

**A fluorescent protein exchange (FPX) strategy for detecting protein-protein interactions and discovering proteinaceous inhibitors in**

***Escherichia coli***

by

Jhon Ralph De La Peña Enterina

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

Master of Science

Department of Chemistry

University of Alberta

©Jhon Ralph De La Peña Enterina, 2015

## Abstract

Green fluorescent protein (GFP), along with other GFP-like proteins, is one of the fundamental tools that are enabling the exciting current era of biological studies. The utility of GFP-like proteins as genetically encoded markers and sensors has provided us with powerful tools for tracking proteins and monitoring biochemical events important to health and diseases. While fluorescent proteins (FPs) are now considered indispensable in the field of bioimaging, protein engineers are striving to apply and/or evolve FPs for a more diverse range of novel and exciting applications. One of the products of these engineering efforts is fluorescent protein exchange (FPX) technology. This strategy relies on competitive interactions between two different colors of dimerization-dependent FPs (ddFPs). Specifically, a red “A” and a green “A” FP compete for binding to a dark “B” partner. In an effort to expand utility of FPX in detecting protein-protein interactions (PPIs), we describe the development of a new method that can be used to detect association and disruption of interacting proteins in bacterial colonies. This approach enables high-throughput screening in validating interacting partners and engineering better protein or peptide-based inhibitors for disrupting aberrant PPIs.

In this thesis we describe the construction and characterization of several bacterial polycistronic vectors which provided us a simpler and more efficient system for co-expressing interacting proteins in *Escherichia coli*. This work

yielded two bicistronic vectors and two tricistronic vector which are able to express two and three separate proteins simultaneously. Characterization revealed that all vectors maintained the inherent features of their parent pBAD vector, including the tightly-regulated expression. We also describe the construction of pFPX, a tricistronic vector derived plasmid that can co-express potentially interacting proteins as ddFP fusions. Using pFPX and the custom fluorescence imaging system, we have successfully monitored associations of E1/K1, p53 transactivation domain/HDM 2 and HRas<sub>WT</sub>/Raf-1<sub>WT</sub> RBD. We also established the applicability of our strategy in screening affinity-engineered interactions by library generation and screening. First, we randomly mutated the 89<sup>th</sup> amino acid of the nonbinding Raf-1<sub>R89L</sub> RBD mutant and screened for clones which rescued its binding of HRas<sub>WT</sub>. Second, we panned libraries of HRas extragenic suppressors and selected clones that gained affinity to Raf-1<sub>R89L</sub> RBD. And third, we designed and optimized 7-mer inhibitory peptides that selectively bind to the LPA<sub>2</sub> interacting-PDZ domains of either NHERF-2 (PDZ 2) or MAGI-3 (PDZ 6).

## Preface

A version of Chapter 1 is being prepared for submission as a review for publication as **J. R. Enterina**, L. Wu and R. E. Campbell, “Emerging fluorescent protein technologies”. I was responsible in the manuscript composition. L. Wu assisted on the revision and enhancement of graphic figures for publication purposes. Y. Ding prepared the cell images in **Figure 1.7** which depicts the utility of ddFPs and FPX technology in live cell imaging.

My supervisor Dr. Robert Campbell conceived the idea of using the FPX strategy in screening and discovering PPIs in bacteria. I conceptualized the design of co-expressing ddFP protein fusions as well as established its suitability in detecting and discovering PPI in *Escherichia coli* colonies. Specifically, I constructed all polycistronic vectors including pFPX, carried out all the proof-of-concept experiments and screened small-to-medium gene libraries of affinity engineered protein-protein and protein-peptide interactions. L. Wu assisted in cloning the 7-mer peptide inhibitor library for MAGI-3 PDZ 3. A version of **Figure 3.5** has been included in a recently published article, Ding, Y.; Li, J.; **Enterina, J. R.**; Shen, Y.; Zhang, I.; Tewson, P. H.; Mo, G. C. H.; Zhang, J.; Quinn, A. M.; Hughes, T. E.; Maysinger, D.; Alford, S. C.; Zhang, Y.; Campbell, R. E. Ratiometric biosensors based on dimerization-dependent fluorescent protein exchange. *Nat. Methods*. DOI: 10.1038/nmeth.3261. Published Online: Jan 26, 2015.

## Table of Contents

|  |             |
|--|-------------|
| <b>Abstract</b> .....  | <b>ii</b>   |
| <b>Preface</b> .....   | <b>iv</b>   |
| <b>List of Tables</b> .....  | <b>ix</b>   |
| <b>List of Figures</b> .....                                       | <b>x</b>    |
| <b>List of Abbreviation</b> .....                                  | <b>xiii</b> |
| <b>Chapter 1. Introduction</b> .....                               | <b>1</b>    |
| 1.1 Background.....  | 1           |
| 1.2 Established applications of FPs.....                           | 3           |
| 1.2.1 FPs for whole-cell labeling.....                             | 3           |
| 1.2.2 Development of red-shifted FPs.....                          | 6           |
| 1.2.3 FP-based biosensors.....                                     | 9           |
| 3.4.5.1 FRET-based biosensors.....                                 | 9           |
| 3.4.5.2 Single FP-based biosensors.....                            | 12          |
| 3.4.5.1 Split FP reconstitution.....                               | 14          |
| 1.2.4 FPs as temporal markers.....                                 | 15          |
| 1.3 Emerging applications of FPs.....                              | 18          |
| 1.3.1 FP-based biosensors incorporating unnatural amino acids..... | 19          |
| 1.3.2 Assembling an FP from 3 pieces.....                          | 21          |

|  |   |           |
|--|---|-----------|
| 1.3.3  | Light-driven FP biosensor .....                           | 24        |
| 1.3.4  | FPs as optogenetic actuator .....                         | 25        |
| 1.3.5  | FP-dependent transcription .....                          | 27        |
| 1.3.6  | Fluorescent protein exchange (FPX) .....                  | 28        |
| 1.4  | Scope of the thesis .....                                 | 32        |
| <b>Chapter 2. Construction and optimization of a polycistronic bacterial</b> |   |           |
| <b>vector for expression of multi-protein complexes .....</b>                |   |           |
|  |   | <b>35</b> |
| 2.1  | Introduction .....  | 35        |
| 2.2  | Results and discussion .....                              | 40        |
| 2.2.1  | Construction of bacterial polycistronic vectors .....     | 40        |
| 2.2.2  | Characterization of bacterial polycistronic vectors ..... | 45        |
| 2.3  | Conclusion .....  | 49        |
| 2.4  | Materials and methods .....                               | 50        |
| 2.4.1  | General materials and reagents .....                      | 50        |
| 2.4.2  | Recombinant DNA techniques .....                          | 53        |
| 2.4.3  | DNA sequencing .....                                      | 55        |
| 2.4.4  | Construction of bacterial bicistronic vectors .....       | 55        |
| 2.4.5  | Construction of bacterial tricistronic vectors .....      | 56        |
| 2.4.6  | Test expression of polycistronic pBAD vectors .....       | 58        |

|  |           |
|--|-----------|
| <b>Chapter 3. Screening of protein-protein interactions and peptide-based inhibitors by fluorescent protein exchange (FPX)</b> .....         | <b>60</b> |
| 3.1 Introduction .....   | 60        |
| 3.2 Results and discussion.....  | 70        |
| 3.2.1 FPX-based PPI screening design and strategy .....  | 70        |
| 3.2.2 FPX proof-of-concept.....  | 76        |
| 3.2.3 Reversion mutation screening for Raf <sub>R89L</sub> RBD-RA.....   | 82        |
| 3.2.4 HRas extragenic suppressors for Raf <sub>R89L</sub> RBD .....  | 85        |
| 3.4.5 Selective inhibition of NHERF 2 PDZ 2 (N2P2) and MAGI 3 PDZ 6 (M3P6) towards lysophosphatidic acid receptor 2 (LPA <sub>2</sub> )..... | 88        |
| 3.4.5.1 Screening of peptide inhibitors for N2P2-LPA <sub>2</sub> association .....  | 92        |
| 3.4.5.2 Screening of peptide inhibitors for M3P6-LPA <sub>2</sub> association .....  | 95        |
| 3.3 Conclusion.....  | 102       |
| 3.4 Materials and methods .....  | 105       |
| 3.4.1 General methods and reagents.....  | 105       |
| 3.4.2 DNA sequencing .....   | 116       |
| 3.4.3 Protein purification of ddFPs.....   | 116       |
| 3.4.4 Dissociation constant ( $K_d$ ) determination .....  | 117       |
| 3.4.5 FPX plasmid and proof-of-concept design.....   | 118       |
| 3.4.6 HRas and Raf-1 RBD library design .....  | 119       |

|                   |   |            |
|-------------------|---|------------|
| 3.4.7             | N2P2-LPA <sub>2</sub> -M3P6 plasmid and library design .....      | 119        |
| 3.4.8             | In colony fluorescence screening using a digital microscope ..... | 121        |
| <b>Chapter 4.</b> | <b>Conclusion and future directions .....</b>                     | <b>122</b> |
| 4.1               | Summary of thesis .....   | 122        |
| 4.2               | Immediate future works.....                                       | 127        |
| 4.2.1             | In vitro characterization of Raf-1 RBD and HRas mutants .....     | 127        |
| 4.2.2             | Characterization of N2P2 and M3P6 peptide inhibitors .....        | 128        |
| 4.3               | Future applications .....   | 130        |
| <b>References</b> | .....   | <b>135</b> |

## List of Tables

|           |   |     |
|-----------|---|-----|
| Table 1.1 | Spectral properties of far-red fluorescent proteins .....   | 8   |
| Table 2.1 | Oligonucleotide used for cloning and sequencing .....   | 51  |
| Table 2.2 | Spectral properties of RFP, EGFP and mPapaya.....   | 59  |
| Table 3.1 | Clones isolated from Raf <sub>R89L</sub> RBD reversion mutagenesis .....                                  | 84  |
| Table 3.2 | Substitution identified in HRas extragenic suppressors. ....  | 86  |
| Table 3.3 | Substitution identified at positions $P^6$ and $P^5$ of $P^6P^5$ SNTKF .....                              | 94  |
| Table 3.4 | Substitutions identified at positions $P^3$ , $P^2$ and $P^1$ of potential M3P6 inhibitory peptides ..... | 96  |
| Table 3.5 | Substitution identified on the second round of M3P6 inhibitory peptide optimization .....                 | 102 |
| Table 3.6 | List of oligonucleotides used for cloning and sequencing .....  | 105 |
| Table 3.7 | Gene constructs and cloning strategy described in Chapter 3 .....   | 111 |

## List of Figures

|            |   |    |
|------------|---|----|
| Figure 1.1 | Graphical representations of FUCCI and Brainbow-1 technologies  | 5  |
| Figure 1.2 | General fluorescent protein biosensor designs   | 10 |
| Figure 1.3 | Expansion of GFP chromophores by unnatural amino acid incorporation   | 21 |
| Figure 1.4 | GFP strategies based on $\beta$ -strand reconstitution  | 23 |
| Figure 1.5 | General schematic representations of FLIPs and T-DDOG strategies used in controlling biological activities in living cells. | 26 |
| Figure 1.6 | General graphic illustration of fluorescent protein exchange (FPX) strategy   | 30 |
| Figure 1.7 | Detection of caspase activity by FPX strategy   | 31 |
| Figure 2.1 | Role of programmed cell death-1 (PD-1) and programmed cell death-ligand 1 in tumor evasion                                  | 36 |
| Figure 2.2 | Popular screening methods for protein-protein interactions  | 38 |
| Figure 2.3 | Linear maps of the cloning region of pBAD-based polycistronic vectors   | 42 |
| Figure 2.4 | Test co-expression of fluorescent proteins using polycistronic pBAD vectors   | 44 |

|             |   |    |
|-------------|---|----|
| Figure 2.5  | Saturation plots of fluorescent proteins expressed cell cultures with varying concentrations of L-arabinose .....   | 48 |
| Figure 3.1  | Structures of p53-HDM X and its inhibitor RG7112.....   | 62 |
| Figure 3.2  | General high-throughput screening (HTS) assay schemes using FRET and PCA systems .....  | 63 |
| Figure 3.3  | A simplified model for LPA2-induced colon cancer pathway .....  | 69 |
| Figure 3.4  | General strategy for evolving binding-deficient protein variants or protein-based inhibitors using FPX strategy.....  | 70 |
| Figure 3.5  | Saturation curves for RA and GB against three ddFP-B copies namely GB, RB and YB .....  | 74 |
| Figure 3.6  | In colony PPI screening design (pFPX).....  | 75 |
| Figure 3.7  | Graphical representation of FPX based on E1 and K1 interaction  | 77 |
| Figure 3.8  | Docking and FPX screening results for helical coiled-coil E1 and K1.....  | 78 |
| Figure 3.9  | Graphical representation of FPX based screening of p53WT and HDM 2 interactions along with binding-deficient p53W23A and HDM 2 peptide inhibitor (PMI)..... | 80 |
| Figure 3.10 | Emission ratios acquired from FPX colony screening of p53-HDM 2 and Raf-1 RBD-HRas <sub>WT</sub> association .....  | 80 |

|             |   |    |
|-------------|---|----|
| Figure 3.11 | Graphical representation of FPX based screening of Raf-1WT RBD and HRasWT interactions along with binding-deficient Raf-1R89L RBD and HRas peptide inhibitor, ECCAVFRLH ..... | 82 |
| Figure 3.12 | A simplified representation of the FPX screening assay of reverted mutants of Raf-1 RBD L89X.....   | 84 |
| Figure 3.13 | A general scheme to screen a library of HRas extragenic suppressors using FPX.....  | 85 |
| Figure 3.14 | Binding interaction of Raf-1 RBD (WT or R89L) and HRas <sub>WT</sub> at the interface region .....  | 88 |
| Figure 3.15 | Superimposed models of N2P2 (green) and M3P6 (cyan) .....   | 90 |
| Figure 3.16 | General library screening protocol for protein or peptide inhibitor discovery.....  | 91 |
| Figure 3.17 | Colony-based emission ratios of 11 clones expressing M3P6 inhibitory peptides (first round of peptide optimization).....  | 97 |
| Figure 3.18 | Simulated M3P6-peptide complex .....  | 99 |

## List of Abbreviations

|         |  |
|---------|--|
| Å       | angstrom   |
| Ala     | alanine  |
| Arg     | arginine   |
| ASAP1   | accelerated sensor for action potential 1                  |
| Asn     | asparagine   |
| Asp     | aspartate  |
| ATP     | adenosine triphosphate                                     |
| BCA     | bicinchoninic acid assay                                   |
| BFP     | blue fluorescent protein                                   |
| BiFC    | bimolecular fluorescence complementation                   |
| B-PER   | bacterial protein extraction reagent                       |
| cAMP    | cyclic adenosine monophosphate                             |
| CiVSP   | <i>Ciona intestinalis</i> voltage sensitive phosphatase    |
| Co-IP   | co-immunoprecipitation                                     |
| cpFP    | circularly permuted fluorescent protein                    |
| cpGFP   | circularly permuted green fluorescent protein              |
| cpsfGFP | circularly permuted super folder green fluorescent protein |
| CyPet   | cyan fluorescent protein for energy transfer               |
| Cys     | cysteine   |
| ddFP    | dimerization-dependent fluorescent protein                 |
| ddFP-A  | dimerization-dependent fluorescent protein-A copy          |

|                |  |
|----------------|--|
| ddFP-B         | dimerization-dependent fluorescent protein-B copy      |
| ddGFP          | dimerization-dependent green fluorescent protein       |
| ddRFP          | dimerization-dependent red fluorescent protein         |
| ddYFP          | dimerization-dependent yellow fluorescent protein      |
| DMSO           | dimethyl sulfoxide                                     |
| DNA            | deoxyribonucleic acid                                  |
| dNTPs          | deoxynucleotide triphosphates                          |
| DsbC           | <i>E. coli</i> protein disulfide isomerase C           |
| DsRed          | <i>Discosoma sp.</i> red fluorescent protein           |
| dTomato        | dimeric Tomato red fluorescent protein                 |
| <i>E. coli</i> | <i>Escherichia coli</i>                                |
| EBFP           | enhanced blue fluorescent protein                      |
| EGF            | epidermal growth factor                                |
| EGFP           | enhanced green fluorescent protein                     |
| ER             | emission ratio   |
| FACS           | fluorescence activated cell sorting                    |
| FP             | fluorescent protein                                    |
| FPX            | fluorescent protein exchange                           |
| FRET           | Förster resonance energy transfer                      |
| FUCCI          | fluorescence ubiquitination-based cell cycle indicator |
| GA             | dimerization-dependent green fluorescent protein-A     |
| GB             | dimerization-dependent green fluorescent protein-B     |
| GFP            | green fluorescent protein                              |

|        |  |
|--------|--|
| Gln    | glutamine  |
| Glu    | glutamate  |
| Gly    | glycine  |
| GPCR   | G-protein coupled receptor                                 |
| GTP    | guanosine triphosphate                                     |
| GTPase | guanosine triphosphatase                                   |
| HDM 2  | human double minute 2 homolog                              |
| HeLa   | cervical cancer cell line originating from Henrietta Lacks |
| His    | histidine  |
| HTS    | high-throughput screening                                  |
| Ile    | isoleucine   |
| ITC    | isothermal titration calorimetry                           |
| kb     | kilobase   |
| $K_d$  | dissociation constant                                      |
| kDA    | kilodalton   |
| LB     | luria bertani  |
| Leu    | leucine  |
| LPA2   | Lysophosphatdic acid receptor 2                            |
| Lys    | lysine   |
| M3P6   | MAGI-3 PDZ 6   |
| mAB    | monoclonal antibody  |
| mAG    | monomeric Azami green                                      |
| MAGI-3 | membrane-associated guanylate kinase, WW and PDZ domain-   |

|             |  |
|-------------|--|
|             | containing protein 3   |
| Mant-GppNHp | 2'/3'-O-(N-Methyl-anthraniloyl)-guanosine-5'-[( $\beta$ , $\gamma$ )-imido]triphosphate, Triethylammonium salt |
| MBP         | maltose binding protein  |
| MBSU        | Molecular Biology Service Unit   |
| mCherry     | monomeric Cherry red fluorescent protein   |
| mEGFP       | monomeric enhanced green fluorescent protein   |
| Met         | methionine   |
| MHC         | major histocompatibility complex   |
| mg          | milligram  |
| mGluR       | metabotropic glutamate receptor  |
| mL          | milliliter   |
| mm          | millimeter   |
| mmol        | millimol   |
| mOrange2    | monomeric Orange 2 orange fluorescent protein  |
| mPapaya     | monomeric papaya yellow fluorescent protein  |
| mRFP 1      | monomeric red fluorescent protein 1  |
| mRNA        | messenger ribonucleic acid   |
| mRuby       | monomeric Ruby red fluorescent protein   |
| MS          | mass spectrometry  |
| ms          | millisecond  |
| MW          | molecular weight   |
| MWCO        | molecular weight cut-off   |

|                  |  |
|------------------|--|
| N2P2             | NHERF-2 PDZ 2  |
| ng               | nanogram   |
| NHERF-2          | Na <sup>+</sup> /H <sup>+</sup> exchange regulatory cofactor |
| Ni-NTA           | nickel-nitrilotriacetic acid                                 |
| nm               | nanometer  |
| nMuMG cells      | normal <i>Mus musculus</i> mammary gland cell line           |
| <i>Npu</i> DnaE  | split intein derived from <i>Nostoc punctiforme</i>          |
| Nrf2             | nuclear factor (erythroid-derived 2)-like 2                  |
| OD600            | optical density at 600 nm                                    |
| OFP              | orange fluorescent protein                                   |
| PAC 1, 2 and 3   | procaspase activating compound 1, 2 and 3                    |
| PALM             | photoactivated localization microscopy                       |
| PBS              | phosphate buffer saline                                      |
| PCA              | protein complementation assay                                |
| PCR              | polymerase chain reaction                                    |
| PD-1             | programmed cell death-1                                      |
| PDB              | protein data bank  |
| PD-L1            | programmed cell death-ligand 1                               |
| PDZ              | PSD95, Dlg1 and Zo-1   |
| Phe              | phenylalanine  |
| PIP <sub>2</sub> | phosphatidylinositol 4,5-biphosphate                         |
| PLC              | phospholipase C  |
| PPI              | protein-protein interaction                                  |

|             |  |
|-------------|--|
| Pro         | proline  |
| RA          | dimerization-dependent red fluorescent protein-A                 |
| RB          | dimerization-dependent red fluorescent protein-B                 |
| RBD         | Ras binding domain   |
| RBS         | ribosome binding site  |
| RFP         | red fluorescent protein  |
| RNA         | ribonucleic acid   |
| rRNA        | ribosomal ribonucleic acid                                       |
| RTHS        | reverse two-hybrid system  |
| ScFv        | single-chain variable fragment of an antibody                    |
| SDS-PAGE    | sodium dodecyl sulfate-polyacrylamide gel electrophoresis        |
| SE-pHluorin | super ecliptic pHluorin  |
| Ser         | serine   |
| sfGFP       | super folder green fluorescent protein                           |
| SICLOPPS    | split intein-mediated circular ligation of peptides and proteins |
| Ssp DnaE    | split intein derived from <i>Synechocystis sp.</i> PCC6803       |
| SubAna      | substitutional analysis  |
| SUMO 1      | small ubiquitin-related modifier 1                               |
| T7-RBS      | ribosome binding site from gene 10 of bacteriophage T7           |
| TCF7        | T-cell specific transcription factor 7                           |
| Thr         | threonine  |
| Trp         | tryptophan   |
| Tyr         | tyrosine   |

|                |   |
|----------------|---|
| Val            | valine  |
| WT             | wild type   |
| YA             | dimerization-dependent yellow fluorescent protein-A |
| YB             | dimerization-dependent yellow fluorescent protein-B |
| YFP            | yellow fluorescent protein                          |
| YPet           | yellow fluorescent protein for energy transfer      |
| $\Delta F$     | change in fluorescence                              |
| $\lambda$      | wavelength, nm                                      |
| $\lambda_{em}$ | emission wavelength, nm                             |
| $\lambda_{ex}$ | excitation wavelength, nm                           |
| $\mu g$        | microgram   |
| $\mu L$        | microliter  |
| $\mu M$        | micromolar  |

# Chapter 1

## Introduction

### 1.1 Background

The story of how fluorescent proteins (FPs) rose to their current position of prominence as research tools began with the discovery and isolation of *Aequorea victoria* jellyfish green FP (GFP) in early 1960s (1). It was not until 1994 when Martin Chalfie and co-workers (2) reported the first elegant application of *Aequorea* GFP as a fluorescent reporter. Specifically, they used it to monitor the activation and relative distribution of *mec-7* gene among touch receptor neurons of *Caenorhabditis elegans*. This first demonstration provided the scientific community with the first hints of the many exciting applications these proteins would enable.

To facilitate the myriad of possible applications of FPs, ongoing efforts have been applied to discovering new FPs and improving the existing ones. For example, genetic manipulation of *Aequorea* GFP resulted in several variants including EGFP (enhanced brightness,  $\lambda_{em}$  max. at 508 nm), EBFP (blue-shifted,  $\lambda_{em}$  max. at 440 nm), CFP (cyan fluorescent,  $\lambda_{em}$  max. at 477 nm) and YFP (red-shifted,  $\lambda_{em}$  max. at 527 nm) (3-5). One thing that laboratory-based protein engineering has, to this day, failed to achieve, is the development of a useful red fluorescent version of *Aequorea* GFP. Fortunately, natural evolution solved this protein engineering problem long ago, as revealed with the report of a red FP

homolog (DsRed) from a mushroom anemone *Discosoma sp.* in 1999 (6). Although, DsRed provided a desirable red-shifted fluorescence that peaks at around 583 nm, its oligomerization tendency and slow maturation rate complicated initial attempts to use it in many of the applications where *Aequorea* GFP had proven effective. Extensive efforts aimed at improving DsRed's spectral properties and decreasing its oligomerization tendency ultimately led to the development of the so-called mFruit series. Representative members of this series include mOrange ( $\lambda_{em}$  max. at 562 nm), mCherry ( $\lambda_{em}$  max. at 610 nm), and mPlum ( $\lambda_{em}$  max. at 649 nm) (7, 8). Analogous efforts with a red FP cloned from the bubble-tip anemone, *Entacmaea quadricolor*, led to the variant known as mKate and a number of far-red FPs including tagRFP657 ( $\lambda_{em}$  max. at 659 nm) and mNeptune ( $\lambda_{em}$  max. at 650 nm) (9-11).

The properties of FPs for live cell and deep tissue imaging have steadily improved over the years, with the introduction of ever-brighter and more red-shifted variants. Although advances in alternative fluorophore technologies, specifically organic dyes and quantum dots, has continued apace, FPs will continue to be the fluorophores of choice for most live cell imaging applications for the foreseeable future. The tremendous advantage of FPs relative to these alternative technologies is that they are genetically encoded fluorophores. This unique feature of FPs means that they can be non-invasively introduced into cells in the form of their corresponding gene. In addition, the gene for the FP can be fused with the gene for practically any protein of interest. The chimeric gene can be introduced to live cells, tissues, or transgenic organisms, and the localization

and dynamics of the protein of interest visualized by virtue of the fused FP. The genetically encoded nature of FP is the key feature that allows us to precisely monitor spatial activation of promoters, visualize protein dynamics, and label subcellular compartments. Many of these applications are impractical using traditional small molecule tags or quantum dots (12, 13).

In this Chapter I will first provide an overview of representative “established” applications of FPs (Section 1.2). I will then move on to reviewing “emerging” applications (Section 1.3) with an emphasis on examples that represent unprecedented new directions in the field of FPs.

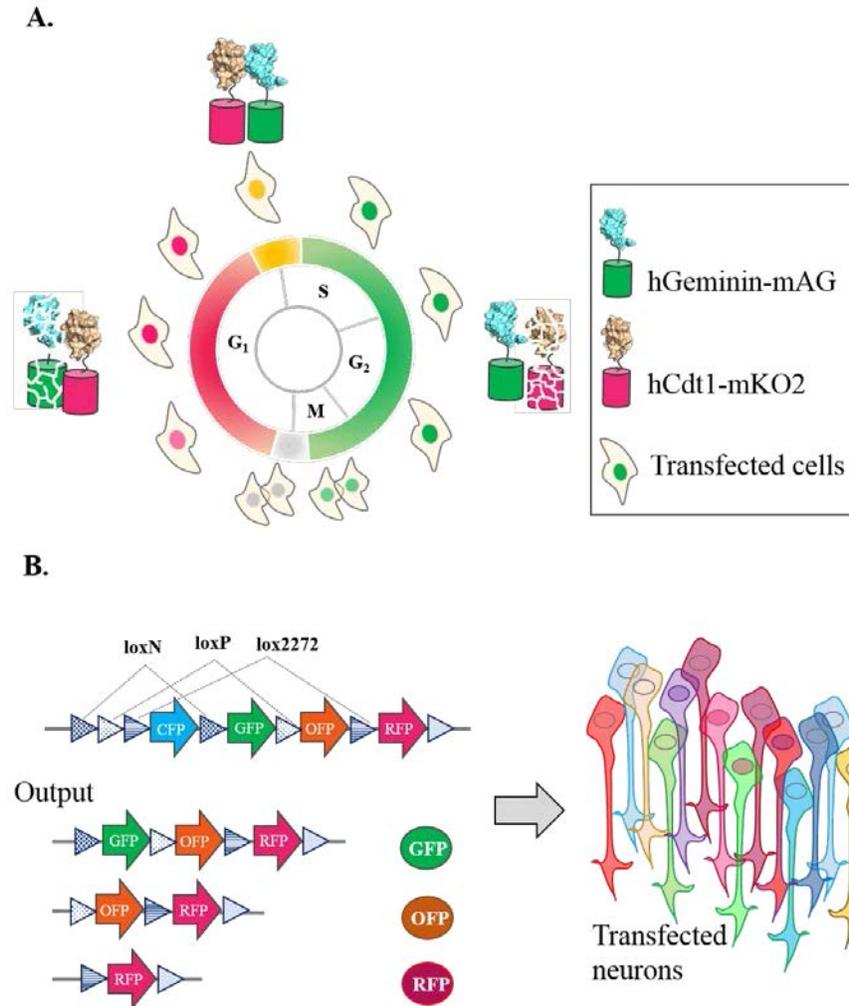
## **1.2 Established applications of FPs**

This section does not aim to be an exhaustive review of the history of FP development and applications, but rather aims to highlight a number of representative examples. For detailed discussion of previous work in established areas of FP development and biosensor engineering, the reader is directed to previous reviews (14-16).

### **1.2.1 FPs for whole cell labeling**

There are numerous examples of using FPs for whole cell labeling, so here we just provide two representative and elegant examples. In 2008, Miyawaki and co-workers (17) used a combination of monomeric green and orange FPs for real-time imaging of cell cycle progression. These probes, collectively known as FUCCI (Fluorescence Ubiquitination Cell Cycle Indicator), were generated by

fusing monomeric Kubasira Orange 2 (mKO2) and monomeric Azami-Green (mAG) FPs to truncated forms of the human cell cycle regulators hCdt1 and hGeminin, respectively. The concentrations of these two regulatory proteins are correlated with progression through the cell cycle. Specifically, hGeminin is targeted for proteosomal degradation in G<sub>1</sub>, while Cdt1 is degraded during S, G<sub>2</sub> and M phases (**Figure 1.1 A**). Accordingly, the intensities of green and orange fluorescence in a particular cell reveal the state of the cell, with respect to the cell division cycle. FUCCI was used to visualize cell cycle dynamics during the epithelial-mesenchymal transition in NMuMG cells, differentiation of neural progenitor cells, and tumor development in live mice. FUCCI is now established as a useful FP-based technology, as evidenced by numerous reported applications and demonstrations (18, 19). Also, a number of new FUCCI variants were introduced recently, including FUCCI 2 (20) and Fly-FUCCI (21). FUCCI 2 system replaces mKO2 and mAG FPs with mCherry and mVenus for improved imaging contrast, while Fly-FUCCI is a set of FP-based cell cycle reporters specifically designed for *Drosophila*.



**Figure 1.1 Graphical representations of FUCCI and Brainbow 1 technologies.** Shown here are the schematic illustrations of *A.* FUCCI and *B.* Brainbow-1 designs. For FUCCI, stable cell transformants tightly control the degradation of either hGeminin-mAG or hCdt1-mKO2 at different phases of the cell cycle. On the other hand, transfection of neurons with Brainbow-1 produces random Cre-recombinase excision patterns, giving a stochastic co-expression of FPs. As a result, each transfected neuron cell can be differentiated from each other due to its unique fluorescent hue.

Another FP-based technology for whole cell labeling is the Brainbow strategy for multicolour labeling of neuronal tissue (**Figure 1.1 B**). First reported in 2007 by Lichtman and co-workers, Brainbow is a multicolor labeling strategy in which individual neurons can be painted in one of ~100 different hues, and thereby easily distinguished from other neurons in the brain (22). This is achieved by randomly co-expressing varied ratios of FPs (CFP, GFP, OFP and RFP) in every transfected cell, giving many different combinations in an RGB-like scheme. Brainbow technology relies on the Cre/lox recombination system and a DNA plasmid construct encoding different genes of FPs, separated by copies of both mutant and canonical loxP sites. Stochastic expression occurs when Cre recombinase excises, inverts or translocates an FP-encoding DNA in between two identical lox sites. Brainbow, together with its latest improved version, has been used for visualization of neuron structure and interactions, which can provide a map of synaptic wiring and potentially identify defective connections relevant to many neurodegenerative disorders (23-25).

### **1.2.2 Development of red-shifted FPs**

Deep-tissue or whole animal optical imaging is most effective using wavelengths of light that fall within the near-infrared wavelength window which ranges from ~600 nm to ~1000 nm. In this wavelength range, light can penetrate deeper into tissues due to a relatively low absorption from myoglobin and hemoglobin. Water is also relatively transparent in this wavelength region, however, this effect attenuates at wavelengths beyond ~950 nm owing to its

increasing extinction coefficient (26, 27). In addition, tissue auto-fluorescence from co-factors like flavins and NADPH is negligible in this range (28, 29). Thus, FPs with excitation and emission wavelengths in the near-infrared region are highly valued as genetically-encoded tags for non-invasive tissue imaging. Extensive engineering and evolution of red FP templates has produced several FP variants with excitation and emission above 600 nm (9, 30-32). Unfortunately, the longer wavelength emission is typically associated with drastic reductions in fluorescent brightness (31). **Table 1.1** summarizes the fluorescent spectral properties of relevant far-red FPs.

Fortunately, several recently introduced far-red FPs, including mNeptune2, mNeptune2.5 and mCardinal (33), appear to offer substantial improvements relative to previously identified far-red FPs. Specifically, these new variants promise increased brightness and photostability compared to monomeric TagRFP657 (9), and dimeric E2-Crimson (30) and eqFP670 (32). Based on their molar absorptivity at 635 nm, the new variants can be efficiently excited by an ~633 nm laser, as is often used for deep-tissue imaging. Of these three, mCardinal appears particularly promising due to its long Stokes shift, red-shifted emission, and good brightness. Non-invasive imaging of mouse esophagus and far-red FP expressing-myocytes in TA muscles (tibialis anterior) revealed that mCardinal outperformed other mNeptune derivatives and phytochrome-based infrared fluorescent protein, iRFP (34), in terms of brightness and contrast.

**Table 1.1** Fluorescent spectral properties of far-red FPs

| Fluorescent protein       | Excitation nm          | Emission nm | Molar $\epsilon$ at peak ( $\text{mM}^{-1} \text{cm}^{-1}$ ) | $\Phi$ , total | Molar $\epsilon$ at 635 nm ( $\text{mM}^{-1} \text{cm}^{-1}$ ) | $\Phi > 635 \text{ nm}$ | Brightness at peak <sup>a</sup> | Photostability <sup>b</sup> |
|---------------------------|------------------------|-------------|--|----------------|--|-------------------------|---------------------------------|-----------------------------|
| mCherry <sup>e</sup>      | 587                    | 610         | 72   | 0.22           | 0.62   | 0.094                   | 16                              | 68                          |
| mKate <sup>e</sup>        | 585                    | 635         | 42   | 0.30           | ND   | ND                      | 13                              | 82                          |
| mGrape3 <sup>e</sup>      | 608                    | 646         | 40 <sup>c</sup>  | 0.03           | ND   | ND                      | 1.2                             | 5                           |
| Neptune <sup>e</sup>      | 600                    | 650         | 72   | 0.18           | ND   | ND                      | 13                              | 185                         |
| E2-Crimson <sup>f</sup>   | 605(611 <sup>d</sup> ) | 646         | 59   | 0.12           | 19   | 0.10                    | 7.1                             | ND                          |
| TagRFP657 <sup>g</sup>    | 611                    | 659         | 29   | 0.10           | 13   | 0.09                    | 2.9                             | 110                         |
| eqFP670 <sup>f</sup>      | 605                    | 670         | 70   | 0.06           | 20   | 0.055                   | 4.2                             | ND                          |
| mNeptune <sup>g</sup>     | 600                    | 651         | 75   | 0.23           | 9.6  | 0.18                    | 17                              | 160                         |
| mNeptune 2 <sup>g</sup>   | 599                    | 651         | 89   | 0.24           | 12   | 0.19                    | 21                              | 373                         |
| mNeptune 2.5 <sup>g</sup> | 599                    | 643         | 95   | 0.28           | 11   | 0.20                    | 27                              | 506                         |
| mCardinal <sup>g</sup>    | 604                    | 659         | 87   | 0.19           | 18   | 0.16                    | 17                              | 730                         |

ND = not determined

<sup>a</sup> Calculated as the product of molar extinction coefficient and total quantum yield ( $\epsilon \cdot \Phi$ ), unit is in  $\text{mM}^{-1} \text{cm}^{-1}$ .<sup>b</sup> Predicted time for fluorescence to photobleach by 50% under arc lamp illumination with excitation intensity adjusted to produce 1000 emission photons per molecule per s.<sup>c</sup> After photoactivation by 470 nm light.<sup>d</sup> Datum taken from ref. (30).<sup>e</sup> Data taken from ref. (10).<sup>f</sup> Data taken from ref. (32).<sup>g</sup> Data taken from ref. (33).

### 1.2.3 FP-based biosensors

In addition to using FPs as relatively static markers for labeling of cells or subcellular structures, there is also tremendous interest in using FP-based biosensors for monitoring of dynamic biochemical events in live cells and tissues. Accordingly, there are numerous examples of using FP-based biosensors for real-time imaging of enzyme activities (e.g., kinases, transferases and proteases), second messenger concentrations (e.g.,  $\text{Ca}^{2+}$  and cAMP), and a wide variety of other dynamic biochemical processes (35-41).

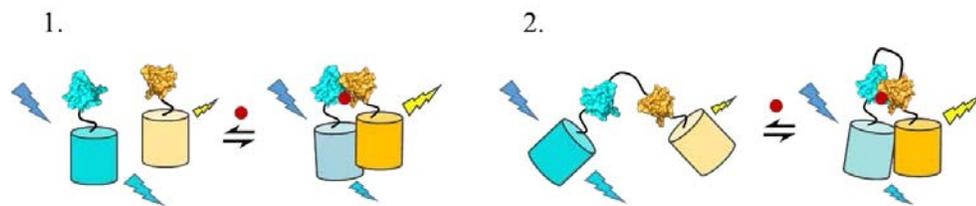
Established designs for FP-based biosensors tend to fall into 3 distinct classes: biosensors based on Förster Resonance Energy Transfer (FRET) between two different colors of FP; biosensors based on modulations of the chromophore environment that affect the brightness of a single FP; and biosensors based on reconstitution of a split FP (**Figure 1.2**). While FRET-based biosensors tend to be to have ratiometric responses, and the other two categories tend to have intensimetric responses, there are a growing number of exceptions to this general rule (35, 42, 43).

#### 1.2.3.1 FRET-based biosensors

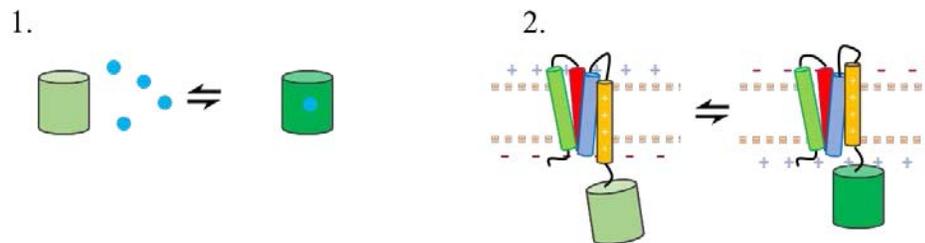
FRET is the radiationless transfer of energy from a higher energy donor fluorophore to a lower energy acceptor chromophore (44). Efficient transfer of energy depends on intrinsic photophysical properties (i.e., higher donor quantum yield, higher acceptor extinction coefficient, greater spectral overlap), relative dipole orientation, and distance between the two fluorophores (45). As FRET acts

like a spectroscopic ruler over distances of less than 10 nm, it can be usefully exploited to design biosensors in which the distance between two FPs is expected to change. The combination of a cyan FP (CFP) and a yellow FP (YFP) has traditionally offered the most robust performance for FRET imaging, with one notable example being the CyPet (cyan FP for energy transfer) and YPet (yellow FP for energy transfer) pair (46, 47). Yet other promising FRET pairs include the green-red Clover-mRuby2 (48) pair and the orange-red mOrange2-mCherry (49) pair.

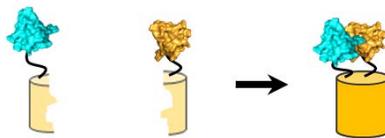
#### A. FRET-based FP biosensors



#### B. Single FP-based biosensors



#### C. Split FP-based biosensor



**Figure 1.2 General FP biosensor designs.** A. Förster resonance energy transfer (FRET)-based FP biosensors. Shown here are two typical designs, which

include intermolecular FRET (1) and intramolecular FRET (2). *B.* Single FP-based biosensors are generally single FPs where signal output depends on the two possible schemes: inherent sensitivity of FP to environmental changes (1) and conformational change of a fused sensing domain (2). *C.* Split FP reconstitution of the reporter's two non-fluorescent fragments.

A variety of different FRET-based biosensors have been developed, with most of them following one of two generally used assembly schemes. The most direct approach is to create an intermolecular biosensor in which both donor and acceptor FPs are fused to protein partners and expressed as separate fusion proteins in a host cell. This strategy was exploited for studying the kinetics of metabotropic glutamate receptors (mGluR) (50) and the oligomerization of dopamine D1 and D2 receptors (51). One disadvantage of this design is the possibility of false negative results. That is, even if true interaction occurs, the termini where the FP pair are fused might still be too far apart for effective FRET to occur. In addition, the expression level of the two biosensor components may differ between cells, causing to cell-to-cell variation in FRET measurements.

The second design, in which FRET occurs intramolecularly, is achieved by fusing the FP pair and the interacting proteins into a single polypeptide chain. Conformational changes, typically induced by posttranslational modification or small molecule binding, brings the donor and acceptor FPs into closer proximity such that the FRET efficiency increases. Alternatively, the FPs can be linked by the peptide substrate for a protease, with cleavage of this sequence resulting in a

loss of FRET. A large number of FP-based FRET biosensors have been constructed in this manner, including biosensors for small molecule messenger and metabolites (35, 36), enzyme activity of proteases (39) and kinases (40), and changes in membrane potential (52). Two notable examples are the cameleon series for imaging of  $\text{Ca}^{2+}$  ion concentrations (53) and the VSFP series for imaging of membrane voltage changes (52).

