SPLIT-PROTEIN REASSEMBLY METHODS FOR THE DETECTION AND INTERROGATION OF BIOMOLECULAR INTERACTIONS AND MODULATORS THEREOF

By

Jason R. Porter

A Dissertation Submitted to the Faculty of the

DEPARTMENT OF CHEMISTRY AND BIOCHEMISTRY

In Partial Fulfillment of the Requirements
For the Degree of

DOCTOR OF PHILOSOPHY

In The Graduate College

THE UNIVERSITY OF ARIZONA

2009
As members of the Dissertation Committee, we certify that we have read the dissertation prepared by Jason R. Porter entitled Split-Protein Reassembly Methods for the Detection and Interrogation of Biomolecular Interactions and Modulators Thereof and recommend that it be accepted as fulfilling the dissertation requirement for the Degree of Doctor of Philosophy.

Indraneel Ghosh  
Date: 8/18/09  
William R. Montfort  
Date: 8/18/09  
Bogdan Z. Olenyuk  
Date: 8/18/09  
Vicki H. Wysocki  
Date: 8/18/09

Final approval and acceptance of this dissertation is contingent upon the candidate’s submission of the final copies of the dissertation the Graduate College.

I hereby certify that I have read this dissertation prepared under my direction and recommend that it be accepted as fulfilling the dissertation requirement.

Dissertation Director: Indraneel Ghosh  
Date: 8/18/09
STATEMENT BY AUTHOR

This dissertation has been submitted in partial fulfillment of requirements for and advanced degree at the University of Arizona and is deposited in the University Library to be made available to borrowers under rules of the Library.

Brief quotations from this dissertation are allowable without special permission, provided that accurate acknowledgment of source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the author.

SIGNED: Jason R. Porter
I would like to thank my advisor Dr. Indraneel Ghosh for his diligent guidance, seemingly infinite patience, and constant encouragement. I thank both past and present members of the Ghosh research group, specifically Dr. Cliff Stains, Dr. Scott Meyer, Dr. Srivats Rajagopal, Jennifer Furman, Ben Jester, and Ahmed Badran for their helpful advice and stimulating conversations concerning matters both scientific and colloquial.

This dissertation is dedicated to my family and friends, for without their support this accomplishment would not have been possible.
TABLE OF CONTENTS

LIST OF FIGURES................................................................................................... 9

LIST OF TABLES..................................................................................................... 13

LIST OF SCHEMES..................................................................................................14

ABSTRACT............................................................................................................... 15

CHAPTER 1: APPLICATIONS OF SPLIT-PROTEIN REASSEMBLY............... 17
  1.1 Explanation of Dissertation Format............................................................... 17
  1.2 Introduction................................................................................................... 17
  1.3 Applications of Split-Protein Reassembly..................................................... 22
    1.3.1 Applications of Split-Protein Reassembly Towards Protein
          Engineering.............................................................................................. 22
    1.3.2 Examining Protein-Protein Interactions and Protein Stability.............. 27
    1.3.3 High Throughput Screening of Protein-Protein Interactions............... 29
    1.3.4 Protein Interactions with Small-Molecules........................................... 36
    1.3.5 Detection of Protein-Nucleic Acid Interactions.................................... 45
    1.3.6 Split-Protein Reassembly in Model Organisms.................................... 50
  1.4 New Directions in Split-Protein Reassembly................................................ 52

CHAPTER 2: APPLICATION OF SEER TO THE DETECTION OF TRANSGENES
          AND CYTOSINE METHYLATION................................................................. 53
  2.1 Introduction................................................................................................... 53
  2.2 Creation and Evaluation of a CryIA-SEER-β-Lac Sensor............................ 57
    2.2.1 Optimization of CryIA-SEER-β-Lac Sensor Conditions..................... 60
  2.3 Sequence Specific Detection of Cytosine Methylation................................. 63
    2.3.1 Design of a mCpG-β-lactamase SEER Sensor..................................... 66
    2.3.2 Initial mCpG-SEER-β-Lac Activity.................................................... 68
    2.3.3 mCpG-SEER-β-Lac Binding Site Specificity...................................... 69
    2.3.4 mCpG-SEER-β-Lac Activity Dependence on Target Site Spacing..... 72
    2.3.5 dsDNA Concentration Dependence of mCpG-SEER-β-Lac Activity.. 74
    2.3.6 Detection of Methylated dsDNA in the Presence of Excess Non-Specific
          Genomic DNA......................................................................................... 76
  2.4 Conclusion..................................................................................................... 78
  2.5 Experimental Procedures............................................................................... 84
    2.5.1 Cloning Expression, and Purification of CryIA-SEER-β-Lac and mCpG-
          SEER-β-Lac Proteins.............................................................................. 84
    2.5.2 SDS-PAGE Analysis of CryIA-SEER-β-Lac and mCpG-SEER-β-Lac
          Proteins.................................................................................................... 90
    2.5.3 Preparation of dsDNA Targets............................................................. 90
    2.5.4 Absorbance Measurements................................................................ 90
    2.5.5 Fluorescence Measurements............................................................... 92
    2.5.6 Initial Evaluation of the CryIA-SEER-β-Lac Sensor......................... 92
TABLE OF CONTENTS-Continued

2.5.7 Determination of Optimal CryIA Sensor Conditions........................... 93
2.5.8 Limit of Detection for CryIA-SEER-β-Lac......................................... 93
2.5.9 mCpG-SEER-β-Lac Specificity and Target Site Spacing
Experiments .................................................................................................. 94
2.5.10 Limit of Detection of mCpG-SEER-β-Lac........................................ 94

CHAPTER 3: APPLICATION OF FIREFLY LUCIFERASE TO SEER........... 96
3.1 Introduction ............................................................................................. 96
3.2 Results and Discussion .......................................................................... 99
  3.2.1 Creation of a dsDNA Sequence Specific Split-Firefly Luciferase Sensor (SEER-FFluc) ........................................................................................................ 99
  3.2.2 Denaturation and Refolding of Full Length and SEER-FFluc............. 104
  3.2.3 Creation of MBP-PBSII-CFluc............................................................ 107
  3.2.4 Cell-Free dsDNA-Dependent Firefly Luciferase Reassembly............. 109
  3.2.5 Conclusion ........................................................................................ 111
3.3 Experimental Procedures ....................................................................... 112
  3.3.1 Cloning Expression, and Purification of Full Length Firefly Luciferase and SEER-FFluc Proteins................................. 112
  3.3.2 Determination of Luciferase Concentration Dependent Luminescence.......................................................................................... 119
  3.3.3 Evaluation of dsDNA-Dependent Luminescence Utilizing Soluble Purified His6-NFluc(1-438)-Zif268 and His6-PBSII-CFluc(438-550) ..... 124
  3.3.4 Denaturation and Refolding of Full Length and Split-Firefly Luciferase ...................................................................................................................... 124
  3.3.5 Denaturation and Reassembly of SEER-FFluc Proteins...................... 125
  3.3.6 Determination of dsDNA binding by SEER-FFluc Proteins.............. 125
  3.3.7 Cell-Free dsDNA-Dependent Reassembly of Firefly Luciferase........ 125

CHAPTER 4: DEVELOPMENT OF A GENERAL CELL-FREE SPLIT-FIREFLY LUCIFERASE METHODOLOGY FOR THE RAPID INTERROGATION OF BIOMOLECULAR INTERACTIONS......................................................... 129
4.1 Introduction ............................................................................................. 129
4.2 Results and Discussion .......................................................................... 132
  4.2.1 Initial Evaluation of the Reassembly of Split-Reporter Proteins in Cell-Free Translation Systems............................................................. 132
  4.2.2 Cell-Free Interrogation of Protein and Protein-Nucleic Acid
Interactions .................................................................................................... 134
  4.2.3 Cell-Free Split-Luciferase Reassembly using a Coupled Rabbit Reticulocyte Lysate System................................................................. 137
  4.2.4 Split-Firefly Luciferase Utilizing Purified Transcription and Translation Components ................................................................. 139
  4.2.5 Reversibility of Split-Firefly Luciferase Reassembly......................... 139
  4.2.6 Cell-Free Determination Nucleic Acid-Binding Protein Specificity... 140
  4.2.7 Conclusion ........................................................................................ 143
TABLE OF CONTENTS - Continued

4.3 Experimental Procedures.......................................................................................... 145
  4.3.1 General Materials............................................................................................. 145
  4.3.2 Plasmid Construction and mRNA Production.................................................. 145
  4.3.3 Target DNA Preparation.................................................................................. 147
  4.3.4 Reassembly of the GFP Variant Venus............................................................. 147
  4.3.5 Reassembly of Split β-lactamase-Zinc Finger Fusions........................................ 147
  4.3.6 Reassembly of Split-Firefly Luciferase.............................................................. 148
  4.3.7 Initial Cell-Free Assays................................................................................... 148
  4.3.8 Reassembly of Split-Firefly Luciferase in a Coupled Transcription
      and Translation Rabbit Reticulocyte Lysate System............................................. 149
  4.3.9 Reassembly of Split-Firefly Luciferase in a Purified Transcription
      and Translation System....................................................................................... 150
  4.3.10 Competition Assay to Identify Protein-DNA Target Site
       Specificity.............................................................................................................. 150
  4.3.11 Dissociation of the Reassembled Pum2-NFluc, CFluc-Pum1,
       RNA Ternary Complex....................................................................................... 151

CHAPTER 5: CELL-FREE INTERROGATION OF PROTEIN-PROTEIN
    INTERACTIONS AND INHIBITORS THEREOF........................................................ 153
  5.1 Introduction.......................................................................................................... 153
    5.1.1 Methods for the Development of Small-Molecule Inhibitors of
         Protein-Protein Interactions............................................................................... 154
    5.1.2 Methods for the Interrogation of Protein-Protein Interactions and
         Inhibitors Thereof............................................................................................. 158
  5.2 Results and Discussion.......................................................................................... 162
    5.2.1 Cell-Free Interrogation of Protein-Protein Interactions................................. 162
    5.2.2 A Cell-Free Split-Luciferase Assay for the Detection of Peptide
         and Small-Molecule Antagonists of Protein-Protein Interactions....................... 170
    5.2.3 Cell-Free Split-Luciferase Assays for the Interrogation of Inhibitors
         of Helix-Receptor Interactions......................................................................... 173
    5.2.4 Specificity Determination of Peptide Inhibitors of Protein-Protein
         Interactions........................................................................................................ 177
    5.2.5 Cell-Free Interrogation of Inhibitors of the hDM2/p53 and
         hDM4/p53 Interactions....................................................................................... 181
    5.2.6 Specificity Determination of Small-Molecule Inhibitors of Protein-Protein
         Interactions........................................................................................................ 183
    5.2.7 Conclusion...................................................................................................... 188
  5.3 Materials and Methods.......................................................................................... 189
    5.3.1 General Reagents........................................................................................... 189
    5.3.2 Cloning of Fusion Proteins and In Vitro Transcription.................................... 190
    5.3.3 Cell-Free Interrogation of Protein-Protein Interactions..................................... 192
    5.3.4 Cell-Free Interrogation of Peptide and Small-Molecule Inhibitors of
         Protein-Protein Interactions............................................................................... 192
TABLE OF CONTENTS-Continued

5.3.5 Cell-Free Interrogation of Bcl-2 Family Inhibitors:
   Modified Protocol.......................................................................................... 193
5.3.6 Cell-Free Interrogation of Inhibitor Specificity......................................... 194

CHAPTER 6: NEW DIRECTIONS IN THE CELL-FREE SPLIT-FIREFLY
   LUCIFERASE ASSAY....................................................................................... 196
6.1 Introduction.................................................................................................... 196
   6.1.1 Application of the Cell-Free Split-Firefly Luciferase Methodology
   for New Sensor Development: ssRNA, Amyloid oligomers, Native
   Proteins, and Proteases................................................................................ 196
   6.1.2 Cell-Free Interrogation of Kinase Inhibitor Specificity............................ 199
6.2 Results and Discussion................................................................................. 205
   6.2.1 Design of a Split-Firefly Luciferase Chemically Induced
   Dimerization Assay...................................................................................... 205
   6.2.2 Evaluation of Small-Molecule-Dependent Firefly Luciferase
   Reassembly..................................................................................................... 206
   6.2.3 Evaluation of Alternative Orientations of Split-Firefly Luciferase
   Fusions........................................................................................................... 208
   6.2.4 Conclusion.............................................................................................. 210
6.3 Material and Methods.................................................................................. 211
   6.3.1 Cloning of DHFR and Kinase Split-Firefly Luciferase Fusions and
   In Vitro Transcription................................................................................... 211
   6.3.2 Synthesis of TMP-Staur......................................................................... 212
   6.3.3 Evaluation of TMP-Staur-Dependent Firefly Luciferase Reassembly. 215
   6.3.4 Staurosporine Competition Assay........................................................ 217

APPENDIX A: PLASMID MAPS........................................................................... 218

REFERENCES......................................................................................................... 286
LIST OF FIGURES

Figure 1.1 Principle behind split-protein reassembly.......................................................... 20
Figure 1.2 Split-mDHFR library-versus-library selection.................................................... 23
Figure 1.3 Diagram of the bait and prey selection strategy................................................... 25
Figure 1.4 Split-protein assay for determinants of protein folding........................................ 30
Figure 1.5 Previously described methods for the identification of protein-protein interactions........................................................................................................ 32
Figure 1.6 Split-ubiquitin strategy for the detection of membrane protein interactions................................. 34
Figure 1.7 Diagram of the RePCA screen.............................................................................. 37
Figure 1.8 Schematic of downstream drug effects................................................................. 39
Figure 1.9 Model of the ER-LBD split-rLuc intramolecular folding sensor............................. 42
Figure 1.10 Schematic of the CID three-hybrid strategy......................................................... 44
Figure 1.11 A general strategy for the visualization of mRNA.............................................. 47
Figure 1.12 General schematic of the SEER strategy............................................................. 49
Figure 2.1 The SEER strategy............................................................................................... 55
Figure 2.2 Modularity and design of DNA sequence specific zinc finger proteins.. 56
Figure 2.3 Initial evaluation of the CryIA SEER-β-lac sensor.............................................. 59
Figure 2.4 Optimization of the CryIA SEER-β-lac sensor..................................................... 62
Figure 2.5 Site-specific determination of promoter methylation.......................................... 64
Figure 2.6 The mCpG-SEER-β-Lac system............................................................................ 67
Figure 2.7 Initial mCpG-SEER-β-Lac activity....................................................................... 70
Figure 2.8 mCpG-SEER-β-Lac specificity.............................................................................. 71
Figure 2.9 mCpG-SEER-β-Lac dependence on target site spacing................................. 73
LIST OF FIGURES-Continued

Figure 2.10 mCpG-SEER-β-Lac limit of detection utilizing nitrocefin............................ 75
Figure 2.11 mCpG-SEER-β-Lac limit of detection utilizing CCF2FA............................ 77
Figure 2.12 Evaluation of mCpG-SEER-β-Lac in the presence of excess
HS-DNA.............................................................................................................. 79
Figure 2.13 The DNA and amino acid sequence of LacA-ZFC1-L3........................... 86
Figure 2.14 The DNA and amino acid sequence of ZFC1-L3-LacB............................ 87
Figure 2.15 The DNA and amino acid sequence of MBD2-LacB.................................. 88
Figure 2.16 Schematics of pMAL-c2X plasmids...................................................... 89
Figure 2.17 SDS-PAGE analysis of SEER-β-Lac proteins........................................... 91
Figure 3.1 Various luciferase reporters..................................................................... 98
Figure 3.2 Firefly luciferase mutations responsible for enhanced thermal
stability and altered spectral properties............................................................... 100
Figure 3.3 Structure of firefly luciferase with highlighted fragmentation sites........... 101
Figure 3.4 Initial luminescence measurements....................................................... 103
Figure 3.5 Refolding of denatured full length firefly luciferase and zinc
finger-firefly luciferase fusions........................................................................... 106
Figure 3.6 PAGE gels showing extent of target dsDNA binding by fusion
Proteins................................................................................................................ 108
Figure 3.7 dsDNA-dependent firefly luciferase reassembly in the rabbit
reticulocyte in vitro translation system............................................................... 110
Figure 3.8 DNA and amino acid sequence of His6-Firefly Luciferase....................... 114
Figure 3.9 DNA and amino acid sequence of His6-NFluc(1-437)-Zif268.................... 116
Figure 3.10 DNA and amino acid sequence of His6-PBSII-CFluc(438-550).............. 117
Figure 3.11 DNA and amino acid sequence of MBP-PBSII-CFluc(438-550).......... 120
Figure 3.12 DNA and amino acid sequence of PBSII-NFluc(2-416)....................... 121
LIST OF FIGURES—Continued

Figure 3.13 DNA and amino acid sequence of CFuc(398-550)-Zif268................... 122

Figure 3.14 Schematics of all plasmids used in this study................................. 123

Figure 4.1 Cell-free detection utilizing split-proteins......................................... 133

Figure 4.2 Cell-free detection of a wide variety of biomolecular interactions utilizing split-luciferase starting from mRNA........................................ 135

Figure 4.3 Cell-free interrogation of a small-molecule dependent protein-protein interaction................................................................. 136

Figure 4.4 Cell-free detection of biomolecular interactions with split-luciferase utilizing coupled transcription and translation systems............. 138

Figure 4.5 Interrogation of protein-nucleic acid interactions utilizing our cell-free split-firefly assay................................................................. 141

Figure 5.1 Small-molecule inhibitors of protein-protein interactions.................. 155

Figure 5.2 Strategies for the identification and interrogation of inhibitors of protein-protein interactions................................................................. 160

Figure 5.3 General schematic for the cell-free interrogation of protein-protein interaction inhibitors................................................................. 163

Figure 5.4 Surface interfaces of helix-receptor interactions (Bcl-XL/BAK)........ 165

Figure 5.5 Structures of hDM2 and hDM4 with p53........................................ 166

Figure 5.6 Bcl-2 family interactions.................................................................... 168

Figure 5.7 Specificity of helix-receptor interactions............................................ 169

Figure 5.8 Interrogation of peptide and small-molecule inhibitors of protein-protein interactions utilizing the cell-free split-firefly luciferase assay......................................................... 172

Figure 5.9 Inhibition of the interaction of the p53 activation domain with hDM2 and hDM4................................................................. 174

Figure 5.10 Inhibition of the interaction of the BIM BH3 domain with Bcl-2 family receptors................................................................................. 176

Figure 5.11 Interrogation of peptide inhibitors of protein-protein interactions.... 178
LIST OF FIGURES-Continued

Figure 5.12 Interaction inhibition by select HBS-Helices......................................... 180

Figure 5.13 Inhibition of hDM2/p53 and hDM4/p53 interaction by various p53
derived peptides.................................................................................................. 182

Figure 5.14 Interrogation of small-molecule inhibitors of Bcl-2 family
interactions.......................................................................................................... 184

Figure 5.15 Interrogation of small-molecule inhibitor specificity....................... 187

Figure 6.1 Additional applications of the cell-free split-firefly luciferase assay...... 198

Figure 6.2 Phage-based kinase ATP site competition strategy........................... 202

Figure 6.3 A general schematic for the CID split-firefly luciferase assay for
interrogating ATP site binding small molecules............................................. 204

Figure 6.4 A CID sensor for the interrogation of kinase ATP site binding small-
molecules........................................................................................................ 207

Figure 6.5 Evaluation of alternative orientation DHFR and kinase firefly luciferase
fusions ................................................................................................................ 209
LIST OF TABLES

Table 1.1 A list of currently developed split-protein sensors................................. 21

Table 2.1 dsDNA target oligonucleotides used to test mCpG-SEER-β-Lac
dependence on target site spacing......................................................................... 95

Table 3.1 dsDNA targets used to evaluate the target site spacing dependence of
firefly luciferase reassembly (SEER-FFluc) ................................................................. 128

Table 4.1 Fusion constructs used in the development of the cell-free split-firefly
luciferase assay........................................................................................................... 146

Table 4.2 Sequences of nucleic acid antagonists used in Chapter 4......................... 152

Table 5.1 Fusion constructs used in Chapter 5............................................................ 191
LIST OF SCHEMES

Scheme 6.1 Synthesis of TMP-Staur................................................................. 213
ABSTRACT

The interactions between protein-protein, protein-nucleic acid, and protein-small molecules are central to biological processes and are key for the design of new therapeutics. Rapid and easy to implement methodologies are needed that enable the interrogation of these interactions in a complex cellular context. Towards this goal, I have utilized the concept of split-protein reassembly, also called protein complementation, for the creation of a variety of sensor architectures that enable the interrogation of protein-nucleic acid, protein-protein, and protein-small molecule interactions.

Utilizing the enzymatic split-reporter β-lactamase and existing zinc finger design strategies we applied our recently developed technology termed SEquence-Enabled Reassembly (SEER) towards the creation of a sensor capable of the specific detection of the CryIA transgene. Additionally, the split β-lactamase reporter was utilized for the site-specific determination of DNA methylation at cytosine residues that is involved in epigenetic regulation. This method, dubbed mCpG-SEER, enabled the direct detection of femtomole levels of dsDNA methylation in sequence specific manner.

In a separate endeavor, we have developed and optimized the first cell-free split-reporter systems for GFP, split β-lactamase, and firefly luciferase for the successful dsDNA-dependent reassembly of the various reporters. Our cell free in vitro translation systems eliminates previous bottlenecks encountered in split-reporter technologies such as laborious transfection/cell culture or protein purification. Capitalizing on the ease of use and speed afforded by this new technology we describe the sensitive detection of
protein-protein, protein-nucleic acid, and protein-small molecule interactions and inhibitors thereof.

In a related area, we have applied this rapid cell-free split-firefly luciferase assay to the specific interrogation of a large class of helix-receptor protein-protein interactions. We have built a panel consisting of the clinically relevant Bcl-2 family of proteins, hDM2, hDM4, and p53 and interrogated the specificity of helix-receptor interactions as well as the specificity of peptide and small-molecule inhibitors of these interactions.

Finally, we describe the further applications of our cell-free technology to the development of a large number of general split-firefly luciferase sensors for the detection of ssRNA sequences, the detection of native proteins, the evaluation of protease activity, and interrogation of enzyme-inhibitor interactions. The new methodologies provided in this study provides a general and enabling methodology for the rapid interrogation of a wide variety of biomolecular interactions and their antagonists without the limitations imposed by current in vitro and in vivo approaches.
CHAPTER 1
APPLICATIONS OF SPLIT-PROTEIN REASSEMBLY

1.1 Explanation of Dissertation Format

Excerpts from peer reviewed published manuscripts where I have been a primary contributor are indicated in the thesis. Full reprints of each published or submitted article along with copyright permissions are provided in the appendices.

1.2 Introduction

The interactions between proteins and their ligands are the driving force behind virtually every known cellular process.\textsuperscript{1,2} Broadly classified as biomolecular interactions, these interactions are composed of a seemingly infinite number of protein-protein, protein-nucleic acid, protein-carbohydrate, and protein-small molecule interactions that not only control cellular states but are also emerging targets for therapeutic intervention. The ability to identify, study, and perturb these interactions is essential to understanding and manipulating biological processes, which is central to understanding and treating human disease. Since the sequencing of the human genome and the emergence of proteomics, we are beginning to understand the complexity of the problem at hand, with a multitude of interactions that are also under post-translational control. One example can be found in a recent study estimating the number of protein-protein interactions in the human interactome at a staggering 650,000 interactions.\textsuperscript{3} When combined with the yet undefined possible number of protein-nucleic acid and protein-small molecule interactions, the total number of biomolecular interactions may appear overwhelming, yet a cell carries out these functions faithfully on a day-to-day basis. In order to fully
understand these events, methods are being developed that allow both the identification and analysis of the biomolecular interactome. To truly understand these interactions, methods must be available that allow the study and interrogation of these interactions in a cellular context. One such recently developed method, split-protein reassembly or protein complementation, has proven itself particularly useful, not only providing the ability to detect interactions between two proteins, but also providing a general strategy for the interrogation of protein-nucleic acid and protein-small molecule interactions as well as small-molecule modulators of these interactions.\textsuperscript{4-6} Moreover, this methodology also allows for discovering and optimizing new biomolecular interactions, which are central to the emerging field of synthetic biology.

