Expanding the applications of dimerization-dependent fluorescent proteins

by

Yan Tiffany Lai

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

Master of Science

Department of Chemistry

University of Alberta

Abstract

The expansion of the fluorescent protein (FP) colour palette through discovery of more species possessing FPs and by subjecting them to intensive protein engineering efforts has greatly aided the effectiveness of FP-based biosensor designs. Currently, the main FP-based methods for detecting protein-protein interactions, enzymatic activity, and changes in small molecule concentration in live cells, are Förster resonance energy transfer (FRET) and bimolecular fluorescence complementation (BiFC). A new biosensor strategy has recently been introduced by our group: dimerization-dependent FPs (ddFPs). This novel platform is based on the change in fluorescence intensity resulting from the reversible heterodimeric association of two distinct ddFP monomers. We believe that ddFPs can overcome some challenges in live cell imaging currently imposed by traditional biosensor platforms. For example, ddFPs may facilitate multicolour imaging of biosensors in live cells due to the fact that they have a narrower spectral profile than a FRET pair.

In this thesis, we describe our efforts to expand the range of applications of ddFP in live cells. We report the creation of systems for the sensing of global O-linked β-*N*-acetylglucosamine (O-GlcNAc) modifications in live cells by O-GlcNAc transferase (OGT) and validation of OGT protease activity to host-cell factor-1 which may be the link between nutrient levels and the cell cycle. In addition, we apply ddFPs for sensing of tomosyn-1 SUMOylation by small ubiquitin modifier 1 (SUMO-1) in pancreatic islet cells. The creation of these plasmids is a proof of the versatility and modularity of the ddFP and its related FP exchange (FPX) biosensor design strategy.

We also report the engineering of a dark "A" copy (DA) for ddFP applications. By a simple rational mutation of the chromophore of dimerization-dependent green A copy

(GA), DA was generated. We confirmed that DA competes with GA for binding to the B copy. Specifically, we demonstrated the utility of DA in combination with GA and B as an intensiometric “turn on” biosensor for monitoring caspase-3 activity. Using both ddFPs and DA we endeavour to develop a facile multiparameter imaging platform.

Table of Contents

Abstract ..... **II**
Table of Contents ..... **IV**
List of Figures ..... **VI**
List of Tables ..... **VIII**
List of Appendices ..... **IX**
List of Abbreviations ..... **X**
Chapter 1 Introduction **1**
1.1 Background **1**
1.1.1 FP-based FRET reporters 3
1.2 General introduction to protein complementation assays **6**
1.2.1 FP complementation strategies 8
1.2.1.1 Traditional approaches to FP complementation: Bimolecular Fluorescence Complementation (BiFC) 10
1.2.1.1.1 The expansion of the BiFC colour palette and its versatility 11
1.2.1.1.1.1 YFP and CFP variants 11
1.2.1.1.1.2 GFP variants 12
1.2.1.1.1.3 RFP variants 14
1.2.1.1.1.4 Bacterial phytochrome photoreceptors for near infrared BiFC 16
1.2.1.1.2 BiFC in RNA-protein interaction 18
1.2.1.1.3 BiFC for genome-wide PPI screening 22
1.2.1.1.4 BiFC for single-molecule localization-based superresolution imaging 24
1.2.1.1.5 BiFC applications for labelling protein dynamics 27
1.2.1.2 Tripartite split-GFP 29
1.2.1.3 Alternate FP beta strand modifications inspired by split sfGFP 32
1.2.1.4 Dimerization-dependent fluorescent proteins (ddFPs) 33
1.3 Scope of Thesis **35**
Chapter 2 Biosensing Strategies for O-GlcNAcylation and SUMOylation **36**
2.1 Introduction **36**
2.2 Results and Discussion **39**
2.2.1 Construction of plasmids for OGT glycosylation activity 39
2.2.2 Construction of plasmids for OGT protease activity 42
2.2.3 Construction of plasmids for SUMOylation of tomosyn-1 by SUMO-1 44
2.3 Conclusion **44**
2.4 Materials and Methods **45**

2.4.1 General methods and materials 45
2.4.2 Site-directed mutagenesis 46
2.4.3 Construction of mammalian expression plasmid for live cell imaging 47
2.4.4 Mammalian cell culture and imaging for OGT glycosylation activity 48
Chapter 3 Engineering a dark accessory protein for ddFPs 50
3.1 Introduction 50
3.2 Results and Discussion 52
3.2.1 Development and characterization of a Dark A copy (DA) 52
3.2.2 Live cell applications 63
3.2.3 Discussion of Tyr67Trp-DA properties 63
3.3 Conclusion 68
3.4 Materials and Methods 69
3.4.1 General methods and materials 69
3.4.2 Site-directed mutagenesis 70
3.4.3 Protein purification and quantification 70
3.4.4 Determination of secondary structure 72
3.4.5 Determination of heterodimer formation 72
3.4.6 Determination of IC_{50} 72
3.4.7 Determination of K_d 73
3.4.8 Live cell imaging applications: Construction of mammalian expression plasmid co-expressing DA-DEVD-GB, GA and $mCherry^{CaaX}$ 73
3.4.9 Mammalian cell culture and imaging 74
3.4.10 Image analysis 75
Chapter 4 Conclusions and future directions 76
4.1 Summary of thesis 76
4.2 Future directions 78
4.3 Concluding remarks 80
References 82
Appendices 98

List of Figures

Figure 1.1. Intra- and intermolecular FP-based FRET reporter design strategies 4

Figure 1.2. Schematic summary of the yeast-two hybrid assay and protein complementation assay 7

Figure 1.3. Flow chart presenting the categorization of the different protein complementation assays 9

Figure 1.4. Scheme of split sfGFP 13

Figure 1.5. Design of iSplit 17

Figure 1.6. Schematic of live-cell visualization of RNA using the bacteriophage MS2 coat protein (MCP)-GFP construct and trimolecular fluorescence complementation (TriFC).... 19

Figure 1.7. Principle of TimeSTAMP for tracking protein fate 28

Figure 1.8. Schematic illustration of tripartite split-GFP and tetramolecular fluorescence complementation (TetFC)..... 30

Figure 1.9. General graphic illustration of the β-strand displacement system 32

Figure 1.10. Schematic of fluorescent protein exchange (FPX) strategy and applications in live cells 34

Figure 2.1. O-GlcNAcylation by O-GlcNAc transferase (OGT) 36

Figure 2.2. The SUMOylation cycle 38

Figure 2.3. OGT O-GlcNAcylation biosensor strategy 41

Figure 2.4. Construction of expression plasmids harbouring DNA sequences encoding p2A peptides flanked by multiple cloning sites 42

Figure 2.5. OGT protease activity sensor 43

Figure 2.6. Map of multiple expression plasmid encoding RA-SUMO-1, m-Tomosyn-1-RB and $EGFP^{CaaX}$ in modified pBAD/His B harbouring p2A peptide and multiple cloning sites. 44

Figure 3.1. Absorption and emission spectra of DA variants (Ex. 450 nm) 53

Figure 3.2. CD spectra of DA variants to PBS, GA and GB control 54

Figure 3.3. Emission spectra of (A) GA, (B) Tyr67Trp-DA and (C) Tyr67Phe-DA at 0, 5, 10, 25, 50, 100, 200 μM (Ex. 455 nm) 55

Figure 3.4. Determination of oligomerization state of Tyr67Trp-DA and GB by size exclusion chromatography 57

Figure 3.5. Determination of Tyr67TrpDA IC_{50} using stopped-flow technique 58

Figure 3.6. Competitive inhibition of GA+GB complex by Tyr67Trp-DA 62

Figure 3.7. K_{DA} determination 63

Figure 3.8. Imaging caspase-3 activity with ddGFP and Tyr67Trp-DA 64

Figure 3.9. Estimation of the chromophore cavity environment of Tyr67Trp-DA 66

List of Tables

Table 2.1 Details of the gene construct created in this work 47
Table 3.1 Details of the gene construct used in this work 75

List of Appendices

Appendix A: Oligonucleotide sequence supplement 98

Appendix B: Sequence of pBAD-*Xho*I-EGFP-*Mlu*I-p2Ap2A-*Xba*I-*Sal*I-p2Ap2A-*EcoR*I-*Hind*III 100

Appendix C: Non-linear and linear regression analysis of Figure 3.6A,B 101

List of Abbreviations

ε	extinction coefficient
ϕ	quantum yield
A	acceptor fluorescent protein
Abs	absorbance
A copy	dimerization-dependent fluorescent protein A
AD	activating domain
AdHTS	adenovirus high-throughput system
Ala	alanine
AP-1	activator protein
Arg	arginine
Asn	asparagine
Asp	aspartate
avGFP	GFP derived from *Aequeorea victoria*
B copy	dimerization-dependent fluorescent protein B
BiFC	bimolecular fluorescence complementation
BONCAT	bio-orthogonal non-canonical amino acid tagging
BphPs	bacterial phytochrome photoreceptors
BV	biliverdin
bJun	basic region leucine zipper domain of c-Jun
bFos	basic region leucine zipper domain of c-Fos
C	carboxyl
CaaX	cysteine-aliphatic residue-aliphatic residue-C-terminal amino acid depending on specificity

CET	cysteine-glutamate-threonine
CFP	cyan fluorescent protein
CHO	Chinese hamster ovary
CD	circular dichroism
cm	centimeter
Cys	cysteine
D	donor fluorescent protein
DAB	diaminobenzidine
DA	dimerization-dependent dark fluorescent protein A
DBD	DNA-binding domain
ddFP	dimerization-dependent fluorescent protein
ddGFP	dimerization-dependent green fluorescent protein
ddRFP	dimerization-dependent red fluorescent protein
ddYFP	dimerization-dependent yellow fluorescent protein
DEVD	Aspartate-Glutamate-Valine-Aspartate
DMEM	Dulbecco's modified eagle media
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
dNTPs	Deoxynucleotide triphosphates
DsRed	*Discosoma* species red fluorescent protein
e1F4A	eukaryotic initiation factor 4A
EB1	end-binding 1
E. coli	*Escherichia coli*

EGF	epidermal growth factor
EGFR	epidermal growth factor receptor
em	emission
ERK	extracellular signal-regulated kinase
ex	excitation
FACS	fluorescence-activated cell sorting
FBS	fetal bovine serum
FMRP	fragile X mental retardation protein
FPALM	fluorescent PALM
FP	fluorescent protein
FPX	fluorescent protein exchange
FRET	Förster Resonance Energy Transfer
FT	fluorescent timer
FUNCAT	fluorescent non-canonical amino acid tagging
GA	dimerization-dependent green fluorescent protein A copy
GAF	cGMP phosphodiesterase/adenylate cyclase/Fh1A transcriptional activator
GAF_m	mutated GAF
GFP	green fluorescent protein
GFP 1-9	β-strands of the barrel GFP1-9
GFP 1-10 OPT	optimized GFP *β*-strands 1-10
GFP10	*β*-strand 10 of the GFP
GFP11	*β*-strand 11 of the GFP

GG	glycine-glycine
Gln	glutamine
Glu	glutamate
GPCR	G protein-coupled receptor
GRASP	GFP Reconstitution Across Synaptic Partners
HCF-1	host cell factor-1
$HCF\text{-}1_C$	C-terminal portion of cleaved host cell factor-1
$HCF\text{-}1_N$	N-terminal portion of cleaved host cell factor-1
HEK293	human embryonic kidney 293
HeLa	cervical cancer cell line originating from Henrietta Lacks
HHBSS	HEPES-buffered Hank's balanced salt solution
HIV-1	human immunodeficiency virus 1
Hsp47	heat shock protein 47
ID	intrinsically disordered
IFP1.4	Infrared FP version 1.4
IMP1	insulin-like growth factor II mRNA-binding protein 1
ITC	isothermal titration calorimetry
K	Kelvin
kb	kilobase
kDa	kilodalton
LB	lysogeny broth/ Luria Bertani
LEDGF/p75	lens epithelium-derived growth factor

Leu	leucine
Lys	lysine
MBD2	mCpG dinucleotide binding domain
MBS	MS2 binding site
MCP	MS2 coat protein
mCpG	methyl cytosine-guanine
MCS	multiple cloning site
mDHFR	murine dihydrofolate reductase
Met	methionine
mKG	monomeric Kusabira-Green
mL	millimeter
mRFP1	monomeric red fluorescent protein
mRNA	messenger ribonucleic acid
mtRNA	mitochondrial ribonucleic acid
MaMTH	mammalian-membrane two-hybrid assay
MYTH	membrane yeast two-hybrid
N	amino
ND6	NADH dehydrogenase subunit 6
NES	nuclear exclusion signal
ng	nanogram
NHEJ	non-homologous end joining pathway
NIR	near infrared
NLS	nuclear localization signal

NTA	nitrilotriacetic acid
nm	nanometer
OGA	O-GlcNAcase
O-GlcNAc	O-linked N-Acetyl-D-glucosamine
OGT	O-GlcNAc transferase
PAFP	photoactivatable fluorescent proteins
PA-GFP	photoactivatable GFP
PALM	photoactivated localization microscopy
PAmCherry	photoactivatable mCherry
PAS	Per-ARNT-Sim repeats
PBS	phosphate-buffered saline
PCA	protein complementation assay
PCM	photosensory core module
PCP	PP7 bacteriophage coat protein
PCR	polymerase chain reaction
PDB	protein data bank
Pen Strep	Penicillin Streptomycin
Phe	phenylalanine
PHY	phytochromes-specific domain
PIP_2	phosphatidylinositol 4,5-biphosphate
PKA	protein kinase A
POI	protein-of-interest
PP2A	protein phosphatase 2A

PPD	proteolytic processing domain
PPI	protein-protein interaction
ppm	parts per million
Pro	proline
PTB	polypyrimidine-tract-binding protein
PTM	post-translational modification
PUM-HD	pumilio homology domain
PUM-Var1	pumilio mutated variant
PUM-WT	PUM-HD wild type
Q	glutamate
QY	quantum yield
RA	dimerization-dependent red fluorescent protein A copy
RNA	ribonucleic acid
RBP	RNA binding protein
rpm	revolutions per minute
RRE	Rev response element
$s10^{203T}$	β-strand 10 with threonine at residue 203
$s10^{203Y}$	β-strand 10 modified at residue 203 from threonine to tyrosine
SATD	serine-alanine-threonine-aspartate
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
SEER	sequence enabled reassembly
SENP	Sentrin/SUMO-specific protease

Ser	serine
sfGFP	superfolder GFP
SILAC	stable isotope labelling with amino acids in cell culture
SNAP25	synaptosomal-associated protein 25 kDa
STAT5	signal transducers and activators of transcription
STORM	stochastic optical reconstruction microscopy
SUMO	small ubiquitin-like modifier
TAP	Tip-associated protein
TAR	trans-activation response element
TB	terrific broth
TetFC	tetramolecular fluorescence complementation
Thr	threonine
TPA	12-*O*-tetradecanoylphorbol-13-acetate
TPRs	tricorpeptide repeats
TriFC	trimolecular fluorescence complementation system
Trp	tryptophan
Tyr	tyrosine
UDP-GlcNAc	uridine diphosphate-β-*N*-acetyl glucosamine
µg	microgram
µL	microliter
µM	micromolar
UTR	untranslated region

UV	ultraviolet
Val	valine
VAMP2	vesicle-associated membrane protein 2
Y2H	yeast two-hybrid
YFP	yellow fluorescent protein
ZF	zinc finger

Chapter 1: Introduction

1.1 Background

The introduction of fluorescent proteins (FP) as genetically encoded reporters has given researchers a superb platform for non-invasive real-time imaging of biomolecules in live cells. The cloning and first recombinant expression of the green fluorescent protein (GFP) derived from the bioluminescent jellyfish *Aequorea victoria* occurred two decades ago [1-3]. Today, FPs are extensively used as molecular biosensors to detect bioanalyte flux, protein localization and protein interaction in cells [4, 5]. Increased knowledge on cellular dynamics, particularly protein interactions which play a critical role in many cellular processes, is essential to combat diseases such as cancer. FPs are one of the most powerful class of tools in the molecular biologist's toolbox.

The ability of FPs to absorb and emit light is attributed to its intrinsic chromophore. The FP sequence encodes a tripeptide which, through a process of a self-sufficient cyclization, oxidation and loss of a water molecule, is converted into the chromophore [6]. The eleven-stranded β-barrel structure of FPs wraps around the central alpha-helix where the chromophore resides thus ensuring a rigid environment to protect the chromophore from bulk solvent and collisional quenching by oxygen species [7]. The central location of the chromophore in the barrel places it in the immediate environment of a number of amino acid residue side chains which the chromophore interacts with [8]. Mutation of interacting residues can confer dramatic changes to the fluorescent properties of an FP [9].

Substantial protein engineering efforts on the *Aequorea* GFP further expanded the available colour spectrum from blue to yellow [10, 11]. However, a red FP was still much

sought after, as lower excitation energy (i.e., longer wavelength) will reduce phototoxicity during imaging of live cells [12]. Furthermore, with a greater range of colour variants multiparameter imaging of multiple structures or processes in a cell is facilitated. At lower wavelength excitation light associated with red FPs tissue autofluorescence, scattering and absorbance by hemoglobin, melanin and water is decreased [12, 13]. This allows deeper tissue penetration which is of particular importance for *in vivo* and deep tissue imaging in small animals such as mice and rats. The discovery of DsRed derived from *Discosoma* sp. finally provided a potentially useful red FP, but unfortunately it suffered from some limitations [14]. DsRed is an obligate tetramer and has been found to emit green fluorescence during its slow maturation [15]. Imaging of DsRed that is genetically fused to certain proteins in cells have also revealed mislocalization due to the inherent tetrameric character of the protein [16]. Understanding these complications and knowing the tremendous potential of a practically useful monomeric variant led to an intensive protein engineering effort. Disruption of the protein-protein contacts on the interface of DsRed yielded a dimeric intermediate [17]. A total of 33 substitutions resulted in monomeric RFP1 (mRFP1; λ_{ex} = 584 nm λ_{em} = 607 nm), the first-generation red monomeric protein [17]. However, it is important to note that loss of oligomerization greatly affected fluorescence. Further rounds of directed-evolution, to rescue photostability, quantum yield and extinction coefficient of mRFP1 yielded a new range of FPs, known as the mFruit series which introduced emission hues extending from green to far red [18, 19].

Using FPs as genetically encoded molecular tools has greatly facilitated both the visualization and perturbation of a variety of biochemical events in living cells [4]. Of continuing importance is the discovery, validation, and detailed investigation of protein

interaction in cells. It is important to note that the category of protein interactions is not only limited to protein-protein interactions, but encompasses a very wide range of biological activities including enzymatic activity, protein-DNA, protein-RNA, and protein interaction with small molecules. An intricate network of these protein interactions governs the everyday life and eventual death of the cell. Undeniably, tremendous efforts have been placed in the study of these protein dynamics which have yielded breakthroughs, such as the discovery of compartmentalized kinase signalling in cardiomyocytes [20]. However, the cell is a universe unto itself. The vast majority of protein interactions have yet to be studied or even discovered. This fact reveals the necessity of expanding and improving the range of FP-based biosensors to further accommodate new protein interaction studies, as each biosensor has its shortcomings and may not always be appropriate for the application in question. Currently, the two standard FP-based methods for detecting protein interactions in live cells are Förster Resonance Energy Transfer (FRET) and FP complementation.

1.1.1 FP-based FRET reporters

The seminal paper by Chalfie *et al.* presented the first demonstration of a FP as a molecular tool for labeling of the neurons of the worm *Caenorhabditis elegans* [2]. This triggered the explosion of interest in genetically encoded biosensors, first exemplified by the FRET-based strategy [21]. FP-based FRET reporters consist of two FPs with appropriate spectral overlap between the donor emission spectrum and acceptor absorbance spectrum [22]. The efficiency of nonradiative transfer of excited-state energy from the donor to the acceptor depends on the orientation and proximity between them, which ultimately affects the ratio of the donor fluorescent intensity to that of the acceptor's [22].

There are two general types of FP-based FRET biosensor designs, intramolecular and intermolecular [21]. The intramolecular design has a molecular recognition element sandwiched between the two FPs and the presence of an analyte will result in a binding event that changes the distance between the two FPs (**Figure 1.1**). Intramolecular FRET-based biosensor designs are useful in determining protein and analyte interactions as well as enzymatic activity and function. The intermolecular designs have the FPs separated into individual constructs. This design is best suited for the study of protein-protein interactions in which the interacting proteins-of-interest are each fused to each of the donor and acceptor FPs [21].

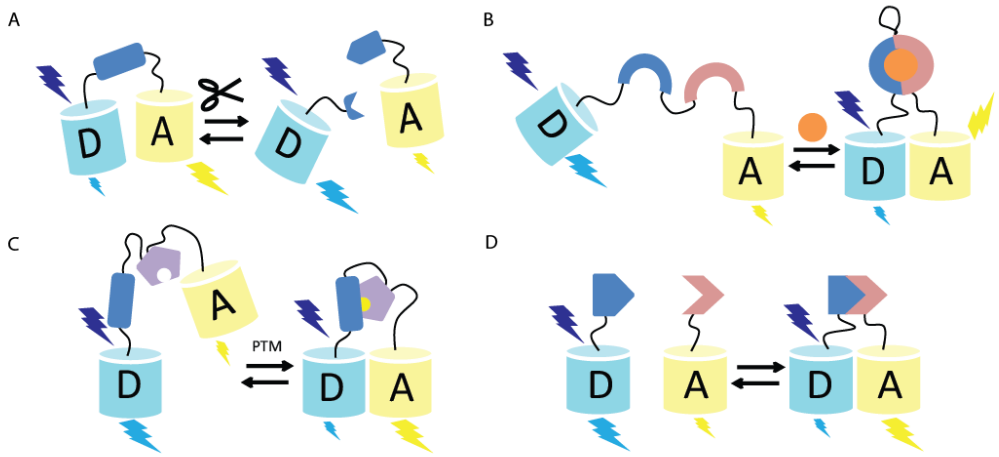


Figure 1.1. Intra- and intermolecular FP-based FRET reporter design strategies. (A) Protease reporters. (B) Intramolecular reporters of small molecules. (C) Post-translational modification (PTM) reporter. (D) Intermolecular reporter of protein-protein interaction. Donor FP is notated as D, and Acceptor FP as A.

Using the two general design strategies, a plethora of FP-based FRET reporters have been developed to detect numerous biological events, such as voltage sensing, analyte flux, enzymatic activity, protease activity, and protein-protein interactions [23], hence, making it the largest class of FP-based biosensors. It is also useful for high throughput screening applications [24]. FP-based FRET biosensors enable real-time monitoring of

cellular dynamics. The ratiometric readout has the advantages of being easily calibrated and providing reduced cell-to-cell variability compared to probes with intensiometric output signals, in which readout can be affected by out-of-focus movements [21, 23].

Although FP-based FRET biosensors are conceptually well suited for a wide variety of applications, there are some drawbacks that must be considered. FRET-based reporters are limited by the dynamic range and the sensitivity of the donor-acceptor pair [25, 26]. Moreover, implementation with multiple fluorescent probes poses a challenge since two spectral windows are already occupied to image one cellular event. For example, the FRET pair of cyan FP (CFP) plus yellow FP (YFP), popularly used due to its good spectral overlap, has two distinct excitation profiles and emission profiles in the blue and yellow region, thus, leaving only the red region of the spectrum open for sensing another activity of interest [23, 24]. In this case a maximum of two events can be imaged, which greatly undermines the potential and effectiveness of simultaneous imaging with multiple probes which can reveal spatiotemporal relationships between complex signalling pathways.

It is important to note that design of FRET fusion protein constructs based on either an intra- or intermolecular strategy cannot be generalized and must be treated on a case-by-case scenario. Linker length must be optimized in order to achieve optimal changes in FRET efficiency and allow sufficient freedom for the interacting domains to orient themselves [27]. An extra challenge associated with intermolecular FRET studies is that the two expressed protein chimeras in the cell can vary in stoichiometry.

In response to the drawbacks found in FP-based FRET reporters and demand for an expanded variety of reporters for detecting protein interactions, additional biosensing strategies have been developed. In this introduction, we will briefly discuss the protein

complementation assays for the study of protein interactions and focus specifically on FP complementation strategies.

1.2 General introduction on protein complementation assays

The introduction of the "classic" yeast two-hybrid (Y2H) system popularized and inspired a new generation of genetic screening and selection platforms aimed at investigation of protein interaction networks. Specifically, these platforms can be applied towards validation of hypothesized protein interactions, screening for inhibitors, dissecting protein function, and high-throughput screening of genomic libraries for protein-protein interactions (PPIs) [28]. Although robust and highly utile, Y2H assays typically cannot address the kinetics and spatiotemporal dynamic aspects of PPIs (**Figure 1.2A)** [29, 30]. Furthermore, the necessity of the protein complex to enter the nucleus to initiate the transcription event effectively prohibits its application to extracellular interactions, membrane-bound proteins, and proteins that exhibit a strong localization signal for exclusion from the nucleus. As a result, only a subset of the complete interactome can be detected via most two-hybrid assays [29, 31]. This is where protein fragment reconstitution assays (also known as protein complementation assays (PCAs)) are useful for the investigation of protein interaction networks.

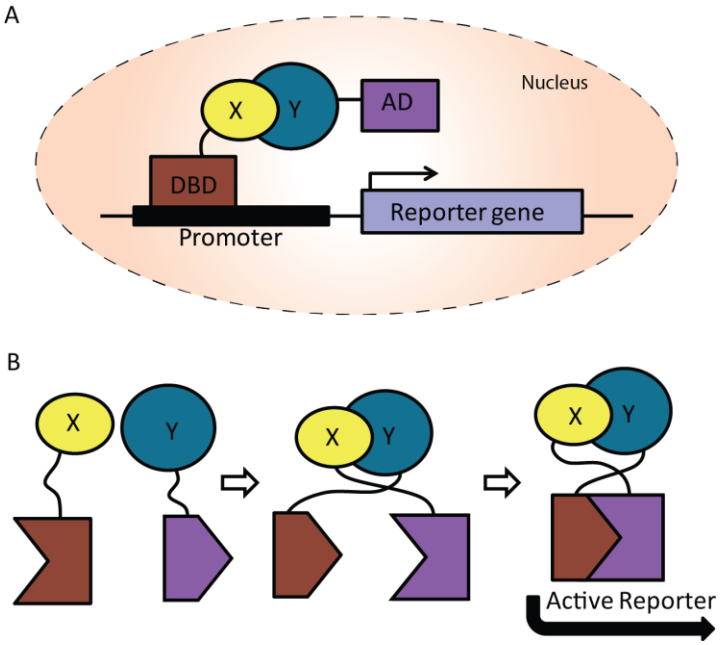


Figure 1.2. Schematic summary of the yeast-two hybrid (Y2H) assay and protein complementation assay (PCA). (A) Y2H system, Protein X is tethered to a DNA-binding domain (DBD) and Y to a transcription activating domain (AD). Co-localization of proteins X and Y due to association between the two, brings the AD in proximity to interact with the RNA polymerase II complex to initiate transcription of the reporter gene such as *lac*Z. (B) PCA, interaction of proteins X and Y leads to reconstitution of the reporter.

Protein fragment complementation-based assays operate on a straightforward principle. Similar to the Y2H, proteins-of-interest ('X' and 'Y' or more often designated as 'bait' and prey') are tethered respectively to the two fragments of the effector protein which have been "split" or separated into two fragments (**Figure 1.2B**) [32]. Reassembly of the fragments, hence rescue of reporter activity, occurs upon interaction of the bait and prey. Reporter fragments have been engineered not to spontaneously associate, since this would elicit false-positive signals [32]. When developing a PCA system, the fused proteins-of-interest must be given enough freedom to orient and associate as hindrance may lead to false-negative results. It is also important to note that formation of reporter protein is irreversible in the large majority of cases [33]. This effectively traps transient oligomeric

protein interactions which are of particular interest [30]. However, these irreversible complexes may perturb protein-complex localization or turnover, though this is not generally observed [32]. If reconstitution occurs, the reporter readout will be provided as colony survival, a fluorescent signal, luminescence, or a change in colour [32]. Many reporter proteins have been manipulated or "split" for PCA purposes. Of these, ubiquitin [30], TEM1 β-lactamase [34], murine dihydrofolate reductase (mDHFR) [35], luciferases [36] and notably, FPs [37] are the more prominent reporters used.

1.2.1 FP complementation strategies

FP complementation techniques are one of the most widespread PCAs used in research today. They are highly transferable between a wide variety of organisms such as *Escherichia coli*, *Saccharomyces cerevisiae*, *Caenorhabditis elegans*, *Drosophila melanogaster,* plants, and mammalian cells [32]. Some general concerns of FP PCA include photobleaching, phototoxicity to the cells, and high autofluorescence during *in vitro* and *in vivo* imaging. Despite this, the ability of FP PCA to detect weak interactions with subcellular resolution offers information on the precise localization of cellular components and rapid dynamics of interacting proteins in their endogenous environment [29, 38].This important advantage greatly overshadows its shortcomings. Another advantage of FP complementation, relative to techniques such as split mDHFR, split luciferase and split β-lactamase, is the lack of exogenous reagents required to obtain an output signal [34-36]. Exogenous agents can be unstable and introduce toxicity to the host organism.