Despite the versatility and proven utility of FRET biosensors for imaging of biochemical events, researchers are increasingly turning to alternative designs of FP-based biosensors when possible, most notably single FP-type designs. Although the changes for these biosensors tend to be intensimetric rather than ratiometric, the magnitude of the changes is typically greater than for FRET-based biosensors. These large intensimetric changes make single FP-based biosensors somewhat more robust and easy to use than their FRET counterparts, while providing satisfactory signal-to-noise ratios.

#### *1.2.3.2 Single FP-based biosensors*

The earliest examples of single FP biosensors relied on the inherent sensitivity of an FP's fluorescence to certain environmental changes. In such FPs, the environmental change alters the chromophore environment, leading to intensimetric or ratiometric changes in fluorescent brightness. Examples of such environmental changes include changes in pH, concentration of halides, or changes in redox potential (54-56). One of the classic examples of a single FP-based pH sensor is the green fluorescent pHluorins (54, 57). Although no red

fluorescent pH indicator can yet match the performance of the pHlourins, a recently reported red indicator, designated as pHuji, can be used in conjunction with the pHlourins for two-color pH imaging applications (58).

By genetically linking an extrinsic sensing domain to an appropriately engineered FP, the scope of analytes that a single FP-based biosensor can detect can be increased. One well-established biosensor of this type is the GCaMP series of  $\text{Ca}^{2+}$  ion indicators (59, 60). These biosensors are composed of a circularly permuted FP (cpFP) fused to a  $\text{Ca}^{2+}$  ion sensing domain, calmodulin (CaM) and M13. Circular permutation is a genetic technique of joining the native N- and C-termini of a protein together with polypeptide linker and introducing new N- and C-termini at an alternative site. In the most effective cpFP-based biosensors, the new N- and C-termini are installed near the vicinity of the chromophore. In this way, conformation changes in the sensing domain have the greatest influence on the equilibrium between the protonated (typically non-fluorescent) and anionic states (typically fluorescent) of the chromophore. For GCaMP-type  $\text{Ca}^{2+}$  indicators, CaM and M13 are flexible in a  $\text{Ca}^{2+}$  free environment and the chromophore of the cpFP is quenched due to its exposure to solvent. Upon binding to  $\text{Ca}^{2+}$ , CaM and M13 associate which modifies the chromophore environment, giving an increase in fluorescent intensity (59, 60).

An emerging new direction for single FP-based biosensors is the construction of biosensors of membrane potential. The prototypical example is ArcLight (61), a single FP-based voltage sensor developed by fusing the voltage-sensing domain of *Ciona intestinalis* voltage sensitive phosphatase (CiVSP) to the

A227D mutant of super ecliptic pHluorin (SE-pHluorin). This voltage sensor exhibits a 14-fold increase in fluorescence ( $\Delta F/F$ ) for +100 mV depolarization, which is approximately three times higher than previously reported CiVSP FRET-based voltage sensors (52).

Recently, Lin and co-workers introduced a further improved single FP-based voltage sensor (62). This voltage sensor, named ASAP1 (Accelerated Sensor of Action Potentials 1), is a chimeric protein composed of the chicken (*Gallus gallus*) voltage-sensitive phosphatase domain and a circularly permuted sfGFP (superfolder GFP). This cpsfGFP is inserted in the extracellular loop between S3 and S4 transmembrane helices, which is believed to undergo a conformational change following depolarization. ASAP1 has faster on and off kinetics and superior spike resolution than ArcLight Q239. In addition, ASAP1 performed well in acute cortex slices of mouse brain by detecting current-induced action potentials even without the need for signal averaging or filtering.

#### 1.2.3.3 Split FP reconstitution

Yet another intensimetric biosensor FP-based biosensor design, popularized in the early 2000s, is based on the reconstitution of a genetically split FP. This biosensing design depends on the effective reassembly of two non-fluorescent FP fragments, leading to the formation of a highly fluorescent intact FP. Since the first report of this design by Regan and co-workers in 2000 (63), split FP-based biosensors have become an established technology for visualization of protein-protein interactions (PPIs) and even DNA/RNA-protein

interactions (64-68). One limitation of this strategy is the irreversible nature of the FP reconstitution. Accordingly, this approach is not useful for real-time visualization of PPI dynamics (69, 70).

#### **1.2.4 FPs as temporal markers**

The intracellular concentration of a particular protein depends on a number of factors, including the relative rates of synthesis, trafficking and degradation. Understanding and visualizing the interplay of these factors is of particular importance to researchers working on neurodegenerative disorders (71, 72). Alterations of these processes, such as the impairment of the degradation pathway (autophagy or ubiquitin-proteasome system), can lead to accumulation and aggregation of an aberrant protein in the cell. Detrimental effects of such accumulations are manifested in severe ailments such as Parkinson's and Huntington's diseases. A number of techniques have been developed to study the dynamics and trafficking of proteins, including cell fractionation and pulse-chase metabolic labeling. However, these techniques are not suitable for monitoring the dynamics of abnormal protein in living cells. To supplement these methods, a FP-based strategy known as "fluorescent timers" (FT) was introduced. Over a period of time these FTs, which are FPs, change their fluorescent hue. This distinct time-dependent change in colour is attributed to the temporary accumulation of fluorescent intermediates during the slow maturation of the chromophore in certain FP variants.

The concept of FTs first surfaced in 2000 when Terskikh *et al.* (73) mutated tetrameric DsRed red FP to change color from green to red fluorescence over a period of 18 h. This mutant, known as DsRed-E5, contains two substitutions compared to DsRed: V105A and S197T. Replacing Val with Ala at position 105 has been reported to increase the quantum yield by 2-fold. The S197T mutation most likely affects chromophore maturation and gives this FP its fluorescent timer characteristics. DsRed-E5 has been used to monitor biogenesis, gene activities, and organelle and fiber bundle formation (73-76). However, the widespread utility of DsRed-E5 in protein dynamics studies has been hindered by the protein's tetrameric nature, which can perturb the subcellular localization and physiological activity of a tagged protein.

To create an FT more suitable for protein tagging, Verkhusha and co-workers (77) converted the monomeric red FP mCherry (7) into three FP timer monomeric variants that change their fluorescence from blue to red over time. These three variants display either a fast, medium or slow chromophore maturation rate. Experimentally, medium-fluorescent timer (medium-FT) was able to track the trafficking of lysosome-associated membrane protein type 2 (LAMP-2A). Tagging LAMP-2A with medium-FT confirmed the prevalence of the direct trafficking mechanism where newly synthesized LAMP-2A proteins are transported from the Golgi to early and recycling endosomes, and eventually to lysosomal compartments. These monomeric FP timers were found to be superior to DsRed-E5 for tracking and timing the dynamics of proteins under different physiological conditions.

An addition to the family of FTs is the tandem fluorescent timers (tFTs), which were reported by Knop and co-workers in 2012 (78). This strategy relies on the difference in maturation kinetics between two monomeric FPs with well-separated emission peaks. Fusing a tFT to a protein-of-interest and imaging the fluorescence ratio over time provides a means to effectively track a protein's fate in time and space. This concept was demonstrated by fusing a tandem of monomeric FP mCherry ( $t_{1/2}$ , ~ 40 min.) and sfGFP ( $t_{1/2}$ , ~ several minutes) to yeast related proteins to study their mobility, stability and inheritance in dividing cells. Knop and co-workers also performed high-throughput screens in an attempt to identify regulators of protein turnover. Using mCherry-sfGFP tandem as a FT permitted ratiometric monitoring of proteins with turnover rates that span from minutes to a few hours. Creating new tFTs by replacing mCherry with other red FPs with faster or slower maturation kinetics could provide an expanded repertoire of timers to investigate biological events happening over different time scales ranging from signaling (few minutes) to cellular differentiation (few days).

Another FP-based method for determining the age of a synthesized protein is TimeSTAMP. In the initial design, the protein-of-interest was genetically tagged with nonstructural protein 3 (NS3) protease from Hepatitis C Virus (HCV) and an epitope tag at the C-terminus (79). Flexible polypeptide linkers, which connect NS3 protease to the cell surface, also contained the protease's cognate cleavage sequence. By default, HCV NS3 protease cleaves itself and the tag off the cell surface. However, this excision can be inhibited by introducing cell-permeable HCV NS3 protease inhibitor. Coupled with immunostaining, this

strategy has been utilized to image the distribution of newly synthesized proteins in fixed cells and the brain. To extend its application in live cells, Butko *et al.* (80) redesigned the original TimeSTAMP to replace the epitope tags with split Venus FP. Using this strategy, they have shown that a postsynaptic protein PSD95 is synthesized upon localized activation of mGluR (metabotropic glutamate receptor). New copies of the protein also accumulate preferentially on stimulated synapses and dendritic regions in rat neurons.

Fluorescent TimeSTAMP differs from the typical fluorescent timers described earlier since it relies on a drug-controlled reconstitution of a split FP rather than a blue to green to red conversion of a single FP or a ratiometric measurement of a pair of fast and slow maturing FPs. Another difference is that this strategy is designed to enable visualization of the spatial distribution of a newly synthesized protein rather than a long term monitoring of protein turnover offered by other FP timers.

### **1.3 Emerging applications of FPs**

As outlined in the preceding sections, FPs are now an indispensable tool available to life scientists. As genetically encoded tags and the basis for construction of live cell biosensors, FPs have enabled a plethora of applications that would be impractical or impossible using fluorescent reporter technologies (i.e., organic dyes and quantum dots) (12, 81). Aside from the many established applications of FPs, many protein engineers have been devoting their efforts to restructuring and redesigning FPs to gain novel properties for more advanced

applications. This section will introduce a number of innovative GFP-related technologies that have emerged in recent years. We highlight the innovative engineering and design that gave rise to these new tools, and describe some of the representative applications.

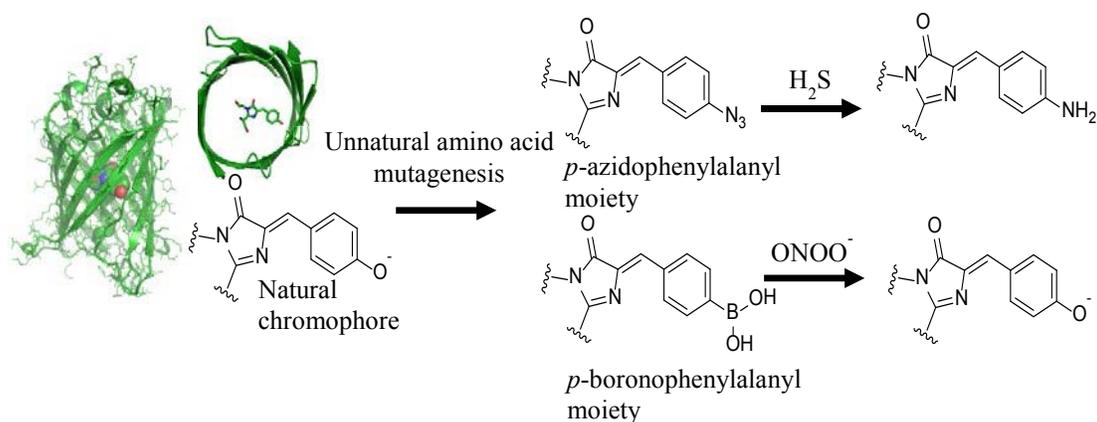
### **1.3.1 FP-based biosensors incorporating unnatural amino acids**

Some of the earliest GFP-based reporters relied on the inherent sensitivity of the chromophore to its environment. This sensitivity could be due to either a direct interaction of the species with the phenolate moiety of the chromophore (i.e.,  $H^+$  in the case of a pH reporter), or by interacting with neighboring residues (i.e.,  $Cl^-$  in the case of a halide concentration reporter) (54-56). Unfortunately, this approach to biosensor design is relatively limited due to the limited range of analytes that can interact with, and affect the fluorescence of, the wild-type GFP chromophore.

One approach to expanding the scope of FP-based biosensors would be to modify the chromophore such that it gained sensitivity to new analytes of interest. With this goal in mind, Ai and co-workers turned to unnatural amino acid mutagenesis to create single FP-based biosensors that incorporate chemically modified FP chromophores (**Figure 1.3**). For example, to create a biosensor of hydrogen sulfide gas ( $H_2S$ ), Ai and co-workers mutated Tyr66 of the chromophore-forming tripeptide to the unnatural amino acid *p*-azidophenylalanine (pAzF) (82).  $H_2S$  has been identified as a gas mediator involved in regulating inflammation, vasorelaxation and cardiac response (83).

Introduction of pAzF, both in *E. coli* and mammalian cells, was performed following the previously reported strategy developed by Schultz and co-workers (84, 85). Specifically, they substituted the codon for Tyr66 of a cpGFP with the TAG amber stop codon and co-expressed it with a cognate tRNA and tRNA-synthetase. The resulting mutant, cpGFP-pAzF, showed only a modest increase in fluorescence when incubated with buffered H<sub>2</sub>S both *in vitro* and in live cells. It also exhibited poor selectivity and sensitivity, and low signal-to-noise ratio (SNR) making it impractical for detecting dynamic changes in biological H<sub>2</sub>S.

Ai and co-workers have reported two additional genetically encoded FP biosensors based on FP chromophores derived from unnatural amino acids. One such biosensor was designed to selectively detect the peroxynitrite (ONOO<sup>-</sup>) ion, which is a highly reactive intracellular redox species linked to Alzheimer's diseases, arthritis and inflammatory disorders (86). The biosensor was prepared by substituting Tyr66 from a superfolder circularly permuted GFP (cpGFP2) with *p*-boronophenylalanine (pBoF, **Figure 1.3**). This sensor, cpGFP2-pBoF, exhibited better selectivity, SNR and linearity compared to its H<sub>2</sub>S biosensor counterpart. Inspired by the large fluorescence response of cpGFP2-pBoF to its analyte ONOO<sup>-</sup>, the same authors also replaced the current GFP template of cpGFP-pAzF with a superfolder cpGFP in an attempt to improve the original H<sub>2</sub>S biosensor design. In contrast to the original cpGFP-pAzF, this new sensor (named hsGFP) is highly selective and considerably brighter and more responsive to H<sub>2</sub>S when tested *in vitro* and in live cells (87).



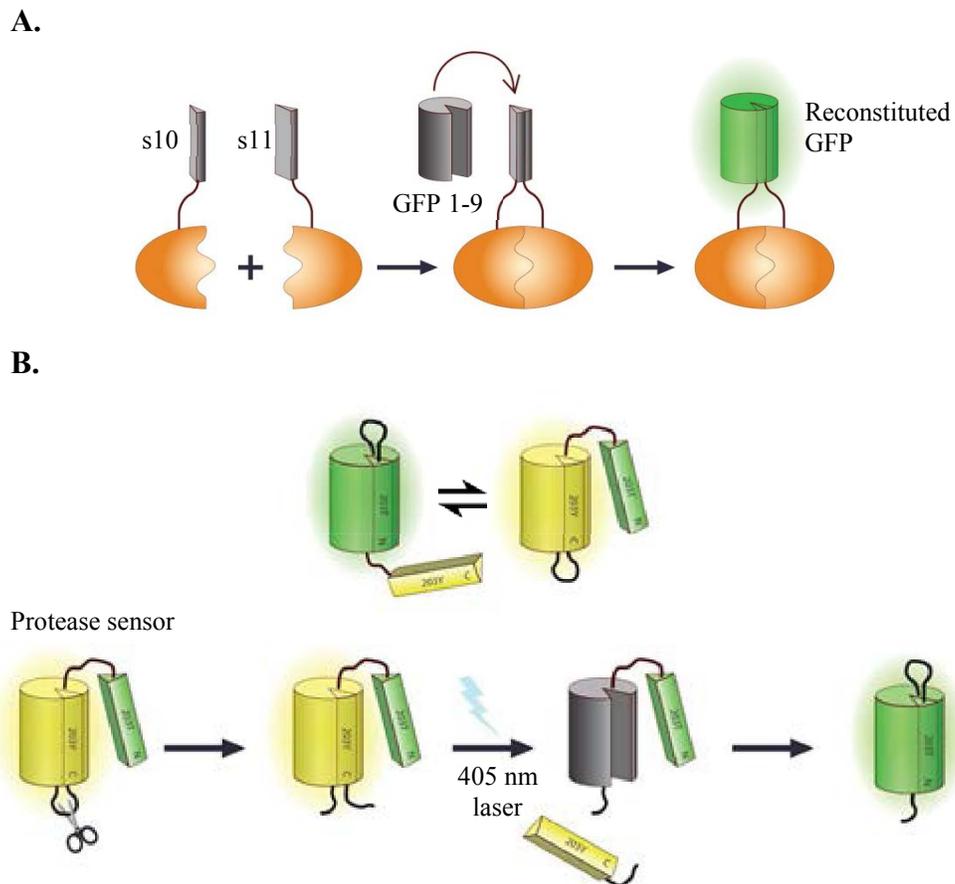
**Figure 1.3 Expansion of GFP chromophores by unnatural amino acid incorporation.** Conversion of Tyr66 in GFP into *p*-azidophenylalanine and *p*-boronophenylalanine produces chromophores with increased sensitivity to several relevant redox species such as hydrogen sulfide ( $\text{H}_2\text{S}$ ) and peroxynitrite ion ( $\text{ONOO}^-$ ).

Drawbacks of biosensors that incorporate unnatural amino acids include a slower maturation rate, the possibility of cross-reactivity with non-target analytes, and an irreversible response that hampers their use in monitoring dynamic changes in ion concentrations. However, this engineering strategy has opened new avenues to broaden the analyte sensitivity of single FP biosensors, especially for biologically important species with no known proteinaceous sensing domains.

### 1.3.2 Assembling an FP from 3 pieces

The FP complementation strategy (refer to **Section 1.2.3.3**) has been used for over a decade now. Although this strategy has proven effective for detecting

and discovering various PPIs, its further applications are hindered by several shortcomings which include poor folding and undesirable background self-assembly (69). Recently, Cabantous *et al.* (88) developed a tripartite split GFP-based reporter that addresses some of the deficiencies of traditional bipartite split FP designs (**Figure 1.4 A**). This strategy relies on the tripartite reconstitution of GFP  $\beta$ -strands 10 and 11 with the large fragment of GFP composed of  $\beta$ -strands 1-9. Reassembly of shorter polypeptides rather than longer fragments may significantly reduce the accumulation and self-association of the split FP, thereby decreasing the number of false positives. Cabantous *et al.* demonstrated the reassembly of functional GFP by fusing  $\beta$ -strands 10 and 11 to an interacting protein pair (FRB/FKBP12 and K1/E1 coiled-coils, respectively) and co-expressing them with the large GFP fragment ( $\beta$ -strands 1-9) in *E. coli*. In comparison to traditional bipartite complementation, this new system has the advantages of higher solubility for fusion proteins, lower aggregation, and lower background fluorescence from self-assembly.



**Figure 1.4 GFP strategies based on  $\beta$ -strand reconstitution.** *A.* Tripartite GFP biosensor. In this design,  $\beta$ -strands 10 and 11 of GFP are fused separately to an interacting protein pair. Upon interaction, both  $\beta$ -strands reassemble to the truncated GFP  $\beta$ -strands 1-9, which results in the formation of an intact and functional GFP. *B.* Light-driven GFP reporter. This strategy relies on the light catalyzed displacement of GFP's  $\beta$ -strand 10. Protease cleavage and illumination by a 405 nm laser rapidly displaces s10<sub>T203Y</sub> from the GFP barrel, leading to replacement with the canonical s10 containing Thr203.

### 1.3.3 Light-driven FP-based biosensor

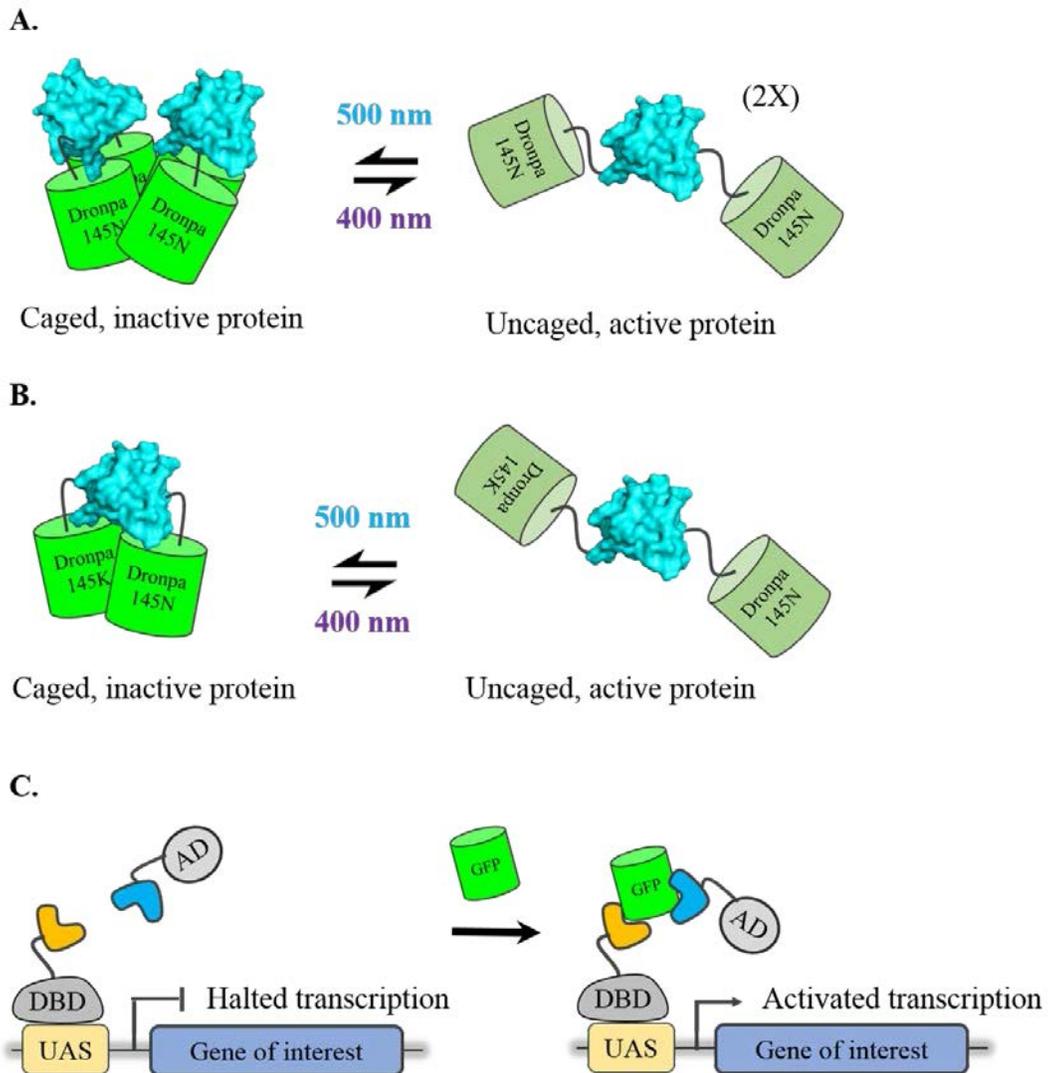
Boxer and co-workers (89, 90) have demonstrated that GFP can undergo photo-induced displacement and exchange of  $\beta$ -strand 10. This exchange can be used to shift GFP's emission from green to yellow fluorescence due to the presence of Tyr203 (associated with yellow emission) on one copy of  $\beta$ -strand 10 (s10<sub>T203Y</sub>), but not the one that is exchanged with (s10<sub>WT</sub>). To exploit this novel property for biosensing applications, Boxer and co-workers constructed a ratiometric and light-driven protease biosensor (91). This was achieved by circularly permuting GFP and flanking the new terminals with either of the two alternating  $\beta$ -strands: s10<sub>WT</sub> or s10<sub>T203Y</sub>. The polypeptide linker connecting the barrel and s10<sub>T203Y</sub> was engineered to include a cleavage site for the protease thrombin (**Figure 1.4 B**). The relative proportion of the two mature barrel isoforms could be tuned to bias one fluorescent hue (green or yellow) to maximize its biosensing potential. This was done by manipulating the length of the two flexible linkers between the barrel and the alternating  $\beta$ -strands. Exposure of s10<sub>T203Y</sub> dominating isoform (90% bound) to thrombin and light *in vitro* gradually shifted its spectral profile from that of YFP to GFP in a span of ~20 min. This biosensor showed a ratiometric fluorescence change of >100-fold, which is much greater than the changes typically observed for FRET-based biosensors.

### 1.3.4 FPs as an optogenetic actuator

The last decade has seen an explosion of interest and enthusiasm for so-called ‘optogenetic actuators’: molecular tools that can be used to artificially manipulate biological activities with spatial and/or temporal precision. Such tools can accelerate studies to elucidate protein functions under physiological conditions and determine the effects of gene activation in development and diseases. One class of optogenetic actuator are proteins that can dimerize or change their conformation upon illumination. These proteins have been exploited to optically control cellular processes both *ex vivo* and *in vivo*.

In 2012, Lin and co-workers (95) reported that the green fluorescent Dronpa FP variant underwent an illumination-dependent change in oligomerization state. Upon illumination with cyan light (500 nm), the tetrameric Dronpa 145N variant (**Figure 1.5 A**) dissociates to form a non-fluorescent and monomeric protein. This protein can be converted back to its fluorescent tetrameric state with violet light (400 nm) illumination. This photoswitching process occurs even when two copies of Dronpa are fused as an intramolecular tandem dimer (**Figure 1.5 B**). This interesting finding led to the development of light-inducible proteins known as FLIPs, the first FP-based optogenetic actuators. To demonstrate the utility of FLIPs for controlling protein activity in mammalian cells, Lin and co-workers caged Cdc42 GEF intersectin and Hepatitis C virus (NCV) NS3-4A protease using two different FLIP designs. In both cases, the caging strategy led to a decrease in the activity of the target enzymes. Light-induced uncaging of intersectin led to extension of existing filopodia. Uncaging of

HCV protease was detected by imaging the release of mCherry linked to the membrane via an HCV protease substrate. The FLIP strategy has enabled a new range of applications for FPs and is likely to serve as a versatile alternative to previously developed optogenetic tools.



**Figure 1.5** General schematic representations of FLIPs and T-DDOG strategies used in controlling biological activities in living cells. Light-induced proteins, popularly called FLIPs, are Dronpa-derived photoswitchable proteins that dissociate from their oligomeric states upon exposure to cyan light (500 nm).

FLIPs can be designed to control protein activity either by *A.* tetrameric caging or *B.* dimeric caging. *C.* Another strategy known as T-DDOG (transcription devices dependent on GFP) uses GFP as a scaffold to control gene expression.

### 1.3.5 FP-dependent transcription

Tang *et al.* have recently reported a strategy for using GFP to regulate transcription activity (96). This strategy was designated Transcription Devices Dependent on GFP (T-DDOG) and relies not on the optical control of GFP, but on the presence of the protein in a given tissue. Essentially, this strategy aims to take advantage of the large number of GFP transgenic animals currently available, and use the presence of GFP to drive other genes of interest. To achieve this aim, Tang *et al.* created a transcriptional system that respond to the presence of GFP by specifically activating certain genes, which could be used to reprogram the development and behavior of transfected cells (**Figure 1.5 C**). To construct the components of T-DDOG, two optimized anti-GFP nanobodies, derived from the V<sub>HH</sub> domains of camelid single-chain antibodies, were separately fused to the Gal4 DNA binding domain (DBD) and p65 activation domain (AD). GFP, serving as a dimerizer, simultaneously binds to DBD-nanobody fusion and recruits the AD-nanobody in the nucleus; hence, forming a biologically active assembly which activates gene expression. The nanobody components of T-DDOG are highly specific to *Aequorea* GFP, making this approach compatible with many fluorescent red-shifted FPs for multicolor imaging. An element of temporal control was introduced by replacing the DBD with the drug-regulated rTetR DNA

binding domain. Application of this strategy *in vivo* directed functional perturbations in specific cell-types in mouse retina and brain.

Although these two emerging strategies, FLIPs and T-DDOG, are still in their infancy, they herald an exciting new era in which the utility of FPs extends beyond its role as a fluorophore. Their development establishes the potential of FPs to be used to manipulate both protein and gene activities in cells.

### **1.3.6 Fluorescent protein exchange (FPX)**

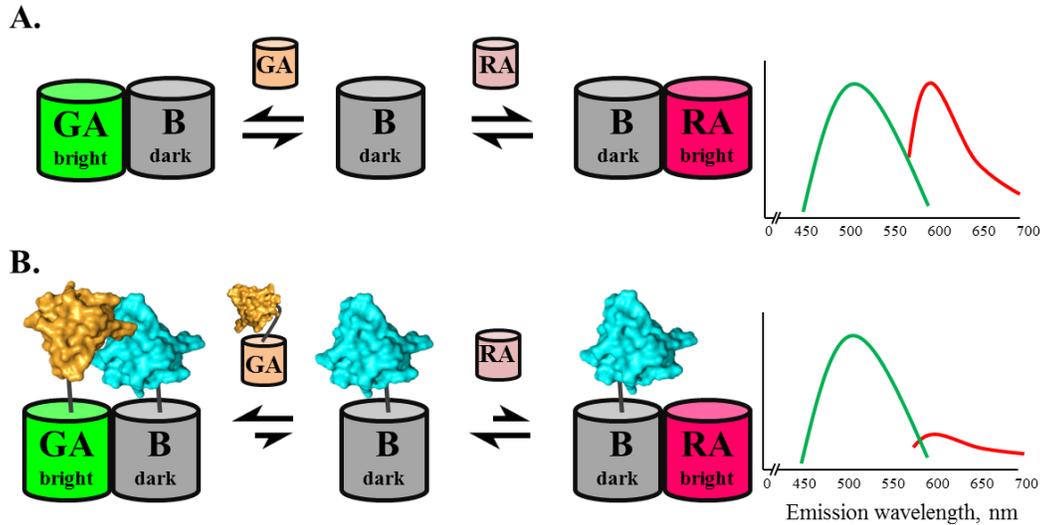
As described in **Section 1.2.3**, established strategies for converting FPs into active reporters of cellular processes such as protein-protein interactions (PPIs), kinase activities, and small molecule messenger dynamics are relatively few in number. For detecting PPI, two of the commonly used methods are FRET between two FPs, and induced-reassembly of two FP fragments (split FP). Despite their popularity, these strategies suffer from a number of limitations. For example, FRET-based reporters often have relatively small signal changes and correspondingly lower signal-to-noise ratios. Likewise, the slow kinetics and irreversible nature of most split FP-based biosensors impede their use in imaging reversible PPIs.

In an effort to address the shortcomings of the current FP-based methods, Alford *et al.* (97, 98) recently introduced dimerization-dependent FPs (ddFPs). The development of ddFPs was inspired by the oligomerization-dependent enhancement of red fluorescence of DsRed and other oligomeric Anthozoan-derived FPs. A ddFP pair consists of a quenched fluorogenic FP (A copy) and a

non-fluorogenic FP (B copy) that can associate to form a brightly fluorescent heterodimer complex (97, 98). The A copy monomer was engineered from dTomato and harbors a fully matured chromophore that is quenched at the monomeric state. The B copy, which was also engineered from dTomato, does not form a chromophore itself but acts as a fluorescent enhancer for A copy upon formation of the AB heterodimer complex. Currently, there are three spectrally distinct pairs of ddFP that have been developed: green, yellow and red. Each of these ddFP pairs have been utilized for intensimetric biosensing of various biochemical activities including  $\text{Ca}^{2+}$  dynamics, protease activity, and mitochondria-endoplasmic reticulum (ER) membrane proximity. This strategy combines many of the positive attributes of the other methods including the reversible nature of FRET and large fluorescence changes of split FPs.

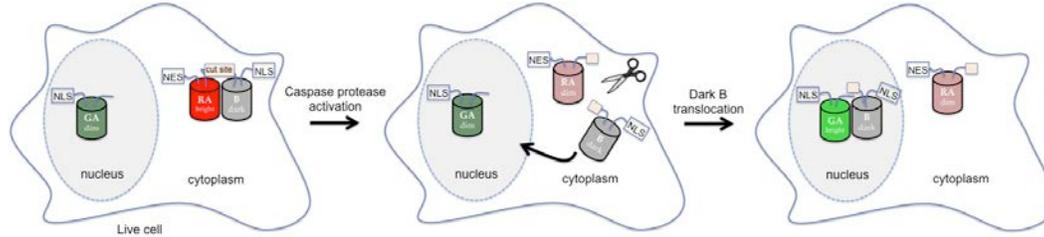
A fortuitous discovery revealed that versions of the B copy optimized to bind to green A copy (GA) and red A copy (RA) partners can pair and induce bright fluorescence with the “wrong A” partner. That is, a green B copy (GB) can associate with RA and enhance its fluorescence and, similarly, red B copy (RB) can bind and increase the fluorescence of GA. This insight led to the development of the fluorescent protein exchange (FPX) biosensor strategy (**Figure 1.6**). This strategy, as recently described by Ding *et al.* (99), is based on the swapping of a B copy from RA to GA or vice versa in response to changes in cellular events like PPIs, enzyme activity, and second messenger signaling. As the fluorescence intensity changes from predominately green when GA interacts with B, to primarily red when RA interacts with B, large ratiometric changes in fluorescence

can be realized. Accordingly, implementation of FPX for biosensing requires the expression of three ddFP monomers (GA, RA and B) in cells either as separate proteins or on one polypeptide chain.

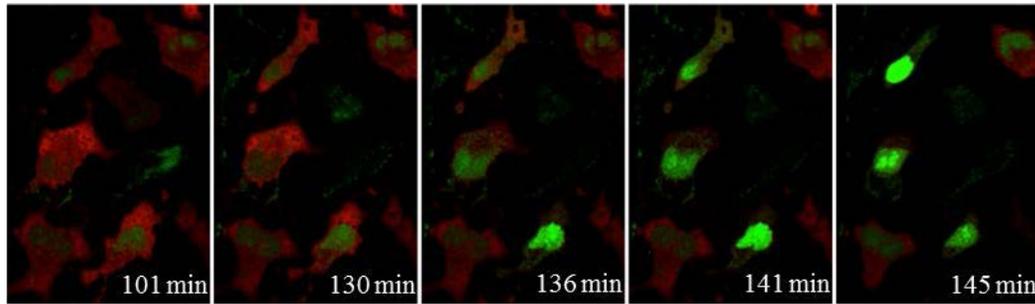


**Figure 1.6 General graphic illustration of fluorescent protein exchange (FPX) strategy.** *A.* The ideal implementation of FPX strategy is the ‘B copy swap’, in which a B copy is co-expressed with two distinctly different fluorogenic A copies (GA and RA as shown in the scheme). Swapping of B copy from a GA-bound state to a RA-bound state or vice versa can be monitored through changes in green-to-red (or vice versa) fluorescent emission ratios. *B.* A putative model for screening protein-protein interaction using FPX strategy. Here, both GA and B are linked to a PPI of interest and expressed as chimeras along with free RA. As the two proteins interact, the ratio of green-to-red fluorescent ratios will substantially increase.

A. FPX translocation strategy for detecting protease activity.



B. Time lapse cell images of caspase activity.



**Figure 1.7 Detection of caspase activity by FPX strategy.** A. Graphical illustration of detecting caspase activity using FPX (B copy translocation). Upon activation, caspase cuts the protease site on the flexible linker between RA and B harboring a nuclear-localization sequence. The B copy then translocates to the nucleus and binds with GA, forming a bright green fluorescence. B. Cell images taken from a time-lapse monitoring of caspase activity in a staurosporine-stimulated HeLa cell undergoing apoptosis (credits to Yidan Ding for the cell images).

The FPX strategy has been used for imaging of dynamic changes in  $\text{Ca}^{2+}$  concentration, cAMP-dependent protein kinase A (PKA) activation, and  $\text{PIP}_2$  hydrolysis into diacylglycerol and 1,4,5-triphosphate ( $\text{IP}_3$ ). It was also applied to

real-time monitoring of caspase activity during apoptosis with a whole cell change in fluorescence color, either with or without translocation of the B copy from the cytoplasm to the nucleus (**Figure 1.7**). These examples establish FPX as a robust and versatile strategy that offers comparable or better qualitative performance than FRET for similar applications. One drawback relative to FRET is that ddFPs are not readily amenable to quantitative measurements. Compared to use of ddFPs alone, FPX exhibits reduced sensitivity to cell-to-cell variations in the concentration of ddFP monomer units, and is amenable to ratiometric-based biosensing.

#### **1.4 Scope of the thesis**

The primary goal of this thesis was to explore applications of FPX technology that do not directly relate to the detection and imaging of biochemical events and PPIs in mammalian cells. Rather, I explored the use of the FPX strategy to detect PPIs in bacteria (*Escherichia coli*), with the expectation that this approach would complement, or overcome certain limitations, associated with existing methods. We expected that the competitive nature of FPX, that is, both green A and red A copies can bind interchangeably with one B copy, would result in more reliable and robust results compared to FRET and split FP strategies. We also benefit from the ease in performing gene manipulation and expressing exogenous proteins in *E. coli*, which is a suitable host for screening libraries of engineered PPIs. With the use of the FPX strategy, and a colony-imaging system, we can validate suspected interacting proteins and develop new therapeutics like

antagonistic proteins and peptides to block aberrant enzymes, receptors or transcription factors.

This thesis describes our ongoing efforts to develop a simple and powerful *E. coli*-based screening strategy for detecting PPIs and discovering potential protein or peptide antagonists. To facilitate its development, this strategy requires bacterial expression plasmids that can efficiently co-express FPX protein fusions at approximately equal quantities. Accordingly, in Chapter 2, we describe the stepwise construction and characterization of four pBAD/His B derived polycistronic vectors. Polycistronic vectors are circular plasmids that are able to express multiple polypeptide chains simultaneously. This can be achieved by installing a defined number of ribosome binding sites (RBS) downstream of the transcription promoter DNA sequence. Of the four vectors we built, two can express two proteins (bicistronic) simultaneously, and two can co-express three separate proteins (tricistronic). Test expression experiments using FPs (i.e., mRFP1, mEGFP and mPapaya) as reporters showed that these modified vectors maintained the tightly regulated expression of their parent vector, pBAD/His B.

Chapter 3 describes the use of the FPX strategy for detecting association or disruption of interacting proteins in *E. coli* colonies. Here, we describe the construction of a bacterial tricistronic plasmid called pFPX which can co-express three proteins (an interacting protein pair and a mutant or protein antagonist) as ddFP fusions. Using pFPX and our custom-made fluorescent colony imaging system, we have successfully monitored the association of E1 and K1 coiled-coils, transactivation domain of p53 and its antagonist HDM 2, and HRas<sub>WT</sub> and

Raf-1<sub>WT</sub> RBD against their binding deficient mutants. To evaluate the strategy's suitability in screening libraries of affinity proteins, we implemented a reversion experiment to rescue the binding of Raf-1<sub>R89L</sub> RBD towards HRas<sub>WT</sub> by screening a small library of randomly mutated Raf-1<sub>R89L</sub> RBD at the 89<sup>th</sup> position. We also screened a library of HRas mutants for variants with restored binding to the R89L mutant of Raf-1 RBD. Moreover, we also demonstrated the suitability of our in-colony FPX strategy for optimizing peptide inhibitor sequences for selective inhibition of NHERF-2 PDZ 2 and MAGI-3 PDZ 6 against their partner GPCR, lysophosphatidic acid receptor 2 (LPA<sub>2</sub>).

Lastly, Chapter 4 summarizes all important results and observations from the preceding Chapters. We thoroughly describe the recommended *in vitro* experiments to validate the results of our in-colony strategy, especially the relative affinities of HRas/Raf-1 RBD mutants and inhibitor peptides identified for both NHERF-2 PDZ 2 and MAGI-3 PDZ 6. In addition, we also include some suggestions on how to utilize our strategy for screening less soluble interacting proteins (specifically receptors and disulfide bond-containing proteins) and optimizing different biologic-based inhibitors other than linear peptides.