Split-protein reassembly involves the rational or systematic dissection of a protein into two fragments, followed by the fusion of those fragments to two proteins of interest. This rational dissection can be achieved by a number of methods including analysis of sequence and structural information\textsuperscript{7-10} or from more biochemical methods such as the development of cyclically permuted variants,\textsuperscript{11,12} peptide insertion scanning,\textsuperscript{13} library selection,\textsuperscript{14,15} and sequence based molecular modeling.\textsuperscript{16,17} Alone, each of the fragments possesses no native activity; however, when the fragments are brought into close proximity, either by direct binding of the attached proteins or through a simultaneous interaction with a third molecule, the two fragments reassemble and produce native activity (Figure 1.1a). Given the universal forces behind protein folding, this strategy can and has been applied to a variety of “reporter” proteins providing a number of readouts. Since this approach was first reported for the split-protein sensor ubiquitin in 1994\textsuperscript{7} by Varshavsky and co-workers, numerous proteins have been fragmented to create a family
of split-protein sensors,\textsuperscript{4,8} including dihydrofolate reductase (DHFR),\textsuperscript{9} the Green Fluorescent Protein (GFP) and its analogues,\textsuperscript{11,18,19} β-lactamase,\textsuperscript{20} and firefly,\textsuperscript{21} Renilla,\textsuperscript{10} and Gaussia\textsuperscript{17} luciferases (Figure 1.1b, see Table 1.1 for a comprehensive list of currently developed split-protein sensors).

One appealing aspect of split-protein reassembly assays is their use in both the detection of biomolecular interactions as well as their potential for the study of dynamic or reversible interactions. For example, assays utilizing GFP are generally considered to be irreversible under native conditions, allowing the trapping and visualization of rare or weak interactions.\textsuperscript{22} Though useful for potentially trapping transient intermediates, this aspect of the split-GFP reassembly assay cannot provide a window for observing dynamic interactions inside a cell. Compared to split-GFP, other split-protein reporters are generally reversible, such as β-lactamase along with firefly, gaussia, and Renilla luciferases\textsuperscript{23} and can be utilized for measuring stronger but dynamic interactions. Unlike GFP-based assays, some of these reversible reporters are well suited to immediately report upon absence or presence of protein-protein interactions that occur within the time-frame of seconds. Furthermore, a number of split-protein assays have also been developed which rely upon the post-translational modification of target proteins. For example, split-protein assays have been utilized to monitor phosphorylation-dependent protein-protein interactions as well as for the creation of kinase activity sensors.\textsuperscript{14,24,25} Additionally, split-protein reassembly has been used to monitor post-translational protein splicing.\textsuperscript{26}
Figure 1.1 Principle behind split-protein reassembly. a) Rational dissection of a reporter protein (R) into two inactive fragments (F1_R and F2_R) followed by fusion to proteins of interest A and B. Hetero-dimerization of proteins A and B facilitates reassembly of the split reporter allowing the detection of the interaction between A and B. b) Structures of the previously described split-protein reporters ubiquitin, dihydrofolate reductase (DHFR), green fluorescent protein (GFP), β-lactamase, and firefly luciferase. N- and C-terminal protein fragments are shown in blue and red respectively with the overlapping portions of firefly luciferase indicated in yellow.
<table>
<thead>
<tr>
<th>Split-Protein Reporter</th>
<th>Organisms Tested</th>
<th>Detected Readout</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dehydrofolate Reductase</td>
<td><em>E. coli, S. cerevisiae</em>, plants, mammalian cells</td>
<td>- Resistance to trimethoprim or methotrexate</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Complements in <em>Dhf</em> knockout</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Binding to fluorescent molecule</td>
</tr>
<tr>
<td>Green Fluorescent Protien</td>
<td><em>In vitro, E. coli, S. cerevisiae</em>, plants, <em>C. elegans</em>, <em>Drosophila melanogaster</em>, mammalian cells</td>
<td>- Fluorescence</td>
</tr>
<tr>
<td><em>(Acquoria jellyfish)</em> and variants</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Monomeric Red Fluorescent Protein</td>
<td><em>E. coli, plants</em></td>
<td>- Fluorescence</td>
</tr>
<tr>
<td><em>(Discosoma coral)</em></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TEM1 β-Lactamase <em>(E. coli)</em></td>
<td><em>In vitro, E. coli</em>, mammalian cells</td>
<td>- Resistance to ampicillin</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Catalysis with colored or fluorescent products</td>
</tr>
<tr>
<td>Firefly, <em>Gaussia</em>, and <em>Renilla Luciferase</em></td>
<td><em>In vitro, mammalian cells, Drosophila melanogaster, Musculus musculus</em></td>
<td>- Bioluminescence</td>
</tr>
<tr>
<td>Hygromycin B phosphotransferase</td>
<td>Mammalian cells</td>
<td>- Resistance to hygromycin B</td>
</tr>
<tr>
<td><em>(Streptomyces hygroscopicus)</em></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N-(5'-phosphoribosyl)-anthranilate isomerase <em>(Trp1p)</em> <em>(S. cerevisiae)</em></td>
<td><em>S. cerevisiae</em></td>
<td>- Complements in <em>Trp1</em> knockout</td>
</tr>
<tr>
<td>Glycinamide ribonucleotide transformylase</td>
<td><em>E. coli</em></td>
<td>- Complements in GAR transformylase knockout</td>
</tr>
<tr>
<td><em>(GAR) transformylase</em> <em>(E. coli)</em></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aminoglycoside phosphotransferase</td>
<td><em>E. coli, mammalian cells</em></td>
<td>- Resistance to neomycin</td>
</tr>
<tr>
<td><em>(Streptomyces fradiae)</em></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CRE-recombinase <em>(bacteriophage P1)</em></td>
<td>Mammalian cells</td>
<td>- chemically induced gene knockout</td>
</tr>
<tr>
<td>Ubiquitin</td>
<td><em>S. cerevisiae</em></td>
<td>- Reporter gene activation</td>
</tr>
<tr>
<td>Adenylate cyclase CyA <em>(Bordetella pertussis)</em></td>
<td><em>E. coli</em></td>
<td>- Reporter gene activation</td>
</tr>
<tr>
<td>Nla tobacco etch virus <em>(TEV)</em> protease</td>
<td>Mammalian cells</td>
<td>- Reporter gene activation</td>
</tr>
</tbody>
</table>

Table 1.1 A list of currently developed split-protein sensors.
1.3 Applications of Split-Protein Reassembly

Given the broad range of activities and possible readouts (absorbance, fluorescence, and luminescence) afforded by split-protein reporters as well as the ability to rescue cellular growth, this class of sensors has been widely utilized to study different aspects of protein function as well as in the creation of new interacting proteins. I describe some of these applications in the following sections.

1.3.1 Application of Split-Protein Reassembly Towards Protein Engineering

The ability to design and predict protein interactions has been a major focus of protein science. Towards this end a variety of techniques have been developed which allow for the “selection” and identification of new proteins and peptides that can recognize a particular protein or other biopolymer target.27-29 Perhaps the most often utilized technique to achieve the above goal has been phage display, which allows one to link genotypic information to the output protein of interest. Though effective, phage-display currently lacks the ability to evolve proteins with desirable in vivo characteristics in their native milieu inside a cell or for co-evolving both binding partners in an interaction pair simultaneously. One of the first practical applications of split-protein reassembly sought to address this problem through the in vivo selection of optimized heterodimeric leucine-zipper pairs by Pelletier et al. utilizing a split-mDHFR survival assay in E. coli. This approach allowed for the simultaneous screening of two semi-randomized libraries of complementary designed leucine-zipper peptides resulting in a testable library size reaching 1 x 10^6 interactions pairs30 (Figure 1.2). Term a library-versus-library selection, this work lead to the identification of an optimized leucine-
Figure 1.2. Split-mDHFR library-versus-library selection. DNA constructs code for fragments of mDHFR fusions to coiled-coil libraries A or B. Plasmids DNA corresponding to library fusion constructs are cotransformed into *E. coli* and plated on selective media. Cells containing library members which result in favorable binding result in cell survival while unfavorable interactions fail to grow on selective media.
zipper pair found not only to be more stable than the parent pair, but also possessed increased hetero- versus homo-specificity.\textsuperscript{30,31} 

Though an effective reporter for the detection of protein-protein interactions the split-mDHFR sensor is biased towards strong interactions. While this bias is useful for the selection of optimal binding partners, this selection platform may be ineffective when it comes to the interrogation of weaker interactions. In a similar library-based selection Magliery \textit{et. al.} applied a split version of the GFP variant sg100 to the selection of heterodimeric leucine-zipper pairs of varying interaction strengths. Unlike split-mDHFR, which allows the selection of interactions within the nanomolar regime, the reassembly of split-GFP is essentially irreversible under native conditions, allowing for the trapping and identification of interactions possessing binding constants as weak as 1 mM.\textsuperscript{22} 

In addition to providing a method for the selection and characterization of optimized binding pairs, split-protein reassembly methods also lend themselves very well to the screening of protein binders against therapeutically relevant targets. Typically, many selections of this nature are performed utilizing a bait and prey strategy (Figure 1.3). In general, the target protein against which binders are selected (the bait protein), is expressed as a fusion to one fragment of a split-reporter while the potential binder or library of binders is expressed as a fusion to the second fragment of the split-reporter. Favorable bait and prey interactions result in reassembly of the split-reporter and are identified either by fluorescence or cell growth on selective media. One example of this approach was the use of a split-mDHFR cell survival selection strategy in conjunction with ribosome display, a more traditional \textit{in vitro} selection method. This recent selection, described by Amstutz \textit{et. al.} utilized a protein scaffold comprised of designed ankryn-
Figure 1.3 Diagram of the bait and prey selection strategy. A bait vector is created by fusing a protein of interest to one fragment of a split-reporter while a prey protein or library of bait protein binders is fused to the other fragment of the split-reporter. Co-transfection of bait and prey proteins results in a population of cells co-expressing both the bait protein and prey protein library members. Prey proteins capable of bait protein binding result in the reassembly of an active split-reporter resulting in either fluorescence or growth on selective media.
repeat proteins (DARPins) that selectively bind specific MAP kinase isoforms with high affinity. This approach, which combines the best aspects of these two selection strategies resulted in the rapid enrichment of high affinity binders specific for c-Jun N-terminal kinase 2 (JNK2) over JNK1, a related MAP kinase sharing over 80% sequence identity.

The exploitation of antibody-antigen interactions has been an invaluable method for the study of proteins and other gene products. Typically, antibodies are developed using either animal models or by more recently described display methods utilizing available scaffold libraries from which binders are selected against an immobilized target. Though effective, these techniques have a number of drawbacks, including the inherent problems associated with the use of animals, the need for purified and stable antigen molecules \textit{in vitro}, and the selection of antigen binders within a non-native context. Recently split-protein reassembly utilizing split-mDHFR and split-\(\beta\)-lactamase have been utilized to engineer and select highly specific and stable antibodies. In one example, the sole use of a split-mDHFR selection resulted in well-expressed and stable antibodies against bacteriophage lambda capsid protein D (gpD) and JNK2. Significantly, the resulting antibodies were capable of high expression in \textit{E. coli} and were able to retain antigen binding regardless of redox state, showing tolerance to loss of disulfide bridge formation. On the other hand, though functional, the resulting antibodies possessed only moderate binding affinities (30 – 100 \(\mu\)M). To alleviate this problem a similar split-protein based selection utilizing a split-\(\beta\)-lactamase survival assay was utilized in conjunction with a pre-selected phage display library to identify specific antibodies against the intracellular portion of the human tyrosine-kinase receptor RON. In
addition to being stable and capable of expression in *E. coli*, the resulting antibodies were also capable of high affinity binding to their antigen with binding affinities ranging from 5-60 nM.

1.3.2 Examining Protein-Protein Interactions and Protein Stability

Instead of identifying members of a randomized library for specific and high affinity binding, split-protein reassembly can also be used to examine the structural requirements of a particular protein-protein interaction of interest. The ability to selectively manipulate specific positions of interest or segments of sequence is a powerful tool in the study of residue specific contributions to protein stability and interaction strength. Combination of these capabilities with split-protein reassembly provides a rapid and powerful way to interrogate the individual contributions of specific residues to a given complex. This approach was utilized recently to identify functional determinants of the interaction between the peroxisome proliferative activated receptor, gamma (PPARγ), with two LXXLL motif containing peptides derived from the nuclear receptor co-activators SRC1 and TRAP220. A modified split-mDHFR selection in conjunction with shotgun alanine scanning of 14 specific interface residues was used to identify six crucial interface residues. Further analysis using a GFP-based screen for intracellular stability identified that five of these six residues were critical determinants of PPARγ stability and when combined with the information obtained from the split-protein analysis provides a striking example of how split-protein reassembly can serve as an important component of a biochemical toolbox.
The study of cancer related mutations with respect to structural and interaction stability has provided a wealth of information in regards to mechanisms of cancer progression. With the intent of conducting an exhaustive study of interaction determinants Sarkar and Magliery have applied split-protein reassembly to the study of known cancer mutations within the BARD1/BRCA1 interaction interface. Utilizing a split-GFP sensor an initial panel of six known BRCA1 mutations identified by various in vitro and in vivo methods was used to assess the ability of the split-GFP sensor to report upon the difference between mutations responsible for interaction versus BRCA1 structural stability. Of the six mutations examined, three were found to show no reduction in interaction stability while three were found to either reduce or abrogate BARD1 binding by BRCA1. Additional pull-down studies revealed that two mutations showing no split-GFP reassembly resulted in background levels of soluble protein, suggesting BRCA1 structural instability as a plausible mechanism for BARD1/BRCA1 interaction abrogation. These results contradict previous yeast two-hybrid assays which suggest the ability of these same interactions to allow BARD1 binding by mutated BRCA1. Taken together with other in vitro and in vivo methods, this suggests differing oncogenic mechanisms for the tested mutations such as interfering with E2 ligase complex formation. These studies also highlight that data from yeast two-hybrids, if combined with data from split-protein approaches, can be of general value for dissecting interactions within a cellular context.

Another application of split-protein reassembly has been in the study of protein topology and sequence determinants of protein folding. Though methods such as multiple sequence alignments (MSA) exist which allow correlation of structural
similarities to sequence diversity, they lack breadth due to the limited number of selected folds for which an adequate number of divergent sequences are available. An alternative to MSA would be the massive perturbation of a given sequence followed by analysis to determine the overall structural or functional effects. To achieve this, split-protein reassembly in the form of an *E. coli*-based split-mDHFR selection has been applied to study the folding of the *ras* binding domain (RBD) of the Ser/Thr kinase Raf in response to systematic randomization of short contiguous segments of sequence. The use of split-protein reassembly allowed for the rapid and high-throughput determination of structural integrity by monitoring the ability of mutant RBDs to bind the protein target *h-ras* (Figure 1.4). The resulting dataset revealed observations of structural diversity that adequately approximates the sequence space observed with traditional MSA analysis, demonstrating the ability of split-protein reassembly approaches to complement MSA methods and provide biophysical insight in the absence of appropriate structural information.

### 1.3.3 High Throughput Screening of Protein-Protein Interactions

The elucidation of a protein’s interaction partners, and thus, the construction of cellular regulatory pathways is a central problem in biology. Towards this end a number of *in vitro* and *in vivo* techniques have been developed which allow the discovery of interaction partners of a given protein of interest. One such powerful method involves the use of tandem affinity purification coupled with mass-spectrometric analysis, a strategy in which target interactions are isolated through affinity chromatography followed by analysis and identification by mass spectrometry (Figure 1.5a). Though
1. Dissect Protein Sequence Into Segments

2. Randomize Each Segment Sequence

   e.g. S9: LHEHKGK → XXXXXX

3. Select For Functional Raf Mutants

**Figure 1.4.** Split-protein assay for determinants of protein folding. Strategy. (1) The h-ras-binding domain of Raf (Raf RBD) is subdivided into 13 segments based on topological elements. (2) Each segment is separately randomized creating 13 separate libraries of Raf RBD mutants. (3) Each library is evaluated for functional (folded) Raf RBD mutants utilizing a split-mDHFR survival assay. Those mutants capable of properly folding and binding Ras result in active DHFR and rescue cell growth of selective media.
effective at identifying stable protein complexes, this technique lacks the ability to interrogate protein-protein interactions within their native cellular environment and limits analysis to relatively strong interactions that survive in vitro purification. On the other hand, many investigators interested in interrogating the protein interactome in a cellular context have utilized the powerful yeast two-hybrid techniques, which have the ability to link interactions to gene expression through the fusion of an interaction pair to a DNA-binding domain (DBD) and an RNA polymerase activation domain (Pol-AD), resulting in the formation of an interaction-dependent artificial transcription factor⁴² (Figure 1.5b). Unlike tandem affinity chromatography, this methodology allows the interrogation of direct binary protein-protein interactions within a cellular context. However, like affinity purification, yeast two-hybrid methods also possess some inherent drawbacks, namely the inability to directly report upon transiently associating protein complexes. Additionally, the requirement of nuclear localization for signal output sometimes biases analysis to protein complexes that are capable of facile nuclear translocation. The use of split-protein reassembly or protein complementation provides an alternative solution to the interrogation of protein-protein interactions within a cellular context where both strong and weak interactions can be measured and visualized with temporal control. Additionally, some split-protein reassembly survival assays, such as those utilizing split-DHFR and split-ubiquitin, are readily applicable to the large-scale and rapid screening of potential protein-protein interaction pairs.

The most exhaustive examples of this type of large-scale analysis have been disclosed for the yeast Saccharomyces cerevisiae, a model eukaryotic organism. One example utilized a split-ubiquitin based two-hybrid yeast survival-selection assay
Figure 1.5. Previously described methods for the identification of protein-protein interactions. a) Tandem affinity purification (TAP) makes use of two consecutive purification strategies to isolate interacting proteins followed by analysis of mass spectrometry. TAP scheme: A TAP tag consisting of *Staphylococcus aureus* Protein A (white) and the calmodulin-binding protein (light blue) is fused to a target protein of interest (dark blue) which is bound to interacting proteins (red and green). A first affinity purification is performed followed by cleavage of Protein A via a protease cleavage site. After cleavage a second purification step is performed to eliminate contaminating proteins (yellow) followed by one-dimensional SDS-PAGE to separate target and interacting proteins. Interacting proteins are then identified by mass spectrometry and database analysis. b) Two-hybrid method for the identification of interacting protein pairs. Fusions are made between proteins of interest A and B with either a DNA-Binding Domain (DBD) or an RNA polymerase activation domain (Pol-AD). Interaction between A and B results in the formation of an interaction dependent transcription factor, facilitating the recruitment of RNA polymerase (RNA Pol) and transcription of a reporter gene.
identifying interactions between integral yeast membrane proteins.\textsuperscript{43} To achieve this, two libraries were constructed in which ORFs of identified yeast integral membrane proteins were fused to either the N-terminal or C-terminal half of ubiquitin. To facilitate interaction-dependent survival, libraries containing C-terminal ubiquitin fusions were also attached to the artificial transcription factor PLV. The interaction of two membrane proteins results in the reassembly of ubiquitin and facilitates the cleavage and release of PLV by ubiquitin-specific proteases. This release allows localization of the PLV transcription factor to the nucleus resulting in transcriptional activation of a \textit{HIS3} reporter gene (Figure 1.6). Using yeast mating strategies in a semiautomatic plating system Miller \textit{et. al.} used this split-ubiquitin strategy to screen approximately 700 integral yeast membrane proteins against each other in an array based format.\textsuperscript{44} Of the 705 proteins tested, 1,985 putative interactions were found among a total of 536 proteins involved in a myriad of biological processes including ion, lipid, and protein transport, vesicle-mediated transport, and protein and lipid metabolism, of which only 34 (1.7\%) had been previously reported in literature.

Though only focusing on membrane associated proteins, the above study demonstrates the ability of split-protein reassembly to address complex biological problems. By eliminating the need for protein purification and by removing the requirement for nuclear localization, split-protein reassembly enables the rapid and high-throughput analysis of large sets of protein-protein interactions. This point has been recently demonstrated in a study utilizing a split-mDHFR survival-selection assay for the genome-wide \textit{in vivo} screening of 5367 \textit{Saccharomyces cerevisiae} ORFs.\textsuperscript{45} In all, 2770 interactions among 1124 endogenously expressed proteins were identified, approximately
Figure 1.6. Split-ubiquitin strategy for the detection of membrane protein interactions. 

a) Two integral membrane protein (red and dark blue) are fused to two halves of a split-ubiquitin reporter, the C-terminal portion of which is fused to the artificial transcription factor PLV. b) Association of the two integral membrane proteins results in the reassembly of ubiquitin resulting in the release of PLV from the membrane followed by translocation to the nucleus. Upon nuclear translocation PLV stimulates transcription of the HIS3 gene facilitating cell growth on selective media.
80% of which were previously unreported. Additionally, 286 (10.3%) interactions were identified which involved one previously uncharacterized protein with a protein of known function or between two previously unknown proteins. In order to be a truly useful and informative methodology, it is necessary to accurately report interactions previously demonstrated by alternative methods. Of the interactions identified in this study utilizing split-mDHFR, 16% were previously observed by affinity purification while 41% were observed using traditional two-hybrid analysis, confirming the ability of split-protein reassembly to accurately confirm and complement existing strategies for the large-scale analysis of protein-protein interactions.45

The above studies represent the large-scale application of split-protein reassembly towards the elucidation of protein interaction networks of any given organism, demonstrating the power of the application of split-protein reassembly methodologies toward solving complex biological problems. In addition to these studies, split-protein reassembly has been used in various forms toward the more focused analysis of distinct biochemical pathways. This includes the use of the split-ubiquitin sensor for the study of the ergosterol synthesis pathway,46 peroxisome,47 endosome,48 and human serpentine receptors in yeast,49 K+ channels in Arabidopsis thaliana,50 and the use of the split-mDHFR sensor for the study of the receptor tyrosine kinase (RTK) pathway in Chinese hamster ovarian (CHO) cells.51

Recently, a split version of the intensely fluorescent protein known as Venus was used in conjunction with the enhanced retroviral mutagen (ERM) vector to interrogate binding partners of the key survival kinase AKT1 in HeLa cells.52 Termed RePCA, this technology applies a bait and prey strategy to the cell specific screening of cDNA
libraries against a bait protein of interest. In general, a stably transfected cell line was created harboring the split-Venus-bait fusion of interest, followed by ERM vector infection to create a library of split-Venus-prey fusions from endogenous genes (Figure 1.7). The advantage of this strategy over conventional split-protein assays is that by utilizing the ERM vector one is able to create a library of split-Venus fusions derived from the genome of the host organism of interest, preserving endogenous folding, post-translational, and alternative splicing effects on the interactions being interrogated. From this analysis, 24 potential interaction partners or substrates of AKT1 were identified, including the previously uncharacterized interaction between α-actinin 4 (ACTN4), potentially linking the PI3K/AKT pathway with motility and invasion which may serve as an underling contributor to focal segmental glomerulosclerosis.