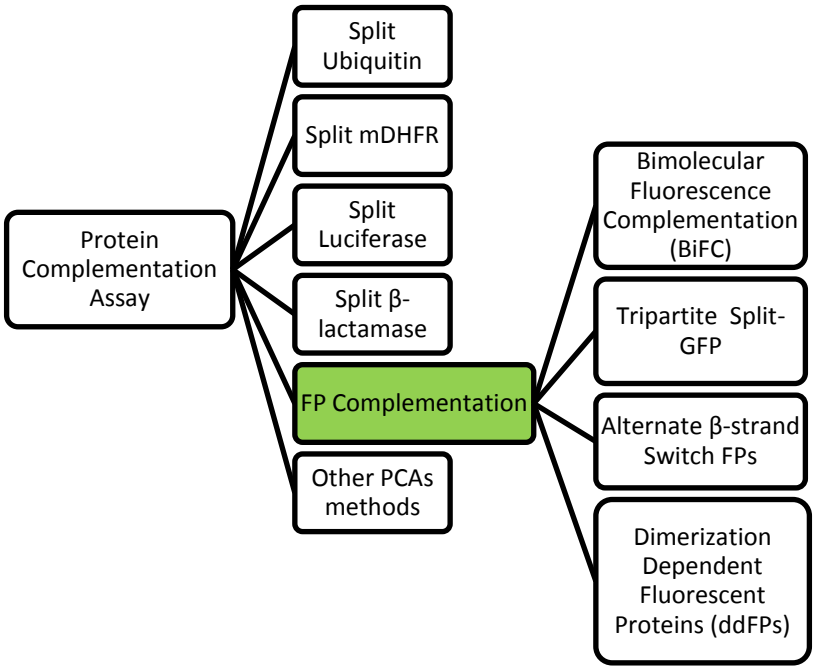


Figure 1.3. Flow chart presenting the categorization of the different protein complementation assays. Note that FP complementation encompasses not only BiFC but also, tripartite split GFP, alternate β-strand switch FPs and ddFPs.

To date, FP complementation techniques have been applied to reveal the subcellular localization of PPIs [39], validate Y2H results [29] and even in the identification of off-target effects of drugs [40]. The popularity and applicability of FP complementation has extended beyond the traditional bimolecular fluorescence complementation (BiFC). Many FP complementation techniques have been developed and cannot be categorized as BiFC, due to fundamental differences in their design and implementation. Here, we will discuss the FP complementation technologies that are currently available and their biological applications. There will be an extensive discussion on BiFC technology as it has been around the longest.

1.2.1.1 Traditional approaches to FP complementation: Bimolecular Fluorescence Complementation (BiFC)

Less than two decades ago, Ghosh *et al.* successfully dissected the 238 amino acid *Aequorea victoria* GFP (avGFP) into two fragments and induced its reassembly via the noncovalent heterodimerization of antiparallel leucine zippers [37]. This initial demonstration opened up a new biosensor platform for FPs as a tool for PPI detection studies. Briefly, the construction of an FP for BiFC is as follows, the FP is genetically divided to yield two fragments. These "split" sites are often introduced at non-conserved amino acid residues within loops [41]. The newly formed amino (N)-terminal fragment and carboxyl (C)-terminal fragments are then each fused to their protein-of-interest, known as the bait and prey. Upon PPI interaction, the unfolded FP fragments reassemble, undergo the autogenic steps of posttranslational modification necessary to form the chromophore, and thereby create the functional reporter that provides a fluorescent signal. After the bait and prey interacts, there is a delay of many minutes before the output signal is detected due to the slow maturation rate of the chromophore [29]. As a result, real-time interaction kinetics for BiFC cannot be detected [41]. Once the reconstituted FP complex is formed, the process is irreversible. Over time, the observed signal intensity increases due to the accumulation of the mature FP which leads to difficulty in distinguishing fluorescence intensity arising from non-specific assembly of the FP fragments versus specific interaction of the protein-of-interests [42]. This is of important consideration when using BiFC technology. Since split avGFP, the spectrum of available BiFC reporters has greatly expanded. The scientific community has embraced this development and they have stretched the limits of BiFC discoveries beyond the imaging of protein interactions in live cells.

1.2.1.1.1 The expansion of the BiFC colour palette and its versatility

1.2.1.1.1.1 YFP and CFP variants

One of the first applications of BiFC was for the investigation of interactions between the bZIP and Rel family of transcription factors in COS-1 cells using a split YFP (λ_{ex} = 500 nm; λ_{em} = 535 nm) [41]. Following this, BiFC multicolour imaging was inaugurated with the split CFP (λ_{ex} = 436 nm, λ_{em} = 470 nm) [39]. Using both split YFP and split CFP two physiological events can be imaged simultaneously without spectral overlap. With regards to BiFC multicolour imaging, its true utility is in the imaging of proteins with several binding partners. The design is as follows, the C-terminal fragment of YFP and CFP are each tethered to their respective proteins-of-interest, we designate those as 'Y' and 'Z', and the N-terminal fragment of the split FP is fused to a protein, designated as 'X' that has the potential to bind to both Y and Z. Importantly, the C-terminal fragment determines the spectral properties of the complex, binding event of X to Y leads to a yellow fluorescence while binding of X to Z leads to cyan [39]. Multicolour BiFC was conducted using this design to study the ability of seven G protein γ subunits binding to the β_1 subunit [43]. Split YFP suffers from sensitivity of chromophore maturation at physiological conditions [44]. This prompted the development of split Venus and split Citrine, which are two improved YFP variants [45, 46]. Use of these improved YFPs meant that the previous inconvenience of preincubating cells at a lower temperature could be avoided. At the same time, Cerulean, a variant with a 2.5-fold increase in fluorescence intensity to its predecessor ECFP, was split, hence, increasing the advantage of utilizing BiFC technology. Pursuing along the lines of the widespread CFP-YFP FRET pair, Shyu *et al.* introduced BiFC-FRET for the study of ternary complex formation of Fos-Jun heterodimers and the NF-κB subunit,

p65 which may be a factor in the life-death determination of the cell [47]. Similar to FRET, BiFC-FRET has a donor, Cerulean, and acceptor, split Venus. Shyu *et al.* attached p65 to Cerulean, and the basic region leucine zipper domain of c-Jun (bJun) and c-Fos (bFos) to the N-terminal and C-terminal fragments of split Venus. Association of bJun-bFos to each other formed the complex known as activator protein 1 (AP-1) and reconstituted the full Venus to allow FRET with Cerulean when AP-1 interacted with p65. Interestingly, this interaction revealed the enhanced binding in the presence of p65 of AP-1 to the AP-1 consensus sequence 12-*O*-tetradecanoylphorbol-13-acetate (TPA) response element which induces target-gene transcription. Recently, BiFC-FRET has assisted in the first known visualization of subunit interactions and subcellular localization of individual Protein Phosphatase 2A (PP2A) holoenzymes in live cells [48].

1.2.1.1.1.2 GFP variants

It is interesting to note that although avGFP was the first FP to be split, its use under physiological conditions is not as widespread as its Venus counterpart [49, 50]. In fact, BiFC applications for avGFP have not been successfully applied in mammalian cells due largely to poor solubility [39]. Notably, tagging and tracking of proteins with just wild-type GFP in cells is not as simple and ideal as postulated because of challenges related to solubility of the protein-of-interest and aggregation [51, 52]. In an effort to overcome these problems, a split GFP was developed from superfolder GFP (sfGFP) (**Figure 1.4**) [53].

Compared to previous BiFC reporters focussed on determining PPI, the split sfGFP is a reporter for protein folding, expression, and promoter activity, while also being useful for protein tagging and localization experiments [54]. To use this system to report on protein folding, β-strand 11 of the GFP (GFP11) is fused to the protein-of-interest and

optimized GFP β-strands 1-10 (GFP1-10 OPT) is expressed separately. If the protein-of-interest is properly folded and soluble, then GFP11 will be accessible and spontaneous reconstitution with GFP1-10 OPT will occur to form the functional sfGFP. This split sfGFP system has been used to explore cell to cell contact and define connectivity of synapses in the nervous system of model organisms such as *C. elegans*. This particular application is known as GFP reconstitution across synaptic partners (GRASP) [55].

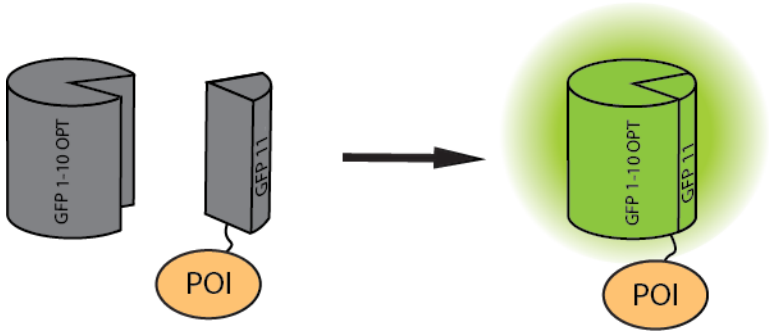


Figure 1.4. Scheme of split sfGFP. As long as the protein-of-interest (POI) is folded properly and soluble, GFP11 will be available to reconstitute to GFP 1-10 OPT to rescue green fluorescence

As a result of the folding issues related to split avGFP fragments, efforts to expand upon the green hued BiFC were redirected to alternative GFP-like variants. Filling up this demand was a split variant of monomeric Kusabira-Green (mKG, λ_{ex} = 494 nm; λ_{em} = 507 nm), an evolved variant of its oligomeric predecessor Kusabira-Orange derived from the stony coral *Fungia concinna* [56, 57]. Split mKG has been used to study the intra- and intermolecular interactions of $p67^{phox}$, $p47^{phox}$, and $p40^{phox}$ during FcγR-mediated phagocytosis [56]. In addition, the mKG system has been applied to study the real time interactions of heat shock protein 47 (Hsp47) and collagen triple helix in live cells [58]. Other green variants developed for BiFC applications include a split version of Dronpa (λ_{ex} = 503nm; λ_{em} = 517nm, $\lambda_{on/off}$ = 488/405). Dronpa is a GFP-like reversibly photoswitchable

FP in which its “on” and “off” fluorescence state can be light controlled [59]. Being photoswitchable is an advantage for repeated measurements of protein behaviour in live cells and useful for superresolution fluorescence imaging [60]. Split Dronpa has been applied to analyze the interaction of hHus1 and hRad1, proteins involved in the cellular response to DNA damage and functions [61].

1.2.1.1.1.3 RFP variants

With the increasing acceptance of BiFC technology in the beginning of the millennium and the monomerization of DsRed to mRFP1 [17], the development of split red FPs was a natural next step. The availability of split red FPs would greatly complement other spectral variants for multicolour BiFC imaging. Furthermore, when imaging RFP variants, there is reduced autofluorescence, lower phototoxicity, lower light scattering, due to the use of longer wavelength light [12]. One of the first red variants to be split was mRFP1-Q66T (λ_{ex} = 549 nm; λ_{em} = 570 nm). The introduction of a Gln66Thr point mutation to mRFP1 conferred faster maturation, greater photostability, and photobleaching properties [62]. Its applications have mainly been in plants where it has been used to reveal the interaction matrix of ten proteins from the plum pox virus in *N. benthamiana* leaf epidermal cells [63]. mCherry (λ_{ex} = 587 nm; λ_{em} = 610 nm) was the successor to mRFP1 and benefited from faster maturation, higher photostability, excellent pH resistance and most importantly, longer wavelength [18]. Accordingly, the split mCherry variant provided a number of improvements relative to split mRFP1-Q66T for these reasons [64]. Beyond the typical PPI studies, the split mCherry has been used as a trimolecular fluorescence complementation system (TriFC) (see section 1.2.1.1.2) [65]. In addition, fusing the two fragments of split mCherry to zinc finger (ZF) DNA-binding domains, ZF Zif268 and ZF

PBSII, a selective double-stranded DNA detection system termed sequence-enabled reassembly (SEER) was created [66]. Swapping of the ZF DNA-binding domains to a methyl cytosine-guanine (mCpG) dinucleotide binding domain (MBD2) created a reporter selective for methylation at CpG sites [66]. TagRFP (λ_{ex} = 555 nm; λ_{em} = 584 nm), a monomeric red FP with 2.8-fold brighter fluorescence than mCherry has also been manipulated for BiFC applications [67, 68]. Another current use of BiFC is to couple it with superresolution imaging techniques such as photoactivated localization microscopy (PALM). The BiFC-PALM technique has been used for imaging PPIs at nanometer resolution (see section 1.2.1.1.4) [69, 70]. Thus far, the only photoswitchable protein to be split is Dronpa. However, split Dronpa has not yet been applied for PALM due to the low number of photons emitted per activation cycle [71]. A photoactivatable mCherry (PAmCherry1, λ_{ex} = 564 nm; λ_{em} = 595 nm) has been split for superresolution imaging of Ras/Raf interactions [71].

Recent advances in the field of BiFC have concentrated on the creation of split far red FP variants which may enable deeper imaging into tissue. Deep tissue imaging in conjunction with BiFC would facilitate efforts to visualize PPIs in living animals. The far-red monomeric FP, mKate (λ_{ex} = 588 nm; λ_{em} = 625 nm) has fast maturation and high photostability but suffers from low brightness [13]. Hence, it was evolved to produce a 2-fold brighter variant, mLumin (λ_{ex} = 587 nm; λ_{em} = 621 nm), which was subsequently split and used to demonstrate the direct interaction of epidermal growth factor receptor (EGFR) with signal transducers and activators of transcription (STAT5) in COS-7 cells [72]. Compared to the previously mentioned split RFP systems, only split mLumin can mature efficiently under physiological conditions of 37 °C [13, 72]. However, the emission

wavelengths of less than 650 nm seen in both variants leave much to be desired for *in vivo* applications. The near infrared (NIR) optical "window" is between 650 nm and 900 nm where tissue is most transparent due to lower absorption of hemoglobin, melanin and water [12].

Additional efforts to create a far red FP-based probe for BiFC applications resulted in development of split mNeptune (λ_{ex} = 600 nm; λ_{em} = 650 nm) [73]. The excitation peak of mNeptune at 600 nm is ideal as excitation light below 600 nm is absorbed by hemoglobin which hinders intravital imaging in live mice. The emission peak of 650 nm just falls into the NIR optical window. The split mNeptune has also been applied to TriFC (see section 1.2.1.1.2) [74].

1.2.1.1.1.4 Bacterial phytochrome photoreceptors for near infrared BiFC

The discovery of GFP-like FPs in reef-building corals was truly a breakthrough that expanded the colour palette of FPs into the red region. Similarly, the latest efforts to expand the colour palette to the NIR region are focussed on bacterial phytochrome photoreceptors (BphPs) found in photosynthetic and non-photosynthetic bacteria [75]. The absorbance and fluorescence of BphPs goes beyond the limited 660/680 nm excitation/emission of current FPs and red FPs developed from plant and cyanobacterial phytochromes [75]. These NIR fluorophores allow better penetration of excitation light for deep-tissue and whole body imaging which is invaluable in stem cell biology, gene therapy, monitoring bacterial and viral infections, monitoring tumour growth, and visualization of metastasis and angiogenesis in cancer [76, 77]. BphPs are comprised of a photosensory core module (PCM) and an output effector module. The PCM includes a Per-ARNT-Sim repeats (PAS) domain, cGMP phosphodiesterase/adenylate cyclase/Fh1A transcriptional activator (GAF) domain

and a phytochromes-specific (PHY) domain which are all connected by α-helix linkers [75] . Biliverdin IXα (BV), a product of heme catabolism is ubiquitously present in mammalian tissue [76] and binds to the GAF domain to create the fluorescent adduct. BphPs from *Deinococcus radiodurans*, an extremophilic bacterium, has been evolved to create the first monomeric NIR FP, IFP1.4 (λ_{ex} = 684 nm; λ_{em} = 708 nm) [76]. IFP1.4 retains only the chromophore-binding domain, PAS and GAF, as they are sufficient for the covalent incorporation of BV.

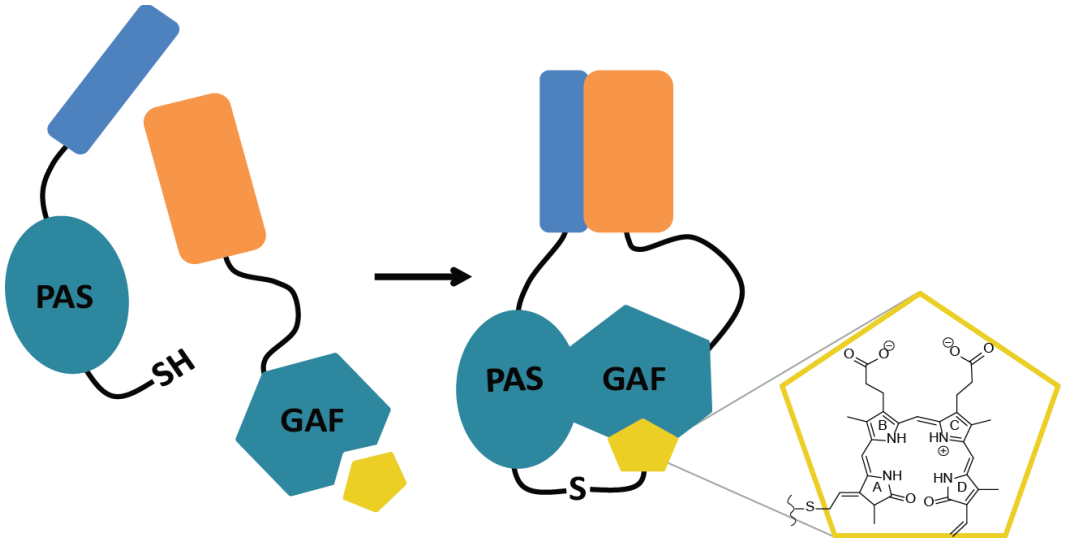


Figure 1.5. Design of iSplit. Interaction of the protein partners (are shown in blue and orange, respectively) causes PAS and GAF to come into close proximity so that GAF with biliverdin IXα bound can interact with the catalytic cysteine on PAS to reform the full iRFP.

Using a similar design, Filonov *et al.* developed infrared FP (iRFP, (λ_{ex} = 690 nm; λ_{em} = 713 nm) from the photosynthetic bacterium *Rhodopseudomonas palustris* with higher molecular brightness, intracellular stability, affinity for BV, and protein expression level compared to IFP1.4 [77]. The higher affinity to BV is the result of randomly mutating GAF, this GAF variant is known as GAF_m. Imaging of iRFP in mouse phantoms showed improved signal-to-background ratio relative to mNeptune [77]. Filonov *et al.* then split

iRFP at the unstructured loop connecting PAS and GAF domain to create iSplit for BiFC applications (**Figure 1.5**) [78]. MTLn3, highly invasive breast adenocarcinoma cell line, was transfected with plasmids for expression of iSplit fragments fused to their respective partners, PAS-FRB and FKBP-GAF_m. These cells were injected into mice that were then injected with rapamycin. Rapamycin-induced association of FRB and FKBP induced strong NIR fluorescence thus demonstrating iSplit's utility for *in vivo* imaging applications to study PPIs [78]. iSplit has been reported to have a tendency to dimerize spontaneously [79] . Another split NIR probe based on iRFP has been reported [80]. It has been applied *in vitro* to visualize lens epithelium-derived growth factor (LEDGF/p75) and HIV-1 integrase interaction, critical for HIV-1 integration and replication, and used to screen for antiviral drug inhibitors for this particular interaction [80]. Recently, a BiFC system has also been evolved from IFP1.4 and subsequently named IFP PCA. Most importantly, this system was reported to be reversible. It has been tested *in vitro* in HeLa cells to visualize SHC1-GRB2 temporal modulation by epidermal growth factor (EGF) receptor. Despite being reversible, which allows quantitative studies on PPI spatiotemporal dynamics, this split version exhibits a lower quantum yield, lower brightness and still requires exogenous BV to boost its fluorescence, just like its parent IFP1.4. *In vivo* imaging applications of IFP PCA have yet to be tested [79].

1.2.1.1.2 BiFC in RNA-protein interaction

Recent achievements in the area of RNA imaging in live cells have largely been led by FPs due to their genetically-encoded properties and the fact that no exogenous reagents are necessary to induce fluorescence. RNA is more easily accessible than DNA and, by studying genes at the RNA level, many genetic and infectious diseases can be targeted.

Single RNA detection for the study of their posttranscriptional dynamics, kinetics and distribution in live cell is of interest as it leads to further understanding of the complex process of gene expression. For RNA-protein detection it is vital to achieve high specificity for the RNA sequence of interest and this fundamentally relates to the specificity of the RNA binding protein (RBP).

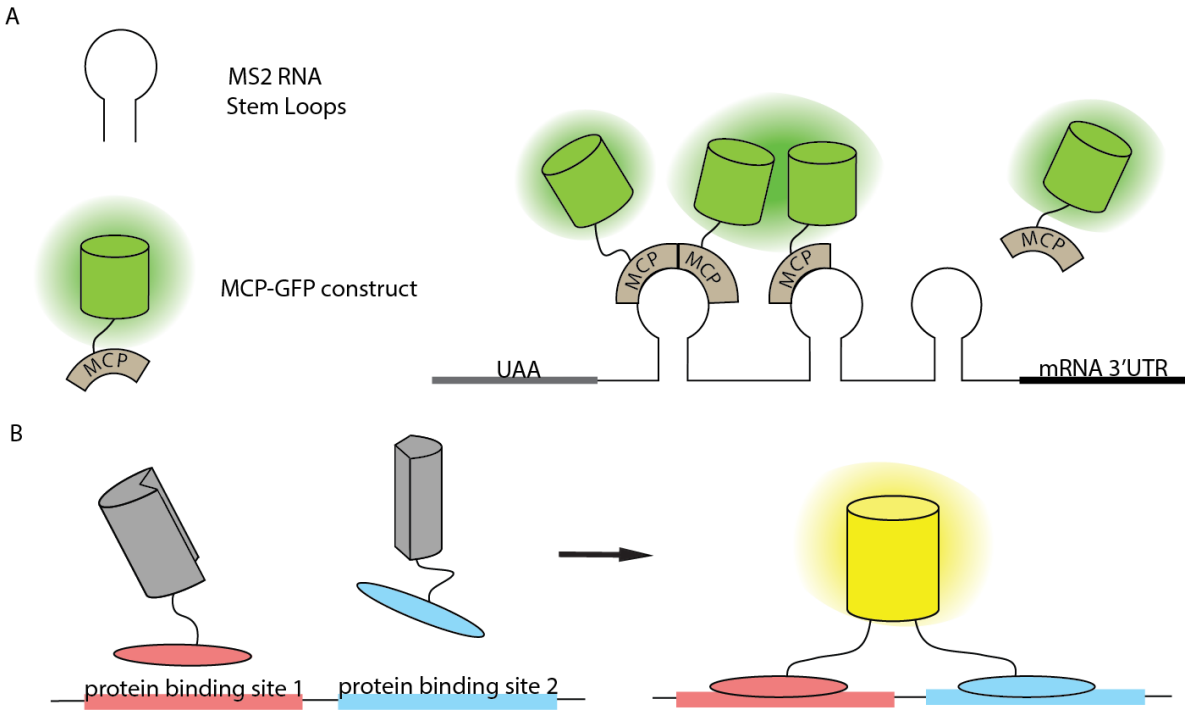


Figure 1.6. Schematic of live-cell visualization of RNA using the bacteriophage MS2 coat protein (MCP)-GFP construct and trimolecular fluorescence complementation (TriFC). (A) A MCP-GFP construct is expressed in cells along with the mRNA of interest with a tandem MS2 binding site comprised of stem loops on the 3' UTR end. A stem loop can bind two MCPs. This may create double the signal intensity than expected. (B) A TriFC system where two RNA binding proteins (RBPs) (red and blue, respectively) are tethered to their respective split FP halves. Association of both RBPs to their respective target RNA sequence allows the mature FP to form. This significantly reduces background. The RNA in this case can be exogenously introduced like the MS2 system or if both RBPs are specific for a particular RNA then no exogenous introduction is needed. However, signal will be comparatively reduced.

The MS2 system pioneered by Singer and co-workers revolutionized the study of RNA in living cells and intact animals at the molecular level **(Figure 1.6A)** [81]. However,

this system has been fraught with problems such as non-specific background signal from free bacteriophage MS2 coat protein (MCP)-FP constructs [81]. To reduce background and increase specificity, the split FP was exploited to create a three component approach known as trimolecular fluorescence complementation (TriFC). Originally developed by Rackham and Brown, this strategy involves a fragment of the split Venus attached to the MCP while the other fragment will be attached to the RBP of interest (**Figure 1.6B**) [65]. The target RNA of interest must encode the MS2 binding site (MBS) at its 3' untranslated region (UTR). Only when the MBS and RNA binding site of interest is bound to the target RNA is a fluorescent signal generated. As such, TriFC allows improved imaging of RNA due to decreased background fluorescence. Nonetheless, transfection of an exogenous RNA is still required for this strategy to work. This may lead to perturbation of cellular dynamics in the host organism. Further perturbation with respect to RNA kinetics can occur since RNA in complexes with RBPs display higher stability [82].

Using the TriFC system, Rackham and co-workers were able to gain insight into the interactions of two RBPs, fragile X mental retardation protein (FMRP) and insulin-like growth factor mRNA-binding protein 1 (IMP1), and their possible link in the mRNA localization process [65]. TriFC has also been extended to GFP-like variants such as split mCherry [64]. In one application, the mCherry FP halves were used to examine the interactions of influenza A viral mRNAs to adaptor proteins such as Aly/REF, UAP56, hnRNP A1, or 9G8 in the cellular Tip-associated protein (TAP) export pathway for viral mRNA nuclear export [83]. Split mNeptune has also been applied for TriFC applications to monitor human polypyrimidine-tract-binding protein (PTB) and HIV-1 mRNA elements in live mice [74]. PTB may play a role in HIV-1 mRNA processing. It is important to note

that the signal generated from a reconstituted split FP binding will be less than 10% of the fluorescence of full-length FP [84]. Another TriFC system utilized PP7 bacteriophage coat protein (PCP) and MCP as the RBP to recognize a tandem array of MBS and PP7 binding sites on an exogenous RNA [85].

In another example, eukaryotic initiation factor 4A (e1F4A), an ATP-dependent RNA helicase important in protein synthesis initiation, was split into two fragments and fused to the fragments of split EGFP [86]. Using this approach, it was possible to visualize the localization of ribosomes in *E. coli* at the cell poles with reduced background signal compared to using full-length EGFP. This demonstrates the ability for correct RNA recognition without using MCP as long as the RBP is specific enough. With this system, real-time visualization of RNA spatiotemporal patterns and quantification of the early stages of RNA synthesis was enabled [87]. In order to reduce the interference on bacterial or eukaryotic cell metabolism, a smaller RNA detection system with respect to the previously mentioned e1F4A, was created by the same research group [88]. A split EGFP was fused to the RNA-binding viral peptides, HIV-1 Rex peptide and bacteriophage λN peptide, which binds to their respective cognate aptamers that are spaced close enough to each other to trigger reassembly of the split EGFP.

Some RNA sequences do not have a specific RBP. In order to expand the diversity of RNA that can be recognized in a similar TriFC design, mutated variants of the pumilio homology domain (PUM-HD) from human PUMILIO1 were each attached to split EGFP fragments [89]. PUM-HD is a RNA-binding domain comprised of eight sequence repeats with each repeat recognizing a specific base [90]. Site-directed mutagenesis can alter the RNA sequence specificity in a rational manner. Using this design, Ozawa *et al.* were able to

tailor the PUM-HD to recognize the two specific RNA sequences flanking the NADH dehydrogenase subunit 6 (ND6) in mitochondrial RNA (mtRNA) [89]. In doing so, ND6 mtRNA spatial and temporal dynamics under oxidative conditions in living mammalian cells were observed. Notably this method requires no exogenous RNA, thus reducing the probability of false localization of the RNA of interest [91]. It is important to keep in mind that reconstitution of the mature FP can take up to thirty minutes which hinders analysis of mRNA dynamics [92].

1.2.1.1.3 BiFC for genome-wide PPI screening

The Y2H assay has already been available for over a quarter of a century. Since the first report of this system, it has and continues to aid scientists in the investigation of proteomic interaction networks [93, 94]. The limitations of the Y2H in detecting integral membrane protein interactions have been rectified with the development of the membrane yeast two-hybrid (MYTH) assay [95-97]. This assay incorporates the fragments of the split-ubiquitin system, which are respectively fused to the potentially interacting integral membrane proteins-of-interest. Interaction initiates recombination of the split ubiquitin which then causes release and translocation of a transcription factor to the nucleus to generate a fluorescence or luminescence signal. Recently the MYTH system been transplanted to mammalian cells to create mammalian-membrane two-hybrid (MaMTH) assay [31]. Moreover, a genome wide screen in yeast using a DHFR survival-selection PCA revealed 2770 interactions among 1124 endogenously expressed proteins of which approximately 80% are previously unknown interactions [33]. This is the first noted use of a PCA for conducting such a large scale analysis of PPI identification. Despite, the wealth

of tools that are available, a large subset of the proteomic interaction networks remains to be explored, particularly in mammalian cells.

In many cases, BiFC is used as an orthogonal assay to verify the PPI identified by more-established assays such as Y2H or MYTH when dealing with integral membrane proteins [98, 99]. However, the advantage of BiFC over other available PCAs lies in its fluorescence output signal and as such, the subcellular localization of the PPI can be identified [29]. One of the main challenges in implementing BiFC or even other PCAs is the development of cDNA or genome-wide prey libraries in various cell types [49]. In yeast, a high throughput screen for SUMO target proteins was completed using BiFC [100]. This was a huge undertaking as a *Saccharomyces cerevisiae* fusion library was created so that 95% of all open reading frames (ORF) listed in the *Saccharomyces* genome database (as of April 2001) was expressed with its appropriate FP fragment. Using BiFC for analysis of this interactome revealed 224 previously unknown SUMO target proteins and their subcellular localization. The authors attribute these findings to the BiFC method, unlike previous efforts, the PPI are measured in the natural cellular context [100].