## CHAPTER 2

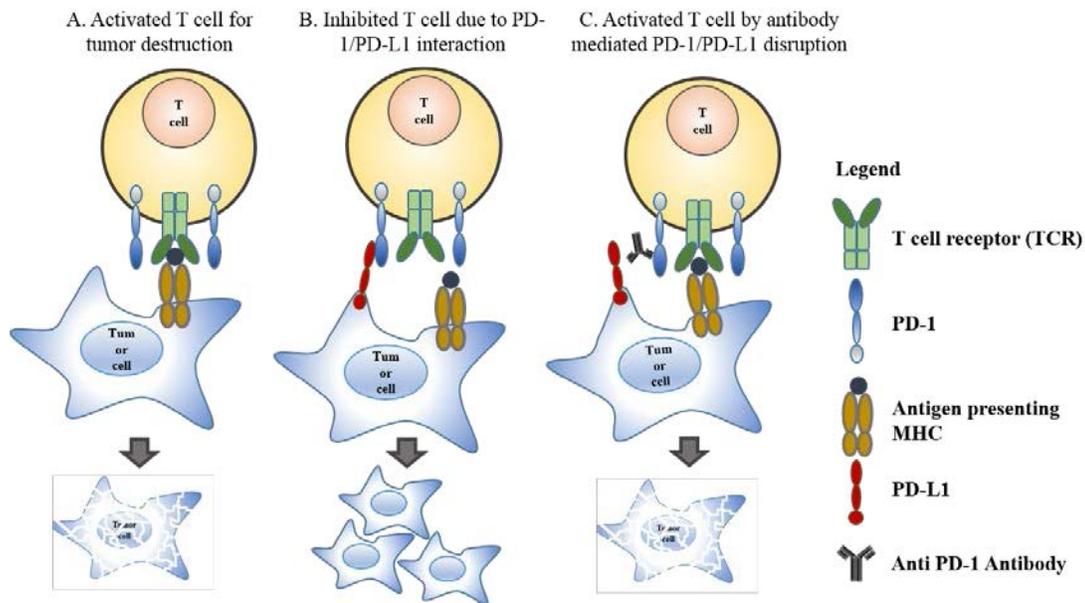
### **Construction and optimization of a polycistronic bacterial vector for the expression of multi-protein complexes**

#### **2.1 Introduction**

Undoubtedly, no other class of macromolecular interactions can compete with the complexity and diversity of interactions between proteins. Protein-protein interactions (PPIs) are involved in many regulatory processes critical to life, such as cell division, immune response, and thousands of biochemical reactions. At a molecular level, these finely tuned protein interactions are facilitated by delicate non-covalent forces, and any changes to their affinity or selectivity brought upon by mutations can drastically impair their functions. These can lead to detrimental changes in cell function, or uncontrolled cell growth that is a key characteristic of cancer (100-102).

For cancer alone, many new small molecule drugs and therapies are designed to target aberrant PPIs. A representative example of a PPI that pharmaceutical companies are exploring as a potential targets for cancer therapy, is the interaction between the surface receptor Programmed Cell Death-1 (PD-1), and its ligand Programmed Cell Death-Ligand 1 (PD-L1) (103,104). These proteins belong to a family of immune checkpoint receptors that inhibit immune cells from eliciting responses to our own tissues and organs. However, some tumors, especially in lung cancer, can effectively deactivate T cells leading to

anti-tumor immunity and evasion. Compelling evidence points to this detrimental immune evasion being caused by an overexpression of PD-L1 on tumor cell surfaces (**Figure 2.1**) (105-108). To date, anti-PD-1 and anti-PD-L1 antibodies are the only therapies currently being administered to patients or undergoing clinical trials (109, 110). The large, relatively flat and featureless nature of the binding interface between PD-1 and PD-L1 binding interface means that it is not readily amenable to being targeted with a small molecule drug that could block the interaction and thereby restore T cell activity (111). Future developments in synthesis strategies and computational drug design will likely be required to arrive at an effective small molecule inhibitor of this interaction (112).



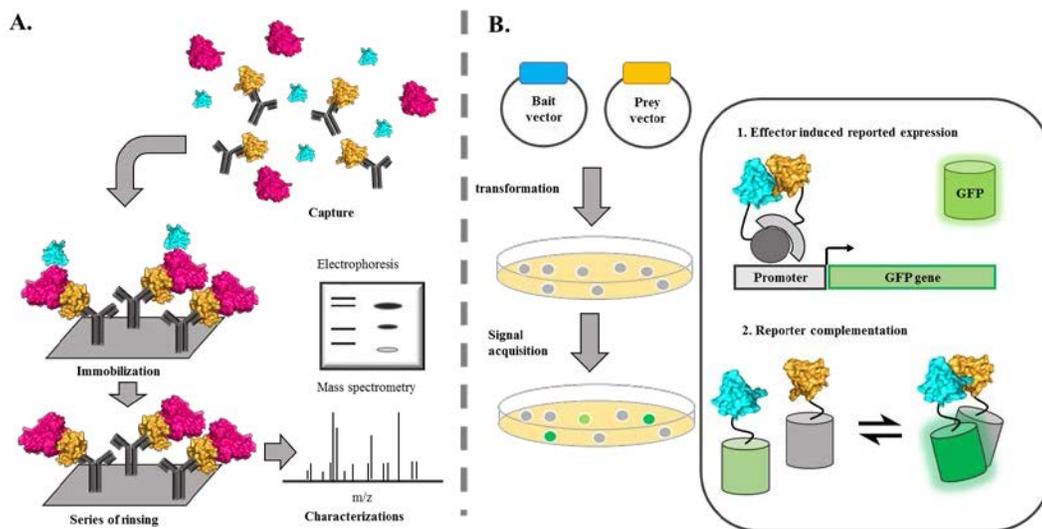
**Figure 2.1 Role of programmed cell death-1 (PD-1) and programmed cell death-ligand 1 in tumor evasion.** A. T cell activation upon binding of its receptor to antigen presenting MHC in tumor cells resulting to tumor cell

destruction. *B.* T cell inhibition through binding of overexpressed PD-L1 on tumor cell surface with PD-1 leading to tumor immune evasion. *C.* Disruption of PD-1/PD-L1 binding with therapeutic antibody reactivates T cell for tumor destruction.

As we identify new proteins, we often find ourselves puzzled with the same question: which other proteins does this protein interact with? It may sound easy, but just answering this question entails a tremendous amount of work that often leads to yet bigger questions. Although this seems a never-ending road, investigating protein interactions will surely reveal many new pathways and networks important to understanding and treatment of disease. Along with the continuous progress in PPI research, screening strategies used to detect and validate suspected interacting proteins are also evolving substantially from the classical co-immunoprecipitation (Co-IP), coupled with electrophoresis or mass spectrometry (MS), to sophisticated genetic assays (*113, 114*). **Figure 2.2** illustrates the general schemes for detecting PPI using Co-IP and a cell-based assay known as a two-hybrid system.

In Co-IP (*115*), an antibody that targets a specific protein is added to clarified cell or tissue lysate. The antibody-protein complex is then precipitated, immobilized on a protein A (or G) coated support and washed several times. The complex is then eluted and characterized by electrophoresis and MS (**Figure 2.2 A**) to identify those proteins that remained associated with the protein of interest. In contrast, protein interactions in two hybrid assays are screened, monitored and

evaluated inside mammalian or yeast host cells, with the help of protein reporters like GFP (116-119). Since the breakthrough development of yeast two-hybrid assay (117), many variations and alternative cell-based PPI assays have been reported. Applications of these assays would include validating suspected PPIs (119), screening genomic libraries (120), and identifying inhibitors of PPI targets (116). Although used in a variety of applications, most common two-hybrid assays generally rely on a similar strategy. As depicted in **Figure 2.2 B**, association of interacting proteins drives the folding or reconstitution of an effector protein or protein complex, which elicits assayable response or triggers the expression of a reporter protein.



**Figure 2.2 Popular screening methods for PPIs.** *A.* Co-IP coupled with electrophoresis and mass spectrometry for identification of interacting partner proteins. *B.* Two-hybrid systems using 1. effector complex reconstitution to induce reporter gene expression and 2. PPI-dependent reporter activation or complementation.

Studying protein assemblies does not end at simply identifying the interacting partners. A detailed mapping of interactions and posttranslational modifications and even structural and biophysical characterizations are necessary to fully comprehend their roles. This entails an expression system that lacks machinery for posttranslational modification, and that can produce large quantities of proteins at a minimal cost. Clearly, *Escherichia coli* expression host still tops all else and remains the most convenient choice (121-123). However, expression of a single subunit of a protein complex in *E. coli* typically results in the protein ending up in insoluble aggregates (inclusion bodies). Proteins may be directed to inclusion bodies due to improper folding or exposure of hydrophobic spots to the solvent. In many cases, folding efficiency can be improved, or hydrophobic spots can be protected from solvent, by co-expressing an interacting partner (124, 125). Established protocols for simultaneous expression of multiple genes include transforming multiple plasmids with compatible replicons and the use of modular polycistronic vectors (122, 126-129). Disadvantages of the former approach include unstable expression due to toxicity, the difficulty of maintaining multiple plasmids, and the added cost of using multiple antibiotics and chemical inducers. Fortunately, the use of polycistronic vectors can alleviate most of these problems. A polycistronic vector contains a defined number of ribosome binding sites (RBS) downstream of the promoter (128, 130, 131). This enables the simultaneous expression of multiple proteins under one promoter and selection marker. Numerous examples of this type of vector have been reported and have been used to successfully express some troublesome proteins. One is example of

such a vector is pST39 designed by Tan and co-workers (128, 129). This T7-based system effectively expressed VHL-elonginB-elonginC, a ternary complex associated with Von Hippel-Lindau syndrome, in *E. coli* at appreciable quantities (128).

In this Chapter, we describe our efforts to construct bicistronic and tricistronic *E. coli* expression vectors that use the tightly controlled pBAD/His-B (Life Technologies) promoter. These polycistronic vectors are expected to provide a better alternative to T7 promoter/lac operator based vectors, especially when expressing toxic proteins. Such a system would facilitate the expression of protein assemblies, co-expression of a protein with chaperones, foldases and lysozymes, or an in-colony two-hybrid screening for improving enzymes and genetically encoded biosensors. We aimed to “tune” these vectors such that the expression system would simultaneously express two or three proteins at approximately equal quantities. The use of these vectors in PPI screening will be described and highlighted in the following Chapter.

## **2.2 Results and discussion**

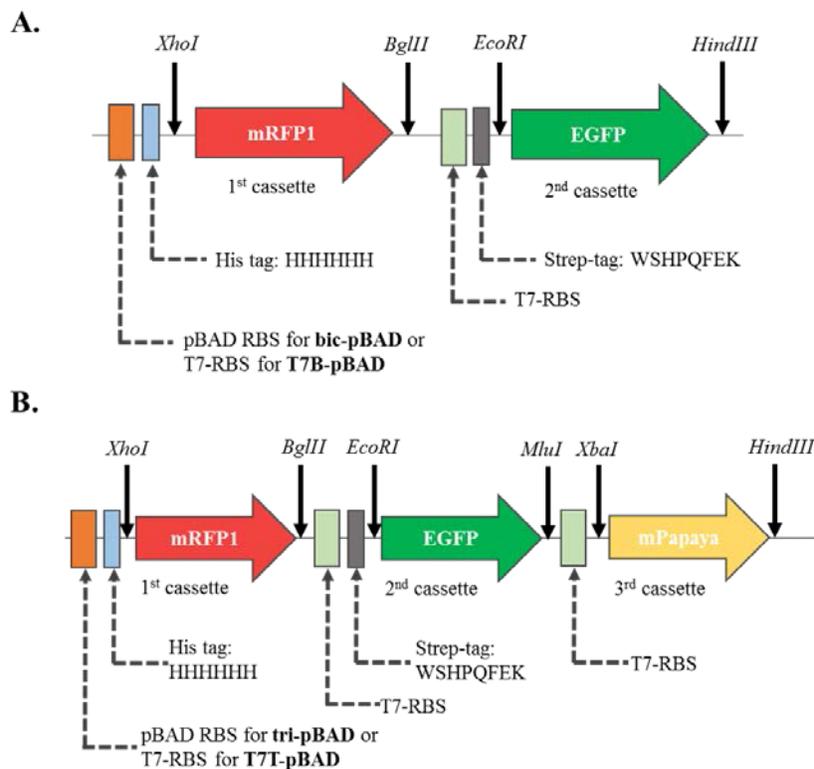
### **2.2.1 Construction of bacterial polycistronic vectors**

Protein complexes are most commonly expressed in *E. coli* through the use of multiple plasmids (122). Although proven to be useful, the strategy suffers from several impracticalities especially when dealing with library construction. First, there is typically a drastic decrease in transformation efficiency when using multiple plasmids, even with highly competent *E. coli* strains such as DH10B.

Second, the use of multiple plasmids complicates routine molecular biology techniques like plasmid purification and sequencing, which is relevant to applications in which genetic libraries are being screened. And third, transformed bacteria grow slower and express lesser quantities of proteins, perhaps due to instability of plasmids and toxicity from using multiple antibiotic selection markers. To avoid these drawbacks during screening of library clones, we opted to design bicistronic and tricistronic vector systems that can co-express two and three proteins, respectively. Although polycistronic vectors have been previously reported, our vectors are derived from pBAD/His B that offer a tighter expression control compared to the pET derived counterparts. Tight control of expression is often relevant when producing toxic proteins.

The bicistronic pBAD vector (bic-pBAD) was constructed via a three-way ligation assembly using pBAD/His B as backbone and two gene fragments encoding for red (mRFP1) and green (EGFP) FPs (**Figure 2.3 A**). We intentionally chose FPs as templates for vector construction in order to facilitate the screening of correctly-ligated clones as well as to assess the expression level of each cassette upon chemical induction. The assembled bic-pBAD vector carries all the important elements of pBAD/His B, including an ampicillin resistance gene, the tightly-controlled pBAD transcription promoter, and an mRNA translation enhancer sequence that specifically initiates the expression of the first cassette (mRFP1 in this case). Translation of EGFP on the second cassette is initiated by a short DNA sequence called T7-RBS located upstream of the gene. T7-RBS is a pyrimidine-rich sequence derived from bacteriophage T7 gene 10

that has been reported to induce protein biosynthesis efficiently in *E. coli*. Although its mechanism still remains unclear, it has been proposed that the T7-RBS enhances the expression by complimentary base-paired interaction with the 460 region of 16S rRNA in *E. coli* (132, 133). In addition to the enhancer element, we maintained the His<sub>6</sub> tag on the first cassette and added a streptavidin-binding peptide, WSHPQFEK (134), on the N-terminal of the second cassette. These features could further simplify the detection and affinity purification of the protein of interest with the use of tag-specific antibodies, and Ni-NTA or Streptactin-II purification system, respectively (134, 135).



T7-RBS sequence: TTAAC TTTAAGAAGGAG

**Figure 2.3 Linear maps of the cloning region of pBAD-based polycistronic vectors.** All relevant features associated with each expression cassette like

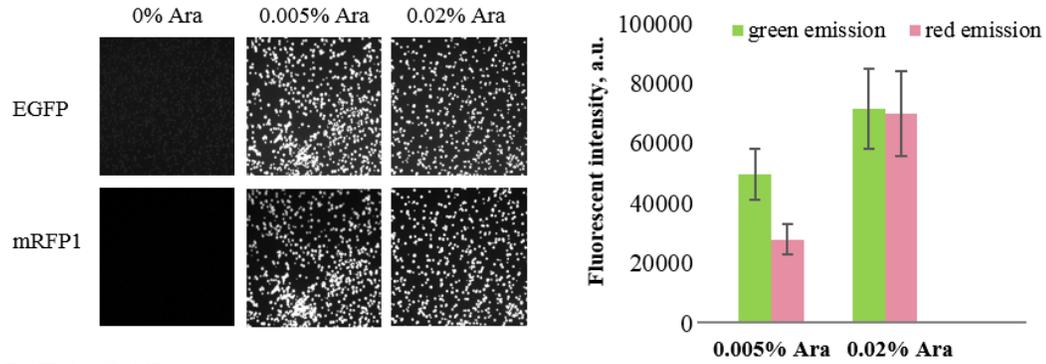
translation element, affinity tags and restriction endonuclease sites are shown on both maps: *A.* Bicistronic pBAD *B.* Tricistronic pBAD.

Building the tricistronic pBAD vector (tri-pBAD) followed a similar strategy to the bicistronic version. We installed the third cassette by inserting two gene fragments, *mRFP1-T7RBS-EGFP* and *T7RBS-mPapaya*, into a modified pBAD/His B (*Mlu I* site removed) via a three-way ligation assembly. We decided to remove the *Mlu I* site on the O<sub>2</sub> binding region of pBAD backbone so we could use this restriction endonuclease site on the second expression cassette. This gene assembly generated tri-pBAD, which could express three separate proteins simultaneously upon inducing a culture of transformed *E. coli* with L-arabinose. As with bic-pBAD, tri-pBAD also retained all the necessary features found in pBAD. Key features of its expression cassettes such as translation enhancer elements, restriction enzyme sites and affinity tags are all illustrated in **Figure 2.3 B**.

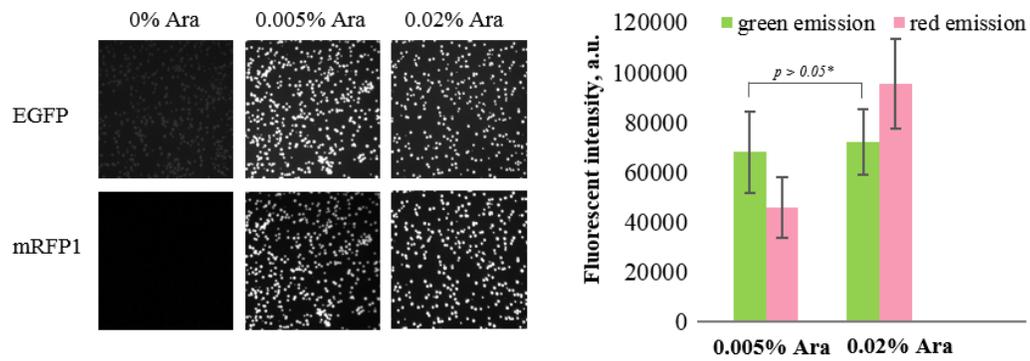
To substantially even out the quantity of protein produced on each cassette, we opted to replace the pBAD translation enhancer element of the first cassette on each polycistronic vectors with T7-RBS. As discussed earlier, T7-RBS produces relatively more protein than any other translation enhancer sequences like the one in pBAD owing to its higher affinity to 16S rRNA in *E. coli*. These new vectors would be useful if uniform cytoplasmic concentration of each protein is required. For simplicity, we named these modified polycistronic vectors as

T7B-pBAD (T7-RBS bicistronic pBAD) and T7T-pBAD (T7-RBS tricistronic pBAD).

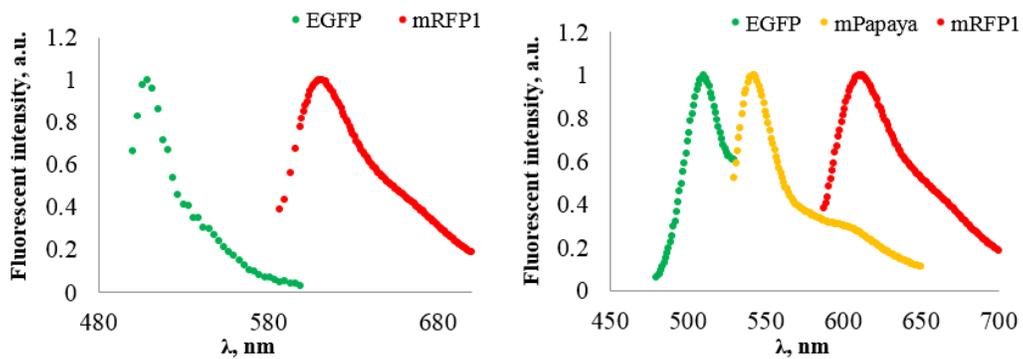
### A. Bic-pBAD



### B. Tri-pBAD



### C. Emission spectra of cell lysate from bic-pBAD (left) and tri-pBAD (right)



## **Figure 2.4 Test co-expression of FPs using polycistronic pBAD vectors.**

*Left panel* on both bic-pBAD (*A*) and tri-pBAD (*B*) shows the representative images of transformed *E. coli* colonies grown on agar plates supplemented with L-arabinose (0%, 0.005%, 0.02%). The *right panel* on both *A* and *B* are the average green and red emission intensities measured from the digital images of fluorescent colonies. *C*. Emission spectra taken from B-PER (Thermo Scientific) soluble fraction of the cell lysate from transformed *E. coli* cultures. The *left panel* is for bic-pBAD (EGFP and mRFP1) and the *right panel* is for tri-pBAD (EGFP, mPapaya and mRFP1).

### **2.2.2 Characterization of polycistronic vectors**

In addition to analytical digestions and DNA sequencing, we performed basic test expression experiments to demonstrate, characterize and verify the suitability of these polycistronic vectors in producing multiple proteins simultaneously (**Figure 2.4**). We transformed the competent DH10B *E. coli* cells with bic-pBAD and tri-pBAD plasmids and induced the expression of FPs on colonies by using agar plates supplemented with various concentrations of L-arabinose. The left panels on both **Figures 2.4 A** and **B** shows the representative fluorescent images of bacterial colonies incubated overnight on plates containing 0%, 0.005% and 0.02% L-arabinose. We observed that colonies grown on 0% L-arabinose plates registered no substantial fluorescence when imaged through a digital microscope. This simple assay confirmed that the introduction of T7-RBS translation elements and other alterations on the cloning region has not affected

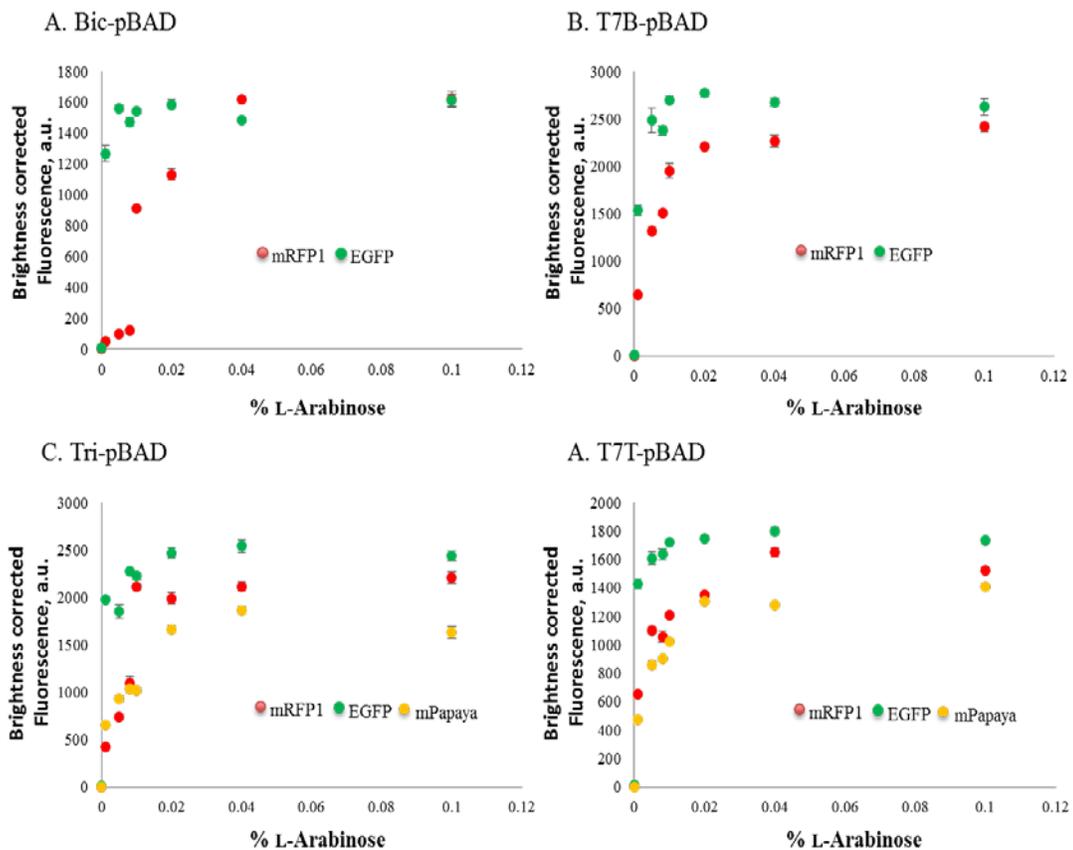
the tightly-regulated expression of any polycistronic vectors we built. However, when colony images were processed using ImageJ Pro software, we recorded weaker green emission signals from colonies as compared to their red emission. We found that the green fluorescence was similar to the autofluorescent background emitted by agar media and colonies not expressing a FP. We observed significant co-expression of mRFP1 and EGFP on bacterial colonies grown on agar plates with 0.005% and 0.02% L-arabinose. The co-expression was detected as the bright red and green fluorescence emission from each colony. Using the macroscope imaging system, it is not possible to visualize the expression of mPapaya, given the spectral bleedthrough from EGFP and mRFP1. It is possible that equipping the imaging system with a set of narrower emission filters could enable us to take a fluorescent image of the mPapaya channel without the interferences from EGFP and mRFP1. Given this limitation, we decided to acquire fluorescence spectral data from the B-PER soluble fractions of each cell culture (bic-pBAD and tri-pBAD transformed cells) to verify the presence of mPapaya. As depicted on **Figure 2.4 C** (right panel), emission spectra of clarified cell lysate from tri-pBAD transformed cell culture confirm that mPapaya ( $\lambda_{em}$  max. at 541 nm) is indeed expressed along with two other FPs, EGFP ( $\lambda_{em}$  max. at 507 nm) and mRFP1 ( $\lambda_{em}$  max. at 607 nm). As expected, the distinct 541 nm emission of mPapaya was not observed from a B-PER soluble fraction of bic-pBAD transformed cells (**Figure 2.4 C**, left panel).

In an effort to assess the induction efficiency of polycistronic vectors, we attempted to correlate the amount of L-arabinose on agar plates with the

expression levels of mRFP1 and EGFP in transformed colonies. For this colony-based screening, we acquired and processed digital images of fluorescent colonies grown on plates supplemented with 0.005% and 0.02% L-arabinose, and quantified and compared the fluorescence intensities of mRFP1 and EGFP between the two plates (**Figures 2.4 A and B**). Emission acquired from both plates revealed that the expression of mRFP1 increased about 2-fold upon going from 0.005% to 0.02% L-arabinose. In contrast, EGFP only displayed an approximately 1.3-fold increase for bic-pBAD and a no significant change in emission ( $p > 0.05$ ) for tri-pBAD transformed colonies. We believe that this observed pattern is likely dependent on the translation enhancer sequences that control the biosynthesis of FPs in *E. coli*. We suggest that the ribosome-binding site (RBS) on the first cassette binds less tightly with 16S rRNA compared to the T7-RBS sequence. Following this hypothesis, increasing the amount of mRNA in the cytoplasm by adding higher concentrations of chemical inducer would benefit the weaker-affinity pBAD RBS more than the higher-affinity T7-RBS.

We also performed a similar expression assay on bacterial cell culture using a wider range of inducer concentrations. This assay provided more quantitative information on the induction profiles of all four polycistronic vectors compared to the colony-based screening described earlier. **Figure 2.5** illustrates the fluorescence saturation plots of bacterial cell cultures transformed with FP-expressing polycistronic plasmids and induced with varying concentrations of the chemical inducer (0%-0.1% L-arabinose). Our results revealed a plateauing of expression starting at 0.02% L-arabinose which is indeed the recommended L-

arabinose concentration for overexpressing proteins in pBAD/His B. In addition, we observed that all proteins expressed under T7-RBS have approached near saturation even with L-arabinose concentration lower than 0.02%. However, mRFP1 expressed under the intrinsic pBAD translation element such as in bic-pBAD and tri-pBAD showed a steady yet slow increase in fluorescence until it reached saturation. This observation suggests that expression of genes controlled by T7-RBS produce more proteins even at lower inducer concentration than those are regulated by pBAD translation enhancer sequence. These also supplemented the results taken by the digital microscope (**Figure 2.4 A and B**).



**Figure 2.5 Saturation plots of FPs expressed cell cultures with varying concentrations of L-arabinose. A. Bic-pBAD B. T7B-pBAD C. Tri-pBAD and D. T7T-pBAD.**

### **2.3. Conclusion**

This Chapter describes our efforts in developing pBAD/His B-based polycistronic bacterial vectors that can effectively co-express two or three proteins upon induction. We constructed a total of four vectors that can be categorized as either bicistronic (bic-pBAD and T7B-pBAD) or tricistronic (tri-pBAD and T7T-pBAD). We also designed expression cassettes on each vector to carry pBAD or T7-RBS translation enhancer elements, distinct restriction enzyme cut sites and affinity tags for effective molecular cloning, induced overexpression and purification. Furthermore, we evaluated their expression profiles and fortunately observed that all polycistronic vectors have maintained the induction characteristics of their predecessor pBAD/His B even after a series of alterations within the cloning region. Creating these vectors would simplify the protocols and minimize the challenges currently encountered by common co-expression strategies to date. These vectors would complement the existing pET polycistronic vectors especially when a tightly-controlled expression is required. Meanwhile, polycistronic pBAD vectors containing all T7-RBS enhancers can be used when equivalent expression levels of all proteins are desired.

## 2.4 Materials and methods

### 2.4.1 General materials and reagents

All synthetic DNA oligonucleotides (**Table 2.1**) used for molecular cloning were purchased from Integrated DNA Technologies (Coralville, IA). Working solutions of deoxyribonucleotide triphosphates (dNTPs) mix were reconstituted from stock solutions provided by Life Technologies. Enzymes for polymerase chain reaction (PCR) like *Pfu* polymerase (Thermo Scientific), *Taq* and Q5 polymerases (New England Biolabs) were used according to manufacturer's recommendations. Routine DNA digestion and ligation of gene products were performed using FastDigest restriction endonucleases and T4 DNA ligase (both from Thermo Scientific), respectively. GeneJet (Thermo Scientific) and EZ-10 Spin Column (Biobasic, Inc.) DNA extraction kits were used to purify DNA either from miniprep or agarose gel electrophoresis. *E. coli* strain ElectroMAX DH10B (Life Technologies) was routinely used for plasmid construction and propagation, and recombinant protein production. For construction of bacterial polycistronic vectors, we used pBAD/His B (Life Technologies) as their plasmid backbones and amplified 3 genes of FPs to complete the assembly of their respective expression cassettes. Monomeric red FP (mRFP1) was copied from pBbA2c-RFP (Addgene, #35326) while genes for EGFP and mPapaya were amplified from plasmids currently in our lab.

**Table 2.1 Oligonucleotide used for cloning and sequencing.**

| <b>Oligo name</b>  | <b>Type<sup>a</sup></b> | <b>DNA sequence</b>  |
|--------------------|-------------------------|--|
| ddRFP_HindIII      | R                       | GTG AAG CTT TTA CTT GTA CAG CTC GTC<br>CAT GCC             |
| EGFP/RA_MluI       | R                       | TGG ACG CGT CCT AGG TTA CTT GTA CAG<br>CTC GTC CAT GCC GAG |
| mpapaya_xbaI       | F                       | CCC TCT AGA AAT GGT GAG CAA GGG CGA<br>GGG GCA A           |
| pBAD_MluI          | F                       | GCC ATG ACA AAA ACG CGT AAC AAA AGT G                      |
| pBAD_MluIremove_QC | F                       | GGA CCA AAG CCA TGA CAA AAG CGC GTA<br>ACA AAA GTG TCT     |
| pBAD_NcoI          | F                       | TGG GCT AAC AGG AGG AAT TAA CCA TGG<br>GG                  |
| pBAD_F             | F                       | ATGCCATAGCATT TTTATCC                                      |
| pBAD_R             | R                       | ACTCAGGAGAGCGTTCAC   |
| RB_xhoI            | R                       | TGG CTC GAG TTA CTT GTA CAG CTC GTC<br>CAT GCC             |
| RFP_BglII          | R                       | AGC AGA TCT TTA AGC ACC GGT GGA GTG<br>ACG                 |
| RFP_xhoI           | F                       | CCA CTC GAG AAT GGC GAG TAG CGA AGA<br>CGT T               |
| Seq_p15A           | R                       | AGT CTT TCG ACT GAG CCT TTC G                              |

---

|                    |   |   |
|--------------------|---|---|
| SeqF_2ndcassette   | F | ACA GGT AGC ACA GGC AGC G   |
| SeqF_3rdcassette   | F | AGATATACCATGGGTCTAGACATG  |
| SeqR_Streptag      | R | CTG CGG GTG GCT CCA GCT A   |
| StreptagMVSK_EcoRI | F | ATG GCT AGC TGG AGC CAC CCG CAG TTC<br>GAA AAA GTG AAT TCA ATG GTG AGC AAG<br>GGC GAG |
| T7RBS_BglII        | F | CCA GAT CTG TAA TTT TGT TTA ACT TTA<br>AGA AGG AGA TAT ACC ATG GCT AGC TGG<br>AGC CAC |
| T7RBS_MluI         | F | CCA CGC GTG TAA TTT TGT TTA ACT TTA<br>AGA AGG AGA TAT ACC ATG GGT CTA GAC<br>ATG ACA |
| T7RBS_pBAD         | R | ACC CAT GGT ATA TCT CCT TCT TAA AGT<br>TAA AAC GGG TAT GGA GAA ACA GTA GAG<br>AGT TGC |

---

<sup>a</sup>Oligo type: F= forward and R = reverse.

## 2.4.2 DNA recombinant techniques

All DNA manipulations including gene amplification, restriction enzyme digestion, ligation and propagation were performed according to Sambrook *et al.* (115). Standard PCR amplification carried out using *Pfu*, *Taq* or Q5 polymerase was executed in a 50  $\mu$ L reaction mixture containing 1 $\times$  reaction buffer (supplemented with MgSO<sub>4</sub>), 1.5  $\mu$ L DMSO, 200  $\mu$ M dNTPs, 200 nM forward

and reverse oligos, 10-100 ng DNA template, nuclease free water and 1.0 unit of polymerase. Cycling parameters for both *Pfu* and *Taq* were as follows: initial denaturation at 95 °C for 60 seconds; 35 cycles of 95 °C for 15 seconds, 54-60 °C for 30 seconds, 72 °C for 60 seconds per kb of gene target; final extension at 72 °C for 5 minutes. However, running protocol recommended for Q5 polymerase was slightly different compared to *Pfu* and *Taq*. It started with an initial denaturation at 98 °C for 30 seconds followed by 35 cycles of 98 °C for 5 seconds, 54-60 °C for 30 seconds, 72 °C for 30 seconds per kb of gene target and then a final extension at 72 °C for 2 minutes.

Typical restriction enzyme digestion was carried out in a 50 µL reaction mixture containing less than 10% v/v of restriction enzymes, 1× green digestion buffer and nuclease free H<sub>2</sub>O. Reaction was performed in a thermocycler set at 37 °C for 4 hours followed by deactivation at 80 °C for 10 minutes and cooling at 10 °C for 30 minutes. However, digestion containing *BamHI* or *EcoRI* restriction endonuclease enzymes was only incubated to a maximum of 2 hours at 37 °C prior to deactivation and cooling. All ligations were completed in a 20 µL solution containing nuclease free water, 1.0 unit of T4 DNA ligase, 1× ligase buffer and 1:6 mole ratio of digested vector and insert. For a three-way ligation assembly, 1:6:6 mole ratio of digested vector and 2 inserts were followed. Ligation mixtures were incubated at 22 °C for 10-15 hours, deactivated at 80 °C for 10 minutes and stored at 4 °C prior to bacterial transformation.

Site-directed mutagenesis was performed using Quikchange Lighting kit from Agilent. Each reaction was carried out using half of the manufacturer's

recommended reaction volume. Typical cycling parameters used on the amplification were as follows: initial denaturation at 95 °C for 2 minutes; 30 cycles of 95 °C for 20 seconds, 58-60 °C for 30 seconds, 72 °C for 30 seconds per kb of plasmid target; final extension at 72 °C for 5 minutes. Digestion of template was executed on the same reaction mixture by spiking 0.5 µL of *DpnI* provided on the kit followed by incubation at 37 °C for 10 minutes and cooling at 10 °C for 10 minutes. Digested mixture was stored at 4 °C prior to bacterial transformation.

Electrocompetent *E. coli* strain DH10B was used for routine plasmid propagation and recombinant protein production. Thawed cells were spiked with 2-3 µL of ligation mixture or 0.5 µL of plasmid. Electroporation was executed using 0.2 cm MicroPulser cuvettes (Bio-Rad) and MicroPulser electroporator (Bio-Rad). Pulsed cells were then incubated and grown on agar plates containing 400 ppm ampicillin and 0.02% L-arabinose overnight at 37 °C.

### **2.4.3 DNA sequencing**

All DNA sequencing were performed either at University of Alberta Molecular Service Unit (MBSU) or DNA Core Services of University of Calgary. First expression cassette of all polycistronic vectors were sequenced using either pBAD-F or pBAD\_NcoI as forward oligo and SeqR\_Streptag as the reverse oligo. Second cassette on all constructs used SeqF\_2<sup>nd</sup>cassette forward primer and pBAD-R reverse primer for bicistronic vector. However, tricistronic vector did not have a specific reverse sequencing oligo for the 2<sup>nd</sup> expression cassette. Third expression cassette of both tricistronic vector were sequenced using

SeqF\_3<sup>rd</sup>cassette as forward oligo and pBAD-R as the reverse oligo. Sequence chromatograms were visualized using Sequence Scanner v1.0 from Applied Biosystems while sequence translation and detailed analysis were performed using the sequence analyzer program found on Addgene webpage.

#### **2.4.4 Construction of bacterial bicistronic vector**

Bicistronic vector (bic-pBAD) was constructed via three-way ligation of an *XhoI/HindIII* digested pBAD/His B vector and 2 inserts, *XhoI-mRFP1-BglII* and *BglII-EGFP-HindIII*. RFP as copied from pBbA2c-RFP using RFP\_choI and RFP\_BglII as forward and reverse oligos, respectively. On the other hand, EGFP was amplified twice using T7RBS\_BglII and StrepMVSK\_EcoRI as forward oligos and ddRFP\_HindIII as reverse oligo. Ligation mixture was transformed in DH10B, grown on agar plate supplemented with 400 ppm ampicillin and 0.02% L-arabinose, and incubated overnight at 37 °C. Colonies were imaged under the digital microscope, a custom-built colony imaging system developed in our lab. *E. coli* colonies exhibiting both green and red fluorescence were picked for further characterization.

Construction of a bicistronic vector with T7-RBS on both cassettes followed the same scheme just mentioned. However, it necessitated a prior modification on the regulatory DNA sequence of the pBAD/His B expression cassette. This was achieved by amplifying the upstream region (130<sup>th</sup>-320<sup>th</sup> base pairs) of the vector using pBAD\_MluI and T7RBS\_pBAD as forward and reverse oligos, respectively. The latter primer encoded a short DNA sequence for T7

translation enhancer and RBS that would replace the inherent translational sequences present in pBAD/His B. The resulting DNA product was digested with restriction enzymes *MluI* and *NcoI* and ligated with the digested vector. The product of this modification, named as T7pBAD vector, was used in constructing the T7-RBS bicistronic vector (T7B-pBAD).

#### **2.4.5 Construction of bacterial tricistronic vector**

The two bacterial tricistronic vectors were built using pBAD/His B and its variant T7pBAD as their backbones. However, these vectors required a removal of *MluI* site found on the upstream sequence of the expression cassette. Site-directed mutation of that restriction endonuclease site was achieved by Quikchange Lighting kit using pBAD\_MluIremove\_QC as the mutagenic oligo. Analytical digestion and DNA sequencing were performed to selected clones to verify the absence of *MluI* site. Tricistronic vectors built from modified pBAD/His B (tri-pBAD) and T7pBAD (T7T-pBAD) can simultaneously overexpress three recombinant proteins. For tri-pBAD, first cassette relies on the translation enhancer inherent from pBAD/His B while the last two cassettes are under T7-RBS translation enhancer. On the other hand, all expression cassettes present in T7T-pBAD are controlled by T7-RBS enhancer sequences.

Construction of tricistronic vectors (tri-pBAD and T7T-pBAD) followed a similar procedure with its bicistronic version. First insert, *XhoI-mRFP1-T7RBS-EGFP-MluI*, was amplified from a bicistronic pBAD vector using RFP\_xhoI and EGFP/RA\_MluI as forward and reverse oligos, respectively. The second insert,

*MluI-T7RBS-mPapaya-HindIII*, was copied from an mPapaya-encoding plasmid using T7RBS\_MluI forward primer and ddRFP\_HindIII reverse primer. This latter PCR product completed the last expression cassette on the tricistronic vector by tagging the gene with T7RBS translation enhancer sequence upstream of mpapaya. The digested PCR fragments and pBAD/His B vector were cloned via three-way ligation and then transformed in DH10B *E. coli* strain. Analytical digestions and DNA sequencing were performed to identify clones with correct plasmid ligation. Finally, recombinant production of 3 FPs was employed to verify the vector's suitability in expressing different proteins simultaneously.

#### **2.4.6 Test expression of polycistronic pBAD vectors.**

To verify the co-expression efficiency of all polycistronic vectors, test expression of FPs as performed in *E. coli*. DH10B strain was transformed with a polycistronic plasmid expressing mRFP1 and EGFP (bic or T7B-pBAD) or mRFP1, EGFP and mpapaya (tri or T7T-pBAD). Colonies were grown on agar plate supplemented with 400 ppm ampicillin and 0.02% L-arabinose (optional). After an overnight incubation at 37 °C, a colony was inoculated in a 10 mL starter medium (LB broth containing 100 ppm ampicillin) and further incubated overnight at 37 °C under rotary shaking (250 rpm). Five milliliter of the overnight starter culture was reconstituted in a fresh 100 mL LB broth (with 100 ppm ampicillin) and incubated for 2-3 hours under 250 rpm rotary shaking. When optical density at 600 nm (OD600) reached 0.8, the culture was divided to different sterile culture tubes and induced with varying volumes of 20% L-

arabinose (final inducer concentration range of 0%-0.1%). After 3 hours of protein production at 37 °C, bacterial cultures were pelleted down, washed and resuspended with 50 mM phosphate buffer (pH 7.5). Characteristic fluorescence of culture samples were measured using Safire2 microplate reader (Tecan). Relevant acquisition parameters such as excitation and emission wavelengths are summarized in **Table 2.2**. For plotting fluorescence saturation curves, emission intensities measured at  $\lambda_{em}$  maximum were normalized against the OD600 of the bacterial suspensions.

