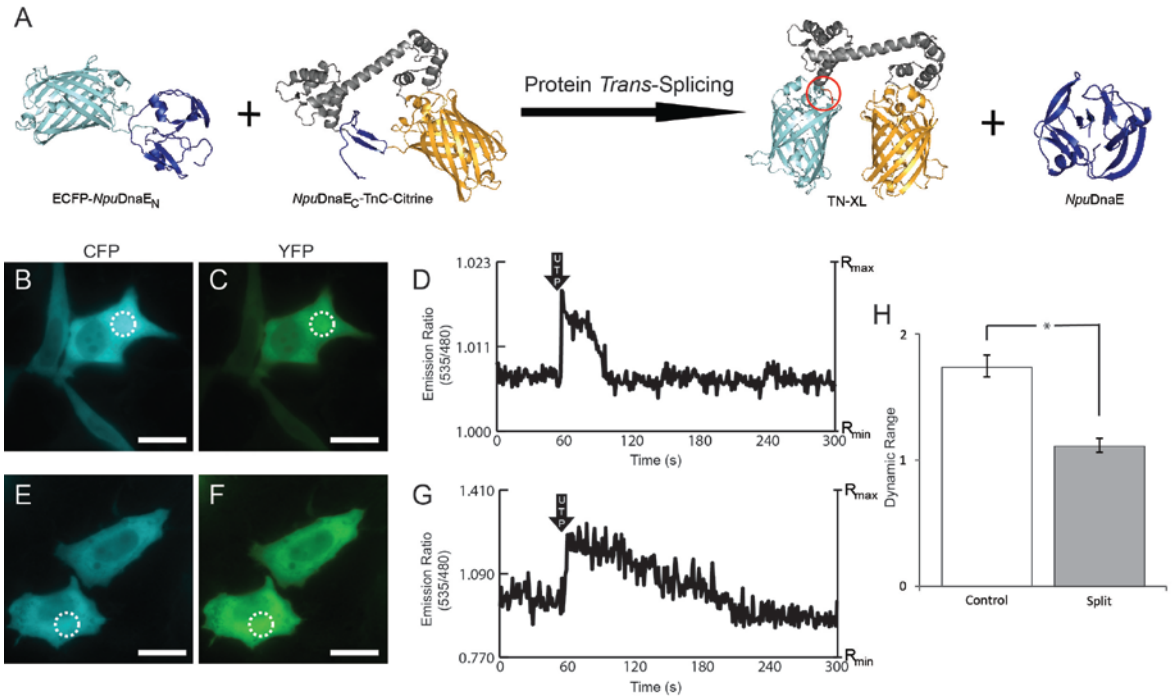

Figure 4.7 Reassembly of split TN-XL Ca^{2+} biosensor.

(**A**) Structural modeling of the reassembly of split TN-XL Ca^{2+} biosensors by PTS activity. TN-XL was artificially split between the first fluorescent protein ECFP (cyan) and chicken skeletal troponin C (TnC, grey) and fused to its respective *Npu*DnaE intein halves (blue). PTS activity reassembled the TN-XL Ca^{2+} biosensor and the associated *Npu*DnaE intein (blue). Of note is the residual amino acids that are incorporated into the reassembled TN-XL biosensor at the splice site (shown and circled in red). Citrine is shown in yellow. HeLa cells co-expression the two precursors and shown under (**B**) CFP and (**C**) YFP fluorescence and (**D**) the profile of the emission ratios of UTP-induced Ca^{2+} transients. Expression of unaltered control TN-XL in HeLa viewed under (**E**) CFP and (**F**) YFP fluorescence and (**G**) the profile of emission ratios of UTP-induced Ca^{2+} transients. (**H**) Comparison of mean dynamic range of the control and split TN-XL biosensor. White dotted circle represent regions where average fluorescence intensity measurements were taken. Error bars represent standard deviation. Cell images are false coloured. Scale bars represent 10 µm.

This reduced FRET dynamic range may be attributed to one of two factors, or more likely, a combination of both. First, PTS process leaves behind a series of residual amino acids (Section 2.2.3) with the spliced products. As a result, the newly reassembled TN-XL biosensor was a mutant form with a longer linker region between ECFP and TnC. Since the efficiencies of FRET are governed by a set of highly defined parameters, for example FRET transfer rates vary inversely to the 6^{th} power of donor-acceptor separation [139], the longer linker likely altered the fluorophore orientation as well as increased the separation distance between them, leading to reduced FRET response. A second more subtle factor is the fact that unconsumed precursors remain fluorescent, and likely contribute significantly to background fluorescence. This was evident when protein constructs were extracted from mammalian cells and separated by SDS-PAGE. In theory, the FRET efficiencies of the reassembled TN-XL can be optimized by removing non-essential amino acids near the split junction that will be replaced by the PTS residual amino acids. However, the issue with unconsumed precursors contributing to background fluorescence remains. This can be remedy by expressing the two precursors at equal stoichiometry. However, the techniques employed are often tedious and difficult at best [140]. The donor-acceptor pairs can be engineered to be non-fluorescent until PTS restores fluorescence, as has been done previously for detecting protein-protein interactions [141]. However this requires the use of two or more different inteins with no cross-reactivity between them, adding unnecessary complexity. It is of interest to note that these issues arise not only for TN-XL but most likely for all FRET-based GECIs. Thus, although conceptually FRET-based GECIs are less sensitive to many of the common artefacts in live cell and animal model studies, the preceding factors make them less than ideal when used in conjunction with inteins.

4.2.4 Reassembling GCaMP2 Ca^{2+} Biosensor by *Npu*DnaE Protein *Trans*-Splicing

An artificially split GCaMP2 biosensor was reassembled by *Npu*DnaE protein *trans*-splicing. GCaMP2 is a variant of the commonly used family of GCaMP Ca^{2+} biosensors [96, 142, 143]. In contrast to FRET-based GECIs, GCaMP2 relies on the changing fluorescence intensities of a single fluorescent protein (GFP) to report on endogenous Ca^{2+} concentrations. It accomplishes this by having a centrally located circularly permutated GFP (cpGFP) moiety flanked by the M13 helix of myosin light chain and the Ca^{2+} sensing domain calmodulin (CaM) at its N- and C-

terminus, respectively [94, 142-144]. CaM is structured similarly to TnC and binds its target peptide M13. The changing fluorescence stems from changing fluorophore efficiency as a result of intramolecular binding of CaM to M13. In the absence of Ca^{2+}, CaM releases M13 which in turn exposes the fluorophore to destabilizing solvents. Increased Ca^{2+} concentration binds CaM to M13, restricting solvent access to the fluorophore and stabilizes it followed by the ensuing increased in fluorescence intensity [144]. Considerable effort went into optimizing the linkers connecting CaM and M13 to cpGFP [96], thus it would be best not to manipulate these areas. Structural examination (PDB ID: 3EVR) and literature [96, 142, 143] revealed that the most promising position to divide GCaMP2 with the least disturbance to its native structural arrangement was within the linker region connecting the cpEGFP. Because this flexible hexapeptide linker, 'GGTGGS', was purposely engineered into the cpEGFP, it follows that this linker could be replaced with another set of flexible linker without adversely affecting GCaMP2 activity. In particular, it was of interest to determine if the PTS residual peptide was able to replace the native linker while minimally disrupting GCaMP2 activity. The native linker is comprised mostly of the small amino acid glycine which provides significant flexibility to cpEGFP whereas the PTS residual peptide, 'CFNGT', is comprised mainly of amino acids with bulky side chains. These bulky side chains may restrict linker movement and in turn cause improper folding of cpEGFP. To account for this, additional amino acids may be required to enhance flexibility of the PTS residual peptide. Since four additional amino acids were introduced by the cassette method (two amino acids on either ends of the PTS residual peptide), no new amino acids were added to the linker. To confirm that this altered GCaMP2 biosensor retained similar response to the unaltered GCaMP2 biosensor, a mutant GCaMP2 was created with the native linker sequence replaced by the new linker sequence (Figure 4.8). The mutant GCaMP2 was well expressed in several mammalian cell lines (HeLa, COS7, CHO) and no protein aggregates or inclusions were observed, which was consistent with unaltered GCaMP2. The mutant was also able to measured UTP (or ATP)-induced Ca^{2+} transients and had measured dynamic ranges that were not statistically different from that of the unaltered GCaMP2 biosensor (2.62 ± 0.29 and 2.75 ± 0.10, respectively, $p = 0.96$, $n = 9$) (Figure 4.8).

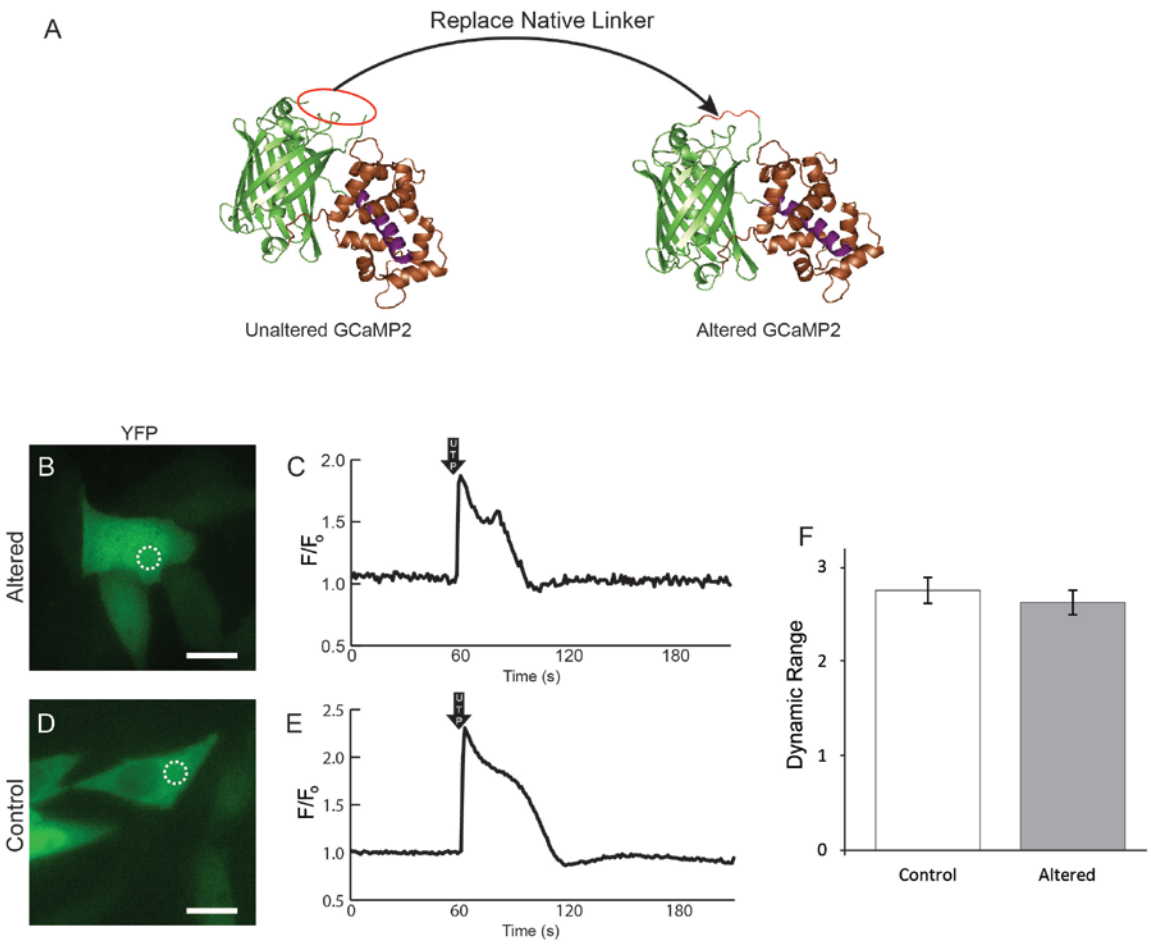

Figure 4.8 Efficacy of mutant GCaMP2.

(**A**) Structural model depicting the replacement of the native GCaMP2 linker (red circle) with the residual peptide linker (shown in red). (**B**) Fluorescent image of HeLa cells expressing the mutant GCaMP2 and (**C**) its measured UTP-induced Ca^{2+} transient profile. (**D**) Fluorescent image of HeLa cells expressing unaltered GCaMP2 and (**E**) its measured UTP-induced Ca^{2+} transient profile. (**F**) Comparison of mean dynamic range of control and mutant (altered) GCaMP2. White dotted circle represent regions where average fluorescence intensity measurements were taken. Cell images are false coloured. Scale bars represent 10 µm.

Drawing from these findings, The N-terminal precursor of the split GCaMP2 intein constructs was created by the tandem fusion of RSET, a plasmid leader sequence required for proper folding at physiological temperature [142-144], M13, and cpEGFP(149-238) directly to

*Npu*DnaE$_N$ to yield GCaMP2$_N$-*Npu*DnaE$_N$ while the C-terminal precursor was a tandem fusion of *Npu*DnaE$_C$ directly to cpEGFP(1-144)-CaM to form *Npu*DnaE$_C$-GCaMP2$_C$ (Figure 4.9).

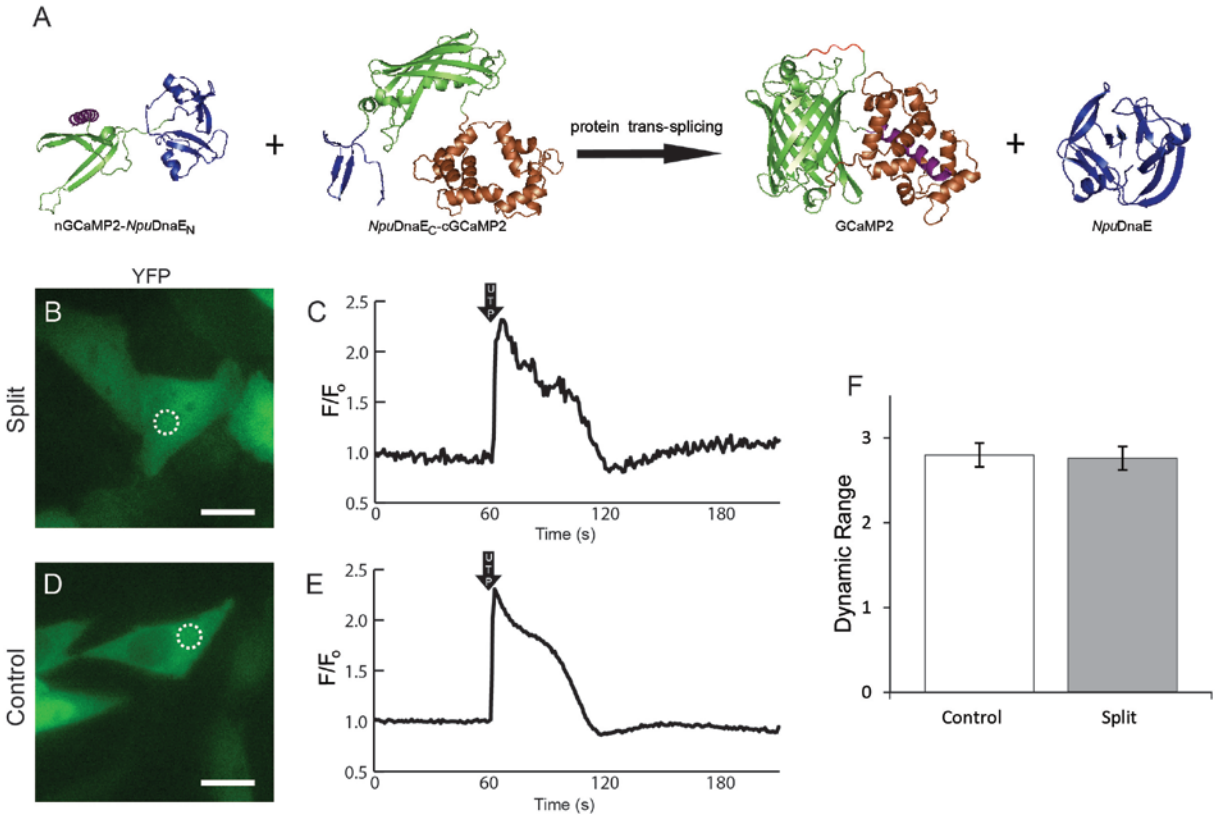

Figure 4.9 Reassembly of split GCaMP2 Ca^{2+} biosensor.

(**A**) Structural modeling of the reassembly of split GCaMP2 Ca^{2+} biosensor by PTS activity. GCaMP2 had the native linker region removed and fused to it respective intein halves (blue). PTS activity reassembled the mutant GCaMP2 with the residual peptide (shown in red). EGFP shown in green; M13 in purple; and CaM in brown. (**B**) Fluorescence image of HeLa cells co-expressing the two precursors and (**C**) its UTP-induced Ca^{2+} transient profile as measured by the reassembled split GCaMP2. Similarly, (**D**) fluorescence image of HeLa cells expression unaltered GCaMP2 and (**E**) its UTP-induced Ca^{2+} transient profile. **(F)** Comparison of mean dynamic range of control and split GCaMP2. White dotted circle represent regions where average fluorescence intensity measurements were taken. Error bars represent standard deviation. Cell images are false coloured. Scale bars represent 10 μm.

The effectiveness of the reassembled GCaMP2 was determined by transfecting HeLa cells with the split GCaMP2 precursors. The reassembled GCaMP2 was able to measure UTP-induced Ca^{2+}

transients and produced similar traces as those measured by a control unaltered GCaMP2 (Figure 4.9). Dynamic ranges were again obtained by loading cells with ionomycin in the presence of $CaCl_2$ to obtained maximum cytosolic Ca^{2+} followed by the loading of excess EDTA to chelate the Ca^{2+}. The reassembled GCaMP2 had a statistically indistinguishable dynamic range compared to the control unaltered GCaMP2 biosensor (2.75 ± 0.41 and 2.79 ± 0.10, respectively; p = 0.93, n = 3 experiments; each experiment had > 3 cells) (Figure 4.9). Similar results were obtained in COS7 and CHO cell lines. Again, reassembly of the split GCaMP2 biosensor was determined to be from the result of PTS activities. Co-expression of non-splicing precursor pairs ($GCaMP2_N$-*Npu*$DnaE_N$ with $GCaMP2_C$; $GCaMP2_N$ with *Npu*$DnaE_C$-$GCaMP2_C$) in several mammalian cells (HeLa, COS7, and CHO) produced no fluorescently labelled cells required for Ca^{2+} imaging experiments. Notably, cells individually transfected with $GCaMP2_N$-*Npu*$DnaE_N$ or *Npu*$DnaE_C$-$GCaMP2_C$ did not display any fluorescence. This is desirable as unpaired and unconsumed GCaMP2 precursors do not contribute to background fluorescence and equal stoichiometric expression of the precursors are unnecessary. These factors taken together, of the two types of GECIs, single fluorophore GECIs such as GCaMP2, offer a more suitable candidates for intein reassembly experiments.

4.2.5 Reassembly of GCaMP2 in Pharyngeal Muscle Cells of *C. Elegans*

The split GCaMP2 biosensor was reassembled and retained functionality in the *C. Elegans* animal model. Split GCaMP2 was chosen over TN-XL for two reasons: first, the dynamics range was as good as the unaltered GCaMP2 biosensor; second, any unpaired precursors were non-fluorescence and thus do not contribute to background fluorescence. To test of the efficacy of the reassembled GCaMP2 in measuring calcium transients in animal models, the pharyngeal muscles of *C. Elegans* was chosen as a target. The pharyngeal muscle is composed of three components: the corpus, isthmus, and terminal bulb [145-147]. These three components together make up a relatively large muscular organ (the terminal blulb alone has a diameter of approximately 20 μm) involved in feeding. During feeding, the vigorous pumping actions of the corpus and isthmus regions direct food into the terminal bulb for grinding [148, 149]. These feeding behaviours can be categorized based on its pumping frequency: very slow pumping (<0.5 Hz), slow pumping (0.5-1.5 Hz), and fast pumping (>1.5 Hz) and has been shown to relate to some distinct

properties [146]. For example, fast pumping frequencies often require the input from pharyngeal neurons, whereas slow pumping can occur independently of neuronal activities [150, 151]. On the other hand, very slow pumping frequencies are often indicative of impending death [146]. Since the pharyngeal muscle components cells are electrically coupled by gap junctions, the pumping actions of the organ gives large synchronized Ca^{2+} transients during feeding and thus facilitates tracking and Ca^{2+} dynamic measurements [150, 152].

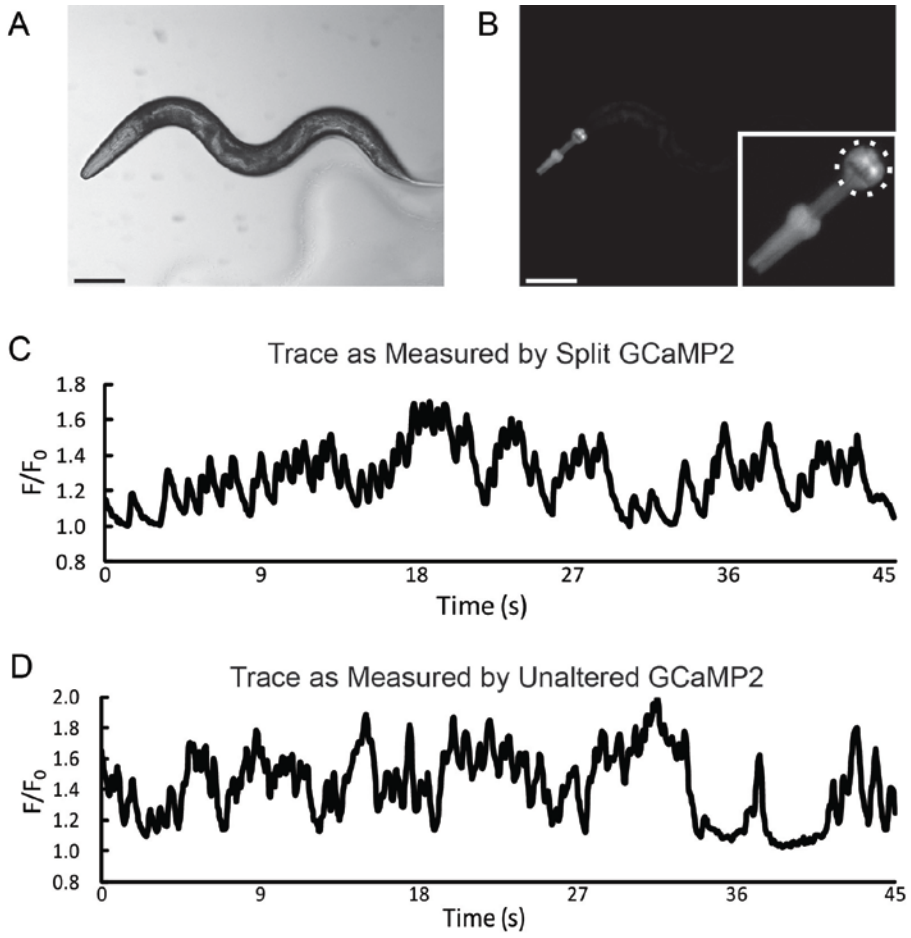

Figure 4.10 Detection of Ca^{2+} transients in the pharynx of freely moving C. Elegans using reassembled split GCaMP2.

The transgenic *C. Elegans* co-expressing the two split GCaMP2 precursors under the myo-2 promoter were imaged in (**A**) bright field and (**B**) fluorescence. The dotted circle (**B**, inlet) indicates the terminal bulb region of the pharyngeal muscle where the time series average fluorescence intensity was measured. (**C**) Normalized intensity of the terminal bulb was plotted as a function time. (**D**) Time series of average intensity of the terminal bulb of control transgenic *C. Elegans* expressing unaltered GCaMP2. Scale bar represent 100 µm.

The split GCaMP2 precursors, GCaMP2$_N$-*Npu*DnaE$_N$ and *Npu*DnaE$_C$-GCaMP2$_C$, were directed under the myo-2 promoter for pharyngeal muscle cells expression and co-transfected into *C. Elegans* to obtain the transgenic worms (Figure 4.10). Transgenic worms successfully expressing the reassembled GCaMP2 were collected and cultivated by looking for emitted fluorescence from the pharyngeal organ. To induce rapid feeding behaviours, the transgenic worms were soaked in a solution of serotonin [153]. Average fluorescence intensity of the terminal bulb was obtained by automated pattern-matching algorithm. A typical temporal profile of the terminal bulb as measured by the reassembled GCaMP2 biosensor had a contraction frequency of 0.5-1.0 Hz (Figure 4.10 and Movie 4.1). This was in agreement with the temporal profile of the terminal bulb as measured with the control unaltered GCaMP2 biosensor (Figure 4.10). Transgenic worms expressing either GCaMP2$_N$-*Npu*DnaE$_N$ or *Npu*DnaE$_C$-GCaMP2$_C$ did not exhibit any fluorescence when stimulated, as expected. Notably, the reassembled split GCaMP2 was evenly expressed across all three pharyngeal muscle regions (the corpus, isthmus, and terminal bulb). This represents an improvement over a previous Ca^{2+} dynamic study where the Ca^{2+} indicator, cameleon, was precluded from the isthmus but expressed elsewhere [146].

4.3 Discussion

An *in vivo* assay was initially developed to provide a simple visual assay for the occurrence of PTS by examining distinct subcellular localizations of fluorescent proteins. PTS would result in the rearrangement of fluorescent proteins where one type would be shuttle from the cytoplasm to the PM and the other from PM to cytoplasm. While naturally split inteins have been traditionally used to ligate two exogenous exteins while leaving the remaining non-splicing ends free [154, 155], here it was demonstrated that the non-splicing ends may be fused to a foreign peptide without inhibiting PTS. This is not surprising as the intein split sites may have been occupied by an endonuclease domain in its ancient ancestors, but over time had shed its endonuclease domain and evolved to splice as two fragments [26, 27]. This phenomenon may be exploited to provide another dimension to naturally split inteins (see Chapter 5).

PTS was then applied to two representative artificially split GECIs. *Npu*DnaE intein was able to reassemble the FRET-based TNXL biosensor. However, due to the highly sensitive nature of

FRET to modifications to the fluorophore separation and orientations, the reassembled TN-XL biosensor had significantly diminished response relative to the control TN-XL biosensor. Though there is no doubt that further optimization will yield a fully functional reassembled TN-XL biosensor, the techniques and factors employed to make this happen makes it impractical for further use in this study. In addition, unconsumed precursors continue to emit fluorescence that contribute to background fluorescence and interfere with accurate signal response measurements.

Next, PTS was used to reassemble the single fluorophore Ca^{2+} biosensor GCaMP2. Here, an altered GCaMP2 was created in order to mitigate the effects of the PTS residual peptide. This mutant GCaMP2 had similar Ca^{2+} response profile and dynamic range as the unaltered GCaMP2 biosensor and was thus used to construct split GCaMP2 precursors throughout this study. The reassembled split GCaMP2 was able to measure induced Ca^{2+} transients with dynamic ranges that were indistinguishable from the control unaltered GCaMP2 biosensor in mammalian cell lines. This was further extended to the *C. Elegans* animal model where the reassembled GCaMP2 was able to detect fluctuating Ca^{2+} dynamics in contracting pharyngeal muscle cells similar to profiles as measured with control GCaMP2 in this study and profiles obtained by other groups using other Ca^{2+} probes [145]. A major advantage of the split GCaMP2 biosensor is that unconsumed precursors are not fluorescent and fluorescence is only restored in those cases where protein splicing was complete. Thus, the effectiveness of the biosensor is not dependent on obtaining equal stoichiometry of the precursor halves and can be more broadly applied, for example in targeting specific cell types using two orthogonal promoters (see Chapter 8).

4.4 Chapter Conclusions

In this chapter, PTS by *Npu*DnaE intein was able to reassemble two representation members of the family of GECIs: TN-XL and GCaMP2. Specific aims set out in the beginning of the chapter and addressed here are:

1. The development of an *in vitro* visual assay to test for PTS by observing distinct subcellular localization of various fluorescent proteins. This was confirmed with by fluorescence SDS-PAGE and non-splicing *Npu*DnaE mutants.