1.3.4 Protein Interactions with Small-Molecules

The interactions of small-molecules with proteins are involved in a variety of biological processes. These include acting as second messengers and critical components of signal transduction pathways, forming direct complexes with enzymes, and serving as cofactors in reaction catalysis. From a disease perspective, small-molecule pharmaceuticals that target proteins represent the backbone of therapeutic strategies targeting aberrant enzyme activity and more recently protein-protein interactions. A number of methods have been developed to interrogate the binding and effects of small-molecules, including affinity chromatography, fluorescence polarization, surface-plasmon resonance, and a myriad of classic biochemical activity assays. Aside from
Figure 1.7. Diagram of the RePCA screen. A stably transfected cell line is created containing one half of a split-reporter protein (C-Venus) fused to a bait protein of interest under control of an exogenous promoter followed by infected with a retrovirus containing an enhanced retroviral mutagen vector. Cells infected with the ERM vector result in the generation of split-reporter fragments fused to endogenous host genes, resulting in a prey library consisting endogenous host proteins. Positive interactions can then be identified via fluorescence.
these techniques, the use of split-protein reassembly-based approaches has become increasingly useful to the study of this important class of interactions.

Given the focus of research upon the creation and implementation of small-molecule inhibitors of biological function, an emerging concern has been the effect of these molecules on biological processes beyond the targets for which they were initially developed. Though effective, the mechanism by which small-molecule drugs affect their therapeutic properties is often unknown, and in some cases this can be particularly detrimental. One example of this can be seen with the drug thalidomide. Originally used as a painkiller and antiemetic agent, it also has been found to have a broad spectrum of anticancer as well as teratogenic properties. To avoid deleterious effects, discover new beneficial effects, and to aid in the elucidation of drug mechanisms, methods are needed which provide the ability to scrutinize the effects of small-molecules on specific biological pathways and the specific components of which they are comprised. Given their previously demonstrated ability to map biochemical pathways and direct analysis to specific interactions within those pathways, split-protein reassembly methods provide an effective tool for the analysis of small-molecule effects on specific biochemical pathways and for the discovery of potentially hidden effects of small-molecule drugs (Figure 1.8).

This application of split-protein reassembly for uncovering hidden phenotypes was recently demonstrated by MacDonald et al. in which split-protein reassembly was used to observe the effects of small-molecule drugs was interrogated against a variety of signal transduction pathways, including insulin, AKT, growth-factor dependent, erythropoietin, TGF-β, and TNF signaling pathways. In this study a split-protein
Figure 1.8. Schematic of downstream drug effects. Components of two pathways (A and B) interact with a third pathway C, facilitating the interaction between components C1 and C2. A drug acting upon component B2 results in the downstream disruption of the interaction between components C1 and C2. Split-protein reassembly assays can be used to interrogate the interactions between various components of these pathways, allowing the determination of downstream or off target effects of a drug on a particular pathway.
reassembly assay based upon the GFP variant Venus and comprising 49 different known interactions was used to screen 107 different drugs for hidden effects or ‘phenotypes’. These assays encompassed known interactions involved in ten cellular processes including cell cycle, apoptosis, mitogenesis, proteolysis, chaperone, cytoskeletal, and receptor mediated pathways. Of the drugs screened, six diverse therapeutic areas were represented, namely cancer, inflammation, cardiovascular disease, diabetes, infectious disease, and neurological disorders, resulting in 127 different measurements for each compound. Utilizing this approach, both expected and unexpected drug effects were found, including four compounds possessing previously undiscovered anti-proliferative activity. Of the 49 assay pairs tested, a 25 pair subset was identified that possibly provide a predictive assay for anti-proliferative drug phenotypes in addition to individual assays which correlate to underlying anti-proliferative mechanisms such as disruption of mitochondrial membrane potential. The results of this study suggest a number of applications of this approach towards understanding polypharmacology. Aside from predicting anti-proliferative drug candidates, the utilization of split protein-based assays focused towards other cellular processes and disease related pathways could provide a means for the discovery of new drugs and new drug applications in addition to providing a means for elucidating therapeutic mechanisms of action.\footnote{53}

The above study represents the large-scale determination of the effects of small-molecules on complex biological pathways and processes. In addition to these types of analyses, split-protein reassembly has also been applied to the study of more specific protein-small-molecule interactions, such as the conformational effects of small-molecules on their protein targets. Small-molecule induced conformational change is a
common mechanism for the regulation of protein function. For example, the binding of specific steroid hormones to the ligand binding domain (LBD) of the estrogen receptor (ER) leads to its dimerization and subsequent binding, through the association of co-activator proteins, to estrogen response elements present in a number of ER target genes.\textsuperscript{54} Crystallographic studies have revealed very specific conformational changes within the ER-LBD in response to binding to agonist and antagonist molecules.\textsuperscript{55-57} Utilizing this structural information, Paulmurugan and Gambhir created a split-luciferase intramolecular folding sensor, which is conceptually related to the fragment based intermolecular approaches, capable of reporting upon ER-LBD binding to various types of ER ligands.\textsuperscript{58} The split-intramolecular sensor placed two halves of a split-\textit{Renilla} luciferase reporter on the N- and C-termini of the ER-LBD, placing them in a conformation resulting in inefficient luciferase reassembly in the absence of ligand binding. Upon ligand binding, alternative ER-LBD conformations are adopted based upon the type of ligand bound, i.e. agonist or antagonist (Figure 1.9). As a result of these ligand-specific conformational changes, varying levels of split-luciferase reassembly were observed, resulting in either high (agonists), low (antagonists), or no luciferase signal. This sensor was capable of accurately distinguishing ER ligand pharmacology for a variety of ligand types in cell lysates, whole cells, and in living animals, demonstrating its use as both a means for high throughput screening of potential ER targeting drugs, and as a tool for the study of ER targeting drugs in living animal model systems. The approach described above demonstrates an interesting strategy for the study of small molecule modulators of protein function and has also been applied to the development of a dual-color ER sensor enabling the study of distinct regulatory mechanisms of action of
Figure 1.9. Model of the ER-LBD split-rLuc intramolecular folding sensor. a) The ER-LBD conformation induced by antagonist binding results in a orientation of split-rLuc fragments which favors complementation resulting in high levels of *Renilla* luciferase activity. b) The ER-LBD conformation induced by agonist binding results in a suboptimal orientation for split-rLuc reassembly resulting in low levels of *Renilla* luciferase activity.
various ER ligands. Additionally, this approach has been applied to the development of a high throughput screen for modulators of Abl kinase activity.

Though useful, the above approach is inherently limited to a subset of interactions which result in a large interaction dependent conformational change, as also necessary for intramolecular FRET and BRET based approaches. More generalized strategies for the interrogation of protein-small-molecule interactions have been developed based upon previously described yeast two-hybrid technologies through the utilization of small-molecule chemical inducers of dimerization (CIDs). Dubbed three-hybrid technologies, these approaches utilize a synthetic bifunctional molecule capable of simultaneously binding two protein moieties fused to DNA and transcriptional activation domains, akin to the two-hybrid system described above, resulting in the activation of a specific reporter gene. Though this technology has proven extremely useful for the proteome-wide screening of potential targets of kinase inhibitors and the development of orthogonal logic gate technologies, it is still limited by the requirement of nuclear localization for readout production and often suffers from false positives and negatives.

Recently, a split-protein reassembly-based variation of this three-hybrid strategy was developed utilizing a split-protein reporter (Figure 1.10). In a proof-of-principle study, Dirnberger et al. successfully reported the creation of a three-hybrid system utilizing fusions of split-ubiquitin to glucocorticoid receptor (GR) and DHFR moieties and monitored for growth of selective media. Introduction of a dexamethosone-methotrexate CID resulted in a specific three-hybrid signal with sensitivities as high as 748-fold signal to noise under optimized conditions. To further demonstrate the applicability of this approach a similar three-hybrid system was constructed utilizing the
Figure 1.10. Schematic of the CID three-hybrid strategy. Two small-molecule binding proteins (A and B) are fused to two inactive fragments of a split-reporter protein. Addition of a synthetic bifunctional chemical inducer of dimerization (CID) results in small-molecule binding by proteins A and B, facilitating the reassembly of the attached split-reporter protein.
human kinase PCTK3 in conjunction with a CID consisting of methotrexate and the kinase inhibitor purvalanol B. As with the control system described above, a specific three-hybrid response was observed upon the addition of the methotrexate-purvalanol B CID, demonstrating the use of this technology for the interrogation of pharmaceutically relevant protein-small-molecule interaction pairs. Not limited to the split-ubiquitin sensor, this three-hybrid approach has also been demonstrated for a number of other split-protein reporters as observed by the well characterized dimerization of FRB and FKBP induced by the small-molecule rapamycin.14,17,20,65,66

1.3.5 Detection of Protein-Nucleic Acid Interactions

In addition to the interaction between two or more proteins, the interactions between proteins and their nucleic acid ligands is an important regulatory mechanism of biological processes.1 Though traditional techniques such as electrophoretic mobility shift assays (EMSAs) allow the interrogation of these interactions, they are limited to the study of proteins amenable to purification and provide information in an optimized in vitro setting. The application of split-protein reassembly-based techniques to the study of protein-nucleic acid interactions has become a powerful tool for the in vivo analysis of these interactions and has also made possible a number of biotechnical and biomedical applications.

An early application of split-protein reassembly to the study of protein-nucleic acid interactions was the visualization of ternary protein-RNA complexes. Utilizing a split-Venus sensor that they termed trimolecular fluorescence complementation (TriFC), Rackham and Brown found that the fragile X mental retardation protein (FMRP) and the
RNA transport factor IMP1 co-localize on common mRNAs resulting in granule formation, suggesting a link between mRNA transport and translational repression. In a similar study, split-protein reassembly was used to visualize the co-localization of nuclear export factor 1 (NXF1) and the splicing factor Y14 on common mRNAs. Additionally, in vivo imaging of these complexes revealed accumulation within and around nuclear speckles, suggesting their involvement in mRNA processing and regulation of gene expression.

In addition to the study of specific protein-RNA complexes, general methods have also been developed, which utilize two proteins capable of recognizing specific RNA motifs, for the detection and analysis of modified mRNA molecules (Figure 1.11). As a proof-of-concept for their TriFC methodology, Rackham and Brown developed a general strategy which allows the visualization of specific mRNA molecules of interest. This was accomplished through the utilization of specific structured RNAs placed within the 3’UTR of a synthetic mRNA construct. To achieve this, the MS2 coat protein-binding motif and a “zip code”-binding motif derived from the 3’ UTR of β-actin mRNA were introduced into an artificial mRNA construct. Coexpression of this mRNA with split-Venus fusions to the MS2 coat protein and a zip code-binding protein facilitated Venus reassembly, resulting in a strong fluorescence signal in COS-7 cells expressing wild-type binding sequences over those harboring mutant RNA constructs. Recently this approach was simplified by the incorporation of an eIF4A binding RNA aptamer into the 3’ end of target RNAs. Structural and experimental studies showed that dissection of the eIF4A protein resulted in two domains which alone are incapable of effective binding, but together are capable of simultaneous binding to the eIF4A aptamer. Fusion of these
Figure 1.11. A general strategy for the visualization of mRNA. Specific RNA motif-binding proteins A and B are fused to inactive fragments of a split-reporter protein. Simultaneous binding by A and B to a synthetic mRNA containing the recognized RNA motif facilitates reassembly of the attached split-reporter protein fragments.
two domains to inactive halves of split-EGFP resulted in EGFP reassembly in vivo within live bacterial cells upon the transcription of LacZ mRNA and 5S sRNA containing repeats of the eIF4A aptamer. In addition to observing enhanced fluorescence, fluctuations in mRNA levels as well as variances in mRNA localization were also correlated with observed fluorescence.70

The use of proteins capable of recognizing specific RNA structures marks a significant advance in the ability to monitor specific RNA molecules, but is limited to the detection of artificial mRNA molecules introduced into model systems. In order to enable the detection of endogenously produced nucleic acids, split-protein reassembly methods have been developed utilizing truly sequence specific binding domains which allow the targeting of virtually any nucleic acid target. Termed SEquence-Enabled Reassembly (SEER), this application of split-protein reassembly was first demonstrated with the direct in vitro detection of specific dsDNA sequences.71 This was achieved through the use of specific dsDNA-binding proteins fused to two inactive halves of the split-reporter protein. In principle, simultaneous recognition and binding of cognate dsDNA sequences by both split-reporter fusions facilitates the reassembly of the reporter in a sequence dependent manner (Figure 1.12). Utilizing purified fusion proteins, significant fluorescence was observed upon incubation of split-GFP-zinc finger fusions with a dsDNA target harboring specific sequences recognized by the dsDNA-binding zinc finger proteins Zif268 and PBSII.71 Utilizing this and mutant dsDNA targets, fluorescence was shown to be concentration dependent and sequence specific, with the elimination of either zinc finger binding site resulting in no observed fluorescence. Though these initial experiments were performed using purified components under
**Figure 1.12.** General schematic of the SEER strategy. Specific dsDNA-binding proteins A and B and fused to inactive fragments of a split-reporter protein. Simultaneous binding by A and B of a specific dsDNA sequence facilitates reassembly of the attached split-reporter protein fragments.
dialysis dependent refolding conditions, supplementary SEER systems have been devised utilizing soluble zinc finger-β-lactamase\textsuperscript{72} fusions \textit{in vitro} in addition to split-firefly and split-\textit{Gaussia} luciferase reporters.\textsuperscript{73}

In addition to dsDNA, split-protein reassembly has also been used to detect specific sequence of RNA inside living cells through the use of the RNA-binding human Pumilio homology domain (PUM-HD). Utilizing known design strategies this protein can be engineered to alter its sequence specificity in a predictable manner.\textsuperscript{74} Using these methods, Ozawa \textit{et. al.} created two PUM-HDs capable of recognizing and binding two adjacent eight-nucleotide stretches of an endogenous mtRNA encoding mitochondrial NADH dehydrogenase subunit 6 (\textit{MT-ND6}). Fusions of these mutant PUM-HDs to split-Venus were used to detect \textit{MT-ND6} mtRNA in the mitochondria of HeLa cells as well as image mtRNA dynamics and monitor mtRNA dispersion and degradation in response to oxidative stress.\textsuperscript{75}

\textbf{1.3.6 Split-Protein Reassembly in Model Organisms}

Numerous studies have shown that split-protein reassembly can be used in virtually any cell line or organism in which reporter constructs can be introduced by plasmid or viral vectors, or through successful genomic incorporation. Such technologies have enabled the use of split-protein reassembly in a whole host of living organisms, demonstrating the broad applicability and potential of its use. For example, implanted cells harboring split-luciferase sensors have been used in mice to monitor a variety of protein-protein interactions as well as visualize small-molecule protein interactions.\textsuperscript{4,76}
the responses of protein interactions to drugs, and nuclear transport of proteins within the brains of live mice.

The nematode worm *Caenorhabditis elegans* and the fruit fly *Drosophila melanogaster* are attractive model organisms due in part to their small size, rapid development, and the ease through which their genomes can be manipulated. A particularly attractive feature of *C. elegans* is the transparent nature of its body, allowing the direct visualization of fluorescent and bioluminescent reporters. For example, Zhang *et al.* demonstrated the use of split-GFP, YFP, and CFP sensors in *C. elegans* to identify subsets of cells displaying overlapping promoter activity. Exploiting the binary nature of split-protein sensors, this study breaks from conventional split-protein reassembly applications of interaction detection, to the study of gene co-expression, the monitoring of changes in gene expression, and as a means to target specific cells for finer genetic analysis. In addition to split-GFP and its variants, a follow-up study utilizing a binary leucine-zipper dependent caspase was used to affected the specific killing of neuronal cells displaying overlapping promoter expression.

Other studies utilizing split-protein sensors have also been performed in transgenic fruit flies. In one study Benton *et al.* utilized a split-YFP sensor to visualize interactions between the transmembrane odorant receptor OR83b and other odorant receptors in live fruit flies. In addition to detecting interactions, this study was further enhanced by the demonstration of unique odorant receptor membrane topology through the exploitation of the requirement of compartmental orientation for the effective reassembly of split-protein sensors. In a separate study utilizing transgenic fruit flies, Schwartz and coworkers demonstrated the modulation of protein post-translational
modifications through the use of chemically-induced protein trans-splicing to reconstitute a split-firefly luciferase sensor in live fruit flies upon the consumption of a small-molecule protein dimerizer.26

1.4 New Directions in Split-Protein Reassembly

In this introductory chapter, recent applications utilizing split-protein reassembly for the creation of new proteins and the study of protein function and their interaction with various types of binding partners have been described. In the following chapters I describe our concurrent efforts to build upon this work through the utilization of split-protein reassembly for the development of new methods enabling the study and interrogation of protein-protein, protein-nucleic acid, and protein-small molecule interactions. Additionally, I describe the utilization of split-protein reporters for the development of a cell-free assay capable of the rapid and high-throughput interrogation of various biomolecular interactions, the evaluation of inhibitors of these interactions, and finally for use in the development of new sensor technologies.
CHAPTER 2
APPLICATION OF SEER TO THE DETECTION OF TRANSGENES AND CYTOSINE METHYLATION

2.1 Introduction

Portions of the work described below are taken from the paper “Split β-Lactamase Sensor for the Sequence-Specific Detection of DNA Methylation,” which can be found in Appendix A. I was responsible for the application of our developed SEER methodology to the creation of a CryIA transgene sensor and the utilization of a split-β-lactamase for the site-specific detection of DNA methylation. The following contains a summary of the most important findings within this document.

The ability to detect specific sequences of dsDNA is of particular importance to a number of biological and biotechnical endeavors. To date, a variety of methods have been developed which enable the detection of specific DNA sequences including hybridization based methods such as PCR, Southern Blot, Northern Blot, and fluorescent in situ hybridization (FISH). Additionally, dsDNA can be detected through the use of synthetic sequence-specific binding agents such as triplex forming oligonucleotides (TFO) and designed \( N \)-methylpyrrole and \( N \)-methylimidazole-based molecules called polyamides. Alternatively, nature utilizes sequence specific DNA-binding proteins in the form of transcription factors to read specific DNA sequences as a means of gene regulation. Often, these DNA binding proteins require the recognition of two adjacent DNA sequences to facilitate binding and signal production. Utilizing this principle in conjunction with split-protein reassembly we have recently described a methodology for the creation of protein-based sensors for the direct detection of specific dsDNA sequences which we have termed SEquence-Enabled Reassembly (SEER). The SEER
approach generally consists of two sequence-specific DNA binding domains which are capable of reassembling two halves of a tethered split-reporter protein only in the presence of a specific dsDNA target of interest, resulting in a “turn-on” sensor capable of visually reporting upon the presence of specific dsDNA sequences. The SEER approach was first demonstrated by the fusion of two Cys2-His2 zinc finger proteins to two rationally dissected inactive fragments of a GFP reporter. Simultaneous binding of a target dsDNA oligonucleotide containing the sequences 5’-GCGTGGGCG-3’ and 5’-GTGTGGAAA-3’ by the zinc fingers Zif268 and PBSII respectively facilitated reassembly of the attached GFP fragments, resulting in dsDNA sequence dependent fluorescence (Figure 2.1).

Conceptually, the use of zinc fingers is particularly advantageous. These proteins consist of a number of tandem repeats of a \( \beta\beta\alpha \) domain which is stabilized by the coordination of a single zinc ion. The \( \alpha\)-helix of this domain binds the major groove of DNA and specifically recognizes a 3 base pair sequence through hydrogen bonding and Van der Waals contacts. The conjugation of three of these zinc finger domains allows the creation of zinc finger proteins capable of recognizing 9 contiguous bases. Importantly, the use of two of these zinc fingers proteins enables the recognition of 18 contiguous bases, allowing for the potential detection of unique DNA sequences within the human genome. Of additional significance is the elucidation of a partial DNA recognition code for zinc fingers which has resulted in the identification of zinc finger domains capable of targeting almost all 3 bp tracts (Figure 2.2). By utilizing this recognition code and employing developed design strategies, zinc finger proteins can be produced which are capable of binding virtually any dsDNA sequence of interest.
Figure 2.1. The SEER strategy. Initial SEquence-Enabled Reassembly (SEER) of GFP consisting of the fusion proteins NGFP(1-157)-Zif268 (cyan and blue) and PBSII-CGFP(158-238) (pink and red). Addition of a dsDNA target containing both Zif268 and PBSII binding sites results in the dsDNA dependent reassembly of split-GFP providing a “turn-on” signal in the presence of the target DNA sequence. Reprinted with permission from the American Chemical Society.
Figure 2.2. Modularity and design of DNA sequence specific zinc finger proteins. a) Representation of the modular nature of zinc finger proteins. Specific residues from individual zinc finger domains make specific contacts within the major groove of DNA, resulting in sequence specific binding. b) Experimentally determine recognition modules which can be incorporated into the α-helix of a zinc finger domain, enabling it to specific bind the indicated 5’ANN-3’, 5’-CNN-3’, or 5’-GNN-3’ DNA sequence.
Subsequently, the SEER methodology was amended to enable the detection of specific sequences of dsDNA by the split-enzymatic reporter β-lactamase. Unlike split-GFP, where in an ideal case we expect one reassembled GFP molecule per ternary complex, the enzymatic nature of split-β-lactamase results in the enzymatic turn-over of substrate upon dsDNA recognition. Thus split-β-lactamase could potentially allow for the detection of lower amounts of a particular DNA target of interest as compared to split-GFP. Having in hand a simple yet modular platform for the creation of sequence-specific DNA sensors, we sought to further test the utility of SEER with the creation of SEER-based sensors capable of detecting a specific naturally occurring DNA target in addition to the detection of DNA chemical modifications, namely the CryIA transgene and DNA methylation.

2.2 Creation and Evaluation of a CryIA-SEER-β-Lac Sensor

The CryIA proteins from Bacillus thuringiensis are pore-forming toxins and are particularly toxic to the larvae of insects. Because of this, CryIA proteins are widely used for the creation of transgenic crops, termed Bt-crops, as a means of producing agriculturally significant crops, such as tobacco, cotton, rice, and maize, with enhanced resistance to pests. Of particular concern is transgene escape, or the migration of transgenes from GM crops through cross-pollination with their wild form or related species, an event which could have undesirable ecological consequences. To monitor this, simple and effective methods are needed which allow the genomic detection of transgenes or the detection of their products. With this in mind, we sought to demonstrate the application of SEER by the creation of a sensor capable of specifically detecting the
CryIA transgene in Bt-maize, the presence of which is typically determined by ELISA-based assays which detect the protein products of the CryIA gene.

To specifically target the CryIA gene we chose to target the 18 base contiguous sequence 5’-GAGGGAGATGTCGATGGG-3’, which corresponds to nucleotides 1237-1254 of the CryIA(b) gene and is unique from the maize genome, a GM-crop often engineered to harbor the CryIA transgene. Utilizing existing design strategies (Figure 2.2b), the zinc finger proteins ZFC1-L3 and ZFC1-R3 were designed to target the 9-base sequences 5’-GAGGGAGAT-3’ and ‘5-GTCGATGGG-3’ and appended to the two inactive fragments of β-lactamase to yield LacA (26-196)-ZFC1-L3 and ZFC1-R3-LacB (198-290). Additionally, each fusion was created with a maltose binding protein affinity tag to ensure solubility and to allow purification by amylose affinity chromatography.