**Table 2.2 Spectral properties of RFP, EGFP and mpapaya.**

| <b>FP</b>     | <b>Excitation (<math>\lambda_{ex}</math>)<br/>maximum, nm</b> | <b>Emission (<math>\lambda_{em}</math>)<br/>maximum, nm</b> |
|---------------|---|---|
| EGFP (136)    | 488   | 507   |
| mPapaya (137) | 530   | 541   |
| mRFP1 (138)   | 584   | 607   |

Emission spectrum of each FP was taken from the cleared cell lysate. After 3 hours of protein production, *E. coli* culture was pelleted down using a centrifuge (13,000  $\times$  g) and lysed with bacterial protein extraction reagent (BPER, Thermo Scientific). Cell debris was centrifuge for 15 minutes at 13,000  $\times$  g and discarded while supernatant solution containing soluble fractions of FPs were characterized using Safire2 microplate reader (Tecan). To capture a fuller emission spectrum, we excited each protein 15-20 nm lower than its  $\lambda_{ex}$  maximum

and acquired emission data points starting from 20-30 nm lower than the  $\lambda_{em}$  maximum.

## Chapter 3

### Screening of protein-protein interactions and peptide-based inhibitors by fluorescent protein exchange (FPX)

#### 3.1 Introduction

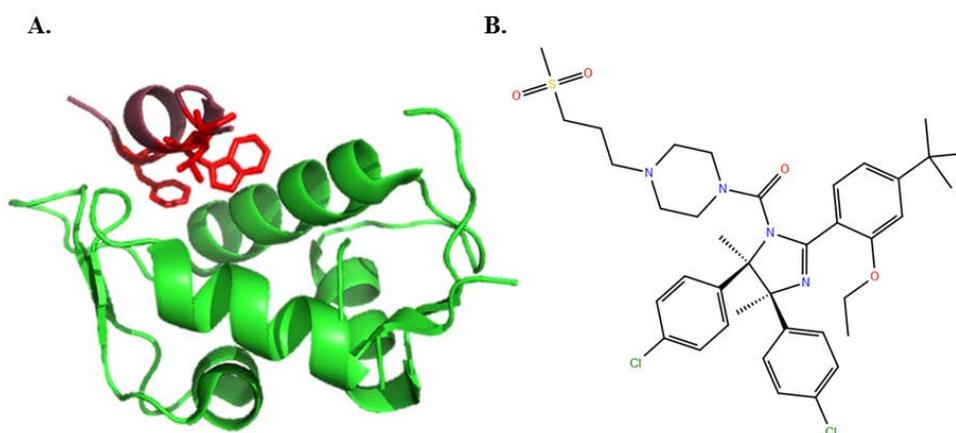
The emergence of rationally targeted therapies (139), together with advanced molecular diagnostics, promises a new era of accurate, powerful and personalized approaches to medical treatments (140-144). Ideally, patients would undergo a thorough profiling of their genetic characteristics to narrow down possible causes of the disorder and determine their response variability to available drugs. Once the target is identified, usually an aberrant protein, a drug that selectively blocks and disables its abnormal activity is administered either as a single agent or in combination with other standard treatments. This strategy offers an unprecedented opportunity for medical practitioners to effectively treat serious ailments such as diabetes, neurodegeneration, and cancer. However, our current information about the activity and interactions of all the molecular targets that could contribute to these complex diseases is very limited. Even if all the targets were known, there remains the immense and expensive hurdle of discovering a drug for each of these targets.

Some of the most common molecular targets identified in many diseases are cell surface receptors and kinases involved in intracellular signaling pathways (145-148). These proteins are often involved in numerous complex yet well-

orchestrated protein-protein interactions (PPIs). However if these interactions are compromised due to mutations, these could result in abnormal metabolism, faulty signaling pathways and uncontrolled cell growth that typify many severe disorders. Accordingly, many pharmaceutical companies invest much of their research efforts in developing new drugs that target aberrant PPIs.

In cancer, targeted therapies that perturb abnormal PPIs are beginning to replace established chemotherapies that are nonspecific and highly toxic (104, 149, 150). This new generation of targeted drug therapies, which include small molecules, peptides and monoclonal antibodies (mABs), are now being commonly used for treating breast, colorectal and lung cancers, as well as chronic lymphoid leukemia and metastatic melanoma. In fact, the United States Food and Drug Administration (FDA) has already approved 15 therapeutic mABs including Genentech/Roche's bevacizumab and ado-trastuzumab emtansine for metastatic colorectal cancer and HER-2 positive metastatic breast cancer, respectively (151, 152). Although blocking PPI interfaces appears to be an exciting and promising area in cancer therapy, achieving this task still presents a number of challenges. The binding hotspots between interacting proteins are normally large, flat, and relatively featureless, compared to other drug targets like kinases and GPCRs (G-protein coupled receptors), which have grooves and deep pocket binding sites (147, 153, 154). Many of the conventional approaches for designing small molecule drugs are poorly suited to finding molecules that can modulate PPIs, leading to the general impression that PPIs are "undruggable". Nonetheless, there have been a number of advances in recent years, and there are currently more than

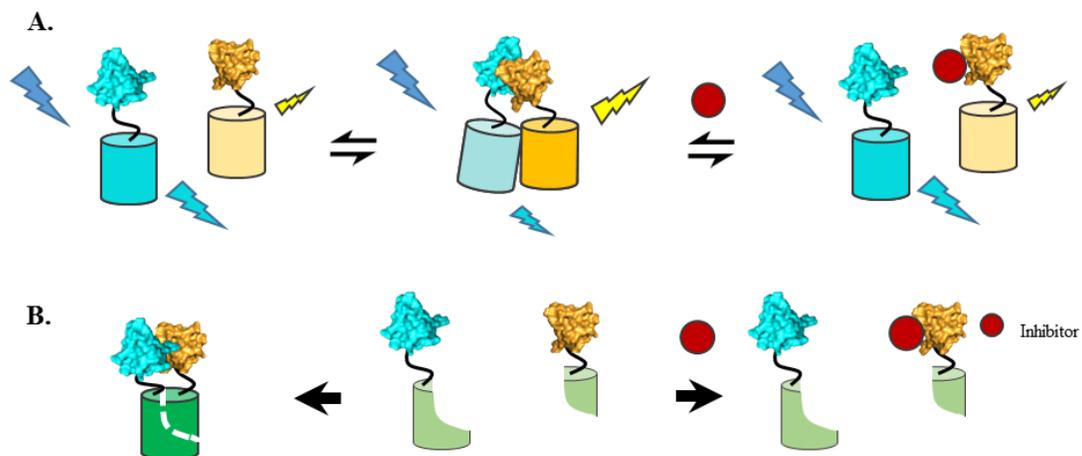
10 PPI modulators in clinical trials. One of these is RG7112 (155), an orthosteric inhibitor for oncogenic interaction of p53-HDM 2 (Figure 3.1). Tumor suppressor p53 is a key player in regulating cell cycle progression and apoptosis. However, in some cancers this protein is inactivated, or deregulated by an overexpression of HDM 2, an E3 ubiquitin protein ligase that targets p53 for proteosomal degradation (156). RG7112 restores the activity of p53 by blocking the binding pocket of HDM 2, thus, deactivating the protein's antagonistic activity against p53.



**Figure 3.1 Structures of p53-HDM X and its inhibitor RG7112.** *A.* Crystal structures (PDB ID 2Z5S) of the transactivation domain of p53 (purple) bound to its antagonist HDM X (green) (157). Phe, Trp and Leu (all in red) are the direct binding residues of p53. *B.* Structure of the inhibitor RG7112.

Another protein of high relevance to cancer is K-Ras. K-Ras, together with other Ras isoforms, is considered the most notorious oncogenic protein and it is mutated in more than 30% of human cancers (158-160). Despite its notoriety,

thirty years of intensive research has not produced any pharmacologically effective antagonist for K-Ras, traditionally placing it near the top of the list of undruggable proteins. However in 2013, Ostrem *et al.* (161) reported small molecule inhibitors that specifically and irreversibly bind to G12C K-Ras, a common oncogenic mutant. These inhibitors drastically diminished K-Ras' binding to GTP, thereby decreasing its activity towards downstream effectors. Another K-Ras signaling inhibitor, named deltarasin, was reported by Zimmerman *et al.* on the same year (162). Although not a direct binder, deltarasin successfully suppressed *in vivo* proliferation of K-Ras dependent human pancreatic ductal adenocarcinoma cells. This inhibitor binds to prenyl-binding protein PDE $\delta$ , which facilitates localization of K-Ras to the plasma membrane.



**Figure 3.2** General high-throughput screening (HTS) assay schemes using FRET and PCA systems. *A.* FRET assay uses donor and acceptor FPs linked to a known or suspected protein partners. FRET is observed upon binding of the interacting proteins, but it is lost or attenuated in the presence of an inhibitor or a competing protein. *B.* PCA utilizes two inactive fragments of a reporter (GFP in

this case). If PPI is not inhibited, the reporter reconstitutes and forms an intact fluorescent protein.

As the interest in validating suspected PPI targets and identifying effective inhibitors increases, a number of high-throughput screening (HTS) methods have been developed. Common strategies include Förster resonance energy transfer (FRET) and protein complementation assays that are either cell-based or in multiwell plate format (**Figure 3.2**).

In a typical FRET assay, a donor-acceptor pair of fluorescent dyes or proteins is linked to interacting partners either chemically or genetically, and emission ratios are measured and interpreted as a function of PPI (44, 45). In 2011, Song *et al.* (163) reported the use of CyPet and YPet for detecting the interaction between SUMO 1 and Ubc9 *in vitro*. SUMO 1 and Ubc9 participate in the sumoylation pathway, a posttranslational modification important for protein trafficking and localization. Upregulation of these proteins is occasionally observed in cancer and neurodegenerative diseases (164-167). Song *et al.* prepared a stable cell line expressing the FRET construct and used it to demonstrate the assay's feasibility in HTS of 2000 compounds in a 384-well plate format. Another FRET-based HTS developed by Schaap *et al.* (168) aimed to identify inhibitors for oncogenic Keap1-Nrf2 interactions. To achieve this goal, they utilized FRET between ECFP and EYFP by fusing it to the kelch domain of Keap1 and a Nrf2 derived 16-mer peptide, AFFAQLQLDEETGEFL,

respectively. A competitive inhibition assay was used to identify a 7-mer peptide, derived from Nrf2, with an  $IC_{50}$  of  $0.12 \pm 0.01 \mu\text{M}$ .

Protein complementation assay (PCA) relies on the interaction-dependent reconstitution of a reporter protein from two fragments. Some representative split reporter proteins used for PCA experiments include ubiquitin,  $\beta$ -lactamase, Cre-recombinase, luciferase, and GFP (63, 169-171). In one example of using a PCA assay for inhibitor discovery, Hashimoto *et al.* (172) used split mKubasira-GFP (mKG) to screen a library of natural products for inhibitors of the interaction between TCF7- $\beta$ -catenin and 20S proteasome related assembly factors PAC1-PAC2 and PAC3. This approach led to discovery of a compound named TB1 with an  $IC_{50}$  of  $0.020 \mu\text{M}$  for PAC3, but no measurable inhibition for TCF7- $\beta$ -catenin and PAC1-PAC2. Yet another useful PCA involves the use of a split Cre-recombinase (170). Cre-recombinase is a DNA editing enzyme that catalyzes site-specific excision of a gene fragment between a pair of lox DNA sequences. In one representative example, *E. coli* cells were co-transformed with a reporter plasmid that contained lox site-RFP-lox site-GFP and split Cre-recombinase fusions (170). Accordingly, cells emit red fluorescence if Cre is not reconstituted, and green fluorescence if Cre is reconstituted (due to excision of the RFP). As a proof-of-concept, the authors attempted to restore the binding of the double mutant FozLZ to its partner JunLZ through reversion mutagenesis. Of the 51 bright green colonies sequenced, 29% had reverted back to wild type residues (Leu/Leu). Next highly isolated variants were Ala/Leu (16%) and Leu/Arg (14%). These results

suggest that split Cre system is an effective method for detection of PPIs in bacterial colonies.

Our lab has introduced a new class of FP-based tools called dimerization-dependent FPs (ddFPs). A ddFP consists of a quenched FP monomer (A copy) and a non-fluorescent FP monomer (B copy) that form a highly fluorescent complex upon dimerization. Both green and red ddFP variants, which are designated as GA-GB and RA-RB, respectively, have been developed and demonstrated to have utility for monitoring protein dynamics in living cells (97, 98). In addition, Ding *et al.* (99) utilized the interchangeable nature of ddFP-B to develop the Fluorescent Protein eXchange (FPX) ratiometric biosensing strategy. This new strategy offers the ease and versatility in designing sensors and monitoring dynamic PPI that are often impractical for FRET- and PCA-based techniques.

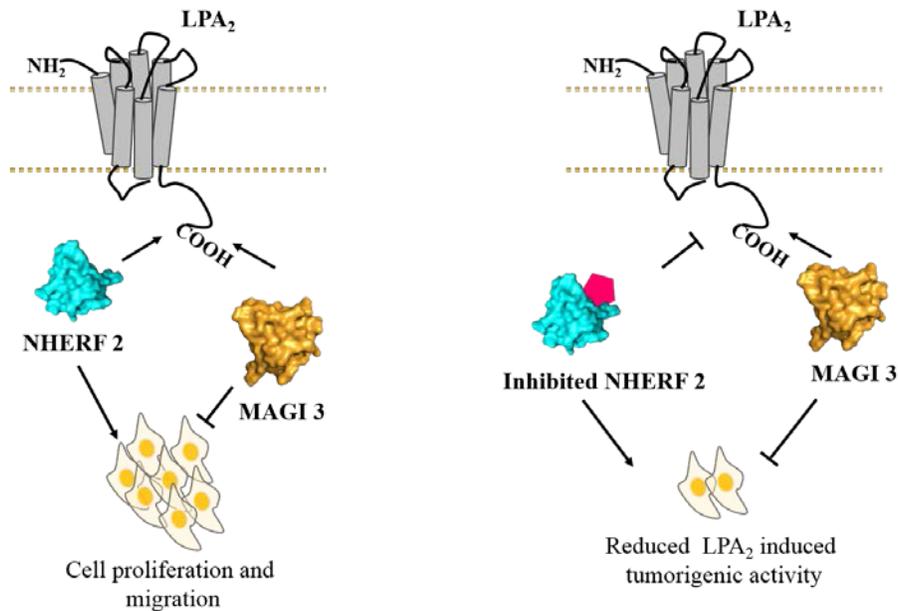
This Chapter describes our efforts to develop a screening method that detects PPIs in *E. coli* colonies. Our goal was to apply the FPX strategy and screen PPI in bacterial colonies by measuring the ratio of green to red fluorescence emission intensities directly on a plate. This approach differs from previously reported assays in that three protein fusions are expressed simultaneously: the two interacting proteins and peptide or third protein that competes for binding to one of the interacting proteins. We expected that this approach could decrease the number of false positives commonly encountered in cell-based FRET and PCA assays (70, 173). We validated our method using some well-characterized PPIs. Such PPIs include the helical coiled-coil E1 and its

partner K1; tumor suppressor p53 and HDM 2; and HRas and the Ras binding domain (RBD) of Raf-1. To demonstrate our strategy's ability to screen gene library of affinity-engineered proteins, we also attempted to rescue the interaction of the binding-deficient Raf RBD mutant R89L to HRas using two methods: reversion mutation in Raf<sub>R89L</sub> RBD and screening of HRas extragenic suppressors of the R89L Raf RBD mutant.

Furthermore using our colony-based screening strategy, we strived to identify peptides that selectively bind to the PDZ domains of the NHERF-2 and MAGI-3 proteins. PDZ domains, named after the three proteins in which they were first found (i.e., PSD95, Dlg1, and ZO-1), constitute one of the largest structurally homologous recognition modules involved in protein transport and trafficking, signal transduction, and complex assembly (*174, 175*). These domains generally recognize short C-terminal sequences of their partner proteins. Due to the conserved structures, PDZ domains often cross-interact with multiple proteins with PDZ binding motifs—thus making them difficult to target with drugs selectively (*176-178*). For this work, we focused on the PDZ domains of the downstream signaling proteins of lysophosphatidic acid receptor-2 (LPA<sub>2</sub>). LPA<sub>2</sub> belongs to the GPCR superfamily and is overexpressed in colon cancer and several types of cancers (*179*) (**Figure 3.3**). In an animal study, researchers discovered that LPA<sub>2</sub> knockout mice showed minimal mucosal damage and tumors, compared to wild type, in a colitis-induced colon tumorigenesis (*180*). LPA<sub>2</sub> contains a PDZ binding motif MDSTL-OH that interacts with several PDZ scaffold proteins, including membrane-associated guanylate kinase with inverted

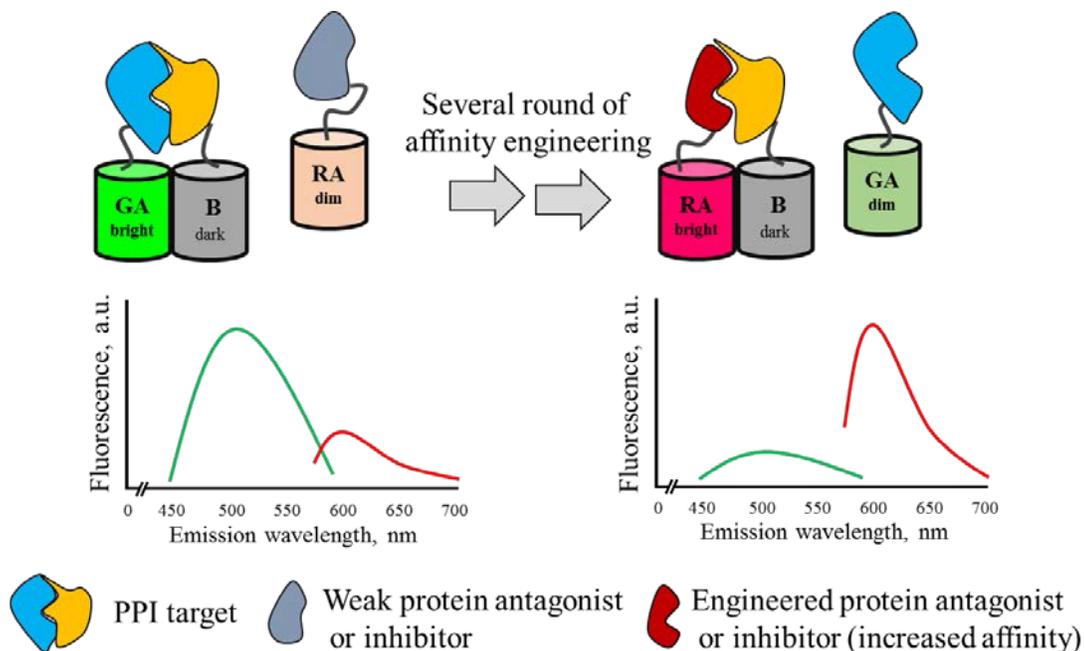
orientation-3 (MAGI-3) and the Na<sup>+</sup>/H<sup>+</sup> exchanger regulatory factor 2 (NHERF-2) (180). Previous reports revealed that NHERF-2, through its second PDZ domain, interacts with LPA<sub>2</sub> and thereby potentiates PLC activity, and enhances cancer cell growth and invasion (181-184). In contrast, the fifth and sixth PDZ domains of MAGI-3 bind with LPA<sub>2</sub>, regulating the receptor's interaction with NHERF-2 (180) and inhibiting cell growth and invasion. Recently, a report by Zheng *et al.* (185) described the use of computational tools to design a 6-mer peptide that selectively binds to NHERF-2 PDZ 2 (N2P2) but not to PDZ 6 of MAGI-3 (M3P6) and vice versa. They identified two inhibitors with more than 100-fold selectivity towards N2P2 but no good inhibitors of M3P6. Our goal was to use the FPX biosensing strategy for colony-based screening of peptide libraries in order to identify inhibitors with even higher selectivity for N2P2 as well as effective inhibitors of M3P6. Shown in **Figure 3.4** is the general scheme for evolving an initially lower affinity second prey to be a more potent inhibitor. Here, the first prey and bait proteins are fused to GA and RB, respectively, and the second prey to RA. Upon expression, a dominant green and weaker red fluorescence is observed (i.e., higher emission ratio). Improvements in the second prey's inhibition potency by rational design, site-directed mutagenesis or directed evolution can be monitored by visual screening, or screening by digital imaging, for colonies with lower emission ratios. This strategy offers a convenient way to generate small-to-medium peptide or protein libraries and screen for improved variants without the time consuming complications of purifying substantial quantities of proteins for affinity characterization. Furthermore, this approach

complements experimental studies on mutational analysis of PPI. Positive or negative effects of mutations introduced can be rapidly assessed through changes in emission ratios relative to a control.



**Figure 3.3** A simplified model for LPA<sub>2</sub>-induced colon cancer pathway.

The *left panel* illustrates the competitive interaction of NHERF 2 and MAGI 3 proteins to the C-terminal residues of LPA<sub>2</sub>. Binding of NHERF 2 activates PLC activity which thereby promotes cell migration and invasion. Meanwhile, MAGI 3 counteracts the aberrant interaction of NHERF 2 by binding to LPA<sub>2</sub> and halting tumorigenesis. The *right panel* shows a model of an approach to regulating LPA<sub>2</sub>-induced tumorigenesis, that is, to identify effective and selective inhibitors to its downstream signaling protein NHERF 2.



**Figure 3.4 General strategy for evolving binding-deficient protein variants or protein-based inhibitors using FPX strategy.** Following the ‘B copy swap’ approach, interacting protein partners can be fused to GA and B copies and a low affinity protein antagonist or inhibitor in RA, generating a high green-to-red fluorescence ratio in colonies. Identification of affinity-engineered antagonist or inhibitor can be screened through a change in fluorescence ratio, that is, from high green-to-red to low green-to-red.

## 3.2 Results and discussion

### 3.2.1 FPX-based PPI screening design and strategy

Our goal was to develop a FPX-based screening strategy (refer to **Section 1.3.6**) that could be used to detect protein-protein or protein-peptide interactions in bacterial colonies. We envisioned that the association of protein partners, and

their disruption by inhibitors, affinity-engineered protein partners, or other interacting protein, would be measured through the intensity ratios of green and red fluorescence in *E. coli* colonies. In order to achieve this goal, two important criteria had to be satisfied. First, a B copy with comparable dissociation constants for both GA and RA was required. Second, a generalizable polycistronic *E. coli* vector that expresses three interacting proteins as ddFP fusions at similar levels had to be constructed.

To satisfy the first criteria, we examined the binding of various B copy proteins to GA and RA. Specifically, we tested the B copy optimized to complement RA (RB), the B copy optimized to complement GA (GB), and the B copy optimized to complement YA (YB). From the saturation binding experiments illustrated in **Figure 3.5**, dissociation constants ( $K_d$ ) of both GA and RA against RB were the highest among the three B copies tested and could be fit using the one-site binding equation. Titrations of 0.5  $\mu\text{M}$  GA and 0.5  $\mu\text{M}$  RA with RB exhibited dissociation constants of 42  $\mu\text{M}$  and 39  $\mu\text{M}$ , respectively (**Figure 3.5 B**). The latter was consistent with the reported  $K_d$  for RA-RB (33  $\mu\text{M}$ ), while the former showed a slightly weaker binding. Although this is the first report of the  $K_d$  for GA-RB, we had anticipated that it would fall around a similar value as RA-RB. During the evolution of RA to GA, the key R153E mutation was retained. This mutation at the interface created a glutamate-glutamate repulsive interaction with RB, which also has the R153E mutation.

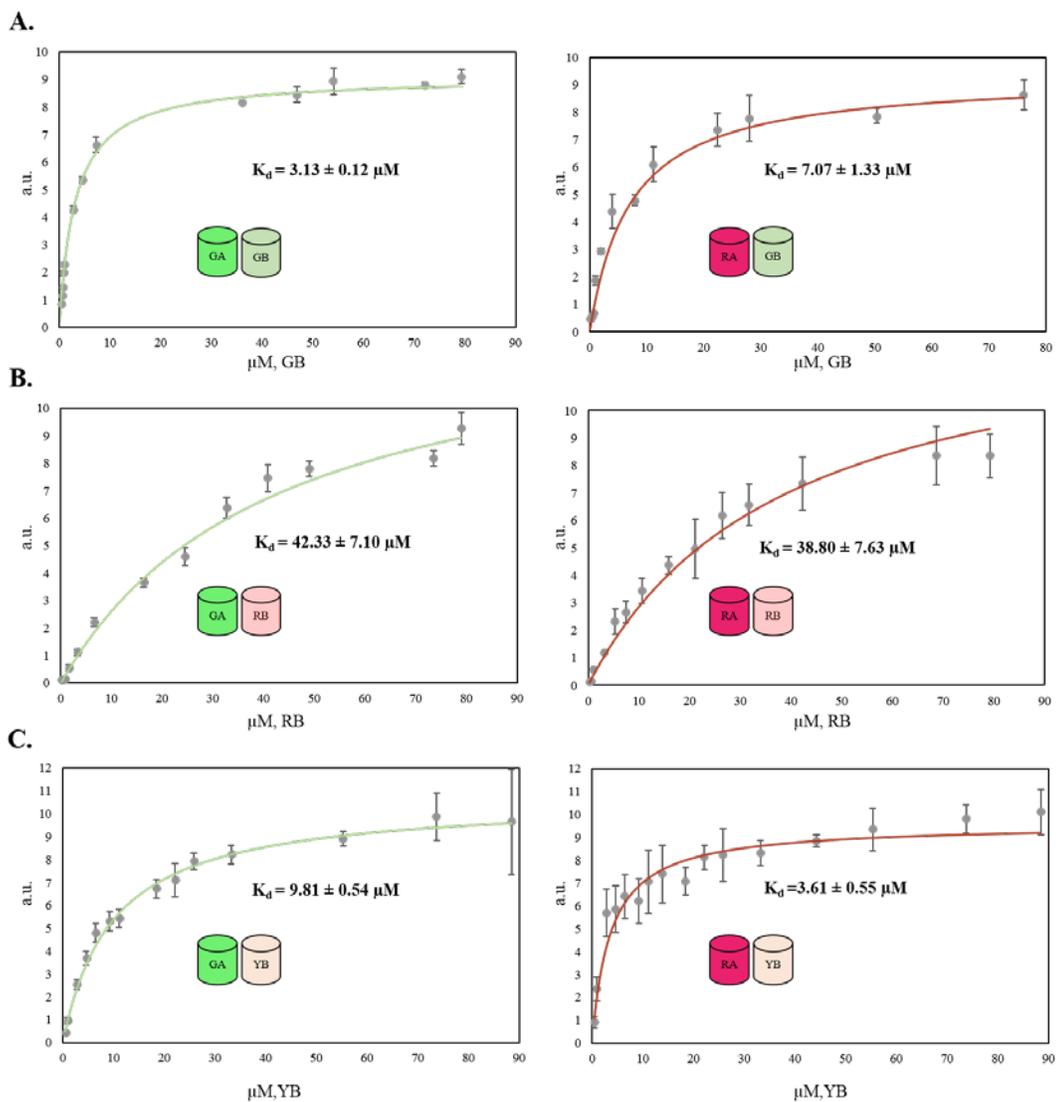
The other two B copies (GB and YB) exhibited lower  $K_d$ s for the various A copy variants (**Figure 3.5 A and C**). Due to the lack of a ddFP crystal

structure, it is unclear on how additional mutations (relative to RB) found in GB and YB contribute to their increased their affinities toward all A copies. However, Spencer *et al.* (97, 98) speculated that the E153K mutation in both GB and YB is involved in favourable electrostatic attractions at the dimer interface.

Overall, these results prompted us to use RB for this colony-based PPI screening. The higher  $K_d$  of the RB variant was expected to substantially reduce the amount of association that occurred independent of interactions between the fused proteins of interest (i.e., the background).

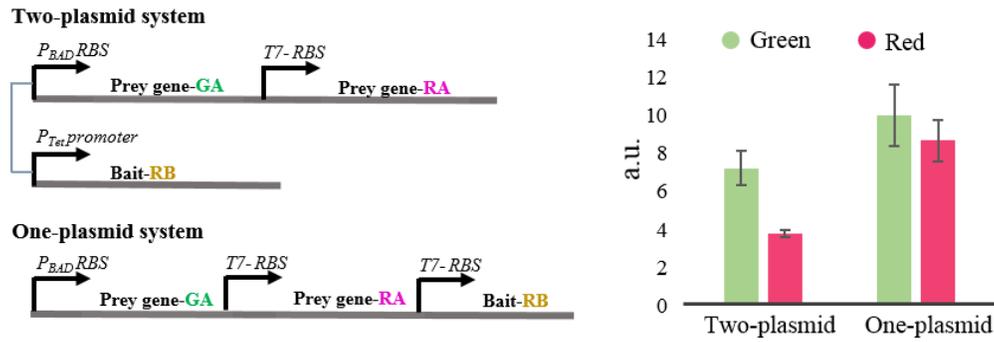
The FPX-based PPI screening necessitates the simultaneous expression of three gene fusions. We used the L-arabinose inducible bicistronic and tricistronic bacterial vectors constructed as described in Chapter 2. All expression cassettes of these vectors are under a tightly regulated  $P_{BAD}$  promoter. For the bicistronic vector, the first expression cassette contains the ribosome binding site (RBS) sequence of pBAD and second cassette has the T7-RBS sequence. For the tricistronic vector, the first expression cassette contains the pBAD RBS sequence, while the second and third cassettes have the T7-RBS. We had originally attempted to pursue this screening strategy using a two-plasmid expression system where GA and RA gene fusions are expressed from the bicistronic vector and the RB fusion from a tetracycline-inducible low expression plasmid that has a compatible replicon p15A, and a chloramphenicol antibiotic marker (**Figure 3.6 A**). However with this two-plasmid system, colonies were small and the fluorescence was quite dim even after two days. This observation could be attributed to the very low expression of RB in *E. coli* colonies and/or toxicity due

to the multiple additives to the growth medium (i.e., both inducers and antibiotics). Fortunately, with the tricistronic vector, colony sizes were uniform across the plate, and substantial fluorescence was observed. This tricistronic vector was further modified to conveniently accommodate any protein partners without complex cloning and gene assembly. As described in **Figure 3.6 B**, this new vector, designated pFPX, was designed to pre-encode the genes of GA, RA and RB on the 1<sup>st</sup>, 2<sup>nd</sup> and 3<sup>rd</sup> cassettes, respectively. It is also designed to link all PPI of interest as N-terminal fusions to ddFPs. The pFPX could one gene of interest (e.g., first prey) on the 1<sup>st</sup> cassette between *XhoI* and *EagI* or *BglII* restriction sites and another on the 2<sup>nd</sup> cassette (e.g., second prey) with *EcoRI* and *Sall* restriction sites before RA. Third cassette was constructed to accommodate a gene between the *XbaI* and *KpnI* sites, to produce a RB gene fusion (e.g., the bait).

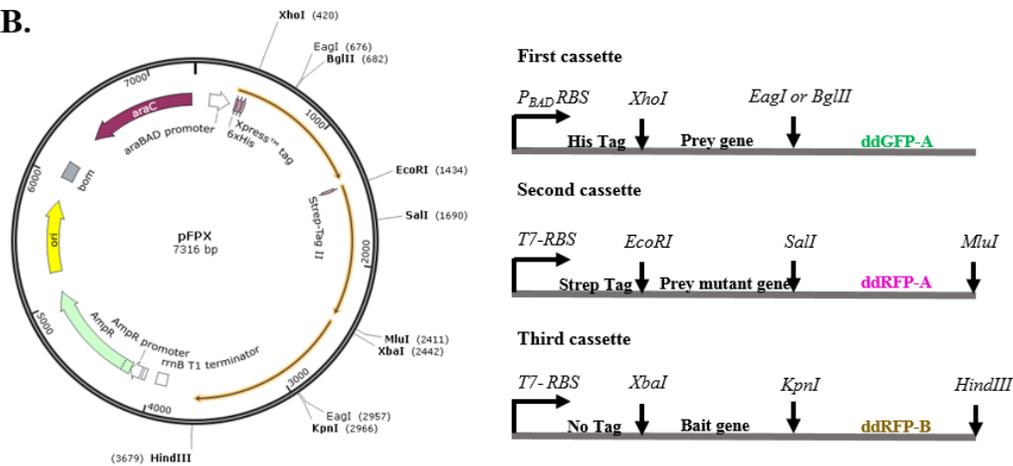


**Figure 3.5 Saturation curves for RA and GB against three ddFP-B copies namely GB, RB and YB.** *A.* GA-GB and RA-GB binding, *B.* GA-RB and RA-RB binding and *C.* GA-YB and RA-YB binding. Steady-state binding was monitored using fluorescence. Error bars are  $\pm$  standard deviation for three independent experiments.

**A.**



**B.**



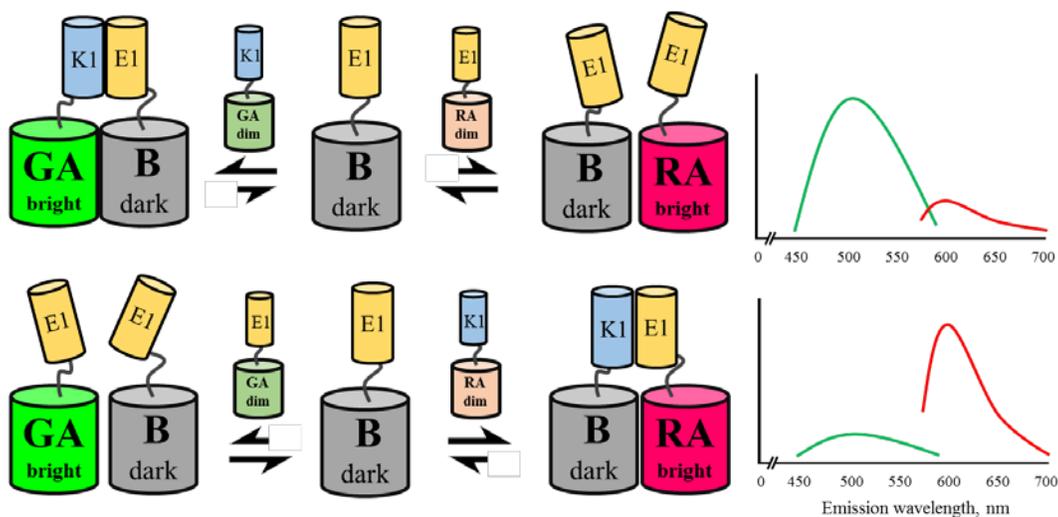
**Figure 3.6 In colony PPI screening plasmid design (pFPX).** *A. (left panel)* Designs of two and one-plasmid systems for FPX and *(right panel)* their relative ddFP green and red fluorescent intensities directly measured on colonies. Error bars are  $\pm$  standard deviation for more than 100 colonies. *B. (left panel)* A Snapgene ([http://www.snapgene.com/products/snapgene\\_viewer/](http://www.snapgene.com/products/snapgene_viewer/)) generated circular plasmid map showing the general vector properties of pFPX. *Right panel* shows the specific RE sites and tags included in each expression cassettes.

This FPX-based strategy relies on the competitive binding of two proteins (first and second prey) to their partner (bait). The first protein is the known interacting partner of the bait (first prey) and the second (second prey) is a biologic antagonist or any protein suspected to competitively interact with the bait. With the pFPX vector, we can express these proteins in *E. coli* as ddFP fusions, which then permits us to thus monitor the relative strength of the interaction of the two prey proteins with the bait, by imaging the ratio of green and red fluorescence emission on colonies. We define the emission ratio as the green intensity divided by the red intensity.

### 3.2.2 FPX proof-of-concept

To validate this screening method, we sought to first apply it to some well-characterized PPIs. These interactions include: the E1 and K1 helical coiled-coil heterodimer designed by Tripet *et al.* (186); the transactivation domain (1-30 aa) of p53 and its antagonist HDM 2 (1-113 aa) (19); and HRas<sub>WT</sub> and the Ras binding domain (RDB) of Raf-1<sub>WT</sub> (187).

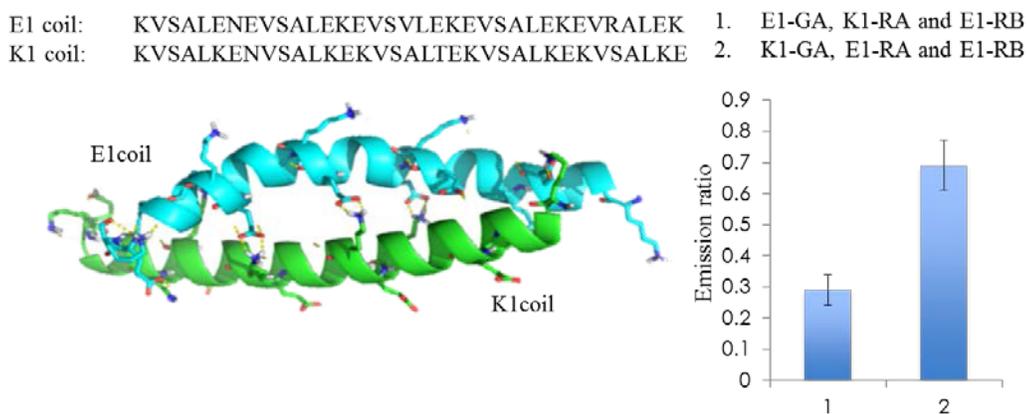
*E1 and K1 coiled-coil* (**Figure 3.7**). Helical coiled-coils are protein motifs that primarily serve as oligomerization subunits (188). Although structurally simple, proteins with these motifs often display a wide range of cellular functions depending on their coiled-coil architectures. One of the most notable coiled-coil pairs is c-Fos and c-Jun of activator protein 1 (AP1) family. Upon dimerization, this pair functions as early response transcription factors important in regulating cell cycle and proliferation (189-192).



**Figure 3.7** Graphical representation of FPX based E1 and K1 interaction.

To demonstrate the FPX strategy, we utilized a *de novo* designed coiled-coil heterodimer known as E1 and K1 (186). This helical coiled-coil pair has increased solubility relative to c-Fos and c-Jun, and does not homodimerize. E1 and K1 have previously been used to optimize PCA-based biosensors such as split-GFP (88). Shown in **Figure 3.8** are the structures of E1 and K1 as predicted by Phyre (193) and docked using ClusPro (194-197). This simulated interaction reveals that E1 and K1 dimerize through a favourable electrostatic interaction arising from a series of hydrogen bonds from glutamate and lysine side chains of E1 and K1 respectively. When expressed in *E. coli* as E1-GA, K1-RA and E1-RB protein fusions, we observed a dimmer green and brighter red fluorescence in colonies, with an emission ratio of  $0.29 \pm 0.05$  (**Figure 3.8**). When E1 and K1 were swapped and E1-RA, K1-GA instead, and E1-RB were expressed together, a 2.3-fold increase in ratio ( $0.67 \pm 0.09$ ) was observed. This result was consistent

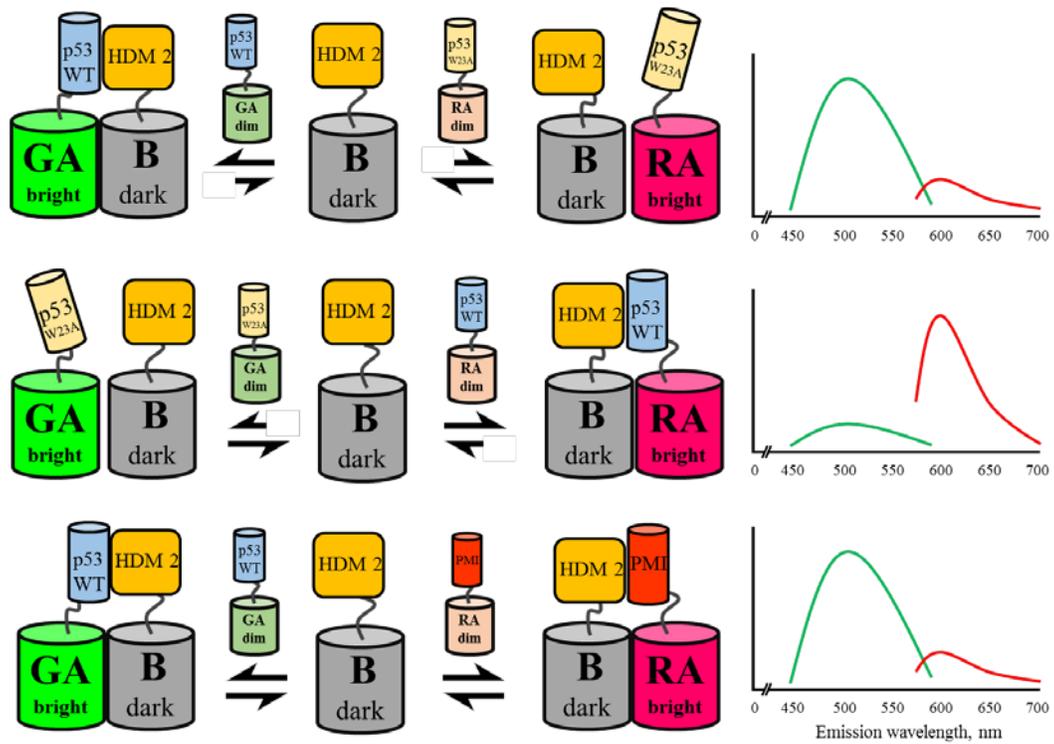
with our expectation that E1-RB should favourably bind to complementary fusions with K1 (i.e., K1-GA or K1-RA), and form highly fluorescent complexes of the corresponding colour.