2. *Npu*DnaE intein was shown to be highly tolerant of proteins attached to the non-splicing ends. This property may be exploited to add increased functionality to the intein.

3. *Npu*DnaE intein was able to reassemble and restore function to TN-XL and GCaMP2, although the reassembled TN-XL had significantly reduced FRET response. In contrast, the split GCaMP2 was able to restore activity in several cell lines and in an animal model.

Research objective 1 as outlined in Chapter 1, was addressed in this chapter. In particular, PTS was used to reform and restore function to two non-functional precursor fragments by ligating the fragments together. In addition, the development of the PTS visual assay alluded to a potentially useful property as outlined in research objective 2.

5 Simultaneous Reassembly of Two Target Proteins

The content of this chapter was modified from the following accepted paper:

Stanley SC Wong, Evan Mills, Kevin Truong. Simultaneous Assembly of Two Target Proteins Using Inteins. *PEDS*. 2012. In Press.

5.1 Chapter Introduction and Aims

During the development of the visual assay in the work from the previous chapter, it was determined that each intein halves (*Npu*DnaE$_N$ and *Npu*DnaE$_C$) could have foreign proteins simultaneously fused to both the non-splicing and splicing ends without hindering PTS activity. Although in the previous work, only one of the two intein halves had foreign proteins attached to both ends while the remaining intein half had its non-splicing end remaining free, there is a possibility that both non-splicing ends on both complementary inteins halves can be occupied by foreign proteins without interfering with PTS activity. If this can be accomplished, it can be used to establish a novel approach of reassembling two target proteins using a single naturally split intein pair. One of the interesting properties of the *Npu*DnaE intein halves used in this study is that it naturally has a very high binding affinity (K_d ~ 3 nM) and seems to remain tightly bound to each other even after PTS [53]. Because of this, in addition to facilitating PTS, *Npu*DnaE intein can also be used as an association pair to reassemble a second target protein such as a fluorescent protein reporter. The following study looks into exploring this property and in turn, further extends the uses of naturally split inteins.

For this chapter, the following three specific aims were established to achieve the overall research goal:

1. Determine if *Npu*DnaE associate in a specific orientation

2. Investigate whether the two target proteins produced are in stoichiometric ratios

3. Explore whether *Npu*DnaE intein can be used to as a fluorescence reporter on the successful PTS process and as a fluorescence reporter for locating cells.

In this chapter study, several engineered fusion proteins were created to address the specific aims and demonstrate the reassembly of two target proteins. All engineered proteins were created in accordance with the gene construction methodologies as presented in Chapter 3 (Figure 5.1).

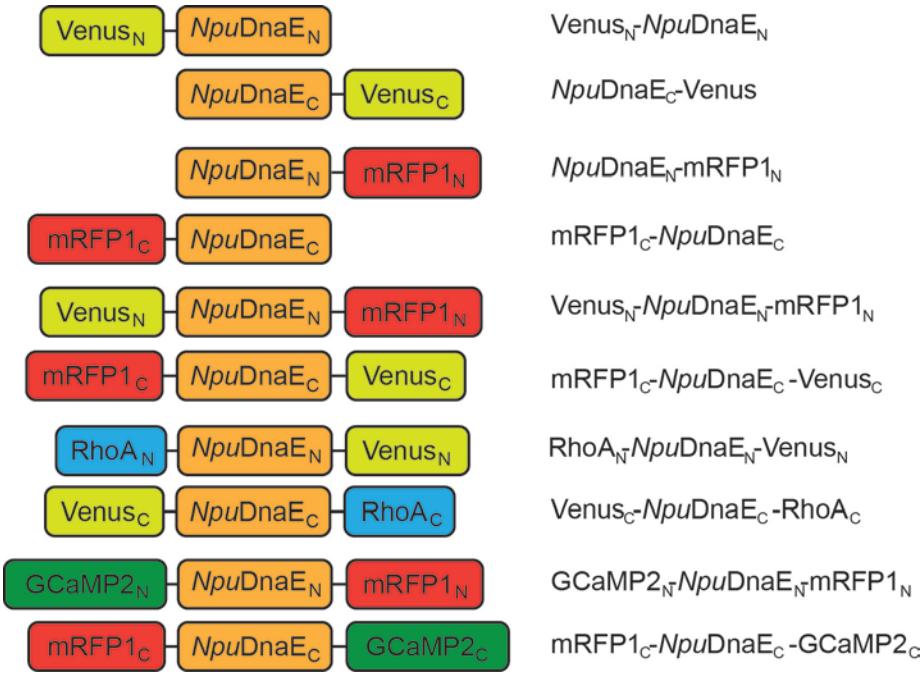

Figure 5.1 Outline of the fusion proteins used in this chapter.

Block diagrams of the fusion protein are shown. The names of the fusion proteins used in this chapter are also given.

5.2 Results

This section will describe the experimental results for the simultaneous reassembly of two proteins. A general description of the strategy will be given (section 5.2.1) followed by the testing of the strategy in mammalian cells (section 5.2.2). This strategy is then validated in reassembling split RhoA (section 5.2.3) and split GCaMP2 (section 5.2.4) with a reporter fluorescent protein.

5.2.1 General Strategy for Reassembling Two Target Proteins Using a Single *Npu*DnaE Intein

The general strategy for the simultaneous reassembly of two target proteins has each *Npu*DnaE intein halves fused with a split extein sequence (i.e. $Extein_N$ and $Extein_C$) while concurrently, the vacant non-splicing ends are fused to complementary halves of another split protein sequence (i.e. $Endo_N$ and $Endo_C$) (Figure 5.2).

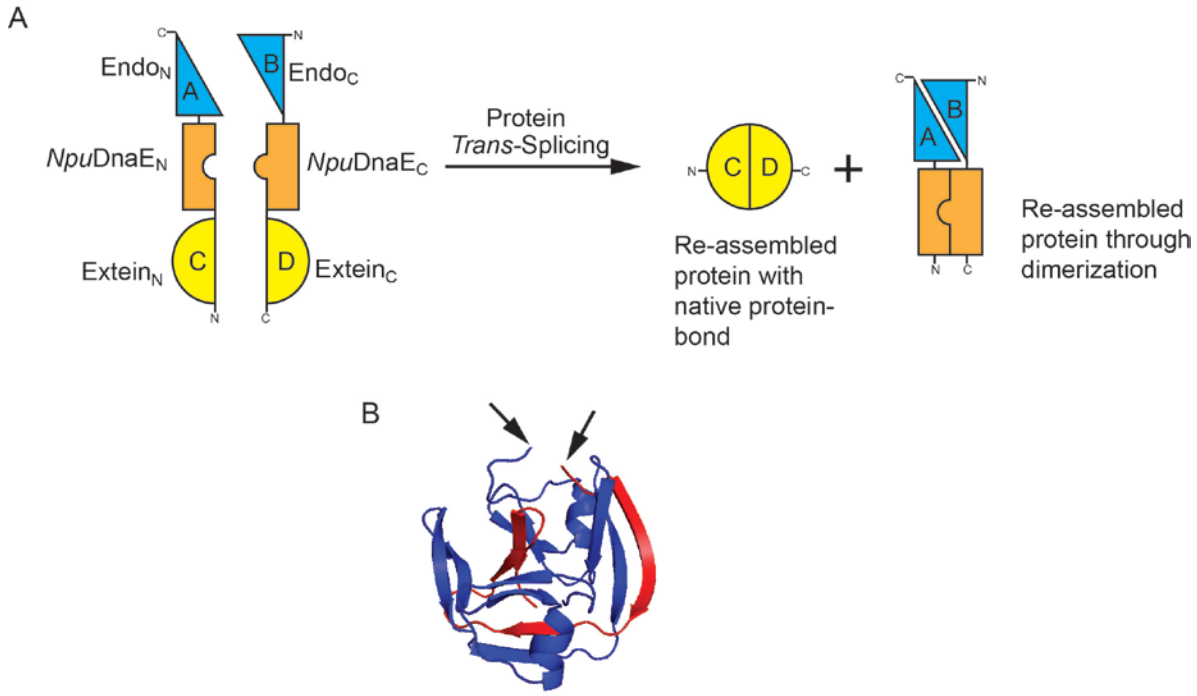

Figure 5.2 General strategy for the simultaneous reassembly of two target proteins.

(**A**) Schematic and nomenclature for the constructs used. The split extein halves are located at the splicing ends at position C and D, while the endo halves are located at the non-splicing ends at positions A and B. The split intein halves self associate and reform the active intein for PTS activity. Exteins C and D are ligated with a peptide bond, while the couple of the intein halves facilitate the non-covalent association of fragments A and B to restore function to the Endo protein. (**B**) Ribbon diagram of the associated *Npu*DnaE$_N$ (blue) and *Npu*DnaE$_C$ (red) halves (PDB ID: 2KEQ). Arrows show the two non-splicing ends.

In the co-expression of the two precursors, two specific processes are expected to occur in concert (Figure 5.2). First, the inteins spontaneously reconstitute and triggers PTS activity. This ligates the two extein sequences with a native peptide bond and reassembles the first target protein. Secondly, the high affinity binding of the *Npu*DnaE intein halves bring the fragments of the second target protein into close proximity. This allows the protein fragments to non-covalently associate to reform its native three-dimensional structure and return function to the second target protein. Often times, the non-covalent association of the protein fragments can be aided by having the protein fragments sufficiently close to each other so as to increase the local concentration or are tethered to linkers to provide adequate flexibility for the fragments to optimally position themselves during protein refolding. Examination of the protein structure of *Npu*DnaE intein revealed that the non-splicing free ends are positioned near each other when they are associated [52]. In addition, the first few residues of both the non-splicing free ends are free of regular secondary structures and they do not appear to be involved with the proper functioning of the inteins (Figure 5.2) [52]. Thus, they can act as short linkers and could assist in tethering and providing flexibility to the protein fragments during reassembly of the second target protein.

5.2.2 *Npu*DnaE Inteins Reassembles Two Target Proteins

Protein *trans*-splicing by *Npu*DnaE inteins was able to simultaneously reassemble two target proteins. In addition, association of the intein halves occur in only one orientation and exhibit stoichiometric reassembly of two target proteins as determined by two different split fluorescent proteins. Fluorescent proteins were initially used as target proteins because they can be easily detected with fluorescence microscopy and non-fluorescent fragments have been shown to restore fluorescence when they were associated by some means [3, 10]. In addition, the split sites of fluorescent proteins were chosen within linker turns to minimize perturbations to the β-barrel structure [156]. The dual reassembly behaviour of *Npu*DnaE intein was tested by attaching split fluorescent proteins to either the splicing or non-splicing ends to determine if the intein coupling can restore fluorescence in both cases. An N-terminal precursor was created by a tandem fusion of $Venus_N$ (i.e. from 1 to 144) to *Npu*$DnaE_N$ to obtain $Venus_N$-*Npu*$DnaE_N$ while the C-terminal

precursor was created by the tandem fusion of *Npu*DnaE$_C$ to Venus$_C$ (i.e. from 145 to 238) to obtain *Npu*DnaE$_C$-Venus$_C$ to test the PTS activity (Figure 5.3).

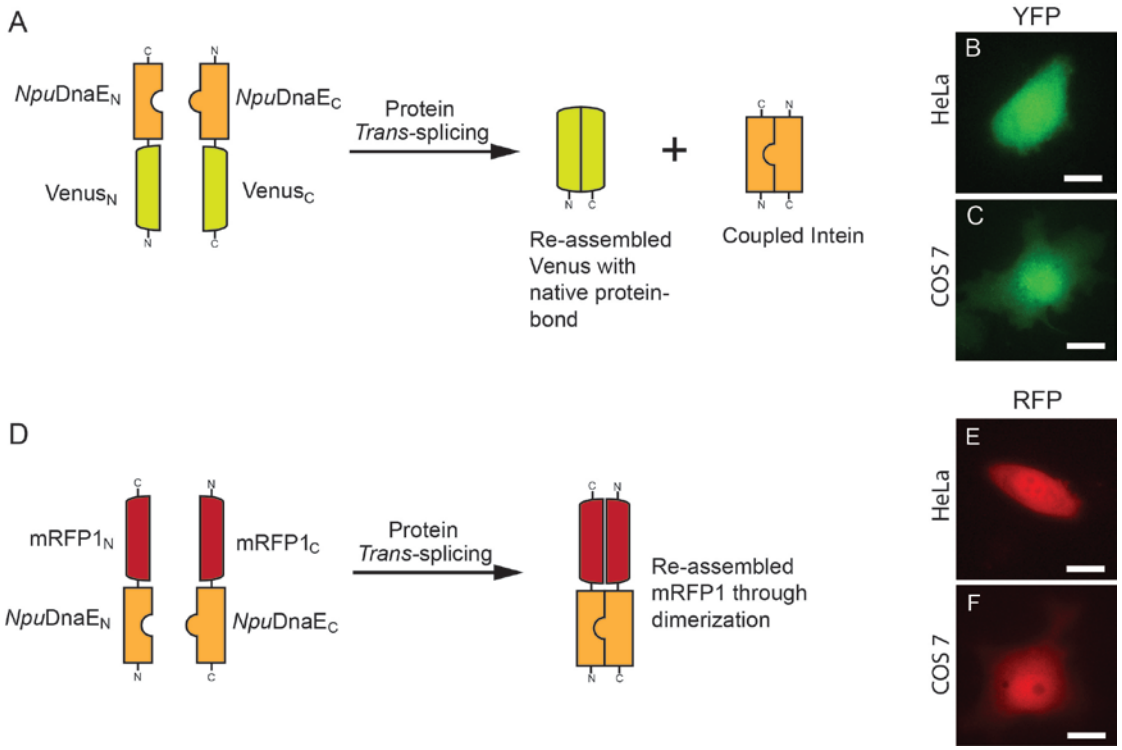

Figure 5.3 Splicing and non-splicing ends restore fluorescence of split fluorescent proteins.

(**A**) Schematic diagram of the restoration of fluorescence from split YFP fused to the splicing end after PTS activity. YFP fluorescence was restored in (**B**) HeLa and (**C**) COS7 cells expressing the two precursors. (**D**) Schematic diagram of the restoration of fluorescence from split mRFP fused to the non-splicing ends of the intein after PTS activity. mRFP fluorescence was restored in (**E**) HeLa and (**F**) COS7 cells expressing the two precursors. Cell images are false coloured. Scale bars represent 10 μm.

As determined in Chapter 4, co-expression of the two precursors in mammalian cells induced PTS activity which restored Venus fluorescence (Figure 5.3). Similarly, attaching split protein fragments to the non-splicing end was also able to restore fluorescence through *Npu*DnaE-mediated dimerization. Here, the N-terminal precursor was created as a tandem fusion of *Npu*DnaE$_N$ to mRFP$_N$ (i.e. from 1 to 139) to yield *Npu*DnaE$_N$-mRFP$_N$ while the C-terminal

precursor was a tandem fusion of $mRFP_C$ (i.e. from 143 to 225) to *Npu*$DnaE_C$ to yield $mRFP_C$-*Npu*$DnaE_C$ (Figure 5.3). Co-expression of the two precursors in mammalian cells restored function to the split mRFP as observed by cytoplasmic emission of red fluorescence (Figure 5.3). As expected, expression of either constructs alone did not yield any apparent fluorescence. These findings together confirm that 1) fusion of foreign to the non-splicing ends of both *Npu*DnaE intein halves do not inhibit PTS activity and 2) the coupling of *Npu*DnaE intein halves were able to bring two protein fragments together to restore protein function through non-covalent association.

Combing the two scenarios from above, the simultaneous reassembly and restoration of fluorescence of two fluorescent proteins can be verified. The N- and C-terminal precursors were created by the tandem fusion of the protein components in the following arrangement: $Venus_N$-*Npu*$DnaE_N$-$mRFP_N$ and $mRFP_C$-*Npu*$DnaE_C$-$Venus_C$, respectively. The co-expression of the two precursors in mammalian cells promoted PTS activity and restored fluorescence via splicing and non-covalent interactions as observed by cytoplasmic emission of Venus and mRFP fluorescence, respectively (Figure 5.4).

Expression of either precursor alone did not result in any apparent fluorescence, as expected. An interesting observation was that cells that expressed stronger Venus fluorescence also emitted stronger red fluorescence, and vice versa. This was not surprising as theoretically, the two split fluorescent proteins should be restored in equal stoichiometry. In mammalian cells co-expressing the two precursors overnight, the YFP/RFP fluorescence intensity ratio of each individual cell was found to be approximately constant (1.25 ± 0.13; $n = 3$ experiments; each experiment had > 20 cells) (Figure 5.4). In contrast, the cells co-expressing Venus and mRFP had widely fluctuating YFP/RFP fluorescence intein ratio from one cell to the next (17.89 ± 15.48; $n = 3$ experiments; each experiment had > 20 cells) (Figure 5.4). Thus, a stronger observed fluorescence in one of the reporter fluorescent protein could potentially indicate an increased reassembly of another non-fluorescent protein.

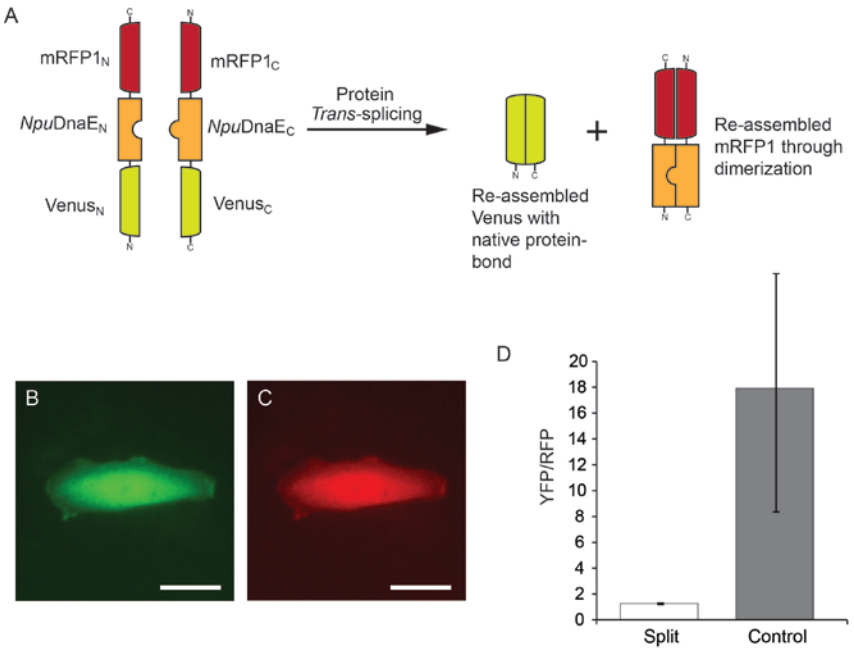

Figure 5.4 Simultaneous reassembly of two fluorescent proteins.

(**A**) Schematic diagram for the simultaneous restoration of fluorescence of two target fluorescent proteins. Split YFP (Venus) is reassembled through the normal splicing pathway by the ligation of the two fragments with a peptide bond. On the other hand, Split mRFP fluorescence is restored by the non-covalent association facilitated by the coupled *Npu*DnaE intein halves. Co-expressing of the two precursors in HeLa cells showing (**B**) YFP and (**C**) RFP fluorescence emanating from the restored fluorescence of the split fluorescent protein via PTS activity. (**D**) Graph of the YFP/RFP average fluorescence intensity ratio showing the wide variability of the transient expression of two fluorescent proteins relative to the fluorescence from the reassembly by *Npu*DnaE inteins. Cell images are false coloured. Scale bars represent 10 μm.

The *Npu*DnaE intein halves also associate in only one orientation. This was demonstrated by creating precursors that had fluorescent protein fragments fused on opposing ends (i.e. one fragment occupies the $Extein_N$ position while the other fragment occupies the $Endo_C$ position). When the following pairs of precursors were expressed in mammalian cells ($Venus_N$-*Npu*$DnaE_N$ with $Venus_C$-*Npu*$DnaE_C$; *Npu*DnaE-$Venus_N$ with *Npu*$DnaE_C$-$Venus_C$), no apparent fluorescence was detected. Likewise, when the precursors, $Venus_N$-*Npu*$DnaE_N$-$mRFP_N$ and $Venus_C$-*Npu*$DnaE_C$-$mRFP_C$, were co-expressed in mammalian cells, there was no detectable fluorescence. If the *Npu*DnaE intein halves had coupled together in any other orientation, fluorescence would

have been detected in all cases. This is not surprising as *Npu*DnaE intein show highly conserved charge segregation with specific acidic residues on *Npu*DnaE$_N$ concentrated on surfaces exposed to *Npu*DnaE$_C$ where conserved basic residues are found [119-121].

For the remainder of the study in the chapter, the coupling of the *Npu*DnaE intein halves (i.e. split protein fragments located on the Endo position) will be used to reassemble a split fluorescent protein reporter for several reasons. First, the restored fluorescence helps in visualizing non-fluorescent proteins formed from PTS. Second, it can be used as a reporter and measure of the expression levels of proteins formed from PTS. Lastly, split fluorescent protein fragments can be non-covalently associated to form function fluorescent proteins and has been extensively used by other groups [3, 10, 157, 158].

5.2.3 *Npu*DnaE PTS Can Restore RhoA Activity and Venus Fluorescence

Mammalian cells expressing a split RhoA and split Venus fused to its respective *Npu*DnaE intein halves successfully underwent PTS to restore RhoA function by splicing and Venus fluorescence through non-covalent association of the inteins. The dominant positive mutant form of RhoA (Q63L) was used because successful restoration of protein function can be visualized by non-apoptotic dynamic blebbing in epithelial-like cells [107-109]. RhoA was split within the small turn between amino acids Asp49 and Gly50. The small turn was chosen as the split site because it projects outward from the protein and thus is surface exposed and it separates RhoA into two modular fragments with well-defined secondary structures so as to minimize disturbance to the RhoA structure and solubility. The split fluorescent protein, Venus, was used as a fluorescence reporter because it emits relatively strong fluorescence as confirmed from the above experiments. The precursors RhoA$_N$-*Npu*DnaE$_N$-Venus$_N$ and Venus$_C$-*Npu*DnaE$_C$-RhoA$_C$ were created, where RhoA$_N$ is the fragment from residues 1 to 49 and RhoA$_C$ from 50 to 193 (Figure 5.5).

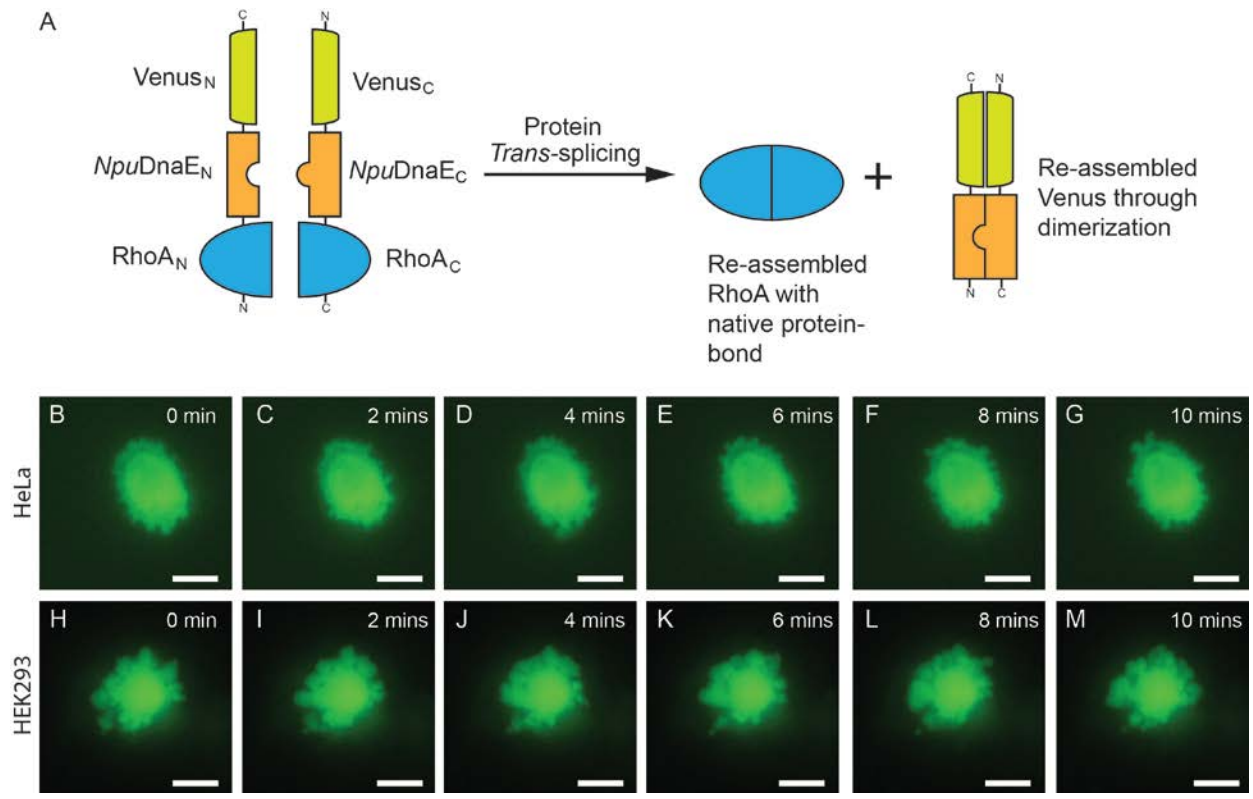

Figure 5.5 Simultaneous reassembly of split RhoA and split Venus.

(**A**) Schematic diagram for the *Npu*DnaE intein mediated reassembly of split RhoA with a peptide bond and the restoration of Venus fluorescence through the coupling of the intein halves. Time course of dynamic blebbing in (**B-G**) HeLa and (**H-M**) HEK293 cells over 10 mins. Cell images are false coloured. Scale bars represent 10 µm.

HeLa cells co-expressing the two precursors were easily located by detecting the restored Venus fluorescence. In addition, co-expressing cells sustained large continuous dynamic blebbing over a 10 minute period (Figure 5.5 and Movie 5.1). Persistent blebbing continued when viewed 4 hours after the initial imaging experiments denoting that cells were still healthy and were not as a result of apoptosis. Mammalian cells expressing either construct alone had no visible blebbing nor exhibited fluorescence. Also, mammalian cells expressing Venus alone did not undergo dynamic blebbing. Results were repeated in HEK293 cells with similar outcomes (Figure 5.5 and Movie 5.2).

5.2.4 *Npu*DnaE PTS Can Restore GCaMP2 Activity and mRFP Fluorescence

Mammalian cells expressing a split GCaMP2 Ca^{2+} biosensor and split mRFP fused to its respective *Npu*DnaE intein halves successfully underwent PTS to restore GCaMP2 function through splicing and mRFP fluorescence through non-covalent association of the inteins. The mutant form of GCaMP2, as described in Chapter 4, was used. The split fluorescent protein, mRFP, was utilized because it expresses fluorescence well and the peak fluorescence emitted from mRFP is sufficiently separated from the peak fluorescence emission of GCaMP2 so as to not contribute to background fluorescence during Ca^{2+} measurements. The N- and C-terminal precursors were constructed as GCaMP2$_N$-*Npu*DnaE$_N$-mRFP$_N$ and mRFP$_C$-*Npu*DnaE$_C$-GCaMP2$_C$, respectively (Figure 5.6). HeLa cells co-expressing the two precursors were located by spotting for emitted red fluorescence from the restored mRFP since GCaMP2 fluorescence is typically low in the absence of Ca^{2+}. The reassembled GCaMP2 biosensor was able to detect UTP-induced Ca^{2+} transients and produced similar Ca^{2+} profiles as those measured by a control unaltered GCaMP2 (Figure 5.6). Dynamic ranges obtained by loading cells with ionomycin in the presence of $CaCl_2$ to obtained maximum cytosolic Ca^{2+} followed by the loading of excess EDTA to chelate the Ca^{2+} generated statistically similar dynamic ranges as those from the control unaltered GCaMP2 biosensor (2.56 ± 0.60 and 2.54 ± 0.16, respectively; $p = 0.91$ $n = 3$ experiments; each experiment had > 3 cells) (Figure 5.6). Similar results were obtained in COS7 cells. As a control, cells expressing either GCaMP2$_N$-*Npu*DnaE$_N$-mRFP$_N$ or mRFP$_C$-*Npu*DnaE$_C$-GCaMP2$_C$ precursors alone did not have GCaMP2 activity nor expressed any measurable fluorescence.