The interaction-induced fluorescence of BiFC lends itself to a powerful combination with flow cytometry. Previous studies have demonstrated the utility of BiFC with flow cytometry in bacteria for the detection of the SH3 domain from c-Abl tyrosine kinase to natural and designed targets [101]. A genome-wide search for telomere regulators was one of the first applications of BiFC combined with flow cytometry in mammalian cells [102]. In this study, a large retroviral array library of 12,212 ORFS tagged to the C-terminal YFP fragment was established. The retroviruses were then created and transfected into bait cells which were each harbouring one of the six core telomeric proteins fused to the N-terminal

fragment of YFP. Of importance in this study was the development of the data analysis platform, CytoArray, for management of the large amount of flow cytometry data from the arrayed screen. Through the stringent selection procedures imposed by this program, followed by validation of bait-prey interaction by co-precipitation, 300 proteins were identified to have association with the six telomeric proteins [102].

High-throughput screening with BiFC for drug discovery has also been explored. Using a simple and robust adenovirus high-throughput system (AdHTS) [103], G protein-coupled receptors (GPCRs) were tagged with the C-terminal Venus fragment for an adenovirus-based β-arrestin BiFC assay to monitor GPCR activation in response to agonists and antagonists [104]. Through use of an adenovirus gene delivery system a more uniform BiFC from a large number of cells was achieved compared to plasmid-mediated BiFC [104]. BiFC was applied to screen for synthetic drug inhibitors targeting the HIV-1 Nef dimerization interface [105]. By another BiFC assay, Eugenol, the major active compound of *Syzgium aromaticum* L., was found to inhibit Beclin1-Bcl2 heterodimer dissociation [106]. This inhibitor caused the arrest of autophagy which led to the impairment of influenza A virus replication [106].

1.2.1.1.4 BiFC for single-molecule localization-based superresolution imaging

It has almost been a decade since the introduction of the closely related single-molecule localization-based superresolution imaging methods which includes stochastic optical reconstruction microscopy (STORM) [107], photoactivated localization microscopy (PALM) [108], fluorescent PALM (fPALM) [109]. These wide-field techniques allow imaging of individual fluorophores within a sample at high precision. Accordingly, ultra-structural imaging, tracking and study on the dynamics of macromolecular assemblies in

live cells can be achieved [107, 110, 111]. Diffraction-limiting images are overcome by stochastically activating different subset of fluorophores at different time points. In this way, the position of otherwise spatially overlapping fluorophores can be resolved. Over a series of imaging cycles, the coordinates of each fluorophores is mapped and recorded to reconstruct an overall image [112]. Since these novel optical techniques have appeared, the development of photoactivatable, photoconvertible, and photoswitchable FPs with suitable characteristics for live-cell superresolution imaging have been playing catch up in order for these microscopy techniques to be practical.

FP properties that are desired or required include: a high intensity contrast between the "on" and "off" states, high photon budget which affects localization precision, high signalling efficiency, and a monomeric structure [108, 110, 113]. During an evaluation of the twelve most commonly used photoactivatable FPs (PAFP), Wang *et al.* found that none of the twelve exhibited optimal scores in all four of the preferred properties [110]. This brings us as to why BiFC has not been widely applied in complementation with STORM/PALM/fPALM. Since the advent of the first useful PAFP, photoactivatable GFP (PA-GFP) (λ_{ac} = 413 nm; λ_{ex} = 488 nm; λ_{em} = 517 nm), there has been continual development to this field and a variety of suitable proteins are now available. mEos3.2 (λ_{ac} = 405 nm; λ_{ex} = 561 nm; λ_{em} = 580 nm) displayed the best photon budgets among current PAFPs, excellent on-off ratio, and no detectable dimerization tendencies [110]. Recently, mEos3.2 was modified for BiFC applications such that high-density PPIs in the bacterial cell can be imaged at the nanometer resolution [69]. This system was used to visualize the interactions between EF-Tu and Mre-B with high specificity and spatial-temporal resolution in *E. coli*. EF-Tu is a prokaryotic translation elongation factor discovered to

interact with MreB, an actin cytoskeleton homologue, in a process that is relevant to *Bacillus subtilis* cell wall synthesis [114]. Both proteins are present in high abundance in *E. coli*. By the BiFC-PALM technique, EFTu-MreB-PPI heterogeneous dynamics and distribution could be observed [69]. In another study, BiFC-PALM was successfully employed in eukaryotic cells to "track" growing microtubule plus ends [70]. A split photoactivatable FP was designed from PA-GFP and retained the spectral characteristics of its parent and exhibited low spontaneous complementation. The fragments were each fused to end-binding 1 (EB1) protein which is known to form homodimers [115, 116]. EB1 protein is part of the family of EB proteins which interacts with growing microtubule plus ends and recruits cell polarity and signalling molecules to microtubule tips [117]. Homodimerization of EB1 indicated reassembly of PA-GFP that enabled visualization of the dynamics of the microtubule plus ends at the leading and tailing edge of the cell and cell body [70]. In the same time frame of the introduction of split PA-GFP for PALM, split PA-mCherry was developed [71]. KRas and Ras-binding domain, CRaf, are proteins involved in signal transduction and often highly deregulated in oncogenesis. Both were individually fused to the N- and C-terminal fragments of split PA-mCherry. From the PALM imaging experiments, clustering of multiple Ras/Raf complexes were observed to be heterogeneously distributed on the cell membrane of U2OS cells [71].

Although both BiFC and PALM have been around for more than a decade, only recently has the combination of the two been explored. With the debut of both split PA-GFP and PA-mCherry, dual colour BiFC-PALM applications will become a reality in the near future. Expanding and improvement of split PAFPs for greater photostability and

brightness with minimized blinking will surely extend the practicality of BiFC-PALM for greater applications in protein interaction imaging [70].

1.2.1.1.5 BiFC applications for labelling protein dynamics

Not only do researchers want to analyze specific protein localization in cells, but they also want to obtain detailed temporal information related to the life cycle of a protein from synthesis to turnover to degradation. Of course, there have been assays made for these purposes such as the mass spectrometry-based approaches of stable isotope labelling with amino acids in cell culture (SILAC), fluorescent non-canonical amino acid tagging (FUNCAT), and bio-orthogonal non-canonical amino acid tagging (BONCAT) [118-120]. These methods are powerful tools for the quantitative analysis of proteomic changes, identification of newly synthesized proteins, and even the visualization of newly synthesized protein localizations. However, this is all undertaken at the global level of the cell. Specific labelling of the protein-of-interest remains a challenge. Proteins-of-interest identified in screens from the above strategies can be used in conjunction with FP-based BiFC.

TimeSTAMP is a novel genetically-encoded technique of which a drug is used to control the maturation of the FP domain [121] in a way that mimics classical pulse-chase experiments [122]. In this design, Hepatitis C NS3 protease is sandwiched between two of its cognate protease cleavage sites, which is in turn placed between the N-terminal and C-terminal fragment of split Venus (**Figure 1.7**). The protein-of-interest is located at the N-terminus of the final construct. In the "pulse" component, addition of BILN-2061 inhibits NS3 thus giving time for the split fragments to form the irreversible, mature Venus which effectively tags the protein-of-interest. If no NS3 inhibitor BILN-2061is continually added,

then the active NS3 will cause protease-mediated fragmentation at both its adjacent cleavage sites and therefore no FP reconstitution can result, the "chase" component. A version of this system that incorporates a split mKO2 orange FP has also been created for simultaneous imaging of two proteins [121]. This method was used to elucidate the stimulus-dependent translation and local synaptic accumulation of PSD95 in neurons. MiniSOG, a genetically encoded singlet oxygen generator was placed into the TimeSTAMP construct to mediate photo-oxidation of diaminobenzidine (DAB) for ultrastructural imaging by electron microscopy. By doing so PSD95 labelling at immature versus mature synapses was made possible [121].

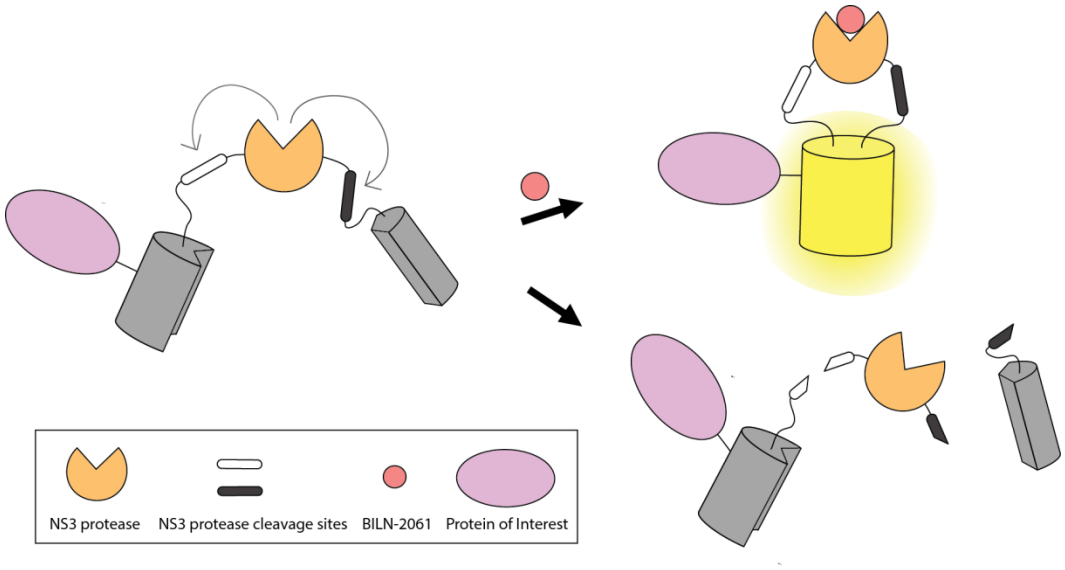


Figure 1.7. Principle of TimeSTAMP for tracking protein fate. The NS3 protease will cleave its protease sites if no inhibitor BILN-2061 has been added leading to a loss of fluorescence and N-end rule mediate degradation of the NS3 protease. Addition of inhibitor blocks NS3 protease activity thus, fluorescence is rescued and the protein-of-interest is tagged.

Fluorescent timers (FT) are another way to label protein dynamics. The underlying theory of FT for tracking proteins is relatively simple. It depends on the FP chromophore maturation to gauge time since spectral changes occur during the oxidation and dehydration

maturation process of the chromophore. For studying protein turnover and degradation, the use of FTs is relatively straightforward compared to traditional techniques such as pulse-chase and radiolabelling [123]. Only transfection of the fusion construct into the cell or organism of interest is necessary to determine protein localization, turnover, and degradation [124, 125]. Despite the obvious practicalities of FTs, it has only recently experienced somewhat of a renaissance. Recently, rational mutations were introduced into mCherry to influence the blue to red chromophore maturation rate. This generated the aptly named Fast and Medium FT with maturation rates of 7.1 and 3.9 hours respectively [126]. These FTs have been fragmented into the BiFC design to detect temporal alternation of the aggregation process of α-synuclein Glu46Lys mutant in living cells [127].

1.2.1.2 Tripartite split-GFP

Despite the versatility of BiFC, there remain issues related to fragment solubility, folding, and spontaneous association that leads to increased background signals. Possible solutions to some of these issues would be to introduce a smaller tag to reduce aggregation and folding interference. This line of reasoning led to the creation of a tripartite split GFP complementation assay [128]. Developed by the same research group involved with split sfGFP, the design bears a strong resemblance. In order to create a smaller tag, the GFP is divided into three fragments: two small tagging peptides, β-strand 10 (GFP10) and β-strand 11 (GFP11) which are approximately 20 amino acids in length and a third fragment with the remaining β-strands of the barrel, GFP1-9 detector (GFP1-9) (**Figure 1.8A**) [128]. Reassembly of a mature FP will only occur if the two tagged proteins come into close proximity to bind and form a complex with GFP1-9. Now, unstable protein complexes that do not complement well with the existing bulky fragments of BiFC can be studied using

these smaller protein tags. Without any interaction partners fused, less self-association and background was observed for this three-body system. The authors demonstrated the interaction of the large protein complexes Ku70 and Ku80, important in the non-homologous end joining pathway (NHEJ) for DNA repair by fusing GFP10 and GFP11 to Ku80 and Ku70 respectively and transfecting them into Chinese hamster ovary (CHO) cells with the detector fragment [128].

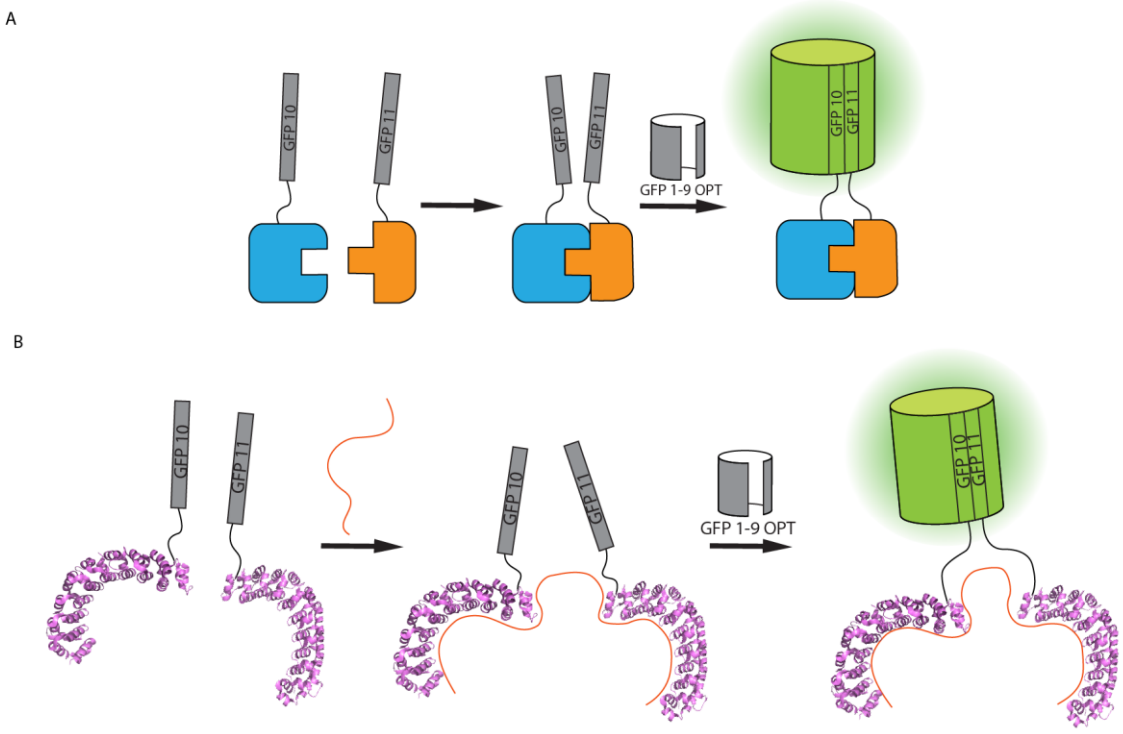


Figure 1.8. Schematic illustration of tripartite split-GFP and tetramolecular fluorescence complementation (TetFC). (A) General design of tripartite split-GFP where a PPI will bring together strands GFP10 and GFP11 thus allowing subsequent complementation with GFP 1-9 OPT to successfully reconstitute the full GFP. (B) TetFC where presence of the RNA strand (red) with corresponding sequences matching the two Pumilio homology domains (purple) brings together GFP10 and GFP11 which allows complementation with GFP 1-9 OPT to release the green fluorescent signal.

The introduction of the tripartite split GFP system for detecting PPIs has inspired a new RNA-protein detection system, tetramolecular fluorescence complementation (TetFC)

[129]. Kellerman *et al.* attached two unique PUM-HD, wild-type PUM-HD (PUM-WT) and a mutated PUM-HD variant (PUM-Var1), to the tagging fragments GFP10 and GFP11 (**Figure 1.8B**) [129]. They were able to detect the signal within ten minutes and distinguish ssRNA with high sequence specificity from total RNA extracted from *E. coli*. Background was also minimal. If no RNA was present, slight fluorescence was observed after ten hours [129]. However, application of this system in both prokaryotes and eukaryotes remains to be tested.

Although Pumilio can be designed for sequence specific recognition of any eight nucleotide stretch of RNA, the RNA target sequence must begin with a 5' UGUR, where R is a purine [130]. To bypass the necessity to engineer Pumilio and to search for a 5' UGUR sequence in the RNA of interest, Rath *et al.* introduced the use of RNA staples to their TetFC system [131]. The RNA staple will bind the complementary RNA sequence of interest and encode a generic PUM-HD recognition site on the other end of its sequence. With this approach, the PUM-HD does not have to be modified at all if the target RNA is changed, only the RNA staple. By implementing two sets of RNA staples with the previous TetFC system, they were able to differentiate spliced and unspliced HIV transcripts where spliced variants represent early stage of infection and unspliced represent the late phase. Maximum signal was achieved after 5 hours using the RNA staple approach [131]. This compares favourably with the previously reported 10 hours without the staples [129]. The fluorescence signal gives a good estimate of the amount of ribonucleoprotein complex formed. However, it remains to be seen whether this system can withstand live cell conditions for single RNA imaging. Some of its limitations are the long time span required for maximum signal output and the intracellular stability of the RNA staple itself.

1.2.1.3 Alternate FP beta strand modifications inspired by split sfGFP

The novelty of the split sfGFP system inspired more experimentation on its characteristics which resulted in a novel FP complementation approach called an alternate FP β-strand displacement system [132]. By circular permutation, $s10^{203T}$ (β-strand 10 with threonine at residue 203) was placed at the N-terminal of the sfGFP sequence while $s10^{203Y}$ (β-strand 10 with residue 203 mutated from threonine to tyrosine) was placed at the C-terminus. The $s10^{203Y}$ strand confers a yellow fluorescence when in complementation with the rest of the GFP β-barrel (**Figure 1.9**) [132]. Through control of the lengths of the two linker sequences joining each s10 strand to the rest of the protein, an initial bias of yellow fluorescence over green was created. A protease cleavage site between the $s10^{203Y}$ strand and the protein creates a light-driven protease sensor. Following cleavage by the protease, 405 nm light irradiation caused accelerated photodissociation of $s10^{203Y}$ [133]. Irreversible intramolecular replacement by $s10^{203T}$ then occurs to produce a green fluorescent signal. This sensor gives a ratiometric dynamic change of over 2000-fold [132].

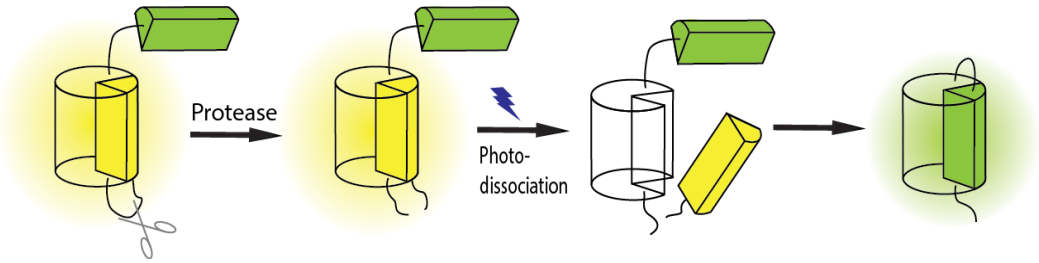


Figure 1.9. General graphic illustration of the β-strand displacement system. The $s10^{203Y}$ strand confers a yellow fluorescence, cleavage at the protease cleavage site followed by photo-dissociation by a 405 nm releases $s10^{203Y}$. This allows $s10^{203T}$ to complement and form the green FP.

There have not been any reported applications of this sensor for live cell applications. It is likely that further understanding of the underlying physical mechanisms

governing the chromophore excitation relation to association and dissociation of the strand will be necessary before further applications can be pursued.

1.2.1.4 Dimerization-dependent Fluorescent Proteins (DdFPs)

The shortcomings of FP-based FRET biosensors and BiFC strategies led our research group to create dimerization-dependent fluorescent proteins (ddFPs) [134]. DdFPs are a new class of FP complementation biosensor and a novel strategy for the detection of PPIs, enzymatic functions, and small molecules. Exploiting the oligomeric properties of DsRed, its dimeric progeny, tdTomato, was evolved to create a red ddFP (ddRFP) [134]. DdFP technology involves two FP monomers known as the A copy and B copy (**Figure 1.10**). The A copy contains a quenched chromophore and heterodimerization with the B copy, which is non-fluorogenic, stabilizes the anionic state of the A copy chromophore leading to an increased fluorescence [135]. The nature of ddFP's "on" fluorescence in heterodimeric state and "off" fluorescence in the monomeric state provides an intensiometric red response that can enable simultaneous live cell imaging of multiple dynamic processes together with other colours of sensors. Since the development of the original ddRFP, the colour palette has been expanded to include green (ddGFP) and yellow (ddYFP) [136].

The greatest advantage of ddFPs over the traditional FP complementation approach is the reversibility of the heterodimer formation. This is particularly advantageous for monitoring real time association and dissociation of PPI in which the two potentially interacting protein-of-interest are each fused to the A and B copies of the ddFP [134]. However, if working with ddFPs at relatively high concentrations (~10 μM), one must be

aware of its tendency to spontaneously dimerize. Comparatively, dimerization of FRET pairs is weak due to the much higher dissociation constants (>100 μM).

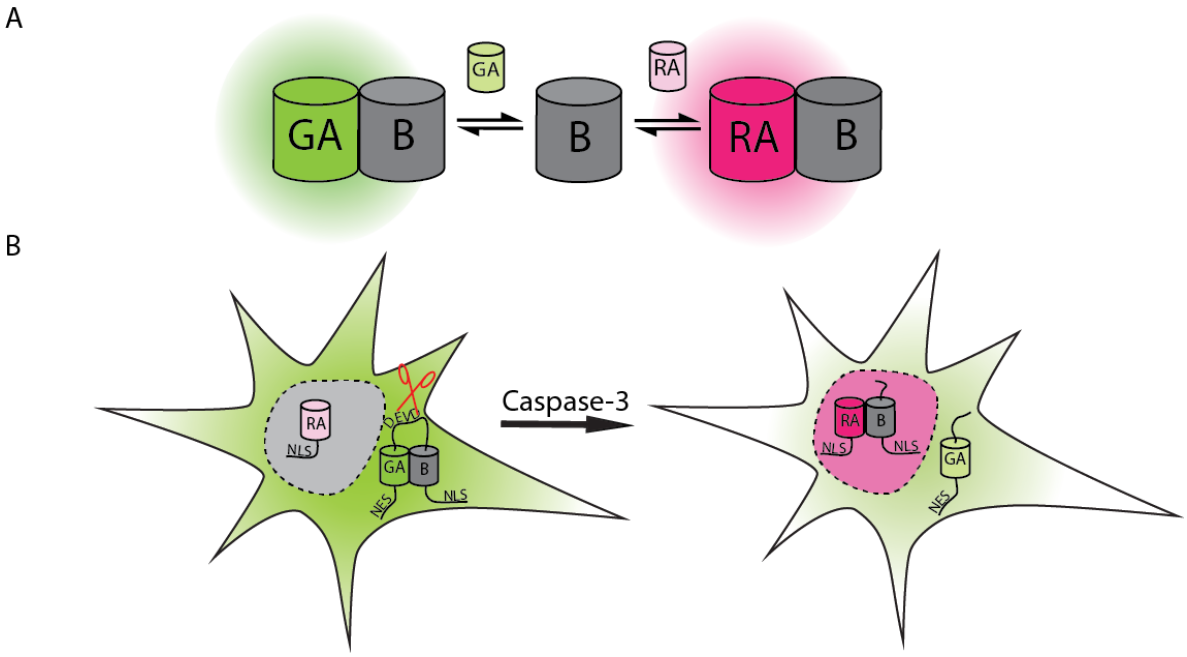


Figure 1.10. Schematic of FPX strategy and applications in live cells. (A) Central to the FPX strategy is the B copy swap where two distinctly fluorescent A copies, GA and RA, are co-expressed with a B copy. Detection of the GA-bound state or RA-bound state of B copy can be determined by the emission ratios. (B) Caspase-3 activity translocation biosensor. Cleavage of the caspase-3 cleavage site (Asp-Glu-Val-Asp, DEVD) between the NES-GA-B-NLS construct by staurosporine induced caspase-3 activity frees B-NLS which can then translocate into the nucleus to bind to RA-NLS, thus, the ratio of red fluorescence over green will become significantly higher as time progresses.

All three sets of ddFPs contain their unique A copy and B copy. A serendipitous discovery demonstrated the interchangeability of the B copies between the unique A copies. This resulted in the development of FP exchange (FPX) biosensor design strategy [5]. By co-expressing a B copy with the green A copy (GA) and red A copy (RA) simultaneously, ratiometric fluorescent changes from green to red can be achieved. By this method, a caspase-3 biosensor was designed of which the caspase-3 cut site was embedded in the linker between GA and B copy with a nuclear exclusion signal (NES) and a nuclear

localization signal (NLS) attached respectively (**Figure 1.10B**) [137]. This NES-GA-B-NLS construct was transfected along with a RA-NLS construct. The former was expressed exclusively in the cytoplasm and the latter in the nucleus. Cells stimulated to undergo apoptosis allowed the release and translocation of the B-NLS into the nucleus to bind to the RA leading to an increase in red fluorescence at the expense of green. The FPX strategy is extremely versatile. The modularity and reversibility of the system presents opportunities for applications previously elusive to the well-established BiFC technology. FPX has been extended for Ca^{2+} monitoring and shown to accurately represent histamine-induced Ca^{2+} oscillations with increases of red intensity juxtaposing the decreases in green. It has been further applied to detect phosphatidylinositol 4,5-biphosphate (PIP_2) hydrolysis, protein kinase A (PKA) activation, and extracellular signal-regulated kinase (ERK) activity [5].

1.3 Scope of Thesis

This thesis will describe our efforts to expand the applications of ddFPs. Chapter 2 will describe the design of plasmids for biosensing of O-GlcNAc Transferase (OGT) O-GlcNAcylation, protease activity, and SUMOylation of Tomosyn-1 by SUMO1. Through various rounds of trial and error we designed a plasmid that is to be tested in cells. Chapter 3 will describe our efforts to characterize a dark 'A' (DA) accessory protein for ddFPs. Preliminary studies in cells show that DA has promise as a competitive inhibitor of GA binding to a B copy. The final chapter is a summary of this thesis and an outlook on the future developments of ddFPs.

Chapter 2: Biosensing Strategies for O-GlcNAcylation and SUMOylation

2.1 Introduction

Coordination of the multitude of protein functions within the human proteome requires an intricate network of protein post-translational modifications (PTM) that ensure proper functioning of the cell. The particular type of PTM can determine a protein's activity state, localization, turnover, and interactions with other proteins [138]. O-GlcNAcylation and SUMOylation are two such PTMs. Both modifications modulate a wide range of targets and are vital at the onset of embryogenesis [139, 140]. Slight aberrations in these PTMs can have implications including cancers, autoimmunity and pathologies of metabolic and neurodegenerative diseases [141].

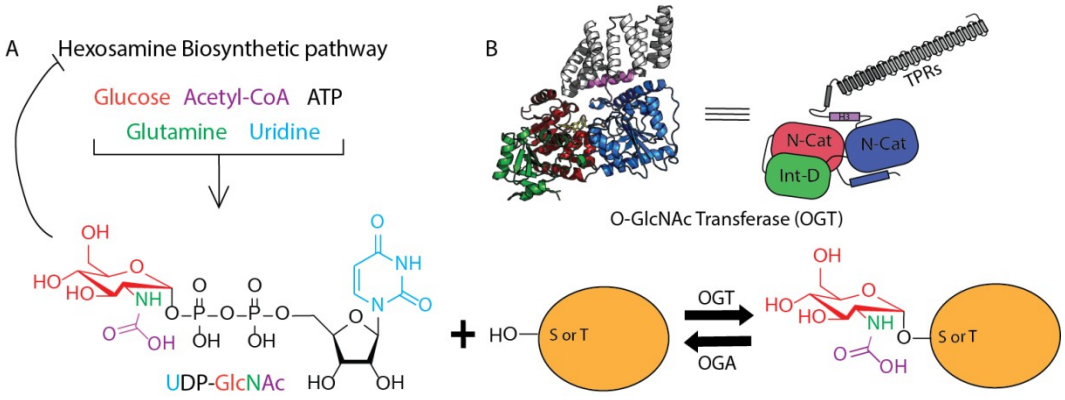


Figure 2.1. O-GlcNAcylation by O-GlcNAc transferase (OGT). (A) Glucose enters the hexosamine biosynthetic pathway and is extensively modified resulting in the endpoint product UDP-GlcNAc. UDP-GlcNAc negatively regulates this pathway. (B) OGT is known to be a nutrient sensing protein. UDP-GlcNAc is taken up by OGT and the GlcNAc moiety is transferred to a serine (Ser) or threonine (Thr) of a protein. Removal of GlcNAc is catalyzed by O-GlcNAcase (OGA). This overall process is known as O-GlcNAcylation cycling. The tricorpeptide repeats (TPRs) domain on the OGT is responsible for recognizing the protein to be glycosylated. There are three isoforms of OGT in humans that differ in the number of TPRs they contain. (OGT, PDB entry 3PE3) (Figures are adapted from [141, 142]).

O-GlcNAcylation is one of the most common post-translational modifications (**Figure 2.1**) [143]. O-GlcNAcylation cycling involves the addition of GlcNAc, derived from its activated form uridine diphosphate-β-*N*-acetyl glucosamine (UDP-GlcNAc), on to a Ser or Thr residue. These Ser or Thr residues are often located at an intrinsically disordered (ID) region of a protein [141]. This addition reaction is catalyzed by O-GlcNAc transferase (OGT), while O-GlcNAcase (OGA) catalyzes the removal of the modification. UDP-GlcNAc holds a crucial position as the endpoint product of the hexosamine biosynthetic pathway. Cells deficient in this molecule have been shown to have decreased O-GlcNAc modifications of cytosolic and nuclear proteins which leads to unresponsiveness to apoptotic stimuli and proliferation defects [144]. Due to the sensitivity of OGT activity to UDP-GlcNAc concentrations, OGT is proposed to regulate cellular responses to nutrient status [143, 145].