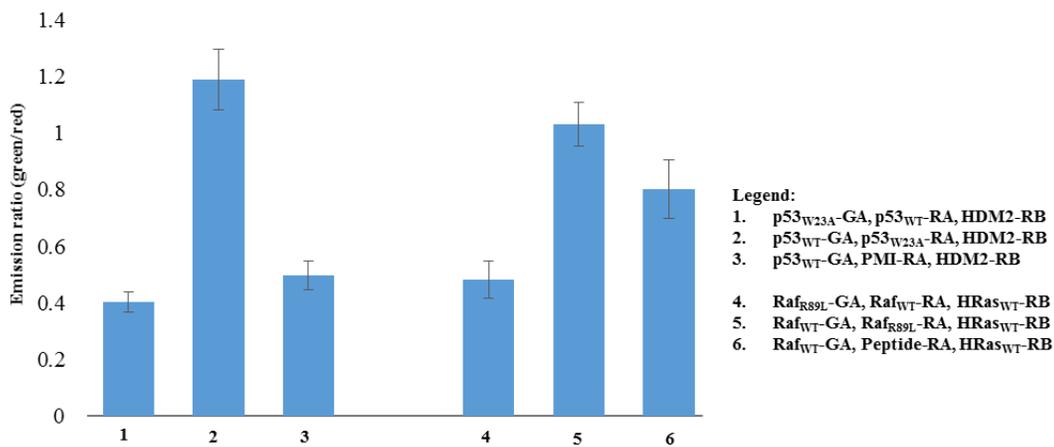
**Figure 3.8 Docking and FPX screening results for helical coiled-coil E1 and K1.** *Left panel* shows the ClusPro predicted binding interaction between E1 (cyan) and K1 (green) coiled-coils. This model also displays a network of hydrogen bonding formed by glutamate-lysine interactions as visualized by PyMol (198). Individual helical coil structure used for docking was predicted using Phyre2 server. The *right panel* displays FPX colony screening results for E1 and K1 interaction. Shown here are the relative emission ratios of *E. coli* colonies co-expressing E1 and K1 as ddFP fusions. Error bars are  $\pm$  standard deviation for more than 100 colonies.

*p53 and HDM 2* (**Figure 3.9**). We constructed three pFPX plasmids to evaluate the interaction of tumor suppressor p53 and HDM 2 along with p53's

lower affinity mutant p53<sub>W23A</sub> and the HDM 2 inhibitor NH<sub>2</sub>-TSFAEYALLSPGG-OH (PMI) discovered by phage display (199, 200). In all constructs, HDM 2 was fused with RB while p53 and its mutant were swapped into either the GA or RA cassettes. After a 2-day growth on agar plates containing L-arabinose (overnight at 37 °C and another day at 4 °C), colonies exhibited distinguishable differences in green and red emissions. Colonies expressing p53-GA, p53<sub>W23A</sub>-RA and HDM 2-RB had an emission ratio of  $1.19 \pm 0.11$ . The ratio was  $0.40 \pm 0.04$  for colonies expressing p53<sub>W23A</sub>-GA, p53-RA and HDM 2-RB (**Figure 3.10**). This result is consistent with previous findings that mutation of Trp23 severely impairs p53's ability to bind to its partner HDM 2. As anticipated, colonies transformed with p53-GA, PMI-RA and HDM 2-RB exhibited a ratio of  $0.49 \pm 0.05$  (**Figure 3.10**). This observation suggested that the PMI peptide inhibitor was effectively competing with p53 for binding to HDM 2.



**Figure 3.9** Graphical representation of FPX based screening of p53<sup>WT</sup> and HDM 2 interactions along with binding-deficient p53<sup>W23A</sup> and HDM 2 peptide inhibitor (PMI).

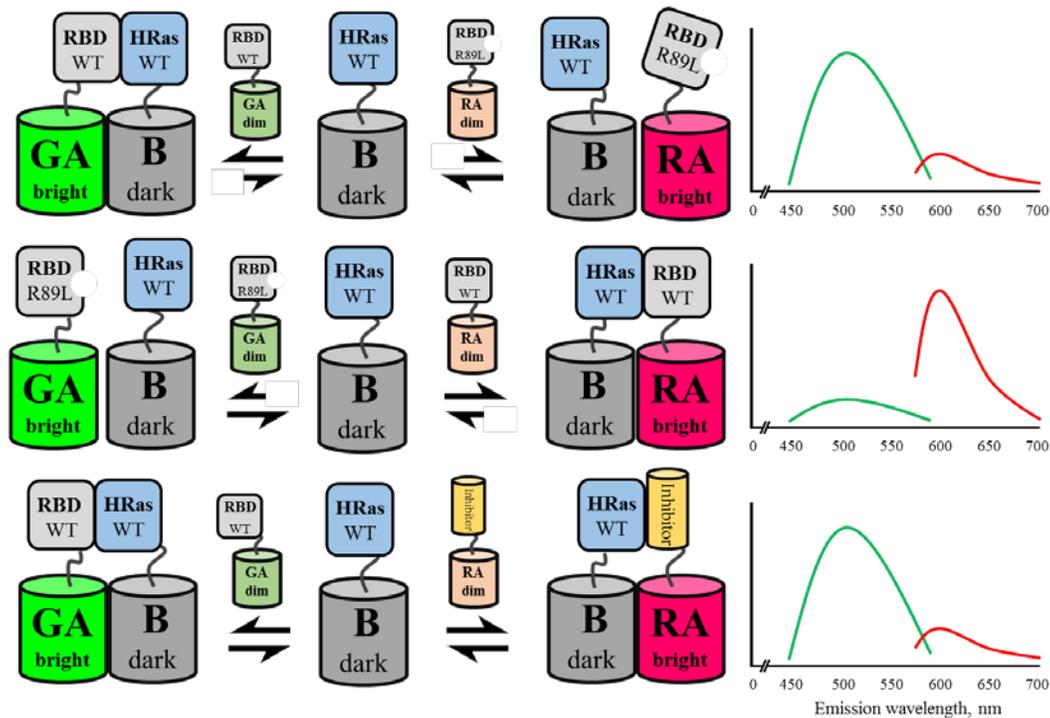


**Figure 3.10** Emission ratios acquired from FPX colony screening of p53-HDM 2 and Raf-1 RBD-HRas<sup>WT</sup> association. The first three ratios represent the

relative binding of HDM 2 towards p53<sub>WT</sub>, its non-binding variant p53<sub>W23A</sub>, and an HDM 2-peptide inhibitor (PMI). The fourth to sixth emission ratios reveal the favored association of HRas<sub>WT</sub> and Raf<sub>WT</sub> RBD relative to the binding deficient Raf<sub>R89L</sub> RBD variant and its peptide inhibitor. Results shown here are mean  $\pm$  standard deviation from more than 100 colonies.

*HRas<sub>WT</sub> and the RDB of Raf-1<sub>WT</sub>* (**Figure 3.11**). As a third validation, we generated a set of pFPX plasmids for evaluating the interaction of HRas<sub>WT</sub> and Raf<sub>WT</sub> RBD. To compete with Raf<sub>WT</sub> RBD binding to HRas<sub>WT</sub> we turned to the binding deficient Raf-1 RBD mutant, R89L, and the decapeptide inhibitor ECCAVFRLH (**Figure 3.10**) (201, 202). The Raf-1 RBD R89L mutation abolishes its nanomolar affinity of RBD ( $K_d = 130$  nm, 203) towards HRas<sub>WT</sub> by destroying the hydrogen bonds formed at the interface. Meanwhile, the decapeptide inhibitor derived from Raf-1 RBD has been reported to inhibit Ras-Raf association, with an IC<sub>50</sub> of 7  $\mu$ M. However, no further cell or *in vivo* studies were conducted to gauge the efficacy of this peptide. Co-expression of Raf<sub>WT</sub> RBD-GA, Raf<sub>R89L</sub> RBD-RA and HRas<sub>WT</sub>-RB gave an in-colony emission ratio of  $1.03 \pm 0.08$ . The emission ratio of colonies expressing Raf<sub>R89L</sub> RBD-GA, Raf<sub>WT</sub> RBD-RA and HRas<sub>WT</sub>-RB was approximately two-fold lower ( $0.48 \pm 0.07$ ). Given the high affinity of the interaction between HRas<sub>WT</sub> and Raf<sub>WT</sub> RBD, we had expected to observe a larger change in change in ratio when swapping the A copies. One possible reason for small change in ratio is the deficiency of guanosine-5'-triphosphate (GTP) in the *E. coli* cytoplasm. GTP is an important

co-factor of HRas<sub>WT</sub>, and the absence of it dramatically decreases the protein's affinity towards its effector Raf-1 RBD. *E. coli* transformed with pFPX encoding Raf<sub>WT</sub> RBD-GA, ECCAVFRLH-RA and HRas<sub>WT</sub>-RB gave an intermediate ratio of  $0.80 \pm 0.10$ , consistent with its modest affinity for HRas<sub>WT</sub>. Results from all three interacting proteins provide solid support for the robustness of the FPX method for detecting PPIs in colonies.

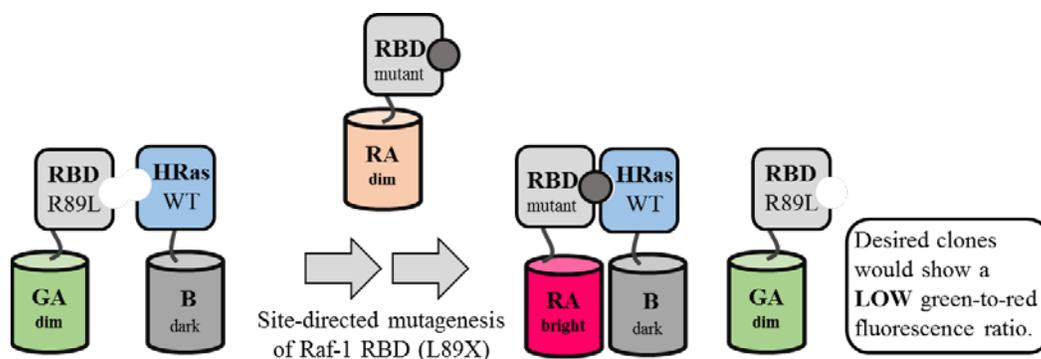


**Figure 3.11** Graphical representation of FPX based screening of Raf-1<sub>WT</sub> RBD and HRas<sub>WT</sub> interactions along with binding-deficient Raf-1<sub>R89L</sub> RBD and HRas peptide inhibitor, ECCAVFRLH.

### 3.2.3 Reversion mutation screening for Raf<sub>R89L</sub> RBD-RA

Given the results of our proof-of-concept experiments, we next attempted to screen genetic libraries for engineering of PPIs in bacterial colonies. We

predicted that we could rescue the binding of Raf<sub>R89L</sub> RBD-RA to HRas-RB by randomly mutating the 89<sup>th</sup> position using an NNK randomized codon (where N = A + G + C + T, and K = G + T). NNK is the smallest nucleotide combination (32 codons) that yields all 20 amino acids. Accordingly, we assembled a pFPX plasmid that encoded Raf<sub>R89L</sub> RBD-GA, HRas<sub>WT</sub>-RB and randomized Raf<sub>L89X</sub> RBD-RA, and screened approximately 5000 transformed *E. coli* colonies (**Figure 3.12**). We picked 20 colonies with low green/red emission ratios, and sequenced the plasmid DNA to identify the amino acid at the 89<sup>th</sup> position. **Table 3.1** summarizes all the mutations at the 89<sup>th</sup> position of Raf<sub>89X</sub> RBD-RA and their number of occurrence. Fifteen percent (3/20) of the colonies were shown to have reverted back to wild type Raf-1 RBD (89R). The remaining 85% were mutants capable of rescuing hydrogen bond networks at the interface. Tyr, Asn and His each occurred in 20% of the clones (4/20), 15% (3/20) were Ser, and the remaining 10% (2/20) were a Cys and a Phe mutant. To acquire accurate emission ratio measurements on all variants, we transformed *E. coli* with each of these pFPX plasmids. His and Asn mutants gave a similar ratio to the wild type, while ratios of Cys, Ser, and Tyr were higher. The Phe mutant gave a ratio that was essentially identical to the negative control (R89L), indicating that this mutation did not rescue Raf-1 RBD's binding to HRas<sub>WT</sub>. This false positive result could be the result of an error in colony picking or a contaminated culture. An additional source of error is the uneven distribution of excitation light on the plate that makes some colonies, particularly those on the side of the plate, dim and thus skews their emission ratio measurements.



**Figure 3.12** A simplified representation of the FPX screening assay of reverted mutants of Raf-1 RBD L89X. A gene library of Raf-1<sub>L89X</sub> RBD is co-expressed with Raf-1<sub>R89L</sub> RBD-GA and HRas<sub>WT</sub>-B in DH10B. Clones in which RBD mutation rescues it binding to HRas<sub>WT</sub>-B would show an evident bright red fluorescence or low green-to-red fluorescence ratio in *E. coli* colonies.

**Table 3.1** Clones isolated from Raf<sub>R89X</sub> RBD reversion mutagenesis.

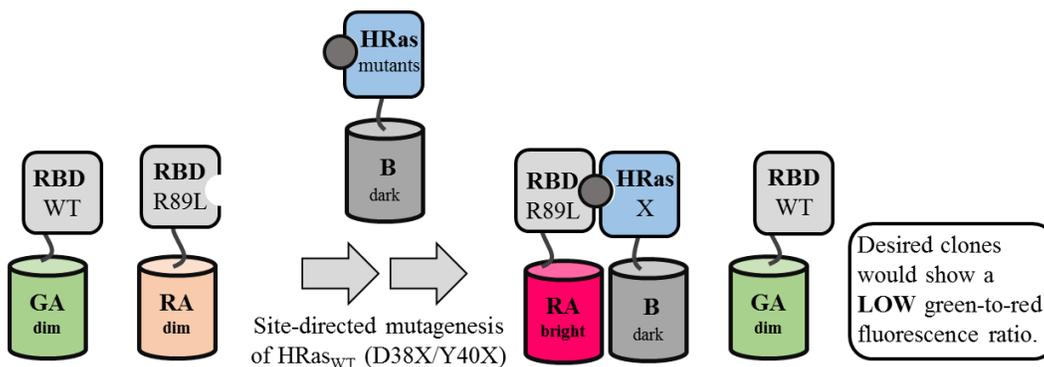
| Raf-1 clones     | Leu89X substitution <sup>a</sup> | Number of occurrences <sup>b</sup> | Emission ratios |
|------------------|----------------------------------|------------------------------------|-----------------|
| 1                | Asn                              | 4                                  | 0.56 ± 0.03     |
| 2                | His                              | 4                                  | 0.48 ± 0.02     |
| 3                | Tyr                              | 4                                  | 0.75 ± 0.09     |
| 4                | Arg (wild type)                  | 3                                  | 0.52 ± 0.04     |
| 5                | Ser                              | 3                                  | 0.69 ± 0.03     |
| 6                | Cys                              | 1                                  | 0.66 ± 0.09     |
| 7                | Phe                              | 1                                  | 0.94 ± 0.08     |
| Negative control | Leu                              | -                                  | 0.92 ± 0.04     |

<sup>a</sup>Abbreviations correspond to the standard three letter amino acid designations.

<sup>b</sup>A total of 20 colonies were selected.

### 3.2.4 HRas extragenic suppressors for Raf<sub>R89L</sub> RBD

Intragenic suppressors that restore signaling activity of Raf<sub>R89L</sub> RBD were first identified close to two decades ago (204). However, to our best knowledge, no HRas extragenic suppressor that fully rescues the binding and activity of Raf<sub>R89L</sub> RBD have yet been reported. Our goal was to generate and screen mutant HRas variants that could potentially form an orthogonal interaction with Raf<sub>R89L</sub> RBD (Figure 3.13). Accordingly, we created a library of HRas variants with two fully randomized positions at the interaction interface. The 38<sup>th</sup> and 40<sup>th</sup> amino acid residues were selected for full randomization since their side chains are within the hydrogen bond distance, 3-4 Å, of Arg89 of wild type Raf-1 RBD. Furthermore, Asp38 of HRas forms a direct hydrogen bond with the guanidinium moiety of Arg89 in the heterodimer (PDB 4G0N) (205, 206).



**Figure 3.13** A general scheme to screen a library of HRas extragenic suppressors using FPX. A pFPX carrying Raf-1<sub>WT</sub> RBD-GA, Raf-1<sub>R89L</sub> RBD-RA and a library of HRas (D38X/Y40X) is transformed in DH10B. Clones which

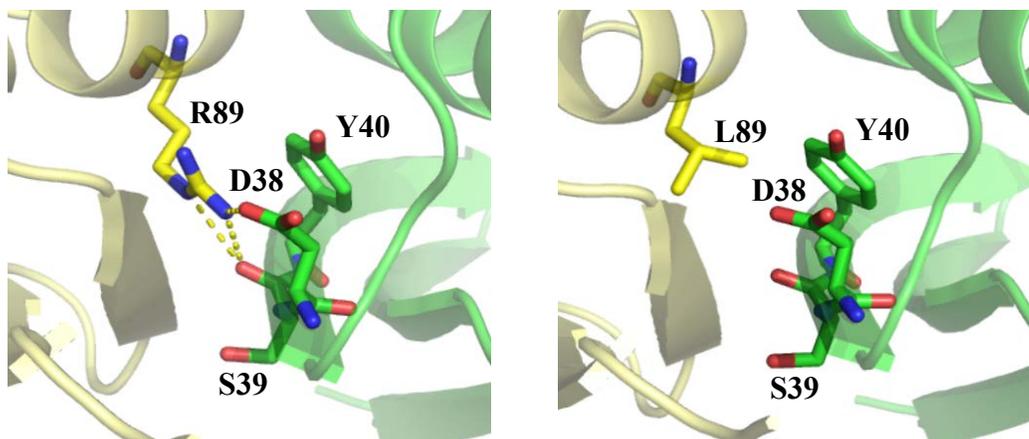
apparently rescued HRas' binding towards Raf-1<sub>R89L</sub> RBD-RA would give an obvious low green-to-red ratio.

**Table 3.2 Substitution identified in HRas extragenic suppressors.**

| HRas suppressor<br>clones | Substitution in suppressors <sup>a</sup> |       |
|---------------------------|--|-------|
|                           | Asp38                                    | Tyr40 |
| 1                         | Ala                                      | Val   |
| 2                         | Pro                                      | Glu   |
| 3                         | Thr                                      | Val   |
| 4                         | Glu                                      | Ile   |
| 5                         | Thr                                      | Asp   |
| 6                         | Ala                                      | Pro   |
| 7                         | Ser                                      | Pro   |
| 8                         | Asp                                      | Lys   |
| 9                         | Tyr                                      | Pro   |
| 10                        | Ser                                      | Gln   |
| 11                        | Val                                      | Cys   |
| 12                        | Pro                                      | Met   |
| 13                        | Gly                                      | Phe   |
| 14                        | Phe                                      | Leu   |
| 15                        | Thr                                      | Arg   |

<sup>a</sup>Abbreviations correspond to the standard three letter amino acid designations.

We assembled a pFPX plasmid that encoded Raf<sub>R89L</sub> RBD-RA, HRas<sub>S38X</sub>,<sub>40X</sub>-RB and Raf<sub>WT</sub> RBD-GA, transformed *E. coli*, and screened approximately 10,000 colonies. In this arrangement, a high green/red ratio would indicate that the HRas mutant was binding to Raf<sub>WT</sub> RBD, while a low ratio would indicate the HRas mutant was binding to Raf<sub>R89L</sub> RBD. This screen led to the identification of 15 promising variants with low ratios. **Table 3.2** outlines all HRas mutations that appeared to have rescued binding to Raf<sub>R89L</sub> RBD based on ratiometric measurements. Sequence wise, we did not see any trend in the amino acid identifies. Due to this lack of consensus, we performed a second round of screening to verify the emission ratios with side-by-side comparison to two controls. As a positive control we used pFPX encoding Raf<sub>R89L</sub> RBD-GA, Raf<sub>WT</sub> RBD-RA and HRas<sub>WT</sub>-RB. As a negative control we used pFPX expressing Raf<sub>WT</sub> RBD-GA, Raf<sub>R89L</sub> RBD-RA and HRas<sub>WT</sub>-RB. This analysis identified clones 7 ( $0.52 \pm 0.07$ ), 12 ( $0.33 \pm 0.02$ ) and 14 ( $0.50 \pm 0.11$ ) which showed emission ratios that were the same as or lower than the positive control ( $0.52 \pm 0.04$ ). With the exception of a Ser at position 38 of clone 7, these sequences are dominated by nonpolar residues. We speculate that these mutations have rescued binding through nonpolar interactions with Leu89 of Raf<sub>R89L</sub> RBD (**Figure 3.14**). The other 12 clones all gave emission ratios higher than the positive control, and appear to be false positives due to personal or systematic errors discussed earlier. HRas binding conformation allowing new interactions to occur. Further characterizations of these HRas variants is essential, but is beyond the scope of this thesis.



**Figure 3.14 Binding interaction of Raf-1 RBD (WT or R89L) and HRas<sub>WT</sub> at the interface region.** *Left panel* displays the favorable hydrogen bond interaction between Arg of Raf<sub>WT</sub> RBD (yellow) and integral  $\beta$ -sheet residues of HRas<sub>WT</sub> (green). *Right panel* shows the mutation resulting in loss of hydrogen bonding between Raf-1 RBD and HRas<sub>WT</sub> upon mutation of Arg to Leu residue. This figure was taken from of an X-ray crystal structure of Ras GppNHp and Raf-1 kinase RBD complex (PBD 4G0N, 206) and visualized by PyMol (198).

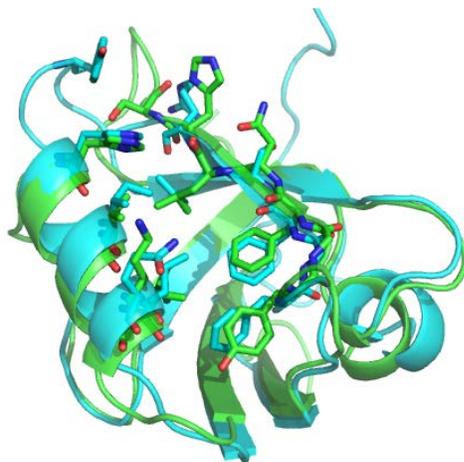
### 3.2.5 Selective inhibition of NHERF-2 PDZ 2 (N2P2) and MAGI-3 PDZ 6 (M3P6) towards lysophosphatidic acid receptor 2 (LPA<sub>2</sub>)

In an attempt to further demonstrate the utility of the FPX screening method, we screened a library of 7-mer peptides to find sequences that could selectively inhibit aberrant interactions of NHERF-2, through PDZ 2 (N2P2), and lysophosphatidic acid receptor-2 (LPA<sub>2</sub>), as commonly observed in colon cancer. Targeting PDZ domains like N2P2 is an ongoing challenge in drug design. Their high structural homology and overlapping recognition sequences make them

difficult to block selectively (**Figure 3.15**). But with the competitive nature of our assay, we hypothesized that this challenge could be overcome. Likewise, we also generated another gene library to identify a set of 7-mer peptides that can effectively inhibit the 6<sup>th</sup> PDZ of MAGI-3, N2P2's competing protein and a recently reported negative regulator of LPA<sub>2</sub> activity (180). Our preliminary design involved a pFPX plasmid that expresses the target protein N2P2 and its antagonist M3P6 as GA and RA fusions, respectively, and a peptide library linked at the C-terminal of RB. Upon transformation of the gene library in DH10B, colonies with higher emission ratios (brightly green) would indicate peptide candidates that have apparently gained selectivity towards N2P2. On the other hand, colonies with lower emission ratios (brightly red) have expressed peptides with increased selectivity towards M3P6. However, using this plasmid design, colony fluorescence was too weak to measure accurately even after 48 hours of incubation (data not shown). We suspected that fusion of the unstructured peptide library to RB had tremendously decreased its solubility in *E. coli*. To overcome this problem, we tagged RB-peptide with the protein SUMO 1 at its N-terminus. We chose SUMO 1, rather than other well-characterized solubility enhancers, due to its small size and no reports of interactions with the proteins of interest. In addition, we decided to replace the RBS of the 1<sup>st</sup> cassette with T7-RBS due to the relatively poor expression of N2P2-GA. Hence, all succeeding library screening was performed using pFPX with T7-RBS upstream of all expression cassettes.

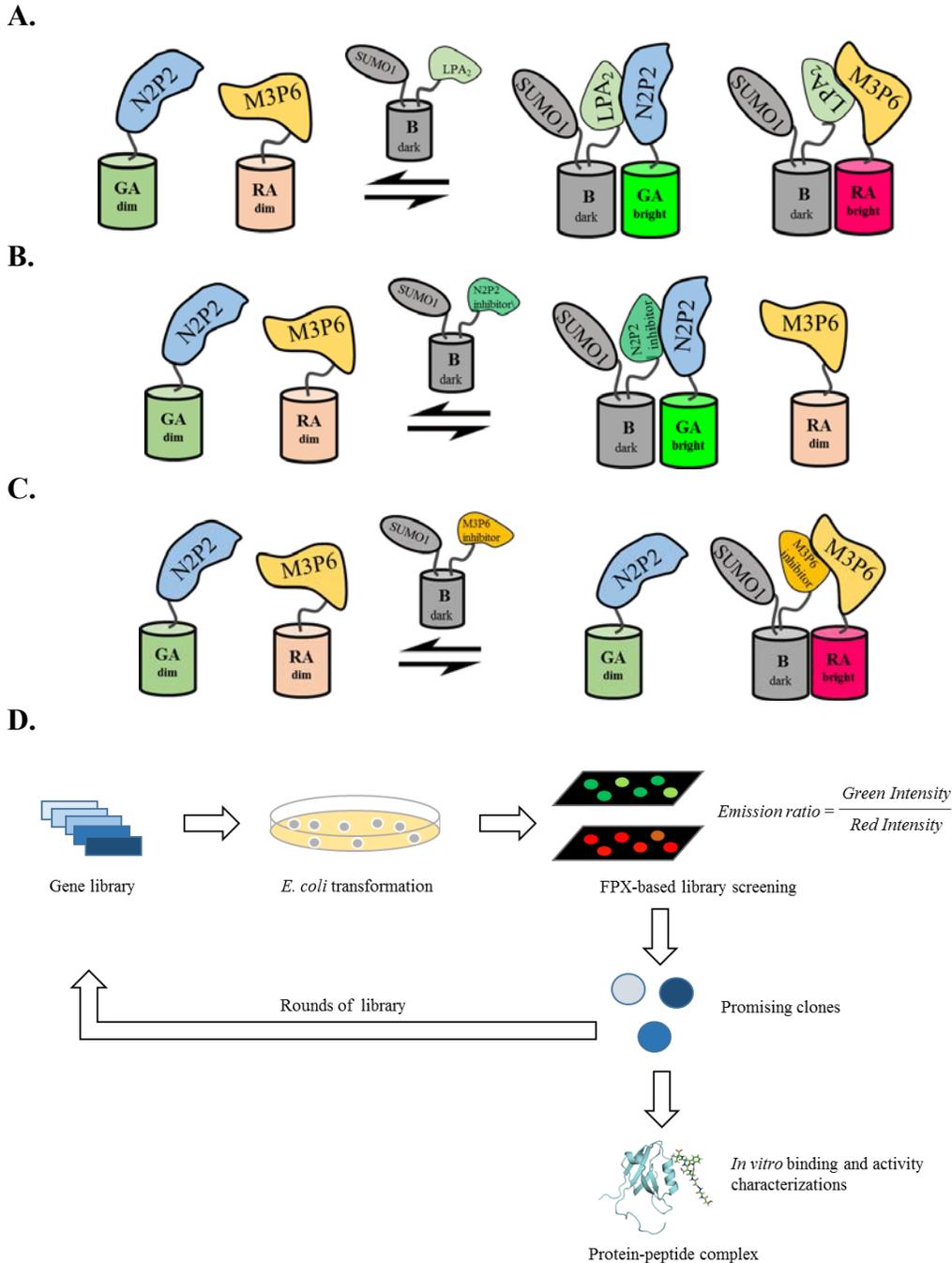
Due to the lack of reported peptides that can selectively bind to either N2P2 or M3P6, we decided to co-express SUMO 1-RB-LPA<sub>2</sub> (311-351 amino

acid residues) together with N2P2-GA and M3P6-RA as our control pFPX. This is neither a positive nor a negative control since the C-terminal of LPA<sub>2</sub> binds to both N2P2 and M3P6. However, if the clone's emission ratio is higher than the emission of the control, it indicates that the lead peptide shows preference to N2P2 relative to the LPA<sub>2</sub> control. If the ratio is lower than control, this means the identified peptide binds selectively towards M3P6. **Figure 3.16** shows the general FPX designs and screening protocol for discovering selective peptide inhibitors for the downstream signaling proteins involved in LPA<sub>2</sub>-induced colon cancer tumorigenesis. The nomenclature for peptide inhibitor design uses standard PDZ motif numbering P<sup>-6</sup>P<sup>-5</sup>P<sup>-4</sup>P<sup>-3</sup>P<sup>-2</sup>P<sup>-1</sup>P<sup>0</sup> (where P<sup>n</sup> denotes to any amino acid residue and its position from the carboxyl end) (185, 207).



**Figure 3.15 Superimposed model of N2P2 (green) and M3P6 (cyan).** This model demonstrates the conserved structural homology of most PDZ domains. Also, important residues on integral binding region (represented in stick model) of most PDZ domains are either identical or highly conserved. Structural alignment

was done by PyMol (198) using the X-ray crystal structure of N2P2 (PDB ID 2HE4, 208) and Bioserf v2.0 modelled structure of M3P6 (209)



**Figure 3.16** General library designs and screening protocol used for discovering selective peptide inhibitors for N2P2 and M3P6. A-C illustrate the

FPX-based approaches in screening and identifying peptide inhibitors for N2P2 and M3P6: *A.* control pFPX with C-terminal residues of LPA<sub>2</sub> fused to B; *B.* pFPX design for screening N2P2 peptide inhibitors; and *C.* pFPX design for screening M3P6 peptide inhibitors. *D.* shows a simplified scheme of the screening protocol. Genes of interacting proteins and inhibitor library are assembled in pFPX vector and transformed in DH10B *E. coli* competent cells. After 24-48 hours of incubation, fluorescence of colonies is imaged using the digital microscope. Plasmids of clones with desired phenotype are then propagated, purified and sequenced. Finally, identified peptides may either proceed to a series *in vitro* characterization or take further rounds of mutagenesis to improve its affinity and selectivity.

### 3.2.5.1 Screening of peptide inhibitors for N2P2-LPA<sub>2</sub> association

We carried out library screening using a pFPX plasmid expressing N2P2-GA, M3P6-RA and SUMO 1-RB-peptide library. We generated the 7-mer peptide library following the template NH<sub>2</sub>-SNTKF-OH which was based on previous information about NHERF PDZ's binding preferences. The choice of Phe at the carboxyl end was based on work by Joo *et al.* that found the motif P<sup>-3</sup>-S/T-P<sup>-1</sup>-F for N1P1 (210). Meanwhile residues at P<sup>-1</sup> to P<sup>-4</sup> were selected by inspection of substitutional analyses (SubAna) described by Vouilleme *et al.* (207). Two fully randomized residues were added at the N-terminal end of the template, forming our first round library sequence of NH<sub>2</sub>-P<sup>-6</sup>P<sup>-5</sup>SNTKF-OH (where P represents a position that was fully randomized with the NNK codon). This construct was

genetically fused to SUMO 1-RB through an 8 residue linker GASNATG and cloned to pFPX.

After screening of approximately 5000 colonies, we identified 7 promising clones that showed a 2-fold increase in ratios relative to control, indicating preferential binding of the peptide to N2P2 rather than M3P6. **Table 3.3** summarizes all these mutations, together with their emission ratios and fold changes relative to the control. We note that there is an apparent lack of consensus in the identified sequences. Based on the X-ray crystal structure of N2P2 (PDB ID 2HE4, 208), these randomized residues could be sitting on the flexible loop region of the domain where diverse functional groups are present. Previous studies have revealed that PDZ domains can accommodate a wide range of residues beyond P<sup>-3</sup> of their PDZ binding motifs. However, the combination of these residues still depends mainly on the C-terminal anchoring motifs downstream. Thus, the screening strategy we developed has successfully identified combinations of residues at P<sup>-5</sup> and P<sup>-6</sup> that appear to improve the affinity and selectivity of the template sequence to N2P2 rather than M3P6.

**Table 3.3 Substitution identified at positions  $P^6$  and  $P^5$  of  $P^6P^5$ SNTKF.**

| N2P2 inhibitory peptide clones | Substitutions in $P^6P^5$ SNTKF <sup>a</sup> |       | Emission ratios | Fold change <sup>b</sup> |
|--------------------------------|--|-------|-----------------|--------------------------|
|                                | $P^6$  | $P^5$ |                 |                          |
| 1                              | Asp  | Lys   | 1.9 ± 0.1       | 1.8                      |
| 2                              | Ser  | Ala   | 2.12 ± 0.07     | 2.04                     |
| 3                              | Leu  | His   | 3.0 ± 0.2       | 2.9                      |
| 4                              | Glu  | Cys   | 2.3 ± 0.2       | 2.2                      |
| 5                              | Gln  | Pro   | 2.43 ± 0.09     | 2.34                     |
| 6                              | Lys  | Ala   | 2.2 ± 0.1       | 2.1                      |
| 7                              | Val  | Pro   | 2.4 ± 0.3       | 2.3                      |
| control                        | -  | -     | 1.04 ± 0.05     | -                        |

<sup>a</sup>Abbreviations correspond to the standard three letter amino acid designations.

<sup>b</sup>Fold changes are normalized values against the control.

We next attempted to further improve the affinity of two variants from the first library by randomizing Asn at position  $P^3$ . For this second round, we selected clones 3 and 7 due to their high emission ratios. Clone 3, expressing a SUMO-RB-LHSNTKF peptide fusion, had the highest ratio of all 7 peptide sequences picked. Further mutagenesis of its anchoring residues could potentially lead to a tighter and more selective N2P2 peptide inhibitor. Clone 7 (SUMO-RB-VPSNTKF) was the only variant with two nonpolar residues at the randomized positions. Accordingly, we speculated that its binding interactinos might be distinct from those in the other variants, and mutating its downstream anchoring sequence might also give rise to a better inhibitor. Unfortunately, the results of

this further mutation effort were disappointing. We did not observe any colonies with improved emission ratios from this round. Most colonies showed high green fluorescence, suggesting that mutating Asn at P<sup>-3</sup> generally weakened the affinity of peptide inhibitor. Two variants derived from clone 3 were observed to have emission ratios higher than 2, but when compared with clone 3 (LHSNTKF), both clones showed at least 25% reduction in emission ratios. Meanwhile, the clone 7 library showed only one colony with ratio higher than 2. This variant however did not show any improvement compared with its template VPSNTKF. Results from the 2<sup>nd</sup> round library support the hypothesis that the C-terminal anchoring sequences (last 4 amino acid residues of the peptide inhibitor or PDZ motif) have unique upstream preferences and any changes on these residues may abolish the binding contributions of the residues that precede it.

### **3.2.5.2 Screening for peptide inhibitors of M3P6-LPA<sub>2</sub> association**

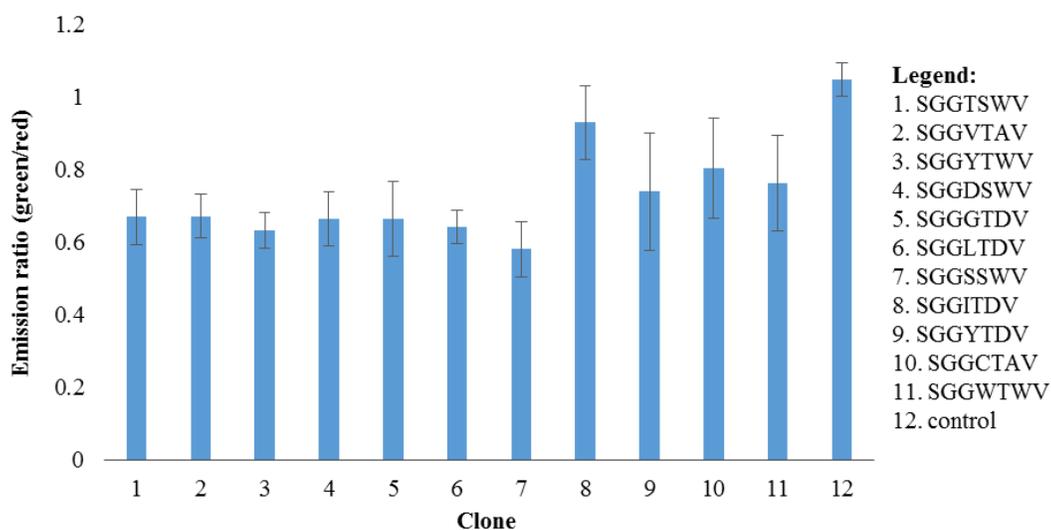
We screened a library to identify peptide inhibitors of the M3P6-LPA interaction using the same pFPX used in the preceding section. Due to the lack of reported binding sequences for M3P6 at the time the library was prepared, we opted to design our template sequence from the canonical PDZ binding motif P<sup>-3</sup>-S/T-P<sup>-1</sup>-V/I/L (211). We installed Val at the carboxyl end of the library due to the prevalence of this residue on most MAGI PDZ partners (212). In addition, SubAna results reported by Vouilleme *et al.* (207) revealed that all NHERF PDZ exhibited no Val affinity at carboxyl end of all peptides tested. Hence, we generated NH<sub>2</sub>-SGGP<sup>-3</sup>S/TP<sup>-1</sup>V-OH (where P represents a position that was fully

randomized with the NNK codon) as our template sequence for the first round of library screening. The upstream SGG residues on the template were intentionally installed and not randomized to limit our library to a size assayable by the current protocol. We envisioned that these residues would be randomized during subsequent rounds of screening.

**Table 3.4 Substitutions identified at positions  $P^3$ ,  $P^2$  and  $P^1$  of potential M3P6 inhibitory peptides.**

| M3P6 inhibitory<br>peptide clones | Substitutions in SGG $P^3$ S/TP $P^1$ V <sup>a</sup> |       |       |
|-----------------------------------|--|-------|-------|
|                                   | $P^3$  | $P^2$ | $P^1$ |
| 1                                 | Thr  | Ser   | Trp   |
| 2                                 | Val  | Thr   | Ala   |
| 3                                 | Tyr  | Thr   | Trp   |
| 4                                 | Asp  | Ser   | Trp   |
| 5                                 | Gly  | Thr   | Asp   |
| 6                                 | Leu  | Thr   | Asp   |
| 7                                 | Ser  | Ser   | Trp   |
| 8                                 | Ile  | Thr   | Asp   |
| 9                                 | Tyr  | Thr   | Asp   |
| 10                                | Cys  | Thr   | Ala   |
| 11                                | Trp  | Thr   | Trp   |

<sup>a</sup>Abbreviations correspond to the standard three letter amino acid designations.



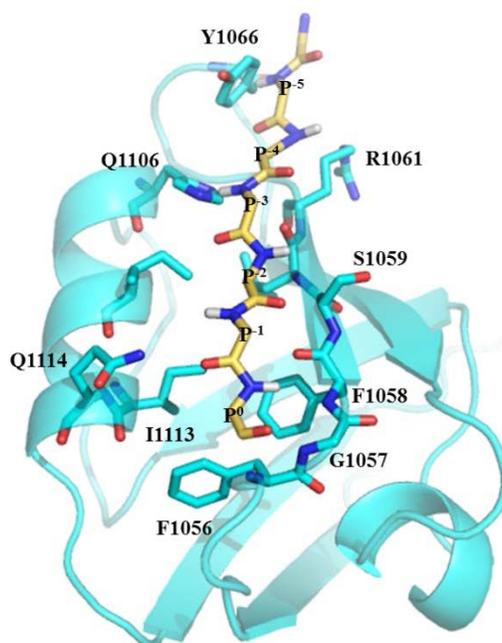
**Figure 3.17 Colony-based emission ratios of 11 clones expressing M3P6 inhibitory peptides (first round of peptide optimization).** Emission ratios were measured in-colony using the digital microscope. Results shown here are mean  $\pm$  standard deviation from more than 100 colonies. Clone number designations are based from **Table 3.4**. Control (clone 12) depicted here is an average emission ratio of colonies co-expressing N2P2-GA, M3P6-RA and SUMO 1-RB-LPA<sub>2</sub>.