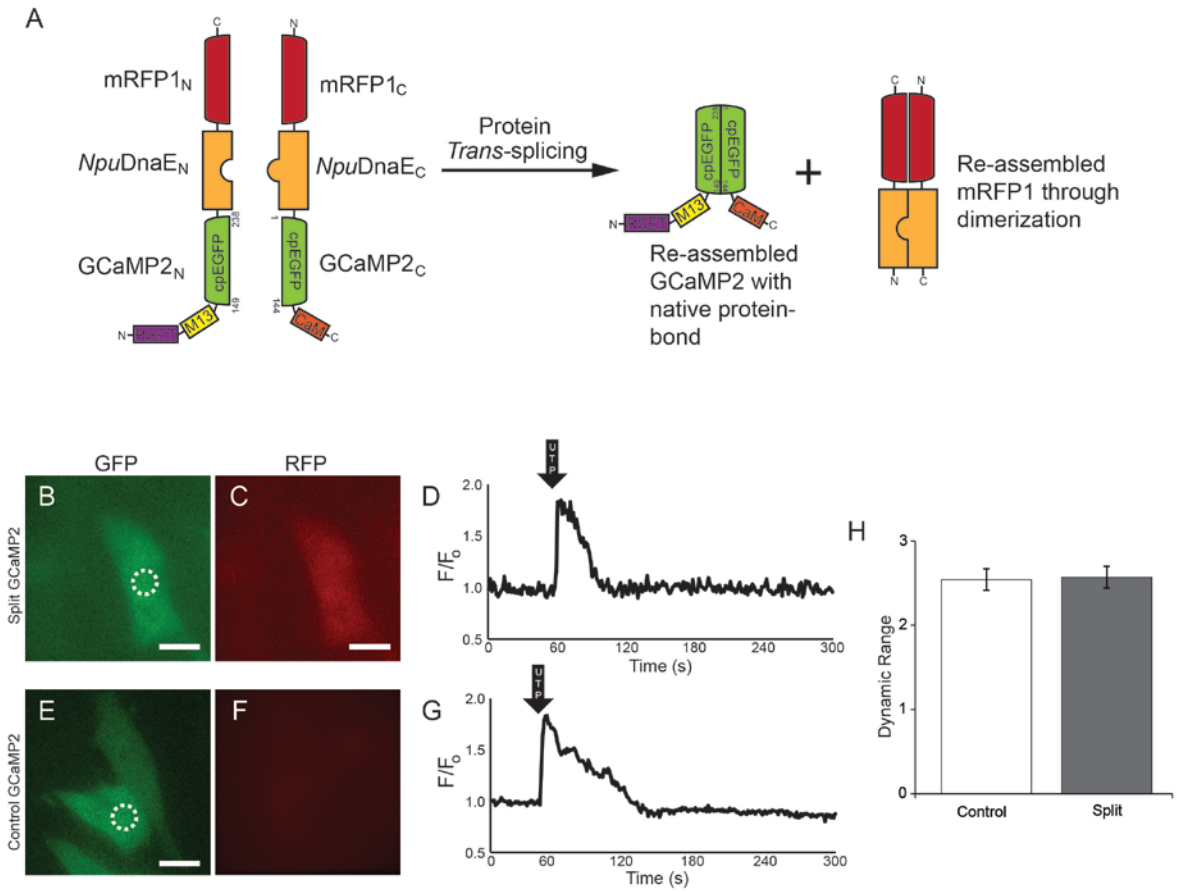

Figure 5.6 Simultaneous reassembly of split GCaMP2 and split mRFP.

(**A**) Schematic diagram for the *Npu*DnaE intein mediated reassembly of split GCaMP2 with a peptide bond and the restoration of mRFP fluorescence through the couple of the intein halves. (**B**) YFP and (**C**) RFP fluorescence images of HeLa cells co-expressing the precursor halves and (**D**) its UTP-induced Ca^{2+} transient profile as measured by the reassembled GCaMP2 Ca^{2+} biosensor. (**E**) YFP and (**F**) RFP fluorescence images of HeLa cells expressing unaltered GCaMP2 and (**G**) its UTP-induced Ca^{2+} transient profile. Note that in this case, no RFP fluorescence is observed. (**H**) Comparison of mean dynamic range of the unaltered (control) and split GCaMP2. White dotted circle represent regions where average fluorescence intensity measurements were taken. Error bars represent standard deviation. Cell images are false coloured. Scale bars represent 10 µm.

5.3 Discussion

Inteins have primarily been used to reassemble a single target protein such as a fluorescent protein reporter or to post-translationally modify target proteins. In particular, naturally split inteins have been exploited because of its fast and efficient means of ligating polypeptides together. In the current study, it was shown that there is a distinct possibility of using naturally split inteins, to restore protein function to another target protein through the high affinity binding of the intein halves. The naturally occurring split *Npu*DnaE intein was used to reassemble several different proteins – fluorescent proteins, a mutant RhoA GTPase, and the GCaMP2 Ca^{2+} biosensor – through PTS activity while simultaneously restoring function to a split fluorescent protein reporter through non-covalent interactions of the inteins. Unlike some heterologous expressed proteins, the excised intein complex has no cross-reactivity with endogenous proteins [37], do not appear to interfere with endogenous cell functions, and seems to linger about within the cell. This lends well to the restoration of the second target proteins as the intein complex is not immediately shuttle off for degradation and thus, the restored secondary target protein can remain within the cell for extended periods (i.e. well long enough for the proper folding and maturation of the fluorescent protein fluorophore). *Npu*DnaE was also shown to associate in a specific orientation. This could be beneficial in applications where two highly related proteins are fused to the intein non-splicing and splicing ends, such as demonstrated above in the restoration of fluorescence of two split fluorescent proteins. Also, knowledge that the inteins associate in one orientation eases the design strategy and process by eliminating the need to account for adverse side reactions and its products. Furthermore, it was determined that the two target proteins are restored in consistent stoichiometric ratios such that the increased of fluorescence of a fluorescent protein reporter could be used as an indicator of an increased expression of another protein.

The strategy of restoring two target proteins was then applied to RhoA and Venus. PTS activity restored protein function to split RhoA mutant. The reassembled RhoA represents the first attempt of restoring function to a protein that affects cell function (i.e. causes dynamic blebbing). In particular, during the construction of the precursors, no attempts were made to replace, or otherwise alter, the small turn flanking the split site to accommodate the PTS residual peptides,

yet protein function was still restored. This suggests that not all proteins require immediate attention to mitigate the effects of the PTS residual peptide and those that do are often highly dependent on a particular orientation or separation (i.e. TN-XL) or are dissected within secondary or tertiary structures. Also, it appears that the current PTS residual peptide is non-reactive and sufficiently flexible as to act as a small peptide turn which may be important in deciding split sites in other proteins.

*Npu*DnaE was also used to restore function to split GCaMP2 and split mRFP. One of the major drawbacks of the reassembly of GCaMP2 in Chapter 4 was that when applied to live *C. Elegans*, it was difficult, if not impossible to separate movement related artefacts from the GCaMP2 measured fluorescence. A reference fluorescence could minimize the effects of these artefacts that are unrelated to changes in free Ca^{2+} concentration. A simple approach would be to express a reference fluorescence in cells around the area. However, when specific cells are targeted using two promoters (see Chapter 8), the surrounding fluorescence could impact the accuracy of the reference fluorescence. Ideally, a reference fluorescence emanating from the cells being measured would provide better accuracy than from surrounding cells. With the current strategy, the restored fluorescence from a secondary fluorescent protein (i.e. mRFP) would provide a source of reference within the cell being measured. In addition, the red fluorescence emanating from the restored mRFP is particular useful in in live animal studies as the emitted fluorescence can be separated from the changing Ca^{2+}-sensitive fluorescence (from GCaMP2).

5.4 Conclusion

In this chapter, a single *Npu*DnaE intein was able to simultaneously reassemble two target proteins – one through the normal PTS activity and the second through non-covalent intein association. This strategy was applied to three different split proteins while exploiting the high affinity binding to restore a split fluorescent protein reporter.

1. It was confirm that the association of the *Npu*DnaE intein halves occur in a specific orientation which is determined by intermolecular ion pair binding.

2. The two protein targets are restored in a consistent stoichiometric ratio. This can be used to judge the expression level of one protein based on the fluorescence intensity of another reporter fluorescent protein.

3. The fluorescence emanating from the non-covalent restoration of the split fluorescent protein can be used to i) locate cells that successfully underwent the PTS process and ii) can be used as a reference fluorescence in live animal studies.

Referring to the research objectives as outlines in Chapter 1, objective 2 was addressed in this chapter. Taking the findings as determined in the previous chapter and merging it with a commonly used split fluorescent protein strategy (BiFC), a naturally split intein pair was shown to be able to simultaneously reassemble two separate proteins. This represents a conceptual improvement over findings from the previous chapter, particular with regards to live animal model studies, and also presents a novel use of naturally split inteins.

6 LOVInC – A Photoactivatable Split Intein

The content of this chapter was modified from the following submitted paper:

Stanley SC Wong, Evan Mills, Kevin Truong. An Engineered Naturally Split Intein for Photoactivated Protein Trans-Splicing.

6.1 Chapter Introduction and Aims

The ability to control the PTS activity is a major step forward in engineering new protein tools for laboratory and therapeutic applications. There are currently no known natural modulators of PTS and thus, there are no natural clues to guide potential engineered approaches. Through a series of novel combination of protein and chemical engineering strategies, artificially split *cis*-splicing inteins, which often have significantly reduced coupling capacities, have been regulated by temperature [56, 57], pH [58], or small molecules [48, 59, 60]. However, these systems have inherent drawbacks that limit their broad application. Light, on the other hand, offers a potentially more desirable stimulus as it causes minimal damage to live cells or animal models and is able to be precisely tuned in both space and time. The naturally split *trans*-splicing inteins represent a far more challenging class of inteins to regulate than artificially split inteins, as the intein halves naturally have a strong tendency to associate. A couple of groups have successfully controlled naturally split inteins using light [61, 62]. However, the two groups chemically cage the intein halves which require the use synthetic protein synthesis to construct the precursors. Here, a genetically encoded LOVInC intein was engineered by exploiting the light sensitive protein domain, LOV2, to generate photoactivatable PTS. This complements existing systems by allowing more complex proteins to be incorporated in a modular way and can be easily expressed in biological systems by genetic means.

For this chapter, the following four specific aims were established to achieve the overall research goal:

1. Determine a minimal C-terminal intein that can still undergo PTS but removing amino acids from the N-terminus.

2. Investigate whether the fusion of LOV2 to the truncated C-terminal intein can inhibit PTS activity.

3. Characterize the photoactivatable PTS activity.

4. Apply design to various protein examples to demonstrate its applicability.

In this chapter study, several engineered fusion proteins were created to address the specific aims and demonstrate photoactivatable PTS activity. All engineered proteins were created in accordance with the gene construction methodologies as presented in Chapter 3 (Figure 6.1).

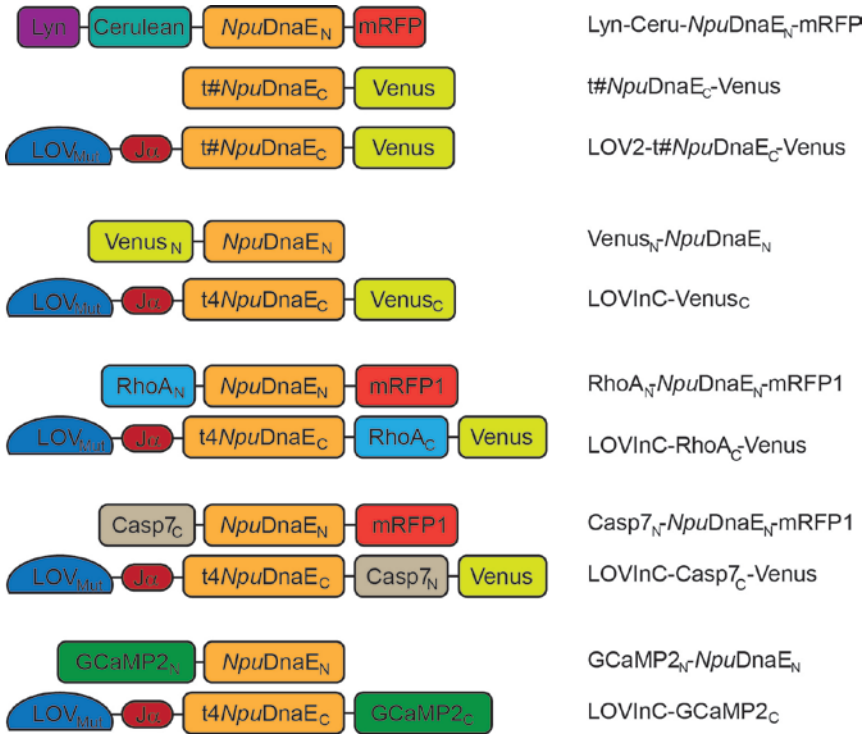

Figure 6.1 Outline of the fusion proteins used in this chapter.

Block diagrams of the fusion protein are shown. The names of the fusion proteins used in this chapter are also given.

6.2 Results

This chapter describes the experimental results in the development of a photoactivatable protein *trans*-splicing by a naturally split intein. A general overview of the strategy will be given (section 6.2.1). This is followed by the determination of the minimal function *trans*-splicing unit (section 6.2.2). The design of the photoactivatable intein will be created and tested (section 6.2.3). Improvements to the design will be made with mutant variants (section 6.2.4). This is followed by light-mediated reassembly split Venus (section 6.2.5), split RhoA (section 6.2.6), split caspase-7 (section 6.2.7), and split GCaMP2 (section 6.2.8) to validate the photoactivatable intein. Finally, the spatiotemporal control of the photoactivatable intein will be demonstrated (section 6.2.9).

6.2.1 General Strategy for Photoactivatable PTS

The general strategy set out to generate a photoactivatable intein was to determine if a light-sensitive effector domain, when attached to *Npu*DnaE intein, can be used as an allosteric switch to trigger PTS activity. Here, the LOV2 domain from *Avena sativa* was chosen as the light-sensitive effector domain as it has been shown to undergo large conformation changes when stimulated by light in the blue spectrum. In a relaxed, unstimulated state (dark-state), LOV2 resumes a compact form where its carboxyl-terminal helical extension (Jα) is docked within the LOV2 core [76, 80, 81]. Illumination (lit-state) drives the formation of an adduct between the FMN chromophore and a conserved Cys residue within the LOV2 core. This induces a large conformational change that facilitates the release and unfolding of the Jα helical extension. By exploiting this unique property, it was anticipated that by attaching one of the intein halves to LOV2, LOV2 in its dark-state closed conformation could allosterically inhibit the spontaneous association of the intein halves and thus block PTS activity. Photostimulation would then release this inhibition and allow re-association of the intein precursors and subsequently allow PTS to proceed (Figure 6.2).

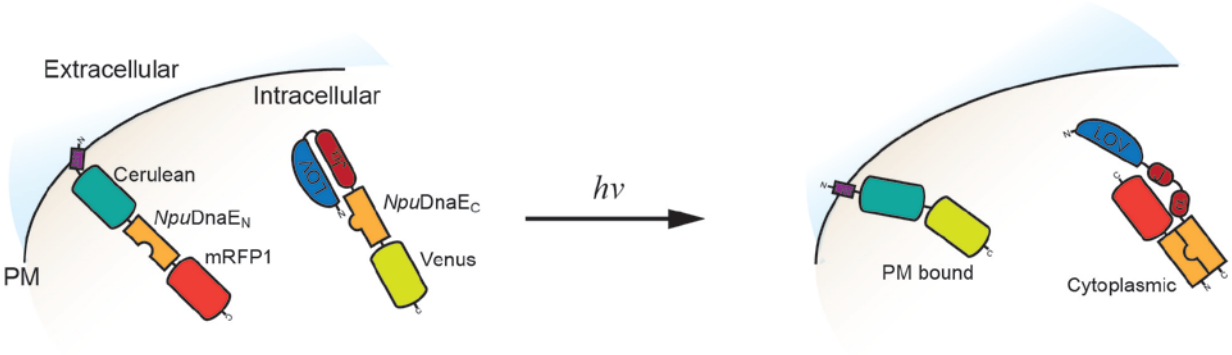

Figure 6.2 General strategy for photoactivatable PTS.

Schematic depiction of strategy employed in generating a photoactivatable PTS intein. The LOV2 domain is used as the light sensing domain. In the absence of photo-stimulation, LOV2 is bound tightly to the Jα helix which is fused directly to the N-terminal end of *Npu*DnaE$_C$. In this conformation, LOV2 allosterically inhibits PTS from occurring by blocking the binding of the intein halves. When photo-stimulated, LOV2 releases the Jα helix releasing this inhibition for PTS activity to proceed. This can be observed by looking for subcellular localizations of different fluorescent proteins (i.e. co-localization of Cerulean and Venus after PTS).

Following previous design strategies where LOV2 domains were attached to the amino-terminus of the affected protein [74, 75, 84, 159], it was decided that the *Npu*DnaE$_C$ intein half would be the better candidate to control for a couple of reasons: 1) the LOV2 domain is attached away from the splicing site and 2) more importantly, if LOV2 had been fused to *Npu*DnaE$_N$, all reassembled proteins resulting from PTS would have the LOV2 domain lodged within its structure, thus negatively affecting proper reassembled protein function. *Npu*DnaE$_N$ intein was not altered in any way in the current study.

6.2.2 Minimal Functional *Npu*DnaE$_C$ Intein

PTS activity was significantly reduced when 6 or more amino acids were removed from the first N-terminal β-strand of *Npu*DnaE$_C$. Studies from the previous two chapters demonstrated that *Npu*DnaE$_C$ can accommodate long chain peptides fused to its non-splicing ends without hindering PTS activity. It follows that the fusion of LOV2 domain (a protein of ~110 amino acids) to the amino-terminus of *Npu*DnaE$_C$ would not significantly block PTS activity. Thus, in

order to maximize the allosteric effects of LOV2 on *Npu*DnaE$_C$, amino acids were removed from the amino-terminus end of *Npu*DnaE$_C$ in hopes that by bringing the LOV2 domain closer to its binding region, the dark-state LOV2 can more effectively prevent intein association from occurring.

Amino acids were systematically truncated from the first N-terminal β-strand of *Npu*DnaE$_C$ to determine the minimal functional unit with efficient PTS activity (Figure 6.3). PTS activity was assayed using a slight variant of the technique developed in Chapter 4 whereby the occurrence of PTS activity can be confirmed by examining distinct subcellular localization of fluorescent proteins before and after PTS. The N-terminal precursor was created as a tandem fusion of Lyn-Ceru-*Npu*DnaE$_N$-mRFP while the C-terminal precursor was comprised of a tandem fusion of truncated variants of *Npu*DnaE$_C$ fused to Venus (i.e. t#*Npu*DnaE$_C$-Venus, where # denotes the number of amino acids removed from the β-strand).

Figure 6.3 Systematic amino acid truncations of *Npu*DnaE$_C$.

Schematic representation of the amino acids that were systematically removed from the first N-terminus β-strand of *Npu*DnaE$_C$. All truncated *Npu*DnaE$_C$ were fused to Venus at the C-terminus end to facilitate visual observations and the PTS assay. Orange arrows denote β-strands of *Npu*DnaE$_C$. Truncations of *Npu*DnaE$_C$ are denoted as t#, where # represents the number of amino acids removed.

In the absence of PTS activity, the PM of mammalian cells would be labelled with cyan and red fluorescence emanating from the membrane bound fusion of Cerulean and mRFP whereas the cytoplasm would be labelled with yellow fluorescence (Figure 6.4). PTS activity would remove the mRFP (along with the coupled intein) from the PM releasing it into the cytoplasm and labelling it with red fluorescence while concurrently, Venus fluorescent proteins would translocate from the cytoplasm to the PM be ligated with the PM-bound Cerulean labelling the PM with cyan and yellow fluorescence (Figure 6.4).

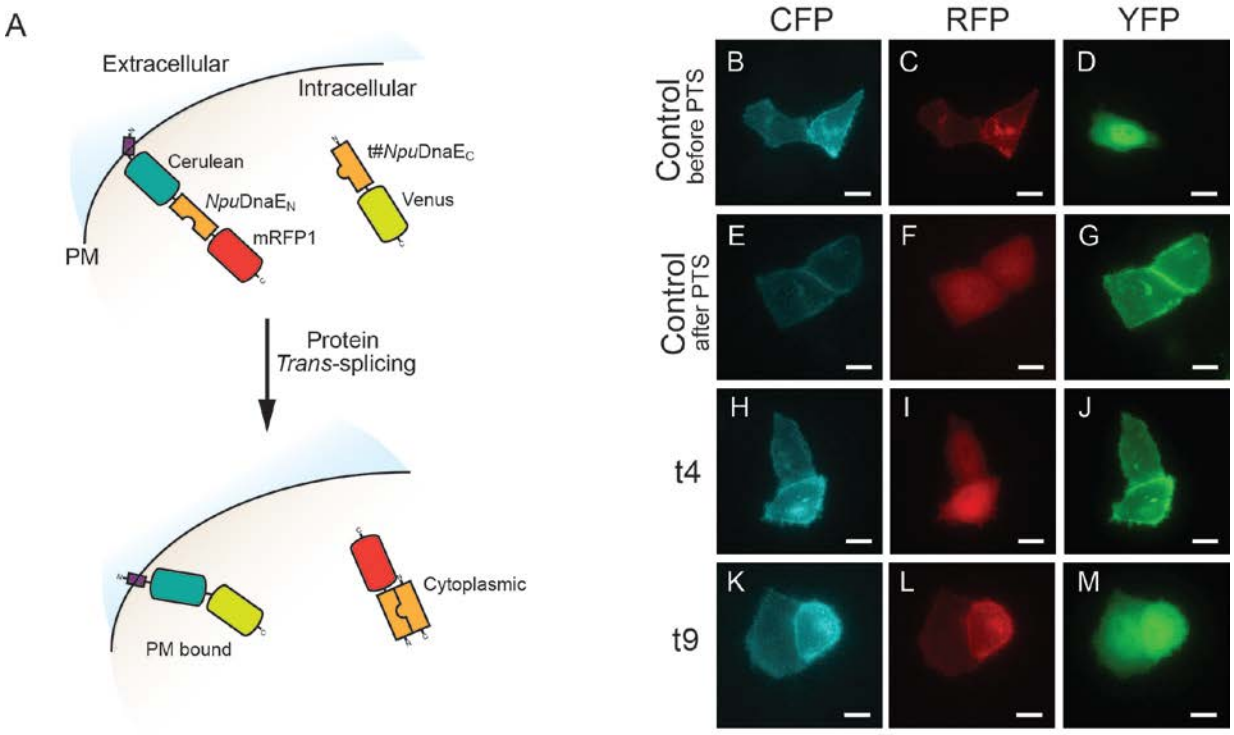

Figure 6.4 Determining minimal functional *Npu*DnaE$_C$ with PTS activity.

(**A**) Schematic diagram of the fusion protein constructs used for determining PTS activity where Lyn is a PM localization sequence (purple). Successful PTS activity would result in PM localization of Cerulean and Venus while mRFP would translocate from the PM to the cytoplasm. HeLa cells expressing Lyn-Ceru-*Npu*DnaE$_N$-mRFP resulted in (**B**) CFP and (**C**) RFP fluorescence localized to the PM. HeLa cells expressing *Npu*DnaE$_C$-Venus resulted in (**D**) cytoplasmic YFP fluorescence. Co-expression of the two precursors resulted in PTS activity with the PM localization of (**E**) CFP and (**G**) YFP fluorescence and (**F**) cytoplasmic RFP fluorescence. Representative *Npu*DnaE$_C$ truncations: (**H-J**) t4 showing complete PTS activity while (**K-M**) t9 showing suppressed PTS activity, under CFP, RFP, and YFP respectively. *$p < 0.001$, **$p < 0.0005$. Cell images are false coloured. Scale bars represent 10 μm.

The efficacy of each truncated *Npu*DnaE$_C$ was determined by co-expressing each C-terminal precursor variant with the N-terminal precursor and counting the number of cells that had successfully undergone PTS activity as described. If all co-expressing cells show the interchange between red and yellow fluorescence, PTS activity was considered efficient and complete. Otherwise, the percentage of cells that underwent PTS activity was determined. As a control, HeLa cells were co-expressed with the N-terminal precursor and the wild type *Npu*DnaE$_C$ fused to Venus (wt*Npu*DnaE$_C$-Venus) and resulted in complete PTS activity (Figure 6.4).

What was found was that the removal of 5 or less amino acids from the β-strand had no effect on PTS activity as all co-expressing cells underwent complete PTS (not statistically different) whereas when 6 or more amino acids were removed, PTS activity decreased substantially (statistically significant with $p < 0.001$) (Figure 6.5). From these results, truncation variants, t1-t5, of *Npu*DnaE$_C$ were used in subsequent studies.

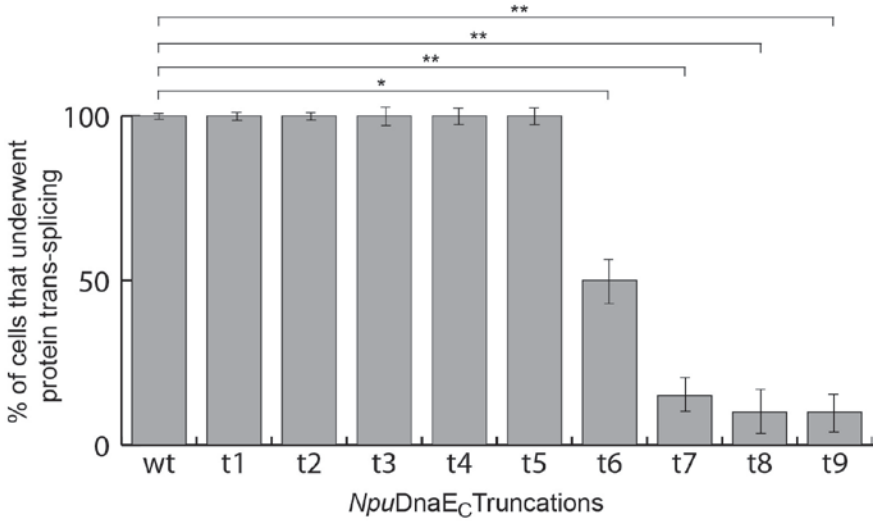

Figure 6.5 Splicing activity of truncated *Npu*DnaE$_C$.

The percentage of cells that underwent protein splicing using truncated mutants of *Npu*DnaE$_C$. Truncations t1-t5 had underwent near complete PTS activity, similar to wild type. When 6 or more amino acids were removed from the β-strand, PTS activity significantly dropped. Comparison between wild-type: *$p < 0.005$, **$p < 0.001$. Error bars represent standard deviation.

6.2.3 Design of LOV *Npu*DnaE$_C$ Intein

Photoactivatable *Npu*DnaE inteins were created by fusing the LOV2 domain to t4 and t5 truncated mutants of *Npu*DnaE$_C$. In theory, a particular fusion of LOV2 to any t1-t5 truncated mutants of *Npu*DnaE$_C$ could allow the dark-state closed conformation of LOV2 to allosterically inhibit the association of intein halves and prevent PTS activity from occurring (Figure 6.6).