As an initial test of our designed CryIA-SEER-β-Lac sensor equimolar amounts (125 nM) of the two purified proteins LacA-ZFC1-L3 and ZFC1-R3-LacB were incubated in the presence (50 nM) or absence of the target dsDNA oligonucleotide L3-R3 containing the CryIA sequence of interest. Under these conditions a clear target dsDNA dependent signal of approximately 5-fold was observed in the presence of L3-R3 over samples containing no target dsDNA (Figure 2.3a). To determine the minimal amount of target dsDNA detectable under these conditions, equimolar amounts (125 nM) of LaA-ZFC1-L3 and ZFC1-R3-LacB were incubated with decreasing amounts of L3-R3 (figure 2.3b). These experiments show that 2.5 nM L3-R3 (500 fmols) is detectable above no DNA.
Figure 2.3. Initial evaluation of the CryIA SEER-β-lac sensor. a) β-lactamase activity of LaA-ZFC1-L3 and ZFC1-R3-LacB (125 nM each) in the presence or absence of 50 nM of the L3-R3 target dsDNA oligonucleotide. b) β-lactamase activity of LaA-ZFC1-L3 and ZFC1-R3-LacB (125 nM each) in the presence of decreasing amounts of the dsDNA target L3-R3. Only 2.5 nM (500 fmols) of DNA is detectable above no dsDNA target under these initial conditions.
2.2.1 Optimization of CryIA-SEER-β-Lac Sensor Conditions

To potentially decrease the amount of dsDNA detectable by our CryIA-SEER-β-Lac sensor, a series of experiments were performed to determine the optimal conditions for maximizing signal-to-noise. To determine this, 125 nM LacA-ZFC1-L3 was incubated with 50 nM L3-R3 and increasing amounts of ZFC1-R3-LacB (Figure 2.4a). These experiments show a ZFC1-R3-LacB dependent signal increase to a concentration of 500 nM, with a similar amount of activity observed at 1 μM. Additionally, β-lactamase activity in the absence of L3-R3 also increased as a function of ZFC1-R3-LacB concentration, with a maximal signal-to-noise of 5-fold being observed with a ratio of 1:4 LacA-ZFC1-L3 to ZFC1-R3-LacB. To further optimize detection conditions, a similar set of experiments were performed in which 125 nM of ZFC1-R3-LacB was incubated with 50 nM target and increasing concentrations of LacA-ZFC1-L3 (Figure 2.4b). Unlike experiments in which LacA-ZFC1-L3 was held constant, no appreciable increase in signal-to-noise was observed with increasing amounts of LacA-ZFC1-L3, suggesting no distinct advantage being gained by using increased concentrations of LacA-ZFC1-L3 over ZFC1-R3-LacB. From this series of experiments, it was determined that a ratio of 1:4 LacA-ZFC1-L3 to ZFC1-R3-LacB provided optimal conditions under which to detect our target dsDNA.

Having optimized conditions in hand, we next sought to determine the minimal amount of dsDNA we could detect with our CryIA-SEER-β-Lac SEER sensor. To determine this, 250 nM LacA-ZFC1-L3 was incubated with 1 μM ZFC1-R3-LacB and decreasing amounts of the dsDNA target L3-R3 (Figure 2.4c.). These experiments show
that 1 nM dsDNA (100 fmols) is detectable above background (0 nM) and that β-lactamase activity scales linearly with L3-R3 concentration (Figure 2.4d). Thus this set of experiments helps validate the application of the SEER approach to a real world problem.
Figure 2.4. Optimization of the *CryLA* SEER-β-lac sensor. a) β-lactamase activity as a function of ZFC1-R3-LacB concentration. 125 nM of LacA-ZFC1-L3 was incubated with 50 nM *L3-R3* and increasing concentrations of ZFC1-R3-LacB. From this concentrations of 125 nM LacA-ZFC1-L3 and 500 nM ZFC1-R3-LacB were determined to provide optimal signal-to-noise in the presence of 50 nM *L3-R3*. b) β-lactamase activity as a function of LacA-ZFC1-L3 concentration. 125 nM of ZFC1-R3-LacB was incubated with 50 nM *L3-R3* and increasing concentrations of LacA-ZFC1-L3. Results of these experiments suggest no distinct advantage of increased concentrations of LacA-ZFC1-L3 over ZFC1-R3-LacB. c) β-lactamase activity of LaA-ZFC1-L3 (250 nM) and ZFC1-R3-LacB (1 μM) in the presence of decreasing amounts of the dsDNA target *L3-R3*. At least 1 nM (100 fmols) of DNA is detectable above no dsDNA target under these initial conditions. d) β-lactamase activity under these optimized conditions scales linearly between the concentrations of 10 nM and 1 nM *L3-R3*. 
2.3 Sequence Specific Detection of Cytosine Methylation

In addition to the myriad of regulatory genes encoded by the human genome, there exists a second and perhaps equally complex level of regulation based on the selective methylation of DNA. This epigenetic control is dependent on the transfer of a methyl group from S-adenosylmethionine to the C5-position of cytosines present in CpG dinucleotides (Figure 2.5a) and is regulated by DNA methyltransferases, Me-CpG binding proteins, such as the mCpG binding protein MBD2, and demethylation via nucleotide excision or other DNA repair mechanisms.\(^{98-100}\) This cytosine-specific methylation is present in ~70% of all CpG dinucleotides found in mammalian somatic cells and contributes to genome stability, repression of transposable elements, and transcriptional silencing through the recruitment of chromatin modification complexes.\(^{101-103}\) Though CpG methylation is distributed throughout the genome, this chemical modification is normally excluded from promoter-associated CpG-rich regions of a sequence known as CpG islands as methylation is associated with translational repression. Thus, the observed aberrant methylation of tumor suppressor gene associated promoter regions likely results in gene silencing and can be linked to a number of human cancers. Several recent studies have shown that methylation of these CpG islands occurs in a sequence and tumor-type specific manner, leading to the identification of gene hypermethylation profiles for a number of human cancer types.\(^{104,105}\)

The ability to directly determine CpG methylation in a promoter-specific manner would undoubtedly provide a powerful tool for the detection and possible determination of specific cancer types. Current methylated CpG detection methods are PCR based and employ the bisulfate modification of DNA followed by DNA sequencing or methylation-
Figure 2.5. Site-specific determination of promoter methylation.  a) C5-Cytosine methylation by DNA methyltransferases (R = deoxyribose).  b) Promoter-specific methylation detection strategy. A methyl-CpG binding protein (MBD2, red) will be used to determine the methylation status of the adjacent CpG site. A sequence-specific DNA binding protein (blue) will direct the sensor to the desired promoter by targeting a unique DNA sequence adjacent to a CpG site being probed for methylation. Simultaneous binding of the two proteins will result in the reassembly of an appended split-signal generating domain and produce a detectable signal. Reprinted with permission from the American Chemical Society.
specific PCR. This treatment provides a basis for the distinction between methylated and nonmethylated cytosines through the conversion of nonmethylated cytosines to uracil while leaving methylated cytosines unchanged.\textsuperscript{106,107} Though effective, a number of problems are associated with these techniques, namely, extensive sample preparation as well as false positives due to incomplete bisulfate reactions.\textsuperscript{108,109} Currently there are no molecular methods available for the direct readout of sequence-specific CpG methylation utilizing antisense-based chemical approaches.

To create a SEER-based system capable of site-specific detection of CpG methylation, termed mCpG-SEER, we envisioned that a sequence-specific DNA binding domain tethered to half of split-GFP and a methyl-CpG recognition domain to the other half of the split-GFP would create a sensor design capable of reading out sequence-specific methylation events. Key to this design are the methyl-CpG binding domain (MBD) family of proteins whose members include the MBD1, MBD2, MBD3, MBD4, and MeCP2 proteins.\textsuperscript{110,111} Previous studies have demonstrated a high degree of structural similarity between these proteins as well as the specific contacts made between them and the major groove of DNA.\textsuperscript{112-114} Thus, our previously reported system based on split-GFP, incorporated the methyl-CpG binding domain, MBD2, in conjunction with the zinc-finger, Zif268, and was able to reassemble the split-GFP in a DNA sequence- and methylation-specific manner. Additionally, this GFP-based mCpG-SEER system was found to be highly specific for methylated CpG sites, showing a greater than 40-fold increase in signal generation in the presence of methylated DNA over nonmethylated DNA.\textsuperscript{115} Thus, mCpG-SEER, presents a general strategy for the development of protein-based sensors capable of interrogating CpG methylation in a promoter-specific manner.
In these systems, a designed sequence-specific DNA binding protein\textsuperscript{87,116,117} can be used to direct analysis to a DNA sequence unique to the promoter of interest, while a methylation-specific protein can be used to recognize an adjacent methylated CpG site. Simultaneous binding and formation of the ternary complex will result in the reassembly of the appended split-reporter resulting in a detectable signal (Figure 2.5b).

2.3.1 Design of a mCpG-\textbeta-lactamase SEER Sensor

Though capable of site-specific detection of DNA methylation, this initial GFP-based system has drawbacks, namely, insoluble component proteins that couple signal generation to a lengthy denaturation and refolding process. This need for refolding, in conjunction with a slow rate of GFP chromophore formation, results in a minimum assay time of 24 h for maximal fluorescence. The current study aims to significantly improve the sensitivity of mCpG-SEER through the incorporation of the enzymatic split-reporter TEM-1 \textbeta-lactamase. \textbeta-lactamase possesses a number of positive attributes for its use in mCpG-SEER, including its small size (29 kDa), ease of expression, and most importantly its prior demonstration as a split-protein reporter for detecting protein-protein interactions.\textsuperscript{20} Additionally, a number of substrates are available which allow both colorimetric and fluorogenic detection of \textbeta-lactamase activity. Finally, we have previously demonstrated that the split \textbeta-lactamase system tethered to zinc-fingers can be utilized to read out normal dsDNA sequence information.\textsuperscript{72} Thus, to construct this new \textbeta-lactamase-based mCpG-SEER system, referred to herein as mCpG-SEER-\textbeta-Lac (Figure 2.6), the DNA binding proteins MBD2 and Zif268 were chosen for use in our proof-of-principle experiment. The use of MBD2, which binds mCpG sites with a
Figure 2.6. mCpG-SEER-β-Lac system. MBD2-LacB (red and pink) and LacA-Zif268 (cyan and blue) are shown in ribbon form with an N-terminally expressed maltose binding protein (MBP) affinity tag (gray sphere) and highlighted TEM-1 β-lactamase active residues (yellow). The mCpG site is shown as spheres with cytosines (cyan) and methyl groups (green) shown as spheres. The Zif268 binding site (5’-GCGTGGGCG-3’) is shown in orange. Reprinted with permission from the American Chemical Society.
reported 2.7 nM affinity while binding analogous nonmethylated CpG sites with a reported 70-fold decrease in affinity,\textsuperscript{118} should provide adequate discrimination between methylated and nonmethylated CpG sites. As previously discussed, in order to uniquely recognize a specific site of DNA methylation, we need to impart sequence specificity to our split protein sensor. The zinc-finger Zif268 (\(K_d = 10^{-14}\) pM)\textsuperscript{119} was chosen to impart sequence specificity as demonstrated in the design of our two previous SEER systems. Each mCpG-SEER-\(\beta\)-Lac component was designed to contain a 15-amino acid linker between the DNA binding and split \(\beta\)-lactamase domains to ensure conformational flexibility to allow for ternary complex formation.

\textit{2.3.2 Initial mCpG-SEER-\(\beta\)-Lac Activity.}

\(\beta\)-lactamase activity assays were conducted using nitrocefin, a chromogenic substrate that undergoes a distinct color change from yellow (\(\lambda_{\text{max}} = 390\) nm) to red (\(\lambda_{\text{max}} = 486\) nm) upon \(\beta\)-lactam hydrolysis. Previous mCpG-SEER-GFP studies showed that a dsDNA target containing a 2-bp separation between the methylated CpG and zinc-finger binding sites provides an adequate signal upon protein binding.\textsuperscript{115} To evaluate the use of split \(\beta\)-lactamase for the detection of methylated CpG sites, the hydrolysis activity of 125 nM MBPLacA-Zif268 plus 125 nM MBP-MBD2-LacB was evaluated in the presence and absence of 5 nM mCpG-2-Zif268, a methylated dsDNA target containing a 2-bp spacing between the mCpG and the zinc-finger binding sites. Proteins were incubated for 30 min followed by the addition of nitrocefin to a final concentration of 100 \(\mu\)M. Absorbance at 486 nm was monitored for 15 min after the addition of nitrocefin. A marked increase in nitrocefin hydrolysis was observed in the presence of the dsDNA
target mCpG-2-Zif268, as opposed to minimal hydrolysis in the absence of dsDNA (Figure 2.7). This initial observation of a marked increase in activity only in the presence of a target methylated dsDNA served to initially validate our design of mCpG-SEER-β-Lac and its ability to address specific sites of DNA methylation.

2.3.3 mCpG-SEER-β-Lac Binding Site Specificity.

In order for mCpG-SEER-β-Lac to serve as a tool for interrogating promoter-specific methylation, the system must be able to distinguish methylated from nonmethylated CpG sites and to differentiate promoter-specific target sequences from nontarget sequences. To interrogate the binding site specificity of the mCpG-SEER-β-Lac system, a series of target dsDNA sequences were designed with mutations intended to abolish binding of either MBD2 or Zif268 to their respective binding sites. The following dsDNA target sequences containing a 2-bp spacing between the mCpG island and the Zif268 binding site were used: (1) nonmethylated CpG (CpG-2-Zif268), (2) Zif268 site absent (mCpG only), (3) a single G to T base mutation in the second finger recognition site of the Zif268 binding site (mCpG-2-Zif268, G to T), and (4) herring sperm DNA (15 ng). Mixtures of the two proteins, MBP-MBD2-LacB (125 nM) and MBP-LacA-Zif268 (125 nM), were allowed to incubate for 30 min in the presence of 5 nM concentrations of each dsDNA target described above followed by hydrolysis rate determination.

The observed rates of hydrolysis demonstrate the ability of mCpG-SEER-β-Lac to distinguish cognate binding sites from nonspecific, nonmethylated, or mutated binding sites (Figure 2.8). In the presence of a nonmethylated CpG site, the observed activity is
**Figure 2.7.** Initial mCpG-SEER-β-Lac activity. Absorbance spectra of MBP-LacA-Zif268 (125 nM) and MBP-MBD2-LacB (125 nM) in the presence (red) and absence (yellow) of 5 nM of the dsDNA target mCpG-2-Zif268 in zinc buffer A. Reprinted with permission from the American Chemical Society.
Figure 2.8. mCpG-SEER-β-Lac specificity. Nitrocefin hydrolysis in the presence of MBP-LacA-Zif268 (125 nM) and MBP-MBD2-LacB (125 nM) in the presence of indicated dsDNA targets (5 nM). A 42-fold decrease in β-lactamase activity is observed in the absence of a methyl group (CpG-2-Zif268). Reprinted with permission form the American Chemical Society.
42-fold less than that observed in the presence of a methylated CpG site, showing the ability of mCpG-SEER-β-Lac to clearly distinguish methylated from nonmethylated CpG sites. In addition to methylation, the need for a cognate Zif268 binding site is apparent in the 170-fold activity decrease in the absence of a Zif268 binding site. As a testament to the sequence specificity of Zif268, the single G to T base mutation resulted in a 93-fold decrease in activity compared to the parent Zif268 site. Last, the 220-fold decrease in activity in the presence of nonspecific herring sperm DNA reaffirms the requirement of both a methylated CpG site and a specific zinc-finger binding site for signal generation.

2.3.4 mCpG-SEER-β-Lac Activity Dependence on Target Site Spacing

To evaluate the effect of distance and proximity on signal generation, a series of dsDNA targets were designed that have an increasing spacing between the mCpG and Zif268 binding sites. In all, 14 dsDNA targets were utilized, which spanned 0-13 bps between the mCpG and Zif268 binding sites. Mixtures of the two mCpG-SEER-β-Lac proteins, MBP-MBD2-LacB (125 nM) and MBP-LacA-Zif268(125 nM), were allowed to incubate for 30 min in the presence of 5 nM concentrations of each dsDNA target. The results from these experiments provide some interesting insight into the signal generation profile of the mCpG-SEER-β-Lac system (Figure 2.9, tan). Virtually no activity is observed when the MBD2 and zinc-finger binding domains are placed next to one another (0-bp spacing). This is most likely due to the fact that the MBD2 binding domain occupies at least 1 bp on the surface of DNA beyond the canonical mCpG recognition site. As the distance between the two binding sites increases to 1 and 2 bps, a sharp increase in activity is observed, with maximal signal being generated when the mCpG
Figure 2.9. mCpG-SEER-β-Lac dependence on target site spacing. Relative signal of mCpG-SEER-β-Lac (125 nM) and mCpG-SEER-GFP in the presence of dsDNA target as a function of increased number of base pairs between the mCpG and Zif268 sites (inset). In both cases maximal reassembly is seen at a spacing of 2 base pairs, which presumably allows the simultaneous binding of both fusion proteins. Reprinted with permission from the American Chemical Society.
and zinc-finger binding sites are separated by 2 bps. As the number of bps increases, a stepwise decrease in activity is observed for spacings of 3-7 bps, with a marked increase occurring again with a second signal maximum (70% of the signal generated at a distance of 2 bps) when the two binding sites are separated by 10 bps. As the distance between the two sites increases from 11 to 13 bps, a decrease in activity is again observed. This pattern serves to graphically recapitulate the helical turn of dsDNA by not only increasing the linear distance between the bound mCpG-SEER-β-Lac proteins but also varying the orientation of each protein relative to each other on the surface of dsDNA. Interestingly a similar but more stringent pattern is observed when GFP is used as the signal generating protein (Figure 2.9, green), but shows significantly less signal generation at separations of 8-10 bps.

2.3.5 dsDNA Concentration Dependence of mCpG-SEER-β-Lac Activity.

The above experiments demonstrated that the maximal signal intensity is obtained when the target dsDNA oligo contains a 2-bp spacing between the mCpG and zinc-finger binding sites. With this in mind, the dsDNA target oligo mCpG-2-Zif268 was used to determine the minimum amount of dsDNA that can be detected by mCpG-SEER-β-Lac. Mixtures of the two mCpG-SEER-β-Lac proteins, MBP-MBD2-LacB (125 nM) and MBP-LacA-Zif268 (125 nM), were allowed to incubate for 30 min in the presence of decreasing concentrations of $mCpG-2-Zif268$ (50-0.2 nM) (Figure 2.10). These experiments show that 200 pM (40 fmol) $mCpG-2-Zif268$ is visible above background and that activity scales linearly from 5 to 0.5 nM when the chromogenic substrate nitrocefin is used (Figure 2.10, inset).
Figure 2.10. mCpG-SEER-β-Lac limit of detection utilizing nitrocefin. Relative hydrolysis of nitrocefin in the presence of MBP-LacA-Zif268 and MBP-MBD2-LacB (125 nM each) in the presence of decreasing concentrations of dsDNA target mCpG-2-Zif268. (Inset) Linear fit from 5 to 0.5 nM mCpG-2-Zif268. Reprinted with permission form the American Chemical Society.
To further increase the sensitivity of mCpG-SEER-β-Lac for dsDNA, the fluorogenic substrate CCF2FA was utilized to optimize the detection capability of mCpG-2-Zif268. CCF2FA is a FRET-based fluorescent β-lactamase substrate and has been previously reported to detect as little as 50 molecules of β-lactamase in a single living cell. Starting with an amount of dsDNA near the lower linear limit of what could be detected with nitrocefin, experiments were performed as described above in the presence of decreasing concentrations of mCpG-2-Zif268 (500-25 pM) (Figure 2.11a). With the use of CCF2FA, mCpG-SEER-β-Lac was capable of detecting target dsDNA concentrations as low as 25 pM (2.5 fmol), resulting in a 16-fold improvement in sensitivity over nitrocefin and a 2000-fold improvement over our previously described GFP-based sensor. Additionally, activity was found to scale linearly from 500 to 25 pM (Figure 2.11b).

2.3.6 Detection of Methylated dsDNA in the Presence of Excess Non-Specific Genomic DNA

Ideally we would like to utilize this and other SEER-based sensors in a genomic context. To mimic genomic conditions we evaluated the ability of our mCpG-SEER-β-Lac sensor to produce a target dsDNA dependent signal in the presence of increasing concentrations of isolated herring sperm DNA (HS-DNA). Mixtures of the two mCpG-SEER-β-Lac proteins, MBP-MBD2-LacB (125 nM) and MBP-LacA-Zif268(125 nM), were allowed to incubate for 30 min in the presence of mCpG-2-Zif268 (5 nM, 17.18 ng) and increasing concentrations of HS-DNA (2.15 – 34.4 μg). Additionally, β-lactamase activity was monitored in the absence of mCpG-2-Zif268 to determine the amount of
Figure 2.11. mCpG-SEER-β-Lac limit of detection utilizing CCF2FA. a) Hydrolysis of CCF2FA in the presence of MBP-LacA-Zif268 and MBP-MBD2-LacB (125 nM each) in the presence of decreasing concentrations of the dsDNA target \( mCpG-2-Zif268 \). Hydrolysis was determined by either the increase in emission at 450 nm (blue) or the decrease in emission at 520 nm (shown in green as the negative decrease). (Inset) Structure of CCF2FA. b) Linear fit of β-lactamase activity as determined by the increase in emission at 450 from 500 to 25 pM \( mCpG-2-Zif268 \). Reprinted with permission form the American Chemical Society.
signal generated in the presence of HS-DNA alone (Figure 2.12a). These results show that background β-lactamase activity increases with the amount of HS-DNA present during incubation with the highest signal being observed for an amount of HS-DNA 2000-fold higher than the amount of mCpG-2-Zif268 used. Surprisingly, in spite of this increase in signal as a function HS-DNA, the total signal for all conditions was relatively equal when 5 nM (17.18 ng) of mCpG-2-Zif268 was present leading to a decrease in sensitivity as the amount of HS-DNA increased. This lead to a steady decrease in signal-over-background as a function of the amount of excess HS-DNA, most likely due to the non-specific binding of mCpG-SEER-β-Lac components to HS-DNA. To decrease non-specific binding, increasing concentrations of the polyamine spermine were added, and the signal-over-background evaluated. Mixtures of the two mCpG-SEER-β-Lac proteins, MBP-MBD2-LacB and MBP-LacA-Zif268 (125 nM each), were incubated for 30 min in the presence or absence of 5 nM mCpG-2-Zif268, 17.18 μg (1000X) HS-DNA, and increasing concentrations of spermine. As the amount of spermine present increased, so did the signal-over-background in the presence of 1000X HS-DNA, with a minimal signal-over-background being achieved with 750 μM spermine (Figure 2.12b).

2.4 Conclusion

A large volume of work has been compiled demonstrating the ability to design and create split-protein biosensors capable of interrogating a wide variety of protein-protein and protein-nucleic acid interactions.\textsuperscript{17,53,68} Additionally, a number of selection and design strategies have been utilized to develop proteins that can selectively bind specific DNA sequences.\textsuperscript{87,116,117} Previously, the split GFP variant sg100 was utilized to
Figure 2.12. Evaluation of mCpG-SEER-β-Lac in the presence of excess HS-DNA. a) β-lactamase activity in the presence of increasing amounts of excess HS-DNA, either with 5 nM mCpG-2-Zif268 or HS-DNA alone. b) β-lactamase activity of MBP-LacA-Zif268 and MBP-MBD2-LacB (125 nM each) with 5 nM mCpG-2-Zif268 and 1000X HS-DNA in the presence of increasing amounts of spermine. From this a maximal signal-over-background is observed in the presence of 750 μM spermine.
develop a new methodology, termed mCpG-SEER, for the site-specific detection of DNA. Here we apply SEER, with the use of a split-β-lactamase reporter and existing zinc finger design strategies, to the creation of a sensor for the detection of the *CryIA* gene within the maize genome. Our experiments show the successful dsDNA-dependent β-lactamase reassembly in the presence of a *CryIA* gene target sequence with the use of designed sequence specific zinc fingers. Additionally, we established optimized conditions consisting of component ratios of 1:4 of LacA-ZFC1-L3 to ZFC1-R3-LacB. Finally, utilizing these conditions, the use of 250 nM LacA-ZFC1-L3 and 1 μM ZFC1-R3-LacB was capable of detecting a concentration of 1 nM of the target dsDNA *L3-R3*, representing 100 fmols of DNA.