Screening for colonies with desired emission profile (low green/red emission ratios) was more challenging for M3P6 than it had been for N2P2. A large number of colonies were intensely green and red fluorescent, suggesting that many residue combinations at randomized positions could cross-interact and rescue N2P2's binding deficiency from Val (at position P<sup>0</sup>). Fortunately, we were able to identify 11 promising clones which displayed ratios lower than the LPA<sub>2</sub> control from around 10,000 *E. coli* colonies screened (**Figure 3.17**). DNA

sequence analysis of these clones, as summarized in **Table 3.4**, revealed a distinct preference at each randomized positions. Interestingly, position P<sup>-1</sup> of all clones was populated by only three amino acids out of the 20 possibilities. Forty five percent (5/11) of them were Trp, 36% (4/11) were Asp, and 18% (2/11) were Ala. The preference for a tryptophan residue identified was also predicted computationally in a recently published work by Zheng *et al.* (185). Their effort predicted three 6-mer peptides with  $\mu\text{M}$ -range inhibition constants. However, all of them suffered from cross-interaction and poor selectivity to M3P6. Position P<sup>-2</sup> which was designed to only be either Ser or Thr, showed a preference for Thr. Seventy three percent (8/11) of clones had Thr mutations including two with Trp and all variants with Asp and Ala at P<sup>-1</sup>. The remaining 27% (3/11) of the clones with Ser were exclusively observed in combination with Trp at P<sup>-1</sup>. Although a bit more diverse, we can still observe a consensus on the residues present at P<sup>-3</sup>. We find that 75% of peptides with Asp at P<sup>-1</sup> had nonpolar residues at P<sup>-3</sup>. These nonpolar residues included Gly, Ile and Leu residues. In contrast, those peptides with Trp at P<sup>-1</sup> had residues capable of hydrogen bonding such as Asp, Ser, Thr, Trp and Tyr.

Structural analysis of manually docked M3P6-peptide complex helped us to rationalize most of the mutations identified on the 1<sup>st</sup> round library. **Figure 3.18** shows the simulated model of M3P6 in complex with a 7-mer peptide. We deliberately removed the side chains of the peptide to simplify our visual inspection. Residue numbers on M3P6 were based on their positions in the context of the entire MAGI-3 protein. From this complex, we identified Ser1059

or Gln-1114 from M3P6 that could be involved in a network of hydrogen bond with Trp or Asp at P<sup>-1</sup> of the peptide. An alanine residue at this position could potentially interact with other nonpolar moieties in the vicinity, especially with Phe-1058 and Ile-1113. Both Ser and Thr at P<sup>-2</sup> would most likely form a hydrogen bond with the imidazole moiety of His (His-1106 for M3P6) residue sitting on an  $\alpha$ -helix of the PDZ domain. This favorable interaction is also commonly observed on other resolved PDZ-peptide complex structures. However, interactions of P<sup>-3</sup> residue were more difficult to predict due to the flexibility of loop in this vicinity. We hypothesized that residues with polar side chains could possibly form a network of hydrogen bonds with Arg1061 and Tyr1066 on the loop, whereas those with nonpolar moieties may have interacted with the nonpolar residues also in close proximity.



**Figure 3.18 Simulated M3P6-peptide complex.** A manually docked M3P6-peptide complex highlights all the potential interacting residues on the binding region of M3P6. To simplify and generalize the visual inspection, side chains of each peptide residue (yellow) was deliberately removed while the interacting residues of M3P6 (cyan) were converted to their stick representations, labeled using one-letter amino acid codes and position number based on the complete MAGI-3 protein. Peptide residues were also labeled using the standard numbering scheme followed for PDZ binding motifs. This modelled structure of M3P6 was predicted by BioSerf v2.0, an automated homology modelling server (209).

Given the results of the 1<sup>st</sup> library, we decided to perform a second round of screening with two of the peptide sequences identified in the first round. These two sequences had either the SWV or TDV tripeptide at the carboxyl end (P<sup>-2</sup>P<sup>-1</sup>P<sup>0</sup>) of the peptide library. Randomization at position P<sup>-3</sup> was restricted to those amino acids prevalent to each anchoring sequences, including Ser, Thr and Asp (STD) for SVW; and Leu, Ile and Tyr (LIY) for TDV. Positions P<sup>-4</sup> and P<sup>-5</sup> were completely randomized with the NNK codon to generate NH<sub>2</sub>-SP<sup>-5</sup>P<sup>-4</sup>(STD)SWV-OH and NH<sub>2</sub>-SP<sup>-5</sup>P<sup>-4</sup>(LIY)TDV-OH as our 7-mer peptide library sequences for round two. Panning about 20,000 *E. coli* colonies (10,000 colonies for each library) led to identification of 6 colonies (3 clones for each template) with improved emission ratios (i.e., ratios that were lower than 1<sup>st</sup> round clones). **Table 3.5** summarizes the mutations identified on all 6 clones together with their emission ratios and relative fold-changes. We observed only a moderate

improvement in emission ratios compared with controls. We postulate three possible explanations for this observation. First, residues beyond P<sup>3</sup> may not have a major role in increasing the peptide affinity, but have more influence on tuning the selectivity of the peptide towards M3P6. For instance, the prevalence of Phe and Tyr residues upstream could be involved in  $\pi$ - $\pi$  stacking with Tyr1066 on the loop region of M3P6 binding site that is lacking on N2P2 loop. Second, overexpression of three cassettes in *E. coli* colonies would lead to some residual SUMO-RB-peptide to bind nonspecifically to N2P2-GA producing a green fluorescent complex. And third, the filter bandwidth used to image the green channel in our digital microscope was somewhat too wide. This resulted in a high background green fluorescence from colonies. Nonetheless, our screening strategy has identified NH<sub>2</sub>-SLVTSWV-OH, NH<sub>2</sub>-STGASWV-OH and NH<sub>2</sub>-SVKYTDV-OH as potential 7-mer peptide inhibitors for MAGI-3 PDZ 6.

**Table 3.5 Substitution identified on the second round of M3P6 inhibitory peptide optimization.**

| M3P6<br>inhibitory<br>peptide<br>clones | Substitutions in SP <sup>-5</sup> P <sup>-4</sup> P <sup>-3</sup> SWV or<br>SP <sup>-5</sup> P <sup>-4</sup> P <sup>-3</sup> TDV <sup>a</sup> |                 |                 |                                 | Emission<br>ratios | Fold<br>change <sup>b</sup> |
|---|---|-----------------|-----------------|---------------------------------|--------------------|-----------------------------|
|   | P <sup>-5</sup>   | P <sup>-4</sup> | P <sup>-3</sup> | P <sup>-2</sup> -P <sup>0</sup> |                    |                             |
|   | 1   | Leu             | Ala             | Asn                             |                    |                             |
| 2                                       | Thr   | Gly             | Ala             | SWV                             | 0.39 ± 0.03        | 2.87                        |
| 3                                       | Leu   | Val             | Thr             | SWV                             | 0.31 ± 0.03        | 3.63                        |
| 4                                       | Asn   | Arg             | Tyr             | TDV                             | 0.50 ± 0.06        | 2.25                        |
| 5                                       | Val   | Lys             | Tyr             | TDV                             | 0.48 ± 0.04        | 2.34                        |
| 6                                       | Met   | Asp             | Phe             | TDV                             | 0.69 ± 0.14        | 1.62                        |
| control                                 |   |                 |                 |                                 | 1.12 ± 0.08        |                             |

<sup>a</sup>Abbreviations correspond to the standard three letter amino acid designations.

<sup>b</sup>Fold changes are normalized values against the control.

### 3.3 Conclusion

In this Chapter, we describe the development, validation, and application, of a new in-colony screening method that can detect PPIs in bacterial colonies. This strategy relied on the construction of bacterial plasmid designated pFPX, which is capable of co-expressing the interacting proteins as GA, RA ad RB fusions. Using pFPX together with the custom-built digital microscope as a colony imaging system, we have successfully imaged the association of E1 and K1 helical coiled-coils, p53 and its antagonist HDM 2, and HRas<sub>WT</sub> and Raf<sub>WT</sub> RBD against their binding deficient variants. We have demonstrated the

suitability of our method for screening small-to-medium sized gene libraries through a reversion experiment of non-binding variant Raf<sub>R89L</sub> RBD towards its partner HRas<sub>WT</sub>. From 20 colonies we picked with desired phenotype, three of them reverted back to wild type Raf-1 RBD (L89R). The other 17 clones were shared by variants with Tyr, His, Ser, Asn and Cys mutations at the 89<sup>th</sup> position of Raf RBD. Although still unverified, we believe that these alternative mutations in Raf RBD may have rescued the hydrogen bonding formed between 89<sup>th</sup> residue of Raf RBD and the corresponding interface residues of HRas<sub>WT</sub>.

We also attempted to design an orthogonal binding partner for Raf<sub>R89L</sub> RBD variant by screening a library of HRas extragenic suppressors. Here, we fully randomized the 38<sup>th</sup> and 40<sup>th</sup> amino acid residues sitting on the integral  $\beta$ -sheet of HRas<sub>WT</sub>. Results from this library screening identified three potential clones that apparently rescued the binding of HRas towards Raf<sub>R89L</sub> RBD. These three variants included clone 7 (D38S, Y40P), clone 12 (D38P, Y40M) and clone 14 (D38F, Y40L). However, cell studies and *in vitro* binding experiments are needed to validate the activities and affinities of these HRas suppressors.

Furthermore, we extended the application of FPX-based PPI screening by generating a set of 7-mer modulators that selectively inhibit either NHERF-2 PDZ 2 (N2P2) or MAGI-3 PDZ 6 (M3P6). Both of these proteins interact competitively through their PDZ domains to the C-terminal residues of lysophosphatidic acid receptor 2 (LPA<sub>2</sub>) and were recently identified as key players in colon tumorigenesis. We intentionally generated all peptide libraries as C-terminal fusions of SUMO1-RB to liberate the carboxyl ends of all library

sequences. For N2P2, we panned a library using NH<sub>2</sub>-P<sup>6</sup>P<sup>5</sup>SNTKF-OH as the preliminary template, which we assembled from substitutional analysis (SubAna) results reported previously (207). Our assay identified seven clones with promising phenotype. One of these peptides was NH<sub>2</sub>-LHSNTKF-OH which gave an approximately three-fold increase in emission ratio compared to the control. We further tried to tune the affinity or selectivity of some peptides generated from the 1<sup>st</sup> round, but no clones with improved phenotype were detected. Next we performed two rounds of library screening for peptide inhibitors of M3P6, starting from NH<sub>2</sub>-SGGP<sup>3</sup>-S/TP<sup>1</sup>V-OH. This template sequence was assembled from the canonical PDZ binding motif P<sup>3</sup>-S/T-P<sup>1</sup>-V/I/L. Valine was also chosen as the terminal residue due to its prevalence on most MAGI protein partners. Two rounds of mutagenesis yielded three promising 7-mer peptides, NH<sub>2</sub>-SLVTSWV-OH, NH<sub>2</sub>-STGASWV-OH and NH<sub>2</sub>-SVKYTDV-OH, which apparently bound with higher affinity to M3P6 relative to our control.

The FPX screening strategy has proven suitable for detecting PPIs, guided-engineering of protein affinities and screening of either protein or peptide inhibitor libraries. For future work, confirming the binding characteristics and biological activities of the engineered proteins and peptides would certainly strengthen our claims.

### 3.4. Materials and methods

#### 3.4.1 General method and reagents

Protocols employed for all molecular biology techniques like polymerase chain reaction (PCR), restriction enzyme digestion, ligation and transformation and DNA gel electrophoresis were from Sambrook *et al.* (115). Overlap extension PCR for making fusion construct were performed as described in Bassette *et al.* (213). All synthetic oligonucleotides (**Table 3.6**) were purchased from Integrated DNA Technologies (Coralville, IA). *Pfu* polymerase, T4 DNA ligase and restriction enzymes used were obtained from Thermo Scientific while Q5 and *Taq* polymerases were bought from New England Biolab (NEB). PCR and digestion products as well as plasmid were purified using gel extraction and miniprep kits (Thermo Scientific or BioBasic Inc.), respectively.

**Table 3.6 List of oligonucleotides used for cloning and sequencing.**

| Oligo name      | Type <sup>a</sup> | DNA sequence <sup>b</sup>   |
|-----------------|-------------------|---|
| 2ndCassette_seq | F                 | ACA GGT AGC ACA GGC AGC G   |
| BglII_Ras       | F                 | CAA AAG ATC TTT TAA GAA GGA GAT ATA<br>CAT ATG GGT CAC CAC CAC CAT CAT CAT<br>GGG AAG |
| ddFP(MVSK)link  | F                 | GGA GGT GGG GGA TCT ATG GTG AGC<br>AAG AGC GAG GAG                                    |
| ddGFP_BamHI     | R                 | CAC GGA TCC TTA GTG GTG GTG GTG GTG<br>GTG GCT ACC GCT GCC TGT GCT                    |
| ddRFP_HindIII   | R                 | GTG AAG CTT TTA CTT GTA CAG CTC GTC   |

---

|                    |   |   |
|--------------------|---|---|
|                    |   | CAT GCC   |
| E1_EagI            | R | TCC CGG CCG TTT CTC CAG TGC TCT GAC<br>CTC CTT CTC  |
| E1_EcoRI           | F | CAC GAA TTC AAT GAA AGT GTC AGC CCT<br>CGA AAA CGA AGT C  |
| E1_KpnI            | R | CCC GGT ACC TTT CTC CAG TGC TCT GAC<br>CTC CTT CTC  |
| E1_SalI            | R | CCC GTC GAC TTT CTC CAG TGC TCT GAC<br>CTC CTT CTC  |
| E1_xbaI            | F | CCA TCT AGA AAT GAA AGT GTC AGC CCT<br>CGA AAA CGA AGT C  |
| E1_xhoI            | F | CCA CTC GAG AAT GAA AGT GTC AGC CCT<br>CGA AAA CGA AGT C  |
| EGFP/RA_MluI       | R | TGG ACG CGT CCT AGG TTA CTT GTA CAG<br>CTC GTC CAT GCC GAG  |
| HDM2_KpnI          | R | TCC GGT ACC CTG CTG ATT GAC TAC TAC<br>CAA GTT CCT GTA GAT CAT  |
| HDM2_xbaI          | F | CCC TCT AGA CAT GTG CAA TAC CAA CAT<br>GTC TGT ACC TAC T  |
| Inhibitorp53_EcoRI | F | CCA GAA TTC AAC CAG TTT CGC GGA GTA<br>CTG GGC CCT GTT GAG TCC CGG AGG TAT<br>GGT GAG CAA GGG CGA GGA GCT |
| K1_EagI            | R | CAC CGG CCG CTC CTT CAG TGC ACT CAC<br>TTT TTC  |
| K1_EcoRI           | F | CCC GAA TTC AAT GAA GGT TAG TGC CTT<br>GAA AGA AAA CGT  |
| K1_SalI            | R | TGG GTC GAC CTC CTT CAG TGC ACT CAC   |

---

---

|                        |   |  |
|------------------------|---|--|
|                        |   | TTT TTC  |
| K1_xhoI                | F | CCC CTC GAG AAT GAA GGT TAG TGC CTT<br>GAA AGA AAA CGT                     |
| LHSXTKF-R              | R | CTC AAG CTT TTA GAA CTT GGT MNN CGA<br>ATG AAG ACC GGT TGC GTT             |
| LPA <sub>2</sub> _link | F | GGT TCT GCA TGC TGC GCT TGC CTG AGG<br>CAG TCC                             |
| LPA <sub>2</sub> MDSTL | R | CCC AAG CTT TTA CAG GGT ACT GTC CAT<br>CAG TGG GTG                         |
| M3P6_SalI.HindIII      | R | CCC AAG CTT TTA GTC GAC TCC CCC AGT<br>TCC CGG TCT AAG                     |
| M3P6_xhoI.EcorI        | F | CCC CTC GAG GAA TTC AAT GCC CGT GGA<br>ATT GGA ACG CGG A                   |
| N2P1_EcorI.xbaI        | F | CCA GAA TTC TCT AGA AAT GCT CAG ACC<br>CAG ACT GTG CAG ACT G               |
| N2P1_SalI.HindIII      | R | TCC AAG CTT TTA GTC GAC GGC CAT TTC<br>CTC TGT GCA TGT CAG                 |
| N2P2_BglII.HindIII     | R | CCC AAG CTT TTA AGA TCT TGG ATC GAC<br>CAC GAG CAG CCT                     |
| N2P2_xhoI              | F | CCA CTC GAG AAT GAG ACT GTG CCA CCT<br>TCG CAA GGG T                       |
| p53_EagI               | R | TCC GTC GAC CGG CCG GGA CAG AAC GTT<br>GTT TTC AGG AAG TAG TTT CCA TAG GTC |
| p53_EcorI              | F | ACC GAA TTC CAT GGA GGA GCC GCA GTC<br>AGA TCC T                           |
| p53_SalI               | F | TCC GTC GAC GGA CAG AAC GTT GTT TTC<br>AGG AAG TAG TTT CCA TAG GTC         |

---

---

|                            |   |   |
|----------------------------|---|---|
| p53_xhoI                   | F | ACC CTC GAG CAT GGA GGA GCC GCA GTC<br>AGA TCC T                                      |
| p53 <sub>w23A</sub> _BglII | R | CCC AGA TCT CAG AAC GTT GTT TTC AGG<br>AAG TAG TTT CGC TAG GTC TGA AAA                |
| p53 <sub>w23A</sub> _SalI  | R | CCC GTC GAC CAG AAC GTT GTT TTC AGG<br>AAG TAG TTT CGC TAG GTC TGA AAA                |
| PeptideRaf_1               | F | GCC GTT TTC CGG TTG CTC CAC GGA GGA<br>CTG CAG GGA GGT GGG GGA TCT ATG GTG<br>AGC AAG |
| PeptideRaf_EcoRI           | F | CAG AAT TCA ATG GGC GGT GAA TGT TGC<br>GCC GTT TTC CGG TTG CTC CAC                    |
| Raf_xhoI                   | F | CCA CTC GAG AAT GCA TGG GAG CAA CAC<br>TAT CCG T                                      |
| Raf-1_EcoRI                | F | CCC GAA TTC AAT GGG AGC AAC ACT ATC<br>CG T   |
| RafL89X                    | F | TGC CTT ATG AAA GCA CTC AAG GTG NNK<br>GGC CTG CAA CCA GAG                            |
| Raflink_EagI/BglII         | R | CTC CTC GCT CTT GCT CAC CAT TCC AGA<br>TCT CGG CCG GCT CCC ATC CAG GAA ATC<br>TAC TTG |
| RafSalIlinkRA              | R | CAC CAT AGA TCC CCC ACC TCC GTC GAC<br>GCT CCC ATC CAG GAA ATC TAC TTG                |
| Ras_HindIII                | F | GGT CAC CAC CAC CAT CAT CAT GGG AAG<br>CTT ATG ACA GAA TAC AAG CTC GTT GTT<br>GTT GGC |
| Ras_xbaI                   | F | TAA GAA GGA GAT ATA CCA TGG GTC TAG<br>ACA TGA CAG AAT ACA AGC TCG TTG TTG            |

---

---

|                    |   |   |
|--------------------|---|---|
|                    |   | TTG GC  |
| RasfR89overlap     | R | CAC CTT GAG TGC TTT CAT AAG GCA   |
| RasKpnRB           | R | CAC CAT AGA TCC CCC ACC TCC GGT ACC<br>GGC CGT TCC ACT GCC  |
| Rassupp_link       | R | CTC TAT AGT GGG GTC GTA TTC GTC CAC<br>AAA ATG  |
| Rassupp_XSX        | F | CAT TTT GTG GAC GAA TAC GAC CCC ACT<br>ATA GAG NNK AGC NNK CGG AAG CAG<br>GTG GTC ATT GAT GGG GAG ACG             |
| RB_link            | R | CAG GCA AGC GCA GCA TGC AGA ACC CTT<br>GTA CAG CTC GTC CAT GCC CAG  |
| RB_linkSUMO        | F | ATG GTG AGC AAG GGC GAG GAG ACC<br>ATC  |
| RB_XS/TXLCOOH      | R | CCC AAG CTT TTA CAG MNN GST MNN CAT<br>CAG CGA GCC ACC GCT TGC GCC CTT GTA<br>CAG CTC GTC CAT GCC CAG             |
| RB_XXSNTKFCOOH     | R | CCC AAG CTT TTA GAA CTT GGT GTT CGA<br>MNN MNN ACC GGT TGC GTT ACC GCT TGC<br>GCC CTT GTA CAG CTC GTC CAT GCC CAG |
| RB_SGGXS/TXVCOOH   | R | CCC AAG CTT TTA CAC MNN GST MNN ACC<br>TCC AGA ACC GGT TGC GTT ACC GCT TGC<br>GCC CTT GTA CAG CTC GTC CAT GCC CAG |
| RB_SXX(STD)SWVCOOH | R | CCC AAG CTT TTA CAC CCA GCT GBY MNN<br>MNN AGA ACC GGT TGC GTT ACC GCT TGC<br>GCC CTT GTA CAG CTC GTC CAT GCC CAG |
| RB_SXX(LIY)TDV     |   | CCC AAG CTT TTA CAC ATC GGT AWD<br>MNN MNN AGA ACC GGT TGC GTT ACC  |

---

---

|              |   |                                     |
|--------------|---|-------------------------------------|
|              |   | GCT TGC GCC CTT GTA CAG CTC GTC CAT |
|              |   | GCC CAG                             |
| SUMO1_linkRB | R | GAT GGT CTC CTC GCC CTT GCT CAC CAT |
|              |   | ACC ACC AAT CTG TTC TCT GTG AGC CTC |
| SUMO1_xbaI   | F | CCG TCT AGA AAT GTC GGA CTC AGA AGT |
|              |   | CAA TCA AGA AGC T                   |
| VPSXTKF-R    | R | CTC AAG CTT TTA GAA CTT GGT MNN CGA |
|              |   | CGG CAC ACC GGT TGC GTT             |

---

<sup>a</sup>Oligo type: F= forward and R = reverse.

<sup>b</sup>Non-standard codes in oligo sequences are mixed bases: B (C, G, T), D(A, G, T), K (G, T), M (A, C), N (A, G, C, T), S (C,G), W (A, T) and Y (C, T).

Genes for E1, K1, NHERF 2 PDZ 2, NHERF 2 PDZ 1, MAGI-3 PDZ 6 and 45 C-terminal amino acid residues of LPA<sub>2</sub> were purchased as *Homo sapiens* codon optimized gblocks from Integrated DNA Technologies (Coralville, IA). Plasmids human p53 (1-393) and pGEX-4T MDM2 WT with accession number 24859 and 16237, respectively, were requested from Addgene. Lastly, plasmids pQE-Ras DHFR, pQE-Raf DHFR, and pQE-R89L used to amplify wild type HRas, Raf-1 RBD and its nonbinding variant R89L were the kind gift of Dr. Stephen Michnick.

Standard PCR amplifications using either *Pfu* or Q5 DNA polymerases were performed in a 50 µL reaction volume. With *Pfu* DNA polymerase, components of the reaction solution were nuclease-free water, 1× reaction buffer, 3% DMSO, 200 µM dNTPs (Life Technologies), 200 nM forward and reverse oligos, 10-50 ng of DNA template and 1.0 unit of enzyme. Typical cycling

parameters were as follows: 1.0 min initial denaturation at 95 °C, 35 cycles of 95 °C for 15 seconds, 54-60 °C for 30 seconds, 72 °C for 60 seconds per kb of target; final extension at 72 °C for 5 minutes. For Q5 DNA polymerase reaction, the 50 µL solution contained nuclease-free water, 1× reaction buffer, 1× GC enhancer, 200 µM dNTPs, 200 nM forward and reverse oligos, 10-50 ng of DNA template and 1.0 unit enzyme. Reaction cycle was carried out using this protocol: 98 °C initial denaturation for 30 seconds, 35 cycles of 98 °C for 10 seconds, 54-60 °C for 25 seconds, 72 °C for 30 seconds per kb of target; final extension at 72 °C for 2 minutes. Cloning for all constructs and library included in this Chapter are outlined in **Table 3.7**. DNA template, relevant oligos, destination plasmids and restriction enzyme sites used to clone each construct were also provided in the table.

**Table 3.7 Gene constructs and cloning strategy described in Chapter 3.**

| <b>Gene construct</b>   |                | <b>PCR Amplification 1</b> |                                     | <b>Gene assembly (overlap PCR)</b> |                |
|-------------------------|----------------|----------------------------|-------------------------------------|------------------------------------|----------------|
|                         | <b>plasmid</b> | <b>Template</b>            | <b>Oligos<sup>a</sup></b>           | <b>Oligos</b>                      | <b>RE</b>      |
|                         |                |                            | <b>(F/R)</b>                        | <b>(F/R)</b>                       | <b>sites</b>   |
|                         |                |                            |                                     |                                    | <b>(5'/3')</b> |
| Raf <sub>WT</sub> RBD   | n/a            | pQE-Raf<br>DHFR            | Raf_xho1/<br>Raflink_<br>EagI.BglII | n/a                                | n/a            |
| Raf <sub>R89L</sub> RBD | n/a            | pQE-R89L                   | Raf-1_EcorI/                        | n/a                                | n/a            |

|                            |               | DHFR                                   | RafSaIIink                        |                                  |                                 |
|----------------------------|---------------|--|-----------------------------------|----------------------------------|---------------------------------|
|                            |               |  | RA                                |                                  |                                 |
| HRas WT                    | n/a           | pQE-Ras                                | Ras_xbaI/                         | n/a                              | n/a                             |
|                            |               | DHFR                                   | RasKpnRB                          |                                  |                                 |
| ddGFP-A                    | n/a           | pddGFP-A <sub>1</sub>                  | ddFP(MVSK)I<br>ink/ddGFP_Ba       | n/a                              | n/a                             |
|                            |               |  | mHI                               |                                  |                                 |
| ddRFP-A                    | n/a           | pddRFP-A <sub>1</sub>                  | ddFP(MVSK)I<br>ink/EGFP-          | n/a                              | n/a                             |
|                            |               |  | RA_MluI                           |                                  |                                 |
| ddRFP-B                    | n/a           | pddRFP-B <sub>1</sub>                  | ddFP(MVSK)I<br>ink/ddRFP_Hi       | n/a                              | n/a                             |
|                            |               |  | ndIII                             |                                  |                                 |
| Raf <sub>WT</sub> RBD-GA   | Tricistronicp | Raf <sub>WT</sub> RBD/<br>BAD          | n/a                               | Raf_xhoI/<br>ddGFP_<br>BamHI     | <i>XhoI</i> /<br><i>BamHI</i>   |
| Raf <sub>R89L</sub> RBD-RA | Tricistronicp | Raf <sub>R89L</sub><br>RBD/<br>ddRFP-A | n/a                               | Raf-1_EcorI/<br>EGFP-<br>RA_MluI | <i>EcoRI</i> /<br><i>MluI</i>   |
| HRas WT-RB                 | Tricistronicp | HRas WT/<br>BAD                        | n/a                               | Ras_xbaI/<br>ddRFP_<br>Hind III  | <i>XbaI</i> /<br><i>HindIII</i> |
| ECCA VFRLH-<br>RA          | n/a           | pddRFP-A <sub>1</sub>                  | PeptideRaf_1/<br>EGFP-<br>RA_MluI | n/a                              | n/a                             |
| ECCA VFRLH<br>(full)-RA    | pFPX          | ECCA VFRL<br>LH-RA                     | PeptideRaf_Ec<br>orI/ EGFP-       | n/a                              | <i>EcoRI</i> /<br><i>MluI</i>   |

|   |      |                       | RA_MluI                    |     |               |
|---|------|-----------------------|----------------------------|-----|---------------|
| E1 (1 <sup>st</sup> cassette)             | pFPX | E1K1                  | E1_xhoI/                   | n/a | <i>XhoI/</i>  |
|   |      | gblocks               | E1_EagI                    |     | <i>EagI</i>   |
| E1 (2 <sup>nd</sup> cassette)             | pFPX | E1K1                  | E1_EcorI/                  | n/a | <i>EcoRI/</i> |
|   |      | gblocks               | E1_SalI                    |     | <i>SalI</i>   |
| E1 (3 <sup>rd</sup> cassette)             | pFPX | E1K1                  | E1_xbaI/                   | n/a | <i>XbaI/</i>  |
|   |      | gblocks               | E1_KpnI                    |     | <i>KpnI</i>   |
| K1 (1 <sup>st</sup> cassette)             | pFPX | E1K1                  | K1_xhoI/                   | n/a | <i>XhoI/</i>  |
|   |      | gblocks               | K1_EagI                    |     | <i>EagI</i>   |
| K1 (2 <sup>nd</sup> cassette)             | pFPX | E1K1                  | K1_EcorI/                  | n/a | <i>EcoRI/</i> |
|   |      | gblocks               | K1_SalI                    |     | <i>SalI</i>   |
| p53 <sub>WT</sub>                         | pFPX | pHuman                | p53 <sub>WT</sub> _xhoI/   | n/a | <i>XhoI/</i>  |
| (1 <sup>st</sup> cassette)                |      | p53(1-393)            | p53 <sub>WT</sub> _EagI    |     | <i>EagI</i>   |
| p53 <sub>W23A</sub>                       | pFPX | pHuman                | p53 <sub>WT</sub> _xhoI    | n/a | <i>XhoI/</i>  |
| (1 <sup>st</sup> cassette)                |      | p53(1-393)            | /p53 <sub>W23A</sub> _     |     | <i>BglII</i>  |
|   |      |                       | BglII                      |     |               |
| p53 <sub>WT</sub>                         | pFPX | pHuman                | p53 <sub>WT</sub> _EcorI/  | n/a | <i>EcoRI/</i> |
| (2 <sup>nd</sup> cassette)                |      | p53(1-393)            | p53 <sub>WT</sub> _SalI    |     | <i>SalI</i>   |
| p53 <sub>W23A</sub>                       | pFPX | pHuman                | p53 <sub>WT</sub> _EcorI   | n/a | <i>EcoRI/</i> |
| (2 <sup>nd</sup> cassette)                |      | p53(1-393)            | /p53 <sub>W23A</sub> _SalI |     | <i>SalI</i>   |
| HDM 2 <sub>1-113</sub>                    | pFPX | pGEX-4T               | HDM2_xbaI/H                | n/a | <i>XbaI/</i>  |
|   |      | HDM 2                 | DM2_KpnI                   |     | <i>KpnI</i>   |
| TSFAEYALLSP                               | pFPX | pddRFP-A <sub>1</sub> | Inhibitorp53_E             | n/a | <i>EcoRI/</i> |
| GG-RA                                     |      |                       | cor1/EGFP-                 |     | <i>MluI</i>   |
|   |      |                       | RA_MluI                    |     |               |
| Raf <sub>189X</sub> _1 <sup>st</sup> half | n/a  | pQE-R89L              | Raf-1_EcorI/               | n/a | n/a           |

|   |      | DHFR                              | RasfR89               |              |                |
|---|------|-----------------------------------|-----------------------|--------------|----------------|
|   |      |                                   | overlap               |              |                |
| Raf <sub>L89X</sub> _2 <sup>nd</sup> half | n/a  | pQE-R89L                          | Raf <sub>L89X</sub> / | n/a          | n/a            |
|   |      | DHFR                              | RafSalIlink           |              |                |
|   |      |                                   | RA                    |              |                |
| Raf <sub>L89X</sub> _lib                  | pFPX | Raf <sub>L89X</sub>               | n/a                   | Raf-1_EcoRI/ | <i>EcoRI</i> / |
| full                                      |      | 1 <sup>st</sup> & 2 <sup>nd</sup> |                       | RafSalIlink  | <i>Sall</i>    |
|   |      |                                   |                       | RA           |                |
| HRassupp_1 <sup>st</sup>                  | n/a  | pFPX-                             | Ras_xbaI/Rass         | n/a          | n/a            |
| half                                      |      | Raf/HRas                          | upp_link              |              |                |
| HRassupp_2 <sup>nd</sup>                  | n/a  | pFPX-                             | Rassupp_XSX           | n/a          | n/a            |
| half                                      |      | Raf/HRas                          | /RasKpn               |              |                |
|   |      |                                   | RB                    |              |                |
| HRassupp_lib                              | pFPX | HRassupp                          | n/a                   | Ras_xbaI/    | <i>XbaI</i> /  |
| XSX full                                  |      | 1 <sup>st</sup> & 2 <sup>nd</sup> |                       | RasKpnRB     | <i>KpnI</i>    |
| N2P2                                      | pFPX | LPA gblock                        | N2P2_xhoI/            | n/a          | <i>XhoI</i> /  |
|   |      |                                   | N2P2_BglII.           |              | <i>BglII</i>   |
|   |      |                                   | HindIII               |              |                |
| M3P6                                      | pFPX | LPA gblock                        | M3P6_xhoI.            | n/a          | <i>EcoRI</i> / |
|   |      |                                   | EcoRI/                |              | <i>Sall</i>    |
|   |      |                                   | M3P6_SalI.            |              |                |
|   |      |                                   | HindIII               |              |                |
| SUMO 1                                    | n/a  | pET-                              | SUMO1_xbaI/           | n/a          | n/a            |
|   |      | champion                          | SUMO1_                |              |                |
|   |      |                                   | linkRB                |              |                |
| RB  | n/a  | pddRFP-B <sub>1</sub>             | RB_link               | n/a          | n/a            |
|   |      |                                   | SUMO/                 |              |                |

|  |      |  | RB_Link  |   |                                 |
|--|------|--|--|---|---------------------------------|
| LPA <sub>2</sub> (311-351)             | n/a  | LPA gblock   | LPA <sub>2</sub> link/<br>LPA <sub>2</sub> MDSTL | n/a                                       | n/a                             |
| SUMO-RB-<br>LPA <sub>2</sub> (311-351) | pFPX | SUMO 1/<br>RB/ LPA <sub>2</sub><br>(311-351)       | n/a  | SUMO1_<br>xbaI/<br>LPA <sub>2</sub> MDSTL | <i>XbaI</i> /<br><i>HindIII</i> |
| SUMO-RB-<br>XXSNTKF                    | pFPX | pFPX-<br>SUMO-RB-<br>LPA <sub>2</sub><br>(311-351) | SUMO1_xbaI/<br>RB_XXSNTK<br>FCOOH                | n/a                                       | <i>XbaI</i> /<br><i>HindIII</i> |
| SUMO-RB-<br>LHSXTKF                    | pFPX | Clone 3<br>pFPX-<br>SUMO-RB-<br>LHSNTKF            | SUMO1_xbaI/<br>LHSXTKF-R                         | n/a                                       | <i>XbaI</i> /<br><i>HindIII</i> |
| SUMO-RB-<br>VPSXTKF                    | pFPX | Clone 7<br>pFPX-<br>SUMO-RB-<br>VPSNTKF            | SUMO1_xbaI/<br>VPSXTKF-R                         | n/a                                       | <i>XbaI</i> /<br><i>HindIII</i> |
| SUMO-RB-<br>SGGXS/TXV                  | pFPX | pFPX-<br>SUMO-RB-<br>LPA <sub>2</sub><br>(311-351) | SUMO1_xbaI/<br>RB_SGGXS/T<br>XVCOOH              | n/a                                       | <i>XbaI</i> /<br><i>HindIII</i> |
| SUMO-RB-<br>SXX(STD)SWV                | pFPX | pFPX-<br>SUMO-RB-<br>LPA <sub>2</sub><br>(311-351) | SUMO1_xbaI/<br>RB_SXX(STD<br>)SWVCOOH            | n/a                                       | <i>XbaI</i> /<br><i>HindIII</i> |

|             |      |                  |             |     |                |
|-------------|------|------------------|-------------|-----|----------------|
| SUMO-RB-    | pFPX | pFPX-            | SUMO1_xbaI/ | n/a | <i>XbaI/</i>   |
| SXX(LIY)SWV |      | SUMO-RB-         | RB_SXX(LIY  |     | <i>HindIII</i> |
|             |      | LPA <sub>2</sub> | )TDVCOOH    |     |                |
|             |      | (311-351)        |             |     |                |

<sup>a</sup>Oligo type: F= forward and R = reverse.

### 3.4.2 DNA sequencing

All DNA sequencing were performed at University of Alberta Molecular Biology Service Unit (MBSU) and DNA Core Services of University of Calgary. The destination plasmid for most of the constructs cloned was pFPX. The pFPX was the derivative of the tricistronic pBAD developed in the preceding Chapter. All constructs on the first and second cassettes were sequenced using pBAD\_NcoI and SeqR\_streptag, and SeqF\_2ndcassette, respectively. Inserts on the third cassette were normally sequenced using the forward oligos of each construct. Analysis of chromatograms and nucleotide sequences were performed using a freeware Sequence Scanner v1.0 (Applied Biosystems) and sequence analyzer software synced on Addgene website.

### 3.4.3 Protein purification of ddFPs

To express and purify ddFP A and B copies, the genes were subcloned into pBAD/His B and were transformed in ElectroMax DH10B *E. coli* (Life Technologies). Colonies were incubated into TB broth supplemented with 100 mg/L ampicillin. After 2 hours, 0.02% L-arabinose were added to induce protein expression. Cultures were then further incubated at 30 °C for 15-20 hours. Cells

were harvested through centrifugation at 7000 r.p.m. (7505 r.c.f.) for 10 min (Beckman), re-suspended with Tris-Cl buffer (pH 7.4) and lysed using high pressure homogenizer (Constant System Ltd.) at 20,000 psi. His6-tagged soluble proteins were purified from cleared cell lysate by Nickel-NTA resin. The lysate-resin mixture was transferred to a column and rinsed with wash buffer (50 mM Tris-HCl, 300 mM NaCl, 20 mM Imidazole, pH 8.0) to clear it from nonbinding proteins from E. coli. Proteins were then eluted from the resin using 50 mM Tris-Cl (pH 7.4) supplemented with 300 mM NaCl and 300 mM imidazole. Proteins were pre-concentrated and cleared from imidazole using filter column with 10 K MWCO (Millipore). Finally, SDS-PAGE and BCA assay were performed to determine purity and concentration.

#### **3.4.4 Dissociation constant ( $K_d$ ) determination**

Dissociation constants of purified ddFP partners were measured by adding and increasing amount of non-fluorescent ddFP-B copy to a fixed amount of ddFP-A thereby generating a fluorescent AB complex. Fluorescence emission spectra were recorded using QuantaMaster spectrofluorometer (Photon Technology International, Inc.) and saturation binding curves were generated by plotting of integrated fluorescence emission as a function of B copy. Experimental data were fit using one-site binding equation (Origin 9.1).

$$[AB] = \frac{[AB]_{max}[B]}{K_d + [B]}$$

### 3.4.5 FPX plasmid and proof-of-concept design

To construct the plasmid, we call pFPX, GA, RA and RB were cloned to the first, second and third cassettes of the tricistronic pBAD-1, respectively. We assembled the entire construct by fusing the three ddFPs to our first proof-of-concept model Raf<sub>WT</sub> RBD-Raf<sub>R89L</sub> RBD-HRas WT. Protein fusions were generated independently by overlap extension PCR producing *XhoI*-Raf<sub>WT</sub> RBD-GA-*BamHI*, *EcoRI*-Raf<sub>R89L</sub> RBD-RA-*MluI* and *XbaI*-HRas WT-RB-*HindIII*. To make pFPX more convenient for succeeding cloning, internal restriction sites were included between the protein of interest and ddFP: *EagI* and *BglII* for GA, *Sall* for RA and *KpnI* for RB. The digested genes were then ligated in a sequential manner. Plasmids with swapped positions of Raf<sub>WT</sub> and Raf<sub>R89L</sub> and a replacement of Raf<sub>R89L</sub> RBD-RA with ECCAVFRLH-RA were also made.

For helical coiled-coil E1 and K1, two plasmids were constructed with GA and RA linked to either E1 or K1 and RB fused to E1. Three plasmids were prepared for screening p53-HDM 2 interactions. All plasmids were cloned with the first 113 amino acid residues of HDM 2 on pFPX's third cassette, therefore expressing HDM 2<sub>1-113</sub>-RB. Genes for the first 30 amino acid residues of p53<sub>WT</sub>, which contains the HDM 2 binding domain, and its variant p53<sub>W23A</sub> were fused to either GA or RA. We also decided cloning the gene for an inhibitor peptide TSFAEYALLSPGG to see whether disruption of p53<sub>WT</sub>-HDM 2 binding is observed in colonies. This is done by replacing p53<sub>W23A</sub>-RA with TSFAEYALLSPGG-RA. Details regarding the templates, oligos and restriction

enzyme sites used for cloning all constructs mentioned above are provided in **Table 3.7**.

### **3.4.6 HRas and Raf-1 RBD library design**

To further validate the efficacy of the screening method, we generated a small library to rescue the binding of RafR89L-1 RBD towards HRas WT by reversion mutation at the 89<sup>th</sup> position of Raf-1 RBD. This was performed by overlap extension PCR with one oligo (RafL89X) carrying the randomized codon at the 89<sup>th</sup> position. The resulting cDNA RafL89X RBD-RA was cloned in a pFPX vector encoding RafWT RBD-GA and HRas WT-RB. A library of HRas extragenic suppressors that could bind to Raf<sub>R89L</sub> RBD were also performed. Here, we randomized the 38<sup>th</sup> and 40<sup>th</sup> of HRas found to interact with WT Raf-1 RBD. Oligo Rassupp\_XSX was used to randomize both position during PCR amplification and the final gene assembly was cloned into pFPX containing genes for RafWT RDB-GA and RafR89L RDB-RA.