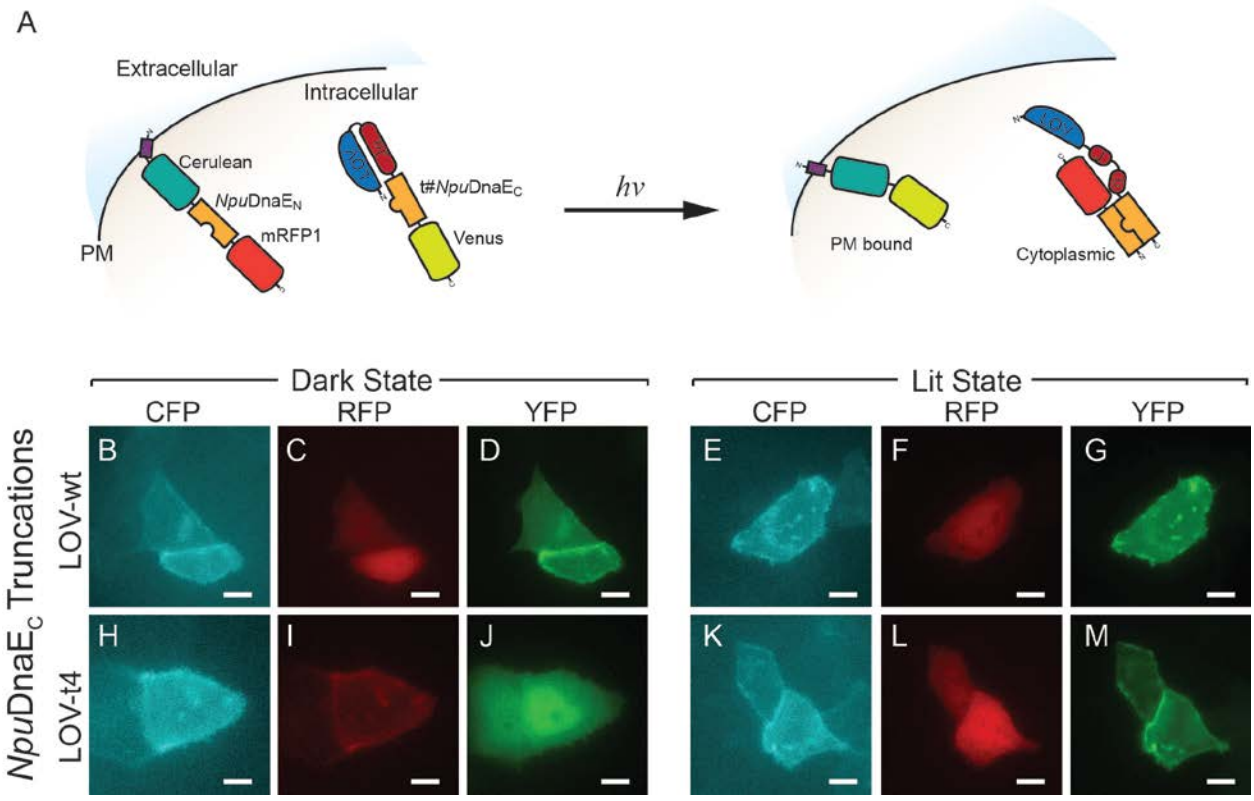

Figure 6.6 Photoactivatable Intein with LOV2 domain.

(**A**) General schematic diagram of photoactivatable *Npu*DnaE$_C$ intein before and after photo-stimulation with blue-light (dark- and lit-state, respectively). Blue-light photo-stimulation releases *Npu*DnaE$_C$ inhibition and allows PTS to proceed. This will result in Cerulean and Venus being PM-bound while mRFP will be translocated from the PM to the cytoplasm. LOV2 fused to wild-type *Npu*DnaE$_C$ exhibited constitutive PTS activity in both the (**B-D**) dark- and (**E-F**) lit-states. On the other hand, LOV2 fused to t4*Npu*DnaE$_C$ did not splice in the (**H-J**) dark-states, whereas (**K-M**) blue-light photo-stimulation induced PTS activity. Images are shown with CFP, RFP, and YFP fluorescence filters, respectively. Cell images are false coloured. Scale bars represent 10 μm.

Blue-light illumination triggers the open form conformation of LOV2 and could release the inhibition and allow intein association to recommence. As expected, the fusion of LOV2 to wild-type *Npu*DnaE$_C$ had no effect on PTS activity as confirmed by HeLa cells co-expressing the N-terminal precursor, Lyn-Ceru-*Npu*DnaE$_N$-mRFP, with the C-terminal precursor LOV2-wt*Npu*DnaE$_C$-Venus (i.e. tandem fusion of LOV2 with wild type *Npu*DnaE$_C$ and Venus) (Figure 6.6 and Figure 6.7). All cells exhibited constitutive PTS activity in both the dark- and lit states as observed by cytoplasmic localization of red fluorescence and PM localization of cyan and yellow fluorescence (Figure 6.6 and Figure 6.7). Likewise, fusion of LOV2 to *Npu*DnaE$_C$ truncations t1 and t2 had no effect on PTS activity and were statistically similar to the wild-type ($p > 0.98$). On the other hand, fusion of LOV2 to *Npu*DnaE$_C$ truncation t3 to t5 affected PTS activity with the greatest change occurring with truncation t4 and t5 (~40% and ~66% increase in PTS activity from dark- to lit-state, respectively; $n = 6$ experiments with at least 20 cells/experiment) (Figure 6.7), whereas truncation t3 was only minimally affected (~18% increase in PTS activity from dark- to lit-state; $n = 6$ experiments with at least 20 cells/experiment) (Figure 6.7). Thus, truncations t4 and t5 of *Npu*DnaE$_C$ were further investigated.

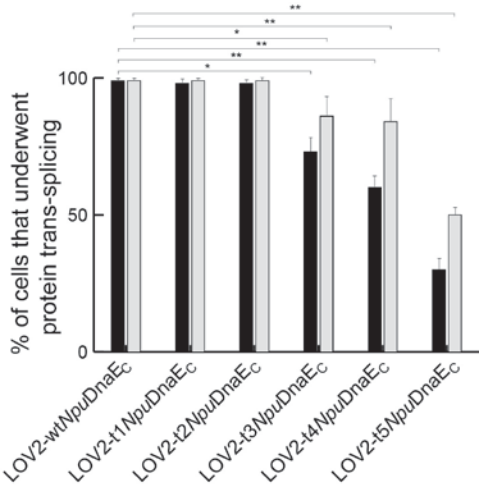

Figure 6.7 Splicing activity of LOV2 fused to truncated *Npu*DnaE$_C$.

Graph showing the percentage of cells that underwent PTS activity between the dark- and lit-states. Truncations t1-t2 did not show appreciable PTS activity between the dark- and lit-states, and were similar to wt. Truncations t4 and t5 showed the largest difference between dark- and lit-states. Dark-states are represented as black bars while lit-states are shown as grey bars. Comparison between wild-type: $*p < 0.01$, $**p < 0.005$. Error bars represent standard deviation.

6.2.4 LOV2 Mutants Improved Photoactivatable *Npu*DnaE$_C$ Intein

The fusion of LOV2$_{C450M}$ mutant to t4*Npu*DnaE$_C$ had the largest change in PTS activity before and after photo-stimulation. Although the fusion of LOV2 to t4*Npu*DnaE$_C$ or t5*Npu*DnaE$_C$ appeared to suppress PTS activity in both cases, the dark-states still suffered from considerable PTS activity before photo-stimulation by blue-light. To improve on this, mutations that preferentially stabilize the LOV2 docked conformation was introduced into LOV2 with the goal of further suppressing PTS activity in the dark-state and potentially increasing the change in PTS activity between the dark- and lit-state. The LOV2$_{I532A}$ mutant, which has been shown to reduce dark-state activation by 75% in other designs [160], was used to replace the LOV2 domain in the two C-terminal precursors to generate LOV2$_{I532A}$-t4*Npu*DnaE$_C$-Venus and LOV2$_{I532A}$-t5*Npu*DnaE$_C$-Venus. However, the co-expression of Lyn-Ceru-*Npu*DnaE$_N$-Venus with either LOV2$_{I532A}$-t4*Npu*DnaE$_C$-Venus or LOV2$_{I532A}$-t5*Npu*DnaE$_C$-Venus did not further suppress dark-state PTS activity nor did it increase the change between dark- and lit-state PTS activity (~49% and ~63% increase in PTS activity from dark- to lit-state, respectively; n = 6 experiments with at least 20 cells/experiment) (Figure 6.8). As controls, the two mutants, LOV2$_{I539E}$ and LOV2$_{C450M}$, were used that purportedly mimic constitutively lit- or dark-state by promoting the docking or undocking of the Jα helical peptide within the LOV2 core, respectively [80-82]. The constitutively active precursor states were generated by replacing the LOV2 domain in the two C-terminal precursors with LOV2$_{I539E}$ to give LOV2$_{I539E}$-t4*Npu*DnaE$_C$-Venus and LOV2$_{I539E}$-t5*Npu*DnaE$_C$-Venus. As expected, HeLa cells co-expressing Lyn-Ceru-*Npu*DnaE$_N$-Venus with either active precursors sustained complete PTS activity irrespective of blue-light photo-stimulation (Figure 6.8). In a similar manner, the dark-state mutant precursors were generated by replacing the LOV2 domain with LOV$_{C450M}$ to yield the precursors LOV2$_{C450M}$-t4*Npu*DnaE$_C$-Venus and LOV2$_{C450M}$-t5*Npu*DnaE$_C$-Venus. In HeLa cells co-expressing Lyn-Ceru-*Npu*DnaE$_N$-Venus and LOV2$_{C450M}$-t5*Npu*DnaE$_C$-Venus, PTS activity was significantly suppressed irrespective of blue-light photo-stimulation, as expected (Figure 6.8). Surprisingly, in HeLa cells co-expressing the precursors Lyn-Ceru-*Npu*DnaE$_N$-Venus and LOV2$_{C450M}$-t4*Npu*DnaE$_C$-Venus, there was diminished PTS activity in the absence of photo-stimulation but had a four-fold increase in PTS activity when photo-stimulated with blue-light (Figure 6.8). Given the strong

affinity of *Npu*DnaE intein halves, even marginal light-sensitivity of the LOV_{C450M} mutant may be sufficient to trigger *Npu*DnaE association and PTS [82, 160].

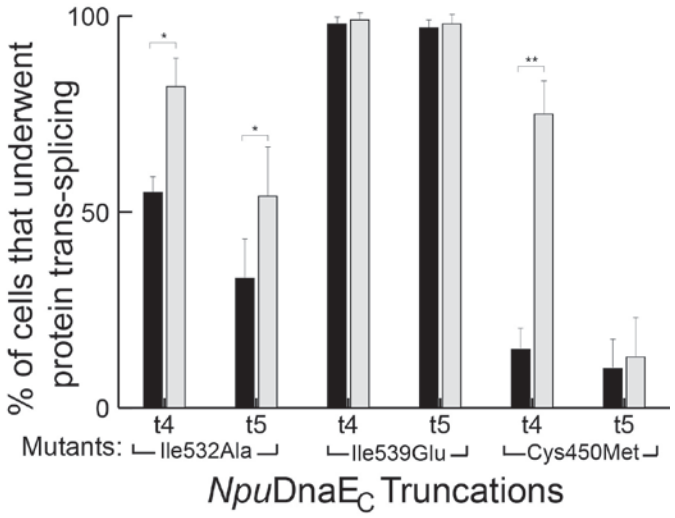

Figure 6.8 Splicing activity with mutant LOV2 fused to truncated *Npu*DnaE$_C$.

Graph showing the percentage of cells that underwent PTS activity between the dark- and lit-states. The I539E mutation did not significantly affect PTS activity between the dark- and lit-states. The I532A mutant did not suppress dark-state PTS activity compared to non-mutated LOV2 (Figure 7.7). Mutation C450M seem to have the greatest effect on *Npu*DnaE$_C$ truncation t4. $*p < 0.01$, $**p < 0.001$.

The length of time for the precursor $LOV2_{C450M}$-t4*Npu*DnaE$_C$-Venus to accumulate sufficient quantities of PTS products after photo-stimulation was determined by tracking the co-localization of cyan and yellow fluorescence and quantified by PC. HeLa cells co-expressing Lyn-Ceru-*Npu*DnaE$_N$-Venus and $LOV2_{C450M}$-t4*Npu*DnaE$_C$-Venus and grown in the absence of light initially had the Cerulean fluorescent protein PM-bound while the Venus was localized to the cytoplasm. Blue-light photo-stimulation (provided by the microscope filters set at 438 nm for 1 second every 30 second interval) initiated PTS activity with the subsequent translocation of Venus from the cytoplasm to the PM. Significant PTS activity occurred approximately 120 ± 23 minutes (n = 6) after the start of photo-stimulation (Figure 6.9). A fluorescent SDS-PAGE confirmed the formation of spliced products after blue-light photo-stimulation. Due to the increase in PTS activity between the dark- and lit-state of $LOV2_{C450M}$-t4*Npu*DnaE$_C$ (LOVInC,

hereafter), it was further applied to engineer photo-activatable reassembly of split proteins such as Venus, RhoA, Caspase-7 and GCaMP2.

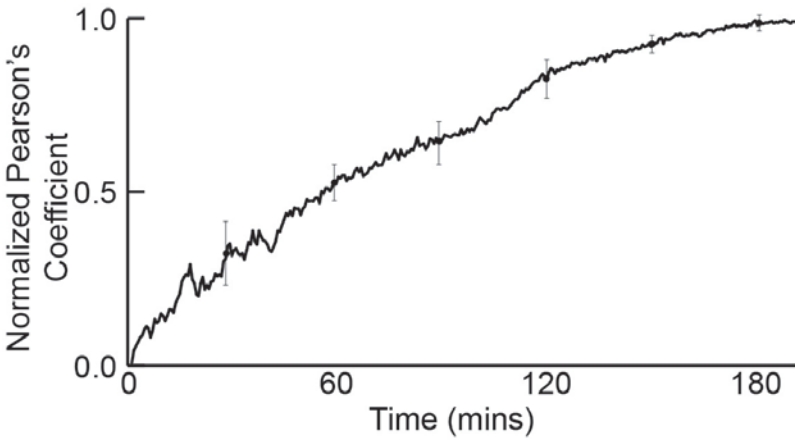

Figure 6.9 LOVInC undergoing PTS activity after photo-stimulation.

Time course showing co-localization of CFP and YFP after blue-light photo-stimulation as measured by normalized Pearson`s coefficient. Significant PTS activity has occurred after approximately 120 minutes after photo-stimulation. Error bars represent standard deviation.

6.2.5 Photoactivatable Reassembly of Split Venus Mediated by LOVInC

Blue-light photo-stimulation induced PTS activity to reassemble a split Venus and restore fluorescence in mammalian cells. The split site for dividing Venus into its fragments was as previously described in Chapter 5. Venus was chosen because it is a representative variant of the family of GFPs that are commonly used as fluorescent reporters. The C-terminal precursor was comprised of a tandem fusion of LOVInC to $Venus_C$ while the N-terminal precursor was a tandem fusion of $Venus_N$ to *Npu*$DnaE_N$ ($Venus_N$-*Npu*$DnaE_N$) (Figure 6.10). Mammalian cells (HeLa and COS7) co-expressing the two precursors and photo-stimulated with blue-light overnight restored Venus fluorescence similar to the restored fluorescence of split Venus without the use of the LOV2 domain in Chapter 5 (n = 6) (Figure 6.10). In contrast, co-expressing cells grown in the absence of photo-stimulation produced no detectable fluorescence (n = 6). Cells

that co-expressed either precursor along have no apparent fluorescence irrespective of overnight photo-stimulation or not.

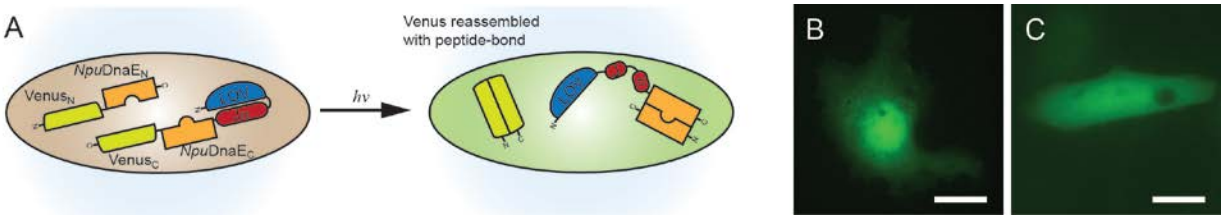

Figure 6.10 Photoactivatable reassembly of split Venus.

(**A**) General schematic depiction of the reassembly and restoration of Venus fluorescence after blue-light photostimulation. (**B**) COS7 and (**C**) HeLa cells co-expressing the two precursors and photo-stimulated with blue-light exhibit Venus fluorescence. Cell images are false coloured. Scale bars represent 10 µm.

6.2.6 Photoactivatable Reassembly of Split RhoA Mediated by LOVInC

Blue-light photo-stimulation induced PTS activity to reassemble a dominant positive split RhoA mutant. The dominant positive variant of RhoA was next tested because it represents endogenous proteins that directly affect cell function. The split site for RhoA fragments are as previously described in Chapter 5. The photo-activatable C-terminal precursor was created as a tandem fusion of LOVInC, $RhoA_C$, and Venus to generate LOVInC-$RhoA_C$-Venus while the N-terminal precursor was comprised of a tandem fusion of $RhoA_N$, *Npu*$DnaE_C$, and mRFP to give $RhoA_N$-*Npu*$DnaE_N$-mRFP (Figure 6.11). The Venus and mRFP fluorescent proteins were added as fluorescent reporters to facilitate imaging. HeLa cells co-expressing the two constructs and grown overnight in the absence of photo-stimulation did not undergo continuous non-apoptotic dynamic blebbling (n = 6). On the other hand, a significant proportion of HeLa cells expressing the two precursors and photo-stimulated overnight with blue-light acquired the continuous dynamic blebbing phenotype (>80%; n = 6 independent experiments with >20 cells each) (Figure 6.11). Cells expressing either precursor alone had normal non-blebbing phenotype (n = 6) (Figure 6.11).

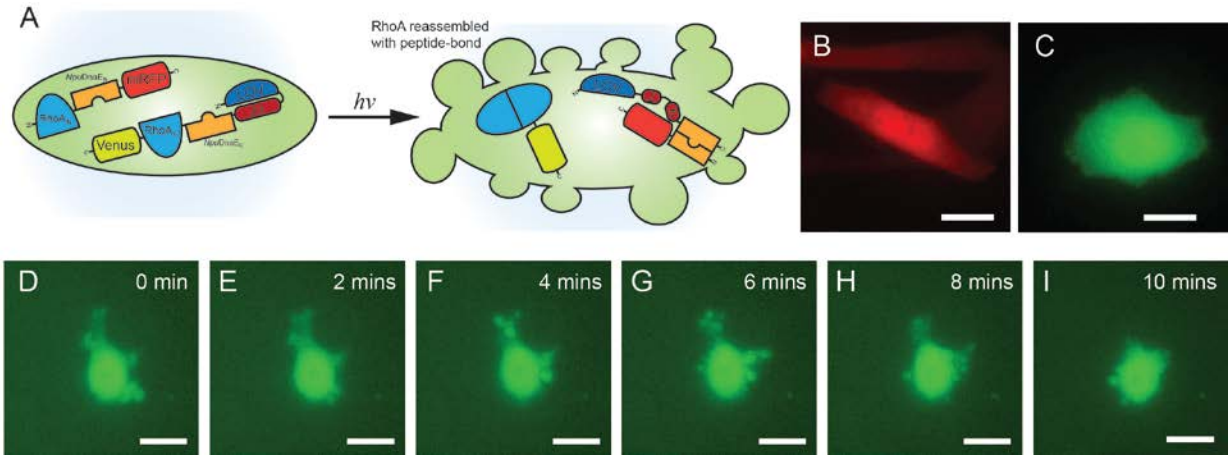

Figure 6.11 Photoactivatable reassembly of split RhoA.

(**A**) General schematic diagram depicting the reassembly and restoration of dominant positive RhoA function after blue-light photo-stimulation. (**B**) HeLa cells expressing only the N-terminal precursor show only RFP fluorescence by no dynamic blebbing phenotype. (**C**) Similarly, HeLa cells expressing only the C-terminal precursor exhibit Venus fluorescence but no dynamic blebbing. (**D-I**) Time course of HeLa cells co-expressing the two precursors showing dynamic blebbing over 10 minutes. Only the YFP fluorescence is shown but the RFP fluorescence would be the same. Cell images are false coloured. Scale bars represent 10 μm.

6.2.7 Photoactivatable Reassembly of Split Caspase-7 Mediated by LOVInC

Blue-light photo-stimulation induced PTS activity to reassemble a split caspase-7 (Casp7) that causes cell death morphologies. Casp7 is a member of the highly conserved family of tightly-regulated proteases that execute the dismantling of the cell function machinery and activate apoptosis. Typically, cells with active Casp7 expressed exhibit morphological hallmarks of cell death such as shrinking and rounding of cell and loss of nuclear membrane integrity [116, 117]. It is these hallmarks that were used to define cells that have successfully reassembled split-Casp7. Casp7 is synthesized as a complex inactive pro-Casp7 where a prodomain caps and protects Casp7 from inadvertently activating apoptosis [112] and requires self-cleavage and association of the two subunits p20 (fragment containing amino acids 57-198) and p11 (207-303) for activation [110, 113-115]. Constitutively active Casp7 can be engineered by tandem fusing the p11 subunit ($Casp7_C$) to the p20 subunit ($Casp7_N$) [110]. The N-terminal precursor was

composed of a tandem fusion of $Casp_C$, *Npu*$DnaE_N$, and mRFP to yield $Casp_C$-*Npu*$DnaE_N$-mRFP while the C-terminal precursor was a tandem fusion of LOVInC, $Casp7_N$, and Venus to give LOVInC-$Casp7_N$-Venus (Figure 6.12). The mRFP and Venus fluorescent proteins were used strictly as fluorescent reporters. The separate expression of the precursors in HeLa cells did not alter cell morphologies (Figure 6.12). Likewise, HeLa cells co-expressing the two precursors and grown in the absence of photo-stimulation did not exhibit considerable cell death morphology changes (16.7 ± 7.8%; n = 6 independent experiments with >20 cells each) (Figure 6.12). On the other hand, co-expressing cells when photo-stimulated overnight by blue-light induced significant cell death morphological changes (87.3 ± 9.1%; n = 6 independent experiments with >20 cells each) (Figure 6.12).

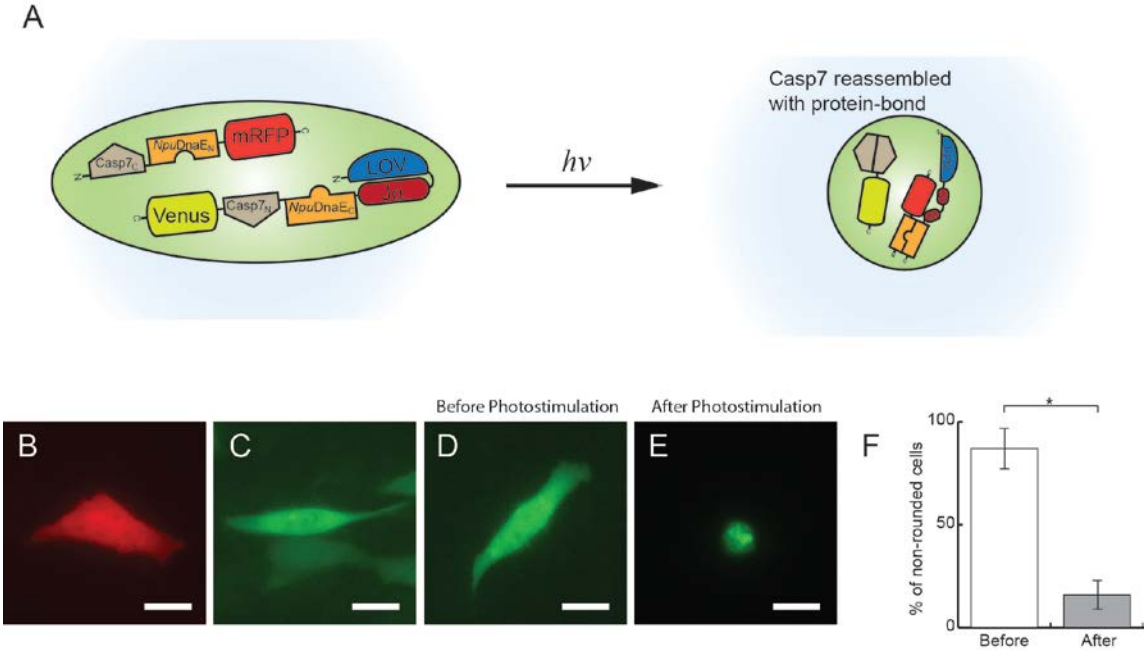

Figure 6.12 Photoactivatable reassembly of split Casp7

(**A**) General schematic diagram depicting the reassembly and restoration of Casp7 function after blue-light photo-stimulation. (**B**) HeLa cells expressing only the N-terminal precursor exhibited RFP fluorescence but did not have apoptotic-like morphology. (**C**) Similar, HeLa cells expressing only the C-terminal precursor expressed YFP fluorescence but had normal cell morphology. (**D**) HeLa cells co-expressing the two precursors and imaged before photo-stimulation. (**E**) After photo-stimulation, cells underwent apoptotic-like morphology such as the shrinking and rounding of cells. Only the YFP fluorescence is shown, but the RFP fluorescence would be the same. (**F**) Percentage of non-rounded cells before and after photo-stimulation. *p < 0.001. Cell images are false coloured. Scale bars represent 10 μm.

6.2.8 Photoactivatable Reassembly of Split GCaMP2 Mediated by LOVInC

Blue-light photo-stimulation induced PTS activity to reassemble a split GCaMP2 Ca^{2+} biosensor that allowed Ca^{2+} measurements to be made. The split site for the GCaMP2 fragments was as previously described in Chapter 5. The N-terminal construct was composed of a tandem fusion of $GCaMP2_N$ and *Npu*$DnaE_N$ to form $GCaMP2_N$-*Npu*$DnaE_N$ while the C-terminal precursor was composed of a tandem fusion of LOVInC and $GCaMP2_C$ to form LOVInC-$GCaMP2_C$ (Figure 6.13).

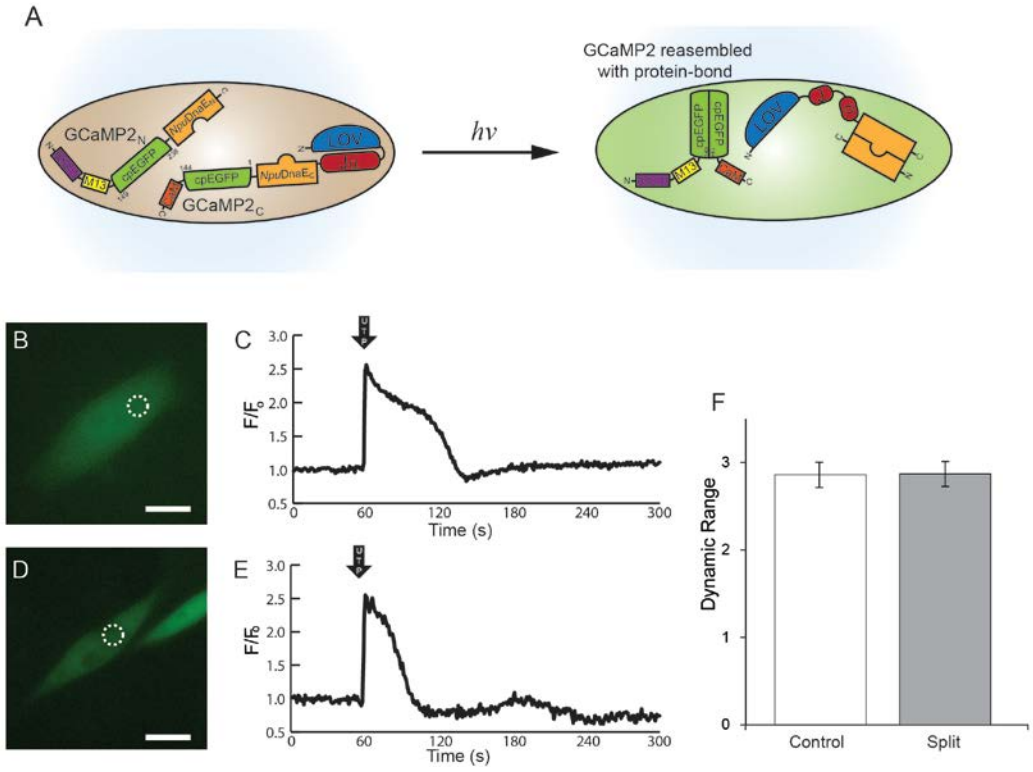

Figure 6.13 Photoactivatable reassembly of split GCaMP2.