Recently, OGT has been discovered to also act as a protease that cleaves the human epigenetic cell-cycle regulator host cell factor-1 (HCF-1) [145-147]. HCF-1 must undergo proteolytic maturation at any one of its six repeat sequences located at its centrally located proteolytic processing domain (PPD) by the nutrient-responsive OGT. The resulting terminal subunits, N-(HCF-1_N) promotes progression through the G1 phase and C-(HCF-1_C) is essential for proper mitosis and cytokinesis in the M phase. Interestingly, the protease active site of OGT is also its glycosylation active site and binding of HCF-1 cleavage sites to OGT is via an ordered bi bi mechanism where UDP-GlcNAc must bind first [148]. Further validation of OGT protease activity on HCF-1 in live cell conditions remains. If confirmed, a link between the cell's nutrient status and cell cycle can be established.

In contrast to O-GlcNAcylation, only a small percentage of proteins are SUMOylated, and most of these proteins are associated with DNA. SUMO (small ubiquitin-like modifier) undergoes proteolytic maturation at its C-terminal end to expose the di-glycine motif which interacts with the lysine residue of the substrate protein to form an isopeptide bond (**Figure 2.2**) [140]. Sentrin/SUMO-specific protease (SENPs) is responsible for both the proteolytic maturation of SUMO, as well as removal of SUMO from its substrate. Compared to lower eukaryotes exhibiting only one SUMO, humans encode four. Among these four are SUMO-1, SUMO-2/3 and SUMO-4. SUMO-4 protein levels are mostly undetectable. Human SUMO-2 and -3 have 97% sequence identity and cannot be distinguished by antibodies, because of this, their nomenclature is as appears, SUMO-2/3, which is consistent with that proposed by Saitoh & Hinchey [149, 150].

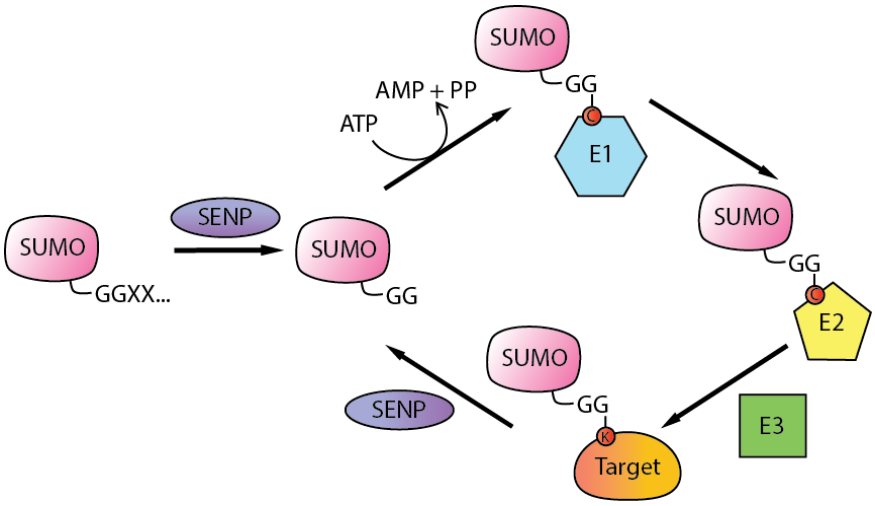


Figure 2.2. The SUMOylation cycle. Maturation of SUMO is by cleavage of its C-terminus end (XX represents the amino acids on the C-terminus) by SUMO specific proteases known as the SENP family. This cleavage reveals the Glycine-Glycine (GG) motif. The matured SUMO is then linked to E1 at the catalytic cysteine (which is denoted as C in this figure) for activation. SUMO is then transferred to E2. E2 catalyzes the formation of the isopeptide bond between SUMO and the lysine of the target with the help of E3 ligase. SUMOylation is reversible with the help of SENP for the deconjugation.

Insulin secretion from pancreatic β-cells necessarily involves the central proteins involved in exocytosis, such as syntaxin-1, SNAP25 and VAMP2 [151]. Tomosyn-1, a syntaxin-1 binding protein, is expressed in pancreatic islets, with m-tomosyn-1 being the predominant isoform [151]. Studies have shown that tomosyn-1 localizes in subcellular compartments enriched in insulin-containing secretory granules [151]. In addition, silencing of tomosyn-1 results in a decrease of insulin exocytosis. Interestingly, overexpression of tomosyn-1 in neurons exerts an inhibitory control of exocytosis [152].

There is both a positive and negative impact from SUMOylation in pancreatic islet cells. Increased expression of SUMO-1 creates insulin secretion dysfunction in the cell, but provides the cell with antiapoptotic properties by reducing activation of NF-κB. In contrast, increased expression of SENP1 leads to induction of apoptosis [153]. We are interested in establishing a relationship between tomosyn-1 and SUMO-1 which have been demonstrated to have possible interactions in pancreatic islet cells. Of particular interest is tomosyn-1's propensity to be modified by SUMO-2/3, but not SUMO-1, in HEK293 cells [154].

In this chapter, we will focus on the construction of plasmids encoding reporters for detection of OGT glycosylation and protease activity, as well as for SUMO-1 binding to tomosyn-1.

2.2 Results and Discussion

2.2.1 Construction of plasmids for OGT glycosylation activity

Previously, nuclear, plasma membrane, and cytoplasm–localized OGT glycosylation intramolecular FRET sensors were developed in efforts to study O-GlcNAc spatial and temporal localization during signalling [155, 156]. The FRET pair of ECFP and Venus was used. These sensors proved moderately useful for imaging of OGT activity in

COS-7 cells, but were limited by a relatively small ratiometric fluorescent response. Here, we decided to create an intermolecular, intensiometric OGT activity biosensor using ddFPs. Similar to the previous sensors developed by Carrillo *et al.,* we utilized GafD as the basis for the sensing activity. GafD, isolated from uropathogenic *E. coli*, is a fimbrial lectin known to bind to GlcNAc [157]. For the OGT O-GlcNAcylation site we used the well-characterized PGGSTPVSSANM casein kinase II peptide substrate, a serine-threonine kinase, known to be O-GlcNAc modified [158]. Intriguingly, there are not many O-GlcNAcylation sites associated with a clear consensus motif, unlike protein phosphorylation sites that can often be identified by their primary sequence alone [141].

We decided to attach GafD to the C-terminal end of RA (RA-GafD) and O-GlcNAcylation substrate to the C-terminal end of RB (RB-Sub) (**Figure 2.3**). O-GlcNAcylation of the Ser residue on the substrate by OGT on the substrate would cause GafD to bind, thus, leading to RA and RB association and red fluorescence. Both these constructs were to be co-expressed in HeLa cells. We anticipated that one challenge of using this reporter would be in properly identifying the transfected cells as this was a "turn on" biosensor and would initially be dim prior to the stimulation of O-GlcNAcylation activity. For this reason, we planned to include a third gene encoding EGFP with a K-Ras CaaX plasma membrane localization signal attached to its C-terminal end ($EGFP^{CaaX}$). To ensure expression of all constructs in the cell we decided to place all three constructs into one plasmid.

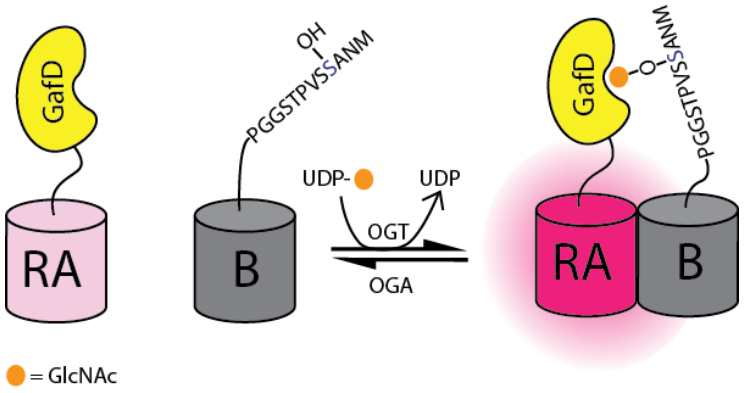


Figure 2.3. OGT O-GlcNAcylation biosensor strategy. The casein kinase II peptide substrate is recognized by OGT and subjected to O-GlcNAc modification on the hydroxyl group of the serine. GafD recognizes this O-GlcNAcylated substrate and binds to it thus allowing the RA and B copy to come into close proximity to each other to allow heterodimer formation and hence, red fluorescence.

In order to get multiple gene products expressed from a single vector, we decided to use the p2A sequence. The 2A peptide is from the porcine teschovirus-1, a foot-and-mouth disease virus, hence its name, p2A. Highly efficient self-cleavage of this peptide occurs after the entire construct has been expressed into its primary protein sequence. An alternative approach for expression of multiple genes is to use an IRES sequence, but due to its size of 500 nucleotides and the difference in translation efficiency for a gene placed after an IRES sequence to the one before [158], we opted for the smaller 19-amino acid p2A peptide. A major advantage of the p2A peptide is that it provides stoichiometric expression of the flanking proteins. Knowing this, we designed a pBAD/His B vector with a multiple cloning site (MCS) to accept three gene constructs with p2A sites in between (**Figure 2.4**). The cleavage efficiency of this p2A peptide is estimated to be approximately 70-80% in cells. To ensure an even higher percentage of cleavage, two p2A sequences were placed one after the other at each site between the gene constructs. The MCS sequence was not immediately placed into a pcDNA3.1(+) mammalian expression vector as leaving it in a

modified pBAD/His B (*Mlu* I removed) vector would facilitate efficient colony selection during gene insertion since we are dealing with FPs.

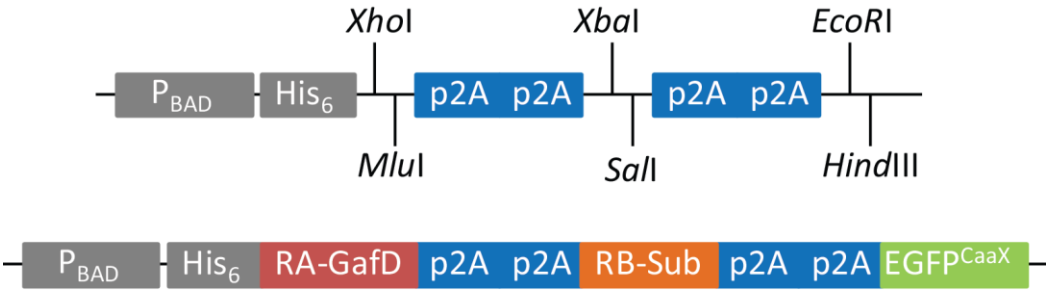


Figure 2.4. Construction of expression plasmids harbouring DNA sequences encoding p2A peptides flanked by multiple cloning sites. Schema showing a map for an expression plasmid harbouring DNA sequences encoding p2A peptide and the construct used in this study encoding p2A peptides flanked by RA-GafD, RB-Sub, EGFP^{CaaX}.

Using a series of ligations, RA-GafD, RB-Sub and EGFP^{CaaX} were introduced into the MCS sequence and subsequently transferred to the pcDNA3.1(+) mammalian expression vector. When we expressed this plasmid in HeLa cells, we saw robust plasma membrane localized green fluorescence. Treatment of these cells to thiamet-G, a potent and selective inhibitor for OGA, and starvation conditions followed by addition of fresh media with serum did not elicit any detectable red fluorescence. We believe that GafD and substrate were not properly expressed, possibly due to the presence of C-terminal attached proteins, which may cause complications in protein folding.

2.2.2 Construction of plasmids for OGT protease activity

For detection of protease activity, we heavily modified the vector pcDNA3.1(+)-*Hind*III-RA-*Kpn*I-DEVD-GB-*Bam*HI-GA-*Xho*I to create pcDNA3.1(+)-*Hind*III-RA-*Kpn*I-SATD-GB-*BamH*I-HCF-1cleavage site-GA-*Xho*I. Addition of new gene sequences was done via ligation. By using this plasmid, we expected to obtain a ratiometric green/red fluorescence readout of HCF-1 cleavage by OGT. Before full induction of OGT activity,

both red and green fluorescence should be detected with equal fluorescence of red and green (**Figure 2.5**). After cleavage, GA would be released and due to the proximity of RA to GB, red fluorescence would be more prominent. GA and GB are expected to bind more tightly due to their higher affinity in comparison to GB for RA. However, with the addition of an HCF-1 cleavage substrate, the linker length between GA and GB has expanded to 44 amino acids which may reduce the change in the ratiometric readout before and after cleavage. This HCF-1 proteolytic repeat is comprised of the cleavage region containing the conserved cut site sequence Cys-Glu-Thr (CET) and a threonine rich region important for OGT recognition. The Glu in the CET sequence is critical for cleavage. Substitution of Glu to Gln, Ala, Asp and Ser has been reported to lead to loss of cleavage [145]. In contrast, Ala and Ser substitutions to the Cys position were tolerated. From this information, we made two additional constructs as controls, one bearing a Glu to Gln mutation and the other Cys to Arg. Our constructs were sent to the Suzanne Walker's lab (Harvard Medical School) for further applications in live cells.

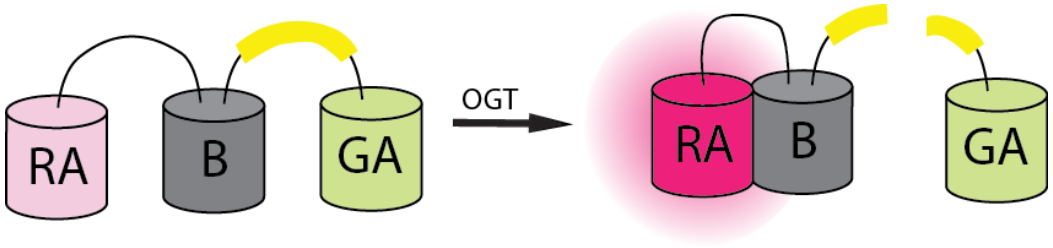


Figure 2.5. OGT protease activity sensor. In yellow is one of the six repeats from the proteolytic processing domain (PPD) of HCF-1. Before the cleavage of OGT, RA and GA both have almost equal probability to associate with B copy to release red or green fluorescence. With the longer linker between B copy and GA compared to RA and B copy, a slightly higher RA and B association is expected. After OGT cleaves the repeat from the PPD of HCF-1, GA is freed and intensity of red fluorescence due more B copy associating with RA results.

2.2.3 Construction of plasmids for SUMOylation of tomosyn-1 by SUMO-1

For the construction of this vector, we utilized the modified pBAD/His B with the MCS and p2A sites. SUMO1 was attached C-terminally to RA and rat m-tomosyn-1 was attached to the N-terminal end of RB. Again, the $EGFP^{CaaX}$ was used as a transfection reporter (**Figure 2.6**). As controls, Lys residues 298 and 854 that are possible SUMOylation sites in m-tomosyn-1 were mutated to Arg in three different combinations. One of the main challenges during the creation of this vector was the sheer size and a number of restriction enzymes sites in the m-tomosyn-1 sequence that were also present in the MCS and the backbone of the plasmid.

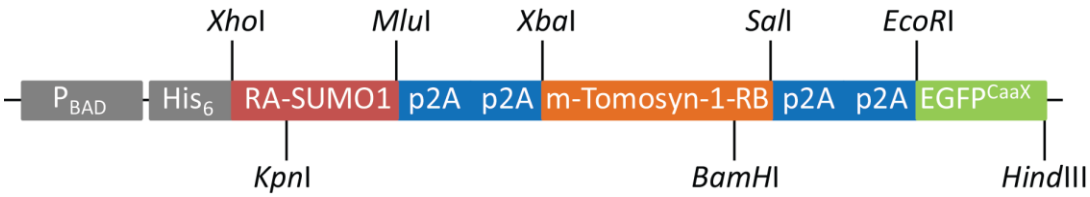


Figure 2.6. Map of multiple expression plasmid encoding RA-SUMO1, m-Tomosyn-1-RB and $EGFP^{CaaX}$ in modified pBAD/His B harbouring p2A peptide and multiple cloning sites.

2.3 Conclusion

Our goal was to create biosensors for OGT glycosylation and protease activity and tomosyn-1 SUMOylation. Undeniably, there still remains a large amount of work before these biosensors are complete. With the OGT glycosylation activity sensor, we are now reversing the placement of GafD and the casein kinase II peptide substrate to the N-terminal end of RA and RB. After creation of the construct, it would be transfected in HeLa cells to determine how well this biosensor detects OGT glycosylation activity. Preliminary imaging results with the OGT protease reporter suggest that the green channel is much

dimmer than expected, leading to less of a change after OGT cleavage occurs. We plan to minimize linker length in efforts to increase the initial ratio of green fluorescence to red. In doing so, we expect to confirm OGT's role as the protease of HCF-1 in mammalian cells. As for the tomosyn-1 SUMOylation reporter, three of the four planned plasmid constructs have been constructed, of which the three are the controls. The WT plasmid construct remains to be created. Once all the plasmid constructs are constructed, they would be transfected to pancreatic islet cells in hopes of validating the SUMOylation of tomosyn-1 by SUMO-1. We look forward to finalizing and refining these plasmid constructs to further elucidate the cell's functions.

2.4 Materials and Methods

2.4.1 General methods and materials.

All synthetic oligonucleotides (**Table 2.1**) were purchased from Integrated DNA Technologies (IDT, Coralville, IA). Pfu polymerase (ThermoScientific), Q5 polymerase (New England Biolabs) and *Taq* polymerase (New England Biolabs) were used for standard PCR reactions. Small scale preparation of plasmid DNA, PCR products and products of restriction digests were purified by gel electrophoresis and extracted by spin column (GeneJET gel extraction kit; Thermo Scientific or BioBasic Inc.) according to the manufacturer's protocols. Restriction enzymes were purchased from Thermo Scientific or New England Biolabs. DNA manipulations such as polymerase chain reaction (PCR), ligation, DNA gel electrophoresis, restriction enzyme digestion, ligation and transformation were performed according to Sambrook *et al.* [159]. Ligations were used to transform electrocompetent DH10B *E. coli*. DNA sequencing was performed either at University of

Alberta Molecular Service Unit (MBSU) or University of Calgary University Core DNA services.

Standard PCR amplifications were performed in 50 µL reaction volumes containing nuclease-free water, 1× reaction buffer, 200 µM dNTPs (Invitrogen), 3% DMSO, 200 nM forward and reverse oligonucleotides, 10-50 ng of template DNA, 1.0 unit of Pfu polymerase and 1.0 unit of Taq polymerase. Typical cycling parameters were as follows: initial denaturation at 95 °C for 60 seconds, 35 cycles of denaturation at 95 °C for 30 seconds, 54-60 °C for 30 seconds, 72 °C for 60 seconds per kilobase (kb) of target; final extension at 72 °C for 5 minutes. PCR amplifications using Q5 were also performed similarly as the standard PCR amplifications with the addition of 1× GC enhancer buffer and the lack of Taq polymerase. Cycling parameters for Q5 are again similar to the standard procedures except initial denaturation and denaturation is at 98 °C, in addition, extension time at 72 °C can be set at 30 seconds per kb of target. Cloning strategies of the constructs described in this chapter are outlined in **Table 2.1**, which details the relevant oligonucleotides, DNA template, destination plasmids and restriction enzyme sites used.

2.4.2 Site-directed mutagenesis

Multi site-directed mutagenesis was completed using the Quikchange Lightning kit (Agilent) following manufacturer's recommendations but at half the reaction volume. Typical cycling parameters were as follows: initial denaturation at 95 °C for 2 minutes; 30 cycles of denaturation at 95 °C for 20 seconds, 58-60 °C for 30 seconds, 72 °C for 30 kb of plasmid target; final extension at 72 °C for 5 minutes. When the reaction was over, 0.5 µL of *DpnI* was added to the reaction mixture and incubated at 37 °C for 10 minutes and then cooled to 12 °C. Propagation of plasmid DNA was done via addition of 1 µL reaction

mixture into thawed electrocompetent DH10B *E. coli*. The mixture was then transferred into 0.2 cm MicroPulser cuvettes (Bio-Rad) and electroporated by a MicroPulser electroporator (Bio-Rad). Pulsed cells were then plated on agar plates containing 100 μg/mL ampicillin and 0.02% L-arabinose. Agar plates were incubated overnight at 37 °C. Colonies were then selected and subcultured in 4 mL lysogeny broth (LB) broth containing 100 μg/mL ampicillin. The subculture was then incubated at 37 °C with shaking at 250 rpm. The next day, DNA was extracted with GeneJet genomic DNA purification kit (Thermo Scientific).

2.4.3 Construction of mammalian expression plasmid for live cell imaging

The construct *Xho*I-*Mlu*I-p2Ap2A-*Xba*I-*Sal*I-p2Ap2A-*EcoR*I-*Hind*III was ordered as a gBlocks Gene Fragments from IDT (Coralville, IA). It was subsequently digested and ligated into a pBAD/His B plasmid modified at its *Mlu*I site in order to expand the restriction enzyme choice in the multiple cloning sites.

Table 2.1 Details of the gene construct created in this work.

Name	Gene (Restriction sites in italics and substrate sequence in bold)	Comments
OGT protease activity biosensor	pcDNA-*Hind*III-RA-*Kpn*I-SATD-GB-*BamH*I-*HCF-1_cleavage_site*-GA-*Xho*I	All assembly was completed by ligation
Cloning Strategy		
Construct	Oligos (Fwd/Rv)	Target Plasmid
*Kpn*I-SATD-GB-*BamH*I	2.1/2.2	pcDNA-*Hind*III-RA-*Kpn*I-mCherry-*BamH*I-GA-*Xho*I
*BamH*I-mCherry-*Xho*I	2.3/2.4	pcDNA-*Hind*III-RA-*Kpn*I-SATD-GB-*BamH*I-GA-*Xho*I
*BamH*I-HCF-1(wt)-GA-*Xho*I	2.5,2.6/2.9	pcDNA-*Hind*III-RA-*Kpn*I-SATD-GB-*BamH*I-HCF-1_cleavage_site-GA-*Xho*I
*BamH*I-HCF-1(E10Q)-GA-*Xho*I	2.5,2.7/2.9	
*BamH*I-HCF-1(C9R)-GA-*Xho*I	2.5,2.8/2.9	

Name	Gene (Restriction sites in italics and substrate sequence in bold)		Comments
SUMO1, tomosyn-1 SUMOylation biosensor	pcDNA-*Xho*I-RA-*Kpn*I-GafD-*Mlu*I-p2Ap2A-*Xba*I-mTomosyn1-*Bgl*II-RB-*Sal*I-p2Ap2A-*EcoR*I-EGFP^{CaaX}-*Hind*III		All assembly was completed by ligation
Cloning Strategy			
Construct	Oligos (Fwd/Rv)	Target Plasmid	
*Xho*I-RA-*Kpn*I	2.10/2.11	pBAD-*Xho*I-EGFP-*Mlu*I-p2Ap2A-*Xba*I-*Sal*I-p2Ap2A-*EcoR*I-*Hind*III	
*Kpn*I-SUMO1-*Mlu*I	2.12/2.13		
*Xba*I-EGFP-*Sal*I	2.14/2.15	pBAD-*Xho*I-RA-SUMO1-*Mlu*I-p2Ap2A-*Xba*I-*Sal*I-p2Ap2A-*EcoR*I-*Hind*III	
*Xba*I-*Bgl*II-RB-*Sal*I	2.16/2.17	pBAD-*Xho*I-RA-SUMO1-*Mlu*I-p2Ap2A-*Xba*I-EGFP-*Sal*I-p2Ap2A-*EcoR*I-*Hind*III	
*Xba*I-mTomosyn1-*Bgl*II	2.18/2.19	pBAD-*Xho*I-RA-SUMO1-*Mlu*I-p2Ap2A-*Xba*I-*Bgl*II-RB-*Sal*I-p2Ap2A-*EcoR*I-*Hind*III	
Transfer *Xho*I-RA-SUMO1-*Mlu*I-p2Ap2A-*Xba*I-mTomosyn1-*Bgl*II-RB-*Sal*I-p2Ap2A-*EcoR*I into already prepared and digested at *Xho*I and *EcoR*I, pcDNA3.1(+)-*Xho*I-*Kpn*I-RA-*Mlu*I-p2Ap2A-*Xba*I-*Bgl*II-RB-*Sal*I-p2Ap2A-*EcoR*I-EGFP^{CaaX}-*Hind*III		pBAD-*Xho*I-RA-SUMO1-*Mlu*I-p2Ap2A-*Xba*I-mTomosyn1-*Bgl*II-RB-*Sal*I-p2Ap2A-*EcoR*I-*Hind*III	

2.4.4 Mammalian cell culture and imaging for OGT glycosylation activity

HeLa cells were maintained in complete media (DMEM with 10% fetal bovine serum (FBS), 1% Penicillin Streptomycin (Pen Strep) (Invitrogen), 5% Glutamax (Invitrogen)) at 37 °C and 5% CO_2 according to standard procedure. Transient transfections of pcDNA3.1 (+) expression plasmids were performed using Turbofect (Thermo Scientific). HeLa cells were grown on 35 mm imaging dishes and grown to approximately 70-80% confluency and incubated with 1 mL of DMEM (FBS free) for 10 min then transfection reagent mixture was added. Generally, 2 μL of transfection reagent was used per 1 μg of DNA transfected and mixed in 0.2 mL of DMEM (FBS free). After 2 hours, the culture

media was changed back to complete media. 24 to 48 hours after post-transfection, thiamet-G (Cayman Chemical) was added to the transfected cells to a final concentration of 100 μM and incubated overnight. The next day, the media was changed to HEPES-buffered Hank's balanced salt solution (HHBSS) and the cells were subjected to imaging. Imaging was completed using an Axiovert 200M (Zeiss) microscope. Axiovert 200M (Zeiss) was equipped with a 75W Xenon-arc lamp, 20× objective lens and a 14-bit CoolSnap HQ2 cooled CCD camera (Photometrics), driven by open source Micro-Manager software.

Chapter 3: Engineering a Dark Accessory Protein for ddFPs

3.1 Introduction

A complete map detailing the complex network of proteins and small molecules or ions involved in intracellular signalling cascades has long been sought after. In order to experimentally determine this interaction web, there has been a substantial investment of effort in the development in fluorescence-based biosensors for simultaneous determination of the spatial and temporal dynamics of more than one event within the same cell, known simply as multiparameter fluorescence imaging [160].

To monitor upstream and downstream effects of a signalling pathway simultaneously, sensors that can be spectrally and/or spatially resolved must be used in combination. Much effort has been invested in the development of FP-based FRET biosensors for simultaneous multiparameter fluorescence imaging [161]. FP-based FRET biosensors can be used in conjunction with a second FP-based FRET biosensors, single FP-based biosensors, and/or BiFC with orthogonal wavelengths to enable multiparameter fluorescence imaging of PPIs, proteolytic cleavage, and small molecule or ion flux [41, 162, 163]. The main challenge associated with all of these approaches is the fact that FPs have very broad excitation and emission spectra and so the fluorescence from one biosensor bleeds into other fluorescence channels [161]. A major contributing factor to this challenge is the fact that FPs have excitation and emission wavelengths that are restricted to the relatively narrow band of visible wavelength (~400-650 nm). A FRET pair such as CFP and YFP has an excitation and emission spectrum spanning greater than 100 nm of the visible spectrum, thus leaving only a small spectral window for measurement of another parameter [160].

A recent innovation in the field of biosensors is ddFPs [134]. This novel strategy is conceptually a blend of FP-based FRET biosensors and BIFC, where two fluorogenic monomeric FPs known as the A and B copy form a fluorescent heterodimer upon association. Unlike BiFC, this process is reversible which makes it ideal for dynamic PPI detection. DdFPs have been used in the FPX biosensor design strategy where the RA and GA both compete for binding to the B copy [5]. In this approach, proteins-of-interest are attached to an A copy and B copy to create a biased binding for each other, and a cellular event will reduce the bias and allow increased binding of the other A copy to B. However, this method still employs two fluorescent channels to image a cellular event.

In this chapter, we will discuss the development, characterization and preliminary applications for a dark accessory protein for ddFPs which we have named the "Dark A copy" (DA). DA is essentially an A copy that is non-fluorogenic both as a monomer and when bound to the B copy. We reasoned that its creation will serve as replacement for one of the A copies in the FPX strategy such that a spectral channel may become "free" and can be used to image another cellular event in a different colour. DA can also serve as a competitive inhibitor for the A copy binding to B to reduce non-specific binding and improve contrast when imaging. An important application of the DA protein would be the creation of 'turn-on' sensors that go from dim to bright in response to a biochemical event such as proteolysis. It has been notoriously difficult to engineer effective turn-on sensors using FPs.

3.2 Results and Discussion

3.2.1 Development and Characterization of a Dark A copy (DA)

The intrinsic chromophore is entirely responsible for the fluorescent properties of an FP. Accordingly, our strategy was to perform site-directed saturation mutagenesis on the Tyr67 of the methionine-tyrosine-glycine (MYG) chromophore of GA to generate a library of "dark" A copies (DA). Changes to the binding properties of GA to GB were expected to be minimal, since no changes would be made to the protein-protein interface. An NNK randomized codon was used to randomly mutate residue 67 of GA where N represents a 25% mix each of adenine, thymine, guanine, and cytosine nucleotides, and K represents a 50% mix of thymine and guanine [164]. The NNK codon encodes all 20 amino acids and one stop codon. Following the site-saturation mutagenesis, non-fluorescent *E. coli* colonies were selected for further characterization. DNA sequencing revealed the following mutations: Tyr67Pro, Tyr67Asn, Tyr67Leu, Tyr67Phe, Tyr67Cys, Tyr67Val and Tyr67Trp. Analysis of their absorption and emission spectra showed no absorbance peaks at wavelengths greater than 300 nm and lack of an emission peak when excited at 450 nm (**Figure 3.1**). However, Tyr67Phe has an absorbance at 315 nm (**Figure 3.1F**) and fluorescence could be visibly detected when excited under UV light. Tyr67Trp also had an absorbance at 428 nm (**Figure 3.1B**) but when excited under UV light no fluorescence was visibly detected. Poor solubility of the Tyr67Asn variant was also noted.