We also apply this SEER methodology, with the use of a split-β-lactamase reporter, to the sequence specific determination of DNA methylation, building upon a previous sensor termed mCpG-SEER-GFP. In this study, the expression vector pMAL-c2X was used, which appends a maltose binding protein domain to the N-terminus of the expressed recombinant proteins, resulting in fusion proteins with increased stability and solubility compared to the GFP-based system. The use of soluble mCpG-SEER-β-Lac proteins, coupled with the enzymatic signal amplification afforded by TEM-1 β-lactamase, allows for the detection of methylated dsDNA in less than 5 min when the chromogenic substrate nitrocefin was used. When the fluorogenic substrate CCF2FA was utilized, low-femtomole quantities of methylated target dsDNA could be detected, though longer assay times were required due to a decreased rate of hydrolysis. This sensor thus provides a greater than 250-fold time improvement over the previously reported GFP-based mCpG-SEER system in terms of assay time. In addition to an
increase in assay time, the currently described β-lactamase-based sensor also provides a 2000-fold improvement in the minimum amount of methylated target dsDNA needed to generate a detectable signal above background.

The crux of the mCpG-SEER methodology is the ability of the designed components to both distinguish methylated from nonmethylated CpG sites and accurately recognize promoter-specific DNA sequences. In the absence of cytosine methylation, a significant 42-fold decrease in activity is seen, showing sensitivity to the methylation status of the CpG site of interest. Additionally, a single bp mismatch in the adjacent zinc-finger binding site results in a 93-fold decrease in activity while the complete elimination of the adjacent zinc-finger binding site results in a 170-fold decrease in activity. These series of control experiments clearly demonstrate that the current mCpG-SEER-β-Lac design shows a remarkable ability to selectively detect methylated CpG sites in a sequence-dependent manner, attesting to the specificity of both MBD2 and zinc-finger DNA binding elements.

The distance-dependence profile of the mCpG-SEER-β-Lac system provides a number of insights into split β-lactamase reassembly and the design and application of the current mCpG-SEER-β-Lac system to promoter methylation analysis. Unlike the previously described GFP-based mCpG-SEER system (Figure 2.9), which shows minimal signal generation at a spacing of 8, 9, and 10 bps, the current mCpG-SEER system shows significant signal generation at a spacing of 8-13 bps. Unlike GFP, which refolds to adopt a single domain β-barrel fold, and is considered to be a single domain protein, β-lactamase comprises two separate domains, between which β-lactamase has been split. As a result, β-lactamase likely possesses a higher degree of conformational
flexibility, allowing catalytic activity even under conditions that subtly distort the interface between these domains. This natural conformational flexibility likely accounts for the increased signal generation observed with β-lactamase over GFP at separations greater than 2 bp. The accurate analysis of methylated promoters will depend on the sequence specificity of the promoter-specific DNA binding domain; thus, the observed activity of mCpG-SEER-β-Lac at different separation distances could potentially provide inaccurate results in the detection of site-specific methylation events. However, this flexibility may be a boon in disguise as it may allow one to potentially increase the amount of targetable sequence space. This is relevant since the tunability of zinc-fingers, though extremely promising, cannot at present target all DNA sequences with high sequence selectivity. Furthermore, the observed flexibility in β-lactamase reassembly will likely afford an avenue to tune the enzymatic activity by redesigning the 15-residue linkers that separate the lactamase halves from the DNA recognition domains.

Previous experiments have demonstrated the detection of as few as 50 molecules of β-lactamase using the fluorogenic substrate CCF2. Using the free acid form of CCF2, our in vitro experiments clearly detect as little as 2.5 fmol of target methylated dsDNA. These results show that the use of β-lactamase provides a 2000-fold decrease in the minimal amount of dsDNA needed to generate a visible signal above background when compared to our split-GFP system. In the current study, fluorescence measurements were carried out using a standard top-read fluorescence plate reader, yielding a detection limit of 2.5 fmol of target dsDNA. Given the limitations of top-read plate readers, it is conceivable that lower detection limits will be achievable by future optimization of optical detection methods combined with reductions in sample volume.
Moreover, modifications of the fluorescent lactamase substrate as well as re-engineering the lactamase can also potentially increase sensitivity. Future efforts will aim to extend this system for detecting DNA methylation by utilizing the split-luciferase system as well as target promoter sequences of genes known to be hypermethylated in cancerous cells with designed zinc-fingers.  

Lastly, the ability of the mCpG-SEER-β-Lac sensor to detect methylated DNA was evaluated in the presence of a large excess of genomic herring sperm DNA (HS-DNA). A signal of >3-fold above background was observed for 17.18 ng of methylated DNA target in the presence of 1860 ng of HS-DNA. Additionally, to reduce the observed non-specific binding of mCpG-SEER-β-Lac to HS-DNA the polyamine spermine was used. The addition of spermine concentrations as high as 1 mM were successful at reducing the amount of non-specific binding to HS-DNA by mCpG-SEER-β-Lac, with the greatest signal-above-background being observed at a spermine concentration of 750 μM.

These studies clearly demonstrate that split-β-lactamase provides an improved split-reporter system domain for the site-specific detection of DNA methylation. With a rapid total assay time and a minimal detection of less than 3 fmol of methylated dsDNA, mCpG-SEER-β-Lac has the potential to provide a new strategy for determining the epigenetic modification of dsDNA. Given the currently available zinc-finger lexicon, new methods for determining sequence preferences of DNA binding proteins, and the modularity of the SEER design, mCpG-SEER-β-Lac provides a new platform for directly interrogating the specific methylation status of promoters of interest. Thus, this
methodology has the potential for providing new diagnostic reagents for understanding epigenetic control as well as aid in the early diagnosis of cancer.

2.5 Experimental Procedures

2.5.1 Cloning Expression, and Purification of CryIA-SEER-β-Lac and mCpG-SEER-β-Lac Proteins

All enzymes were purchased from NEB, dNTPs were purchased from Fermentas. Separate pUC57 plasmids containing *E. coli* optimized genes encoding the designed ZFC1-L3 and ZFC1-R3 CryIA targeting zinc fingers, and a the human MBD2 (residues 147-215) were purchased from GenScript. Zinc buffer A 100 mM Tris at pH 7.5, 90 mM KCl, 1 mM MgCl2, 100 μM ZnCl2, and 5 mM DTT.

The following primers were used to clone ZFC1-L3 into an existing cassette containing an N-terminal MBP tag and the N-terminal portion of TEM-1 β-lactamase (LacB), residues 198-290, followed by a 15-amino acid linker: Fwd, 5’- ATATATGGTA CCCCCGGGAGAAGCCCT-3’; and Rev, 5’- ATATATACCGGTGTGAGTACGTTGG TG-3’.

The following primers were used to clone ZFC1-R3 into an existing cassette containing an N-terminal MBP tag and a 15-amino acid linker followed by the C-terminal portion of TEM-1 β-lactamase (LacB), residues 198-290: Fwd, 5’- ATATATGAATTCC CCCCCGGGAGAAGCCC-3’; and Rev, 5’- ATATATACCGGTGTGAGTACGTTGG TG-3’.

The following primers were used to clone MBD2 into an existing cassette containing an N-terminal MBP tag and a 15-amino acid linker followed by the C-terminal
portion of TEM-1 β-lactamase (LacB), residues 198-290: Fwd, 5’-GCGCGGAATTCTGAAGC
GCAAACGC-3’; and Rev, 5’-CGGTTAACCGGTCATTTGCCGGTACG-3’.

All β-lactamase fusion proteins were purified as follows: The resulting pMAL-c2X plasmids containing either the LacA-ZFC1-L3, ZFC1-R3-LacB, or MBD2-LacB gene (Figures 2.13, 2.14, 2.15, and 2.16) were transformed by electroporation into BL21 (DE3) cells (Novagen). An overnight culture of these cells was used to inoculate a 100-
mL culture of 2xYT media supplemented with 100 μg/mL ampicillin, 100 μM ZnCl2, and 2% glucose at an OD600 of 0.05. Protein expression was induced at an OD600 of 0.8 with 1 mM isopropyl β-D-thiogalactopyranoside for 4 h at 37 °C. Cells were pelleted by centrifugation and frozen overnight at -20 °C. Pelleted cells were thawed on ice and resuspended in zinc buffer A. Resuspended cells were lysed by sonication. The lysate was cleared by centrifugation at 18,000 cfm for 30 min. MBP-tagged β-lactamase fusions were purified over amylose resin (New England Biolabs) using the following protocol. 3 mL of resin was loaded into a PD-10 column and equilibrated with 5 column volumes of zinc buffer A. Following the addition of cleared lysates, the column was washed with 10 column volumes of zinc buffer A supplemented with 2 M NaCl followed by an additional wash with 8 column volumes of zinc buffer A. Fusions were eluted with 5 mL of zinc buffer A supplemented with 10 mM maltose and characterized by SDS-PAGE. Concentrations were determined in 6 M guanidine using absorbance measurements at 280 nm (calculated ε: MBP-LacA-ZFC1-L3 = 82,780 M⁻¹ cm⁻¹; MBP-ZFC1-R3-LacB = 89,185 M⁻¹ cm⁻¹; MBP-MBD2-LacB = 94,310 M⁻¹ cm⁻¹).
<table>
<thead>
<tr>
<th>Amino Acid Sequence</th>
<th>DNA Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>LacA is shown in red, the 15 amino acid linker is shown in green, and ZFC1-L3 is shown in blue.</td>
<td></td>
</tr>
</tbody>
</table>

**Figure 2.13.** The DNA and amino acid sequence of LacA-ZFC1-L3. LacA is shown in red, the 15 amino acid linker is shown in green, and ZFC1-L3 is shown in blue.
Figure 2.14. The DNA and amino acid sequence of ZFC1-L3-LacB. ZFC1-R3 is shown in blue, the 15 amino acid linker is shown in green, and LacB is shown in red.
Figure 2.15. The DNA and amino acid sequence of MBD2-LacB. MBD2 is shown in blue, the 15 amino acid linker is shown in green, and LacB is shown in red.
Figure 2.16. Schematics of pMAL-c2X plasmids showing gene positions and restriction enzyme sites used. a) LacA-ZFC1-L3. b) ZFC1-R3-LacB. c) MBD2-LacB.
The MBP-LacA-Zif268 fusion protein (containing residues 26-196 of TEM-1 β-lactamase) was expressed and purified as described above. MBP-LacA-Zif268 was characterized by SDS-PAGE; concentrations were determined in 6 M guanidine using absorbance measurements at 280 nm (calculated $\varepsilon = 79,300 \text{ M}^{-1} \text{ cm}^{-1}$).

2.5.2 SDS-PAGE Analysis of CryIA-SEER-β-Lac and mCpG-SEER-β-Lac Proteins

Equivalent amounts of purified MBP-LacA-ZFC1-L3 (74.5 kDa) and MBP-ZFC1-R3-LacB (64.5 kDa) (Figure 2.17a) or MBP-LacA-Zif268 (75.0 kDa) and MBP-MBD2-LacB (62.7 kDa) (Figure 2.17b) were loaded on a 12.5% SDS-PAGE gel.

2.5.3 Preparation of dsDNA Targets

DNA oligonucleotides containing target DNA sequences (HPLC-purified) were obtained from IDT, and sheared herring sperm DNA was obtained from Invitrogen. Target DNA oligos in 1X BamHI buffer (NEB) were annealed using a Techne Genius thermocycler with the following protocol: 95 °C for 7 min, decrease to 56 °C at 1 °C/min, with equilibration for 5 min at 56 °C, and finally decrease to 25 °C at 1 °C/min.

2.5.4 Absorbance Measurements

All absorbance spectra were acquired on a μQuant plate reader (Bio-Tek Instruments) at 486 nm over a period of 15 min. Signal generation is defined as the rate of nitrocefin hydrolysis between 0.5 and 2 min after nitrocefin addition. Activity was determined by subtracting background rate of hydrolysis for mCpG-SEER-β-Lac proteins.
Figure 2.17. SDS-PAGE analysis of SEER-β-Lac proteins. a) CryIA-SEER-β-Lac proteins. Lane 1: MW Standards; Lane 2: MBP-LacA-ZFC1-L3; Lane 3: MBP-ZFC1-R3-LacB. b) mCpG-SEER-β-Lac proteins. Lane 1: MW Standards; Lane 2: MBP-LacA-Zif268; Lane 3: MBP-MBD2-LacB.
at the same concentration in the absence of DNA; where indicated, these samples were then normalized to the sample with the highest rate of hydrolysis. All experiments were conducted in 96-well plates (CoStar, Corning).

2.5.5 Fluorescence Measurements.

All fluorescence measurements were acquired on a Spectra Max Gemini (Molecular Devices) plate reader with excitation at 409 nm and monitored emission at 450 and 520 nm. Cutoffs were set at 435 and 495 nm for the 450- and 520-nm emissions, respectively. Given the decreased rate of catalysis of CCF2FA ($k_{cat} = 29 \text{ s}^{-1}$) compared to nitrocefin ($k_{cat} = 930 \text{ s}^{-1}$) a 30-min incubation period after CCF2FA addition was required. After incubation, fluorescence measurements were acquired at 450 and 520 nm every 2 min for 20 min. Activity was determined by subtracting background signal for mCpG-SEER-β-Lac proteins at the same concentration without DNA. All experiments were conducted in 96-well plates (CoStar, Corning).

2.5.6 Initial Evaluation of the CryIA-SEER-β-Lac Sensor

To determine the initial activity of the CryIA-SEER-β-Lac sensor LacA-ZFC1-L3 and ZFC1-R3-LacB (125 nM each) were mixed with the dsDNA target L3-R3 (50 nM) in zinc buffer A to a volume of 180 μL and allowed to incubate for 30 min at room temperature. Following the 30-min incubation, 20 μL of 1 mM nitrocefin in phosphate buffer (100 mM, pH 7.0) was added to a final concentration of 100 μM and a final volume of 200 μL.
To determine the minimum amount of dsDNA needed to produce a detectable signal, LacA-ZFC1-L3 and ZFC1-R3-LacB (125 nM each) were mixed in zinc buffer A, in a volume of 180 μL, with decreasing concentrations of the dsDNA target \( L3-R3 \). Following a 30-min incubation at room temperature, 20 μL of 1 mM nitrocefin in phosphate buffer (100 mM, pH 7.0) was added to a final concentration of 100 μM and a final volume of 200 μL.

2.5.7 Determination of Optimal CryIA Sensor Conditions

To determine the optimal ratios of LacA-ZFC1-L3 to ZFC1-R3-LacB, and visa versa, 125 nM of either LacA-ZFC1-L3 to ZFC1-R3-LacB were mixed with increasing concentrations of either ZFC1-R3-LacB or LacA-ZFC1-L3 and 50 nM dsDNA target \( (L3-R3) \) in zinc buffer A. Following a 30-min incubation at room temperature, 20 μL of 1 mM nitrocefin in phosphate buffer (100 mM, pH 7.0) was added to a final concentration of 100 μM and a final volume of 200 μL.

2.5.8 Limit of Detection for CryIA-SEER-β-Lac

To determine the minimum amount of dsDNA needed to produce a detectable signal under optimized conditions, LacA-ZFC1-L3 (250 nM) and ZFC1-R3-LacB (1 μM) were mixed in zinc buffer A, in a volume of 180 μL, with decreasing concentrations of the dsDNA target \( L3-R3 \). Following a 30-min incubation at room temperature, 20 μL of 1 mM nitrocefin in phosphate buffer (100 mM, pH 7.0) was added to a final concentration of 100 μM and a final volume of 200 μL.
2.5.9 *mCpG-SEER-β-Lac Specificity and Target Site Spacing Experiments*

All target DNA sequences used for determining *mCpG-SEER-β-Lac* specificity and target-site spacing dependence are given in Figure 2.8 and Table 2.1, respectively. All experiments were performed in triplicate.

To determine *mCpG-SEER-β-Lac* dependence on target specificity and target site spacing, MBP-MBD2-LacB and MBP-LacA-Zif268 (125 nM each) were mixed with target dsDNA (5 nM) in zinc buffer A to a volume of 180 μL and allowed to incubate for 30 min at room temperature. Following the 30-min incubation, 20 μL of 1 mM nitrocefin in phosphate buffer (100 mM, pH 7.0) was added to a final concentration of 100 μM and a final volume of 200 μL.

2.5.10 *Limit of Detection of mCpG-SEER-β-Lac*

To determine the minimum amount of dsDNA needed to produce a detectable signal, MBP-MBD2-LacB and MBP-LacA-Zif268 (125 nM each) were mixed in zinc buffer A, in a volume of 180 μL, with decreasing concentrations of a dsDNA target containing a 2-bp spacing between the *mCpG* and Zif268 binding sites. Following a 30 min incubation at room temperature, 20 μL of 1 mM nitrocefin in phosphate buffer (100 mM, pH 7.0) was added to a final concentration of 100 μM and a final volume of 200 μL. For experiments using the fluorescent substrate CCF2FA, 10 μL of 100 μM CCF2FA in PBS was added to a 90 μL of solution containing *mCpG-SEER-β-Lac* proteins (125 nM each) and target dsDNA to a final concentration of 10 μM in a final volume of 100 μL. Generated signal is defined as the rate of CCF2FA hydrolysis between 30 and 50 min after CCF2FA addition.
Table 2.1. dsDNA target oligonucleotides used to test mCpG-SEER-β-Lac dependence on target site spacing. MBD2 and Zif268 sites are shown in red and blue respectively.
CHAPTER 3

APPLICATION OF FIREFLY LUCIFERASE TO SEER

3.1 Introduction

The use of light emitting reporter such as GFP and “luciferases” has greatly enhanced the study of gene expression and biological processes, including the study of viral gene transfer, single cell determination of gene expression, and the monitoring of gene expression and biological events in plants, nematode worms, and living mice. Though an effective reporter protein, GFP falls short with respect to these latter applications due to its need for external excitation, especially within the context of animal subjects where efficient tissue penetrance is essential. The use of bioluminescent reporters such as luciferases have re-emerged as potential luminescent probes as they do not require excitation, are extremely sensitive (with a linear range of eight orders of magnitude), and have low background compared to fluorescence-based measurements.

Luciferases are a family of proteins which produce light through the reaction of molecular oxygen with a chemical substrate, eliminating the need for an external light source and allowing their use in a variety of contexts, including tissue and living animals. To date, a variety of bioluminescent luciferase proteins have been identified and used as reporter genes, including, bacterial luciferase (lux), firefly luciferase, and the Renilla and Gaussia luciferases. Though they all catalyze light emitting reactions, these enzymes, which come from bacteria, beetles, or marine organisms, share no nucleotide homology, and are structurally diverse, and with the exception of the Renilla and Gaussia
luciferases, utilize chemically unrelated substrates\textsuperscript{132,133} (Figure 3.1). This substrate orthogonality is of particular utility, allowing the simultaneous use of multiple luciferases as with different colored GFPs.\textsuperscript{134} Utilizing the various luciferases mentioned above, a variety of sensor technologies have been developed to study numerous biological processes and detect a variety of substrates. For example, engineered bacteria have been designed which place \textit{lux} bacterial luciferase reporter genes under control of various pollutant-responsive catabolic operons, creating bacterial sensors capable of sensing the presence and bioavailability of organic and heavy metal toxins.\textsuperscript{135-137} Additionally, bacterial luciferase has been used to create genetically engineered \textit{salmonella} for the study of bacterial virulence in living mice.\textsuperscript{138} Unlike bacterial luciferase whose use has been limited to bacteria, firefly, \textit{Renilla}, and \textit{Gaussia} luciferases have been used as reporters of gene expression in both cultured cells and live animals, as well as, reporters of protein-protein interactions and protein trafficking.\textsuperscript{139-141}

Given the advantages provided by bioluminescent reporters, particularly low background and large dynamic range, we sought to employ the use of a luciferase reporter protein within our previously described sequence-enabled reassembly (SEER) strategy for the detection of specific dsDNA sequences.\textsuperscript{71,72} Of the reporters available, not all are readily amenable to split-protein applications. Traditionally, split-protein reassembly systems utilize small monomeric proteins or enzymes.\textsuperscript{10} This general quality eliminates the use of bacterial luciferase, for which dimerization is required for luminescence to occur. Of the remaining reporters, we chose firefly luciferase as an initial bioluminescent reporter that could be utilized in our SEER approach.
Figure 3.1. Various luciferase reporters. a) Bacterial luciferase (left). Bacterial luciferase substrates flavin mononucleotide (FMNH$_2$), a long chain aldehyde, and molecular oxygen (right). b) Firefly luciferase (left). Firefly luciferase substrates D-luciferin, adenosine triphosphate (ATP), and molecular oxygen (right). c) Renilla luciferase (left). Renilla luciferase substrates coelenterazine and molecular oxygen (right). d) Amino acid sequence of Gaussia luciferase (16 residue N-terminal secretion signal sequence removed) with predicted $\alpha$-helical and extended sheet structural elements (top). Gaussia luciferase substrates coelenterazine and molecular oxygen (bottom).
At the time of this work, firefly luciferase was the only bioluminescent reporter for which a solved crystal structure was available\textsuperscript{142} and additionally, due to its previous use as a transcriptional reporter, various engineering efforts have created a number of luciferase mutants possessing enhanced stability and altered spectral properties. For example, structural variants of firefly luciferase have been created which enhance the thermal stability by 27-fold at 37 °C, enabling the use of luciferase based technologies at 37 °C.\textsuperscript{143} In addition to thermally stable variants, efforts have resulted in the creation of both red and green firefly luciferase mutants which possess non-overlapping emission spectra whose maxima are separated by 65 nm,\textsuperscript{144,145} potentially enabling the creation of pairs of luciferase sensors for the simultaneous detection of multiple dsDNA sequences. A summary of these mutations is provided in Figure 3.2. Most importantly, at the time this work was initiated, a variety of studies utilizing split firefly luciferase had been reported, providing us with a variety of starting points from which to develop a split-firefly luciferase SEER sensor (SEER-FFluc).

3.2 Results and Discussion

3.2.1 Creation of a dsDNA Sequence Specific Split-Firefly Luciferase Sensor (SEER-FFluc)

As stated above, a variety of studies have utilized split-firefly luciferase for the study of protein-protein interactions, protein dynamics, and protein trafficking. Additionally, a systematic study by Gambhir and coworkers has been recently performed to identify optimal sites within the firefly luciferase primary sequence for the creation of a split-firefly luciferase reporter\textsuperscript{76} (Figure 3.3). Of these, the firefly luciferase fragments
**Figure 3.2.** Firefly luciferase mutations responsible for enhanced thermal stability and altered spectral properties.  
a) Structure of firefly luciferase with enhanced thermal stability (yellow), red emission (red) and green emission (green) mutations highlighted.  
b) Table describing specific residue mutation responsible for thermal stability and emission alterations.