(**A**) General schematic diagram depicting the reassembly and restoration of GCaMP2 function after blue-light photo-stimulation. (**B**) GFP fluorescence image of HeLa cells co-expressing the two precursors and photo-stimulated with blue-light. The faint fluorescence is emanating from the reassembled GCaMP2 biosensor in a low Ca^{2+} environment. (**C**) UTP-induced Ca^{2+} transient profile as measured by the reassembled GCaMP2. (**D**) YFP fluorescence image of HeLa cells expressing control GCaMP2 and its (**E**) measured UTP-induced Ca^{2+} transient profile. (**F**) Comparison of mean dynamic range of control and reassembled GCaMP2. White dotted circle represent regions where average fluorescence intensity measurements were taken. Error bars represent standard deviation. Cell images are false coloured. Scale bars represent 10 μm.

HeLa cells co-expressing the two precursors were grown overnight in the presence or absence of periodic blue-light photo-stimulation. Co-expressing cells that were photo-stimulated had reassembled the split GCaMP2 Ca^{2+} biosensor as detected by the weak fluorescence emitted in the presence of low cytosolic Ca^{2+} concentration (fluorescence was above auto-fluorescence of cells). The UTP-induced Ca^{2+} transients as measured by the reassembled split GCaMP2 Ca^{2+} biosensor was compared to non-altered GCaMP2 biosensors and had similar UTP-induced Ca^{2+} traces and statistically indistinguishable dynamic ranges (2.89 ± 0.18 and 2.85 ± 0.16, respectively; $p = 0.98$; $n > 6$) (Figure 6.13). As expected, co-expressing cells grown overnight in the absence of blue-light photo-stimulation did no yield any reassembled GCaMP2 biosensor (no cells exhibited weak fluorescence) and UTP-induced Ca^{2+} transients could not be measured (i.e. increases in cytosolic Ca^{2+} concentration did not yield any fluorescent cells which should be apparent if GCaMP2 Ca^{2+} biosensors had been reassembled). Cells expressing other precursors alone did not display any GCaMP2 activity or fluorescence regardless of photo-stimulations.

6.2.9 Blue-Light Mediates PTS Activity with Spatiotemporal Control

Blue-light inducible PTS activity can reassemble engineered split proteins in a light-dependent manner *in vitro*. Using the $RhoA_N$-*Npu*$DnaE_N$-mRFP and LOVInC-$RhoA_C$-Venus precursors as an example, PTS activity was restricted to an illuminated area as defined by a light beam. A narrow beam of blue-light emanating from the objective lens (100X) was used to periodically illuminate (i.e. 1 second of blue-light for every 30 seconds interval of no photo-stimulation) a field of HeLa cells co-expressing the two precursors. The narrow beam is small enough to only illuminate a centrally located region of cells in a much larger field of view (20X). Approximately 150 minutes after the start of illumination, cells situated within the illuminated region exhibited continuous dynamic blebbing while cells located outside the region remained unchanged (Figure 6.14). This exemplifies the potential of LOVInC to mediate the reassembly of non-functional split proteins in regions of interest of defined groups of cells.

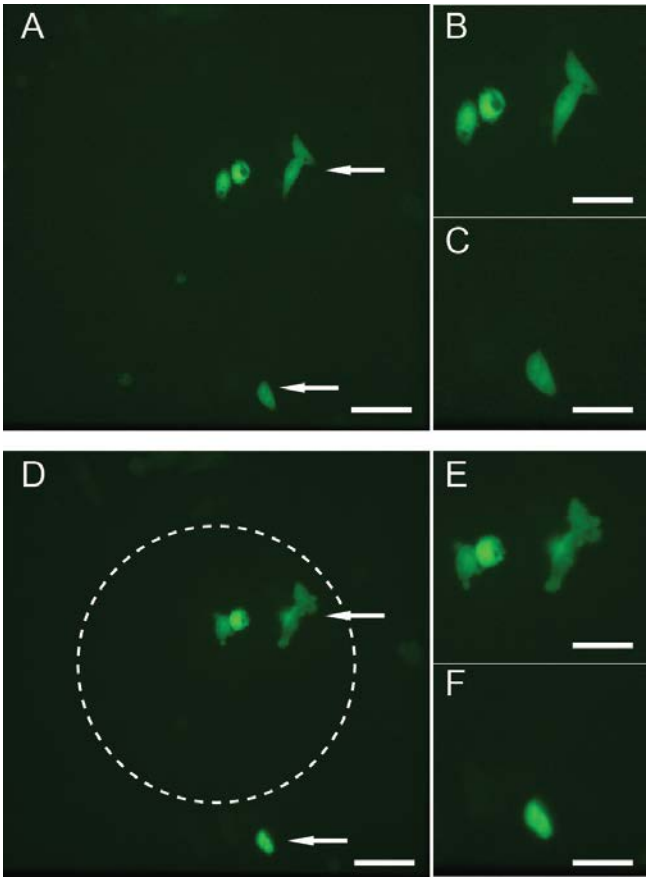

Figure 6.14 Photoactivatable intein has spatial precision

(**A**) HeLa cells co-expressing the $RhoA_N$-*Npu*$DnaE_N$-mRFP and LOVInC-$RhoA_C$-Venus and viewed under low magnification (20x objective). Two groups of cells have been identified by white arrows and shown enlarged in panels (**B**) and (**C**). Cells located near the center of the field of view were photo-stimulated with periodic intervals of blue-light (1 second every 30 seconds). (**D**) Low magnification of the same set of cells after 150 minutes showing cells within the illuminated zone (dotted white circle) have undergone dynamic blebbing while cells outside the illumination zone generally remained unchanged. Again, the two groups of cells are shown enlarged in panels (**E**) and (**F**), respectively. Cell images are false coloured. Scale bars for **A** and **B** are 10 µm and 30 µm for **B**, **C**, **E**, and **F**.

6.3 Discussion

Here, the LOVInC engineered protein represents a first step towards developing genetically encoded regulators of PTS activity. Here, the LOV2 domain was used as a transducer that transmitted an external light stimulus to induce conformation changes to act directly on the PTS activity. This shows that PTS activity can be suppress by allosteric interactions or steric hindrances and may be applied in a similar manner to other homologous naturally split inteins such as Ssp [161-163]. The advantages of such a system are that it does not require exogenous co-factors or energy and does not involve the use of synthetic protein synthesis methods; simple genetic expression is all that is required. The use of light as an input signal also provides exquisite control in both time and space. The LOVInC, the result of combining the LOV2 domain and *Npu*DnaE$_C$ intein half together, was engineered to act as a single modular unit so that it can be simply engineered into proteins of interest.

*Npu*DnaE intein halves are known to bind with high affinity and fast kinetics however, little is known about what drives the association of the inteins [50, 51]. It has been postulated that the intein halves when unbound are in an unfolded and disordered form and undergo a large conformational transition upon association to restore its three dimensional structure [54]. What drives this association has been linked to intermolecular electrostatic interactions between the split intein halves [53, 127]. In addition, the intein halves appear to interact only with each other and do not exhibit non-specific binding with endogenous proteins as shown here and other studies [161, 164]. This is beneficial as the inteins themselves have no effect on endogenous protein function. The C-terminal intein half, *Npu*DnaE$_C$, is comprised of four short β-strands attached together with short linkers. In particular, the first β-strand from the amino-terminal end is the longest and contains highly localized conserved basic amino acids that are presumably involved with intein association. In this study, the removal of 5 or less amino acids, some of which are the conserved amino acids, did not appear to suppress PTS activity significantly. On the other hand, the removal of 6 or more amino acids dramatically decreased PTS activity when tested in mammalian cells. Thus, it appears that the basic amino acids situated in the latter half of the β-strand may be more important to intein association than the first few amino acids, although additional, more detailed studies may be required to tease out the relationships.

The fusion of LOV2 domains to truncated *Npu*DnaE$_C$ appeared to suppress PTS activity by interfering with intein association as a result of allosteric interactions or steric hindrances. The LOV2 domain seems to have a greater effect on *Npu*DnaE$_C$ when 5 amino acids are removed than when 4 amino acids are removed from the β-strand. Again, this may point to the fact that centrally located basic amino acids are more crucial to intein association. A surprising find was that the LOV2$_{C450M}$, a mutant alleged to be light insensitive, was able to induce PTS activity when photo-stimulated. The covalent adduct is formed when the sulphur of the conserved Cys residue is covalently attached to the carbon 4a of the FMN when stimulated by light [165-167]. This initiates the conformational change that unbinds and unwinds the helical Jα peptide. Cysteine is one of the two amino acids that contain sulphur – the other being methionine. When the conserved Cys residue is replaced with Met, the sulphur group in Met may form the covalent adduct, albeit at a significantly reduced rate. Similar to how the attachment of different extein sequences affect PTS efficiency [168-171], the efficacy of LOV2 in affecting proteins may be dependent on the type or manner the proteins are attached [172]. In addition, given high propensity for intein halves to associate, slight exposure of amino acids may trigger association and PTS activity. All these contributing factors may explain, at least in part, the photo-inducible nature of LOV2$_{C450M}$ on t4*Npu*DnaE$_C$.

The engineered photoactivatable LOVInC was used to reassemble several representative proteins in an effort to display its generality with respect to foreign extein sequences. Due to the modularity of LOVInC, it was simply fused to the split protein fragments using simple genetic manipulation methods as described (refer to Chapter 3). The first protein, split Venus, represents the reassembly of exogenous proteins in a physiological context. The ability to reassemble proteins foreign to the cell is important as many tools are comprised of protein parts foreign to the cell. This can allow for better control and interrogation of biological systems. Next, split RhoA and split Casp7 were reassembled when photo-stimulated with blue-light. These two examples represent proteins that directly affect cell function; causing dynamic blebbing or committing cells to obtain cell death morphologies. Proteins of such nature could be used to endow cells with specific functions or in therapeutic applications such as gene therapies [173]. Lastly, the split GCaMP2 Ca^{2+} biosensor was demonstrated to be reassembled after photo-stimulation, a further improvement over existing strategies by allowing biosensors to be

expressed in specific locations and time. LOVInC was also shown to have spatial and temporal control as only cells photo-stimulated underwent PTS activity to reassemble a split protein. Here, it was found that peak PTS occurred at approximately the 120 minutes mark. However, because PTS is a continuous process and PTS products progressively increases with time, depending on the split proteins being reassembled, it is possible that the effects of the split proteins may occur a lot sooner than the peak accumulation time that was determined. Taken together, LOVInC may be used as a novel platform from which protein engineering tools and therapeutic interventions can be developed on.

6.4 Conclusion

In this chapter, a photoactivatable *Npu*DnaE intein was created by fusing a LOV2 domain to the truncated C-terminal intein to create LOVInC. Repeated photo-stimulation by blue-light induced PTS activity that was able to reassemble a variety of split proteins. In addition, the photo-inducible nature of LOVInC allows it to be controlled in both time and space, a property not afforded by other means of stimulus.

1. The removal of 6 or more amino acids from the first β-strand of *Npu*DnaE$_C$ intein substantially reduced PTS activity while the removal of 5 or less animo acids did not appear to affect PTS activity. Besides the systematic truncations, *Npu*DnaE$_C$ was not altered in any way.

2. The fusion of LOV2 to the amino-terminus end of truncated *Npu*DnaE$_C$ was able to suppress PTS activity. Mutants introduced into LOV2 (C450M) improved the range of PTS activity between the dark- and lit-state.

3. The fusion of LOV2$_{C450M}$ to t4*Npu*DnaE$_C$ (LOVInC) exhibited highest range with a four-fold increase in PTS activity from the dark-state to the lit-state. Substantial PTS activity occurred at approximately 120 minutes after photo-stimulation when assayed by observing for distinct co-localizations (see Chapter 4).

4. LOVInC was used to reassemble four different split proteins. The modular nature of LOVInC allowed quick creation of each of the precursors and because the full precursors were genetically encoded, they were easily and directly expressed in cells by transfection.

Referring to the research objectives as outlines in Chapter 1, objective 3 was addressed in this chapter. A photoactivatable *Npu*DnaE intein was engineered by combining LOV2 domain and a truncated variant of *Npu*DnaE$_C$. Light was selected as the input signal because it offers exquisite spatiotemporal control not available with methods such as chemical induction. The photoactivatable *Npu*DnaE intein is wholly defined by its genetic sequence, and thus can be manipulated with common genetic methodologies and easily transfected into cells or animal models. The study presented here also represents the first attempt at modulating PTS activity of naturally split inteins using an effector protein rather than by chemical means.

7 Summary and Conclusion

There are currently only a handful of studies on the naturally occurring split *Npu*DnaE intein. Much of the current studies have focused on elucidating the splicing mechanisms or used *Npu*DnaE intein to generate site-specific modifications on proteins. The reassembly of split proteins is a natural application of *Npu*DnaE intein. The studies presented here demonstrate unique ways *Npu*DnaE intein can be used to reassemble split proteins. It achieves this in a logical, progressive manner presented in three chapters, culminating in the development of a control mechanism for modulating PTS activity. In addition, the three main objectives set out at the beginning of this study were addressed in the three chapters. Some key determinants and potential structural explanations will be provided with regards to *Npu*DnaE intein PTS. The findings in this study can also serve as a platform for other homologous naturally split inteins.

Objective 1 and Chapter 4 - Reassembling artificially split Ca^{2+}-sensitive biosensors: A simple *in vitro* visual assay was developed in this chapter to test for PTS activity by examining distinct subcellular localizations of different fluorescent proteins. This assay, along with slight variations of it, was used throughout this study to verify the presence of PTS activity. The reassembling property of *Npu*DnaE intein was validated in two representative GECIs: TN-XL and GCaMP2. The performance of the reassembled TN-XL was considerably reduced which may be attributable to the residual peptide left behind after PTS activity. For this and other reasons, TN-XL was impractical for further use and was not optimized or developed further. On the other hand, the functionality of the reassembled GCaMP2 was maintained. It was able to measure induced Ca^{2+} transients in HeLa, COS7, and CHO similar to unaltered GCaMP2.

Objective 2 and Chapter 5 – Simultaneous reassembly of two split proteins: In addition to the PTS activity in reassembling split proteins, the couple interaction of the inteins was also shown to be able to non-covalently bring two fragments of a protein together to restore protein function. The coupling interactions of the two intein halves associate in one orientation, thus two similar, but different function proteins can be formed. The proteins reassembled though PTS activity and the protein restored through non-covalent interactions appear to be in stoichiometric ratios. When the non-covalent intein interactions are used to restore fluorescence to a fluorescent

protein, the fluorescence can potentially be used as a measure of the extent of reassembly of a non-fluorescent protein.

Objective 3 and Chapter 6 – Light as an exogenous input signal to modulate PTS activity: The PTS activity of *Npu*DnaE intein was modulated with the use of a LOV2 domain. The LOV2 domain acted as an allosteric switch to control PTS activity in response to light. A minimal functional unit of *Npu*DnaE$_C$ was determined and found to maintain high PTS activity. For LOV2 effectively suppress PTS activity, a truncated variant of the C-terminal half, t4*Npu*DnaE$_C$, was used. The use of the LOV2$_{C450M}$ mutant (LOVInC) improved the range of PTS activity between the dark- and lit-state by four folds. LOVInC was also demonstrated to induce PTS activity in a spatiotemporal manner. The photoactivatable LOVInC used to reassemble a selection of split proteins to show it generality with respect to various extein sequences.

7.1 Key Determinants of PTS by NpuDnaE intein

Chapter 4 and Chapter 5: *Npu*DnaE intein spontaneously undergoes PTS in the presence of the two precursors. In our first study, we exploited this intrinsic intein PTS property to reassemble GECIs. During the course of this study, we developed a PTS assay that had a fluorescent protein attached to the non-splicing end of either of the intein precursors, a property not found in the intein's natural state. Remarkably, what we found was that both ends of each precursor (the splicing and non-splicing ends) can be fused to proteins without markedly reduced PTS activity. Utilizing these findings, we proposed in our second study that in addition to the reassembly of a split protein with a peptide bond by PTS, there is a possibility of forming a second target protein through non-covalent interactions at the non-splicing ends. This is similar to previous strategies such as bimolecular fluorescence complementation using leucine zippers or small chemical molecules such as rapamycin and FKBP12 and FRB. The low dissociation constant (3nM) was enough to drive complementation of the split target protein (in addition to the split proteins at the splicing ends) and maintain association for the target protein to refold and mature (in the case of GFPs). Potential structural explanations are as follows:

1. Although the intein precursors are not found with proteins or protein fragments attached to its non-splicing end, it is not surprising that this is possible, as naturally split inteins most likely evolved from cis splicing inteins with the endonucleases removed during evolution. The associated intein has a compact globular structure without any highly disordered regions. In addition, the split site is situated within a long 14 amino acid flexible linker that is surface exposed and located at the end of one of the horse-shoe fold, away from the catalytic core. These factors together may explain the accommodation of fused proteins or protein fragments at the non-splicing ends without substantial effect on PTS activity.

2. The intein fragments exists as soluble, unfolded free form before intein association and refolding. Although it is known that an N-terminally situated well folded protein can promote folding and enhanced expressing of downstream proteins, a property we exploited in one of our studies for improving on-plate fluorescence screening (Section 3.1.5 and Appendix B), the solubility of the of the unassociated intein precursors may not be strong enough to promote folding of downstream proteins. Thus, target proteins to be split should be chosen such that the split fragments form well defined modular structures such as caspases or fragments are known to be soluble such as GFPs.

3. For efficient PTS to occur, specific amino acids are required to flank the splice junctions, particularly the C-terminus splice junction. Although the exact mechanism remains to be elucidated, it reasonable to postulate the flanking amino acids are involved in the splicing process by contributing hydrogen bonding or van der waals forces to orientate the sessile bonds in optimal positions conductive to protein splicing. In addition, Cys is particularly important as it is involved in thioester formation and trans-thio-esterification process.

4. Interestingly, although we did not directly study it, our studies have alluded to the nature that the excised found intein remains are bound for long durations, enough for fluorescent protein maturation visualization.

Chapter 6: In our third study, we looked at engineering a split intein that can be conditionally controlled by light. This is interesting as *Npu*DnaE intein PTS is a spontaneously process in the presence of both precursors. Although attempts have been made to control naturally split inteins though chemical means, this study is, to the best of our knowledge, the first attempt of controlling naturally split intein PTS activity using genetically encoded methods. The light sensing molecule we used was LOV2 from phototropin 1 of Oat. Truncations of 6 or more amino acids from the C-terminal intein fragment significantly reduced PTS activity presumably due to the decrease in intein association. This is because most of the basic amino acids involved in intein association are removed.

8 Future Directions

Future directions for the technologies developed in the previous chapters will be presented in this chapter. These include the improvement of intein splicing, some *in vivo* applications, how they can be applied to help elucidate biological networks, and finally some potential challenges that may appear.

8.1 Further Development and Improvement of *Npu*DnaE Inteins

A major limitation to the broad application of *Npu*DnaE PTS to all proteins is the requirement of specific amino acids at the N- and C-terminal splice junctions. These amino acids, collective termed the residual sequence in this study, get incorporated in the splice product after PTS and result in the formation of mutant proteins of interest which may adversely affect the proper functioning of the protein [31, 50, 51, 174]. Although the N-terminal splice junction is more tolerant of non-canonical sequences, the C-terminal splice junction requires the addition of the tripeptide sequence 'CFN' for optimal PTS efficiency [19, 31, 50, 174, 175]. The development an *Npu*DnaE intein that splices in a traceless manner would greatly broaden the application of the engineering tools established in this study.

One method to overcome this restriction is to generate mutant split inteins that splice at non-native splice junctions. Here, the canonical 'CFN' sequence is replaced with another set of desired sequence and mutant split inteins are recovered that will splice at the desired sequence. To efficiently obtain the appropriate mutant split intein, protein microarrays can be used to screen potential mutants [176-178]. Large protein libraries of mutant split inteins can be generated by random mutagenesis such as error-prone PCR [179-181]. One half of the split mutant intein can be immobilized on a solid support while the other half can be added and incubated. By attaching fluorescent probes on the free half, favourable split intein mutants can be isolated screening for fluorescence that has been attached to the solid support by PTS activity. Another method of obtaining mutant split inteins that splice at non-native splice junctions is through directed evolution [182, 183]. A selection method using various bacterial resistance genes can be used to rapidly evolve split intein function with desired properties [174].

The reassembly of multiple split proteins within a host cell can be accomplished with the use of two or more separate inteins. Similarly, multiple split inteins can be used to ligate three or more fragments together. This technique has found important applications in generating isotopically labeled proteins for NMR studies where specific segments of large complex proteins are labelled [17, 184]. In addition, the limitation of creating synthetic peptides with methods such as solid-phase peptide synthesis (often limited to peptides of ~50-70 amino acids) can be overcome by the intein-mediated tandem ligation of three of more fragments [14, 185-187]. The success and limitation of these scenarios will depend on the specificity with which a pair of split intein halves interacts with each other in the presence of other intein halves from other inteins. Non-specific interactions between other inteins could result in undesirable ligated products.

The highly efficient and specific association of naturally split intein halves have been speculated to be the result of intermolecular ionic interactions [54, 127]. Charges are segregated on specific residues of the intein halves. For example, acidic residues are concentrated at specific regions of the N-intein half while basic residues are concentrated on complementary regions on the C-intein half of *Npu*DnaE intein [53]. Despite this, cross-splicing between inteins do occur, particular with homologous inteins (i.e. *Npu*DnaE and *Ssp*DnaE inteins) [50]. One method would be to swap the charges on the inteins. For example, the basic residues on the N-terminal intein half would be swapped with acidic residues and vice versa for the C-terminal intein half [53]. Having alternating charges on each intein halves could be another way of increasing the specificity of intein association. When combined with the techniques of generating photo-activated PTS, the activity of multiple proteins can be post-translationally restored with precise spatiotemporal control. This can be used to study complex signalling networks involving multiple proteins or domains and decipher the link between the spatial dynamic effects of protein function to cellular function.

8.2 Potential *In Vivo* Applications

The ability to target bio-molecules (either native or engineered proteins) to specific tissue types is paramount in the study and understanding of cellular and neural systems. Currently, one of the

best methods of achieving this remains the use of cell-specific promoters to target expression of desired genes. In addition, promoters that are specific to malignant cells may offer therapeutic interventions such as specific gene therapies or help in the diagnosis of cancer [188]. While these promoters provide excellent targeting of genes, it remains a challenge to target defined population of cells with greater specificity. For one, these promoters have much lower expression levels than other more commonly used promoters therefore, limiting its applicability [189]. In addition, the specificities of these promoters are often not sufficiently strict. For example, ion targeting particular cancer cells, it was found that the human telomerase reverse transcriptase (hTERT) promoter was strongly active in tumour cells but generally repressed in normal cells [190-193]. However upon further investigation, it was determined that hTERT promoter activity was also detected in hematopoietic stem cells, the lymphohematopoietic system, intestinal cells, and epithelial liver cells [194-196]. These non-specific hTERT promoter activities may cause unknown side effects in these cells and tissues rendering this promoter unsuitable for gene targeted therapeutics. Thus, these wide ranging applications will benefit significantly in one form or another if engineering proteins can be further restricted to particular cells.

One method of overcoming the limitations of using single promoters for expression is to use two orthogonal tissue-specific promoters (Figure 8.1). The incorporation of two orthogonal tissue-specific promoters could improve cell targeting precision since only cells under the activation of both promoters will express the desired protein. The methods developed here utlizing *Npu*DnaE inteins are particularly conductive to this approach. For one, the intein precursors are already separated into two fragments and thus only require the addition of specific promoters for increased tissue targetability. Secondly, proteins are divided into two non-functional fragments, so that the non-specific expression of either precursor will not form functional proteins in cells/tissues that are not targeted. The use of the photoactivatable LOVInC can also add another element of control. For example, one can easily imagine using proteins that induce apoptosis, such as the caspases, to target cancer cells and control cell death [111]. The caspases will be divided into two fragments and attached to the complementary intein halves, *Npu*DnaE$_N$ or LOVInC. The expression of the two precursors under two orthogonal promoters will only target the specific cancer cell type; however cell death would not be induced yet. Only in cells that are

stimulated by light will apoptosis be induced. Thus, the spatiotemporal precision of light can be used as a further control mechanism in addition to specific targeting of cell types. This strategy can be easily extended to the reassembly of two proteins (see Chapter 5), where the GCaMP2 Ca^{2+} biosensor and a reference fluorescence is targeted to particular groups of cells. For example, it can be used in studying neural circuitry of live animals such as *C. Elegans* by targeting a subset of neural cells and measuring the Ca^{2+} dynamics and activity of the cells during different neural events [66, 197-199].

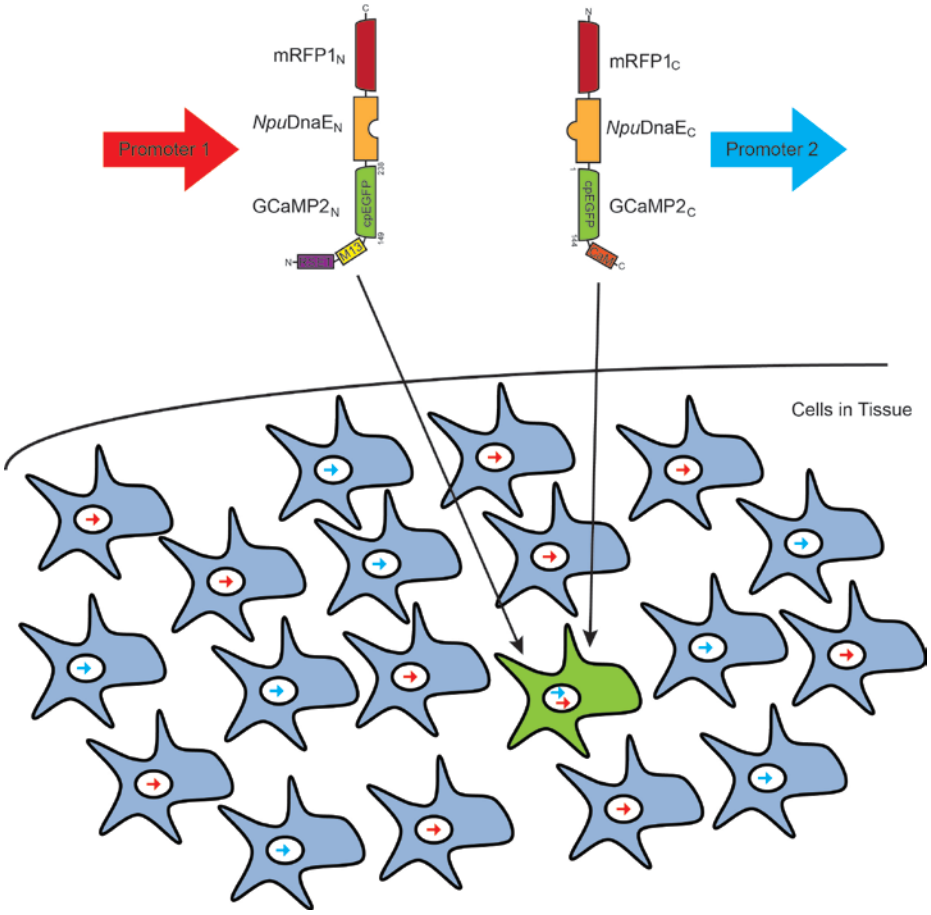

Figure 8.1 Future outlook: Two orthogonal promoters for increased targetability.

A hypothetical group of tissue cells of two kinds: cells that are under the control of one promoter (shown as red arrows) and another group of cells under the control of another promoter (shown as blue arrows). Cells that are only under the control of one of the promoters will only express one-half of the precursors, and thus will not reassemble the split protein (i.e. split GCaMP2 and split mRFP in this scenario). On the other hand, cells under the control of both promoters will express both precursors and allow PTS activity to reassemble and restore function to both GCaMP2 and mRFP (cell shown in green).