We also attempted to create a dark GA variant without a chromophore mutation. Two mutations, Lys83Arg and Lys70Met, were chosen based on Dr. Spencer C. Alford's results during the evolution of ddRFP [135]. Inclusion of both mutations into GA yielded a

dark variant, however, it was found not to be very soluble and hence we did not further pursue this variant.

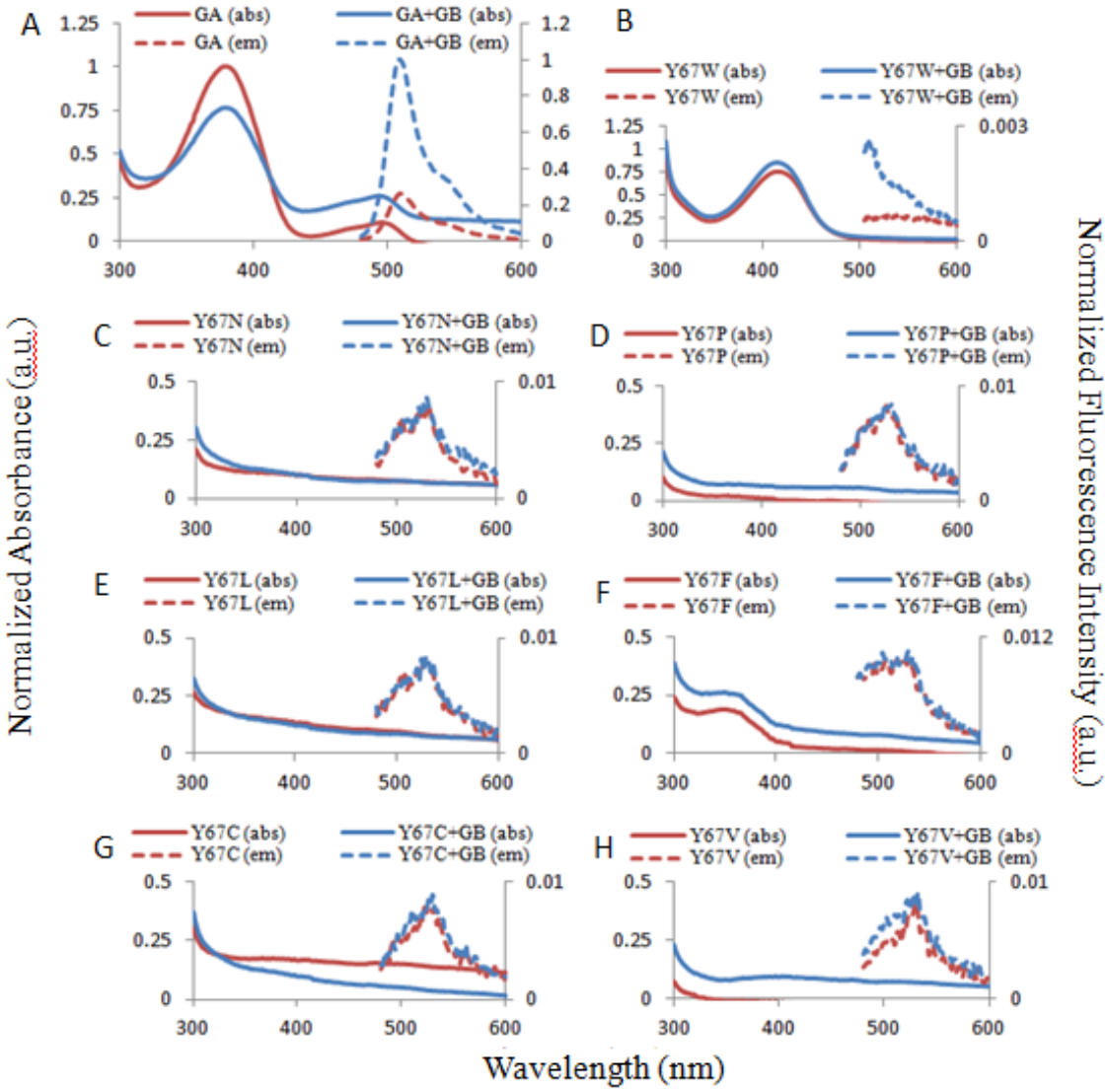


Figure 3.1. Absorption and emission spectra of DA variants (Ex. 450 nm). (A) GA and GA + GB; (B) Y67W; (C) Y67N; (D) Y67P; (E) Y67L; (F) Y67F; (G) Y67C; (H) Y67V; each with and without GB. All notations are standard one letter amino acid codes.

To initially evaluate whether the mutation has affected the secondary structure and folding of the protein, we examined all DA variants with circular dichroism (CD) (**Figure 3.2**). CD offers a rapid evaluation of protein secondary structural composition [165], which is the strength and limitation of this technique. Tertiary structure cannot be well deduced

from a CD data due to complications arising from the contribution of individual aromatic amino acids to the near UV CD spectrum [166].

The CD spectrum of each DA variant show characteristic antiparallel β-sheet content as seen by negative bands at 218 nm and positive bands at 195 nm (**Figure 3.2**) [165]. Of particular interest is the Tyr67Phe-DA that has a CD spectrum that is practically identical to GA. This is not surprising as the structures of Phe and Tyr differs only by a hydroxyl group. The Tyr67Trp-DA exhibits similar qualities. From the CD data, we decided to focus on both Tyr67Phe-DA and Tyr67Trp-DA as the data suggested the least perturbations to the barrel by these mutations and so they are also the most likely to retain binding affinity to GB.

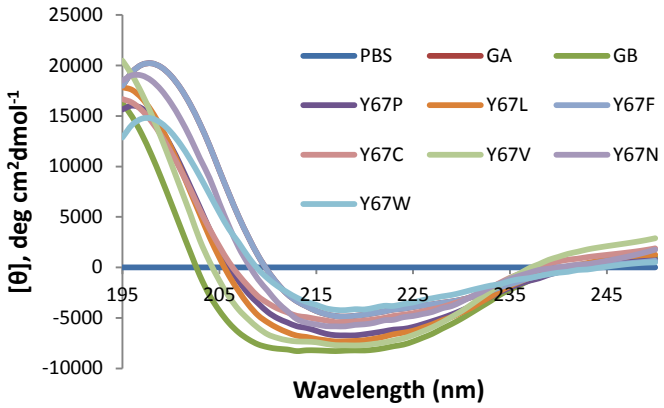


Figure 3.2. CD spectra of DA variants to phosphate-buffered saline (PBS), GA and GB control. All variants were run at 1 mg/mL in 1× PBS, pH 7.4. The GA curve is hidden by the Y67F curve. Note: all notations are standard one letter amino acid codes.

When conducting a competitive displacement assay of GA displacement from GB by Tyr67Trp-DA and Tyr67Phe-DA, we noticed that both DA variants showed emission after a certain threshold concentration. This result confirmed our previous observations that Tyr67Phe-DA emits fluorescence when excited by UV light and showed that Tyr67Trp-DA has detectable fluorescence when at high enough concentration. **Figure 3.3B** and **C** shows

the emission spectra of Tyr67Trp-DA and Tyr67Phe-DA at varying concentrations with respect to GA. Noticeably, Tyr67Phe-DA at 50 µM has approximately the same emission intensity of Tyr67Trp-DA at 200 µM. However, the maximum emission intensity of Tyr67Trp-DA at 200 µM is just approximately 3% of GA's at 200 µM and Tyr67Phe at 200 µM is only 8% of GA's maximum emission intensity at 200 µM (**Figure 3.3A**). Working concentrations in live cell applications should never reach concentrations as high as 200 µM. For *in vitro* characterization experiments such as competitive assays, the emission of Tyr67Phe-DA at 50 µM is already a complication even when a blank at the same concentration is run and subtracted (data not shown). Tyr67Trp-DA was thus selected for further experiments due to its lower level of fluorescence. Although this variant has absorption and slight emission, these attributes come in handy when characterizing its binding to GB.

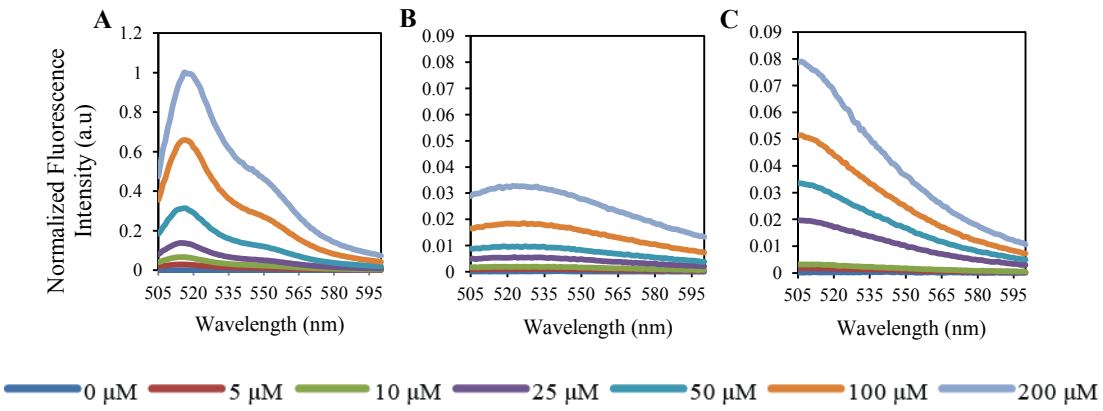


Figure 3.3. Emission spectra of (A) GA, (B) Tyr67Trp-DA and (C) Tyr67Phe-DA at 0, 5, 10, 25, 50, 100, 200 µM (Ex. 455 nm). Both Y67W-DA and Y67F-DA emission spectra are normalized to GA.

We next asked whether the Tyr67Trp-DA variant maintained the ability to bind to GB. Similar to its parent, GA, Tyr67Trp-DA has absorbance at 455 nm. By size exclusion chromatography the association of the Tyr67Trp-DA and GB heterodimer was confirmed

by measuring the elution profile by absorbance of 455 nm (**Figure 3.4D**). The Tyr67Trp-DA monomer eluted at 72 minutes (**Figure 3.4D**) which matches to the elution times of the GA monomer and GB monomer controls, 71 and 72 minutes respectively (**Figure 3.4A,B**). The Tyr67Trp-DA and GB heterodimer eluted at 67 minutes (**Figure 3.4D**) which closely matches the elution time of the GA and GB heterodimer at 66 minutes (**Figure 3.4A,B**). It was difficult to validate Tyr67Trp-DA and GB formation at 280 nm (**Figure 3.4C**). A large fraction of both proteins remained as monomers, similar to GA and GB heterodimer (**Figure 3.4A**), resulting in a broad dimer peak which was very close to that of the GB monomer and Tyr67Trp-DA monomer peaks. However, from these results we can conclude that Tyr67Trp-DA forms a heterodimer with GB.

Taking Tyr67Trp-DA as an antagonist to GA and GB binding, a competitive assay was developed to determine the IC_{50} of Tyr67Trp-DA via stopped-flow technique. Kinetic data pertaining to the k_{on} and k_{off} was not extracted from the stopped-flow data acquired according to the scheme shown in **Figure 3.5A** due to the complexity in derivation of the correct kinetic fitting equation for a competitive assay. As a result, only an equilibrium based analysis was employed. The data obtained from the stopped-flow kinetic curves (**Figure 3.5B**) were fit with a variable slope dose-response curve, Equation (1) where Y_{max} is the maximal response, Y_{min}, is the maximally inhibited response and p is the Hill slope. Extraction of IC_{50} from the fitting revealed the fluorescence response to be reduced in half at 5.30 ± 1.03 μM.

$$y = Y_{min} + \frac{(Y_{max} - Y_{min})}{1 + 10^{(logIC_{50} - x)p}} \qquad \text{Equation (1)}$$

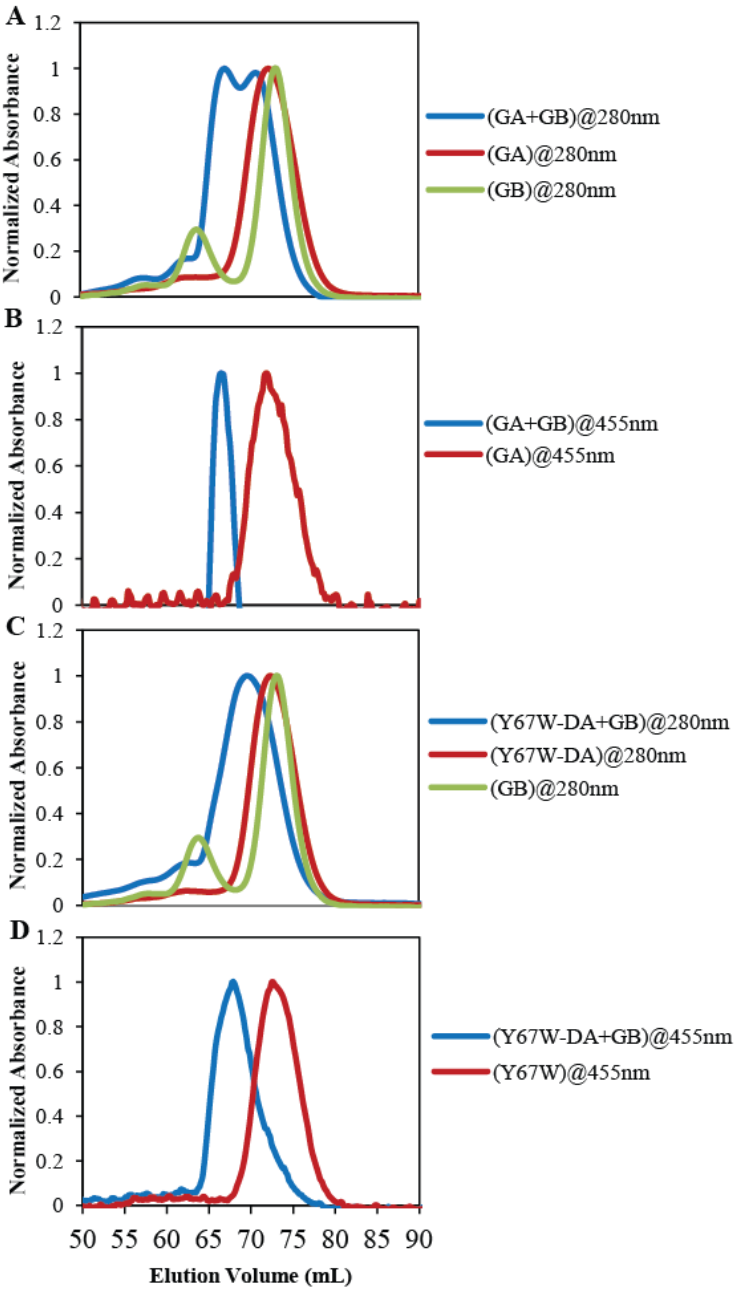
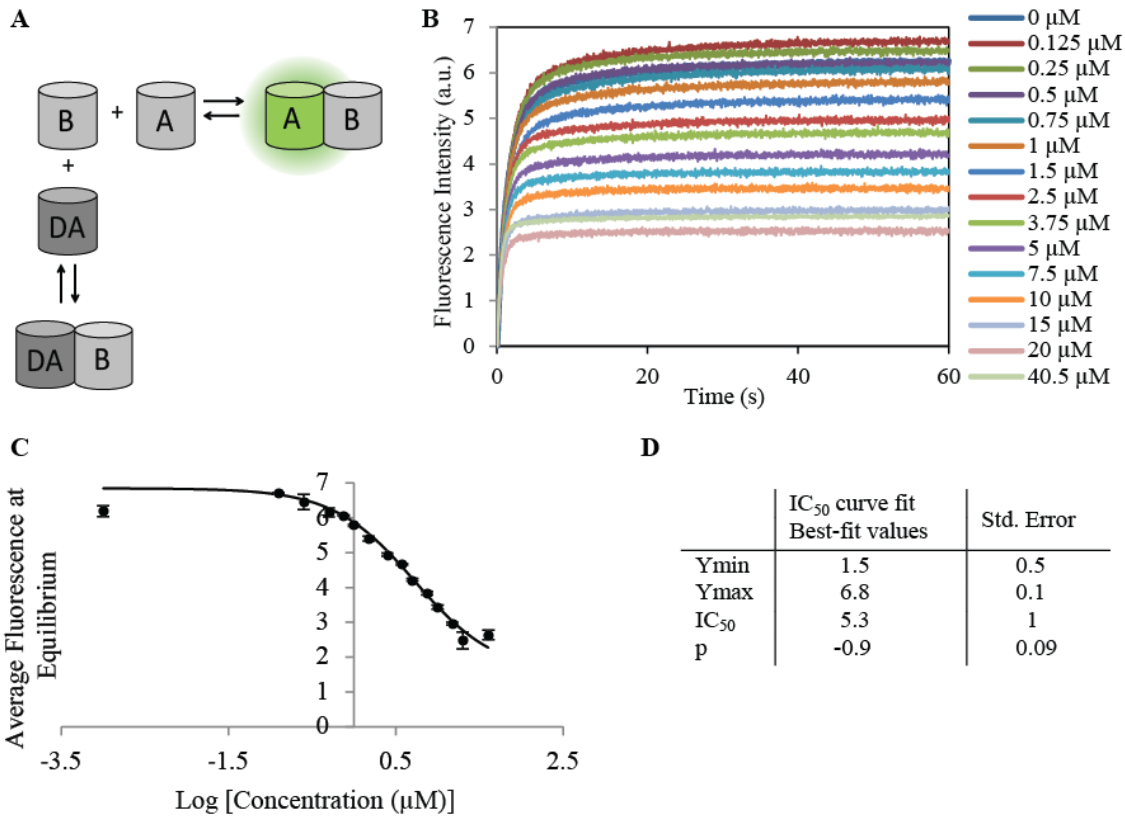


Figure 3.4. Determination of oligomerization state of Tyr67Trp-DA and GB by size exclusion chromatography. Designated proteins were resolved on a HiLoad 16/60 Superdex75 pg gel filtration column using 1× PBS, pH 7.4 at a flow rate of 1 mL/min. (A,B) Elution profiles of GA and GB (loaded at 125 µM each) at absorption detection at 280 nm (A) and 455 nm (B) were used as the oligomeric size standards. (C,D) Elution profiles of Y67W-DA and GB (loaded at 200 µM each) at absorption detection at 280 nm (C) and 455 nm (D). Note: all notations are standard one letter amino acid codes.



	IC_{50} curve fit Best-fit values	Std. Error
Ymin	1.5	0.5
Ymax	6.8	0.1
IC_{50}	5.3	1
p	-0.9	0.09

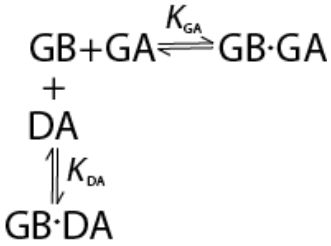
Figure 3.5. Determination of Tyr67Trp-DA IC_{50} using stopped-flow technique. (A) Competitive assay scheme where GA is abbreviated as A, GB as B and Y67W-DA as DA. (B) Raw kinetic traces were measured at a fixed GB concentration of 0.25 µM and GA concentration of 2.5 µM with Y67W-DA concentrations ranging from 0 µM to 40.5 µM in 1× PBS, pH 7.4. The reactions were excited at 455 nm and fluorescence was monitored via emission at 510 nm. Each trace is an average of 4 measurements. (B) An average of the maximal fluorescence intensity from the 50 to 60 second range was calculated and plotted against the logarithm of the Y67W-DA concentrations. Using Origin 2015 software, a variable slope dose-response curve (Eq. (1)) was used to fit the data. (C) Data yielded from B.

Of course in cases such as this spectroscopically silent Tyr67Trp-DA interaction, alternative methods such as isothermal titration calorimetry (ITC) are available to determine the binding constant and to elucidate the reaction mechanism. Indeed, ITC was initially attempted for characterization of the Tyr67Trp-DA interaction. However, we were not able to obtain the characteristic sigmoidal plots often obtained in ITC experiments. Accordingly, we could not extract a meaningful K_d value from our data (data not shown).

One likely reason to explain this negative result is that the heat released or used by the protein system was small relative to the heat of dilution (i.e., the heat associated with injection of the protein solution into the cell containing just the buffer). Subtraction of the heat of dilution obtained from a control run to the total heat of reaction which includes the heat of dilution and heat of protein interaction resulted in negligible results. Furthermore, the high protein concentration required to achieve an ideal c-value may have also led to small amounts of precipitation or aggregation. The c-value is a unitless parameter critical for determining the shape of the binding isotherm of which a c-value of 50 to 500 is ideal for measuring the binding constant [167]. The definition of the c-value is the product of the stoichiometry parameter, multiplied by the initial total macromolecule concentration in the cell, multiplied by the binding constant.

After finding that ITC was not suitable for our system, we tried to find other methods to determine the affinity of Tyr67Trp-DA to GB. Accordingly, we next pursued the use of a gel-shift assay. However, visualization of the Tyr67Trp-DA and GB heterodimer band was not observed in the gel along with the control GA and GB heterodimer band. Only until at very high protein concentrations of 100 μM were the heterodimer bands observed. This situation is similar to the one faced when obtaining the size exclusion column results where the heterodimer peaks were observed only at high concentrations. Furthermore, distinguishing between the monomer and heterodimer band was complicated due to their close proximity to each other. It was for these reasons that we pursued the competitive assay of GA and Tyr67Trp-DA binding to GB via stopped-flow techniques.

Although the IC_{50} of Tyr67Trp-DA has been obtained, it is not exactly representative or equivalent to the K_d for binding of Tyr67Trp-DA to GB. To determine the binding constants for this Tyr67Trp-DA and GB system, we conducted a more thorough analysis of the competitive binding of GA and Tyr67Trp-DA to GB. The competition can be described by the scheme:



The above terms in the above scheme are defined as follows, Tyr67Trp-DA is simply DA and the binding affinities are expressed by their dissociation constants K_{GA} and K_{DA}.

$$K_{GA} = \frac{[GB][GA]}{[GB*GA]} \quad \text{and} \quad K_{DA} = \frac{[GB][DA]}{[GB*DA]} \qquad \text{Equation (2A,B)}$$

The total amount of GB is

$$[GB]_0 = [GB] + [GB * GA] + [GB * DA] \qquad \text{Equation (3)}$$

[GB] and [GB*DA] are replaced by K_{GA} and K_{DA} in [3]:

$$[E]_0 = \frac{K_{GA}[GB * GA]}{[GA]}\left(1 + \frac{[DA]}{K_{DA}}\right) + [GB * GA] \qquad \text{Equation (4)}$$

Isolating for [GB*GA] gives the following expression:

$$[GB * GA] = \frac{[GB]_0[GA]}{[GA] + K_{GA}\left(1 + \frac{[DA]}{K_{DA}}\right)} \qquad \text{Equation (5)}$$

Rearranging this equation to give the fraction of GA bound (v) gives:

$$v = \frac{[GB * GA]}{[GB]_0} = \frac{[GA]}{[GA] + K'_{GA}} \quad \text{Equation (6)}$$

$$K'_{GA} = K_{GA}\left(1 + \frac{[DA]}{K_{DA}}\right) \quad \text{Equation (7)}$$

The double reciprocal relationship is:

$$\frac{1}{v} = 1 + \frac{K'_{GA}}{[GA]} \quad \text{Equation (8)}$$

For v, the fraction of GB saturated with GA varies from 0 to 1. Equation (6) is the Langmuir binding isotherm for our competition system.

We then measured the fluorescent intensity of the GB and GA heterodimer under increasing concentrations of Tyr67Trp-DA (**Figure 3.6A**). The double reciprocal of each trace from **Figure 3.6A** was taken to create the double reciprocal plot shown in **Figure 3.6B**. Applying the double reciprocal relationship from Equation (8), K'_{GA} was extracted from the slopes of each trace and plotted against the concentration of Tyr67Trp-DA to obtain the K_{DA} value of 4.16 ± 0.34 μM, the K_d of Tyr67Trp-DA to GB (**Figure 3.7**). This value is very similar to the K_d for binding of GA to GB of 3.1 ± 0.1 μM [5]. This indicates that the Trp mutation did not greatly perturb the interaction surface with GB.

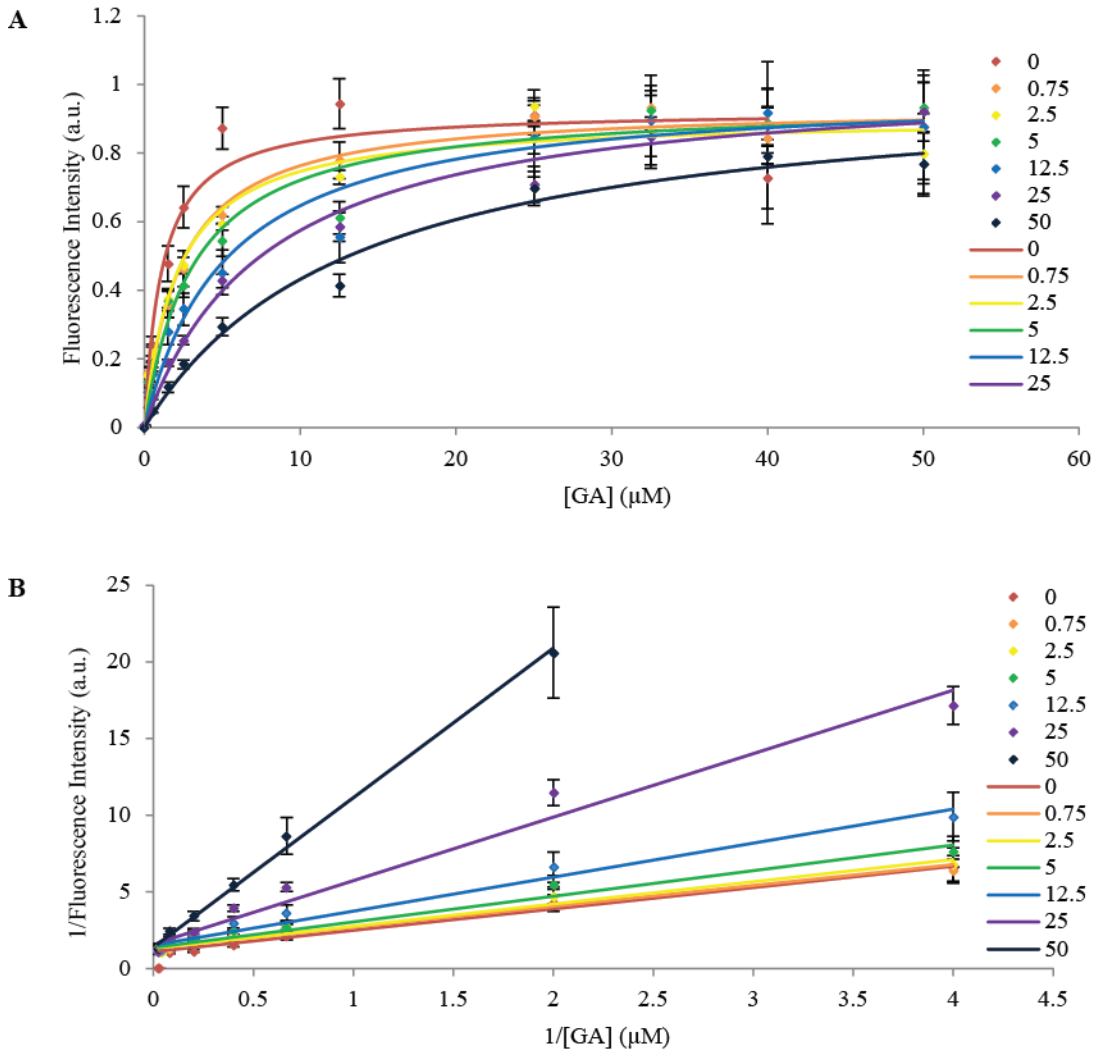
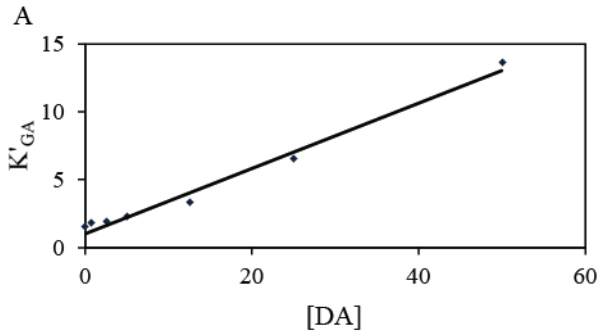


Figure 3.6. Competitive inhibition of GA+GB complex by Tyr67Trp-DA. (A) GB (fixed 0.25 μM for each inhibition trace) was saturated with GA (in increasing concentrations from 0, 0.25, 0.5, 1.5, 2.5, 5, 12.5, 25, 32.5, 40, 50 μM) in the presence of fixed concentrations of Y67W-DA (0, 0.75, 2.5, 5, 12.5, 25, and 50 μM). Each trace was fit with the Langmuir binding isotherm [6]. Three measurements were run for each data point. A GA blank with the corresponding concentration was subtracted from each run to remove the fluorescence contribution by GA. (B) Double reciprocal plot of (A) fit with the double reciprocal relationship [8].



B

Y-intercept	1.0 ± 0.3
Slope	0.2 ± 0.01
R^2	0.98
K_{DA}	4.2 ± 0.3

Figure 3.7. K_{DA} determination (A) Knowing K'_{GA} from **Figure 3.1** the K_{DA} can be calculated via equation [7] to create a K'_{GA} versus DA concentration plot. (B) Values calculated from data represented in (A).