### **3.4.7 N2P2-LPA<sub>2</sub>-M3P6 plasmid and library design**

A control pFPX expressing N2P2-GA, M3P6-RA and SUMO-RB-LPA<sub>2</sub> (311-351) was constructed prior to library design (**Table 3.7**). N2P2 and M3P6 were amplified from a gene block, producing *XhoI*-N2P2-*BglII* and *EcoRI*-M3P6-*Sall*. These genes were sequentially cloned in pFPX with SUMO-RB-LPA<sub>2</sub> (311-351) on the third cassette. On the other hand, SUMO-RB-LPA<sub>2</sub> (311-351) gene was assembled by three fragment-overlap extension PCR.

For N2P2 peptide inhibitor library, we assembled SUMO-RB-linker-heptapeptide by amplifying SUMO-RB from the control pFPX above using SUMO\_xbaI and RB\_XXSNTKFCOOH as forward and reverse oligos, respectively. RB\_XXSNTKFCOOH oligo encoded an 8-residue linker (GASGNATG) followed by XXSNTKF where X stands for fully randomized codons in the heptapeptide library. The 2<sup>nd</sup> round library screening for N2P2 inhibitor were intended to improve the affinities of variants 3 and 7 identified from the 1<sup>st</sup> round. Using clones 3 and 7 as template, we copied SUMO 1-RB with SUMO\_xbaI as forward oligo and LHSXTKF-R or VPSXTKF-R as reverse oligo generating *XbaI-SUMO-RB-LHSP<sup>3</sup>TKF-HindIII* and *XbaI-SUMO-RB-VPSP<sup>3</sup>TKF-HindIII*, respectively. Digested products of these genes were subsequently cloned to pFPX forming two sets of libraries.

Conversely, peptide inhibitor library for M3P6 was generated following the same scheme for N2P2. Here, SUMO\_xbaI and RB\_SGGXS/TXVCOOH oligos were used. The latter which encoded for the heptapeptide library SGGP<sup>3</sup>S/TP<sup>1</sup>V has full randomization at positions P<sup>-1</sup> and P<sup>-3</sup> from valine and a serine or threonine at position P<sup>-2</sup>. Second round of library screening aimed at randomizing positions P<sup>-3</sup> to P<sup>-5</sup> from the promising variants identified from the 1<sup>st</sup> library. We amplified SUMO 1-RB using SUMO\_xbaI as forward oligo and RB\_SXX(STD)SWVCOOH or RB\_SXX(LIY)TDVCOOH as reverse oligo. Digested genes were then cloned to pFPX creating two sets of library for the second round.

### **3.4.8 In colony fluorescence screening using a digital microscope**

Fluorescence of *E. coli* colonies on plates were imaged using the digital microscope that was previously described in detail (214). A digital acquisition protocol was followed for measuring fluorescence of each colony on the plate. For an efficient screening, colonies on the plate were limited to about 100-500. Green and red emissions were measured from 500 ms and 1000 ms exposure times, respectively. Results were then processed and analyzed using Image Pro Plus (Media Cybernetics). Emission ratios (= green intensity/ red intensity) of colonies on plates were identified before selecting them for further characterization.

## Chapter 4

### Conclusion and future directions

#### 4.1 Summary of thesis

Ever since Chalfie's breakthrough demonstration that the gene encoding GFP could be functionally expressed in the unnatural host *C. elegans* (2), GFP and its variants have continued to become increasingly popular imaging tools in cell biology. By genetically fusing a FP with a protein of interest, researchers can track and image protein localization and trafficking in live cells. For example, such approaches enable real-time monitoring of cell division and a wide variety of other intracellular processes, which would be impractical using traditional fluorescent dyes (215, 216). During this past decade, great strides have been made in transforming FPs from mere genetic intracellular tags into versatile tools with a wide range of applications. Some highlights include: photoactivatable FPs for use in superresolution microscopy (217, 218); FP-based actuators for precise control of protein activities (95); biosensors for detection of posttranslational modifications (40); biosensors of membrane potential (61, 62); and biosensors of influx or release of various second messengers and other small molecules (35, 36, 38, 219).

Dr. Spencer Alford, a former Ph.D. student in this lab, developed a new class of FP-based biosensors which complements other FP-based reporter strategies like Förster resonance energy transfer (FRET) and protein

complementation assay (PCA). This approach harnesses the oligomerization-dependent fluorescence of DsRed, the *Discosoma sp.* derived red FP from which mRFP1, mCherry, and dTomato were developed (7, 138, 220). Through iterative rounds of interface mutations and directed evolution, Dr. Alford (97) generated a low affinity RFP heterodimer pair, designated as dimerization-dependent RFP (ddRFP). This ddRFP is composed of a dimly fluorescent A copy and a dark B copy. Upon dimerization, the A copy substantially increases its red fluorescence intensity. Closely related green and yellow ddFP versions were introduced shortly after (98). Together, these ddFPs were shown to be useful for imaging of Ca<sup>2+</sup> ion concentrations and protein-protein interactions (PPIs) in live cells.

To further improve the utility of ddFPs, we recently developed the Fluorescent Protein eXchange (FPX) strategy, which utilizes the ability of A copies from red and green ddFP to bind interchangeably with a dark B copy. This versatile method was used to generate ratiometric reporters for protease activities and small molecule messenger signaling (99).

We have now used FPX technology as the basis for a new in-colony strategy that can detect PPI and screen a library of potential protein or peptide antagonists directly in *E. coli* colonies. This method will complement the currently available techniques for studying PPI and simplify the optimization, design and discovery of biologic antagonists, especially for those molecular targets with no available crystal structures. One of the keys to our strategy is a polycistronic vector that can simultaneously express at least two interacting proteins as ddFP fusions. As described in Chapter 2, we constructed four

pBAD/His B derived polycistronic vectors, which are able to co-express either two (bicistronic vectors) or three (tricistronic vectors) separate proteins. We also characterized the induction efficiency of each expression cassette cloned in polycistronic vectors by monitoring the production of FPs (mRFP1, EGFP and mPapaya). Our test expression assays revealed that all four polycistronic vectors maintained the inherent tight regulated transcription of pBAD/His B. Expression levels of all cassettes plateaued at approximately 0.02% of the L-arabinose inducer. Overall, we conclude that we successfully constructed polycistronic vectors that retained the important features of the parent vector, pBAD/His B.

Starting from two tricistronic pBAD vectors (tri-pBAD or T7T-pBAD), we designed and constructed a bacterial plasmid called pFPX. This plasmid enables the expression of interacting proteins in *E. coli* as GA, RA and RB fusions. Our goal was to use this system for fluorescence detection of the association and disruption of protein partners in the context of bacterial colonies by imaging of the green-to-red fluorescent ratios. Using this strategy, we successfully demonstrated the higher affinity association of E1 and K1 coiled-coils, p53<sub>WT</sub> and HDM 2, and HRas<sub>WT</sub> and Raf-1<sub>WT</sub> RBD, relative to lower affinity mutants in each case.

Our success at applying this strategy to several well-characterized interactions inspired us to use this method to screen small-to-medium sized libraries. As a first example, we screened a library for reversion mutations in a binding impaired variant Raf<sub>R89L</sub> RBD. Out of twenty colonies showing the desired phenotype, three clones (15%) were mutated back to its wild type residue,

Arg; four clones (20%) with Tyr; four clones (20%) with His; four clones (20%) with Asn; three clones (15%) mutated to Ser; and the remaining two clones (10%) were converted to Cys and Phe mutants. Although we only had moderate success in selecting reverted clones, we speculated that other identified residues could have rescued the lost hydrogen bonds formed by Arg89 of Raf-1 RBD and the adjacent positions on the  $\beta$ -strand (Asp38, Ser39 and Tyr40) at the interface region of HRas<sub>WT</sub>. Furthermore, we tried to pan a library of HRas extragenic suppressors to identify an orthogonal binding partner for the R89L variant of Raf-1 RBD. We screened about 10,000 colonies and found 15 clones with promising phenotype. Of these, 3 clones showed emission ratios similar to our positive control. These clones included HRas 7 (D38S, Y40P), HRas 12 (D38P, Y40M) and HRas 14 (D38F, Y40L). Although the results from this strategy were very encouraging, further investigations such as cell studies and *in vitro* binding characterizations of the identified mutants will be necessary to strengthen our claims.

In an effort to apply our technique to a more challenging PPI target, we carried out another in-colony library screen to identify 7-mer peptide inhibitors for two intriguing scaffold proteins, NHERF-2 and MAGI-3. It was recently reported that these proteins, through their PDZ domains, could either promote or halt growth and propagation of colon tumors by interacting competitively with the C-terminal motif of lysophosphatidic acid receptor 2 (LPA<sub>2</sub>) (182). We performed the screening using the pFPX plasmid, which co-expresses NHERF-2 PDZ 2 (N2P2), MAGI-3 PDZ 6 (M3P6) and a peptide library as GA, RA and RB

fusions, respectively. The peptide library was genetically linked to the C-terminal region of SUMO-RB forming a final gene assembly, SUMO-RB-P<sup>6</sup>P<sup>5</sup>P<sup>4</sup>P<sup>3</sup>P<sup>2</sup>P<sup>1</sup>P<sup>0</sup> (P<sup>n</sup> denotes any amino acid residue and its position from the carboxyl end). This design would liberate the carboxyl moiety of P<sup>0</sup> residue, which has been found to be an essential feature of peptide-based PDZ inhibitors.

Initial screening of peptide-based N2P2 inhibitors was carried out using NH<sub>2</sub>-P<sup>6</sup>P<sup>5</sup>SNTKF-OH as the template sequence. After screening of approximately 5000 colonies, we selected seven clones which registered a greater than 2-fold change in fluorescence ratio relative to control. From the seven clones, we have identified three potential inhibitory peptides: NH<sub>2</sub>-LHSNTKF-OH, NH<sub>2</sub>-QPSNTKF-OH and NH<sub>2</sub>-VPSNTKF-OH, with higher selectivities towards N2P2. In parallel, we performed two rounds of library screening for M3P6 inhibitory peptides. Due to the lack of known peptide inhibitors that we could use as a template sequence, we decided to start with NH<sub>2</sub>-SGGP<sup>3</sup>-S/TP<sup>1</sup>V-OH as the initial library template. This sequence follows the rule for a canonical PDZ anchoring motifs, P<sup>3</sup>-S/T-P<sup>1</sup>-V/I/L. We decided to fix a Val at P<sup>0</sup> due to its prevalence in most MAGI interacting partners. Interestingly, sequencing results from eleven selected clones showed obvious residue preference at positions P<sup>2</sup> and P<sup>1</sup> of the peptide, which led us to identify two probable anchoring sequences: NH<sub>2</sub>-P<sup>6</sup>P<sup>5</sup>P<sup>4</sup>P<sup>3</sup>SWV-OH and NH<sub>2</sub>-P<sup>6</sup>P<sup>5</sup>P<sup>4</sup>P<sup>3</sup>TDV-OH. We carried out the second round of optimization using the two sequences we identified from the first in-colony screening. From approximately 20,000 colonies screened, we identified the top three peptides, NH<sub>2</sub>-SLVTSWV-OH, NH<sub>2</sub>-STGASWV-OH and NH<sub>2</sub>-

SVKYTDV-OH, which appear to have increased affinity and selectivity towards M3P6.

## **4.2 Immediate future works**

### **4.2.1 *In vitro* characterization of Raf-1 RBD and HRas mutants**

To support our claims regarding the FPX-based in-colony screening method, we must validate the results of the reversion mutations and HRas extragenic suppressor library *in vitro*. Earlier, we attempted to measure the affinities of purified proteins by steady-state fluorescence polarization using GFP as the fluorescent dye. However, we did not get any reliable data due to very minute changes in polarization. We also tried to determine affinities with the use of isothermal titration calorimetry (ITC), but we failed to detect measurable heat changes from a series of titration experiments. Since ITC-based affinities of Ras and Raf proteins have been reported in the past, we reasoned that our poor results were because of the low heat sensitivity of the particular ITC model we used. Another approach will be to determine their affinities indirectly by monitoring the inhibition of GTP dissociation in HRas upon binding of Raf-1 RBD. Following the published work of Herrmann and coworkers (221), we can use a fluorescently labelled and non-hydrolyzable analogue of GTP, Mant-GppNHp. This molecule will allow us to indirectly quantify the affinities of Raf-1 RBD mutants and HRas suppressors towards HRas<sub>WT</sub> and Raf-1<sub>R89L</sub> RBD, respectively. Measurements of fluorescence decay can be measured either by stopped flow kinetics or by steady-state fluorometer (microplate fluorescent reader or spectrofluorometer).

Alternatively, we can also evaluate the effects of these mutations on their relative interactions in mammalian cells. This can be done by transfecting either HeLa or HEK 293 cell lines with HRas-CAAX and Raf-1 RBD expressing pcDNA plasmids. To visualize the interaction by fluorescence microscopy, we can fuse the cytosolic Raf-1 RBD with a FP (preferably an optimized RFP variant to lessen background fluorescence). We anticipate that true interacting protein partners will show a distinct localization pattern of Raf-1 RBD-FP on the plasma membrane. This experiment might only work on interacting partners with higher affinity. One possible solution for imaging of lower affinity pairs might be to stimulate the activation of HRas in live cells using the human epidermal growth factor (EGF) (222). EGF stimulation increases the recruitment capacity of HRas towards its effector proteins, which include Raf-1 RBD. Using this assay, we can monitor any low affinity Raf-1 RBD and HRas variant, and evaluate the effects of integral  $\beta$ -strand mutations on HRas signaling activity.

#### **4.2.2 Characterization of N2P2 and M3P6 peptide inhibitors**

In Chapter 3, we described the screening of 7-mer inhibitory peptides for both N2P2 and M3P6, which have been recently identified as key players in colon cancer. Our strategy led to the identification of novel peptides with selectivity towards their target PDZ domains and, perhaps, higher affinity compared to the control. However, we failed to demonstrate their antagonistic activities through other well-established techniques. Thus, we recommend validation of the results from our strategy by performing *in vitro* measurements on peptides' kinetics,

inhibition, and hemolytic activities. These investigations will certainly not just evaluate the effectiveness of the in-colony screening method but will also gauge the suitability of the identified peptides as potent and selective biologic inhibitors.

One robust technique for measuring protein-peptide binding is fluorescence polarization (207, 223). A polarization assay relies on the principle that when a fluorescent molecule is excited with a polarized light, a light is emitted at the same polarized plane if the fluorophore remained stationary or is slowly tumbling (224). This classical method can provide us a fast, reliable and sensitive assay to measure the inhibition constants of each peptide. To facilitate the implementation of fluorescent polarization assay, we can synthesize and conjugate each peptide with a fluorescent dye at the N-terminal of the peptide. It is also preferred to use a dye that is red-shifted, photostable in most buffer systems, and has longer excited-state lifetime (e.g. BODIPY and cyanine dyes) (225). This will allow a sensitive fluorescent polarization over a larger molecular weight range. We suggest performing this experiment using a stopped flow kinetic instrument. In this way, we can evaluate both their inhibition constants as well as kinetic profiles (association and dissociation rates) simultaneously.

Other experiments that we can execute to fully characterize these peptides may include NMR footprinting of PDZ-peptide complex, MS binding analysis, and cell-based or *in vivo* inhibition studies.

### 4.3 Future applications

In this work, we have demonstrated the suitability of FPX for detecting PPIs and identifying novel inhibitory peptides in the context of *E. coli* colonies. Our results are very encouraging and we are motivated to use this strategy for additional practical applications. Here, we highlight the possibility of using the strategy for detecting the interaction of a less soluble protein pair and screening of other biologic-based inhibitors aside from linear peptides. In addition, we also propose an improvement of our current screening protocol that will allow for efficient screening of larger libraries and increased sensitivity of colony ratiometric measurements.

We attempted to express the extracellular domains of programmed cell death 1 receptor (PD-1) and its ligand programmed cell death ligand 1 (PD-L1) in *E. coli*. Both proteins showed moderate expressions in bacteria when fused with maltose binding protein (MBP) and GFP. However, when cloned in pFPX and co-expressed as ddFP fusions, we observed a very slow maturation of all protein fusions. Fluorescence was also too weak to precisely measure using our digital microscope. This observed downregulated expression or slow maturation could be attributed to the improper folding of extracellular PD-1 and PD-L1 domains. Both receptors contain disulfide bonds at the inner region of their extracellular domains, which we believe are essential for their correct folding and proper conformation. If these disulfide linkages do form, as expected in the highly reducing *E. coli* cytoplasm, it is likely that unfolded and aggregated polypeptides chains will result. To resolve this challenge, we suggest co-expressing ddFP

fusions of extracellular receptor domains with DsbC. DsbC is a protein disulfide isomerase that primarily interacts with misfolded proteins to correct their non-native disulfide linkages (226). Several published works have shown that co-expressing this protein in either *E. coli* cytoplasm or periplasm can effectively produce proteins with correct disulfide bonds (227-229). This protein can be introduced into bacteria in two different ways: reconstruction of pFPX to introduce a fourth expression cassette specifically for DsbC, or preparation of DH10B electrocompetent cells which are pre-transformed with a DsbC expressing p15A bacterial plasmid. If this approach produces folded and soluble fractions enough to be imaged on colonies, we can proceed with generating protein or peptide inhibitors for the PD-1 and PD-L1 interaction.

To the best of our knowledge, all currently available inhibitors against the PD-1 and PD-L1 interaction are monoclonal antibodies. However, optimizing the affinity of mABs using our strategy would be complicated and somewhat impractical due to expression and solubility issues commonly encountered for expression of full antibodies in bacteria. One interesting alternative is to tune the selectivity and affinity of a single-chain variable fragment (ScFv) of either anti PD-1 or PD-L1 mAB. A ScFv is an antibody-derived fusion protein which is generally prepared by fusing the variable domains of light ( $V_L$ ) and heavy ( $V_H$ ) chains of an antibody with a flexible polypeptide linker (230). Although production of active ScFv in *E. coli* is generally achieved in the periplasm (231, 232), an overexpression of DsbC in cytoplasm will surely assist proper folding and formation of disulfide bond necessary for increased activity and yield. We

can also explore the generation of nanobody-based inhibitors. A nanobody is a small polypeptide (12-15 kDa) derived from V<sub>HH</sub> region of a camelid single-chain immunoglobulins (233, 234). Though small, these proteins harness the full binding capacity of their parent antibodies and contain unique structural features beneficial for tailoring its affinity, specificity and stability. Also, it has been reported that an active nanobody can be expressed in *E. coli* at high concentration (235).

Two other classes of biologic inhibitors that we can investigate for disrupting PD-1/PD-L1 association or any desired PPI are cyclic peptide and plant derived cyclotides. The former generally offers better specificity, affinity and uptake than their linear peptide counterparts. Due to its rigid structure, cyclic peptides are also more stable in solution compared to linear peptides, which undergoes rapid equilibration between different conformational states (236, 237). This fast conformation switching yields higher fractions of linear peptide with reduced activity towards its target. It has been reported that cyclic peptides can be effectively biosynthesized in bacteria with the aid of split inteins (238). Split inteins like *Ssp* DnaE and *Npu* DnaE are naturally occurring transplicing proteins that catalyze the posttranslational conjugation of two inactive peptide domains to form a fully functional protein (239-241). In 1999, Scott and coworkers (238) developed SICLOPPS (split intein-mediated circular ligation of peptides and proteins) strategy that tuned a split intein from *Synechocystis sp.* PCC6803 (*Ssp*) to effectively catalyze the intracellular head-to-tail cyclization of essentially any peptides or proteins. In the past ten years, SICLOPPS technology, coupled with

bacterial based reversed two hybrid system (RTHS), has successfully been used to discover cyclic peptide blockers to a number of PPIs (242, 243). We believe that introducing SICLOPPS to our strategy will offer a better screening method and due to fewer false positives than RTHS.

Cyclotides are naturally occurring cyclic polypeptides commonly found in the Violaceae and Rubiaceae plant families (244). Cyclotides contain an unusual cystine knotted motif that gives them remarkable stability to thermal, chemical, and protease degradation (245). Aside from this unique structural feature, they also display some inherent biological activities which include protease inhibition (246), tumor arrest (245), and antiviral activities (247). Thus, it is not surprising that these proteins are increasingly popular templates for targeted drug development (248). Using our in-colony FPX strategy, we could generate and pan a library of cyclotides to potentially inhibit the PD-1/PD-L1 association or other aberrant protein interaction. A library of randomized cyclotides could be generated in *E. coli* either cyclically or linearly. In order for cyclization to occur in bacteria, we could genetically fuse the library between the I<sub>C</sub> and I<sub>N</sub> domains of a split intein. Upon expression, the split intein could catalytically trans-splice and cyclize the cyclotide-encoding polypeptide inserted to it. Meanwhile, production of linear cyclotides can be achieved by introducing a break at its loop 2 region (249, 250). It was previously reported that, relative to other loops in the protein, loop 2 is more amenable to introduction of breaks without destroying the overall conformation of the protein. We can also fuse a ddFP on the C-terminal of the linearized cyclotide to boost our strategy's sensitivity to the relative affinity of

each variant. We are optimistic that this fusion protein will work in the cytoplasm of *E. coli*, especially when co-expressed with DsbC.

Together with all these exciting applications, we also aim for our strategy to compete with other high-throughput methods in screening larger library sizes. Our current protocol uses the custom-built digital microscope for imaging and measuring fluorescence on colonies. Although the system gives us good sensitivity, it limits the library size that we are able to access to approximately  $10^4$ . To overcome this hurdle, we plan to couple our in-colony strategy with a fluorescence activated cell sorting (FACS) instrument. FACS is well suited for screening very large libraries in *E. coli* cells. It was successfully implemented in the screening of PPIs, optimized protein-based inhibitors and improved FP variants (251-253). Integrating FACS with our in-colony strategy will permit us to maximize the number randomized residues without compromising sensitivity and library coverage.

## 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. Compar. Physl.* **1962**, *59*, 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*, 802-805.
3. Yang, T. T.; Cheng, L.; Kain, S. R. Optimized codon usage and chromophore mutations provide enhanced sensitivity with the green fluorescent protein. *Nucleic Acids Res.* **1996**, *24*, 4592-4593.
4. Ai, H.; 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 (N. Y.)* **2007**, *46*, 5904-5910.
5. Ormo, 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*, 1392-1395.
6. 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*, 969-973.

7. 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*, 1567-1572.
8. Wang, L.; Jackson, W. C.; Steinbach, P. A.; Tsien, R. Y. Evolution of new nonantibody proteins via iterative somatic hypermutation. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 16745-16749.
9. Morozova, K. S.; Piatkevich, K. D.; Gould, T. J.; Zhang, J.; Bewersdorf, J.; Verkhusha, V. V. Far-red fluorescent protein excitable with red lasers for flow cytometry and superresolution STED nanoscopy. *Biophys. J.* **2010**, *99*, L13-L15.
10. Lin, M. Z.; McKeown, M. R.; Ng, H.; 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*, 1169-1179.
11. Shcherbo, D.; Merzlyak, E. M.; Chepurnykh, T. V.; Fradkov, A. F.; Ermakova, G. V.; Solovieva, E. A.; Lukyanov, K. A.; Bogdanova, E. A.; Zaraisky, A. G.; Lukyanov, S. Bright far-red fluorescent protein for whole-body imaging. *Nat. Methods* **2007**, *4*, 741-746.

12. Resch-Genger, U.; Grabolle, M.; Cavaliere-Jaricot, S.; Nitschke, R.; Nann, T. Quantum dots versus organic dyes as fluorescent labels. *Nat. Methods* **2008**, *5*, 763-775.
13. Mérian, J.; Gravier, J.; Navarro, F.; Texier, I. Fluorescent nanoprobe dedicated to in vivo imaging: from preclinical validations to clinical translation. *Molecules* **2012**, *17*, 5564-5591.
14. Chudakov, D. M.; Matz, M. V.; Lukyanov, S.; Lukyanov, K. A. Fluorescent proteins and their applications in imaging living cells and tissues. *Physiol. Rev.* **2010**, *90*, 1103-1163.
15. Ibraheem, A.; Campbell, R. E. Designs and applications of fluorescent protein-based biosensors. *Curr. Opin. Chem. Biol.* **2010**, *14*, 30-36.
16. Lindenburg, L.; Merkx, M. Engineering genetically encoded FRET sensors. *Sensors* **2014**, *14*, 11691-11713.
17. Sakaue-Sawano, A.; Kurokawa, H.; Morimura, T.; Hanyu, A.; Hama, H.; Osawa, H.; Kashiwagi, S.; Fukami, K.; Miyata, T.; Miyoshi, H. Visualizing spatiotemporal dynamics of multicellular cell-cycle progression. *Cell* **2008**, *132*, 487-498.
18. Aiba, Y.; Kometani, K.; Hamadate, M.; Moriyama, S.; Sakaue-Sawano, A.; Tomura, M.; Luche, H.; Fehling, H. J.; Casellas, R.; Kanagawa, O.; Miyawaki, A.; Kurosaki, T. Preferential localization of IgG memory B cells

- adjacent to contracted germinal centers. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 12192-12197.
19. Sakaue-Sawano, A.; Hoshida, T.; Yo, M.; Takahashi, R.; Ohtawa, K.; Arai, T.; Takahashi, E.; Noda, S.; Miyoshi, H.; Miyawaki, A. Visualizing developmentally programmed endoreplication in mammals using ubiquitin oscillators. *Development* **2013**, *140*, 4624-4632.
  20. Abe, T.; Sakaue-Sawano, A.; Kiyonari, H.; Shioi, G.; Inoue, K.; Horiuchi, T.; Nakao, K.; Miyawaki, A.; Aizawa, S.; Fujimori, T. Visualization of cell cycle in mouse embryos with Fucci2 reporter directed by Rosa26 promoter. *Development* **2013**, *140*, 237-246.
  21. Zielke, N.; Korzelius, J.; van Straaten, M.; Bender, K.; Schuhknecht, G. F.; Dutta, D.; Xiang, J.; Edgar, B. A. Fly-FUCCI: a versatile tool for studying cell proliferation in complex tissues. *Cell Rep.* **2014**, *7*, 588-598.
  22. Livet, J.; Weissman, T. A.; Kang, H.; Draft, R. W.; Lu, J.; Bennis, R. A.; Sanes, J. R.; Lichtman, J. W. Transgenic strategies for combinatorial expression of fluorescent proteins in the nervous system. *Nature* **2007**, *450*, 56-62.
  23. Cai, D.; Cohen, K. B.; Luo, T.; Lichtman, J. W.; Sanes, J. R. Improved tools for the Brainbow toolbox. *Nat. Methods* **2013**, *10*, 540-547.

24. Hampel, S.; Chung, P.; McKellar, C. E.; Hall, D.; Looger, L. L.; Simpson, J. H. *Drosophila* Brainbow: a recombinase-based fluorescence labeling technique to subdivide neural expression patterns. *Nat. Methods* **2011**, *8*, 253-259.
25. Pan, Y. A.; Freundlich, T.; Weissman, T. A.; Schoppik, D.; Wang, X. C.; Zimmerman, S.; Ciruna, B.; Sanes, J. R.; Lichtman, J. W.; Schier, A. F. Zebrawow: multispectral cell labeling for cell tracing and lineage analysis in zebrafish. *Development* **2013**, *140*, 2835-2846.
26. Frangioni, J. V. In vivo near-infrared fluorescence imaging. *Curr. Opin. Chem. Biol.* **2003**, *7*, 626-634.
27. Weissleder, R. A clearer vision for in vivo imaging. *Nat. Biotechnol.* **2001**, *19*, 316-316.
28. Stamatas, G. N.; Southall, M.; Kollias, N. In vivo monitoring of cutaneous edema using spectral imaging in the visible and near infrared. *J. Invest. Dermatol.* **2006**, *126*, 1753-1760.
29. Tromberg, B. J.; Shah, N.; Lanning, R.; Cerussi, A.; Espinoza, J.; Pham, T.; Svaasand, L.; Butler, J. Non-invasive in vivo characterization of breast tumors using photon migration spectroscopy. *Neoplasia* **2000**, *2*, 26-40.

30. Strack, R. L.; Hein, B.; Bhattacharyya, D.; Hell, S. W.; Keenan, R. J.; Glick, B. S. A rapidly maturing far-red derivative of DsRed-Express2 for whole-cell labeling. *Biochemistry (N. Y.)* **2009**, *48*, 8279-8281.
31. Chica, R. A.; Moore, M. M.; Allen, B. D.; Mayo, S. L. Generation of longer emission wavelength red fluorescent proteins using computationally designed libraries. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 20257-20262.
32. Shcherbo, D.; Shemiakina, I. I.; Ryabova, A. V.; Luker, K. E.; Schmidt, B. T.; Souslova, E. A.; Gorodnicheva, T. V.; Strukova, L.; Shidlovskiy, K. M.; Britanova, O. V. Near-infrared fluorescent proteins. *Nat. Methods* **2010**, *7*, 827-829.
33. Chu, J.; Haynes, R. D.; Corbel, S. Y.; Li, P.; González-González, E.; Burg, J. S.; Ataie, N. J.; Lam, A. J.; Cranfill, P. J.; Baird, M. A. Non-invasive intravital imaging of cellular differentiation with a bright red-excitable fluorescent protein. *Nat. Methods* **2014**, *11*, 572-578.
34. Filonov, G. S.; Piatkevich, K. D.; Ting, L.; Zhang, J.; Kim, K.; Verkhusha, V. V. Bright and stable near-infrared fluorescent protein for in vivo imaging. *Nat. Biotechnol.* **2011**, *29*, 757-761.
35. Zhao, Y.; Araki, S.; Wu, J.; Teramoto, T.; Chang, Y. F.; Nakano, M.; Abdelfattah, A. S.; Fujiwara, M.; Ishihara, T.; Nagai, T.; Campbell, R. E. An

- expanded palette of genetically encoded  $\text{Ca}^{2+}$  indicators. *Science* **2011**, 333, 1888-1891.
36. Tewson, P.; Westenberg, M.; Zhao, Y.; Campbell, R. E.; Quinn, A. M.; Hughes, T. E. Simultaneous detection of  $\text{Ca}^{2+}$  and diacylglycerol signaling in living cells. *PloS one* **2012**, 7, e42791.
37. Mochizuki, N.; Yamashita, S.; Kurokawa, K.; Ohba, Y.; Nagai, T.; Miyawaki, A.; Matsuda, M. Spatio-temporal images of growth-factor-induced activation of Ras and Rap1. *Nature* **2001**, 411, 1065-1068.
38. Imamura, H.; Nhat, K. P.; Togawa, H.; Saito, K.; Iino, R.; Kato-Yamada, Y.; Nagai, T.; Noji, H. Visualization of ATP levels inside single living cells with fluorescence resonance energy transfer-based genetically encoded indicators. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, 106, 15651-15656.
39. Ai, H.; Hazelwood, K. L.; Davidson, M. W.; Campbell, R. E. Fluorescent protein FRET pairs for ratiometric imaging of dual biosensors. *Nat. Methods* **2008**, 5, 401-403.
40. 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, 4647-4656.

41. Zaccolo, M.; De Giorgi, F.; Cho, C. Y.; Feng, L.; Knapp, T.; Negulescu, P. A.; Taylor, S. S.; Tsien, R. Y.; Pozzan, T. A genetically encoded, fluorescent indicator for cyclic AMP in living cells. *Nat. Cell Biol.* **1999**, *2*, 25-29.
42. Ganesan, S.; Ameer-Beg, S. M.; Ng, T. T.; Vojnovic, B.; Wouters, F. S. A dark yellow fluorescent protein (YFP)-based Resonance Energy-Accepting Chromoprotein (REACH) for Förster resonance energy transfer with GFP. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 4089-4094.
43. Wu, J.; Abdelfattah, A. S.; Miraucourt, L. S.; Kutsarova, E.; Ruangkittisakul, A.; Zhou, H.; Ballanyi, K.; Wicks, G.; Drobizhev, M.; Rebane, A. A long Stokes shift red fluorescent Ca<sup>2+</sup> indicator protein for two-photon and ratiometric imaging. *Nat. Commun.* **2014**, *5*.
44. Förster, T. Zwischenmolekulare energiewanderung und fluoreszenz. *Annalen der physik* **1948**, *437*, 55-75.
45. Campbell, R. E. Fluorescent-protein-based biosensors: modulation of energy transfer as a design principle. *Anal. Chem.* **2009**, *81*, 5972-5979.
46. Nguyen, A. W.; Daugherty, P. S. Evolutionary optimization of fluorescent proteins for intracellular FRET. *Nat. Biotechnol.* **2005**, *23*, 355-360.
47. Vinkenborg, J. L.; Evers, T. H.; Reulen, S. W.; Meijer, E.; Merkx, M. Enhanced Sensitivity of FRET-Based Protease Sensors by Redesign of the GFP Dimerization Interface. *ChemBioChem* **2007**, *8*, 1119-1121.

48. Lam, A. J.; St-Pierre, F.; Gong, Y.; Marshall, J. D.; Cranfill, P. J.; Baird, M. A.; McKeown, M. R.; Wiedenmann, J.; Davidson, M. W.; Schnitzer, M. J. Improving FRET dynamic range with bright green and red fluorescent proteins. *Nat. Methods* **2012**, *9*, 1005-1012.
49. Ouyang, M.; Huang, H.; Shaner, N. C.; Remacle, A. G.; Shiryaev, S. A.; Strongin, A. Y.; Tsien, R. Y.; Wang, Y. Simultaneous visualization of protumorigenic Src and MT1-MMP activities with fluorescence resonance energy transfer. *Cancer Res.* **2010**, *70*, 2204-2212.
50. Marcaggi, P.; Mutoh, H.; Dimitrov, D.; Beato, M.; Knöpfel, T. Optical measurement of mGluR1 conformational changes reveals fast activation, slow deactivation, and sensitization. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 11388-11393.
51. Łukasiewicz, S.; Faron-Górecka, A.; Dobrucki, J.; Polit, A.; Dziejzicka-Wasylewska, M. Studies on the role of the receptor protein motifs possibly involved in electrostatic interactions on the dopamine D1 and D2 receptor oligomerization. *FEBS journal* **2009**, *276*, 760-775.
52. Lundby, A.; Mutoh, H.; Dimitrov, D.; Akemann, W.; Knöpfel, T. Engineering of a genetically encodable fluorescent voltage sensor exploiting fast Ci-VSP voltage-sensing movements. *PloS one* **2008**, *3*, e2514.

53. Miyawaki, A.; Llopis, J.; Heim, R.; McCaffery, J. M.; Adams, J. A.; Ikura, M.; Tsien, R. Y. Fluorescent indicators for  $\text{Ca}^{2+}$  & plus; based on green fluorescent proteins and calmodulin. *Nature* **1997**, *388*, 882-887.
54. Miesenböck, G.; De Angelis, D. A.; Rothman, J. E. Visualizing secretion and synaptic transmission with pH-sensitive green fluorescent proteins. *Nature* **1998**, *394*, 192-195.
55. Galiotta, L. J.; Haggie, P. M.; Verkman, A. Green fluorescent protein-based halide indicators with improved chloride and iodide affinities. *FEBS Lett.* **2001**, *499*, 220-224.
56. Hanson, G. T.; Aggeler, R.; Oglesbee, D.; Cannon, M.; Capaldi, R. A.; Tsien, R. Y.; Remington, S. J. Investigating mitochondrial redox potential with redox-sensitive green fluorescent protein indicators. *J. Biol. Chem.* **2004**, *279*, 13044-13053.
57. Sankaranarayanan, S.; De Angelis, D.; Rothman, J. E.; Ryan, T. A. The use of pHluorins for optical measurements of presynaptic activity. *Biophys. J.* **2000**, *79*, 2199-2208.
58. Shen, Y.; Rosendale, M.; Campbell, R. E.; Perrais, D. pHuji, a pH-sensitive red fluorescent protein for imaging of exo- and endocytosis. *J. Cell Biol.* **2014**, *207*, 419-432.

59. Akerboom, J.; Chen, T. W.; Wardill, T. J.; Tian, L.; Marvin, J. S.; Mutlu, S.; Calderon, N. C.; Esposti, F.; Borghuis, B. G.; Sun, X. R.; Gordus, A.; Orger, M. B.; Portugues, R.; Engert, F.; Macklin, J. J.; Filosa, A.; Aggarwal, A.; Kerr, R. A.; Takagi, R.; Kracun, S.; Shigetomi, E.; Khakh, B. S.; Baier, H.; Lagnado, L.; Wang, S. S.; Bargmann, C. I.; Kimmel, B. E.; Jayaraman, V.; Svoboda, K.; Kim, D. S.; Schreier, E. R.; Looger, L. L. Optimization of a GCaMP calcium indicator for neural activity imaging. *J. Neurosci.* **2012**, *32*, 13819-13840.
60. Zhao, Y.; Araki, S.; Wu, J.; Teramoto, T.; Chang, Y. F.; Nakano, M.; Abdelfattah, A. S.; Fujiwara, M.; Ishihara, T.; Nagai, T.; Campbell, R. E. An expanded palette of genetically encoded Ca<sup>2+</sup> indicators. *Science* **2011**, *333*, 1888-1891.
61. Jin, L.; Han, Z.; Platisa, J.; Wooltorton, J. R.; Cohen, L. B.; Pieribone, V. A. Single action potentials and subthreshold electrical events imaged in neurons with a fluorescent protein voltage probe. *Neuron* **2012**, *75*, 779-785.
62. St-Pierre, F.; Marshall, J. D.; Yang, Y.; Gong, Y.; Schnitzer, M. J.; Lin, M. Z. High-fidelity optical reporting of neuronal electrical activity with an ultrafast fluorescent voltage sensor. *Nat. Neurosci.* **2014**, *17*, 884-889.
63. 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*, 5658-5659.

64. Blakeley, B. D.; Chapman, A. M.; McNaughton, B. R. Split-superpositive GFP reassembly is a fast, efficient, and robust method for detecting protein–protein interactions in vivo. *Mol. BioSyst.* **2012**, *8*, 2036-2040.
65. Hu, C.; Kerppola, T. K. Simultaneous visualization of multiple protein interactions in living cells using multicolor fluorescence complementation analysis. *Nat. Biotechnol.* **2003**, *21*, 539-545.
66. Kerppola, T. K. Bimolecular fluorescence complementation (BiFC) analysis as a probe of protein interactions in living cells. *Annu. Rev. Biophys.* **2008**, *37*, 465-487.
67. Lindman, S.; Hernandez-Garcia, A.; Szczepankiewicz, O.; Frohm, B.; Linse, S. In vivo protein stabilization based on fragment complementation and a split GFP system. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 19826-19831.
68. Sachdeva, G.; Garg, A.; Godding, D.; Way, J. C.; Silver, P. A. In vivo co-localization of enzymes on RNA scaffolds increases metabolic production in a geometrically dependent manner. *Nucleic Acids Res.* **2014**, *42*, 9493-9503.
69. Kodama, Y.; Hu, C. Bimolecular fluorescence complementation (BiFC): A 5-year update and future perspectives. *BioTechniques* **2012**, *53*, 285-298.
70. Horstman, A.; Tonaco, I. A. N.; Boutilier, K.; Immink, R. G. A cautionary note on the use of split-YFP/BiFC in plant protein-protein interaction studies. *Int. J Mol. Sci.* **2014**, *15*, 9628-9643.

71. Tai, H.; Schuman, E. M. Ubiquitin, the proteasome and protein degradation in neuronal function and dysfunction. *Nat. Rev. Neurosci.* **2008**, *9*, 826-838.
72. Sambamurti, K.; Suram, A.; Venugopal, C.; Prakasam, A.; Zhou, Y.; Lahiri, D. K.; Greig, N. H. A partial failure of membrane protein turnover may cause Alzheimer's disease: a new hypothesis. *Curr. Alzheimer Res.* **2006**, *3*, 81-90.
73. Terskikh, A.; Fradkov, A.; Ermakova, G.; Zaraisky, A.; Tan, P.; Kajava, A. V.; Zhao, X.; Lukyanov, S.; Matz, M.; Kim, S.; Weissman, I.; Siebert, P. "Fluorescent timer": protein that changes color with time. *Science* **2000**, *290*, 1585-1588.
74. Duncan, R. R.; Greaves, J.; Wiegand, U. K.; Matskevich, I.; Bodammer, G.; Apps, D. K.; Shipston, M. J.; Chow, R. H. Functional and spatial segregation of secretory vesicle pools according to vesicle age. *Nature* **2003**, *422*, 176-180.
75. Mirabella, R.; Franken, C.; van der Krogt, G. N.; Bisseling, T.; Geurts, R. Use of the fluorescent timer DsRED-E5 as reporter to monitor dynamics of gene activity in plants. *Plant Physiol.* **2004**, *135*, 1879-1887.
76. Verkhusha, V. V.; Otsuna, H.; Awasaki, T.; Oda, H.; Tsukita, S.; Ito, K. An enhanced mutant of red fluorescent protein DsRed for double labeling and developmental timer of neural fiber bundle formation. *J. Biol. Chem.* **2001**, *276*, 29621-29624.

77. 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*, 118-126.
78. Khmelinskii, A.; Keller, P. J.; Bartosik, A.; Meurer, M.; Barry, J. D.; Mardin, B. R.; Kaufmann, A.; Trautmann, S.; Wachsmuth, M.; Pereira, G.; Huber, W.; Schiebel, E.; Knop, M. Tandem fluorescent protein timers for in vivo analysis of protein dynamics. *Nat. Biotechnol.* **2012**, *30*, 708-714.
79. Lin, M. Z.; Glenn, J. S.; Tsien, R. Y. A drug-controllable tag for visualizing newly synthesized proteins in cells and whole animals. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 7744-7749.
80. 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*, 1742-1751.
81. Terai, T.; Nagano, T. Fluorescent probes for bioimaging applications. *Curr. Opin. Chem. Biol.* **2008**, *12*, 515-521.
82. Chen, S.; Chen, Z.; Ren, W.; Ai, H. Reaction-based genetically encoded fluorescent hydrogen sulfide sensors. *J. Am. Chem. Soc.* **2012**, *134*, 9589-9592.