8.3 Studying Biological Networks

The cellular translocation of various proteins is an important process in the normal functioning of complex cellular systems whereby the cell transports specific proteins to particular regions or areas of the cell. These targets include the inner and outer membranes of different organelles and secretory vesicles. Translocation of proteins also plays a role in the complex signalling pathways that cells use to for internal and external communications. In particular, proteins involved in the signalling transduction pathway are often translocated to specific regions of the cell to where they colocalized with other signalling proteins to selective activate downstream functions [200, 201]. In order to study these various signalling pathways and understand the effects of the transported proteins, a method that reversibly moves proteins from one area of the cell to another will greatly facilitate elucidation of these events.

Although the PTS process itself is not a reversible process, the binding of the intein halves could potentially be exploited to create a reversible binding pair inducible by an external stimulus. By inactivating the PTS process without altering the binding regions, a high affinity binding pair can be isolated. If the binding can be modulated in some way, for example, by steric blockage of the binding sites, then the binding association of the pairs can theoretically be made reversible. LOVInC could be used as a starting point since it has been shown to have reduced binding affinity (and PTS activity) in the dark-state and it can be induced by a highly controllable light source. For LOVInC to act solely as a binding pair, the PTS activity must eliminated. This can be accomplished by the introduction of mutations near the splice junctions that inhibit PTS activity (see Chapter 5). In addition, LOVInC has the added advantage that the binding affinity and kinetics of LOVInC and *Npu*DnaE$_N$ can be tailored for specific applications by playing with the charged amino acids on the intein halves such as the removal of certain charged residues and replacing them with neutral amino acids to reduce the binding affinity. The translocation of LOVInCts (LOVInC trans-locating) between two regions can be accomplished by localizing the *Npu*DnaE$_N$ intein half with a targeting tag, such as Lyn for PM localization [130]. When LOVInCts is photo-stimulated, this induces the binding between LOVInCts and *Npu*DnaE$_N$ translocating and co-localizing the LOVInCts to, for example, the PM.

This approach can be used to study a number of biological processes. For example, it can be used to study the cell migration patterns of moving cells. This is important because a number of cellular and pathological problems depend on specific cell migration patterns such as wound healing and malignancy of cancer cells. One family of proteins that play key roles in cell migration by regulating the actin skeleton is the Rho family of GTPases which include the Rac1, RhoA, and Cdc42 members [103, 202, 203]. Recent studies have shown that translocation of these GTPases to the PM can trigger migration patterns such as the formation of lamellipodia and filopodia [130]. The above described approach can be used as an on and off switch to translocate these GTPases to the PM to induce specific migration patterns. For example, this can be accomplished by initially attaching *Npu*DnaE$_N$ to the PM with Lyn. LOVInCts can then be attached to Rho GTPases and expressed in the cytoplasm. Photo-stimulation of the cells would trigger translocation of the Rho GTPases to the membrane and induce general cell migration patterns. For more localized cell migration patterns, LOVInCts could be localized to the PM and *Npu*DnaE$_N$ fused to Rho GTPases and expressed in the cytoplasm. Photo-stimulation of localized regions of the PM would trigger only those proteins to co-localize with *Npu*DnaE$_N$. These induced migration patterns can be easily discontinued by removing the photo-stimulations. Because of the reversibility of LOVInCts, repeated illumination could trigger cells to migrate in a certain direction and allow careful studies of cell migrations under various scenarios.

The reversibility of LOVInCts can also be used to study Ca^{2+} signalling events. Ca^{2+} an ubiquitous second messenger signalling molecule that regulates many diverse cellular events such as cell growth, cell proliferation, apoptosis, cell mobility, and gene transcription[85-89, 204]. Ca^{2+} can enter the cytoplasm through membrane ion channels or through the release of Ca^{2+} stores. One important internal source of Ca^{2+} is from the ER and a series of events induces the release [205, 206]. As mentioned in Chapter 2, an external stimulus is sensed by a membrane bound G protein-coupled receptors [99]. This triggers the G_q subunit to activate phospholipase-Cβ (PLCβ) which cleaves PIP2 to generate IP3. IP3 then activates IP3-receptors and triggers Ca^{2+} release from the ER stores. LOVInCts can be used to reversibly inhibit ER Ca^{2+} release by blocking one of the events in the pathway. For example, *Npu*DnaE$_N$ could be attached to the G protein-coupled receptor such that *Npu*DnaE$_N$ itself would not hinder the functioning of the receptor, but would allow steric blockage of the active site when other proteins are attached

while LOVInCts could be fused to an inhibitor or a large globular domain such as GFP [207]. Photo-stimulation would trigger the translocation of LOVInCts fused to an inhibitor to the receptor, suppressing its function and the subsequent Ca^{2+} release. Removal of the photo-stimulation would release the inhibition and allow Ca^{2+} to be released from the ER. This exquisite control of Ca^{2+} release can be used to simulate oscillatory Ca^{2+} transients or other Ca^{2+} events to better understand how Ca^{2+} is controlled and used for cellular function.

8.4 Potential Challenges

There are several challenges involved in the application of the *Npu*DnaE intein strategies presented in this thesis. Two are briefly described below.

Effect of extein sequences on PTS activity: An on-going challenge is the effect of foreign extein sequences on the PTS activity of the inteins [168]. Larger and more complex proteins could greatly reduce the PTS activity by interfering with intein association or changing the protein structure if not properly designed. The split sites of the foreign proteins must be choosen carefully as well. Ideally, split sites should be within turns or linkers and far from catalytic sites. In addition, split sites should be chosen so that the two fragments fold in well-defined structures. This will minimize exposure of hydrophobic residues which could lead to protein aggregations.

Reactivity of precursors to endogenous proteins: Although the inteins themselves have minimal reactivity with endogenous proteins, care must be taken in choosing the foreign exteins so as to minimize the non-specific interactions with endogenous proteins. Factors that will minimize this are outlined above.

Effects of charge swapping for increased specificity: When optimization the specificity of the intein halves by charge swapping the acidic and basic residues, care must be taken to only change residues that are only involved with the intermolecular ionic interactions and that the residues are isolated from intramolecular ionic interactions.

References

1. Shekhawat, S.S. and I. Ghosh, *Split-protein systems: beyond binary protein-protien interactions.* Current Opinion in Chemical Biology, 2011. **15**: p. 789-797.
2. Villalobos, V., S. Naik, and D. Piwnica-Worms, *Current state of imaging protein-protein interactions in vivo with genetically encoded reporters.* Annual Review of Biomedical Engineering, 2007. **9**: p. 321-349.
3. Kerppola, T.K., *Bimolecular fluorescence complementation (BiFC) analysis as a probe of protein interactions in living cell.* Annual Review of Biophysics, 2008. **37**: p. 465-487.
4. Tian, J., K. Ma, and I. Saaem, *Advancing high-throughput gene synthesis technology.* Molecular BioSystems, 2009. **5**: p. 714-722.
5. Czar, M.J., et al., *Gene synthesis demystified.* Trends in Biotechnology, 2008. **27**(2): p. 63-72.
6. Nandagopal, N. and M.B. Elowitz, *Synthetic biology: Integrated gene circuits.* Science, 2011. **333**: p. 1244-1248.
7. Purnick, P.E.M. and R. Weiss, *The second wave of synthetic biology: from modules to systems.* Nature Reviews Molecular Cell Biology, 2009. **10**: p. 410-422.
8. Negrin, R.S. and C.H. Contag, *In vivo imaging using bioluminescence: a tool for probing graft-versus-host disease.* Nature Reviews Immunology, 2006. **6**: p. 484-490.
9. Gross, S. and D. Piwnica-Worms, *Spying on cancer: Molecular imaging in vivo with genetically encoded reporters.* Cancer Cell, 2005. **7**: p. 5-15.
10. Hu, C.-D. and T.K. Kerppola, *Simultaneous visualization of multiple protein interactions in living cells using multicolor fluorescence complementation analysis.* Nature Biotechnology, 2003. **21**: p. 539-545.
11. Remy, I. and S.W. Michnick, *A highly sensitive protein-protein interaction assay based on Gaussia luciferase.* Nature Methods, 2006. **3**: p. 977-979.
12. Luker, K.E., et al., *Kinetics of regulated protein-protein interactions revealed with firefly luciferase complementation imaging in cells and living animals.* Proceeding of the National Academy of Science of the United States of America, 2004. **101**(33): p. 12288-12293.
13. Muir, T.W., D. Sondhi, and P.A. Cole, *Expressed protein ligation: A general method for protein engineering.* Proceeding of the National Academy of Science of the United States of America, 1998. **95**: p. 6705-6710.
14. Muir, T.W., *Semisynthesis of proteins by expressed protein ligation.* Annual Review of Biochemistry, 2003. **72**: p. 249-289.
15. Mootz, H.D., *Split inteins as versatile tools for protein semisynthesis.* ChemBioChem. **10**: p. 2579-2589.
16. Volkmann, G. and H. Iwai, *Protein trans-splicing and its use in structural biology: opportunities and limitations.* Molecular BioSystems, 2010. **6**: p. 2110-2121.
17. Otomo, T., et al., *Improved segmental isotope labeling of proteins and application to a larger protein.* Journal of Biomolecular NMR, 1999. **14**: p. 105-114.
18. Yamazaki, T., et al., *Segmental isotope labeling for protein NMR using peptide splicing.* Journal of American Chemical Society, 1998. **120**: p. 5591-5592.

19. Zuger, S. and H. Iwai, *Intein-based biosynthetic incorporation of unlabeled protein tags into isotopically labeled proteins for NMR studies.* Nature Biotechnology, 2005. **23**: p. 736-740.
20. Paulmurugan, R. and S.S. Gambhir, *Monitoring protein-protein interactions using split synthetic renilla luciferase protein-fragment-assisted complementation.* Analytical Chemistry, 2003. **75**(7): p. 1584-1589.
21. Massoud, T.F., R. Paulmurugan, and S.S. Gambhir, *A molecularly engineered split reporter for imaging protein-protein interactions with positron emission tomography.* Nature Medicine, 2010. **16**(8): p. 921-926.
22. Lindman, S., et al., *In vivo protein stabilization based on fragment complementation and a split GFP system.* Proceeding of the National Academy of Science of the United States of America, 2010. **PNAS Early Edition**: p. 1-6.
23. Porter, J.R., et al., *A general and rapid cell-free approach for the interrogation of protein-protein, protein-DNA, and protein-RNA interactions and their antagonists utilizing split-protein reporters.* Journal of American Chemical Society, 2008. **130**: p. 6488-6497.
24. Wolfe, S.A., L. Nekludova, and C.O. Pabo, *DNA recognition by Cys_2His_2 zinc finger proteins.* Annual Review in Biophysics and Biomolecular Structure, 1999. **3**: p. 183-212.
25. Choi, J., et al., *Structure of the FKBP12-rapamycin complex interacting with the binding domain of human FRAP.* Science, 1996. **273**(5272): p. 239-242.
26. Paulus, H., *Protein splicing and related forms of protein autoprocessing.* Annual Review of Biochemistry, 2000. **69**: p. 447-496.
27. Gogarten, J.P., et al., *Inteins: Structure, function, and evolution.* Annual Review in Microbiology, 2002. **56**: p. 263-287.
28. Perler, F.B., *Protein splicing mechanisms and applications.* IUBMB Life, 2005. **57**(7): p. 469-476.
29. Perler, F.B., *InBase: the intein database.* Nucleic Acids Research, 2002. **30**(1): p. 383-384.
30. Chevalier, B.S. and B.L. Stoddard, *Homing endonucleases: structural and functional insight into the catalysts of intron/intein mobility.* Nucleic Acids Research, 2001. **29**(18): p. 3757-3774.
31. Dassa, B., et al., *Trans protein splicing of cyanobacterial split inteins in endogenous and exogenous combinations.* Biochemistry, 2007. **46**: p. 322-330.
32. Gogarten, J.P. and E. Hilario, *Inteins, introns, and homing endonucleases: recent revelations about the life cycle of parasitic genetic elements.* BME Evolutionary Biology, 2006. **6**(94): p. 1-5.
33. Liu, X.-Q., *Protein-splicing inteins: Genetic mobility, origin, and evolution.* Annual Review of Genetics, 2000. **34**: p. 61-76.
34. Kanno, A., et al., *Cyclic luciferase for real-time sensing of caspase-3 activities in living mammals.* Angewandte Chemie International Edition, 2007. **46**: p. 7595-7599.
35. Ozawa, T., et al., *A genetic approach to identifying mitochondrial proteins.* Nature Biotechnology, 2003. **21**: p. 287-293.
36. Noren, C.J., J. Wang, and F.B. Perler, *Dissecting the chemistry of protein splicing and its application.* Angewandte Chemie International Edition, 2000. **39**: p. 450-466.
37. Vila-Perello, M. and T.W. Muir, *Biological applications of protein splicing.* Cell, 2010. **143**: p. 191-200.

38. Hackenberger, C.P.R. and D. Schwarzer, *Chemoselective ligation and modification strategies for peptides and proteins.* Angewandte Chemi International Edition, 2008. **47**: p. 10030-10074.
39. Xu, M.-Q., et al., *Protein splicing: an analysis of the branched intermediate and its resolution by succinimide formation.* THE EMBO Journal, 1994. **13**(23): p. 5517-5522.
40. Shao, Y., M.-Q. Xu, and H. Paulus, *Protein splicing: Characterization of the aminosuccinimide residue at the carboxyl terminus of the excised intervening sequence.* Biochemistry, 1995. **34**(34): p. 10844-10850.
41. Anraku, Y., R. Mizutani, and Y. Satow, *Protein splicing: Its discovery and structural insight into novel chemical mechanisms.* IUBMB Life, 2005. **57**(8): p. 563-574.
42. Klabunde, T., et al., *Crystal structure of GyrA intein from Mycobacterium xenopi reveals structural basis of protein splicing.* Nature Structural Biology, 1998. **5**(1): p. 31-36.
43. Hirata, R., et al., *Molecular structure of a gene, VMA1, encoding the catalytic subunit of H^+-translocating adenosine triphosphatase from vacuolar membranes of Saccharomyces cerevisiae.* The Journal of Biological Chemistry, 1990. **265**(12): p. 6726-6733.
44. Saleh, L. and F.B. Perler, *Protein splicing in cis and in trans.* The Chemical Record, 2006. **6**: p. 183-193.
45. Mills, K.V., et al., *Protein splicing in trans by purified N- and C-terminal fragments of the Mycobacterium tuberculosis RecA intein.* Proceeding of the National Academy of Science of the United States of America, 1998. **95**: p. 3543-3548.
46. Southworth, M.W., et al., *Control of protein splicing by intein fragment reasembly.* THE EMBO Journal, 1998. **17**(4): p. 918-926.
47. Brenzel, S., T. Kurpiers, and H.D. Mootz, *Engineering artificially split inteins for applications in protein chemistry: Biochemical characterization of the split Ssp DnaB intein and comparison to the split Sce VMA intein.* Biochemistry, 2006. **45**: p. 1571-1578.
48. Mootz, H.D., et al., *Conditional protein splicing: A new tool to control protein structure and function in vitro and in vivo.* Journal of American Chemical Society, 2003. **125**(35): p. 10561-10569.
49. Gross, M., et al., *Reconstitution of active octameric mitochondrial creatine kinase from two genetically engineered fragments.* Protein Science, 1996. **5**: p. 320-330.
50. Iwai, H., et al., *Highly efficient protein trans-splicing by a naturally split DnaE intein from Nostoc punctiform.* FEBS Letters, 2006. **580**(7): p. 1853-1858.
51. Zettler, J., V. Schutz, and H.D. Mootz, *The naturally split Npu DnaE intein exhibits an extraordinary high rate in the protein trans-splicing reaction.* FEBS Letters, 2009. **583**(5): p. 909-914.
52. Oeemig, J.S., et al., *Solution structure of DnaE intein from Nostoc punctiforme: Structural basis for the design of a new split intein suitable for site-specific chemical modification.* FEBS Letters, 2009. **538**(9): p. 1451-1456.
53. Shah, N.H., M. Vila-Perello, and T.W. Muir, *Kinetic control of one-pot trans-splicing reactions by using a wild-type and designed split intein.* Angewandte Chemi International Edition, 2011. **50**: p. 6511-6515.
54. Zheng, Y., et al., *Mutual synergistic protein folding in split intein.* Bioscience Reports, 2012. **32**: p. 433-442.
55. Aranko, A.S., et al., *In vivo and in vitro protein ligation by naturally occurring and engineered split DnaE inteins.* PLoS ONE, 2009. **4**(4): p. e5185.

56. Zeidler, M.P., et al., *Temperature-sensitive control of protein activity by conditionally splicing inteins.* Nature Biotechnology, 2004. **22**(7): p. 871-876.
57. Adam, E. and F.B. Perler, *Development of a positive genetic selection system for inhibition of protein splicing using mycobacterial inteins in Escherichia coli DNA gyrase subunit A.* Journal of Microbiology and Biotechnology, 2002. **4**: p. 479-487.
58. Wood, D.W., et al., *A genetic system yields self-cleaving inteins for bioseparations.* Nature Biotechnology, 1999. **17**: p. 889-892.
59. Buskirk, A.R., et al., *Directed evolution of ligand dependence: Small-molecule-activated protein splicing.* Proceeding of the National Academy of Science of the United States of America, 2004. **101**(29): p. 10505-10510.
60. Skretas, G. and D.W. Wood, *Regulation of protein activity with small-molecule-controlled inteins.* Protein Science, 2004. **14**: p. 523-532.
61. Vila-Perello, M., et al., *Activation of protein splicing by protease- or light-triggered O to N Acyl Migration.* Angewandte Chemi International Edition, 2008. **47**: p. 7764-7767.
62. Berrade, L., Y. Kwon, and J.A. Camarero, *Photomodulation of protein trans-splicing through backbone photocaging of the DnaE split intein.* ChemBioChem, 2010. **11**: p. 1368-1372.
63. Binschik, J., J. Zettler, and H.D. Mootz, *Photocontrol of protein activity mediated by the cleavage reaction of a split intein.* Angewandte Chemi International Edition, 2011. **50**: p. 3249-3252.
64. Tyszkiewicz, A.B. and T.W. Muir, *Activation of protein splicing with light in yeast.* Nature Methods, 2008. **5**(4): p. 303-305.
65. Fenno, L., O. Yizhar, and K. Deisseroth, *The development and application of optogenetics.* Annual Review of Neuroscience, 2011. **34**: p. 389-412.
66. Yizhar, O., et al., *Optogenetics in neural systems.* Neuron, 2011. **71**(9-34).
67. Deisseroth, K., *Optogenetics.* Nature Methods, 2011. **8**(1): p. 26-29.
68. Zhang, F., et al., *Circuit-breakers: optical technologies for probing neural signals and systems.* Nature Reviews Neuroscience, 2007. **8**: p. 577-581.
69. Deisseroth, K., et al., *Next-generation optical technologies for illuminating genetically targeted brain circuits.* The Journal of Neuroscience, 2006. **26**(41): p. 10380-10386.
70. Bae, G. and G. Choi, *Decoding of light signals by plant phytochromes and their interacting proteins.* Annual Review of Platn Biology, 2008. **59**: p. 281-311.
71. Kato, H.E., et al., *Crystal structure of the channelrhodopsin light-gated cation channel.* Nature, 2012. **482**: p. 369-374.
72. Lin, J.Y., *A user's guide to channelrhodopsin variants: features, limitations and future development.* Experimental Physiology, 2010. **96**(1): p. 19-25.
73. Strickland, D., et al., *TULIPS: tunable, light-controlled interacting protein tags for cell biology.* Nature Methods, 2012. **9**(4): p. 379-384.
74. Wu, Y.I., et al., *A genetically encoded photoactivatable Rac controls the motility of living cells.* Nature, 2009. **461**(3): p. 104-108.
75. Strickland, D., K. Moffat, and T.R. Sosnick, *Light-activated DNA binding in a designed allosteric protein.* Proceeding of the National Academy of Science of the United States of America, 2007. **105**(31): p. 10709-10714.
76. Christie, J.M., *Phototropin blue-light receptors.* Annual Review of Plant Biology, 2007. **58**: p. 21-45.

77. Crosson, S., S. Rajagopal, and K. Moffat, *The LOV Domain Family: Photoresponsive signaling modules coupled to diverse output domains.* Biochemistry, 2002. **42**(1): p. 2-10.
78. Huala, E., et al., *Arabidopsis NPH1: A protein kinase with a putative redox-sensing domain.* Science, 1997. **278**(2120-2123).
79. Peter, E., B. Dick, and S.A. Baeurle, *Mechanism of signal transduction of the LOV2-Ja photosensor from Avena sativa.* Nature Communications, 2010. **1**: p. 122.
80. Harper, S.M., J.M. Christie, and K.H. Gardner, *Disruption of the LOV-Ja helix interaction activates phototropin kinase activity.* Biochemistry, 2004. **43**: p. 16184-16192.
81. Harper, S.M., L.C. Neil, and K.H. Gardner, *Structural basis of a phototopin light switch.* Science, 2003. **301**: p. 1541-1544.
82. Salomon, M., et al., *Photochemical and mutational analysis of the FMN-binding domains of the plant blue light receptor, phototropin.* Biochemistry, 2000. **39**: p. 9401-9410.
83. Christie, J.M., et al., *LOV (light, oxygen, or voltage) domains of the blue-light photoreceptor phototropin (nph1): Binding sites for the chromophore flavin mononucleotide.* Proceeding of the National Academy of Science of the United States of America, 1999. **96**(15): p. 8779-8783.
84. Pham, E., E. Mills, and K. Truong, *A synthetic photoactivated protein to generate local of global Ca^{2+} signals.* Chemistry & Biology, 2011. **18**: p. 880-890.
85. Feske, S., *Calcium signalling in lymphocyte activation and disease.* Nature Reviews Immunology, 2007. **7**: p. 690-702.
86. Clapham, D.E., *Calcium Signaling.* Cell, 2007. **131**: p. 1047-1058.
87. Berridge, M.J., M.D. Bootman, and H.L. Roderick, *Calcium signalling: Dyanimcs, homeostasis and remodelling.* Nature Reviews Molecular Cell Biology, 2003. **4**: p. 517-529.
88. Berridge, M.J., M.D. Bootman, and P. Lipp, *Calcium - a life and death signal.* Nature, 1998. **395**: p. 645-648.
89. Berridge, M.J., P. Lipp, and M.D. Bootman, *The versatility and universality of calcium signalling.* Nature Reviews Molecular Cell Biology, 2000. **1**: p. 11-21.
90. Miyawaki, A., et al., *Dynamic and quantitative Ca^{2+} measurements using improved cameleons.* Proceeding of the National Academy of Science of the United States of America, 1999. **96**: p. 2135-2140.
91. Miyawaki, A., et al., *Fluorescent indicators for Ca^{2+} based on green fluorescent proteins and calmodulin.* Nature, 1997. **388**: p. 882-887.
92. Truong, K., et al., *FRET-based in vivo Ca^{2+} calmodulin-GFP fusion molecule.* Nature Structural Biology, 2001. **8**: p. 1069-1073.
93. Mank, M., et al., *A FRET-based calcium biosensor with fast signal kinetics and high fluorescence change.* Biophysical Journal, 2006. **90**: p. 1790-1796.
94. Nagai, T., et al., *Circularly permuted green fluorescent proteins engineered to sense Ca^{2+}.* Proceeding of the National Academy of Science of the United States of America, 2001. **98**(6): p. 3197-3202.
95. Zhao, Y., et al., *AN expanded palette of genetically encoded Ca^{2+} indicators.* Science, 2011. **333**: p. 1888-1891.
96. Nakai, J., M. Ohkura, and K. Imoto, *A high signal-to-noise Ca^{2+} probe composed of a single green fluorescent protein.* Nature Biotechnology, 2001. **19**: p. 137-141.
97. North, R.A., *Molecular physiology of P2X receptors.* Physiology Review, 2002. **82**: p. 1013-1067.

98. Ralevic, V. and G. Burnstock, *Receptors for purines and pyrimidines.* Pharmacological Reviews, 1998. **50**(3): p. 413-492.
99. Abbracchio, M.P., et al., *Internation Union of Pharmacology LVIII: Undate on the P2Y G protein-coupled nucleotide receptors: From molecular mechanisms and pathophysiology to therapy.* Pharmacological Reviews, 2006. **58**(3): p. 281-341.
100. Zhou, Y., et al., *STIM1 gates the store-operated calcium channle ORAI1 in vitro.* Nature Structural & Molecular Biology, 2010. **17**(1): p. 112-116.
101. Varnai, P., L. Hunyady, and T. Balla, *STIM and Orai: the long-awaited constituents of store-operated calcium entry.* Trends in Pharmacological Science, 2009. **30**(3): p. 118-128.
102. Morgan, A.J. and R. Jacob, *Ionomycin enhances Ca^{2+} influx by stimulating store-regulated cation entry and not by a direct action at the plasma membrane.* Journal of Biochemistry, 1994. **300**: p. 665-672.
103. Heasman, S.J. and A.J. Ridley, *Mammalian Rho GTPases: new insights into their functions from in vivo studies.* Nature Reviews Molecular Cell Biology, 2008. **9**: p. 690-701.
104. Jaffe, A.B. and A. Hall, *Rho GTPases: Biochemistry and biology.* The Annual Review of Cell and Developmental Biology, 2005. **21**: p. 247-269.
105. Yeh, B.J., et al., *Rewiring cellular morphology pathways with synthetic guanine nucleotide exchange factors.* Nature, 2007. **447**: p. 596-600.
106. Pertz, O., et al., *Spatiotemporal dynamics of RhoA activity in migrating cells.* Nature, 2006. **440**: p. 1069-1072.
107. Mills, E., E. Pham, and K. Truong, *Structure based design of a Ca^{2+}-sensitive RhoA protein that controls cell morphology.* Cell Calcium, 2010. **48**: p. 195-201.
108. Fackler, O.T. and R. Grosse, *Cell motility through plasma membrane blebbing.* The Journal of Cell Biology, 2008. **181**(6): p. 879-884.
109. Tinevez, J.-Y., et al., *Role of cortical tension in bleb growth.* Proceeding of the National Academy of Science of the United States of America, 2009. **106**(44): p. 1-6.
110. Walsh, J.G., et al., *Executioner caspase-3 and capase-7 are functionally distinct proteases.* Proceeding of the National Academy of Science of the United States of America, 2008. **105**(35): p. 12815-12819.
111. Riedl, S.J. and Y. Shi, *Molecular mechanisms of caspase regulation during apoptosis.* Nature Reviews Molecular Cell Biology, 2004. **5**: p. 897-907.
112. Fan, T.-J., et al., *Caspase family proteases and apoptosis.* Acta Biochimica et Biophysica Sinica, 2005. **37**(11): p. 719-727.
113. Erener, S., et al., *Inflammasome-activated caspase 7 cleases PARP1 to enhance the expression of a subset of NF-kB target genes.* Molecular Cell, 2012. **46**: p. 200-211.
114. Zhou, Q. and G.S. Salvesen, *Activation of pro-caspase-7 by serine proteases includes a non-canonical specificity.* Journal of Biochemistry, 1997. **324**: p. 361-364.
115. Twiddy, D., et al., *Capase-7 is directly activated by the ~700-kDa apoptosome complex and is released as a stable XIAP-Caspase-7 ~200-kDa complex*. The Journal of Biological Chemistry, 2006. **281**(7): p. 3876-3888.
116. Saraste, A. and K. Pulkki, *Morphologic and biochemical hallmarks of apoptosis.* Cardiovascular Research, 2000. **45**: p. 528-537.
117. Hacker, G., *The morphology of apoptosis.* Cell Tissue Research, 2000. **301**: p. 5-17.