3.2.2 Live cell applications

As a proof-of-concept experiment to demonstrate competitive binding of Tyr67Trp-DA and GA to GB, we undertook a cell-based assay where both A copies and GB would be introduced into the cell. To facilitate visualization of the change in the cell we introduced an initial bias for Tyr67Trp-DA by fusing it in tandem with GB with a caspase-3 substrate sequence, Aspartate-Glutamate-Valine-Aspartate (DEVD) in the linker region. The resulting caspase-3 biosensor construct was designated, DA-DEVD-GB. GA would be co-expressed throughout the cell along with mCherry with a plasma membrane localization signal acting as a non-participating reporter of transfection efficiency. All three components are expressed in one plasmid. Inclusion of inserted p2A sequences between the various components ensured separation of each at the protein sequence level.

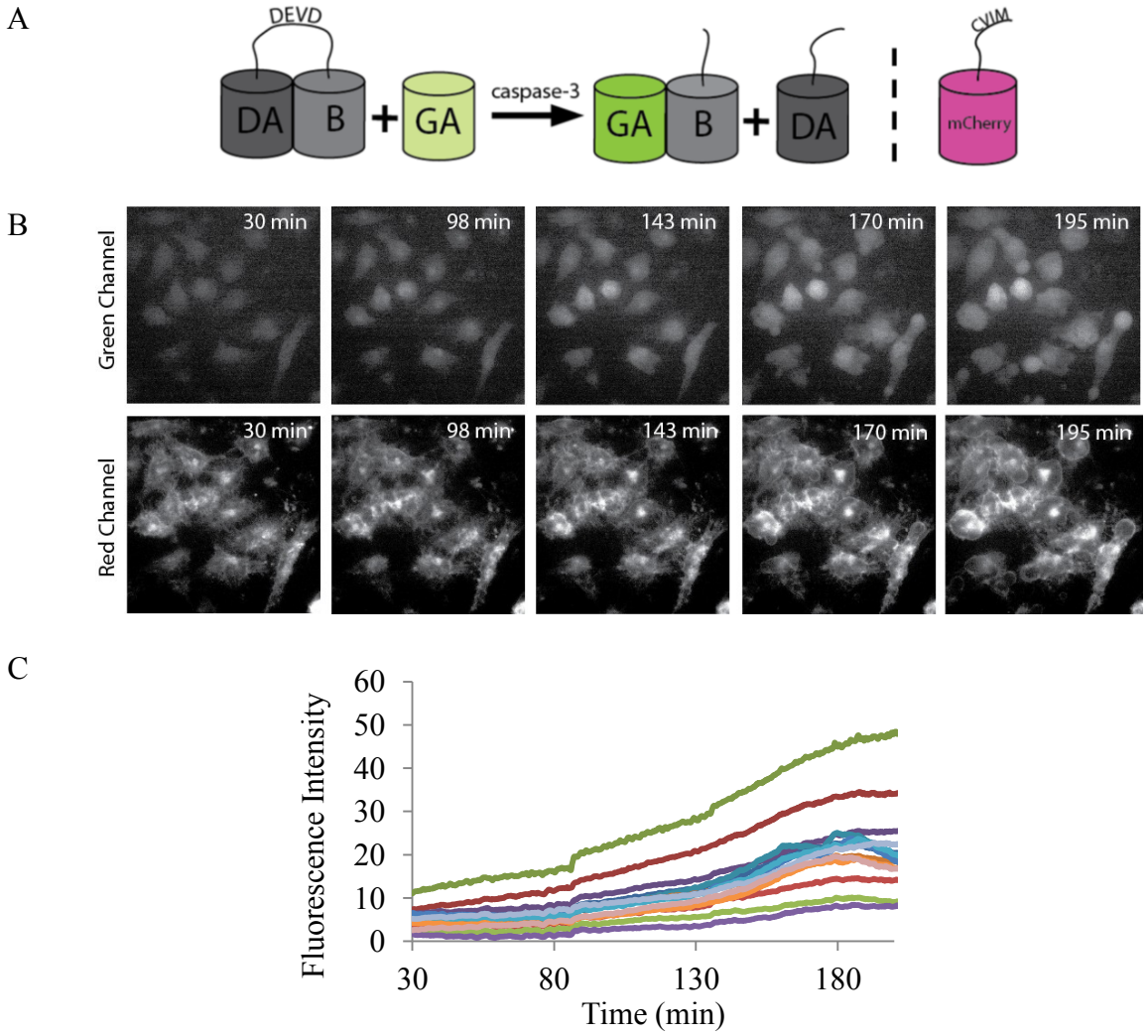


Figure 3.8. Imaging caspase-3 activity with ddGFP and Tyr67Trp-DA. (A) Schematic illustration of caspase-3 biosensor DA-DEVD-GB expressed with GA. Cleavage of the DEVD site allows GA to bind to GB. (B) Selected frames from imaging staurosporine-treated HeLa cells co-expressing DA-DEVD-GB, GA, and mCherry^{CaaX}. Images were collected from Axiovert 200M (Zeiss) equipped with a 20× objective lens. (C) Green fluorescence intensity versus time for the transfected HeLa cells shown in B.

This plasmid was transfected into HeLa cells which were then treated with staurosporine to induce caspase-3 activation and imaged through time [168]. With the initial predisposition of DA binding to GB the cells are initially "dark". Using the mCherry^{CaaX} reporter we were able to easily locate the transfected cells. Cleavage of the DEVD site released DA and GB and allowed GA to bind to the GB hence creating a

maximum of green fluorescence intensity upon cell shrinkage and blebbing which is indicative of apoptosis [169], as seen starting at 143 minutes in the representative images shown in **Figure 3.8B**. The fluorescence fold change was calculated to be, 4.9 ± 1.3 (N=14). We have created an intensiometric caspase-3 "turn on" biosensor using ddFPs without the need for translocation signals. The rate of transition from a dim to bright state is gradual as the onset of apoptosis begins (**Figure3.8 B,C**).

For experimental monitoring of caspase-3, this biosensor can still be improved. It is possible that even after cleavage, DA continues to compete with the free GA for binding to GB. This may explain the comparatively sub-optimal fluorescence detected during apoptosis relative to previous sensors employing ddGFP [137].

3.2.3 Discussion of Tyr67Trp-DA properties

At high concentrations of Tyr67Trp-DA and Tyr67Phe-DA, low intensity fluorescence emission observed (**Figure 3.3B, C**). This indicates that both these DA variants have intact chromophores. Exchange of Tyr66 in avGFP variants, which is homologous to Tyr67 in DsRed variants, to Phe or Trp resulted in variants with blue-shifted absorbance and fluorescence [9]. Substitution of Phe in avGFP at position 66 resulted in a variant with low quantum yield (QY) and an emission wavelength of 442 nm [6]. This observation is similar to the low fluorescence emission observed in the Tyr67Phe-DA variant. Evolving the Tyr66Phe avGFP variant followed by random mutagenesis resulted in the ultramarine FP named Sirius (λ_{abs} = 355 nm; λ_{em} = 424 nm) with efficient post-translational maturation in the intracellular environment at 37 °C and an improved QY [170]. The mutations were directed at the amino acids surrounding the chromophore and those that promoted chromophore maturation. If similar mutagenesis was attempted with

the Tyr67Phe-DA variant, improvement of the fluorescence intensity is likely. Also to note is the substitution of Phe at the Tyr67 position in a mCherry variant followed by directed evolution to generate mBlueberry (λ_{abs} = 398 nm; λ_{em} = 452 nm) [166]. The Tyr66Trp mutation in GFP resulted in the popular blue-shifted variant CFP [11]. Introduction of the Tyr67Trp along with five other mutations in close proximity to the chromophore of mRFP1.1 created mHoneydew (λ_{abs} = 487/504 nm; λ_{em} = 537/562 nm), a variant with broad yellow fluorescence [18]. It is interesting that in both these cases of the Tyr to Trp mutation strong fluorescence of the FP was observed while the Tyr67Trp-DA exhibited fluorescence only at high protein concentrations. This indicates that the interactions of the amino acids to the chromophore in mRFP1.1 are very different from the chromophore environment provided by the GA barrel.

Steric clash of the bulky Trp side chain with the side chains pointing into the chromophore cavity probably contributes to the extremely low QY observed for Tyr67Trp-DA. This is reminiscent of the initial introduction of Tyr66Trp mutation in GFP which led to creation of a blue variant with low QY [10]. Improvements to accommodate the bulk of Trp in this blue variant rescued the QY and led to the widely used ECFP variant and, most recently, the much brighter mTurquoise2 variant [11, 171].

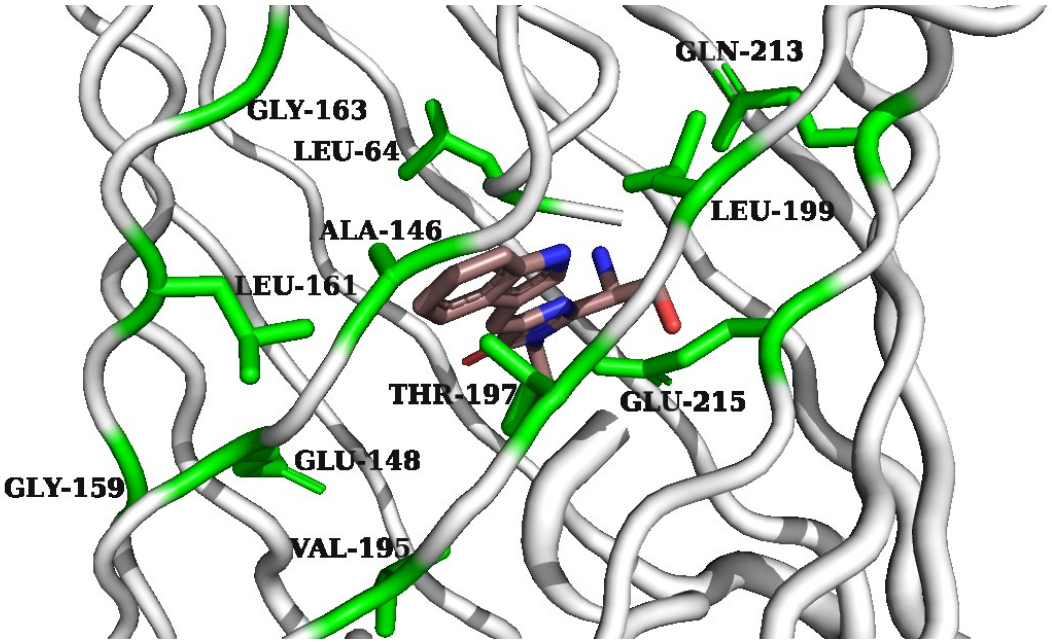


Figure 3.9. Estimation of the chromophore cavity environment of Tyr67Trp-DA. Aligning mTurquoise 2 (PDB entry 3ZTF) Ser-Trp-Gly chromophore with the DsRed barrel (PDB entry 1GGX) which has been modified to GA by substituting the amino acid mutations created the above model of Tyr67Trp-DA. The chromophore is coloured in tan. Coloured in green are amino acid residues of interest in close proximity to the chromophore. Blue represents Nitrogen and red represents oxygen.

For a rough estimation of the chromophore cavity environment in Tyr67Trp-DA, we implemented the GA mutations in a molecular model of the DsRed structure (PDB entry 1GGX [172]) and aligned the structure with mTurquoise2 (PDB entry 3ZTF [171]) as it contains a Ser-Trp-Gly chromophore. Although this is a slight deviation from the Tyr67Trp-DA chromophore of Met-Trp-Gly, we will focus principally on the Trp and its surrounding components. Analysis of this Tyr67Trp-DA model shows the immediate hydrophobicity of the Trp surrounding's with the exception of charged residues Thr197 and Glu215 directly below Trp (**Figure 3.9**). We postulate that Trp is not stabilized by hydrogen bonds due to the distance and orientation from these charged residues and thus the electron density is shifted to the imidazolinone moiety [173]. Surprisingly, besides Leu199, there does not seem to be large steric clash between Trp and its neighbouring

residues due to a number of residues pointing into the chromophore cavity being Gly and Ala. Comparison with the Met-Tyr-Gly chromophore of GA shows the Tyr is not rigidly packed which concurs with the previous suggestion that the GA chromophore exists in a protonated and non-coplanar conformations which favours excited state deactivation via radiationless decay [135]. Similar to GA, we predict that the Tyr67Trp-DA chromophore also exists largely in a non-coplanar conformation that explains its low QY.

Energy dissipation will occur via non-radiative pathways such as vibrational relaxation as the chromophore is not held rigidly. Due to necessary mutations to create ddFPs which resulted in a larger chromophore pocket and lack of hydrogen bonds, greater distortion of the chromophore can be accommodated [174, 175]. We postulate a large majority of the chromophore is in a *cis* configuration, as a *trans* configuration would lead to a steric collision between N-H on the Trp indole ring and the carbonyl group in the imidazolinone ring [174]. *Cis-trans* isomerisation as a mechanism for radiationless decay for this chromophore seems improbable considering the size of the Trp sidechain [176].

3.3 Conclusion

Through a single Tyr67Trp chromophore mutation we have engineered a “dark” A copy (DA) for FPX applications. Whether as a monomer or when associated with GB, little to no fluorescence emission is detected. Tyr67Trp-DA and GB exhibits a K_d similar to the K_d of GA and GB. We have demonstrated the possibility of DA in combination with other ddFPs as an intensiometric biosensor for protease activity monitoring. Although an intensiometric biosensor does not offer quantitative information like ratiometric biosensors, they offer an opportunity for multiparameter imaging. Thus far, intensiometric biosensors have been associated with single FPs biosensors or BiFC. Both of these strategies are

mainly applied to small molecule detection or PPI, rarely protease detection with the exception of a reported intein-based split Venus caspase-3 biosensor [177]. Nevertheless, if this DA intensiometric biosensor is to be widely used, the contrast between the "off" and the "on" state must be drastically improved.

3.4 Materials and Methods

3.4.1 General methods and materials.

All synthetic oligonucleotides (**Table 3.1**) were purchased from Integrated DNA Technologies (Coralville, IA). Pfu polymerase (ThermoScientific) and *Taq* polymerase (New England Biolabs) were used for standard PCR reactions. Small scale preparation of plasmid DNA, PCR products and products of restriction digests were purified by gel electrophoresis and extracted by spin column (GeneJET gel extraction kit; Thermo Scientific or BioBasic Inc.) according to the manufacturer's protocols. Restriction enzymes were purchased from Thermo Scientific or New England Biolabs. DNA manipulations such as polymerase chain reaction (PCR), ligation, DNA gel electrophoresis, restriction enzyme digestion, ligation and transformation were performed according to Sambrook *et al.*[139]. Ligations were transformed into electrocompetent DH10B *E. coli*. DNA sequencing was performed either at University of Alberta Molecular Service Unit (MBSU) or University of Calgary University Core DNA services.

Standard PCR amplifications were performed in 50 μL reaction volumes containing nuclease-free water, 1× reaction buffer, 200 μM dNTPs (Invitrogen), 3% DMSO, 200 nM forward and reverse oligonucleotides, 10-50 ng of template DNA, 1.0 unit of Pfu polymerase and 1.0 unit of Taq polymerase. Typical cycling parameters were as follow: initial denaturation at 95 °C for 60 seconds, 35 cycles of: denaturation at 95 °C for 30

seconds, 54-60 °C for 30 seconds, 72 °C for 60 seconds per kb of target; final extension at 72 °C for 5 minutes. Cloning strategies of the constructs described in this chapter are outlined in **Table 3.2**, which details the relevant oligonucleotides, DNA template, destination plasmids and restriction enzyme sites used.

3.4.2 Site-directed mutagenesis

Site-directed mutagenesis of residue 67 was completed by Quikchange Lightning kit (Agilent) following manufacturer's recommendations but at half the reaction volume. Typical cycling parameters were as follows: initial denaturation at 95 °C for 2 minutes; 30 cycles of : denaturation at 95 °C for 20 seconds, 58-60C for 30 seconds, 72 °C for 30 kb of plasmid target; final extension at 72 °C for 5 minutes. When the reaction was over, 0.5 µL of *DpnI* was added to the reaction mixture and incubated at 37 °C for 10 minutes and cooled at 12 °C. Propagation of plasmid DNA was done via addition of 1 µL reaction mixture into thawed electrocompetent DH10B *E. coli*. The mixture was then transferred into 0.2 cm MicroPulser cuvettes (Bio-Rad) and electroporated by a MicroPulser electroporator (Bio-Rad). Pulsed cells were then plated on agar plates containing 400 µg/mL ampicillin and 0.2% L-arabinose overnight at 37 °C. Plates were analyzed for fluorescence using a colony screener with appropriate filters (λ_{exc} = 470/40 nm; λ_{em} = 480/40 nm).

3.4.3 Protein Purification and Quantification

The genes already in pBAD/His B vector were transformed into electrocompetent DH10B *E. coli* and plated on 100 µg/mL ampicillin and 0.2% L-arabinose agar plates overnight at 37 °C. Colonies were then picked and subcultured in LB media. The volume of subculture to be prepared is in a 1:100 ratio to the volume of the large-scale culture media.

Terrific Broth (TB) was the media used following Cold Spring Harbor Online Protocol. Subculture was added into the TB broth along with ampicillin to a final concentration of 800 ppm. This was then incubated at 37 °C with shaking at 250 rpm until the optical density was at 0.5-0.7. At this point the culture was induced with 0.2% L-arabinose and allowed to incubate for 12 hours or up to 3 days (depending on culture volume and needs of the experiment) at 28 °C with shaking at 250 rpm. Culture was then pelleted at 10 000 rpm for 10 minutes. Cell pellets were then resuspended in 1× PBS (137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4) at pH 7.4 and lysed using a cell disrupter (Constant Systems Ltd.). Cleared lysates were obtained by centrifugation at 8,000 rpm for 40 minutes. Lysates were incubated with Nickel-NTA resin on ice with agitation for 1 hour and then batch purified through a fritted column using a vacuum manifold. At least 10 column volumes of 1× PBS with 20 mM imidazole to a final pH of pH 8.0 was used to wash the protein-bound resin at a slow flow rate of approximately 20 drops per minute. Proteins were eluted using filtered 1× PBS with 250 mM imidazole at pH 8.0.

Purity assessment of the purified protein was done by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) with a (5% stacking gel/12% resolving gel) according to reported methods [178]. Protein samples were prepared with 1× sample buffer (62.5 mM Tris-Cl pH 6.8, 2% SDS, 7.5% glycerol, 0.02% bromophenol blue). Protein bands were identified by staining with a standard coomassie stain and subsequent washing with 10% acetic acid.

Protein concentration was determined via bicinchoninic acid (BCA) protein assay (Pierce) according to manufacturer's microplate format protocols in 96-well plates

(Corning). Absorbance values of developed assay were read at 562 nm by the TECAN Safire^2 microplate reader.

According to the requirements of the experiment, proteins were concentrated using Amicon Ultra-4 mL or Ultra-15 mL Centrifugal Filters (EMD Millipore).

3.4.4 Determination of secondary structure

Circular Dichroism experiments were performed on an Olis DSM 17 Circular Dichroism. The far UV spectra (195-250 nm) were obtained in a 0.2 mm path length quartz cuvette and filled with 100 μL of protein sample at a concentration of approximately 1 mg/mL in 1× PBS, pH 7.4. Five spectra with a resolution of 1 nm were accumulated and averaged. The temperature of all experiments was 293 K. All samples were freshly prepared prior to measurements.

3.4.5 Determination of heterodimer formation

To determine dimerization, purified proteins were resolved by gel filtration chromatography. Proteins were resolved over a HiLoad 16/60 Superdex 75 pg gel filtration column using an AKTA basic liquid chromatography system (GE Healthcare) with a multiple wavelength absorbance detector. Proteins were resolved with filtered and degassed 1× PBS, pH 7.4 mobile phase and 1.0 mL/min flow rate. Eluted fractions were monitored at 280 nm and 455 nm.

3.4.6 Determination of IC_{50}

A SX20 stopped-flow spectrometer (Applied Photophysics) was used to measure the kinetic traces to obtain the IC_{50} value. Excitation was set at 455 nm and emission detection at 510 nm. 2000 points were read for each run which lasted 60 seconds each. Proteins samples of a fixed concentration of GB was rapidly mixed at room temperature

with a mixture containing a concentration of Tyr67Trp-DA which concentrations' are varied between runs and fixed concentrations of GA. Please note that the final concentrations of the protein samples when mixed are halved hence the reported concentrations in **Figure 3.5** reflect this. Each measurement was done four times and then averaged to obtain the final kinetic trace.

3.4.7 Determination of K_d

K_d determination was completed via a competition assay. Each measurement was run with GB, Tyr67Trp-DA and GA at their corresponding concentrations for each well in a 384 well plate (Nunc) to a final volume of 40 μL and data was read with a Tecan Safire^2 microplate reader. Excitation was set at 488 nm and emission at 510 nm. Each data point was subtracted against its corresponding GA emission blank. All runs were completed three times each.

3.4.8 Live cell imaging applications construction of mammalian expression plasmid co-expressing DA-DEVD-GB, GA and mCherry^{CaaX}

DA and GB were linked by the sequence SGTASGDEVDGATI to create the DA-DEVD-GB construct. This was PCR amplified with a 3' primer which appended a *Mlu*I site to it. This construct was then ligated into a modified pBAD vector encoding pBAD-*Xho*I-EGFP-*Mlu*I-p2Ap2A-*Xba*I-*Sal*I-p2Ap2A-*EcoR*I-*Hind*III (p2A, GSGATNFSLLKQAGDVEENPGP). GA template was PCR amplified to have a *Xba*I site on its 5' end and *Sal1* site on its 3' end and then digested and ligated into the above vector containing DA-DEVD-GB but with EGFP between the *Xba*I/*Sal*I sites. This gene sequence now containing *Xho*I-DA-DEVD-GB-*Mlu*I-p2Ap2A-*Xba*I-GA-*Sal*I-p2Ap2A-*EcoR*I-HindIII was transferred to a digested *Xho*I/*Hind*III pcDNA3.1(+) vector. The plasma

membrane localization sequence (KMSKDGKKKKKKSKTKCVIM), a modified version of the KRAS membrane localization sequence favouring greater palmitoylation, was appended at the 3' end of the mCherry sequence with a *Hind*III site and an *EcoR1* site on its 5' end.

3.4.9 Mammalian cell culture and imaging

HeLa cells were maintained in complete media (DMEM with 10% fetal bovine serum (FBS), 1% Penicillin Streptomycin (Pen Strep) (Invitrogen), 5% Glutamax (Invitrogen)) at 37 °C and 5% CO_2 according to standard procedure. Transient transfections of pcDNA3.1 (+) expression plasmids were performed using Turbofect (Thermo Scientific). HeLa cells were grown on 35 mm imaging dishes and grown to approximately 70-80% confluency and incubated with 1 mL of DMEM (FBS free) for 10 min then transfection reagent mixture was added. Generally, 2 μL of transfection reagent was used per 1 μg of DNA transfected and mixed in 0.2 mL of DMEM (FBS free). After 2 hours, the culture media was changed back to complete media. 24 to 48 hours after post-transfection, staurosporine was added to the transfected cells to a final concentration of 2 μM and incubated for 30 min. The media was then changed to HEPES-buffered Hank's balanced salt solution (HHBSS) and the cells were subjected to imaging at 1 to 2 min intervals for 4-6 hours. Imaging was completed using an Axiovert 200M (Zeiss) microscope. Axiovert 200M (Zeiss) was equipped with a 75W Xenon-arc lamp, 20× objective lens and a 14-bit CoolSnap HQ2 cooled CCD camera (Photometrics), driven by open source Micro-Manager software.

Table 3.1 Details of the gene construct used in this work.

Name	Gene (Restriction sites in italics and substrate sequence in bold)	Comments
DA-DEVD-GB,GA-mCherry^{CaaX}	pcDNA-*Xho*I-DA-*Kpn*I-**DEVD**-GB-*Mlu*I-p2Ap2A-*Xba*I-GA-*Sal*I-p2Ap2A-*EcoR*I-mCherry^{CaaX}-*Hind*III	All assembly was completed by ligation

Cloning Strategy		
Construct	Oligos (Fwd/Rv)	Target Plasmid
*Xho*I-DA-*Kpn*I	3.1/3.2	pBAD-*Xho*I-EGFP-*Mlu*I-p2Ap2A-*Xba*I-*Sal*I-p2Ap2A-*EcoR*I-*Hind*III
*Kpn*I-DEVD-GB-*Mlu*I	3.3,3.4/3.5	
*Xba*I-EGFP-*Sal*I	3.6/3.7	pBAD-*Xho*I-DA-DEVD-GB-*Mlu*I-p2Ap2A-*Xba*I-*Sal*I-p2Ap2A-*EcoR*I-*Hind*III
*Xba*I-GA-*Sal*I	3.6/3.8	pBAD-*Xho*I-DA-DEVD-GB-*Mlu*I-p2Ap2A-*Xba*I-EGFP-*Sal*I-p2Ap2A-*EcoR*I-*Hind*III
*EcoR*I-mCherry^{CaaX}-*Hind*III	3.9/3.10,3.11	pcDNA-*Xho*I-DA-DEVD-GB-*Mlu*I-p2Ap2A-*Xba*I-GA-*Sal*I-p2Ap2A-*EcoR*I-*Hind*III

3.4.10 Image Analysis

The images were assembled using Adobe Illustrator and analyzed using freeware ImageJ.

Chapter 4: Conclusions and future directions

4.1 Summary of thesis

The initial implementation of FPs as genetically encoded reporters kick-started a revolution in molecular biosensing [2]. Now, scientists can use a wide variety of FP-based reporters to observe the inner workings of a live cell non-invasively in real-time. The continuous development of improved biosensing platforms for detection of PPIs, enzymatic activity, enzyme functions, and protein interaction with small molecules, has brought innovative solutions to the table. These new technologies continue to make up for the insufficiencies of more traditional technologies such as ones based on FP-based FRET biosensors and BiFC. The BiFC design itself has extended beyond the visible range spectrum and into the near-infrared range to facilitate *in vivo* imaging [39, 78]. It has also been creatively modified and applied for detection of RNA, genome-wide PPI screening, drug discovery and superresolution microscopy [49, 65, 69, 106]. Development of novel FP complementation platforms such a tripartite split GFP, alternate β-strand switch FP, and ddFPs has enabled a broader breadth of detection of cellular function [128, 133, 134]. Nevertheless, despite these great strides in biosensing tools, there is still a lot room to improve their utility and make them more appealing to a broader audience of researchers who would benefit from implementing these tools in their research.

Limitations of the available biosensing strategies (i.e., FP-based FRET, BiFC and single FP-based biosensors), precipitated the conception of ddFPs. DdFP is a recently developed novel biosensing strategy which exploits the obligate oligomerization of DsRed [134]. Through iterative rounds of evolution and interface modifications, a low affinity heterodimeric ddRFP was generated. The FPs in this system are individually named A copy

and B copy and association of the two leads to greater fluorescence emission of the A copy. Very low fluorescence is observed when the two copies are apart. Later expansion of the colour palette resulted in yellow and green hued ddFP variants [136].

Each ddFP colour variant was individually evolved with their respective B copies. However, knowledge of the fact that the B copies can be interchanged between RA, GA and YA and still confer an increase in fluorescence resulted in FPX applications of this system [5]. Thus far, the FPX strategy has been applied for detection of Ca^{2+} dynamics, caspase-3 activity, PIP_2 hydrolysis, PKA activation and ERK activity. We endeavour to increase the applications of ddFPs via the FPX approach in order to demonstrate the versatility of this system. To this end, I have designed OGT glycosylation and protease activity sensors and a tomosyn-1 SUMOylation sensor. In addition, a 'dark' accessory protein was developed for ddFPs.

In chapter 2, I described the construction of three plasmids for sensing OGT glycosylation and protease activity and the SUMOylation of tomosyn-1. During the construction of these plasmids we were challenged with constructing some relatively large DNA constructs. These efforts were further complicated by a number of restriction enzyme sites present in both the gene of interest (i.e., tomosyn-1) and the target plasmid backbone and construct. Furthermore, to minimize the possibility of variable expression of multiple transfection products between cells, the biosensing systems involving more than one construct were inserted into a vector harbouring p2A sequences so that expression of each construct would be stoichiometric. Small details pertaining to the OGT glycosylation and SUMOylation of tomosyn-1 still remain to be fixed before the application of these plasmids can be attempted in live cells. We endeavour to work towards this goal.

Beyond the direct applications of new ddFP-based biosensors in live cells, we created a dark A copy accessory protein for ddFPs we named as DA. We successfully characterized the DA to reveal it exists predominantly as a monomer and its K_d is 4.16 ± 0.34 μM. Live cell applications demonstrated the potential of the implementation of DA with GA and GB to create an intensiometric "turn-on" caspase-3 biosensor. Further modifications of this sensor are necessary if it is to be made practical for live-cell imaging purposes.

4.2 Future directions

With the expansion of ddFP applications to the detection of global modification events such as O-GlcNAcylation and SUMOylation we hope to demonstrate the practicality of ddFPs in the study of reversible phenomenon in live cells. Of course, if ddFPs are to be embraced by the scientific community, then the brightness and the contrast between the "off" and "on" state must be significantly improved to facilitate straightforward detection. As of now, the DsRed crystal structure can still be marginally used for the identification and prediction of rational mutation sites that may bestow variants with improved brightness. Obtaining the crystal structures of all three ddFP variants is imperative for further enhancement of this biosensing strategy. Immediate methods to improve the brightness and contrast are to increase the solubility of the B copy and stability of the A copy. From our observations, B copy has a tendency to precipitate after 2-3 days. A copy undergoes chromophore hydrolysis to create a 19 kDa and 11 kDa band when run in a 10% SDS PAGE [135]. Increasing the solubility increases the stability of the FP and hence greater fluorescence.