<table>
<thead>
<tr>
<th>Property</th>
<th>Mutations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Enhanced Thermal Stability</td>
<td>Thr214Ala, Ala215Leu, Ile232Ala, Phe295Leu, Glu354Lys</td>
</tr>
<tr>
<td>Red Emission (616 nm maxima)</td>
<td>Ser284Thr</td>
</tr>
<tr>
<td>Green Emission (551 nm maxima)</td>
<td>Val241Ile, Gly246Ile, Phe250Ser</td>
</tr>
</tbody>
</table>
Figure 3.3. Structure of firefly luciferase with highlighted fragmentation sites analyzed by Gambhir and coworkers.
NFluc(1-437) and CFluc(438-550) were chosen for our initial SEER-FFluc reporter fragments due to the reported high signal sensitivity afforded by this pair. To facilitate dsDNA dependent luciferase reassembly the well characterized zinc fingers Zif268 and PBSII were appended to the C-terminus and N-terminus of NFluc and CFluc respectively creating the zinc finger-firefly luciferase fusions NFluc(1-437)-Zif268 and PBSII-CFluc (438-550) (Figure 3.4a).

Previously described GFP and β-lactamase based SEER systems utilized recombinant proteins purified from *E. coli* either by either His₆ or maltose binding protein (MBP) affinity tags.⁷¹,⁷²,¹¹⁵,¹⁴⁶ To verify appropriate expression conditions for firefly luciferase, the firefly luciferase gene was incorporated into the His-tag containing pETDuet vector allowing for the expression of full length firefly luciferase in the *E. coli* cell line BL-21 DE3 and purification via Ni-NTA affinity chromatography. Initially, standard protein expression conditions were used, i.e. induction of protein expression at 37 °C for 6 hours. These initial conditions failed to produce active luciferase, likely due to the decreased thermal stability of firefly luciferase at 37 °C. Subsequent reduction to 22 °C upon induction of protein expression resulted in soluble and active firefly luciferase by visual inspection of crude lysates upon addition of D-luciferin and ATP. Full length firefly luciferase was purified using Ni-NTA affinity chromatography and was found to produce concentration dependent luminescence upon the addition of 100 μM of both D-luciferin and ATP at a pH of 7.8 (Figure 3.4b).

Having established firefly luciferase expression conditions, we next sought to demonstrate the dsDNA dependent reassembly of the split-firefly luciferase constructs. Both NFluc (1-437)-Zif268 and PBSII-CFluc (438-550) expressed within the soluble
Figure 3.4. Initial luminescence measurements. a) Top: Initial firefly luciferase-zinc finger fusions used. Bottom: schematic representation of the dsDNA-dependent firefly luciferase reassembly of NFluc(1-437)-PBSII and Zif268-CFluc (438-550). b) Concentration dependent luminescence of purified full length firefly luciferase. c) Luminescence of NFluc(1-437)-PBSII and Zif268-CFluc (438-550), 5 μM each, in the presence of 2.5 μM of the dsDNA targets Zif268-0-PBSII and Zif268-10-PBSII and no dsDNA target.
fraction, were purified using Ni-NTA affinity chromatography, and verified by polyacrylamide gel electrophoresis. To evaluate dsDNA dependent reassembly equimolar amounts (5 \( \mu \)M) of the two purified proteins, NFluc(1-437)-Zif268 and PBSII-CFluc(438-550), were incubated with 2.5 \( \mu \)M of either Zif268-0-PBSII or Zif268-10-PBSII, two dsDNA targets which separate the 9-base pair recognition sites for Zif268 and PBSII by either zero or ten bases, followed by luminescence measurement. These initial attempts produced no appreciable signal above background (Figure 3.4c). Subsequent attempts to optimize experimental conditions, such as systematically varying ratios of protein to target dsDNA also failed to produce any significant dsDNA-dependent luminescence.

3.2.2 Denaturation and Refolding of Full Length and SEER-FFluc

Previously, the denaturation and subsequent refolding of fusion proteins was used to facilitate the dsDNA dependent reassembly of GFP.\textsuperscript{71} Given the successful application of this approach to the dsDNA-dependent reassembly of GFP we next sought to establish conditions under which firefly luciferase could be denatured and refolded to yield active enzyme. Given its size, tendency to aggregate, and straightforward readout of activity, the folding of firefly luciferase has been extensively studied and used as a tool to analyze the activity of protein chaperones and the refolding ability of cellular lysates.\textsuperscript{147-149} A variety of methods have been utilized to aid in the refolding of firefly luciferase, including the use of antibodies\textsuperscript{150} and immobilization on agarose beads.\textsuperscript{151} Additionally, firefly luciferase can be refolded by rapid dilution at low concentrations (100 nM) and low temperatures (10 °C) at a pH of 7.8. Utilizing this strategy, 14.7 \( \mu \)M firefly
luciferase was denatured in 5 M GdmCl followed by rapid dilution to a final concentration of 147 nM. Under these conditions significant levels of refolded and active firefly luciferase were obtained, with 73% activity as compared to native non-denatured luciferase (147 nM) being obtained 24 hours after dilution (Figure 3.5a). Having in hand conditions under which denatured luciferase could be refolded, we next sought to determine if firefly luciferase refolding could be achieved under condition favorable to zinc finger protein refolding, such as 200 μM Zn\(^{2+}\). Initially, the addition of 200 μM Zn\(^{2+}\) was found to be detrimental to both refolded and native luciferase, with background levels of luminescence being observed upon the incubation of native firefly luciferase with 200 μM Zn\(^{2+}\) as well as with firefly luciferase refolded in the presence of 200 μM Zn\(^{2+}\). Subsequent titrations determined that refolding with low levels of zinc were tolerated, resulting in only a 15% reduction in firefly luciferase activity when refolded in the presence of 10 μM Zn\(^{2+}\) (Figure 3.5b).

Using the above conditions (denaturation, rapid dilution, 10 μM Zn\(^{2+}\)) we attempted to denature and refold the split-firefly luciferase fusion proteins NFluc (1-437)-Zif268 and PBSII-CFluc (438-550) in the presence of a dsDNA target in order to dsDNA-dependent luminescence. Equimolar amounts (300 nM final) of the two purified fusion proteins were denatured and refolded in the presence of either 300 nM (final) Zif268-0-PBSII or no dsDNA in the presence of increasing concentrations of Zn\(^{2+}\). Unlike full length firefly luciferase, no appreciable luminescence above background was observed 12 hours after dilution for any concentration of Zn\(^{2+}\) tested (Figure 3.5c). Subsequent attempts to systematically optimize refolding conditions utilizing varying concentrations of proteins and dsDNA target also proved unsuccessful.
Figure 3.5. Refolding of denatured full length firefly luciferase and zinc finger-firefly luciferase fusions.  a) Luminescence of native and refolded firefly luciferase 5 and 24 hours after refolding was initiated by rapid dilution of 5 M GdmHCl denatured firefly luciferase.  b) Luminescence of denatured firefly luciferase 5 hours after refolding in the presence of increasing concentrations of Zn$^{2+}$.  c) Luminescence of denatured NF_Luc(1-437)-PBSII and Zif268-CFluc (438-550) 12 hours after refolding in the presence of 300 nM dsDNA (Zif268-0-PBSII) and varying concentrations of Zn$^{2+}$. 
3.2.3 Creation of MBP-PBSII-CFluc

To determine if the fusion proteins NFluc (1-437)-Zif268 and PBSII-CFluc (438-550) were capable of binding the target Zif268-0-PBSII, varying concentrations of each protein were incubated with 1 μM Zif268-0-PBSII followed by PAGE analysis to determine the amount, if any, of dsDNA bound by each fusion protein (Figure 3.6a). From this analysis it was determined that purified PBSII-CFluc (438-550) was unable to bind Zif268-0-PBSII even when present in a 10-fold molar excess while the same excess of NFluc (1-437)-Zif268 bound 1 μM Zif268-0-PBSII almost completely. In light of this it was determined that our inability to observe dsDNA-dependent firefly luciferase reassembly may be a result of insufficient DNA binding by PBSII under our experimental conditions.

Previously we have employed the use of maltose binding protein (MBP) fusions to create a SEER-β-lactamase dsDNA sensor. Additionally, attachment of this protein has been utilized to help create soluble and stable zinc-finger fusions in previous studies. In an attempt to obtain a stable and effective zinc finger fusion PBSII-CFluc (438-550) was cloned into the MBP containing expression vector pMAL-c2x, creating the fusion protein MBP-PBSII-CFluc (438–550). As with the previous construct, MBP-PBSII-CFluc (438-550) was soluble, but PAGE analysis determined that MBP-PBSII-CFluc (438–550) was unable to bind 1 μM Zif268-0-PBSII even when present in a 10-fold molar excess (Figure 3.6b).
Figure 3.6. PAGE gels showing extent of target dsDNA binding by fusion proteins. a) PAGE gel showing the extent of Zif268-0-PBSII target dsDNA binding by His-CFluc(438-550)-PBSII and His-NFluc(1-437)-Zif268. Lane 1: 10 μM His-CFluc(438-550)-PBSII plus 1 μM Target, Lane 2: 7.5 μM His-CFluc(438-550)-PBSII plus 1 μM Target, Lane 3: 5 μM His-CFluc(438-550)-PBSII plus 1 μM Target, Lane 4: 10 μM His-CFluc(438-550)-PBSII, Lane 5: 1 μM Target, Lane 6: 10 μM His-NFluc(1-437)-Zif268 plus 1 μM Target, Lane 7: 7.5 μM His-NFluc(1-437)-Zif268 plus 1 μM Target, Lane 8: 5 μM His-NFluc(1-437)-Zif268 plus 1 μM Target, Lane 9: 10 μM His-NFluc(1-437)-Zif268. * denotes His-NFluc(1-437)-Zif268 bound Zif268-0-PBSII dsDNA target. b) PAGE gel showing the extent of Zif268-0-PBSII target dsDNA binding by MBP-CFluc(438-550)-PBSII. Lane 1: 10 μM MBP-CFluc(438-550)-PBSII, Lane 2: 1 μM Target, Lane 3: 1 μM MBP-CFluc(438-550)-PBSII plus 1 μM Target, Lane 4: 2.5 μM MBP-CFluc(438-550)-PBSII plus 1 μM Target, Lane 5: 5 μM MBP-CFluc(438-550)-PBSII plus 1 μM Target, Lane 6: 7.5 μM MBP-CFluc(438-550)-PBSII plus 1 μM Target, Lane 7: 10 μM MBP-CFluc(438-550)-PBSII plus 1 μM Target.
3.2.4 Cell-Free dsDNA-Dependent Firefly Luciferase Reassembly

Given the inability of previously utilized purification and refolding methods to yield dsDNA-dependent firefly luciferase reassembly, our next approach sought to eliminate these factors completely through the use of available in vitro translation methodologies. As an alternative to the luciferase fragments utilized above, we chose to append the zinc fingers Zif268 and PBSII to a pair of luciferase fragments identified by Luker and coworkers, creating the fusions PBSII-NFluc(2-416) and CFluc(398-550)-Zif268 (Figure 3.7a). These firefly luciferase fusions were then evaluated for their ability to reassemble in the presence of target dsDNA utilizing in vitro transcribed mRNA in a commercially available rabbit reticulocyte lysate in vitro translation system (Figure 3.7b). Unlike previous efforts, this approach resulted in significant and concentration dependent luminescence in the presence of the dsDNA target Zif268-10-PBSII (Figure 3.7c), demonstrating the successful dsDNA-dependent reassembly of split-firefly luciferase. Finally, having a viable strategy in hand we sought to further demonstrate the dsDNA-dependent reassembly of firefly luciferase and to evaluate the effect of target site spacing on the efficiency of firefly luciferase reassembly. To test this, six dsDNA targets were tested with spacings of 0, 1, 2, 3, 6, and 10 base pairs separating the Zif268 and PBSII zinc finger binding sites. mRNA corresponding to both fusion proteins PBSII-NFluc (2-416) and CFluc (398-550)-Zif268 was used to generate fusion proteins in vitro in the presence of 25 nM of each target dsDNA in addition to a no dsDNA control (Figure 3.7d). For each target a significant amount of luminescence was observed compared to no dsDNA, with a distinct decrease in signal when six base pairs are present between the Zif268 and PBSII target sites.
Figure 3.7. dsDNA-dependent firefly luciferase reassembly in the rabbit reticulocyte *in vitro* translation system. a) Firefly luciferase-zinc finger fusions used to demonstrate dsDNA-dependent firefly luciferase reassembly in a rabbit reticulocyte *in vitro* translation system. b) Schematic of the cell-free dsDNA dependent reassembly of firefly luciferase. *In vitro* transcribed mRNA corresponding to the fusion proteins PBSII-NFluc(2-416) and CFluc(398-550)-Zif268 is incubated for 90 minutes in rabbit reticulocyte lysates in the presence of the dsDNA target Zif268-0-PBSII followed by luminescence measurement. c) Luminescence of *in vitro* translated PBSII-NFluc(2-416) and CFluc(398-550)-Zif268 in the presence of varying concentrations of the dsDNA target Zif268-0-PBSII. d) Luminescence of *in vitro* translated PBSII-NFluc(2-416) and CFluc(398-550)-Zif268 in the presence of target dsDNA (25 nM each) as a function of increasing spacing between Zif268 and PBSII target sites. e) Luminescence of *in vitro* translated NFluc(1-437)-Zif268 and PBSII-CFluc(438-550) in the presence of the dsDNA targets Zif268-0-PBSII and Zif268-10-PBSII (50 nM each).
In addition to the alternative zinc finger-firefly luciferase fusions described above, *in vitro* translation was also used to evaluate our original fusion proteins NFluc (1-437)-Zif268 and PBSII-CFluc (438-550) for dsDNA-dependent firefly luciferase reassembly in the presence of the dsDNA targets Zif268-0-PBSII and Zif268-10-PBSII (50 nM) (Figure 3.7e). *In vitro* translation of this initial protein pair produced no significant luminescence in the presence of either dsDNA target.

### 3.2.5 Conclusion

In conclusion, I sought to demonstrate the dsDNA-dependent reassembly of split-firefly luciferase. Previously, purified fusion proteins from recombinant sources had been used to achieve the dsDNA-dependent reassembly of GFP and β-lactamase. These efforts took the form of either denatured components which were refolded in the presence of a target dsDNA, or the use of MBP-tagged proteins which could be purified in their soluble form which were capable of dsDNA binding and dsDNA-dependent reassembly upon co-incubation with a specific dsDNA target. In order to demonstrate the dsDNA-dependent reassembly of firefly luciferase both of these strategies were attempted with an initial pair of zinc finger-firefly luciferase fusions, neither of which was successful. Subsequent PAGE analysis of dsDNA binding revealed that one fusion construct, namely PBSII-CFluc, was unable to bind dsDNA in any appreciable amount under purified conditions, possibly due to misfolding as a consequence of purification or PBSII-attachment. To circumvent the need for protein purification a cell-free methodology was utilized which relied upon the use of *in vitro* translation within rabbit reticulocyte lysates as a means of protein synthesis. Additionally, the alternative pair of zinc finger-firefly luciferase fusion
proteins, PBSII-NFluc (2-416) and CFluc (398-550)-Zif268, was utilized. Ultimately, the use of in vitro translation proved successful, resulting in a distinct dsDNA-dependent luminescent signal. Surprisingly, our original protein constructs NFluc(1-437)-Zif268 and PBSII-CFluc(438-550) also failed to produce a distinct dsDNA-dependent signal under these conditions, suggesting that this split-protein pair is not robust under any of the conditions we have utilized. Significantly, the use of in vitro translation greatly reduces the time and effort needed in order to produce a dsDNA-dependent signal, allowing the detection of specific dsDNA sequences without the need for purified components and laborious refolding procedures.

3.3 Experimental Procedures

3.3.1 Cloning Expression, and Purification of Full Length Firefly Luciferase and SEER-FFluc Proteins

All enzymes were purchased from NEB, dNTPs were purchased from Fermentas. A pET plasmid containing the E. coli optimized firefly luciferase gene was kindly provided by Craig Aspinwall. Zinc buffer A is 100 mM Tris at pH 7.5, 90 mM KCl, 1 mM MgCl2, 100 μM ZnCl2, and 5 mM DTT. Buffer I is 100 mM Tris-HCl, 200 mM KCl, 0.5 mM EDTA, and 0.5 mM DTT at pH 7.8.

The following primers were used to clone full length luciferase the 1st multiple cloning site of the pETDuet-1 vector: Fwd, 5’-GCGCGCGATCCGAAGACGCCAACATA-3’; and Rev, 5’- GCGCGCAAGCTTTTACAATTTGGACTTTCCGCC -3’.

The following primers were used to clone NFluc, residues 1-437, into an existing cassette containing a 15-amino acid linker followed by the zinc finger Zif268: Fwd, 5’-
The following primers were used to clone CFluc, residues 438-550, into an existing cassette containing the zinc finger PBSII and a 15-amino acid linker: Fwd, 5’-GCGCGCGGTACCCTTTGAAAGTCTTTAATTTAAA-3’; and Rev, 5’- GCGCGCAAGCTT TTACAATTTGGACCTTTCCGCC-3’.

The following primers were used to clone CFluc, residues 398-550, into an existing cassette containing a 15-amino acid linker followed by the zinc finger Zif268: Fwd, 5’-GCGCGCCCATGGGTATGTCCGGTTATGTA-3’; and Rev, 5’- ATATATCC TGCAGGGCAATTTGGACTTTCC-3’.

The following primers were used to clone NFluc, residues 2-416, into an existing cassette containing the zinc finger PBSII and a 15-amino acid linker: Fwd, 5’- GCGCGCGGTACCCTTTGAAAGTCTTTAATTTAAA-3’; and Rev, 5’- ATATATGCGGCGCTTATCAT CCATCCTTTGCAAT-3’.

The following primers were used to clone NFluc, residues 2-416, into an existing cassette containing an N-terminal MBP tag and a 15-amino acid linker followed by the zinc finger PBSII: Fwd, 5’-GCGCGCGGTACCCTTTGAAAGTCTTTAATTTAAA-3’; and Rev, 5’- GCGCGCAAGCTT TTACAATTTGGACCTTTCCGCC-3’.

Purification of full length firefly luciferase: The pETDuet plasmid containing the His6-firefly luciferase gene (Figures 3.8 and 3.14) was transformed into electro-competent BL21(DE3) cells (Novagen). An over night of this culture was used to inoculate a one-liter culture of 2XYT with 100 μg/mL ampicillin at an OD600 of 0.05. Protein expression was induced at an OD600 of 0.8 with 1 mM isopropyl β-D-
Figure 3.8. DNA and amino acid sequence of His6-Firefly Luciferase. Firefly luciferase is shown in red.
thiogalactopyranoside for 6 h. at 22 °C. Cells were pelleted and frozen over night at -20 °C. Pelleted cells were thawed on ice and resuspended in phosphate buffered saline (PBS). Resuspended cells were lysed by sonication. The lysate was cleared by centrifugation at 18,000 cfm for 30 min. SDS-PAGE analysis showed that both proteins expressed in the soluble fraction. His$_6$-tagged proteins were purified over Ni-NTA agarose resin (Qiagen). Lysate was added to 1 mL of resin and allowed to equilibrate for 1 hour at 4°C. Protein was eluted from the resin by the addition of PBS containing increasing concentrations of imidazole (2, 10, 20, 50, and 500 mM respectively) and characterized by SDS-PAGE. Fractions containing protein were dialyzed overnight into PBS to reduce concentrations of imidazole. Concentrations were determined in 6 M guanidine using absorbance measurements at 280 nm (calculated $\varepsilon = 39,560 \text{ M}^{-1} \text{ cm}^{-1}$).

Purification of NFluc(1-437)-Zif268 and PBSII-CFluc(438-550): The pETDuet plasmids containing either His$_6$-NFluc(1-437)-Zif268, or His$_6$-PBSII-CFluc(438-550) genes (Figures 3.9, 3.10, and 3.14) were transformed into electro-competent BL21(DE3) cells (Novagen). An over night of this culture was used to inoculate a one-liter culture of 2XYT with 100 µg/mL ampicillin and 100 µM ZnCl$_2$ at an OD$_{600}$ of 0.05. Protein expression was induced at an OD$_{600}$ of 0.8 with 1 mM isopropyl $\beta$-D-thiogalactopyranoside for 6 h at 22 °C. Cells were pelleted and frozen over night at -20 °C. Pelleted cells were thawed on ice and resuspended in zinc buffer A. Resuspended cells were lysed by sonication. The lysate was cleared by centrifugation at 18,000 cfm for 30 min. SDS-PAGE analysis showed that both proteins expressed in the soluble fraction. His$_6$-tagged proteins were purified over Ni-NTA agarose resin (Qiagen). Lysate was added to 1 mL of resin and allowed to equilibrate for 1 hour at 4°C. Protein was eluted
Figure 3.9. DNA and amino acid sequence of His<sub>6</sub>-NF<sub>Luc</sub>(1-437)-Zif268. NF<sub>Luc</sub> is shown in red, the 15 amino acid linker is shown in green, and Zif268 is shown in blue.
Figure 3.10. DNA and amino acid sequence of His<sub>e</sub>-PBSII-CFluc(438-550). PBSII is shown in blue, the 15 amino acid linker is shown in green, and CFluc is shown in red.
from the resin by the addition of zinc buffer A containing increasing concentrations of imidazole (2, 10, 20, 50, and 500 mM respectively) and characterized by SDS-PAGE. Fractions containing protein were dialyzed overnight into zinc buffer A to reduce concentrations of imidazole. Concentrations were determined in 6 M guanidine using absorbance measurements at 280 nm (calculated $\varepsilon$: NFuc (1-437)-Zif268 = 38,445 M$^{-1}$ cm$^{-1}$; PBSII-CFluc (438-550) = 9,315 M$^{-1}$ cm$^{-1}$).

Purification of MBD-PBSII-CFluc(438-550): The pMAL-c2X plasmid containing the MBD-PBSII-CFluc(438-550) gene (Figures 3.11 and 3.14) was transformed by electroporation into BL21 (DE3) cells (Novagen). An overnight culture of these cells was used to inoculate a 100-mL culture of 2xYT media supplemented with 100 μg/mL ampicillin, 100 μM ZnCl$_2$, and 2% glucose at an OD$_{600}$ of 0.05. Protein expression was induced at an OD$_{600}$ of 0.8 with 1 mM isopropyl β-D-thiogalactopyranoside for 6 h at 22 °C. Cells were pelleted by centrifugation and frozen overnight at -20 °C. Pelleted cells were thawed on ice and resuspended in zinc buffer A. Resuspended cells were lysed by sonication. The lysate was cleared by centrifugation at 18,000 cfm for 30 min. MBP-tagged β-lactamase fusions were purified over amylose resin (New England Biolabs) using the following protocol. 3 mL of resin was loaded into a PD-10 column and equilibrated with 5 column volumes of zinc buffer A. Following the addition of cleared lysates, the column was washed with 10 column volumes of zinc buffer A supplemented with 2 M NaCl followed by an additional wash with 8 column volumes of zinc buffer A. Fusions were eluted with 5 mL of zinc buffer A supplemented with 10 mM maltose and characterized by SDS-PAGE. Concentrations were determined in 6 M guanidine using absorbance measurements at 280 nm (calculated $\varepsilon = 75,665$ M$^{-1}$ cm$^{-1}$).
In vitro translation of PBSII-NFluc(2-416) and CFluc(398-550)-Zif268: Fragments encoding PBSII-NFluc(2-416) and CFluc(398-550)-Zif268 (Figures 3.12, 3.13, and 3.14) were generated by PCR starting from the specific plasmids. The fusion protein constructs were generated using standard cloning techniques and verified by sequencing. The mRNA necessary for cell-free assays was generated as follows: PCR fragments corresponding to the desired fusion constructs were generated using a forward primer containing a T7 RNA polymerase promoter and Kozak sequence and a reverse primer containing a 3’ hairpin loop. The purified PCR products were subsequently used as template for in vitro transcription using the RiboMAX Large Scale RNA Production System-T7 (Promega) following the manufacturer’s protocols. These generated mRNA molecules were then used as templates for protein synthesis utilizing the Flexi-Rabbit Reticulocyte Lysates (Promega) according to the manufacturer’s protocol using 2 pmol of each mRNA encoding for either PBSII-NFluc(2-416) and CFluc(398-550)-Zif268.