118. Wong, S.S. and K. Truong, *Fluorescent protein-based methods for on-plate screening of gene insertion.* Public Library of Science, 2010. **5**(12): p. 1-5.
119. Truong, K., A. Khorchid, and M. Ikura, *A fluorescent cassette-based strategy for engineering multiple domain fusion proteins.* BMC Biotechnology, 2003. **3**: p. 1-8.
120. Pedelacq, J.-D., et al., *Engineering and characterization of a superfolder green fluorescent protein.* Nature Biotechnology, 2006. **24**(1): p. 79-88.
121. Waldo, G.S., et al., *Rapid protein-folding assay using green fluorescent protein.* Nature Biotechnology, 1999. **17**: p. 691-695.
122. Makrides, S.C., *Strategies for achieving high-level expression of genes in Escherichia coli.* Microbiological Reviews, 1996. **60**(3): p. 512-538.
123. Cabantous, S., T.C. Terwilliger, and G.S. Waldo, *Protein tagging and detection with engineered self-assembling fragments of green fluorescent protein.* Nature Biotechnology, 2005. **23**(1): p. 102-107.
124. Zimmer, M., *Green fluorescent protein (GFP): Applications, structure, and related photophysical behavior.* Chemical Reviews, 2002. **102**: p. 759-781.
125. Tsien, R.Y., *The green fluorescent protein.* Annual Review of Biochemistry, 1998. **67**: p. 509-544.
126. Shine, J. and L. Dalgarno, *Determinant of cistron specificity in bacterial ribosomes.* Nature, 1975. **254**: p. 34-38.
127. Shi, J. and T.W. Muir, *Development of a tandem protein trans-splicing system based on native and engineered split inteins.* Journal of American Chemical Society, 2005. **127**(17): p. 6198-6206.
128. Bolte, S. and F.P. Cordelieres, *A guided tour into subcellular colocalization analysis in light microscopy.* Jounral of Microscopy, 2006. **224**(3): p. 213-232.
129. Wong, S.S., et al., *Split-intein mediated re-assembly of genetically encoded Ca^{2+} indicators.* Cell Calcium, 2012. **51**: p. 57-64.
130. Violin, J.D., et al., *A genetically encoded fluorescent reporter reveals oscillatory phosphorylation by protein kinase C.* The Journal of Cell Biology, 2003. **161**(5): p. 899-909.
131. Rizzo, M.A., et al., *An improved cyan fluorescent protein variant useful for FRET.* Nature Biotechnology, 2004. **22**(4): p. 445-449.
132. Campbell, R.E., et al., *A monomeric red fluorescent protein.* Proceeding of the National Academy of Science of the United States of America, 2002. **99**(12): p. 7877-7882.
133. Nagai, T., et al., *A variant of yellow fluorescent protein with fast and efficient maturation for cell-biological applications.* Nature Biotechnology, 2002. **20**: p. 87-90.
134. Chong, S., et al., *Single-column purification of free recombinant proteins using a self-cleavable affinity tag derived from a protein splicing element.* Gene, 1997. **192**: p. 271-281.
135. Chong, S., et al., *Protein splicing involving the Saccharomyces cerevisiae VMA intein.* The Journal of Biological Chemistry, 1996. **271**(36): p. 22159-22168.
136. Heim, N. and O. Griesbeck, *Genetically encoded indicators of cellular calcium dynamics based on troponin C and green fluorescent protein.* The Journal of Biological Chemistry, 2004. **279**(14): p. 14280-14286.
137. Vassylyev, D.G., et al., *Crystal structure of troponin C in complex with troponin I fragment at 2.3-A resolution.* Proceeding of the National Academy of Science of the United States of America, 1998. **95**: p. 4847-4852.

138. Mank, M. and O. Griesbeck, *Genetically encoded calcium indicators.* Chemical Reviews, 2008. **108**(5): p. 1550-1564.
139. Jares-Erijman, E.A. and T.M. Jovin, *FRET imaging.* Nature Biotechnology, 2003. **21**: p. 1387-1395.
140. Barski, O.A., et al., *Human aldehyde reductase promoter allows simltaneous expression of two genes in oppostie direcitons.* BioTechniques, 2004. **36**: p. 382-388.
141. Ozawa, T., et al., *A fluorescent indicator for detecting protein-protein interactions in vivo based on protein splicing.* Analytical Chemistry, 2000. **72**(21): p. 5151-5157.
142. Tallini, Y.N., et al., *Imaging cellular signals in the heart in vivo: Cardiac expression of the high-signal Ca^{2+} indicator GCaMP2.* Proceeding of the National Academy of Science of the United States of America, 2006. **103**(12): p. 4753-4758.
143. Lee, M.Y., et al., *Local subplasma membrane Ca^{2+} signals detected by a tethered Ca^{2+} sensor.* Proceeding of the National Academy of Science of the United States of America, 2006. **103**(35).
144. Wang, Q., et al., *Structural basis for calcium sensing by GCaMP2.* Structure, 2008. **16**: p. 1817-1827.
145. Shimozono, S., et al., *Slow Ca^{2+} dynamics in pharyngeal muscles in Caenorhabditis elegans during fast pumping.* European Molecular Biology Organization, 2004. **5**(5): p. 521-526.
146. Kerr, R., et al., *Optical imaging of calcium transients in neurons and pharyngeal muscle of C. elegans.* Neuron, 2000. **26**: p. 583-894.
147. Albertson, D.G. and J.N. Thomson, *The pharynx of Caenorhabditis elegans.* Philosophical Transaction of the Royal Society of London, 1975. **275**: p. 299-325.
148. Niacaris, T. and L. Avery, *Serotonin regulates repolarization of the C. elegans pharyngeal muscle.* The Journal of Experimental Biology, 2003. **206**: p. 223-231.
149. Horvitz, H.R., et al., *Serotonin and octopamine in the nematode Caenorhabditis elegans.* Science, 1982. **216**(4549): p. 1012-1014.
150. Avery, L. and H.R. Horvitz, *Pharyngeal pumping continues after laser killing of the pharyngeal nervous system of C. elegans.* Neuron, 1989. **3**: p. 473-485.
151. Raizen, d.M., R.Y.N. Lee, and L. Avery, *Interacting genes required for pharyngeal excitation by motor neuron MC in Caenorhabditis elegans.* Genetics, 1995. **141**: p. 1365-1382.
152. Starich, T.A., et al., *eat-5 and unc-7 represent a multigene family in Caenorhabditis elegans involved in cell-cell coupling.* The Journal of Cell Biology, 1996. **134**(2): p. 537-548.
153. Avery, L. and H.R. Horvitz, *Effects of starvation and neuroactive drugs on feeding in Caenorhabditis elegans.* The Journal of Experimental Zoology, 1990. **253**: p. 263-270.
154. Dhar, T. and H.D. Mootz, *Modification of transmembrane and GPI-anchored proteins on living cells by efficient protein trans-splicing using the Npu DnaE intein.* Chemical Communications, 2011. **47**: p. 3063-3065.
155. Charalambous, A., et al., *Split-inteins for simultaneous, site-specific conjugation of quantum dots to muliple protein targets in vivo.* Journal of Nanobiotechnology, 2011. **9**(37): p. 1-14.
156. Baird, G.S., D.A. Zacharias, and R.Y. Tsien, *Circular permutation and receptor insertion within green fluorescent proteins.* Proceeding of the National Academy of Science of the United States of America, 1999. **96**: p. 11241-11246.

157. Michnick, S.W., et al., *Universal strategies in research and drug discovery based on protein-fragment complementation assays.* Nature Reviews Drug Discovery, 2007. **6**: p. 569-582.
158. Kerppola, T.K., *Visualization of molecular interactions by fluorescence complementation.* Nature Reviews Molecular Cell Biology, 2006. **7**: p. 449-456.
159. Lee, J., et al., *Surface sites for engineering allosteric control of proteins.* Science, 2008. **322**: p. 438-442.
160. Strickland, D., et al., *Rationally improving LOV domain-based photoswitches.* Nature Methods, 2010. **7**(8): p. 623-626.
161. Wu, H., Z. Hu, and X.-Q. Lui, *Protein trans-splicing by a split intein encoded in a split DnaE gene of Synechocystis sp. PCC6803.* Proceeding of the National Academy of Science of the United States of America, 1998. **95**: p. 9226-9231.
162. Thomas C. Evans, J., et al., *Protein trans-splicing and cyclization by a naturally split intein from the dnaE gene of Synechocystis species PCC6803.* The Journal of Biological Chemistry, 2000. **275**(13): p. 9091-9094.
163. Martin, D.D., M.-Q. Xu, and J. Thomas C. Evans, *Characterization of a naturally occurring trans-splicing intein from Synechocystis sp. PCC6803.* Biochemistry, 2001. **40**: p. 1393-1402.
164. Giriat, I. and T.W. Muir, *Protein semi-synthesis in living cells.* Journal of American Chemical Society, 2003. **125**(24): p. 7180-7181.
165. Crosson, S. and K. Moffat, *Photoexcited structure of a plant photoreceptor domain reveals a light-driven molecular switch.* The Plant Cell, 2002. **14**: p. 1067-1075.
166. Crosson, S. and K. Moffat, *Structure of a flavin-binding plant photoreceptor domain: Insights into light-mediated signal transduction.* Proceeding of the National Academy of Science of the United States of America, 2001. **98**(6): p. 2995-3000.
167. Swartz, T.E., et al., *The photocycle of a flavin-binding domain of the blue light photoreceptor phototropin.* The Journal of Biological Chemistry, 2001. **276**(39): p. 36493-36500.
168. Amitai, G., et al., *Modulation of intein activity by its neighboring extein substrates.* Proceeding of the National Academy of Science of the United States of America, 2009. **PNAS Early Edition**: p. 1-6.
169. Pearl, E.J., et al., *Preceding hydrophobic and B-branched amino acids attenuate splicing by the CnePRP9 intein.* Biochimica et Biophysica Acta 2007. **1774**: p. 995-1001.
170. Chong, S., et al., *Modulation of protein splicing of the Saccharomyces cerevisiae vacuolar membrane ATPase intein.* The Journal of Biological Chemistry, 1998. **273**(17): p. 10567-10577.
171. Chong, S., et al., *Utilizing the C-terminal cleavage activity of a protein splicing element to purify recombinant proteins in a single chromatographic step.* Nucleic Acids Research, 1998. **26**(22): p. 5109-5115.
172. Lungu, O.I., et al., *Designing photoswitchable peptides using the AsLOV2 domain.* Chemistry & Biology, 2012. **19**: p. 507-517.
173. Li, J., et al., *Protein trans-splicing as a means for viral vector-mediated in vivo gene therapy.* Human Gene Therapy, 2008. **19**: p. 958-964.
174. Lockless, S.W. and T.W. Muir, *Traceless protein splicing utilizing evolved split inteins.* Proceeding of the National Academy of Science of the United States of America, 2009. **106**(27): p. 10999-11004.

175. Muona, M., A.S. Aranko, and H. Iwai, *Segmental isotopic labelling of a multidomain protein by protein ligation by protein trans-splicing.* ChemBioChem, 2008. **9**: p. 2958-2961.
176. Hultschig, C., et al., *Recent advances of protein microarrays.* Current Opinion in Chemical Biology, 2006. **10**: p. 4-10.
177. Liotta, L.A., et al., *Protein microarrays: Meeting analytical challenges for clinical applications.* Cancer Cell, 2003. **3**: p. 317-325.
178. MacBeath, G., *Protein microarrays and proteomics.* Nature Genetics Supplement, 2002. **32**: p. 526-532.
179. Abou-Nader, M. and M.J. Benedik, *Rapid generation of random mutant libraries.* Bioengineered Bugs, 2010. **1**(5): p. 337-340.
180. McCullum, E.O., et al., *Random mutagenesis by error-prone PCR.* In Vitro Mutagenesis Protocols: Third Edition, Methods in Molecular Biology, 2010. **634**(2): p. 103-109.
181. Cirino, P.C., K.M. Mayer, and D. Umeno, *Generating mutant libraries using error-prone PCR.* Directed Evolution Library Creation Methods in Molecular Biology, 2003. **231**(1): p. 3-9.
182. Romero, P.A. and F.H. Arnold, *Exploring protein fitness landscapes by directed evolution.* Nature Reviews Molecular Cell Biology, 2009. **10**: p. 866-876.
183. Jackel, C., P. Kast, and D. Hilvert, *Protein design by directed evolution.* Annual Review of Biophysics, 2008. **37**: p. 153-173.
184. Otomo, T., et al., *NMR observation of selected segments in a larger protein: Central-segment isotope labeling through intein-mediated ligation.* Biochemistry, 1999. **38**: p. 16040-16044.
185. Flavell, R.R. and T.W. Muir, *Expressed protein ligation (EPL) in the study of signal transduction, ion conduction, and chromatin biology.* Accounts of Chemical Research, 2008. **42**(1): p. 107-116.
186. Merrifield, R.B., *Solid phase peptide synthesis I. The synthesis of a tetrapeptide.* Journal of American Chemical Society, 1963. **85**(14): p. 2149-2154.
187. David, R., M.P.O. Richter, and A.G. Beck-Sickinger, *Expressed protein ligation method and applications.* European Journal of Biochemistry, 2004. **271**: p. 663-677.
188. Kanegae, Y., et al., *High-level expression by tissue/cancer-specific promoter with strict specificity using a single-adenoviral vector.* Nucleic Acids Research, 2010. **39**(2): p. e7.
189. Sato, Y., et al., *Enhanced and specific gene expression via tissue-specific production of Cre recombinase using adenovirus vector.* Biochemical and Biophysical Research Communications, 1998. **244**: p. 455-462.
190. Kanaya, T., et al., *hTERT is a critical determinant of telomerase activity in renal-cell carcinoma.* International Journal of Cancer, 1998. **78**: p. 539-543.
191. Fukazawa, T., et al., *Development of a cancer-targeted tissue-specific promoter system.* Cancer Research, 2004. **64**: p. 363-369.
192. Blasco, M.A., et al., *Functional characterization and developmental regulation of mouse telomerase RNA.* Science, 1995. **269**(5228): p. 1267-1270.
193. Ito, H., et al., *Expression of human telomerase subunits and correlation with telomerase activity in urothelial cancer.* Clinical Cancer Research, 1998. **4**: p. 1603-1608.
194. Forsyth, N.R., W.E. Wright, and J.W. Shay, *Telomerase and differentiation in multicellular organisms: Turn it off, turn it on, and turn it off again.* Differentiation, 2002. **69**: p. 188-197.

195. Lee, H.-W., et al., *Essential role of mouse telomerase in highly proliferative organs.* Nature, 1998. **392**: p. 569-574.
196. Thomson, J.A., et al., *Embryonic stem cell lines derived from human blastocytes.* Science, 1998. **282**: p. 1145-1147.
197. Hobert, O., R.J.J. Jr, and S. Chang, *Left-right asymmetry in the nervous system: The Caenorhabditis elegans model.* Nature Review Neuroscience, 2002. **3**: p. 629-640.
198. Jorgensen, E.M. and S.E. Mango, *The art and design of genetic screens: Caenorhabditis elegans.* Nature Reviews Genetics, 2002. **3**: p. 356-369.
199. Bargmann, C.I., *Genetic and cellular analysis of behavior in C. elegans.* Annual Review of Neuroscience, 1993. **16**: p. 47-71.
200. Ankers, J.M., et al., *Spatio-temporal protein dynamics in single living cells.* Current Opinion in Chemical Biology, 2008. **19**: p. 375-380.
201. Teruel, M.N. and T. Meyer, *Translocation and reversible localization of signaling proteins: A dynamic future for signal transduction.* Cell, 2000. **103**: p. 181-184.
202. Raftopoulou, M. and A. Hall, *Cell migration: Rho GTPases lead the way.* Developmental Biology, 2004. **265**: p. 23-32.
203. Ridley, A.J., *Rho GTPases and cell migration.* Journal of Cell Science, 2001. **114**: p. 2713-2722.
204. Berridge, M.J., M.D. Bootman, and H.L. Roderick, *Calcium signalling: Dynamics, homeostasis and remodelling.* Nature Reviews Molecular Cell Biology, 2003. **4**: p. 517-529.
205. Graves, T.K. and P.M. Hinkle, *Endoplasmic reticulum calcium storage and release in cells expressing misfolded growth hormone.* Growth Hormone & IGF Research, 2003. **13**: p. 36-43.
206. Koch, G.L.E., *The endoplasmic reticulum and calcium storage.* BIoEssays, 1990. **12**(11): p. 527-531.
207. Harris, G.L., et al., *In vitro and in vivo antagonism of a G protein-coupled receptor (S1P3) with a Novel blocking monoclonal antibody.* PLoS ONE, 2012. **7**(4): p. e35129.

Appendices

The appendices contain supplemental material that may be of interest to the reader. DNA and protein sequences of the main proteins used in this study are also presented along with additional controls.

The appendices are divided into the following sections:

Appendix A: DNA and Protein Sequence
Appendix B: Strategies for Improving Fluorescent Screening of Gene Inserts
Appendix C: Supplemental Co-localization Analysis of PTS Fluorescence Assay
Appendix D: Fluorescence Bleed-Through in Other Channels
Appendix E: Additional Control Experiments for the Mutant GCaMP2

Appendix A: DNA and Protein Sequence

The DNA and protein sequences of the main proteins used in this study are given below. This is followed by the DNA and protein sequences used in improving the fluorescent screening of gene inserts (Section 3.1.5). Protein sequences are shown in standard one letter code. Sequences of fluorescent proteins that were used as reporter tags are not shown.

*Npu*DnaE$_N$

DNA Sequence:

```
TGTTTAAGCTATGAAACGGAAATATTGACAGTAGAATATGGATTATTACCGATTGGTAAAATTG
TAGAAAAGCGCATCGAATGTACTGTTTATAGCGTTGATAATAATGGAAATATTTATACACAACC
TGTAGCACAATGGCACGATCGCGGAGAACAAGAGGTGTTTGAGTATTGTTTGGAAGATGGTTCA
TTGATTCGGGCAACAAAAGACCATAAGTTTATGACTGTTGATGGTCAAATGTTGCCAATTGATG
AAATATTTGAACGTGAATTGGATTTGATGCGGGTTGATAATTTGCCGAAT
```

Amino Acid Sequence:

```
CLSYETEILTVEYGLLPIGKIVEKRIECTVYSVDNNGNIYTQPVAQWHDRGEQEVFEYCLEDGS
LIRATKDHKFMTVDGQMLPIDEIFERELDLMRVDNLPN
```

*Npu*DnaE$_C$ (including the first three native extein sequences)

DNA Sequence:

TATGATCAAAATAGCCACACGTAAATATTTAGGCAAACAAAATGTCTATGACATTGGAGTTGAGCGCGAC
CATAATTTTGCACTCAAAAATGGCTTCATAGCTTCTAATTGTTTCAAT

Amino Acid Sequence:

MIKIATRKYLGKQNVYDIGVERDHNFALKNGFIASNCFN

Lyn

DNA Sequence:

ATGGGCTGCATCAAGAGCAAGGGCAAGGACAGCGCCACTAGT

Amino Acid Sequence:

MGCIKSKGKDSATS

TN-XL

DNA Sequence:

CTCAGCGAGGAGATGATTGCTGAGTTCAAAGCTGCCTTTGACATGTTTGATGCGGACGGTGGTG
GGGACATCAGCACCAAGGAGTTGGGCACGGTGATGAGGATGCTGGGCCAGAACCCCACCAAAGA
GGAGCTGGATGCCATCATCGAGGAGGTGGACGAGGATGGCAGCGGCACCATCGACTTCGAGGAG
TTCCTGGTGATGATGGTGCGCCAGATGAAAGAGGACGCCAAGGGCAAGTCTGAGGAGGAGCTGG
CCAACTGCTTCCGCATCTTCGACAAGGACGCTAACGGGTTCATCGACATCGAGGAGCTGGGTGA
GATTCTCAGGGCCACTGGGGAGCACGTCATCGAGGAGGACATAGAAGACCTCATGAAGGATTCA
GACAAGGACAATAACGGCCGCATTGACTTCGATGAGTTCCTGAAGATGATGGAGGGTGTGCAG

Amino Acid Sequence:

LSEEMIAEFKAAFDMFDADGGGDISTKELGTVMRMLGQNPTKEELDAIIEEVDEDGSGTIDFEE
FLVMMVRQMKEDAKGKSEEELANCFRIFDKDANGFIDIEELGEILRATGEHVIEEDIEDLMKDS
DKDNNGRIDFDEFLKMMEGVQ

GCaMP2

DNA Sequence:

GCCACCATGCGGGGTTCTCATCATCATCATCATCATGGTATGGCTAGCATGACTGGTGGACAGC
AAATGGGTCGGGATCTGTACGACGATGACGATAAGGATCTCGCCACCATGGTCGACTCATCACG
TCGTAAGTGGAATAAGACAGGTCACGCAGTCAGAGCTATAGGTCGGCTGAGCTCACTCGAGAAC
GTCTATATCATGGCCGACAAGCAGAAGAACGGCATCAAGGCGAACTTCAAGATCCGCCACAACA
TCGAGGACGGCGGCGTGCAGCTCGCCTACCACTACCAGCAGAACACCCCCATCGGCGACGGCCC
CGTGCTGCTGCCCGACAACCACTACCTGAGCACCCAGTCCAAACTTTCGAAAGACCCCAACGAG
AAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACG
AGCTGTACAAGGGCGGTACCGGAGGGAGCATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGT
GGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAG
GGTGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGC
CCGTGCCCTGGCCCACCCTCGTGACCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCC
CGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACATCCAGGAGCGC
ACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACA
CCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCA
CAAGCTGGAGTACAACACGCGTGACCAACTGACTGAAGAGCAGATCGCAGAATTTAAAGAGGCT
TTCTCCCTATTTGACAAGGACGGGGATGGGACAATAACAACCAAGGAGCTGGGGACGGTGATGC
GGTCTCTGGGGCAGAACCCCACAGAAGCAGAGCTGCAGGACATGATCAATGAAGTAGATGCCGA
CGGTAATGGCACAATCGACTTCCCTGAGTTCCTGACAATGATGGCAAGAAAAATGAAAGACACA
GACAGTGAAGAAGAAATTAGAGAAGCGTTCCGTGTGTTTGATAAGGATGGCAATGGCTACATCA
GTGCAGCAGAGCTTCGCCACGTGATGACAAACCTTGGAGAGAAGTTAACAGATGAAGAGGTTGA
TGAAATGATCAGGGAAGCAGACATCGATGGGGATGGTCAGGTAAACTACGAAGAGTTTGTACAA
ATGATGACAGCGAAG

Amino Acid Sequence:

ATMRGSHHHHHHGMASMTGGQQMGRDLYDDDDKDLATMVDSSRRKWNKTGHAVRAIGRLSSLEN
VYIMADKQKNGIKANFKIRHNIEDGGVQLAYHYQQNTPIGDGPVLLPDNHYLSTQSKLSKDPNE
KRDHMVLLEFVTAAGITLGMDELYKGGTGGSMVSKGEELFTGVVPILVELDGDVNGHKFSVSGE
GEGDATYGKLTLKFICTTGKLPVPWPTLVTTLTYGVQCFSRYPDHMKQHDFFKSAMPEGYIQER
TIFFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNTRDQLTEEQIAEFKEA
FSLFDKDGDGTITTKELGTVMRSLGQNPTEAELQDMINEVDADGNGTIDFPEFLTMMARKMKDT
DSEEEIREAFRVFDKDGNGYISAAELRHVMTNLGEKLTDEEVDEMIREADIDGDGQVNYEEFVQ
MMTAK

Venus$_N$

DNA Sequence:

CTGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCG
ACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCT
GACCCTGAAGCTGATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACC
CTGGGCTACGGCCTGCAGTGCTTCGCCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCA
AGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTA
CAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGC
ATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAAC

Amino Acid Sequence:

LVSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKLICTTGKLPVPWPTLVTT
LGYGLQCFARYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLVNRIELKG
IDFKEDGNILGHKLEYN

Venus$_C$

DNA Sequence:

AGTTACAACAGCCACAACGTCTATATCACCGCCGACAAGCAGAAGAACGGCATCAAGGCCAACTTCAAGA
TCCGCCACAACATCGAGGACGGCGGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGA
CGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCTACCAGTCCGCCCTGAGCAAAGACCCCAACGAG
AAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGCTAGT

Amino Acid Sequence:

SYNSHNVYITADKQKNGIKANFKIRHNIEDGGVQLADHYQQNTPIGDGPVLLPDNHYLSYQSAL
SKDPNEKRDHMVLLEFVTAAAS

mRFP$_N$

DNA Sequence:

ATGGCCTCCTCCGAGGACGTCATCAAGGAGTTCATYCGCTTCARGGTGCGCATGGAGGGCTCCG
TGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGCACCCAGAC
CGCCAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATCCTGTCCCCTCAG
TTCCAGTACGGCTCCAAGGCCTACGTGAAGCACCCCGCCGACATCCCCGACTACTTGAAGCTGT
CCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAGGACGGCGGCGTGGTGACCGT
GACCCAGGACTCCTCCCTGCAGGACGGCGAGTTCATCTACAAGGTGAAGCTGCGCGGCACCAAC
TTCCCCTCCGACGGCCCCGTAATGCARAAGAAG

Amino Acid Sequence:

MASSEDVIKEFIRFXVRMEGSVNGHEFEIEGEGEGRPYEGTQTAKLKVTKGGPLPFAWDILSPQ
FQYGSKAYVKHPADIPDYLKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGEFIYKVKLRGTN
FPSDGPVMQKK

mRFP$_C$

DNA Sequence:

TGGGAGGCCTCCACCGAGCGGATGTACCCCGAGGACGGCGCCCTGAAGGGCGAGATCAAGATGA
GGCTGAAGCTGAAGGACGGCGGCCACTACGACGCCGAGGTCAAGACCACCTACATGGCCAAGAA
GCCCGTGCAGCTGCCCGGCGCCTACAAGACCGACATCAAGCTGGACATCACCTCCCACAACGAG
GACTACACCATCGTGGAACAGTACGAGCGCGCCGAGGGCCGCCACTCCACCGGCGCC

Amino Acid Sequence:

WEASTERMYPEDGALKGEIKMRLKLKDGGHYDAEVKTTYMAKKPVQLPGAYKTDIKLDITSHNE
DYTIVEQYERAEGRHSTGA

RhoA$_N$

DNA Sequence:

GCTGCCATCCGRAARAAACTGGTGATTGTTGGTGATGGAGCCTGTGGAAAGACATGCTTGCTCATAGTCT
TCAGCAAGGACCAGTTCCCARAGGTGTATGTGCCCACAGTGTTTGARAACTATGTGGCARATATCGAGGT
GGAT

Amino Acid Sequence:

AAIRKKLVIVGDGACGKTCLLIVFSKDQFPXVYVPTVFENYVAXIEVD

RhoA$_C$

DNA Sequence:

GGAAAGCAGGTAGAGTTGGCTTTGTGGGACACAGCTGGGCTGGAARATTATGATCGCCTGAGGCCCCTCT
CCTACCCARATACCGATGTTATACTGATGTGTTTTTCCATCGACAGCCCTGATAGTTTARAAAACATCCC
ARAAAAGTGGACCCCARAAGTCAAGCATTTCTGTCCCAACGTGCCCATCATCCTGGTTGGRAATAAGAAG
GATCTTCGGAATGATGAGCACACAAGGCGGGAGCTTGCCAAGATGAAGCAGGAGCCGGTGAAACCTGAAG
AAGGCAGAGATATGGCAAACAGGATTGGCGCTTTTGGGTACATGGAGTGTTCAGCAAAGACCAAAGATGG
AGTGAGAGAGGTTTTTGAAATGGCTACGAGAGCTGCTCTGCAAGCT

Amino Acid Sequence:

GKQVELALWDTAGLEXYDRLRPLSYPXTDVILMCFSIDSPDSLXNIPXKWTPXVKHFCPNVPII
LVGNKKDLRNDEHTRRELAKMKQEPVKPEEGRDMANRIGAFGYMECSAKTKDGVREVFEMATRA
ALQA

Casp7$_N$ (p20)

DNA Sequence:

ACATATCAGTACAACATGAATTTTGAAAAGCTGGGCAAATGCATCATAATAAACAACAAGAACTTTGATA
AAGTGACAGGTATGGGCGTTCGAAACGGAACAGACAAAGATGCCGAGGCGCTCTTCAAGTGCTTCCGAAG
CCTGGGTTTTGACGTGATTGTCTATAATGACTGCTCTTGTGCCAAGATACAAGATCTGCTTAAAAAAGCT
TCTGAAGAGGACCATACAAATGCCGCCTGCTTCGCCTGCATCCTCTTAAGCCATGGAGAAGAAAATGTAA
TTTATGGGAAAGATGGTGTCACACCAATAAAGGATTTGACTGCCCACTTTAGGGGGGATAGATGCAAAAC
CCTTTTAGAGAAACCCAAACTCTTCTTCATTCAGGCTTGCCGAGGGACCGAGCTTGATGATGGCATCCAG
GCCGAC