83. Gadalla, M. M.; Snyder, S. H. Hydrogen sulfide as a gasotransmitter. *J. Neurochem.* **2010**, *113*, 14-26.
84. Ryu, Y.; Schultz, P. G. Efficient incorporation of unnatural amino acids into proteins in *Escherichia coli*. *Nat. Methods* **2006**, *3*, 263-265.
85. Wang, L.; Xie, J.; Schultz, P. G. Expanding the genetic code. *Annu. Rev. Biophys. Biomol. Struct.* **2006**, *35*, 225-249.
86. Chen, Z.; Ren, W.; Wright, Q. E.; Ai, H. Genetically encoded fluorescent probe for the selective detection of peroxynitrite. *J. Am. Chem. Soc.* **2013**, *135*, 14940-14943.
87. Chen, Z.; Ai, H. A highly responsive and selective fluorescent probe for imaging physiological hydrogen sulfide. *Biochemistry (N. Y.)* **2014**, *53*, 5966-5974.
88. Cabantous, S.; Nguyen, H. B.; Pedelacq, J.; 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*.
89. Kent, K. P.; Boxer, S. G. Light-activated reassembly of split green fluorescent protein. *J. Am. Chem. Soc.* **2011**, *133*, 4046-4052.

90. Do, K.; Boxer, S. G. Thermodynamics, kinetics, and photochemistry of  $\beta$ -strand association and dissociation in a split-GFP System. *J. Am. Chem. Soc.* **2011**, *133*, 18078-18081.
91. Do, K.; Boxer, S. G. GFP variants with alternative  $\beta$ -strands and their application as light-driven protease sensors: A Tale of Two Tails. *J. Am. Chem. Soc.* **2013**, *135*, 10226-10229.
92. Culler, S. J.; Hoff, K. G.; Smolke, C. D. Reprogramming cellular behavior with RNA controllers responsive to endogenous proteins. *Science* **2010**, *330*, 1251-1255.
93. Tischer, D.; Weiner, O. D. Illuminating cell signalling with optogenetic tools. *Nat. Rev. Mol. Cell Biol.* **2014**, *15*, 551-558.
94. Pathak, G. P.; Vrana, J. D.; Tucker, C. L. Optogenetic control of cell function using engineered photoreceptors. *Biol. Cell* **2013**, *105*, 59-72.
95. Zhou, X. X.; Chung, H. K.; Lam, A. J.; Lin, M. Z. Optical control of protein activity by fluorescent protein domains. *Science* **2012**, *338*, 810-814.
96. Tang, J. C.; Szikra, T.; Kozorovitskiy, Y.; Teixeira, M.; Sabatini, B. L.; Roska, B.; Cepko, C. L. A nanobody-based system using fluorescent proteins as scaffolds for cell-specific gene manipulation. *Cell* **2013**, *154*, 928-939.

97. Alford, S. C.; Abdelfattah, A. S.; Ding, Y.; Campbell, R. E. A fluorogenic red fluorescent protein heterodimer. *Chem. Biol.* **2012**, *19*, 353-360.
98. Alford, S. C.; Ding, Y.; Simmen, T.; Campbell, R. E. Dimerization-dependent green and yellow fluorescent proteins. *ACS Synth. Biol.* **2012**, *1*, 569-575.
99. Ding, Y.; Li, J.; Enterina, J. R.; Shen, Y.; Zhang, I.; Tewson, P. H.; Mo, G. C. H.; Zhang, J.; Quinn, A. M.; Hughes, T. E.; Maysinger, D.; Alford, S. C.; Zhang, Y.; Campbell, R. E. Ratiometric biosensors based on dimerization-dependent fluorescent protein exchange. *Nat. Methods* **2015**, *12*, 195-198.
100. Nero, T. L.; Morton, C. J.; Holien, J. K.; Wielens, J.; Parker, M. W. Oncogenic protein interfaces: small molecules, big challenges. *Nat. Rev. Cancer* **2014**, *14*, 248-262.
101. Jones, S.; Thornton, J. M. Principles of protein-protein interactions. *Proc. Natl. Acad. Sci. U. S. A.* **1996**, *93*, 13-20.
102. Ivanov, A. A.; Khuri, F. R.; Fu, H. Targeting protein-protein interactions as an anticancer strategy. *Trends Pharmacol. Sci.* **2013**, *34*, 393-400.
103. DiGiulio, S. Keytruda Approved by FDA for Advanced Melanoma, Called 'Game Changer'. *Oncology Times* **2014**.

104. Powles, T.; Eder, J. P.; Fine, G. D.; Braiteh, F. S.; Loria, Y.; Cruz, C.; Bellmunt, J.; Burris, H. A.; Petrylak, D. P.; Teng, S. MPDL3280A (anti-PD-L1) treatment leads to clinical activity in metastatic bladder cancer. *Nature* **2014**, *515*, 558-562.
105. Naidoo, J.; Page, D.; Wolchok, J. Immune modulation for cancer therapy. *Br. J. Cancer* **2014**, *111*, 2214–2219.
106. Okazaki, T.; Honjo, T. PD-1 and PD-1 ligands: from discovery to clinical application. *Int. Immunol.* **2007**, *19*, 813-824.
107. Velcheti, V.; Schalper, K. A.; Carvajal, D. E.; Anagnostou, V. K.; Syrigos, K. N.; Sznol, M.; Herbst, R. S.; Gettinger, S. N.; Chen, L.; Rimm, D. L. Programmed death ligand-1 expression in non-small cell lung cancer. *Lab. Invest.* **2013**, *94*, 107-116.
108. Keir, M. E.; Butte, M. J.; Freeman, G. J.; Sharpe, A. H. PD-1 and its ligands in tolerance and immunity. *Annu. Rev. Immunol.* **2008**, *26*, 677-704
109. Topalian, S. L.; Hodi, F. S.; Brahmer, J. R.; Gettinger, S. N.; Smith, D. C.; McDermott, D. F.; Powderly, J. D.; Carvajal, R. D.; Sosman, J. A.; Atkins, M. B. Safety, activity, and immune correlates of anti-PD-1 antibody in cancer. *N. Engl. J. Med.* **2012**, *366*, 2443-2454.
110. Philips, G. K.; Atkins, M. Therapeutic uses of anti-PD-1 and anti-PD-L1 antibodies. *Int. Immunol.* **2015**, *27*, 39-46.

111. Lin, D. Y.; Tanaka, Y.; Iwasaki, M.; Gittis, A. G.; Su, H. P.; Mikami, B.; Okazaki, T.; Honjo, T.; Minato, N.; Garboczi, D. N. The PD-1/PD-L1 complex resembles the antigen-binding Fv domains of antibodies and T cell receptors. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 3011-3016.
112. Sasikumar, P. G.; Ramachandra, M.; Vadlamani, S. K.; Vemula, K. R.; Satyam, L. K.; Subbarao, K.; Shrimali, K. R.; Kandepu, S. *Immunosuppression modulating compounds* **2011**. U.S. Patent Application 13/168,453
113. Mann, M.; Hendrickson, R. C.; Pandey, A. Analysis of proteins and proteomes by mass spectrometry. *Annu. Rev. Biochem.* **2001**, *70*, 437-473.
114. Hamdi, A.; Colas, P. Yeast two-hybrid methods and their applications in drug discovery. *Trends Pharmacol. Sci.* **2012**, *33*, 109-118.
115. Sambrook, J.; Fritsch, E. F.; Maniatis, T. *Molecular cloning*; Cold Spring Harbor Laboratory Press New York: 1989; Vol. 2.
116. Barreto, K.; Bharathikumar, V.; Ricardo, A.; DeCoteau, J. F.; Luo, Y.; Geyer, C. R. A genetic screen for isolating “lariat” peptide inhibitors of protein function. *Chem. Biol.* **2009**, *16*, 1148-1157.
117. Fields, S.; Song, O. A novel genetic system to detect protein protein interactions. *Nature* **1989**, *340*, 245 – 246.

118. Lievens, S.; Lemmens, I.; Tavernier, J. Mammalian two-hybrids come of age. *Trends Biochem. Sci.* **2009**, *34*, 579-588.
119. Yamazaki, M.; Zhang, Y.; Watanabe, H.; Yokozeki, T.; Ohno, S.; Kaibuchi, K.; Shibata, H.; Mukai, H.; Ono, Y.; Frohman, M. A.; Kanaho, Y. Interaction of the small G protein RhoA with the C terminus of human phospholipase D1. *J. Biol. Chem.* **1999**, *274*, 6035-6038.
120. Lalonde, S.; Sero, A.; Pratelli, R.; Pilot, G.; Chen, J.; Sardi, M. I.; Parsa, S. A.; Kim, D. Y.; Acharya, B. R.; Stein, E. V.; Hu, H. C.; Villiers, F.; Takeda, K.; Yang, Y.; Han, Y. S.; Schwacke, R.; Chiang, W.; Kato, N.; Loque, D.; Assmann, S. M.; Kwak, J. M.; Schroeder, J. I.; Rhee, S. Y.; Frommer, W. B. A membrane protein/signaling protein interaction network for Arabidopsis version AMPv2. *Front. Physiol.* **2010**, *1*, 24.
121. Baneyx, F. Recombinant protein expression in Escherichia coli. *Curr. Opin. Biotechnol.* **1999**, *10*, 411-421.
122. Tolia, N. H.; Joshua-Tor, L. Strategies for protein coexpression in Escherichia coli. *Nat. Methods* **2006**, *3*, 55-64.
123. Rosano, G. L.; Ceccarelli, E. A. Recombinant protein expression in Escherichia coli: advances and challenges. *Front. Microbiol.* **2014**, *5*, 172.

124. Liu, Z.; Olejniczak, E. T.; Fesik, S. W. Over-expression of the human MDM2 p53 binding domain by fusion to a p53 transactivation peptide. *Protein Expr. Purif.* **2004**, *37*, 493-498.
125. McNally, E. M.; Goodwin, E. B.; Spudich, J. A.; Leinwand, L. A. Coexpression and assembly of myosin heavy chain and myosin light chain in *Escherichia coli*. *Proc. Natl. Acad. Sci. U. S. A.* **1988**, *85*, 7270-7273.
126. Henricksen, L. A.; Umbricht, C. B.; Wold, M. S. Recombinant replication protein A: expression, complex formation, and functional characterization. *J. Biol. Chem.* **1994**, *269*, 11121-11132.
127. Li, C.; Schwabe, J. W.; Banayo, E.; Evans, R. M. Coexpression of nuclear receptor partners increases their solubility and biological activities. *Proc. Natl. Acad. Sci. U. S. A.* **1997**, *94*, 2278-2283.
128. Tan, S. A modular polycistronic expression system for overexpressing protein complexes in *Escherichia coli*. *Protein Expr. Purif.* **2001**, *21*, 224-234.
129. Tan, S.; Kern, R. C.; Selleck, W. The pST44 polycistronic expression system for producing protein complexes in *Escherichia coli*. *Protein Expr. Purif.* **2005**, *40*, 385-395.

130. Ringquist, S.; Shinedling, S.; Barrick, D.; Green, L.; Binkley, J.; Stormo, G. D.; Gold, L. Translation initiation in *Escherichia coli*: sequences within the ribosome-binding site. *Mol. Microbiol.* **1992**, *6*, 1219-1229.
131. Vimberg, V.; Tats, A.; Remm, M.; Tenson, T. Translation initiation region sequence preferences in *Escherichia coli*. *BMC Mol. Biol.* **2007**, *8*, 100.
132. Olins, P. O.; Rangwala, S. H. A novel sequence element derived from bacteriophage T7 mRNA acts as an enhancer of translation of the *lacZ* gene in *Escherichia coli*. *J. Biol. Chem.* **1989**, *264*, 16973-16976.
133. O'Connor, M.; Dahlberg, A. E. Enhancement of translation by the epsilon element is independent of the sequence of the 460 region of 16S rRNA. *Nucleic Acids Res.* **2001**, *29*, 1420-1425.
134. Schmidt, T. G.; Skerra, A. The Strep-tag system for one-step purification and high-affinity detection or capturing of proteins. *Nat. Protoc.* **2007**, *2*, 1528-1535.
135. Crowe, J.; Henco, K. The QIAexpressionist. *Qiagen Inc, Chatsworth, CA, USA* **1992**.
136. Shaner, N. C.; Steinbach, P. A.; Tsien, R. Y. A guide to choosing fluorescent proteins. *Nat. Methods* **2005**, *2*, 905-909.

137. Hoi, H.; Howe, E. S.; Ding, Y.; Zhang, W.; Baird, M. A.; Sell, B. R.; Allen, J. R.; Davidson, M. W.; Campbell, R. E. An engineered monomeric Zoanthus sp. yellow fluorescent protein. *Chem. Biol.* **2013**, *20*, 1296-1304.
138. 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. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 7877-7882.
139. Gerber, D. E. Targeted therapies: a new generation of cancer treatments. *Am. Fam. Physician* **2008**, *77*, 311-319.
140. Hamburg, M. A.; Collins, F. S. The path to personalized medicine. *N. Engl. J. Med.* **2010**, *363*, 301-304.
141. Hanash, S. M.; Baik, C. S.; Kallioniemi, O. Emerging molecular biomarkers—blood-based strategies to detect and monitor cancer. *Nat. Rev. Clin. Oncol.* **2011**, *8*, 142-150.
142. Hayes, D. F.; Markus, H. S.; Leslie, R. D.; Topol, E. J. Personalized medicine: risk prediction, targeted therapies and mobile health technology. *BMC medicine* **2014**, *12*, 37.
143. Sechler, M.; Cizmiciu, A. D.; Avasarala, S.; Van Scoyk, M.; Brzezinski, C.; Kelley, N.; Bikkavilli, R. K.; Winn, R. A. Non-small-cell lung cancer: molecular targeted therapy and personalized medicine—drug resistance, mechanisms, and strategies. *Pharmacogenomics Pers. Med.* **2013**, *6*, 25.

144. van't Veer, L. J.; Bernards, R. Enabling personalized cancer medicine through analysis of gene-expression patterns. *Nature* **2008**, *452*, 564-570.
145. Innamorati, G.; Valenti, M. T.; Giovinazzo, F.; Dalle Carbonare, L.; Parenti, M.; Bassi, C. Molecular approaches to target GPCRs in cancer therapy. *Pharmaceuticals* **2011**, *4*, 567-589.
146. Johnson, F. M.; Gallick, G. E. SRC family nonreceptor tyrosine kinases as molecular targets for cancer therapy. *Anticancer Agents Med. Chem.* **2007**, *7*, 651-659.
147. Smrcka, A. V. Molecular targeting of G $\alpha$  and G $\beta\gamma$  subunits: a potential approach for cancer therapeutics. *Trends Pharmacol. Sci.* **2013**, *34*, 290-298.
148. Tamura, K.; Fukuoka, M. Molecular target-based cancer therapy: tyrosine kinase inhibitors. *Int. J. Clin. Oncol.* **2003**, *8*, 207-211.
149. DiGiulio, S. Keytruda Approved by FDA for Advanced Melanoma, Called 'Game Changer'. *Oncology Times* **2014**.
150. Ivanov, A. A.; Khuri, F. R.; Fu, H. Targeting protein–protein interactions as an anticancer strategy. *Trends Pharmacol. Sci.* **2013**, *34*, 393-400.
151. Scott, A. M.; Wolchok, J. D.; Old, L. J. Antibody therapy of cancer. *Nat. Rev. Cancer* **2012**, *12*, 278-287.

152. US Food and Drug Administration. Hematology/oncology (cancer) approvals & safety notifications: 2014.<http://www.fda.gov/Drugs/InformationOnDrugs/ApprovedDrugs/ucm279174.htm>. (Accessed March 10, 2015).
153. Chen, J.; Sawyer, N.; Regan, L. Protein–protein interactions: General trends in the relationship between binding affinity and interfacial buried surface area. *Protein Sci.* **2013**, *22*, 510-515.
154. Hollenstein, K.; de Graaf, C.; Bortolato, A.; Wang, M.; Marshall, F. H.; Stevens, R. C. Insights into the structure of class B GPCRs. *Trends Pharmacol. Sci.* **2014**, *35*, 12-22.
155. Vu, B.; Wovkulich, P.; Pizzolato, G.; Lovey, A.; Ding, Q.; Jiang, N.; Liu, J.; Zhao, C.; Glenn, K.; Wen, Y. Discovery of RG7112: a small-molecule MDM2 inhibitor in clinical development. *ACS Med. Chem. Lett.* **2013**, *4*, 466-469.
156. Moll, U. M.; Petrenko, O. The MDM2-p53 interaction. *Mol. Cancer. Res.* **2003**, *1*, 1001-1008.
157. Popowicz, G. M.; Czarna, A.; Rothweiler, U.; Szwagierczak, A.; Krajewski, M.; Weber, L.; Holak, T. A. Molecular basis for the inhibition of p53 by Mdmx. *Cell Cycle* **2007**, *6*, 2386-2392.
158. Eser, S.; Schnieke, A.; Schneider, G.; Saur, D. Oncogenic KRAS signalling in pancreatic cancer. *Br. J. Cancer* **2014**, *111*, 817–822.

159. Laghi, L.; Orbetegli, O.; Bianchi, P.; Zerbi, A.; Di Carlo, V.; Boland, C. R.; Malesci, A. Common occurrence of multiple K-RAS mutations in pancreatic cancers with associated precursor lesions and in biliary cancers. *Oncogene* **2002**, *21*, 4301-4306.
160. Tan, C.; Du, X. KRAS mutation testing in metastatic colorectal cancer. *World J. Gastroenterol.* **2012**, *18*, 5171.
161. Ostrem, J. M.; Peters, U.; Sos, M. L.; Wells, J. A.; Shokat, K. M. K-Ras (G12C) inhibitors allosterically control GTP affinity and effector interactions. *Nature* **2013**, *503*, 548-551.
162. Zimmermann, G.; Papke, B.; Ismail, S.; Vartak, N.; Chandra, A.; Hoffmann, M.; Hahn, S. A.; Triola, G.; Wittinghofer, A.; Bastiaens, P. I. Small molecule inhibition of the KRAS-PDE [dgr] interaction impairs oncogenic KRAS signalling. *Nature* **2013**, *497*, 638-642.
163. Song, Y.; Madahar, V.; Liao, J. Development of FRET Assay into quantitative and high-throughput screening Technology platforms for protein–protein interactions. *Ann. Biomed. Eng.* **2011**, *39*, 1224-1234.
164. Sarge, K. D.; Park-Sarge, O. Sumoylation and human disease pathogenesis. *Trends Biochem. Sci.* **2009**, *34*, 200-205.
165. Dorval, V.; Fraser, P. E. SUMO on the road to neurodegeneration. *Mol. Cell Res.* **2007**, *1773*, 694-706.

166. Kim, K. I.; Baek, S. H. SUMOylation code in cancer development and metastasis. *Mol. Cells* **2006**, *22*, 247-253.
167. Seeler, J.; Bischof, O.; Nacerddine, K.; Dejean, A. In *SUMO, the three Rs and cancer*; Acute Promyelocytic Leukemia; Springer: 2007; pp 49-71.
168. Schaap, M.; Hancock, R.; Wilderspin, A.; Wells, G. Development of a steady-state FRET-based assay to identify inhibitors of the Keap1-Nrf2 protein-protein interaction. *Protein Sci.* **2013**, *22*, 1812-1819.
169. Galarneau, A.; Primeau, M.; Trudeau, L.; Michnick, S. W.  $\beta$ -Lactamase protein fragment complementation assays as in vivo and in vitro sensors of protein-protein interactions. *Nat. Biotechnol.* **2002**, *20*, 619-622.
170. O'Brien, S. P.; DeLisa, M. P. Split-Cre recombinase effectively monitors protein-protein interactions in living bacteria. *Biotechnol. J.* **2014**, *9*, 355-361.
171. Paulmurugan, R.; Gambhir, S. Monitoring protein-protein interactions using split synthetic renilla luciferase protein-fragment-assisted complementation. *Anal. Chem.* **2003**, *75*, 1584-1589.
172. Hashimoto, J.; Watanabe, T.; Seki, T.; Karasawa, S.; Izumikawa, M.; Seki, T.; Iemura, S.; Natsume, T.; Nomura, N.; Goshima, N.; Miyawaki, A.; Takagi, M.; Shin-Ya, K. Novel in vitro protein fragment complementation

- assay applicable to high-throughput screening in a 1536-well format. *J. Biomol. Screen.* **2009**, *14*, 970-979.
173. Broussard, J. A.; Rappaz, B.; Webb, D. J.; Brown, C. M. Fluorescence resonance energy transfer microscopy as demonstrated by measuring the activation of the serine/threonine kinase Akt. *Nat. Protoc.* **2013**, *8*, 265-281.
174. Garner, C. C.; Nash, J.; Haganir, R. L. PDZ domains in synapse assembly and signalling. *Trends Cell Biol.* **2000**, *10*, 274-280.
175. Tsunoda, S.; Sierralta, J.; Sun, Y.; Bodner, R.; Suzuki, E.; Becker, A.; Socolich, M.; Zuker, C. S. A multivalent PDZ-domain protein assembles signalling complexes in a G-protein-coupled cascade. *Nature* **1997**, *388*, 243-249.
176. Amacher, J. F.; Cushing, P. R.; Bahl, C. D.; Beck, T.; Madden, D. R. Stereochemical Determinants of C-terminal Specificity in PDZ Peptide-binding Domains. *J. Biol. Chem.* **2012**, *2013*, 5114-5126.
177. Lee, H.; Zheng, J. J. Review PDZ domains and their binding partners: structure, specificity, and modification. *Cell Commun. Signal.* **2010**, *8*, 8.
178. Tonikian, R.; Zhang, Y.; Sazinsky, S. L.; Currell, B.; Yeh, J.; Reva, B.; Held, H. A.; Appleton, B. A.; Evangelista, M.; Wu, Y. A specificity map for the PDZ domain family. *PLoS biology* **2008**, *6*, e239.

179. Shida, D.; Watanabe, T.; Aoki, J.; Hama, K.; Kitayama, J.; Sonoda, H.; Kishi, Y.; Yamaguchi, H.; Sasaki, S.; Sako, A. Aberrant expression of lysophosphatidic acid (LPA) receptors in human colorectal cancer. *Lab. Invest.* **2004**, *84*, 1352-1362.
180. Lin, S.; Wang, D.; Iyer, S.; Ghaleb, A. M.; Shim, H.; Yang, V. W.; Chun, J.; Yun, C. C. The Absence of LPA<sub>2</sub> Attenuates Tumor Formation in an Experimental Model of Colitis-Associated Cancer. *Gastroenterology* **2009**, *136*, 1711-1720.
181. Meneses-Morales, I.; Tecalco-Cruz, A. C.; Barrios-Garcia, T.; Gomez-Romero, V.; Trujillo-Gonzalez, I.; Reyes-Carmona, S.; Garcia-Zepeda, E.; Mendez-Enriquez, E.; Cervantes-Roldan, R.; Perez-Sanchez, V.; Recillas-Targa, F.; Mohar-Betancourt, A.; Leon-Del-Rio, A. SIP1/NHERF2 enhances estrogen receptor alpha transactivation in breast cancer cells. *Nucleic Acids Res.* **2014**, *42*, 6885-6900.
182. Lee, S.; Zhang, H.; Shim, H.; Yun, C. MAGI-3 and NHERF2, reciprocally regulate the oncogenic effects of LPA<sub>2</sub> receptor. *Gastroenterology* **2010**, *138*, S-249-S-250.
183. Oh, Y. S.; Jo, N. W.; Choi, J. W.; Kim, H. S.; Seo, S. W.; Kang, K. O.; Hwang, J. I.; Heo, K.; Kim, S. H.; Kim, Y. H.; Kim, I. H.; Kim, J. H.; Banno, Y.; Ryu, S. H.; Suh, P. G. NHERF2 specifically interacts with LPA<sub>2</sub> receptor

- and defines the specificity and efficiency of receptor-mediated phospholipase C-beta3 activation. *Mol. Cell. Biol.* **2004**, *24*, 5069-5079.
184. Voltz, J. W.; Weinman, E. J.; Shenolikar, S. Expanding the role of NHERF, a PDZ-domain containing protein adapter, to growth regulation. *Oncogene* **2001**, *20*, 6309-6314.
185. Zheng, F.; Jewell, H.; Fitzpatrick, J.; Zhang, J.; Mierke, D. F.; Grigoryan, G. Computational Design of Selective Peptides to Discriminate between Similar PDZ Domains in an Oncogenic Pathway. *J. Mol. Biol.* **2015**, *427*, 491-510.
186. Tripet, B.; Yu, L.; Bautista, D. L.; Wong, W. Y.; Irvin, R. T.; Hodges, R. S. Engineering a de novo-designed coiled-coil heterodimerization domain off the rapid detection, purification and characterization of recombinantly expressed peptides and proteins. *Protein Eng.* **1996**, *9*, 1029-1042.
187. Fabian, J. R.; Vojtek, A. B.; Cooper, J. A.; Morrison, D. K. A single amino acid change in Raf-1 inhibits Ras binding and alters Raf-1 function. *Proc. Natl. Acad. Sci. U. S. A.* **1994**, *91*, 5982-5986.
188. Mason, J. M.; Arndt, K. M. Coiled coil domains: stability, specificity, and biological implications. *ChemBioChem* **2004**, *5*, 170-176.
189. Eferl, R.; Wagner, E. F. AP-1: a double-edged sword in tumorigenesis. *Nat. Rev. Cancer* **2003**, *3*, 859-868.

190. Karin, M.; Liu, Z.; Zandi, E. AP-1 function and regulation. *Curr. Opin. Cell Biol.* **1997**, *9*, 240-246.
191. Edwards, J.; Krishna, N. S.; Mukherjee, R.; Bartlett, J. The role of c-Jun and c-Fos expression in androgen-independent prostate cancer. *J. Pathol.* **2004**, *204*, 153-158.
192. Ransone, L. J.; Visvader, J.; Sassone-Corsi, P.; Verma, I. M. Fos-Jun interaction: mutational analysis of the leucine zipper domain of both proteins. *Genes Dev.* **1989**, *3*, 770-781.
193. Kelley, L. A.; Sternberg, M. J. Protein structure prediction on the Web: a case study using the Phyre server. *Nat. Protoc.* **2009**, *4*, 363-371.
194. Comeau, S. R.; Gatchell, D. W.; Vajda, S.; Camacho, C. J. ClusPro: a fully automated algorithm for protein-protein docking. *Nucleic Acids Res.* **2004**, *32*, W96-9.
195. Comeau, S. R.; Gatchell, D. W.; Vajda, S.; Camacho, C. J. ClusPro: an automated docking and discrimination method for the prediction of protein complexes. *Bioinformatics* **2004**, *20*, 45-50.
196. Kozakov, D.; Brenke, R.; Comeau, S. R.; Vajda, S. PIPER: An FFT-based protein docking program with pairwise potentials. *Proteins: Struct., Funct., Bioinf.* **2006**, *65*, 392-406.

197. Lensink, M. F.; Méndez, R.; Wodak, S. J. Docking and scoring protein complexes: CAPRI 3rd Edition. *Proteins: Struct., Funct., Bioinf.* **2007**, *69*, 704-718.
198. Schrödinger LLC. The PyMOL Molecular Graphics System, Version 1.3 r1.
199. Hu, B.; Gilkes, D. M.; Chen, J. Efficient p53 activation and apoptosis by simultaneous disruption of binding to MDM2 and MDMX. *Cancer Res.* **2007**, *67*, 8810-8817.
200. Liu, Z.; Olejniczak, E. T.; Fesik, S. W. Over-expression of the human MDM2 p53 binding domain by fusion to a p53 transactivation peptide. *Protein Expr. Purif.* **2004**, *37*, 493-498.
201. Barnard, D.; Sun, H.; Baker, L.; Marshall, M. S. In vitro inhibition of Ras–Raf association by short peptides. *Biochem. Biophys. Res. Commun.* **1998**, *247*, 176-180.
202. Fabian, J. R.; Vojtek, A. B.; Cooper, J. A.; Morrison, D. K. A single amino acid change in Raf-1 inhibits Ras binding and alters Raf-1 function. *Proc. Natl. Acad. Sci. U. S. A.* **1994**, *91*, 5982-5986.
203. Tchekanda, E.; Sivanesan, D.; Michnick, S. W. An infrared reporter to detect spatiotemporal dynamics of protein-protein interactions. *Nat. Methods* **2014**, *11*, 641-644.

204. Cutler, R. E.; Morrison, D. K. Mammalian Raf-1 is activated by mutations that restore Raf signaling in *Drosophila*. *EMBO J.* **1997**, *16*, 1953-1960.
205. Nassar, M.; Horn, G.; Herrmann, C.; Scherer, A.; McCormick, F.; Wittinghofer, A. The 2.2-Angstrom Crystal-Structure of the Ras-Binding Domain of the Serine Threonine Kinase C-Raf1 in Complex with Rap1a and a Gtp Analog. *Nature* **1995**, *375*, 554-560.
206. Fetis, S. K.; Kearney, B. M.; Buhrman, G.; Mattos, C. Allosteric effects of oncogenic Ras Q61L mutant of Raf-RBD. *Structure* **2015**, *23*, 505-516.
207. Vouilleme, L.; Cushing, P. R.; Volkmer, R.; Madden, D. R.; Boisguerin, P. Engineering peptide inhibitors to overcome PDZ binding promiscuity. *Angew. Chem., Int. Ed.* **2010**, *49*, 9912-9916.
208. Elkins, J. M.; Papagrigoriou, E.; Berridge, G.; Yang, X.; Phillips, C.; Gileadi, C.; Savitsky, P.; Doyle, D. A. Structure of PICK1 and other PDZ domains obtained with the help of self-binding C-terminal extensions. *Protein Sci.* **2007**, *16*, 683-694.
209. Buchan, D. W.; Minneci, F.; Nugent, T. C.; Bryson, K.; Jones, D. T. Scalable web services for the PSIPRED Protein Analysis Workbench. *Nucleic Acids Res.* **2013**, *41*, W349-57.

210. Joo, S. H.; Pei, D. Synthesis and screening of support-bound combinatorial peptide libraries with free C-termini: determination of the sequence specificity of PDZ domains. *Biochemistry (N. Y.)* **2008**, *47*, 3061-3072.
211. Songyang, Z.; Fanning, A. S.; Fu, C.; Xu, J.; Marfatia, S. M.; Chishti, A. H.; Crompton, A.; Chan, A. C.; Anderson, J. M.; Cantley, L. C. Recognition of unique carboxyl-terminal motifs by distinct PDZ domains. *Science* **1997**, *275*, 73-77.
212. Lee, C.; Laimins, L. A. Role of the PDZ domain-binding motif of the oncoprotein E6 in the pathogenesis of human papillomavirus type 31. *J. Virol.* **2004**, *78*, 12366-12377.
213. Bessette, P. H.; Mena, M. A.; Nguyen, A. W.; Daugherty, P. S. In *Construction of designed protein libraries using gene assembly mutagenesis; Directed Evolution Library Creation*; Springer: 2003; pp 29-37.
214. Ai, H.; Baird, M. A.; Shen, Y.; Davidson, M. W.; Campbell, R. E. Engineering and characterizing monomeric fluorescent proteins for live-cell imaging applications. *Nat. Protoc.* **2014**, *9*, 910-928.
215. Ehrhardt, D. GFP technology for live cell imaging. *Curr. Opin. Plant Biol.* **2003**, *6*, 622-628.
216. Stephens, D. J.; Allan, V. J. Light microscopy techniques for live cell imaging. *Science* **2003**, *300*, 82-86.

217. McEvoy, A. L.; Hoi, H.; Bates, M.; Platonova, E.; Cranfill, P. J.; Baird, M. A.; Davidson, M. W.; Ewers, H.; Liphardt, J.; Campbell, R. E. mMaple: a photoconvertible fluorescent protein for use in multiple imaging modalities. *PloS one* **2012**, *7*, e51314.
218. Lee, S. H.; Shin, J. Y.; Lee, A.; Bustamante, C. Counting single photoactivatable fluorescent molecules by photoactivated localization microscopy (PALM). *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 17436-17441.
219. Cicchetti, G.; Biernacki, M.; Farquharson, J.; Allen, P. G. A ratiometric expressible FRET sensor for phosphoinositides displays a signal change in highly dynamic membrane structures in fibroblasts. *Biochemistry (N. Y.)* **2004**, *43*, 1939-1949.
220. Baird, G. S.; Zacharias, D. A.; Tsien, R. Y. Biochemistry, mutagenesis, and oligomerization of DsRed, a red fluorescent protein from coral. *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97*, 11984-11989.
221. Linnemann, T.; Geyer, M.; Jaitner, B. K.; Block, C.; Kalbitzer, H. R.; Wittinghofer, A.; Herrmann, C. Thermodynamic and kinetic characterization of the interaction between the Ras binding domain of AF6 and members of the Ras subfamily. *J. Biol. Chem.* **1999**, *274*, 13556-13562.

222. Mochizuki, N.; Yamashita, S.; Kurokawa, K.; Ohba, Y.; Nagai, T.; Miyawaki, A.; Matsuda, M. Spatio-temporal images of growth-factor-induced activation of Ras and Rap1. *Nature* **2001**, *411*, 1065-1068.
223. Moerke, N. J. Fluorescence Polarization (FP) Assays for Monitoring Peptide-Protein or Nucleic Acid-Protein Binding. *Curr. Protoc. Chem. Biol.* **2009**, 1-15.
224. Owicki, J. C. Fluorescence polarization and anisotropy in high throughput screening: perspectives and primer. *J. Biomol. Screen.* **2000**, *5*, 297-306.
225. Banks, P.; Gosselin, M.; Prystay, L. Impact of a red-shifted dye label for high throughput fluorescence polarization assays of G protein-coupled receptors. *J. Biomol. Screen.* **2000**, *5*, 329-334.
226. Lilie, H.; Schwarz, E.; Rudolph, R. Advances in refolding of proteins produced in *E. coli*. *Curr. Opin. Biotechnol.* **1998**, *9*, 497-501.
227. Kong, B.; Guo, G. L. Soluble Expression of Disulfide Bond Containing Proteins FGF15 and FGF19 in the Cytoplasm of *Escherichia coli*. *PloS one* **2014**, *9*, e85890.
228. Nozach, H.; Fruchart-Gaillard, C.; Fenaille, F.; Beau, F.; Ramos, O.; Douzi, B.; Saez, N. J.; Moutiez, M.; Servent, D.; Gondry, M. High throughput screening identifies disulfide isomerase DsbC as a very efficient partner for

- recombinant expression of small disulfide-rich proteins in *E. coli*. *Microb Cell Fact* **2013**, *12*, 10.1186.
229. Nakamoto, H.; Bardwell, J. C. Catalysis of disulfide bond formation and isomerization in the *Escherichia coli* periplasm. *BBA-Mol. Cell Res.* **2004**, *1694*, 111-119.
230. Bird, R. E.; Walker, B. W. Single chain antibody variable regions. *Trends Biotechnol.* **1991**, *9*, 132-137.
231. Miller, K. D.; Weaver-Feldhaus, J.; Gray, S. A.; Siegel, R. W.; Feldhaus, M. J. Production, purification, and characterization of human scFv antibodies expressed in *Saccharomyces cerevisiae*, *Pichia pastoris*, and *Escherichia coli*. *Protein Expr. Purif.* **2005**, *42*, 255-267.
232. Duenas, M.; Vazquez, J.; Ayala, M.; Soderlind, E.; Ohlin, M.; Perez, L.; Borrebaeck, C. A.; Gavilondo, J. V. Intra- and extracellular expression of an scFv antibody fragment in *E. coli*: effect of bacterial strains and pathway engineering using GroES/L chaperonins. *BioTechniques* **1994**, *16*, 476-7, 480-3.
233. Revets, H.; De Baetselier, P.; Muyldermans, S. Nanobodies as novel agents for cancer therapy. *Expert Opin. Biol. Th.* **2005**, *5*, 111-124.

234. Van Bockstaele, F.; Holz, J. B.; Revets, H. The development of nanobodies for therapeutic applications. *Curr. Opin. Investig Drugs* **2009**, *10*, 1212-1224.
235. Harmsen, M.; De Haard, H. Properties, production, and applications of camelid single-domain antibody fragments. *Appl. Microbiol. Biotechnol.* **2007**, *77*, 13-22.
236. Bogdanowich-Knipp, S. J.; Chakrabarti, S.; Siahhan, T. J.; Williams, T. D.; Dillman, R. K. Solution stability of linear vs. cyclic RGD peptides. *J. Pep. Res.* **1999**, *53*, 530-541.
237. Roxin, Á; Zheng, G. Flexible or fixed: a comparative review of linear and cyclic cancer-targeting peptides. *Future Med. Chem.* **2012**, *4*, 1601-1618.
238. Scott, C. P.; Abel-Santos, E.; Wall, M.; Wahnou, D. C.; Benkovic, S. J. Production of cyclic peptides and proteins in vivo. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96*, 13638-13643.
239. Evans, T. C., Jr; Martin, D.; Kolly, R.; Panne, D.; Sun, L.; Ghosh, I.; Chen, L.; Benner, J.; Liu, X. Q.; Xu, M. Q. Protein trans-splicing and cyclization by a naturally split intein from the dnaE gene of *Synechocystis* species PCC6803. *J. Biol. Chem.* **2000**, *275*, 9091-9094.

240. Iwai, H.; Züger, S.; Jin, J.; Tam, P. Highly efficient protein trans-splicing by a naturally split DnaE intein from *Nostoc punctiforme*. *FEBS Lett.* **2006**, *580*, 1853-1858.
241. Shah, N. H.; Eryilmaz, E.; Cowburn, D.; Muir, T. W. Naturally split inteins assemble through a “capture and collapse” mechanism. *J. Am. Chem. Soc.* **2013**, *135*, 18673-18681.
242. Horswill, A. R.; Savinov, S. N.; Benkovic, S. J. A systematic method for identifying small-molecule modulators of protein-protein interactions. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 15591-15596.
243. Tavassoli, A.; Benkovic, S. J. Split-intein mediated circular ligation used in the synthesis of cyclic peptide libraries in *E. coli*. *Nat. Protoc.* **2007**, *2*, 1126-1133.
244. Gruber, C. W.; Elliott, A. G.; Ireland, D. C.; Delprete, P. G.; Dessein, S.; Goransson, U.; Trabi, M.; Wang, C. K.; Kinghorn, A. B.; Robbrecht, E.; Craik, D. J. Distribution and evolution of circular miniproteins in flowering plants. *Plant Cell* **2008**, *20*, 2471-2483.
245. Lindholm, P.; Goransson, U.; Johansson, S.; Claeson, P.; Gullbo, J.; Larsson, R.; Bohlin, L.; Backlund, A. Cyclotides: a novel type of cytotoxic agents. *Mol. Cancer Ther.* **2002**, *1*, 365-369.

246. Avrutina, O.; Schmoldt, H.; Gabrijelcic-Geiger, D.; Le Nguyen, D.; Sommerhoff, C. P.; Diederichsen, U.; Kolmar, H. Trypsin inhibition by macrocyclic and open-chain variants of the squash inhibitor MCoTI-II. *Biol. Chem.* **2005**, *386*, 1301-1306.
247. Bokesch, H. R.; Pannell, L. K.; Cochran, P. K.; Sowder, R. C.; McKee, T. C.; Boyd, M. R. A Novel Anti-HIV Macrocyclic Peptide from *Palicourea condensata*. *J. Nat. Prod.* **2001**, *64*, 249-250.
248. Ji, Y.; Majumder, S.; Millard, M.; Borra, R.; Bi, T.; Elnagar, A. Y.; Neamati, N.; Shekhtman, A.; Camarero, J. A. In vivo activation of the p53 tumor suppressor pathway by an engineered cyclotide. *J. Am. Chem. Soc.* **2013**, *135*, 11623-11633.
249. Getz, J. A.; Rice, J. J.; Daugherty, P. S. Protease-resistant peptide ligands from a knottin scaffold library. *ACS Chem. Biol.* **2011**, *6*, 837-844.
250. Daly, N. L.; Craik, D. J. Acyclic permutants of naturally occurring cyclic proteins. Characterization of cystine knot and beta-sheet formation in the macrocyclic polypeptide kalata B1. *J. Biol. Chem.* **2000**, *275*, 19068-19075.
251. Jackrel, M. E.; Cortajarena, A. L.; Liu, T. Y.; Regan, L. Screening libraries to identify proteins with desired binding activities using a split-GFP reassembly assay. *ACS Chem. Biol.* **2010**, *5*, 553-562.

252. Telford, W. G.; Hawley, T.; Subach, F.; Verkhusha, V.; Hawley, R. G. Flow cytometry of fluorescent proteins. *Methods* **2012**, *57*, 318-330.
253. Piatkevich, K. D.; Subach, F. V.; Verkhusha, V. V. Far-red light photoactivatable near-infrared fluorescent proteins engineered from a bacterial phytochrome. *Nature Commun.* **2013**, *4*.