3.3.2 Determination of Luciferase Concentration Dependent Luminescence

To determine the minimal amount of luciferase detectable decreasing concentrations of full length firefly luciferase in PBS was mixed with 50 μL of 0.2 mM ATP, 10 μL of 1 mM D-luciferin, and PBS to a final volume of 100 μL in 96-well plates (CoStar, Corning). Luminescence measurements were collected every 30 seconds for 5 minutes. Data is represented as the luminescence after 5 minutes.
Figure 3.11. DNA and amino acid sequence of MBP-PBSII-CIuc(438-550). PBSII is shown in blue, the 15 amino acid linker is shown in green, and CIuc is shown in red.
Figure 3.12. DNA and amino acid sequence of PBSII-NFluc(2-416). PBSII is shown in blue, the 15 amino acid linker is shown in green, and NFluc is shown in red.
Figure 3.13. DNA and amino acid sequence of CFLuc(398-550)-Zif268. CFLuc is shown in red, the 15 amino acid linker is shown in green, and Zif268 is shown in blue.
3.3.3 Evaluation of dsDNA-Dependent Luminescence Utilizing Soluble Purified His$_6$-NFluc(1-438)-Zif268 and His$_6$-PBSII-CFluc(438-550).

To determine the initial activity of the SEER-FFluc sensor His$_6$-NFluc(1-438)-Zif268 and His$_6$-PBSII-CFluc(438-550) (5 μM) were mixed with the dsDNA target Zif268-0-PBSII or Zif268-10-PBSII (2.5 μM) in zinc buffer A to a volume of 160 μL. Following a 60 minute incubation 20 μL of 1 mM D-luciferin and 20 μL of 1 mM ATP to a final volume of 200 μL. Luminescence measurements were collected every 30 seconds for 5 minutes in 96-well plates. Data is represented as the luminescence after 5 minutes.

3.3.4 Denaturation and Refolding of Full Length and Split-Firefly Luciferase

To refold firefly luciferase purified full length was diluted to a concentration of 14.7 μM in Buffer I containing 5 M guanadinium chloride (GdmCl) to denature. Following a 3 hour incubation at 20 °C the denatured luciferase diluted 1:100 to a final concentration of 147 nM firefly luciferase and 50 mM GdmCl. At 5 and 24 hours 160 μL of refolded firefly luciferase was mixed with 20 μL of 1 mM D-luciferin and 20 μL of 1 mM ATP to a final volume of 200 μL. Luminescence measurements were collected every 30 seconds for 5 minutes in 96-well plates. Data is represented as the luminescence after 5 minutes.

Utilizing this protocol firefly luciferase was refolded at a final concentration of 147 nM in zinc buffer A in the presence of increasing concentrations of Zn$^{2+}$ (0, 10, 25, 50, 100, and 200 μM). After 5 hours 160 μL of refolded firefly luciferase from each Zn$^{2+}$ concentration was mixed with 20 μL of 1 mM D-luciferin and 20 μL of 1 mM ATP to a
final volume of 200 μL. Luminescence measurements were collected every 30 seconds for 5 minutes in 96-well plates. Data is represented as the luminescence after 5 minutes.

3.3.5 Denaturation and Reassembly of SEER-FFluc Proteins

Equimolar amounts of His$_6$-NFluc(1-438)-Zif268 and His$_6$-PBSII-CFluc(438-550) were denatured and refolded at a final concentration of 300 nM each in zinc buffer A containing 300 nM Zif268-0-PBSII and increasing concentrations of Zn$^{2+}$. As a comparison full length firefly luciferase was denatured and refolded at a final concentration of 300 nM in zinc buffer A. After 12 hours 160 μL of refolded firefly luciferase and SEER-FFluc components from each Zn$^{2+}$ concentration was mixed with 20 μL of 1 mM D-luciferin and 20 μL of 1 mM ATP to a final volume of 200 μL. Luminescence measurements were collected every 30 seconds for 5 minutes in 96-well plates. Data is represented as the luminescence after 5 minutes.

3.3.6 Determination of dsDNA binding by SEER-FFluc Proteins

5, 7.5 and 10 μM of His$_6$-NFluc(1-438)-Zif268 and His$_6$-PBSII-CFluc(438-550), and 1, 2.5, 5, 7.5, and 10 μM MBP-PBSII-CFluc(438-550) were incubated with 1 μM Zif268-0-PBSII for 60 minutes in zinc buffer A. 15 μL of each sample was analyzed via PAGE and stained with ethidium bromide to determine DNA binding.

3.3.7 Cell-Free dsDNA-Dependent Reassembly of Firefly Luciferase

Duplicate 25 μL translations were carried out in Flexi-Rabbit Reticulocyte Lysate (Promega) according to the manufacturer’s protocol using 2 pmols of each mRNA
encoding for PBSII-NFluc(2-416) and CFluc(398-550)-Zif268, 10 μM ZnCl₂, 0.5 μL RNasin Plus (Promega), and either 100 nM or 250 nM Zif268-10-PBSII dsDNA target or no target. Translations were incubated at 30 °C for 90 minutes and assayed by combining 20 μL of translation solution with 80 μL of Steady-Glo Luciferase Assay System (Promega). The final concentration of Zif268-10-PBSII in these assays was 20 nM and 50 nM. Light emission was monitored 1 minute after Steady-Glo addition using a Turner TD-20e luminometer with a 3 second delay and a 10 second integration time.

To determine the effect of target site spacing on firefly luciferase reassembly duplicate 25 μL reaction were carried out as described above in the presence of 250 nM target dsDNA containing either 0, 1, 2, 3, 6, or 10 bases between the Zif268 and PBSII binding sites (Table 3.1). Translations were incubated at 30 °C for 90 minutes and assayed by combining 20 μL of translation solution with 80 μL of Steady-Glo Luciferase Assay System (Promega). The final concentration of target dsDNA in these assays was 50 nM. Light emission was monitored 1 minute after Steady-Glo addition using a Turner TD-20e luminometer with a 3 second delay and a 10 second integration time.

To evaluate the fusion proteins NFluc(1-437)-Zif268 and PBSII-CFluc(438-550) for their ability to reassemble under cell-free in vitro conditions duplicate 25 μL translations were carried out in Flexi-Rabbit Reticulocyte Lysate (Promega) according to the manufacturer’s protocol using 2 pmols of each mRNA encoding for NFluc(1-437)-Zif268 and PBSII-CFluc(438-550), 10 μM ZnCl₂, 0.5 μL RNasin Plus (Promega), and 250 nM Zif268-0-PBSII or Zif268-10-PBSII dsDNA target or no target. Translations were incubated at 30 °C for 90 minutes and assayed by combining 20 μL of translation solution with 80 μL of Steady-Glo Luciferase Assay System (Promega). The final
concentration of Zif268-0-PBSII and Zif268-10-PBSII in these assays was 50 nM. Light emission was monitored 1 minute after Steady-Glo addition using a Turner TD-20e luminometer with a 3 second delay and a 10 second integration time.
<table>
<thead>
<tr>
<th>Spacing</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zif268-0-PBSII</td>
<td>GCGTACGCGTGGCGGTGTGGAAACACCG</td>
</tr>
<tr>
<td>Zif268-1-PBSII</td>
<td>GCGTACGCGTGGCGGTGTGGAAACACCG</td>
</tr>
<tr>
<td>Zif268-2-PBSII</td>
<td>GCGTACGCGTGGCGTGGTGAAACACCG</td>
</tr>
<tr>
<td>Zif268-3-PBSII</td>
<td>GCGTACGCGTGGCGTGTGGAAACACCG</td>
</tr>
<tr>
<td>Zif268-6-PBSII</td>
<td>GCGTACGCGTGGCGTGTTAGTGTAAGTGAAACACCG</td>
</tr>
<tr>
<td>Zif268-10-PBSII</td>
<td>GCGTACGCGTGGCGTGGGAGTGTAAGTGAAACACCG</td>
</tr>
</tbody>
</table>

**Table 3.1.** dsDNA targets used to evaluate the target site spacing dependence of firefly luciferase reassembly (SEER-FFluc). Blue and Red indicate the Zif268 and PBSII sites respectively.
CHAPTER 4
DEVELOPMENT OF A GENERAL CELL-FREE SPLIT-FIREFLY LUCIFERASE METHODOLOGY FOR THE RAPID INTERROGATION OF BIOMOLECULAR INTERACTIONS

4.1 Introduction

Having demonstrated the dsDNA-dependent reassembly of firefly luciferase using cell-free in vitro translation systems, we sought to investigate the use of this strategy to develop a rapid and general assay system for the interrogation of protein-protein, protein DNA, and protein-RNA interactions. The entirety of this paper entitled “A General and Rapid Cell-Free Approach for the Interrogation of Protein-Protein, Protein-DNA and Protein-RNA Interactions and their Antagonists Utilizing Split-Protein Reporters” is presented in Appendix B. I was responsible for the development and application of the split-luciferase reporter strategy for the interrogation of a large range of biomolecular interactions, Cliff Stains was responsible for the initial utilization of cell-free translation systems and the development of the RNA detection strategy utilizing pumilio domains, and Ben Jester assisted us in the construction of the protein-protein interaction pairs. The following is a summary of the most important findings within the document.

Protein-protein\textsuperscript{153} and protein-nucleic acid\textsuperscript{154} interactions are central to cellular function and are also emerging targets for pharmacological intervention when implicated in a particular disease pathway. Thus numerous in vitro and in vivo methods have been developed to target\textsuperscript{155-159} and study these biomolecular interactions. Widely utilized in vitro methods for interrogating protein-protein and protein-DNA interactions and their antagonists include variations of enzyme linked immunosorbent assays (ELISAs), surface
plasmon resonance (SPR), and fluorescence polarization (FP), which either require the use of antibodies or purified proteins in addition to the use of chemically modified components. On the other hand powerful in vivo methods such as the yeast two-hybrid \cite{42} assays have the advantage of speed by eliminating the need for protein purification but can be subject to false positives and negatives due to the multifactorial nature of signal generation. \cite{160} In between these two extremes lies the protein fragment based methods, where a specific biomolecular interaction drives the reassembly of a previously split reporter protein. \cite{4}

The reconstitution of a functional protein from split-peptide fragments was first demonstrated for ribonuclease in 1959. \cite{161} Since then “split-protein reassembly” or “protein complementation” has been applied to the in vivo detection of a wide variety of protein-protein interactions utilizing numerous split-protein hosts including ubiquitin, \cite{7} beta-galactosidase, \cite{162} dihydrofolate reductase, \cite{9} β-lactamase, \cite{20} GFP, \cite{163} GFP-variants, \cite{164,165} firefly luciferase \cite{21} and Gaussia luciferase. \cite{17} Recently, we and others have also described methods for detecting nucleic acids and their chemical modification by the reassembly of ternary complexes of split-GFP and split-β-lactamase tethered to nucleic acid binding proteins. \cite{71,166-168} Thus split-protein systems or “protein complementation assays” (PCAs) can directly image most biomolecular interactions. Though the current methods are of great utility, all of the current split-protein methods have certain limitations for interrogating protein-protein and protein-nucleic acid interactions and their inhibitors in a very rapid and high-throughput fashion. For example, current in vitro methods require extensive protein purification \cite{163,169} and chemical modification in addition to relying on proper folding of recombinant proteins, while in vivo methods
require lengthy transfection and propagation of cellular cultures prior to analysis, both approaches being time intensive.\textsuperscript{168,170} The current methods are also prone to problems arising from potential proteolysis of intracellularly expressed proteins and peptides as well as a lack of control over interfering co-expressed cellular factors as is also the case with yeast n-hybrid methods.

To provide a rapid and general method that circumvents many of the limitations discussed above, we hypothesized that fragmented reporter proteins fused to functional proteins could be rapidly generated directly from mRNA utilizing cell-free translation methodologies and immediately interrogated for biomolecular interaction dependent signal generation. The use of split-proteins in cell-free translation takes advantage of fast protein synthesis rates, from 60 to 90 minutes, and easy adaptation to homogeneous assays avoiding immobilization and washing protocols. Herein we demonstrate how this cell-free approach provides a general platform for rapidly detecting protein-protein, protein-small molecule, protein-DNA, protein-methylated DNA, and protein-RNA interactions starting from mRNA or directly from DNA corresponding to the desired interaction pair in less than two hours. Additionally we demonstrate the reversibility of the split-luciferase reporter and its use in interrogating antagonists of protein-nucleic acid interactions and in the determination of sequence specificity of nucleic acid-binding domains.
4.2 Results and Discussion

4.2.1 Initial Evaluation of the Reassembly of Split-Reporter Proteins in Cell-Free Translation Systems

Initially, we evaluated the ability of our previously reported split-GFP\textsuperscript{71} and split-β-lactamase\textsuperscript{167,169} systems appended to specific zinc-fingers, to reassemble in the presence of target DNA utilizing in vitro transcribed mRNA in a purified wheat germ extract translation system (Figure 4.1a). Signal from the DNA dependent reassembled GFP\textsuperscript{71} was too low to observe over background using standard fluorescence measurements (Figure 4.1b), while DNA dependent β-lactamase activity\textsuperscript{169} yielded measurable but low signal-to-background ratios (Figure 4.1c). Thus we turned to recently reported \textit{in vivo} split-luciferase systems which have the significant advantage of negligible background from the translation system due to the generation of a bioluminescent signal (Figure 4.1d). We chose to first examine the fragmented firefly luciferase (Fluc) reported by Luker \textit{et. al}.\textsuperscript{14} which when appended to our zinc fingers showed significant signal over background luminescence upon addition of target DNA (Figure 4.1e). This comprises the first demonstration of the bioluminescent read-out of a specific nucleic acid sequence and this split-luciferase system was chosen for further studies in cell-free systems. The very recently described split-\textit{Gauussia} luciferase\textsuperscript{17} and alternatively split-firefly luciferase (Paulmurugan \textit{et. al}.\textsuperscript{14,171}) complementation systems, were also tethered to our zinc-fingers and displayed extremely good signal over background bioluminescence (Figure 4.1f and g) in a DNA dependent fashion and are currently being further explored for evaluating their relative merits in cell-free assays.
Figure 4.1. Cell-free detection utilizing split-proteins. a) Cartoon representation of a split protein-system with zinc-fingers tethered to the split-protein in the presence of target dsDNA oligonucleotide. Different split-protein reporters tethered to sequence specific zinc-fingers in the presence and absence of target dsDNA; b) split-Venus (a green fluorescent protein variant); c) split-β-lactamase; d) Cartoon representation of dsDNA-dependent firefly luciferase reassembly; e) Split-firefly luciferase as described by Luker et al.; f) split-firefly luciferase as described by Paulmurugan et al. g) split-Gaussia luciferase as described by Remy et al. Reprinted with permission from the American Chemical Society.
4.2.2 Cell-Free Interrogation of Protein and Protein-Nucleic Acid Interactions

To test the generality of the cell free split-luciferase approach we chose to investigate five well characterized biomolecular interactions (Figure 4.2); (a) the catalytic subunit of cAMP-dependent protein kinase (PKA) with its inhibitor PKI (PKA/PKI);\textsuperscript{172-174} (b) and the interaction between hypoxia inducible factor-1\(\alpha\) (HIF-1\(\alpha\)) and the CH1 domain of the transcriptional coactivator p300 (HIF-1\(\alpha\)/p300).\textsuperscript{175,176} Akin to the yeast three hybrid systems,\textsuperscript{177} we also investigated the ternary association of (c) two sequence specific zinc-fingers\textsuperscript{178} with a target DNA (Zif268/PBSII); (d) a zinc-finger and methyl CpG-binding domain with a target CpG-methylated DNA (Zif268/MBD2);\textsuperscript{179-182} and finally (e) two RNA-specific pumilio domains\textsuperscript{74} with a target RNA (Pum1/Pum2). To complement the interacting protein pairs described above, we also tested the ability of this cell-free assay to report upon the small-molecule dependent interaction between two proteins (Figure 4.3a), namely the rapamycin dependent interaction between the human FK506-binding protein 12 (FKBP) and the FKBP12-rapamycin binding (FRB) domain of human mTOR (FKBP/FRB)\textsuperscript{183-185} (Figure 4.3b). Additionally, A rapamycin concentration dependant increase in luminescence was observed as expected from the cell-free translations of the split reporters (Figure 4.3c).\textsuperscript{14} The overall sensitivity (signal/background) of these systems were excellent and varied from 22 to 1800-fold, while the total assay time from translation to analysis was less than two hours.
Figure 4.2. Cell-free detection of a wide variety of biomolecular interactions utilizing split-luciferase starting from mRNA. Detection of (a) protein-protein interaction between PKI-NFluc and CFImc-PKA; (b) protein-protein interaction between p300-NFluc and CFImc-HIF-1α; (c) protein-DNA interaction between PBSII-NFluc, CFImc-Zif268, and a target dsDNA oligonucleotide; (d) methylation dependent protein-DNA interaction between MBD2-NFluc, CFImc-Zif268, and a target methylated CpG dsDNA oligonucleotide; and (e) protein-RNA interaction between Pum2-NFluc, CFImc-Pum1, and a target RNA oligonucleotide. Reprinted with permission from the American Chemical Society.
Figure 4.3. Cell-free interrogation of a small-molecule dependent protein-protein interaction.  
a) Schematic of the cell-free interrogation of a small-molecule dependent protein-protein interaction.  
b) Detection of the rapamycin-dependent interaction between FRB-NFluc and CFluc-FKBP.  
c) Concentration dependent association of FRB-NFluc and CFluc-FKBP mediated by the small-molecule rapamycin (inset).  
Reprinted with permission from the American Chemical Society.
4.2.3 Cell-Free Split-Luciferase Reassembly using a Coupled Rabbit Reticulocyte Lysate System

In addition to the cell-free interrogation of biomolecular interactions starting from \textit{in vitro} transcribed mRNA we also investigated the use of a coupled rabbit reticulocyte lysate system for the reassembly of split-firefly luciferase.\textsuperscript{186} In this system fusion protein transcription and translation occur \textit{in situ} from a DNA template containing a T7 Promoter and a Kozak sequence, which may eventually allow for eliminating the need for a separate \textit{in vitro} transcription step, and further simplify our assay format. Utilizing this approach, one could imagine the use of this assay format for the initial high-throughput interrogation of protein-protein and protein-nucleic acid interactions starting from a collection of plasmids encoding split-firefly luciferase fusions in coupled transcription and translation systems. In the case of protein-protein interactions, plasmids corresponding to positively identified interactions from this initial screen can then be readily transfected into appropriate cell lines for further study.

To test this, we evaluated the ability of a rabbit reticulocyte coupled system to report upon two previously studied interactions, namely the small-molecule dependent protein-protein interaction between FRB and FKBP, and the interactions between the zinc fingers Zif268 and PBSII with a target dsDNA oligonucleotide. In each case, the addition of the appropriate molecule, either rapamycin or the dsDNA target Zif268-0-PBSII (Z0P) resulted in split-firefly luciferase reassembly as observed by a significant increase in luminescence (Figure 4.4a and b). These results show the use of coupled lysates systems for the interrogation of protein-protein as well as specific protein-DNA
Figure 4.4. Cell-free detection of biomolecular interactions with split-luciferase utilizing coupled transcription and translation systems: a) the ternary protein-DNA interaction between PBSII-NFluc, CFLuc-Zif268, and 5nM target dsDNA oligonucleotide (Z0P) and b) the rapamycin-induced (5 nM) interaction between FRB-NFluc and CFLuc-FKBP. Detection of biomolecular interactions utilizing purified DNA in the PURESYSTEM classic II system consisting of completely purified transcriptional and translational components c) the ternary protein-DNA interaction between PBSII-NFluc, CFLuc-Zif268, and 5nM target dsDNA oligonucleotide (Z0P) and d) rapamycin-induced (5 nM) interaction between FRB-NFluc and CFLuc-FKBP. Reprinted with permission from the American Chemical Society.
interactions, demonstrating the potential application of this assay format to the high-throughput interrogation of protein-protein and protein-DNA interactions.

4.2.4 Split-Firefly Luciferase Utilizing Purified Transcription and Translation Components

Recently, Shimizu et al. has describe their effort to individually purify all 36 prokaryotic transcription and translation components, the recombination of which with purified ribosomes and necessary energy sources has lead to the creation of a completely purified coupled transcription and translation system. This PURE translation system allows for the production of target proteins within a chemically defined environment where all components are known. To determine whether or not the above observed split-firefly luciferase reassembly is effected by unknown lysates components we sought to utilize this PURE system to demonstrate split-firefly luciferase reassembly. Again we tested the rapamycin-induced interactions between FRB and FKBP in addition to the interaction between the zinc fingers Zif268 and PBSII with a target dsDNA target (Figure 4.4c and d). In each case either rapamycin- or target dsDNA-dependent luminescence was observed, showing the potential use of this chemically defined PURE translation system for the interrogation of protein-protein and protein-DNA interactions.

4.2.5 Reversibility of Split-Firefly Luciferase Reassembly

In addition to the detection of protein-nucleic acid interactions we also sought to evaluate the ability of our cell-free split-firefly luciferase assay format to report upon the activity of modulators of these interactions. In order to detect antagonists of protein-
nucleic acid interactions, we first needed to demonstrate the thermodynamic reversibility of the ternary complex consisting of reassembled firefly luciferase fragments tethered to two zinc fingers (PBSII and Zif268) and a target dsDNA target oligonucleotide (Figure 4.5a). Towards this goal, translations using mRNA encoding PBSII-NFluc and CFluc-Zif268 were initiated in the presence of the target dsDNA oligonucleotide \( \text{Zif268-0-PBSII} \). Post DNA-dependent firefly luciferase reassembly, a hairpin DNA (\( \text{hpDNA-Zif268} \)), which is a competitor for only Zif268 binding, was added at increasing concentrations followed by equilibration for 30 min. A concentration dependent decrease in luminescence was observed, clearly demonstrating that the ternary complex of the firefly luciferase was reversible and could be inhibited by addition of the dominant-negative (\( \text{hpDNA-Zif268} \)) oligonucleotide (Figure 4.5b, TGG containing hpDNA). The generality of utilizing the cell-free system for probing protein-nucleic acid inhibition was further demonstrated with translations containing mRNA encoding Pum2-NFluc and CFluc-Pum1 in the presence of target RNA. As with the dsDNA-dependent reassembly above, a concentration dependent decrease in luminescence was observed only upon the addition of increasing amounts of a competitor half-site RNA target that is known to selectively bind one of the pumilio domains\(^7\) (Figure 4.5e).