Amino Acid Sequence:

TYQYNMNFEKLGKCIIINNKNFDKVTGMGVRNGTDKDAEALFKCFRSLGFDVIVYNDCSCAKIQ
DLLKKASEEDHTNAACFACILLSHGEENVIYGKDGVTPIKDLTAHFRGDRCKTLLEKPKLFFIQ
ACRGTELDDGIQAD

Casp7$_C$ (p11)

DNA Sequence:

GCTAATCCTCGATACAAGATCCCAGTGGAAGCTGACTTCCTCTTCGCCTATTCCACGGTTCCAGGCTATT
ACTCGTGGAGGAGCCCAGGAAGAGGCTCCTGGTTTGTGCAAGCCCTCTGCTCCATCCTGGAGGAGCACGG
AAAAGACCTGGAAATCATGCAGATCCTCACCAGGGTGAATGACAGAGTTGCCAGGCACTTTGAGTCTCAG

TCTGATGACCCACACTTCCATGAGAAGAAGCAGATCCCCTGTGTGGTCTCCATGCTCACCAAGGAACTCT
ACTTCAGTCAA

Amino Acid Sequence:

ANPRYKIPVEADFLFAYSTVPGYYSWRSPGRGSWFVQALCSILEEHGKDLEIMQILTRVNDRVA
RHFESQSDDPHFHEKKQIPCVVSMLTKELYFSQ

LOV2-Jα Domain

DNA Sequence:

TTGGCTACTACACTTGAACGTATTGAGAAGAACTTTGTCATTACTGACCCAAGATTGCCAGATA
ATCCCATTATATTCGCGTCCGATAGTTTCTTGCAGTTGACAGAATATAGCCGTGAAGAAATTTT
GGGAAGAAACTGCAGGTTTCTACAAGGTCCTGAAACTGATCGCGCGACAGTGAGAAAAATTAGA
GATGCCATAGATAACCAAACAGAGGTCACTGTTCAGCTGATTAATTATACAAAGAGTGGTAAAA
AGTTCTGGAACCTCTTTCACTTGCAGCCTATGCGAGATCAGAAGGGAGATGTCCAGTACTTTAT
TGGGGTTCAGTTGGATGGAACTGAGCATGTCCGAGATGCTGCCGAGAGAGAGGGAGTCATGCTG
ATTAAGAAAACTGCAGAAAATATTGATGAGGCGGCAAAAGAACTT

Amino Acid Sequence:

LATTLERIEKNFVITDPRLPDNPIIFASDSFLQLTEYSREEILGRNCRFLQGPETDRATVRKIR
DAIDNQTEVTVQLINYTKSGKKFWNLFHLQPMRDQKGDVQYFIGVQLDGTEHVRDAAEREGVML
IKKTAENIDEAAKEL

LOVInC

DNA Sequence:

TTGGCTACTACACTTGAACGTATTGAGAAGAACTTTGTCATTACTGACCCAAGATTGCCAGATA
ATCCCATTATATTCGCGTCCGATAGTTTCTTGCAGTTGACAGAATATAGCCGTGAAGAAATTTT
GGGAAGAAACATGAGGTTTCTACAAGGTCCTGAAACTGATCGCGCGACAGTGAGAAAAATTAGA
GATGCCATAGATAACCAAACAGAGGTCACTGTTCAGCTGATTAATTATACAAAGAGTGGTAAAA

AGTTCTGGAACCTCTTTCACTTGCAGCCTATGCGAGATCAGAAGGGAGATGTCCAGTACTTTAT
TGGGGTTCAGTTGGATGGAACTGAGCATGTCCGAGATGCTGCCGAGAGAGAGGGAGTCATGCTG
ATTAAGAAAACTGCAGAAAATATTGATGAGGCGGCAAAAGAACTTGCTAGTGGATCCAAATATT
TAGGCAAACAAAATGTCTATGACATTGGAGTTGAGCGCGACCATAATTTTGCACTCAAAAATGG
CTTCATAGCTTCTAATTGTTTCAAT

Amino Acid Sequence:

LATTLERIEKNFVITDPRLPDNPIIFASDSFLQLTEYSREEILGRNMRFLQGPETDRATVRKIR
DAIDNQTEVTVQLINYTKSGKKFWNLFHLQPMRDQKGDVQYFIGVQLDGTEHVRDAAEREGVML
IKKTAENIDEAAKELASGSKYLGKQNVYDIGVERDHNFALKNGFIASNCFN

CaM

DNA Sequence:

GCTGATCAGCTGACCGAARAACAGATTGCTGAATTCAAGGAAGCCTTCTCCCTATTTGATAAAG
ATGGCGATGGCACCATCACAACAAAGGAACTTGGAACTGTCATGAGGTCACTGGGTCARAACCC
AACAGAAGCTGAATTGCAGGATATGATCAATGAAGTGGATGCTGATGGTAATGGCACCATTGAC
TTCCCCGAATTTTTGACTATGATGGCTAGAAAAATGAAAGATACAGATAGTGAARAAGAAATCC
GTGAGGCATTCCGAGTCTTTGACAAGGATGGCAATGGTTATATCAGTGCAGCARAACTACGTCA
CGTCATGACAAACTTAGGAGAAAAACTAACAGATGAAGAAGTAGATGAAATGATCAGAGAAGCA
GATATTGATGGAGACGGACAAGTCAACTATGAAGAATTCGTACAGATGATGACTGCAAAA

Amino Acid Sequence:

ADQLTEXQIAEFKEAFSLFDKDGDGTITTKELGTVMRSLGQNPTEAELQDMINEVDADGNGTID
FPEFLTMMARKMKDTDSEXEIREAFRVFDKDGNGYISAAXLRHVMTNLGEKLTDEEVDEMIREA
DIDGDGQVNYEEFVQMMTAK

Appendix B: Strategies for Improving Fluorescent Screening of Gene Inserts

Approach #1: N-terminally situated well-folded and expressed protein

One method to enhance fluorescence screening is to improve the folding or enhanced expression of a gene insert by an N-terminal protein that is well-folded and expressed such as CaM. The cassette vector with CaM upstream of gene insertions (Figure 3.6) was tested with the insertion of a random gene obtained from PCR. In this case, the fluorescent protein hcRed was chosen because it expresses fluorescence and can act as an additional reporter for successful gene insertion, although this is not strictly necessary as the vector intrinsically contains Venus as a fluorescent protein reporter. When transformed *E.Coli* colonies were plated on agar plates and grown overnight at 37^{o}C, a substantial increase in Venus fluorescence intensity was observed (Figure B.1). Enhanced Venus fluorescence was also observed when other genes of varying length and folding efficiencies were inserted downstream of the CaM. For example, the gene encoding RhoA of ~330 amino acids and the human p21 protein of ~220 amino acids that is known to be poorly expressed in bacteria (Figure B.1).

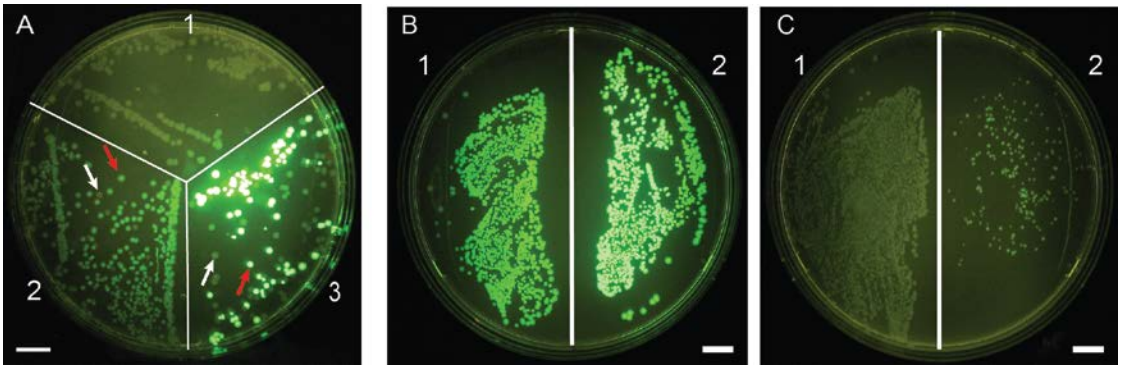

Figure B.1 Comparison of fluorescent images of bacterial platings.

(**A**) Insertion of the gene encoding hcRed into the two different vectors in quadrant (QD) 2 and 3. QD1 – *E. Coli* with no fluorescent proteins (control); QD2 – the standard cassette vector; QD3 – cassette vector with CaM. (**B**) Gene insertion of RhoA cloned in the CaM vector (QD2) relative to the standard cassette vector control (QD1). (**C**) Likewise, gene insertion of p21 in the CaM vector (QD2) relative to the control (QD1). Red arrows indicate fluorescent colonies and white arrows indicate non-fluorescent colonies. Scale bar represents 1 cm.

Fluorescent colonies from each type were chosen, grown overnight, and culture growth adjusted to obtain equal cell densities. The fluorescence intensity increase as measured with a fluorometer determined that the expression vector with CaM showed ~30x increased in fluorescence relative to non-fluorescent colonies, whereas the standard cassette vector only showed a ~3x increase in fluorescence over non-fluorescent colonies (Figure B.2). This increase in fluorescence greatly facilitates the discrimination of bacterial colonies with successful gene insertions from unsuccessful colonies.

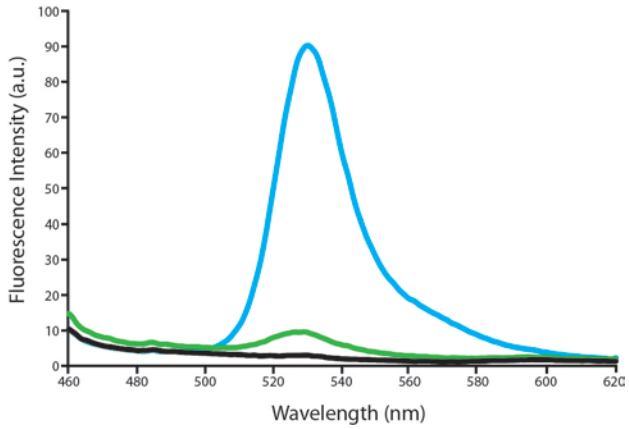

Figure B.2 Relative fluorescence intensities of reporting YFP.

Emission spectrum of YFP fluorescence of the cassette vector with CaM (blue), the standard cassette vector (green), and a non-fluorescent plasmid (black) showing varying fluorescence intensities.

Serial dilutions of a gene insert were inserted into the CaM vector to mimic sub-optimal ligation conditions. Under low expressed fluorescence, it would be difficult to discern and separate the successful colonies from non-successful ones. However, the increased fluorescence using the CaM vector facilitated this (Figure B.3).

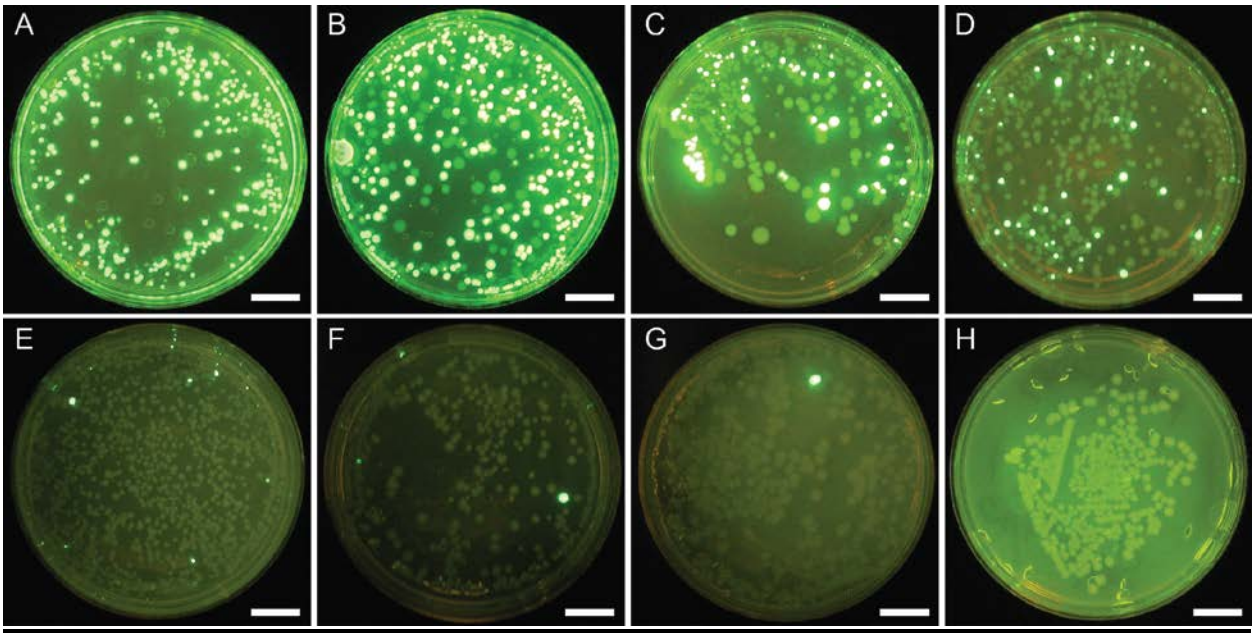

Figure B.3 Platings of serial diluted sub-optimal ligations.

E. Coli platings of sub-optimal ligation reactions of serial diluted inserts of hcRed gene fragment into the cassette vector with CaM. Transfection efficiencies: (**A**) 100% (positive control); (**B**) 50%; (**C**) 25%; (**D**) 12.5%; (**E**) 6.25%; (**F**) 3.125%; (**G**) 1%; (**H**) 0% (negative control). Scale bar represents 1 cm.

Approach #2: Reconstitution of Truncated FP

The second method involves reconstituting a truncated non-fluorescent FP to restore fluorescence after gene insertion and thus allow for screening of successfully inserted genes. Here, rather than have the reporter FP downstream of the gene insert, a truncated FP is situated upstream of the gene insert. By designing PCF fragments to include the missing portions of the FP, the inserted PCR fragment will complete the β-barrel structure of the truncated FP and express fluorescence (Figure 3.6). Care must be taken to ensure that the inserted gene is in-frame with the FP as there as a possibility that the downstream protein can be inserted with a frame shift after rescuing the truncated FP. To test this approach, mRFP was chosen as the FP for truncation as it is stable and matures quickly. Fluorescence was abolished after 10 amino acids were removed (Figure B.4).

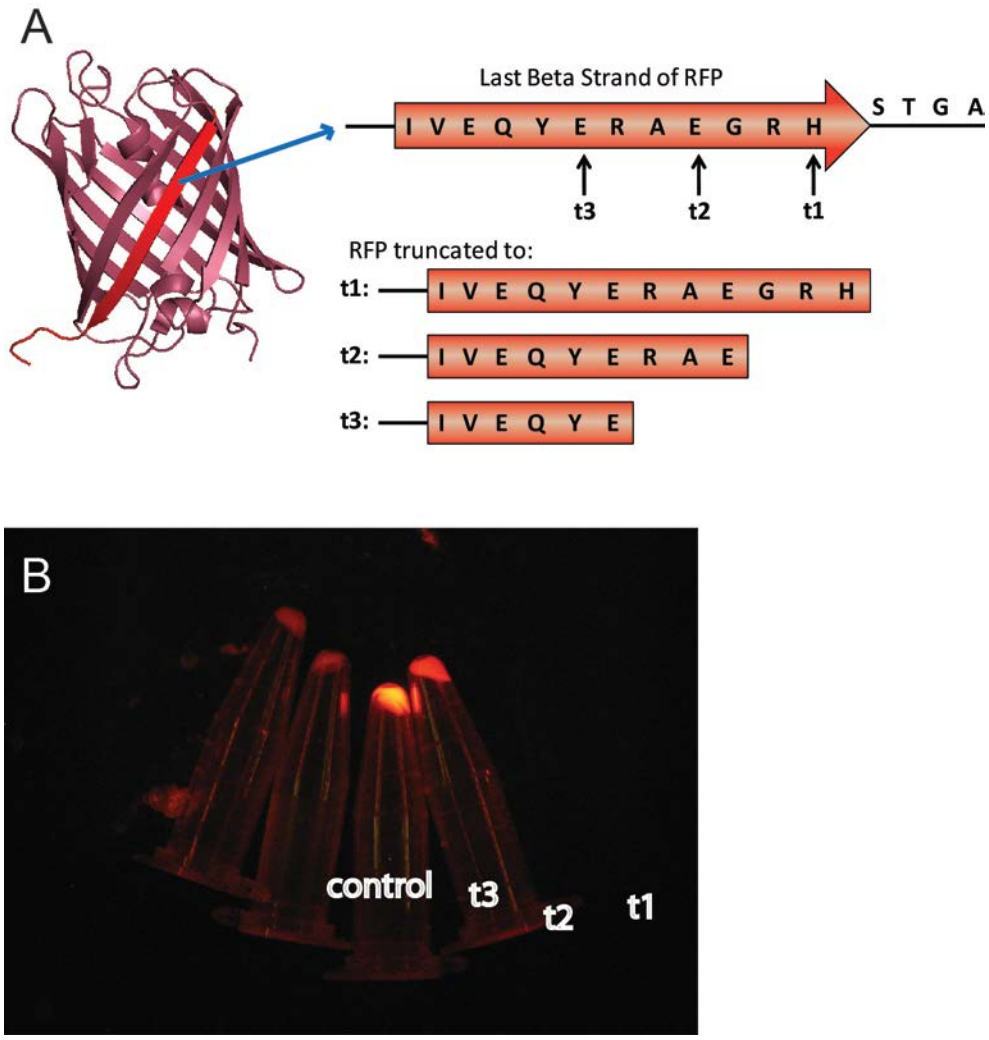

Figure B.4 C-terminal truncations of mRFP.

(**A**) Schematic diagram of the truncations of the last β-strand of mRFP. (**B**) Fluorescence image of pelleted *E. Coli* cells of a select truncated mRFP showing fluorescence was abolished after 10 amino acids were removed. Of note is that the fluorescence intensity increased when 7 amino acids were removed.

Surprisingly, red fluorescence increased in intensity after 7 amino acids were removed. This is ideal as shorter complementary strands reduce cost and increase PCF efficiency while still allowing for fluorescence. Therefore, this shorter complementary strand was used. To test this approach, PCR primers for Cerulean were designed to include the complementary strand to methodically complete the truncated mRFP. The insertion of the PCR fragment yielded colonies with bright fluorescence (Figure B.5).

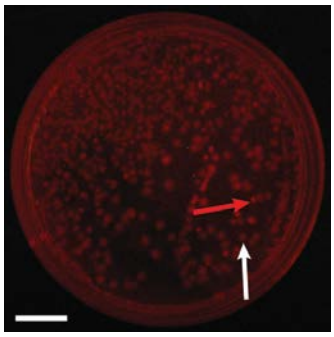

Figure B.5 Platings of the completed truncated mRFP.

Truncated mRFP was completed upon gene insert was able to show RFP fluorescence under a fluorescence plate reader. Red arrows indicate fluorescent colonies and white arrows indicate non-fluorescent colonies. Scale bar represents 1 cm.

In addition, because the expression vector backbone contains a mammalian promoter, the final construct can be directly used in cell imaging studies. As a demonstration, the rescued mRFP fused with Cerulean was directly transfected into three different cell lines (Figure B.6).

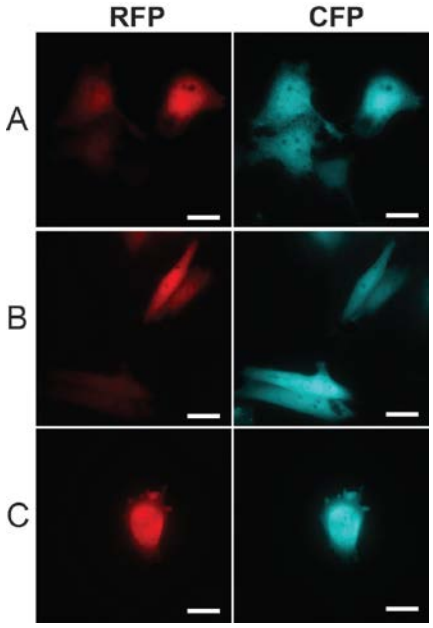

Figure B.6 Cell cultures expressing the rescued mRFP fused with Cerulean.

The plasmid expressing vector expressing the rescued truncated mRFP was directed transfected into three different cell lines. (**A**) COS7; (**B**) CHO; and (**C**) HeLa. In all three cell lines, transfected cells expressed both red and cyan fluorescence. Scale bar represents 10 μm.

Approach #3: Utilizing Shine-Dalgarno Sequence

The dependence of the reporter FP can be isolated from the inserted gene by exploiting the Shine-Dalgarno sequence. The Shine-Dalgarno sequence allows bacterial expression to start at a position other than the built-in T7 promoter. In designing the PCR primers of the gene of interest, a stop codon, a Shine-Dalgarno sequence, and a start codon is included in the 3'- primer sequence (Figure 3.6). Using this method, mRFP was cloned into the standard cassette vector and transformed cells were plated. On-plate colonies with successful gene insertion showed YFP as well as RFP fluorescence (Figure B.7).

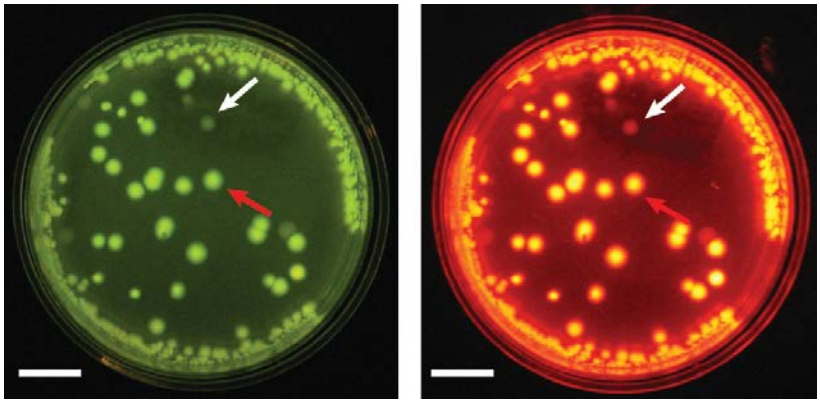

Figure B.7 Fluorescent images of bacterial platings using the Shine-Dalgarno Sequence.

Bacterial platings showing both YFP and RFP fluorescence indicating that the gene was successfully inserted. Note that YFP fluorescence of successful colonies is strong enough to easily differentiate it from non-fluorescence colonies. Red arrows indicate fluorescent colonies and white arrows indicate non-fluorescent colonies. Scale bar represents 1 cm.

Appendix C: Supplemental Co-localization Analysis of PTS Fluorescence Assay

The co-localization of the FP in the PTS fluorescence assay was also determined using PC with Costes' automatic thresholding and Van Steensel's cross-correlation function (CFF) obtained from the JACoP plug-in [110]. The two scenarios are presented below.

Precursors	PTS	PC with Costes' Thresholding	Van Steensel's CCF	Ratio
Lyn-Ceru-*Npu*DnaE$_N$-mRFP + *Npu*DnaE$_C$-Venus	Before	0.934	0.946	CFP:RFP
	After	0.483	0.770	
Lyn-Ceru-*Npu*DnaE$_N$ + mRFP-*Npu*DnaE$_C$-Venus	Before	0.964	0.965	YFP:RFP
	After	0.653	0.716	

Table C.1 Comparison of co-localization of PF before and after PTS.

The first set of results is comparing CFP:RFP as the two fluorescence are initially co-localized at the PM. After PTS, mRFP is translocated from the PM to the cytoplasm. The second set of results is comparing YFP:RFP as the two FP are initially co-localized in the cytoplasm before PTS. After PTS, YFP is translocated to the PM to be bound there while RFP remains in the cytoplasm. In both cases, there should be a drop in PC values as is seen.

Appendix D: Fluorescence Bleed-Through in Other Channels

Significant fluorescence bleed-through in other fluorescent channels could significantly affect the accuracy of the results obtained. To verify that the filters used in this study are not affected by bleed-through and sufficiently separate of the emitted fluorescence, cells expressing one of the FP proteins (i.e. either CFP, YFP, or RFP) were examined for emitted fluorescence in the other channels. Average fluorescence intensity were obtained within the region of interest from all three channels and compared to the control (the expressed FP and the emitted fluorescence in that channel). Results are shown below.

	Fluorescence Channel		
Expressed FP	**CFP**	**YFP**	**RFP**
CFP	**1.000**	0.039	0.032
YFP	0.072	**1.000**	0.041
RFP	0.042	0.068	**1.000**

Table D.1 Comparison of fluorescence bleed-through

The 'Expressed FP' column represents the FP expressed in the cell. Control ratios are bolded. All other ratios were obtained relative to the control. In all cases, the bleed-through is minimal.

Appendix E: Additional Control Experiments for the Mutant GCaMP2

The efficacy of the mutant GCaMP2 was also tested in two other cell lines, COS7 and CHO. The results are presented below. The mutant GCaMP2 was able to detect ATP-induced Ca^{2+} transients that had a similar profile as the unaltered GCaMP2.

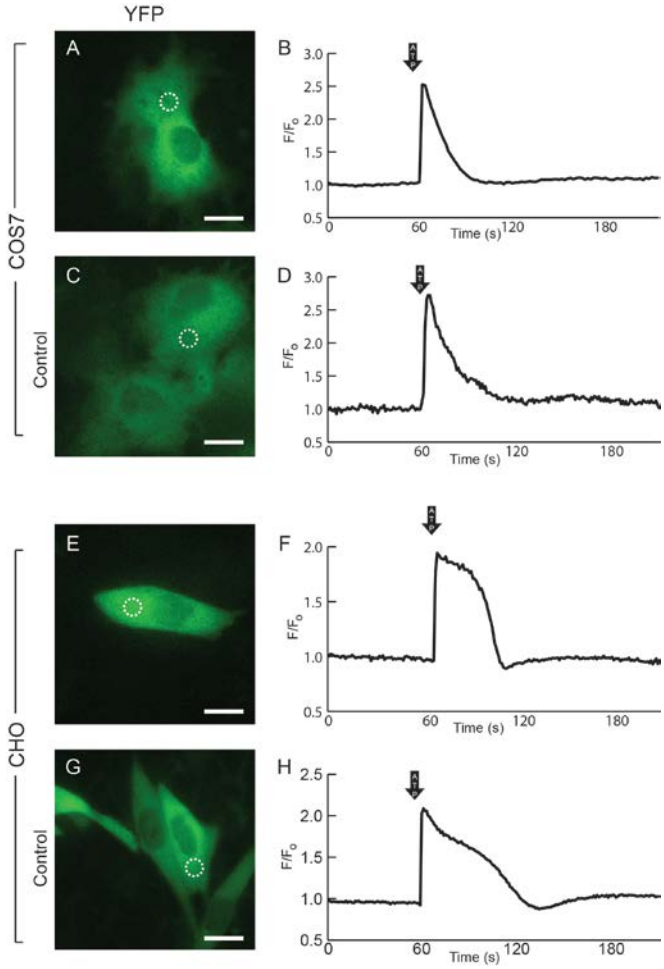

Figure E.1 Substitution of "GGTGGS" linker with "CFNGT" did not affect the behaviour of the GCaMP2 biosensor.

An altered GCaMP2 biosensor was created by replacing the native linker "GGTGGS" with "CFNGT". Expression of this altered GCaMP2 biosensor in COS7 cells exhibited (**A**) YFP fluorescence and (**B**) recorded an ATP-induced Ca^{2+} transient profile. This was similar to the recorded profiles of unaltered GCaMP2 (**C-D**). (**E-H**) Similar results were obtained in CHO cells. White dotted circle represent regions where average fluorescence intensity measurements were taken. Images are in false colour. Scale bars represent 10 μm.