Alternative methods to enhance the brightness and contrast is to begin with a new template such as the tetrameric LanYFP (λ_{ex} = 513 nm; λ_{em} = 524 nm) derived from the cephalochordate *Branchiostoma lanceolatum* [179]. LanYFP has exceptional optical properties, a high QY (0.95) and extinction coefficient (150,000 $M^{-1}cm^{-1}$), making it an attractive candidate for future generations of ddFP [179]. The brightness exceeds that of the brightest YFP variant, YPet (ϕ = 0.77, ε = 104,000 M^{-1} cm^{-1}) [24]. Structure prediction software has predicted LanYFP's closest tetrameric structural relative to be a cyan fluorescent protein from *Discosoma* sp., the same coral of which DsRed is derived from [179]. Moreover, using the engineering of mRFP1 as a guide, a dimeric LanYFP possessing near identical optical properties as its parent was achieved. In addition, a monomeric green variant, mNeonGreen (λ_{ex} = 506 nm; λ_{em} = 517 nm), has also been created with reduced brightness that is still approximately three times greater than mEGFP. This is all very encouraging. Appropriating LanYFP as a ddFP certainly guarantees some loss of brightness. However, with respect to the case of mNeonGreen, we can estimate that the brightness will remain above or close to that of popularly used FPs.

Meanwhile, it is foreseeable that the expansion of ddFP and its FPX strategy is to follow that of BiFC's. Many of the BiFC designs can be easily moulded for ddFP purposes, for example, TriFC for RNA detection using Pum-HD. ddFPs are also a powerful tool for high-throughput screening of PPIs and for drug discovery. Their high affinity (< 10 μM) for each and residual fluorescence of the A copy may complicate screening [5]. However, BiFC which is reversible has already been used in mammalian cells and coupled with FACs to identify new regulators for telomere signalling in human cells [102]. Taking this as a precedent and the fact that ddFPs are reversible, ddFPs can easily be modified for a

simplified MaMTH screening for integral membrane PPI without the need for a translocation of a transcription factor into the nucleus to initiate a response. As for ddFP application for superresolution microscopy, the ddFP must firstly be modified to be photoactivatable and have a high photon budget. It might be more practical to create a ddFP from a new FP template already possessing photoswitchable characteristics such as the tetrameric EosFP ($\lambda_{max,\ green}$, ex/em = 506/516 nm; $\lambda_{max,\ red}$ ex/em = 571/581 nm) exhibiting green-to-red photoconversion [180].

The DA was created for further extension of ddFP applications. Thus far, in a preliminary proof-of-concept demonstration, DA was used in combination with GA and GB for the sensing of caspase-3 activation in live cells. A gradual dark to green fluorescence was observed in the cells over time. We have yet to test this system with a control design with a mutated cleavage site. We envision similar application of DA for the sensing of other protease activities in combination with other FPs for true multiparameter imaging in live cells.

4.3 Concluding remarks

The research presented in this thesis described our efforts to expand the applications of ddFPs and the development of the DA for ddFP applications. We were able to design and construct biosensors for the detection of OGT glycosylation and protease activity and the SUMOylation of tomosyn-1. Minor changes are necessary for some constructs before live cell applications. Through a simple single mutation at the chromophore of the GA were able to successfully create the DA which does form an intact chromophore but has very low emission intensity. The DA was created to competitively bind with the A copy for the B copy. In doing so, we aimed to reduce non-specific binding of A copy for the B copy when

not undergoing PPI, in this way, contrast between the "on" and "off" state for the ddFPs may be improved. Development of the DA also led to the creation of an intensiometric "turn on" biosensor for protease activity sensing without the need for translocation of any of the FP components into a specific cellular compartment. We desire to further improve this preliminary application of DA and to conduct multiparameter imaging with this biosensor to simultaneously detect multiple cellular events. Many exciting opportunities for ddFPs still await. We look forward to the firm establishment of ddFPs in every molecular biologist's toolbox.

References

1. Shimomura, O.; Johnson, F. H.; Saiga, Y., Extraction, purification and properties of aequorin, a bioluminescent protein from the luminous hydromedusan, Aequorea. *J. Cell. Physiol.* **1962,** *59* (3), 223-239.

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

3. Prasher, D. C.; Eckenrode, V. K.; Ward, W. W.; Prendergast, F. G.; Cormier, M. J., Primary structure of the Aequorea victoria green-fluorescent protein. *Gene* **1992,** *111* (2), 229-233.

4. Sample, V.; Mehta, S.; Zhang, J., Genetically encoded molecular probes to visualize and perturb signaling dynamics in living biological systems. *J. Cell Sci.* **2014,** *127* (6), 1151-1160.

5. Ding, Y.; Li, J.; Enterina, J. R.; Shen, Y.; Zhang, I.; Tewson, P. H.; Mo, G. C.; Zhang, J.; Quinn, A. M.; Hughes, T. E., Ratiometric biosensors based on dimerization-dependent fluorescent protein exchange. *Nat. Methods* **2015,** *12* (3), 195-198.

6. Cubitt, A. B.; Heim, R.; Adams, S. R.; Boyd, A. E.; Gross, L. A.; Tsien, R. Y., Understanding, improving and using green fluorescent proteins. *Trends Bioch. Sci.* **1995,** *20* (11), 448-455.

7. Yang, F.; Moss, L. G.; Phillips Jr, G. N., The molecular structure of green fluorescent protein. *Nat. Biotechnol.* **1997,** *14 (10), 1246-1251.*

8. Ormö, M.; Cubitt, A. B.; Kallio, K.; Gross, L. A.; Tsien, R. Y.; Remington, S. J., Crystal structure of the Aequorea victoria green fluorescent protein. *Science* **1996,** *273* (5280), 1392-1395.

9. Tsien, R. Y., The green fluorescent protein. *Annu. Rev. Biochem.* **1998,** *67* (1), 509-544.

10. Heim, R.; Prasher, D. C.; Tsien, R. Y., Wavelength mutations and posttranslational autoxidation of green fluorescent protein. *P. Natl. Acad. Sci.* **1994,** *91* (26), 12501-12504.

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

12. Shen, Y.; Lai, T.; Campbell, R. E., Red fluorescent proteins (RFPs) and RFP-based biosensors for neuronal imaging applications. *Neurophotonics* **2015,** *2* (3), 031203-031203.

13. Shcherbakova, D. M.; Subach, O. M.; Verkhusha, V. V., Red fluorescent proteins: advanced imaging applications and future design. *Angew. Chem. Int. Edit.* **2012,** *51* (43), 10724-10738.

14. Matz, M. V.; Fradkov, A. F.; Labas, Y. A.; Savitsky, A. P.; Zaraisky, A. G.; Markelov, M. L.; Lukyanov, S. A., Fluorescent proteins from nonbioluminescent Anthozoa species. *Nat. Biotechnol.* **1999,** *17* (10), 969-973.

15. Baird, G. S.; Zacharias, D. A.; Tsien, R. Y., Biochemistry, mutagenesis, and oligomerization of DsRed, a red fluorescent protein from coral. *P. Natl. Acad. Sci.* **2000,** *97* (22), 11984-11989.

16. Mizuno, H.; Sawano, A.; Eli, P.; Hama, H.; Miyawaki, A., Red fluorescent protein from Discosoma as a fusion tag and a partner for fluorescence resonance energy transfer. *Biochemistry* **2001,** *40* (8), 2502-2510.

17. Campbell, R. E.; Tour, O.; Palmer, A. E.; Steinbach, P. A.; Baird, G. S.; Zacharias, D. A.; Tsien, R. Y., A monomeric red fluorescent protein. *P. Natl. Acad. Sci.* **2002,** *99* (12), 7877-7882.

18. Shaner, N. C.; Campbell, R. E.; Steinbach, P. A.; Giepmans, B. N.; Palmer, A. E.; Tsien, R. Y., Improved monomeric red, orange and yellow fluorescent proteins derived from Discosoma sp. red fluorescent protein. *Nat. Biotechnol.* **2004,** *22* (12), 1567-1572.

19. Shaner, N. C.; Lin, M. Z.; McKeown, M. R.; Steinbach, P. A.; Hazelwood, K. L.; Davidson, M. W.; Tsien, R. Y., Improving the photostability of bright monomeric orange and red fluorescent proteins. *Nat. Methods* **2008,** *5* (6), 545-551.

20. Oldach, L.; Zhang, J., Genetically Encoded Fluorescent Biosensors for Live-Cell Visualization of Protein Phosphorylation. *Chem. Biol.* **2014,** *21* (2), 186-197.

21. Ibraheem, A.; Campbell, R. E., Designs and applications of fluorescent protein-based biosensors. *Curr. Opin.Chem. Biol.* **2010,** *14* (1), 30-36.

22. Campbell, R. E., Fluorescent-protein-based biosensors: modulation of energy transfer as a design principle. *Anal. Chem.* **2009,** *81* (15), 5972-5979.

23. Alford, S. C.; Wu, J.; Zhao, Y.; Campbell, R. E.; Knöpfel, T., Optogenetic reporters. *Biol. Cell* **2013,** *105* (1), 14-29.

24. Nguyen, A. W.; Daugherty, P. S., Evolutionary optimization of fluorescent proteins for intracellular FRET. *Nat. Biotechnol.* **2005,** *23* (3), 355-360.

25. Rizzo, M. A.; Springer, G. H.; Granada, B.; Piston, D. W., An improved cyan fluorescent protein variant useful for FRET. *Nat. Biotechnol.* **2004,** *22* (4), 445-449.

26. Jensen, K. K.; Martini, L.; Schwartz, T. W., Enhanced fluorescence resonance energy transfer between spectral variants of green fluorescent protein through zinc-site engineering. *Biochemistry* **2001,** *40* (4), 938-945.

27. Komatsu, N.; Aoki, K.; Yamada, M.; Yukinaga, H.; Fujita, Y.; Kamioka, Y.; Matsuda, M., Development of an optimized backbone of FRET biosensors for kinases and GTPases. *Mol. Biol. Cell* **2011,** *22* (23), 4647-4656.

28. Ear, P. H.; Michnick, S. W., A general life-death selection strategy for dissecting protein functions. *Nat. Methods* **2009,** *6* (11), 813-816.

29. Stynen, B.; Tournu, H.; Tavernier, J.; Van Dijck, P., Diversity in genetic in vivo methods for protein-protein interaction studies: from the yeast two-hybrid system to the mammalian split-luciferase system. *Microbiol. Mol. Biol. R.* **2012,** *76* (2), 331-382.

30. Johnsson, N.; Varshavsky, A., Split ubiquitin as a sensor of protein interactions *in vivo*. *P. Natl. Acad. Sci.* **1994,** *91* (22), 10340-10344.

31. Petschnigg, J.; Groisman, B.; Kotlyar, M.; Taipale, M.; Zheng, Y.; Kurat, C. F.; Sayad, A.; Sierra, J. R.; Usaj, M. M.; Snider, J., The mammalian-membrane two-hybrid assay (MaMTH) for probing membrane-protein interactions in human cells. *Nat. Methods* **2014,** *11* (5), 585-592.

32. Michnick, S. W.; Ear, P. H.; Manderson, E. N.; Remy, I.; Stefan, E., Universal strategies in research and drug discovery based on protein-fragment complementation assays. *Nat. Rev. Drug Discov.* **2007,** *6* (7), 569-582.

33. Tarassov, K.; Messier, V.; Landry, C. R.; Radinovic, S.; Molina, M. M. S.; Shames, I.; Malitskaya, Y.; Vogel, J.; Bussey, H.; Michnick, S. W., An in vivo map of the yeast protein interactome. *Science* **2008,** *320* (5882), 1465-1470.

34. Galarneau, A.; Primeau, M.; Trudeau, L.-E.; Michnick, S. W., β-Lactamase protein fragment complementation assays as in vivo and in vitro sensors of protein–protein interactions. *Nat. Biotechnol.* **2002,** *20* (6), 619-622.

35. Remy, I.; Michnick, S. W., Clonal selection and in vivo quantitation of protein interactions with protein-fragment complementation assays. *P. Natl. Acad. Sci.* **1999,** *96* (10), 5394-5399.

36. Kaihara, A.; Kawai, Y.; Sato, M.; Ozawa, T.; Umezawa, Y., Locating a protein-protein interaction in living cells via split Renilla luciferase complementation. *Anal. Chem.* **2003,** *75* (16), 4176-4181.

37. Ghosh, I.; Hamilton, A. D.; Regan, L., Antiparallel leucine zipper-directed protein reassembly: application to the green fluorescent protein. *J. Am. Chem. Soc.* **2000,** *122* (23), 5658-5659.

38. Welsh, D. K.; Noguchi, T., Cellular bioluminescence imaging. *Cold Spring Harb. Protoc.* **2012,** *2012*.

39. Hu, C.-D.; Kerppola, T. K., Simultaneous visualization of multiple protein interactions in living cells using multicolor fluorescence complementation analysis. *Nat. Biotechnol.* **2003,** *21* (5), 539-545.

40. MacDonald, M. L.; Lamerdin, J.; Owens, S.; Keon, B. H.; Bilter, G. K.; Shang, Z.; Huang, Z.; Yu, H.; Dias, J.; Minami, T., Identifying off-target effects and hidden phenotypes of drugs in human cells. *Nat. Chem. Biol.* **2006,** *2* (6), 329-337.

41. Hu, C.-D.; Chinenov, Y.; Kerppola, T. K., Visualization of interactions among bZIP and Rel family proteins in living cells using bimolecular fluorescence complementation. *Mol. Cell* **2002,** *9* (4), 789-798.

42. Shyu, Y. J.; Hu, C.-D., Fluorescence complementation: an emerging tool for biological research. *Trends Biotechnol.* **2008,** *26* (11), 622-630.

43. Mervine, S. M.; Yost, E. A.; Sabo, J. L.; Hynes, T. R.; Berlot, C. H., Analysis of G protein βγ dimer formation in live cells using multicolor bimolecular fluorescence complementation demonstrates preferences of β1 for particular γ subunits. *Mol. Pharmacol.* **2006,** *70* (1), 194-205.

44. Shyu, Y. J.; Liu, H.; Deng, X.; Hu, C., Identification of new fluorescent protein fragments for biomolecular fluorescence complementation analysis under physiological conditions. *Biotechniques* **2006,** *40* (1), 61.

45. Nagai, T.; Ibata, K.; Park, E. S.; Kubota, M.; Mikoshiba, K.; Miyawaki, A., A variant of yellow fluorescent protein with fast and efficient maturation for cell-biological applications. *Nat. Biotechnol.* **2002,** *20* (1), 87-90.

46. Griesbeck, O.; Baird, G. S.; Campbell, R. E.; Zacharias, D. A.; Tsien, R. Y., Reducing the environmental sensitivity of yellow fluorescent protein mechanism and applications. *J. Biol. Chem.* **2001,** *276* (31), 29188-29194.

47. Shyu, Y. J.; Suarez, C. D.; Hu, C.-D., Visualization of AP-1–NF-κB ternary complexes in living cells by using a BiFC-based FRET. *P. Natl Acad. Sci.* **2008,** *105* (1), 151-156.

48. Mo, S.; Chiang, S.; Lai, T.; Cheng, Y.; Chung, C.; Kuo, S.; Reece, K.; Chen, Y.; Chang, N.; Wadzinski, B., Visualization of Subunit Interactions and Ternary Complexes of Protein Phosphatase 2A in Mammalian Cells. *PLOS ONE.* **2014,** *9* (12), e116074.

49. Miller, K. E.; Kim, Y.; Huh, W.-K.; Park, H.-O., Bimolecular fluorescence complementation (BiFC) analysis: Advances and recent applications for genome-wide interaction studies. *J. Mol. Biol.* **2015**.

50. Parsons, M. J.; Rehm, M.; Bouchier-Hayes, L., Imaging-Based Methods for Assessing Caspase Activity in Single Cells. *Cold Spring Harb. Protoc.* **2015,** *2015* (1), pdb. top070342.

51. Cormack, B. P.; Valdivia, R. H.; Falkow, S., FACS-optimized mutants of the green fluorescent protein (GFP). *Gene* **1996,** *173* (1), 33-38.

52. Crameri, A.; Whitehorn, E.; Tate, E.; Stemmer, W.; Crameri, A.; Kitts, P.; Kitts, P., Improved green fluorescent protein by molecular evolution using. *Nat. Biotechnol.* **1996,** *14* (3), 315-319.

53. Pédelacq, J.-D.; Cabantous, S.; Tran, T.; Terwilliger, T. C.; Waldo, G. S., Engineering and characterization of a superfolder green fluorescent protein. *Nat. Biotechnol.* **2006,** *24* (1), 79-88.

54. Cabantous, S.; Terwilliger, T. C.; Waldo, G. S., Protein tagging and detection with engineered self-assembling fragments of green fluorescent protein. *Nat. Biotechnol.* **2005,** *23* (1), 102-107.

55. Feinberg, E. H.; VanHoven, M. K.; Bendesky, A.; Wang, G.; Fetter, R. D.; Shen, K.; Bargmann, C. I., GFP Reconstitution Across Synaptic Partners (GRASP) defines cell contacts and synapses in living nervous systems. *Neuron* **2008,** *57* (3), 353-363.

56. Ueyama, T.; Kusakabe, T.; Karasawa, S.; Kawasaki, T.; Shimizu, A.; Son, J.; Leto, T. L.; Miyawaki, A.; Saito, N., Sequential binding of cytosolic Phox complex to phagosomes through regulated adaptor proteins: evaluation using the novel monomeric Kusabira-Green System and live imaging of phagocytosis. *J. Immunol.* **2008,** *181* (1), 629-640.

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

58. Ono, T.; Miyazaki, T.; Ishida, Y.; Uehata, M.; Nagata, K., Direct in vitro and in vivo evidence for interaction between Hsp47 protein and collagen triple helix. *J. Biol. Chem.* **2012,** *287* (9), 6810-6818.

59. Ando, R.; Mizuno, H.; Miyawaki, A., Regulated fast nucleocytoplasmic shuttling observed by reversible protein highlighting. *Science* **2004,** *306* (5700), 1370-1373.

60. Zhou, X. X.; Lin, M. Z., Photoswitchable fluorescent proteins: ten years of colorful chemistry and exciting applications. *Curr. Opin.Chem. Biol.* **2013,** *17* (4), 682-690.

61. Lee, Y. R.; Park, J.-H.; Hahm, S.-H.; Kang, L.-W.; Chung, J. H.; Nam, K.-H.; Hwang, K. Y.; Kwon, I. C.; Han, Y. S., Development of bimolecular fluorescence complementation using Dronpa for visualization of protein–protein interactions in cells. *Mol. Imaging Biol.* **2010,** *12* (5), 468-478.

62. Jach, G.; Pesch, M.; Richter, K.; Frings, S.; Uhrig, J. F., An improved mRFP1 adds red to bimolecular fluorescence complementation. *Nat. Methods* **2006,** *3* (8), 597-600.

63. Zilian, E.; Maiss, E., Detection of plum pox potyviral protein–protein interactions in planta using an optimized mRFP-based bimolecular fluorescence complementation system. *J. Gen. Virol.* **2011,** *92* (12), 2711-2723.

64. Fan, J.-Y.; Cui, Z.-Q.; Wei, H.-P.; Zhang, Z.-P.; Zhou, Y.-F.; Wang, Y.-P.; Zhang, X.-E., Split mCherry as a new red bimolecular fluorescence complementation system for visualizing protein–protein interactions in living cells. *Biochem. Bioph. Res. Co* **2008,** *367* (1), 47-53.

65. Rackham, O.; Brown, C. M., Visualization of RNA–protein interactions in living cells: FMRP and IMP1 interact on mRNAs. *EMBO J.* **2004,** *23* (16), 3346-3355.

66. Furman, J. L.; Badran, A. H.; Shen, S.; Stains, C. I.; Hannallah, J.; Segal, D. J.; Ghosh, I., Systematic evaluation of split-fluorescent proteins for the direct detection of native and methylated DNA. *Bioorgan. Med. Chem. Lett.* **2009,** *19* (14), 3748-3751.

67. Qin, L.; Chu, J.; Zheng, Y.; Luo, Q.; Zeng, S.; Zhang, Z., *A new red bimolecular fluorescence complementation based on TagRFP*, SPIE BiOS: Biomedical Optics, International Society for Optics and Photonics: 2009; pp 71910G-71910G-6.

68. Merzlyak, E. M.; Goedhart, J.; Shcherbo, D.; Bulina, M. E.; Shcheglov, A. S.; Fradkov, A. F.; Gaintzeva, A.; Lukyanov, K. A.; Lukyanov, S.; Gadella, T. W., Bright monomeric red fluorescent protein with an extended fluorescence lifetime. *Nat. Methods* **2007,** *4* (7), 555-557.

69. Liu, Z.; Xing, D.; Su, Q. P.; Zhu, Y.; Zhang, J.; Kong, X.; Xue, B.; Wang, S.; Sun, H.; Tao, Y., Super-resolution imaging and tracking of protein–protein interactions in sub-diffraction cellular space. *Nat. Commun.* **2014,** *5*.

70. Xia, P.; Liu, X.; Wu, B.; Zhang, S.; Song, X.; Yao, P. Y.; Lippincott-Schwartz, J.; Yao, X., Superresolution imaging reveals structural features of EB1 in microtubule plus-end tracking. *Mol. Biol. Cell* **2014,** *25* (25), 4166-4173.

71. Nickerson, A.; Huang, T.; Lin, L.-J.; Nan, X., Photoactivated localization microscopy with bimolecular fluorescence complementation (BiFC-PALM) for nanoscale imaging of protein-protein interactions in cells. *PLOS ONE* **2014,** *9* (6), e100589.

72. Chu, J.; Zhang, Z.; Zheng, Y.; Yang, J.; Qin, L.; Lu, J.; Huang, Z.-L.; Zeng, S.; Luo, Q., A novel far-red bimolecular fluorescence complementation system that allows for efficient visualization of protein interactions under physiological conditions. *Biosens. Bioelectron.* **2009,** *25* (1), 234-239.

73. Lin, M. Z.; McKeown, M. R.; Ng, H.-L.; Aguilera, T. A.; Shaner, N. C.; Campbell, R. E.; Adams, S. R.; Gross, L. A.; Ma, W.; Alber, T., Autofluorescent proteins with excitation in the optical window for intravital imaging in mammals. *Chem. Biol.* **2009,** *16* (11), 1169-1179.

74. Han, Y.; Wang, S.; Zhang, Z.; Ma, X.; Li, W.; Zhang, X.; Deng, J.; Wei, H.; Li, Z.; Zhang, X.-E., In vivo imaging of protein–protein and RNA–protein interactions using novel far-red fluorescence complementation systems. *Nucleic Acids Res.* **2014,** *42* (13), e103-e103.

75. Piatkevich, K. D.; Subach, F. V.; Verkhusha, V. V., Engineering of bacterial phytochromes for near-infrared imaging, sensing, and light-control in mammals. *Chem. Soc. Rev.* **2013,** *42* (8), 3441-3452.

76. Shu, X.; Royant, A.; Lin, M. Z.; Aguilera, T. A.; Lev-Ram, V.; Steinbach, P. A.; Tsien, R. Y., Mammalian expression of infrared fluorescent proteins engineered from a bacterial phytochrome. *Science* **2009,** *324* (5928), 804-807.

77. Filonov, G. S.; Piatkevich, K. D.; Ting, L.-M.; Zhang, J.; Kim, K.; Verkhusha, V. V., Bright and stable near-infrared fluorescent protein for in vivo imaging. *Nat. Biotechnol.* **2011,** *29* (8), 757-761.

78. Filonov, G. S.; Verkhusha, V. V., A near-infrared BiFC reporter for in vivo imaging of protein-protein interactions. *Chem. Biol.* **2013,** *20* (8), 1078-1086.

79. Tchekanda, E.; Sivanesan, D.; Michnick, S. W., An infrared reporter to detect spatiotemporal dynamics of protein-protein interactions. *Nat. Methods* **2014,** *11* (6), 641-644.

80. Chen, M.; Li, W.; Zhang, Z.; Liu, S.; Zhang, X.; Zhang, X.-E.; Cui, Z., Novel near-infrared BiFC systems from a bacterial phytochrome for imaging protein

interactions and drug evaluation under physiological conditions. *Biomaterials* **2015,** *48*, 97-107.

81. Bertrand, E.; Chartrand, P.; Schaefer, M.; Shenoy, S. M.; Singer, R. H.; Long, R. M., Localization of ASH1 mRNA particles in living yeast. *Mol. Cell* **1998,** *2* (4), 437-445.

82. Golding, I.; Paulsson, J.; Zawilski, S. M.; Cox, E. C., Real-time kinetics of gene activity in individual bacteria. *Cell* **2005,** *123* (6), 1025-1036.

83. Yin, J.; Zhu, D.; Zhang, Z.; Wang, W.; Fan, J.; Men, D.; Deng, J.; Wei, H.; Zhang, X.-E.; Cui, Z., Imaging of mRNA–Protein Interactions in Live Cells Using Novel mCherry Trimolecular Fluorescence Complementation Systems. *PLOS ONE* **2013,** *8* (11), e80851.

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

85. Wu, B.; Chen, J.; Singer, R. H., Background free imaging of single mRNAs in live cells using split fluorescent proteins. *Sci. Rep.* **2014,** *4*.

86. Valencia-Burton, M.; McCullough, R. M.; Cantor, C. R.; Broude, N. E., RNA visualization in live bacterial cells using fluorescent protein complementation. *Nat. Methods* **2007,** *4* (5), 421-427.

87. Valencia-Burton, M.; Shah, A.; Sutin, J.; Borogovac, A.; McCullough, R. M.; Cantor, C. R.; Meller, A.; Broude, N. E., Spatiotemporal patterns and transcription kinetics of induced RNA in single bacterial cells. *P. Natl. Acad. Sci.* **2009,** *106* (38), 16399-16404.

88. Yiu, H.-W.; Demidov, V. V.; Toran, P.; Cantor, C. R.; Broude, N. E., RNA detection in live bacterial cells using fluorescent protein complementation triggered by interaction of two RNA aptamers with two RNA-binding peptides. *Pharm.* **2011,** *4* (3), 494-508.

89. Ozawa, T.; Natori, Y.; Sato, M.; Umezawa, Y., Imaging dynamics of endogenous mitochondrial RNA in single living cells. *Nat. Methods* **2007,** *4* (5), 413-419.

90. Cheong, C.-G.; Hall, T. M. T., Engineering RNA sequence specificity of Pumilio repeats. *P. Natl. Acad. Sci.* **2006,** *103* (37), 13635-13639.

91. Llopis, P. M.; Jackson, A. F.; Sliusarenko, O.; Surovtsev, I.; Heinritz, J.; Emonet, T.; Jacobs-Wagner, C., Spatial organization of the flow of genetic information in bacteria. *Nature* **2010,** *466* (7302), 77-81.

92. Yoshimura, H.; Inaguma, A.; Yamada, T.; Ozawa, T., Fluorescent probes for imaging endogenous β-actin mRNA in living cells using fluorescent protein-tagged Pumilio. *ACS Chem. Biol.* **2012,** *7* (6), 999-1005.

93. Yu, H.; Braun, P.; Yıldırım, M. A.; Lemmens, I.; Venkatesan, K.; Sahalie, J.; Hirozane-Kishikawa, T.; Gebreab, F.; Li, N.; Simonis, N., High-quality binary protein interaction map of the yeast interactome network. *Science* **2008,** *322* (5898), 104-110.

94. Boruc, J.; Van den Daele, H.; Hollunder, J.; Rombauts, S.; Mylle, E.; Hilson, P.; Inzé, D.; De Veylder, L.; Russinova, E., Functional modules in the Arabidopsis core cell cycle binary protein–protein interaction network. *The Plant Cell Online* **2010,** *22* (4), 1264-1280.

95. Stagljar, I.; Korostensky, C.; Johnsson, N.; Te Heesen, S., A genetic system based on split-ubiquitin for the analysis of interactions between membrane proteins in vivo. *P. Natl. Acad. Sci.* **1998,** *95* (9), 5187-5192.

96. Snider, J.; Kittanakom, S.; Damjanovic, D.; Curak, J.; Wong, V.; Stagljar, I., Detecting interactions with membrane proteins using a membrane two-hybrid assay in yeast. *Nat. Protoc.* **2010,** *5* (7), 1281-1293.

97. Iyer, K.; Burkle, L.; Auerbach, D.; Thaminy, S.; Dinkel, M.; Engels, K.; Stagljar, I., Utilizing the split-ubiquitin membrane yeast two-hybrid system to identify protein-protein interactions of integral membrane proteins. *Sci. Signal.* **2005,** *2005* (275), pl3.

98. Snider, J.; Hanif, A.; Lee, M. E.; Jin, K.; Analyn, R. Y.; Graham, C.; Chuk, M.; Damjanovic, D.; Wierzbicka, M.; Tang, P., Mapping the functional yeast ABC transporter interactome. *Nat. Chem. Biol.* **2013,** *9* (9), 565-572.

99. Lam, M. H. Y.; Snider, J.; Rehal, M.; Wong, V.; Aboualizadeh, F.; Drecun, L.; Wong, O.; Jubran, B.; Li, M.; Ali, M., A Comprehensive Membrane Interactome Mapping of Sho1p Reveals Fps1p as a Novel Key Player in the Regulation of the HOG Pathway in S. cerevisiae. *J. Mol. Biol.* **2015**.

100. Sung, M.-K.; Lim, G.; Yi, D.-G.; Chang, Y. J.; Yang, E. B.; Lee, K.; Huh, W.-K., Genome-wide bimolecular fluorescence complementation analysis of SUMO interactome in yeast. *Genome Res.* **2013,** *23* (4), 736-746.

101. Morell, M.; Espargaro, A.; Avilés, F. X.; Ventura, S., Detection of transient protein–protein interactions by bimolecular fluorescence complementation: The Abl-SH3 case. *Proteomics* **2007,** *7* (7), 1023-1036.

102. Lee, O.-H.; Kim, H.; He, Q.; Baek, H. J.; Yang, D.; Chen, L.-Y.; Liang, J.; Chae, H. K.; Safari, A.; Liu, D., Genome-wide YFP fluorescence complementation screen

identifies new regulators for telomere signaling in human cells. *Mol. Cell Proteomics* **2011,** *10* (2), M110. 001628.

103. Choi, E.-W.; Seen, D.-S.; Song, Y. B.; Son, H.-S.; Jung, N.-C.; Huh, W.-K.; Hahn, J.-S.; Kim, K.; Jeong, J.-Y.; Lee, T. G., AdHTS: a high-throughput system for generating recombinant adenoviruses. *J. Biotechnol.* **2012,** *162* (2), 246-252.

104. Song, Y. B.; Park, C. O.; Jeong, J.-Y.; Huh, W.-K., Monitoring G protein-coupled receptor activation using an adenovirus-based β-arrestin bimolecular fluorescence complementation assay. *Anal. Biochem.* **2014,** *449*, 32-41.

105. Poe, J. A.; Vollmer, L.; Vogt, A.; Smithgall, T. E., Development and Validation of a High-Content Bimolecular Fluorescence Complementation Assay for Small-Molecule Inhibitors of HIV-1 Nef Dimerization. *J. Biomol. Screen.* **2014,** *19* (4), 556-565.

106. Dai, J.-P.; Zhao, X.-F.; Zeng, J.; Wan, Q.-Y.; Yang, J.-C.; Li, W.-Z.; Chen, X.-X.; Wang, G.-F.; Li, K.-S., Drug screening for autophagy inhibitors based on the dissociation of Beclin1-Bcl2 complex using BiFC technique and mechanism of eugenol on anti-influenza A virus activity. *PLOS ONE* **2013,** *8* (4), e61026.

107. Rust, M. J.; Bates, M.; Zhuang, X., Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). *Nat. Methods* **2006,** *3* (10), 793-796.

108. Betzig, E.; Patterson, G. H.; Sougrat, R.; Lindwasser, O. W.; Olenych, S.; Bonifacino, J. S.; Davidson, M. W.; Lippincott-Schwartz, J.; Hess, H. F., Imaging intracellular fluorescent proteins at nanometer resolution. *Science* **2006,** *313* (5793), 1642-1645.

109. Hess, S. T.; Girirajan, T. P.; Mason, M. D., Ultra-high resolution imaging by fluorescence photoactivation localization microscopy. *Biophys. J.* **2006,** *91* (11), 4258-4272.

110. Wang, S.; Moffitt, J. R.; Dempsey, G. T.; Xie, X. S.; Zhuang, X., Characterization and development of photoactivatable fluorescent proteins for single-molecule–based superresolution imaging. *P. Natl. Acad. Sci.* **2014,** *111* (23), 8452-8457.

111. Shroff, H.; Galbraith, C. G.; Galbraith, J. A.; Betzig, E., Live-cell photoactivated localization microscopy of nanoscale adhesion dynamics. *Nat. Methods* **2008,** *5* (5), 417-423.

112. Huang, B.; Bates, M.; Zhuang, X., Super resolution fluorescence microscopy. *Annu. Rev. Biochem.* **2009,** *78*, 993.

113. Zhang, M.; Chang, H.; Zhang, Y.; Yu, J.; Wu, L.; Ji, W.; Chen, J.; Liu, B.; Lu, J.; Liu, Y., Rational design of true monomeric and bright photoactivatable fluorescent proteins. *Nat. Methods* **2012,** *9* (7), 727-729.

114. Soufo, H. J. D.; Reimold, C.; Linne, U.; Knust, T.; Gescher, J.; Graumann, P. L., Bacterial translation elongation factor EF-Tu interacts and colocalizes with actin-like MreB protein. *P. Natl. Acad. Sci.* **2010,** *107* (7), 3163-3168.

115. Buey, R. M.; Mohan, R.; Leslie, K.; Walzthoeni, T.; Missimer, J. H.; Menzel, A.; Bjelić, S.; Bargsten, K.; Grigoriev, I.; Smal, I., Insights into EB1 structure and the role of its C-terminal domain for discriminating microtubule tips from the lattice. *Mole. Biol. Cell* **2011,** *22* (16), 2912-2923.

116. Komarova, Y.; De Groot, C. O.; Grigoriev, I.; Gouveia, S. M.; Munteanu, E. L.; Schober, J. M.; Honnappa, S.; Buey, R. M.; Hoogenraad, C. C.; Dogterom, M., Mammalian end binding proteins control persistent microtubule growth. *J. Cell Biol.* **2009,** *184* (5), 691-706.

117. Slep, K. C.; Rogers, S. L.; Elliott, S. L.; Ohkura, H.; Kolodziej, P. A.; Vale, R. D., Structural determinants for EB1-mediated recruitment of APC and spectraplakins to the microtubule plus end. *J. Cell Biol.* **2005,** *168* (4), 587-598.

118. Cohen, L. D.; Zuchman, R.; Sorokina, O.; Müller, A.; Dieterich, D. C.; Armstrong, J. D.; Ziv, T.; Ziv, N. E., Metabolic turnover of synaptic proteins: kinetics, interdependencies and implications for synaptic maintenance. *PLOS ONE* **2013,** *8* (5), e63191.

119. Hinz, F.; Dieterich, D.; Schuman, E., Teaching old NCATs new tricks: using non-canonical amino acid tagging to study neuronal plasticity. *Curr. Opin. Chemical Biol.* **2013,** *17* (5), 738-746.

120. Ong, S.-E.; Blagoev, B.; Kratchmarova, I.; Kristensen, D. B.; Steen, H.; Pandey, A.; Mann, M., Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol. Cell. Proteomics* **2002,** *1* (5), 376-386.

121. Butko, M. T.; Yang, J.; Geng, Y.; Kim, H. J.; Jeon, N. L.; Shu, X.; Mackey, M. R.; Ellisman, M. H.; Tsien, R. Y.; Lin, M. Z., Fluorescent and photo-oxidizing TimeSTAMP tags track protein fates in light and electron microscopy. *Nat. Neurosci.* **2012,** *15* (12), 1742-1751.

122. Fritzsche, S.; Springer, S., Pulse-Chase Analysis for Studying Protein Synthesis and Maturation. *Curr. Protoc. Protein Sci.* **2014**, 30.3. 1-30.3. 23.

123. Yewdell, J. W.; Lacsina, J. R.; Rechsteiner, M. C.; Nicchitta, C. V., Out with the old, in with the new? Comparing methods for measuring protein degradation. *Cell Biol. Int.* **2011,** *35* (5), 457-462.

124. Knop, M.; Edgar, B. A., Tracking protein turnover and degradation by microscopy: photo-switchable versus time-encoded fluorescent proteins. *Open Biol.* **2014,** *4* (4), 140002.

125. Khmelinskii, A.; Keller, P. J.; Bartosik, A.; Meurer, M.; Barry, J. D.; Mardin, B. R.; Kaufmann, A.; Trautmann, S.; Wachsmuth, M.; Pereira, G., Tandem fluorescent protein timers for in vivo analysis of protein dynamics. *Nat. Biotechnol.* **2012,** *30* (7), 708-714.

126. Subach, F. V.; Subach, O. M.; Gundorov, I. S.; Morozova, K. S.; Piatkevich, K. D.; Cuervo, A. M.; Verkhusha, V. V., Monomeric fluorescent timers that change color from blue to red report on cellular trafficking. *Nat. Chem. Biol.* **2009,** *5* (2), 118-126.

127. Takamura, A.; Hattori, M.; Yoshimura, H.; Ozawa, T., Simultaneous time-lamination imaging of protein association using a split fluorescent timer protein. *Anal. Chem.* **2015,** *87* (6), 3366-3372.

128. Cabantous, S.; Nguyen, H. B.; Pedelacq, J.-D.; Koraïchi, F.; Chaudhary, A.; Ganguly, K.; Lockard, M. A.; Favre, G.; Terwilliger, T. C.; Waldo, G. S., A new protein-protein interaction sensor based on tripartite split-GFP association. *Sci. Rep.* **2013,** *3*.

129. Kellermann, S. J.; Rath, A. K.; Rentmeister, A., Tetramolecular fluorescence complementation for detection of specific RNAs in vitro. *ChemBioChem* **2013,** *14* (2), 200-204.

130. Wang, Y.; Wang, Z.; Tanaka Hall, T. M., Engineered proteins with Pumilio/fem-3 mRNA binding factor scaffold to manipulate RNA metabolism. *FEBS J.* **2013,** *280* (16), 3755-3767.

131. Rath, A. K.; Kellermann, S. J.; Rentmeister, A., Programmable design of functional ribonucleoprotein complexes. *Chem. Asian J.* **2014,** *9* (8), 2045-2051.

132. Do, K.; Boxer, S. G., GFP Variants with Alternative β-Strands and Their Application as Light-driven Protease Sensors: A Tale of Two Tails. *J. Am. Chem. Soc.* **2013,** *135* (28), 10226-10229.

133. Do, K.; Boxer, S. G., Thermodynamics, Kinetics, and Photochemistry of β-Strand Association and Dissociation in a Split-GFP System. *J. Am. Chem. Soc.* **2011,** *133* (45), 18078-18081.

134. Alford, S. C.; Abdelfattah, A. S.; Ding, Y.; Campbell, R. E., A fluorogenic red fluorescent protein heterodimer. *Chem. Biol.* **2012,** *19* (3), 353-360.

135. Alford, S. C. Development of fluorogenic fluorescent protein heterodimers. University of Alberta, Edmonton, 2012.

136. Alford, S. C.; Ding, Y.; Simmen, T.; Campbell, R. E., Dimerization-dependent green and yellow fluorescent proteins. *ACS Synth. Biol.* **2012,** *1* (12), 569-575.

137. Ding, Y. Design and applications of fluorescent protein-based biosensors for live cell imaging. University of Alberta, Edmonton, 2014.

138. Mann, M.; Jensen, O. N., Proteomic analysis of post-translational modifications. *Nat. Biotechnol.* **2003,** *21* (3), 255-261.

139. Shafi, R.; Iyer, S. P. N.; Ellies, L. G.; O'Donnell, N.; Marek, K. W.; Chui, D.; Hart, G. W.; Marth, J. D., The O-GlcNAc transferase gene resides on the X chromosome and is essential for embryonic stem cell viability and mouse ontogeny. *P. Natl. Acad. Sci.* **2000,** *97* (11), 5735-5739.

140. Sarangi, P.; Zhao, X., SUMO-mediated regulation of DNA damage repair and responses. *Trends Biochem. Sci.* **2015,** *40* (4), 233-242.

141. Bond, M. R.; Hanover, J. A., A little sugar goes a long way: The cell biology of O-GlcNAc. *J. Cell Biol.* **2015,** *208* (7), 869-880.

142. Lazarus, M. B.; Nam, Y.; Jiang, J.; Sliz, P.; Walker, S., Structure of human O-GlcNAc transferase and its complex with a peptide substrate. *Nature* **2011,** *469* (7331), 564-567.

143. Hart, G. W.; Housley, M. P.; Slawson, C., Cycling of O-linked β-N-acetylglucosamine on nucleocytoplasmic proteins. *Nature* **2007,** *446* (7139), 1017-1022.

144. Boehmelt, G.; Wakeham, A.; Elia, A.; Sasaki, T.; Plyte, S.; Potter, J.; Yang, Y.; Tsang, E.; Ruland, J.; Iscove, N. N., Decreased UDP-GlcNAc levels abrogate proliferation control in EMeg32-deficient cells. *EMBO J.* **2000,** *19* (19), 5092-5104.

145. Lazarus, M. B.; Jiang, J.; Kapuria, V.; Bhuiyan, T.; Janetzko, J.; Zandberg, W. F.; Vocadlo, D. J.; Herr, W.; Walker, S., HCF-1 is cleaved in the active site of O-GlcNAc transferase. *Science* **2013,** *342* (6163), 1235-1239.

146. Capotosti, F.; Guernier, S.; Lammers, F.; Waridel, P.; Cai, Y.; Jin, J.; Conaway, J. W.; Conaway, R. C.; Herr, W., O-GlcNAc transferase catalyzes site-specific proteolysis of HCF-1. *Cell* **2011,** *144* (3), 376-388.

147. Daou, S.; Mashtalir, N.; Hammond-Martel, I.; Pak, H.; Yu, H.; Sui, G.; Vogel, J. L.; Kristie, T. M.; Affar, E. B., Crosstalk between O-GlcNAcylation and proteolytic cleavage regulates the host cell factor-1 maturation pathway. *P. Natl. Acad. Sci.* **2011,** *108* (7), 2747-2752.

148. Janetzko, J.; Walker, S., The making of a sweet modification: structure and function of O-GlcNAc transferase. *J. Biol. Chem.* **2014,** *289* (50), 34424-34432.

149. Flotho, A.; Melchior, F., Sumoylation: a regulatory protein modification in health and disease. *Annu. Rev. Biochem.* **2013,** *82*, 357-385.

150. Saitoh, H.; Hinchey, J., Functional heterogeneity of small ubiquitin-related protein modifiers SUMO-1 versus SUMO-2/3. *J. Biol. Chem.* **2000,** *275* (9), 6252-6258.

151. Cheviet, S.; Bezzi, P.; Ivarsson, R.; Renström, E.; Viertl, D.; Kasas, S.; Catsicas, S.; Regazzi, R., Tomosyn-1 is involved in a post-docking event required for pancreatic β-cell exocytosis. *J. Cell Sci.* **2006,** *119* (14), 2912-2920.

152. Williams, A. L.; Bielopolski, N.; Meroz, D.; Lam, A. D.; Passmore, D. R.; Ben-Tal, N.; Ernst, S. A.; Ashery, U.; Stuenkel, E. L., Structural and functional analysis of tomosyn identifies domains important in exocytotic regulation. *J. Biol. Chem.* **2011,** *286* (16), 14542-14553.

153. Hajmrle, C.; Ferdaoussi, M.; Plummer, G.; Spigelman, A. F.; Lai, K.; Fox, J. E. M.; MacDonald, P. E., SUMOylation protects against IL-1β-induced apoptosis in INS-1 832/13 cells and human islets. *Am. J. Physiol. Endoc. M.* **2014,** *307* (8), E664-E673.

154. Geerts, C. J.; Jacobsen, L.; van de Bospoort, R.; Verhage, M.; Groffen, A. J. A., Tomosyn Interacts with the SUMO E3 Ligase PIAS gamma. *PLOS ONE* **2014,** *9* (3).

155. Carrillo, L. D.; Krishnamoorthy, L.; Mahal, L. K., A cellular FRET-based sensor for β-O-GlcNAc, a dynamic carbohydrate modification involved in signaling. *J. Am. Chem. Soc.* **2006,** *128* (46), 14768-14769.

156. Carrillo, L. D.; Froemming, J. A.; Mahal, L. K., Targeted in vivo O-GlcNAc sensors reveal discrete compartment-specific dynamics during signal transduction. *J. Biol. Chem.* **2011,** *286* (8), 6650-6658.

157. Merckel, M. C.; Tanskanen, J.; Edelman, S.; Westerlund-Wikström, B.; Korhonen, T. K.; Goldman, A., The structural basis of receptor-binding by Escherichia coli associated with diarrhea and septicemia. *J. Mol. Biol.* **2003,** *331* (4), 897-905.

158. Tarrant, M. K.; Rho, H.-S.; Xie, Z.; Jiang, Y. L.; Gross, C.; Culhane, J. C.; Yan, G.; Qian, J.; Ichikawa, Y.; Matsuoka, T., Regulation of CK2 by phosphorylation and O-GlcNAcylation revealed by semisynthesis. *Nat. Chem. Biol.* **2012,** *8* (3), 262-269.

159. Sambrook, J., E.F. Fritsch, and T. Maniatis, *Molecular cloning: a laboratory manual*. Cold Spring Laboratory Press: New York, 1989; Vol. 2.

160. Schultz, C.; Schleifenbaum, A.; Goedhart, J.; Gadella, T. W., Multiparameter imaging for the analysis of intracellular signaling. *ChemBioChem* **2005,** *6* (8), 1323-1330.

161. Carlson, H. J.; Campbell, R. E., Genetically encoded FRET-based biosensors for multiparameter fluorescence imaging. *Curr. Opin. Biotechnol.* **2009,** *20* (1), 19-27.

162. Ai, H.-w.; Hazelwood, K. L.; Davidson, M. W.; Campbell, R. E., Fluorescent protein FRET pairs for ratiometric imaging of dual biosensors. *Nat. Methods* **2008,** *5* (5), 401-403.

163. Zhao, Y.; Araki, S.; Wu, J.; Teramoto, T.; Chang, Y.-F.; Nakano, M.; Abdelfattah, A. S.; Fujiwara, M.; Ishihara, T.; Nagai, T., An expanded palette of genetically encoded Ca2+ indicators. *Science* **2011,** *333* (6051), 1888-1891.

164. Krumpe, L. R.; Schumacher, K. M.; McMahon, J. B.; Makowski, L.; Mori, T., Trinucleotide cassettes increase diversity of T7 phage-displayed peptide library. *BMC Biotechnol.* **2007,** *7* (1), 65.

165. Greenfield, N. J., Using circular dichroism spectra to estimate protein secondary structure. *Nat. Protoc.* **2006,** *1* (6), 2876-2890.

166. Ai, H.-w.; Shaner, N. C.; Cheng, Z.; Tsien, R. Y.; Campbell, R. E., Exploration of new chromophore structures leads to the identification of improved blue fluorescent proteins. *Biochemistry* **2007,** *46* (20), 5904-5910.

167. MicroCal, L., ITC Data Analysis in Origin: Tutorial Guide. *MicroCal, LLC, Northampton, MA* **2004**.

168. Belmokhtar, C. A.; Hillion, J.; Ségal-Bendirdjian, E., Staurosporine induces apoptosis through both caspase-dependent and caspase-independent mechanisms. *Oncogene* **2001,** *20* (26), 3354-3362.

169. Bortner, C.; Cidlowski, J., Apoptotic volume decrease and the incredible shrinking cell. *Cell Death Differ.* **2002,** *9* (12), 1307-1310.

170. Tomosugi, W.; Matsuda, T.; Tani, T.; Nemoto, T.; Kotera, I.; Saito, K.; Horikawa, K.; Nagai, T., An ultramarine fluorescent protein with increased photostability and pH insensitivity. *Nat. Methods* **2009,** *6* (5), 351-353.

171. Goedhart, J.; von Stetten, D.; Noirclerc-Savoye, M.; Lelimousin, M.; Joosen, L.; Hink, M. A.; van Weeren, L.; Gadella Jr, T. W.; Royant, A., Structure-guided

evolution of cyan fluorescent proteins towards a quantum yield of 93%. *Nat. Commun.* **2012,** *3*, 751.

172. Wall, M. A.; Socolich, M.; Ranganathan, R., The structural basis for red fluorescence in the tetrameric GFP homolog DsRed. *Nat. Struct. Mol. Biol.* **2000,** *7* (12), 1133-1138.

173. Shu, X.; Shaner, N. C.; Yarbrough, C. A.; Tsien, R. Y.; Remington, S. J., Novel chromophores and buried charges control color in mFruits. *Biochemistry* **2006,** *45* (32), 9639-9647.

174. Quillin, M. L.; Anstrom, D. M.; Shu, X.; O'Leary, S.; Kallio, K.; Chudakov, D. M.; Remington, S. J., Kindling fluorescent protein from Anemonia sulcata: dark-state structure at 1.38 Å resolution. *Biochemistry* **2005,** *44* (15), 5774-5787.

175. Mizuno, H.; Mal, T. K.; Wälchli, M.; Kikuchi, A.; Fukano, T.; Ando, R.; Jeyakanthan, J.; Taka, J.; Shiro, Y.; Ikura, M., Light-dependent regulation of structural flexibility in a photochromic fluorescent protein. *P. Natl. Acad. Sci.* **2008,** *105* (27), 9227-9232.

176. Weber, W.; Helms, V.; McCammon, J. A.; Langhoff, P. W., Shedding light on the dark and weakly fluorescent states of green fluorescent proteins. *P. Natl. Acad.Sci.* **1999,** *96* (11), 6177-6182.

177. Zhang, J.; Wang, X.; Cui, W.; Wang, W.; Zhang, H.; Liu, L.; Zhang, Z.; Li, Z.; Ying, G.; Zhang, N., Visualization of caspase-3-like activity in cells using a genetically encoded fluorescent biosensor activated by protein cleavage. *Nat. Commun.* **2013,** *4*.

178. Laemmli, U. K., Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **1970,** *227* (5259), 680-685.

179. Shaner, N. C.; Lambert, G. G.; Chammas, A.; Ni, Y.; Cranfill, P. J.; Baird, M. A.; Sell, B. R.; Allen, J. R.; Day, R. N.; Israelsson, M., A bright monomeric green fluorescent protein derived from Branchiostoma lanceolatum. *Nat. Methods* **2013,** *10* (5), 407-409.

180. Wiedenmann, J.; Ivanchenko, S.; Oswald, F.; Schmitt, F.; Röcker, C.; Salih, A.; Spindler, K.-D.; Nienhaus, G. U., EosFP, a fluorescent marker protein with UV-inducible green-to-red fluorescence conversion. *P. Natl Acad.Sci. USA* **2004,** *101* (45), 15905-15910.

Appendices

Appendix A: Oligonucleotide sequence supplement

Table A1. List of oligonucleotides used in this work

OLIGO #	Lay ID	SEQUENCE (5' TO 3')
2.1	Kpn1-SATDGGIATV-GB(K153E)	CAGGGTACCGCCTCCGGCAGCGCCACCGATGGAGGTATCGCAACTGTGGGCGGAGCCACC
2.2	BamH1-GB(K153E)-RV	CTGGGATCCCTTGTACCGCTCGTCCATGCCC
2.3	BamH1-mCherry-FW	CAGGGATCCATGGTGAGCAAGGGCGAGGAG
2.4	Xho1-mCherry RV	CTGCTCGAGCTTGTACAGCTCGTCCATGCCG
2.5	BamH1-HCF1-repFW	CAGGGATCCGTGAGAGTGTGCAGTAACCCTCCT
2.6	HCF1rep2-3bp FW	GTGAGAGTGTGCAGTAACCCTCCTTGCGAAACCCACGAGACCGGCACCACAAATACTGCC ACAACTGCGACATCCAATATGGCCGGACAA
2.7	HCF1-rep2-E10Q FW	GTGAGAGTGTGCAGTAACCCTCCTTGCCAGACCCACGAGACCGGCACCACAAATACTGCCACAACTGCGACATCCAATATGGCCGGACAA
2.8	HCF1-rep2-C9R	GTGAGAGTGTGCAGTAACCCTCCTCGGGAAACCCACGAGACCGGCACCACAAATACTGCC ACAACTGCGACATCCAATATGGCCGGACAA
2.9	GA Xho1 RV	CTGCTCGAGTTAATGCCCCAGGTGGTGGC
2.10	Xho1-RA-FW	CAGCTCGAGAATGGTGAGCAAGAGCGAGGAG
2.11	RA-Kpn1-RV	GTTGAAGGTTTTGCCTCCTGGTCAGACATGGTACCCTTGTACAGCTCGTCCATGCCCAG
2.12	Kpn-SUMO1-FW	AAGGGTACCATGTCTGACCAGGAGGCAAAACCTTCAAC
2.13	SUMO1-MluI-RV	CCCACGCGTAACTGTTGAATGACCCCCCGTTTGTTC

2.14	Xba1-EGFP FW	CAGTCTAGAATGGAGCTGTTCACCGGGG TGGTG
2.15	EGFP-Sal1 RV	CTGGTCGACCTTGTACAGCTCGTCCATGC CG
2.16	Xba1-5AA-BglII-RB	CAGTCTAGAGGGGTCTCAATCAGCAGAT CT ACCATCAAAGAGTTCATGCGCTTCAAGG
2.17	RB-Sal1-RV	CTGGTCGACCTTGTACCGCTCGTCCATGC C
2.18	Xba1-mTomosyn1-FW	CAGTCTAGAATGAGGAAATTCAACATCA GGAAGGTGCTGGAC
2.19	mTomosyn1-BglII-RV	CTGAGATCTCTTCTTATCTTTGTATTTCA GCATCATCTCATGAGCATGTTTGGA
3.1	XhoI-GA FW	CAGCTCGAGAATGGTGAGCAAGAGCGA GGAG
3.2	DA-L-Kpn1Ov RV	GGTACCTGAACCAATATCCTTCCCGGCT CCAGACAGGGCTCCTGTGTGGCCCAGGT GGTG
3.3	Ov L-Kpn1-L-DEVD-L-GB FW	CGGGAAGGATATTGGTTCAGGTACCGCC TCCGGCGATGAGGTGGATGGAGCCACCA TCAAA
3.4	DEVD-L-GB-FW	TGGATGGAGCCACCATCAAAGAGTTCAT GCGCTTCAAGGT
3.5	GB-Mlu1-T-HindIII RV	CTGAAGCTTAACGCGTCTTGTACCGCTC GTCCATGCCC
3.6	MVSK-P2A-FW	CTCTCTGTTAAAGCAAGCAGGAGACGTG GAAGAAAACCCCGGTCCTTCTAGAATGG TGAGCAAGAGCGAGGAGG
3.7	EGFP-Sal1-RV	CTGGTCGACCTTGTACAGCTCGTCCATGC CG
3.8	GA-Sal1-RV	CTGGTCGACGCTACCTGTACCATGCCCC AGGTG
3.9	p2Ap2A-EcoRI-MVSK	CTGCTCAAGCAGGCCGGGGACGTGGAGG AGAACCCAGGACCCGAATTCATGGTGAG CAAGAGCGAGGAGG
3.10	mCherry-CaaX1/2-RV	TTTTCTTTTTCTTTTTCTTTCCATCCTTAC TCATTTTCTTGTACAGCTCGTCCATGCCG C
3.11	mCherry-CaaX 2/2-RV	CTGAAGCTTTTACATTATTACACACTTTG TTTTGCTTTTCTTTTTCTTTTTCTTTCCATC CTTACTCATTTT

Appendix B: Sequence of pBAD-*Xho*I-EGFP-*Mlu*I-p2Ap2A-*Xba*I-*Sal*I-p2Ap2A-*EcoR*I-*Hind*III

5'-
GTCCTCGAGAATGGGGTCAGTAATCACCTCTGGGTTGAAGGGGACGCGT*GGGA*
GCGGCGCTACAAATTTTAGTCTTCTGAAACAGGCCGGCGATGTTGAGGAGAATCCCG
GGCCAGGATCCGGAGCCACGAACTTCTCTCTGTTAAAGCAAGCAGGAGACGTGGAA
*GAAAACCCCGGTCCT*TCTAGAGTCGGGAATACATCAGGTATCGCCGTCAGCGTC
GAC*GGGTCAGGCGCTACGAAT*
TTTAGCCTCCTTAAACAAGCTGGAGATGTCGAAGAAAATCCCGGGCCAGGTAGCGGA
GCCACTAACTTCTCCCTGCTCAAGCAGGCCGGGGACGTGGAGGAGAACCCAGGACC
*C*GAATTCGGTATGGCCGGTATAGTATCTCTGGACGGATAAAAGCTTCAG-3'

Underlined portions represent restriction sites and in italics are the p2Ap2A sites. Although codon-optimized, the p2A sites all follow the amino acid sequence as listed by Kim *et al.* GSGATNFSLLKQAGDVEENPGP [126].

Appendix C: Non-linear and linear regression analysis of Figure 3.6A,B

Table C1. Data extracted from Figure 3.6A. Data obtained from Langmuir isotherm (Eq. [6]) fitting of data points using Origin 2015 software.

	[DA]=0 µM	[DA]=0.75 µM	[DA]=2.5 µM	[DA]=5 µM	[DA]=12.5 µM	[DA]= 25 µM	[DA]= 50 µM
Langmuir isotherm Best-fit values							
v when [GA]→∞	0.926	0.937	0.903	0.945	0.989	1.027	1.014
K'_{GA}	1.14	2.28	2.11	3.13	5.30	7.90	13.5
Std. Error							
v when [GA]→∞	0.06	0.0254	0.0306	0.0435	0.054	0.0392	0.0657
K'_{GA}	0.338	0.296	0.347	0.648	1.16	1.074	2.47

Table C2. Data extracted from double reciprocal plot. Slope and Y-intercept data obtained from **Figure 3.6B** was used to calculate for K'_{GA} values using [8] to fit data points with Origin 2015 software.

	[DA]=0 µM	[DA]=0.75 µM	[DA]=2.5 µM	[DA]=5 µM	[DA]=12.5 µM	[DA]= 25 µM	[DA]=50 µM
Best-fit values							
Slope	1.39±0.06	1.36±0.10	1.45±0.08	1.67±0.10	2.22±0.13	4.14±0.22	9.75±0.23
Y-intercept	1.10±0.10	1.33±0.15	1.32±0.11	1.37±0.15	1.52±0.19	1.59±0.32	1.40±0.18
95% Confidence Intervals							
Slope	1.24 to 1.55	1.12 to 1.6	1.27to 1.02	1.43 to 1.91	1.91 to 2.53	3.63 to 4.66	9.17 to 10.3
Y-intercept when X=0.0	0.857 to 1.34	0.991 to 1.67	1.06to 1.58	1.02 to 1.71	1.08 to 1.96	0.844 to 2.33	0.962 to 1.84
Goodness of Fit							
r squared	0.986	0.951	0.975	0.966	0.968	0.974	0.996
Residual sum of squares	0.319	1.27	0.729	1.31	2.16	6.05	1.05
Langmuir Isotherm Best-fit values							
K'_{GA}	1.54±0.12	1.81±0.21	1.91±0.16	2.29±0.22	3.38±0.30	6.57±0.51	13.65±0.25