### 4.2.6 Cell-Free Determination Nucleic Acid-Binding Protein Specificity

Building upon the reversibility discussed above we envisioned that our split-firefly luciferase based cell-free system could be readily utilized to analyze the relative target site specificity of nucleic acid-binding proteins through competition binding experiments. A number of methods have been developed to interrogate the relative
Figure 4.5. Interrogation of protein-nucleic acid interactions utilizing our cell-free split-firefly assay. a) Cartoon showing dissociation of dsDNA dependent firefly luciferase ternary complex by the addition of a competitor hairpin DNA containing one of the two zinc-finger binding sites. b) Dissociation of the reassembled PBSII-NFluc, CFLuc-Zif268, dsDNA ternary complex by the addition of zif268 hairpin DNA targets containing TGG (wild type), TAG, TTG, and TCG triplet basepairs. c) and d) Previously reported relative affinities of target oligonucleotides with zif268 with IC50 values derived from the cell-free firefly luciferase reassembly assay and their correlation. e) Dissociation of the Pum2-NFluc, CFLuc-Pum1, RNA ternary complex by the addition of an RNA target containing a Pum1 binding site. Reprinted with permission from the American Chemical Society.
affinity of DNA-binding proteins for their target site, including traditional EMSAs and more recently DNA microarrays. Though powerful, these techniques require the use of purified components, specialized equipment, or radioactive materials. Having established that ternary zinc finger-DNA complexes can be disrupted by a competitor oligonucleotide added in trans, we next attempted to correlate the known binding affinities of Zif268 for single nucleotide changes in its binding site to IC$_{50}$ values obtained from our cell-free split-firefly luciferase assay in a 96-well format. To demonstrate this separate translation reactions of PBSII-NFluc and CFluc-Zif268 mRNA in the presence of the dsDNA target oligonucleotide, Zif268-0-PBSII were initiated. Duplicate experiments containing ternary complexes were allowed to assemble for 90 minutes followed by the addition of increasing concentrations of competitor hpDNA containing one of four different Zif268 binding sites having either A, T, C, or G at the central position. In each case a competitor hpDNA concentration dependent decrease in luminescence was observed (Figure 4.5b) within 30 min. IC$_{50}$ values for each competitor hpDNA (Figure 4.5c) were shown to correlate extremely well ($R^2 = 0.996$) (Figure 4.5d) with previously reported relative affinities of these target sites. These results serve to validate the application of the split-luciferase cell-free system for the determination of relative binding affinities of nucleic acid-binding proteins for their target sites and more generally in studying inhibitors of protein-nucleic acid interactions. Thus this cell-free system provides an attractive alternative to current methods for interrogating protein-nucleic acid binding as they can be performed in a simple, rapid, high-throughput and homogeneous format without having to purify or refold the protein of interest.
4.2.7 Conclusion

Numerous *in vitro* and *in vivo* methods are being developed to interrogate the interactions between proteins and nucleic acids, including strategies utilizing the reassembly of split-protein reporters. Current cell-based and *in vitro* strategies though powerful, depend on cumbersome steps which can include transfection, cell culture, purification, washing steps, and/or covalent modification yielding overall experimental times in excess of 12-48 hours starting from appropriate clones. Herein we have described a general platform for interrogating biomolecular interactions in homogeneous assays based on cell-free split-protein systems within two hours. This cell-free assay is capable of utilizing a variety of split protein reporters providing both fluorescent (β-lactamase) and bioluminescent (firefly and *Gaussia* luciferase) signal outputs. One drawback of the current method as also found in ELISA and *in vivo* approaches is that only relative affinities and IC₅₀ values can be determined unlike methods such as FP and SPR which are capable of providing accurate kinetic binding information. However, the speed and ease of implementation of this cell-free approach which does not require cell culture, protein purification, or chemical derivatization can be used to rapidly address biological and chemical questions with appropriate controls, as we have demonstrated with either dominant negative inhibition or known small molecule ligand.

Our experiments demonstrate the ability to detect a wide variety of protein-protein interactions, including the well studied interaction between the protein kinase PKA and its inhibitor PKI, and the small molecule-dependent interaction between FKBP and FRB. Additionally, we provide the first example for a rapid method for interrogating the interaction between HIF-1α and p300, an emerging protein-protein target implicated
in cancer progression. Furthermore, we detail the first examples of sensitive split-luciferase mediated detection of a wide range of protein-nucleic acid interactions, including zinc finger domains with specific dsDNA, a methyl CpG-binding domain with specific methylated DNA, and RNA binding pumilio domains with target RNA. We have also demonstrated that this methodology can be utilized to interrogate the relative binding affinities of nucleic-acid binding proteins for their target sites. In addition to using purified lysates and wheat germ extracts, we have demonstrated the detection of protein-protein and protein-nucleic acid using a system composed entirely of purified components that minimize non-specific interactions from cellular components and allow control over the translational machinery that may find utility in numerous applications involving unnatural amino acid incorporation.190-193

We envision that this cell-free format could be potentially used in a wide-variety of applications that include screening of DNA or RNA target sites for nucleic acid-binding proteins and the determination of target site preference. More importantly perhaps, this approach can be applied to the screening of small-molecules, nucleic acids, peptides, even or proteins for inhibition of specific protein-protein or protein-nucleic acid interactions.153,154 Moreover, since other split-protein approaches have been widely utilized within a cellular context, the initial hits from the rapid cell-free system can be rapidly applied in vivo within specific cell lines of interest.4,21 Thus we envision that this rapid, sensitive, and homogeneous assay system will be widely utilized for interrogating user-defined natural and unnatural biomolecular interactions and for evaluating agonists and antagonists of these interactions.
4.3 Experimental Procedures

4.3.1 General Materials

All materials were obtained from Sigma-Aldrich unless otherwise noted. ZnCl₂ was obtained from EM Sciences. Restriction enzymes were obtained from NEB and in vitro translational products from Promega. Oligonucleotide primers and targets were from IDT.

4.3.2 Plasmid Construction and mRNA Production.

The fusion protein constructs used in this study are shown in Table 4.1. Fragments coding for reporter protein fragments (GFP, β-Lactamase, and three Luciferases) were generated by PCR with appropriate primers and subsequently cloned into either the pETDuet-1 vector (Novagen) or the pMAL-c2x vector (NEB) using standard techniques and verified by sequencing. Fragments encoding the nucleic acid-binding proteins or associating proteins used in this study were generated by PCR starting from the specific plasmids. The fusion protein constructs were generated using standard cloning techniques and verified by sequencing. The mRNA necessary for cell-free assays was generated as follows: PCR fragments corresponding to the desired fusion constructs were generated using a forward primer containing a T7 RNA polymerase promoter and Kozak sequence and a reverse primer containing a 3’ hairpin loop. The purified PCR products were subsequently used as template for in vitro transcription using the RiboMAX Large Scale RNA Production System-T7 (Promega) following the manufacturer’s protocols.
<table>
<thead>
<tr>
<th>Reassembly Pairs</th>
<th>Split-Protein Fusion</th>
<th>Agonists</th>
</tr>
</thead>
<tbody>
<tr>
<td>NVenus-Zif268</td>
<td>Venus (1-157)</td>
<td>GCGTAGCGTGGGCGGTTGGAACACCCG</td>
</tr>
<tr>
<td>PBSII-CVenus</td>
<td>Venus (158-238)</td>
<td></td>
</tr>
<tr>
<td>NBetaLac-Zif268</td>
<td>Beta-Lactamase (26-196)</td>
<td>GCGTAGCGTGGGCGGTTGGAACACCCG</td>
</tr>
<tr>
<td>PBSII-CBetaLac</td>
<td>Beta-Lactamase (198-290)</td>
<td></td>
</tr>
<tr>
<td>PBSII-NFluc</td>
<td>Firefly Luciferase (2-416)</td>
<td>GCGTAGCGTGGGCGGTTGGAACACCCG</td>
</tr>
<tr>
<td>CFOuc-Zif268</td>
<td>Firefly Luciferase (398-550)</td>
<td></td>
</tr>
<tr>
<td>NFluc-zif268</td>
<td>Firefly Luciferase (2-398)</td>
<td>GCGTAGCGTGGGCGGTTGGAACACCCG</td>
</tr>
<tr>
<td>PBSII-CFouc</td>
<td>Firefly Luciferase (394-550)</td>
<td></td>
</tr>
<tr>
<td>NGluc-Zif268</td>
<td>Gaussis Luciferase (17-109)</td>
<td>GCGTAGCGTGGGCGGTTGGAACACCCG</td>
</tr>
<tr>
<td>PDSII-CGfouc</td>
<td>Gaussia Luciferase (110-165)</td>
<td></td>
</tr>
<tr>
<td>Pum2-NFluc</td>
<td>Firefly Luciferase (2-416)</td>
<td>CAUGGUGUAUAGUCUUUGUAUAGCGGC</td>
</tr>
<tr>
<td>CFOuc-Pum1</td>
<td>Firefly Luciferase (398-550)</td>
<td></td>
</tr>
<tr>
<td>MBD2-NFluc</td>
<td>Firefly Luciferase (2-416)</td>
<td>GCGTA_CGTACGCCACGCACCG</td>
</tr>
<tr>
<td>CFOuc-zif268</td>
<td>Firefly Luciferase (398-550)</td>
<td></td>
</tr>
<tr>
<td>PKI-NFluc</td>
<td>Firefly Luciferase (2-416)</td>
<td>GCGTAGCGTGGGCGGTTGGAACACCCG</td>
</tr>
<tr>
<td>CFOuc-PKA</td>
<td>Firefly Luciferase (398-550)</td>
<td></td>
</tr>
<tr>
<td>FRB-NFluc</td>
<td>Firefly Luciferase (2-416)</td>
<td>Rapamycin</td>
</tr>
<tr>
<td>CFOuc-FKBP</td>
<td>Firefly Luciferase (398-550)</td>
<td></td>
</tr>
<tr>
<td>p300-NFluc</td>
<td>Firefly Luciferase (2-416)</td>
<td></td>
</tr>
<tr>
<td>CFOuc-HIF-1alpha</td>
<td>Firefly Luciferase (398-550)</td>
<td></td>
</tr>
</tbody>
</table>

Table 4.1. Fusion constructs used in the development of the cell-free split-firefly luciferase assay. Agonists of each respective interaction are shown. Reprinted with the permission of the American Chemical Society.
4.3.3 Target DNA Preparation

All nucleic acid targets were obtained from IDT. A dsDNA target containing a zero base pair separation between the Zif268 and PBSII zinc finger sites (Z0P) was annealed as previously described. Hairpin DNA targets were annealed in 1x BamHI buffer by heating at 95°C for 7 minutes followed immediately by cooling on ice.

4.3.4 Reassembly of the GFP Variant Venus

Duplicate 150 μL translations were carried out in Wheat Germ Plus extracts (Promega) according to the manufacturer’s protocol using 4 pmol of each mRNA encoding for NVenus(residues 1-157)-Zif268 and PBSII-CVenus(residues 158-238), 10 μM ZnCl₂, 0.5 μL of RNasin Plus (Promega), and either 50 nM Z0P target DNA or no DNA. Translations were incubated at 25 °C for 2 hours (no fluorescence was observed) or alternatively interrogated for fluorescence followed by a 20 hour incubation at room temperature. Fluorescence spectra were acquired by exiting at 515 nm and monitoring emission at 528 nm.

4.3.5 Reassembly of Split β-lactamase-Zinc Finger Fusions

Four duplicate 25 μL translations were carried out in wheat germ plus extracts (Promega) according to the manufacturer’s protocol using 0.5 pmol of each mRNA encoding for NβLac(residues 26-196)-Zif268 and PBSII-CβLac(residues 198-290), 10 μM ZnCl₂, 0.5 μL of RNasin Plus (Promega), and either 20 nM Z0P target DNA or no DNA. Translations were incubated at 25 °C for 2 hours and assayed by adding 25 μL of translation to 75 μL of PBS containing a final concentration of 10 μM Fluorocilin Green.
soluble β-Lactamase substrate (Invitrogen). The final concentration of DNA in the assay was 5 nM. The rate of Fluorocillin Green hydrolysis was determined by exciting at 495 nm and monitoring emission at 525 nm with a 515 nm emission cutoff using a Spectra Max Gemini plate reader. Emission was read every 30 seconds for 10 minutes.

4.3.6 Reassembly of Split-Firefly Luciferase

Duplicate 25 μL translations were carried out in Flexi-Rabbit Reticulocyte Lysates (Promega) according to the manufacturer’s protocol using 2 pmol of each mRNA encoding for either PBSII-NFluc (residues 2-416) and CFluc-Zif268 (residues 398-550) or NFluc(residues 2-398)-Zif268 and PBSII-CFluc(residues 394-550), 10 μM ZnCl₂, 0.5 μL of Rnasin Plus (Promega), and either 25 nM Z0P target DNA or no DNA. Translations were incubated at 30 °C for 90 minutes and assayed by adding 20 μL of translation to 80 μL of Steady-Glo Luciferase Assay System (Promega). The final concentration of DNA in the assay was 5 nM. Light emission was monitored 1 minute after Steady-Glo addition using a Turner TD-20e luminometer with a 3 second delay and a 10 second integration time.

4.3.7 Initial Cell-Free Assays

Duplicate 25 μL translations were carried out in Flexi-Rabbit Reticulocyte Lysates (Promega) according to the manufacturer’s protocol using 2 pmol of each mRNA encoding the fusion proteins being analyzed, and 0.5 μL of Rnasin Plus (Promega). For translations containing zinc finger proteins 10 μM ZnCl₂ was also added to the translation mixture. Translations were incubated at 30 °C for 90 min and assayed by adding 20 μL
of translation mix to 80 μL of Steady-Glo Luciferase Assay System (Promega). In the case of nucleic acid-binding proteins target oligonucleotides were either present or absent during translation. For the rapamycin induced interaction between FRB and FKBP either 5 nM rapamycin or control, DMSO, was added after translation followed by a 30 minute incubation at room temperature. Light emission was monitored 1 minute after Steady-Glo addition using a Turner TD-20e luminometer with a 3 second delay and a 10 second integration time.

4.3.8 Reassembly of Split-Firefly Luciferase in a Coupled Transcription and Translation Rabbit Reticulocyte Lysate System

Coupled transcription/translation reactions were carried out in TNT T7 Coupled Rabbit Reticulocyte Lysates (Promega) according to the manufacturer’s protocol. Coupled reactions using split firefly luciferase-zinc finger fusions contained 0.5 pmols of each DNA encoding PBSII-NFluc and CFluc-Zif268, 10 μM ZnCl2, 1 μL of RNasin Plus (Promega), and either 100 nM Z0P target DNA or no DNA in a total of 25 μL. Coupled reactions using split firefly luciferase-FKBP and FRBP fusions contained 0.5 pmol of DNA encoding FRB-NFluc and CFluc-FKBP, and 1 μL of RNasin Plus (Promega) in a total of 25 μL. Solutions were incubated at 30 °C for 90 minutes. Reactions were diluted at a 1:4 ratio into PBS containing 1% BSA (1% BSA and either 25 nM rapamycin or DMSO in the case of FRB/FKBP) and equilibrated at room temperature for 30 minutes. Samples were assayed for luciferase activity by mixing 20 μL of lysate with 80 μL of Steady-Glo Luciferase Assay System (Promega). Luminescence readings were taken on a Turner TD20e luminometer using a 3 second delay and 10 second integrations, the
average of replicate experiments is shown. The final concentration of Z0P or rapamycin in the assay was 5 nM.

4.3.9 Reassembly of Split-Firefly Luciferase in a Purified Transcription and Translation System

Coupled transcription/translation reactions were carried out using the PURESYSTEM classic II system (Post Genome Inst. Co. Ltd.) according to the manufacturer’s protocol. Coupled reactions using split firefly luciferase-zinc finger fusions contained 0.5 pmols of each DNA encoding PBSII-NFluc and CFluc-Zif268, 10 μM ZnCl₂, 1 μL of RNasin Plus (Promega), and either 100 nM Z0P target DNA or no DNA in a total of 25 μL. Coupled reactions using split firefly luciferase-FKBP and FRBP fusions contained 0.5 pmol of DNA encoding FRB-NFluc and CFluc-FKBP, and 1 μL of RNasin Plus (Promega) in a total of 25 μL. Solutions were incubated at 37 °C for 60 minutes, followed by the addition of either 25 nM rapamycin or DMSO in the case of FKBP/FRB. Samples were assayed for luciferase activity by mixing 20 μL of lysate with 80 μL of Steady-Glo Luciferase Assay System (Promega). Luminescence readings were taken on a Turner TD20e luminometer using a 3 second delay and 10 second integrations, the average of replicate experiments is shown. The final concentration of Z0P or rapamycin in the assay was 5 nM.

4.3.10 Competition Assay to Identify Protein-DNA Target Site Specificity

Duplicate 25 μL translations were carried out in Rabbit Reticulocyte Lysates (Promega) according to the manufacturer’s protocol using 0.05 pmols of mRNA
encoding PBSII-NFluc and CFluc-Zif268, 10 μM ZnCl₂, and 0.5 μL of RNasin Plus (Promega) and allowed to incubate for 90 minutes at 30 °C in the presence of 750 pM Z0P dsDNA target. Following translation and firefly luciferase reassembly, increasing concentrations of each Zif268 hairpin DNA being tested (Table 4.2) were added followed by a 30 minute incubation at room temperature. Light emission was monitored 1 minute after Steady-Glo addition using a Wallac 1420 VICTOR 3™ V luminometer with a 1 second integration time.

4.3.11 Dissociation of the Reassembled Pum2-NFluc, CFluc-Pum1, RNA Ternary Complex

Duplicate 25 μL translation reactions were carried out in Rabbit Reticulocyte Lysates (Promega) according to the manufacturer’s protocol using 0.1 pmols of mRNA encoding Pum2-NFluc and CFluc-Pum1 and 0.5 μL of RNasin Plus (Promega) and allowed to incubate for 90 minutes at 30 °C in the presence of 2.5 nM RNA oligonucleotide target. Following translation and firefly luciferase reassembly, increasing concentrations of a competitor RNA oligonucleotide (Table 4.2) were added followed by a 30 minute incubation at room temperature. Light emission was monitored 1 minute after Steady-Glo addition using a Turner TD-20e luminometer with a 3 second delay and a 10 second integration time.
<table>
<thead>
<tr>
<th>Zif268 DNA Hairpin Antagonists</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Wild Type TGG:</strong></td>
</tr>
<tr>
<td>5’-GCGTAGCTGGGCGTACGCTCCCTGCACGCTACGC-3’</td>
</tr>
<tr>
<td><strong>TAG:</strong></td>
</tr>
<tr>
<td>5’-GCGTAGCTGCTGGGCGTACGCTCCCTGCACGCTACGC-3’</td>
</tr>
<tr>
<td><strong>TTG:</strong></td>
</tr>
<tr>
<td>5’-GCGTAGCTGCCGCGTACGCTCCCTGCACGCTACGC-3’</td>
</tr>
<tr>
<td><strong>TCG:</strong></td>
</tr>
<tr>
<td>5’-GCGTAGCTCGCCGCGTACGCTCCCTGCACGCTACGC-3’</td>
</tr>
<tr>
<td><strong>AAA:</strong></td>
</tr>
<tr>
<td>5’-GCGTAGCGAAGGCACGTACGCTCCCTGCACGCTACGC-3’</td>
</tr>
<tr>
<td><strong>RNA Antagonist</strong></td>
</tr>
<tr>
<td>5’-CCAGAAUUGAUUAUUCG-3’</td>
</tr>
</tbody>
</table>

Table 4.2. Sequences of nucleic acid antagonists used in this study. Zif268 DNA Hairpin Antagonists: Zif268 site is highlighted in blue, mutations are highlighted in orange.
5.1 Introduction

Protein-protein interactions are essential regulatory components of a number of biological processes including cell proliferation, growth, signal transduction, and programmed cell death.\textsuperscript{194} As a result, the development of interaction-specific reagents capable of disrupting these interactions is crucial to our understanding of complex biochemical pathways. Additionally, aberrant protein-protein interactions represent an increasingly growing class of human therapeutic targets for which the discovery of interaction inhibitors is a growing focus of pharmaceutical intervention.\textsuperscript{2,195} To date, a number of approaches have been developed which allow the specific targeting and inhibition of protein-protein interactions including the use of designed antibodies, peptides and peptide mimetics, and more recently small-molecules. Though effective inhibitors and useful reagents for the study of these interactions, antibody and peptide-based approaches lack key features that are desirable \emph{in vivo} drug candidates such as cell-permeability, oral bioavailability, and stability. On the other hand, the development of small-molecule inhibitors of protein-protein interactions is by no means an easy task and has been faced with a number of challenges. First, the surface contact area present amongst protein-protein interactions are generally large ($\sim$1,500-3,000 Å$^2$)\textsuperscript{196,197} compared to those of protein-small-molecule interactions (300-1,000 Å$^2$).\textsuperscript{198,199} Second, these surfaces tend to be relatively flat and featureless, lacking the well defined pockets conducive to small-molecule binding such as those found in enzyme active sites, making the design of selective inhibitors difficult. Third, contributors of interaction recognition
are often non-contiguous, limiting the effectiveness of mimicry-based approaches. Lastly, unlike enzyme inhibitors, for which small-molecule substrate starting points exist in nature, similar small-molecule starting points for the development of protein-protein interaction inhibitors are generally unavailable.

5.1.1 Methods for the development of small-molecule inhibitors of protein-protein interactions

Despite the challenges of targeting protein-protein interfaces, a number of recent examples do exist which demonstrate the successful development of small-molecule inhibitors of protein-protein interactions. To achieve this, a number of approaches have been employed to identify small-molecules capable of disrupting non-contiguous protein-protein interfaces. One approach, broadly defined as proteomimetics, involves the design of small molecules capable of mimicking the structures of extended regions of a protein-protein interface. For example, Hamilton and co-workers have developed a series of molecules utilizing a 3,2',2'"-substituted terphenyl scaffold, a molecular scaffold capable of mimicking the spatial orientation of the $i$, $i+4$, and $i+7$ residues of an $\alpha$-helix (Figure 5.1a). By decorating this scaffold with the appropriate functionalities responsible for binding, a small-molecule inhibitor of the interaction between calmodulin (CaM) and smooth muscle myosin light chain kinase (smMLCK) was developed, and was capable of inhibiting the CaM/smMLCK interaction with an IC$_{50}$ value of 9 nM (Figure 5.1b). Additionally, other appropriately decorated terphenyl scaffolds have been shown to inhibit the assembly of gp41, the interaction between p53 and hDM2, and the interaction between the BH3 domain of Bak and the anti-apoptotic protein Bcl-X$_L$. In a similar vein,
Figure 5.1. Small-molecule inhibitors of protein-protein interactions. a) Analogy between the $\alpha$-helix and the terphenyl scaffold used by Hamilton and co-workers. b) Terphenyl based small-molecule inhibitor of the interaction between CaM and smMLCK. c) Structure of the Bcl-2 family inhibitor HA14-1. d) Structure of the Bcl-2 family inhibitor ABT-737. e) Structure of the hDM2/p53 inhibitor Nutlin-3. f) Structure of the Bcl-2 family inhibitor GX15-070 (Obatoclax).
Gadek and co-workers utilized β-sheet mimicry to develop a high potency small-molecule (IC\textsubscript{50} = 1.4 nM) with the ability to inhibit the interaction between leukocyte functional agent-1 (LFA-1) and intercellular adhesion molecule-1 (ICAM-1).

As an alternative approach to surface mimicry, \textit{in silico} screening has also been utilized for the discovery of small-molecule inhibitors of protein-protein interactions. In one example, virtual screening of a diverse compound library against the modeled Bcl-2/Bak BH3 domain interaction was used to identify potential inhibitors\textsuperscript{200}. Of the 193,833 compounds screened, 53 were selected from the \textit{in silico} screening, of which 28 were assayed for inhibitory activity \textit{in vitro}. Of these, only one molecule, HA14-1 was found to possess desirable Bcl-2 binding activity (IC\textsubscript{50} = 9 μM) and was capable of inducing apoptosis in HL-60 cells (Figure 5.1c). In a similar approach, Enyedy \textit{et al.} utilized virtual screening of the National Cancer Institute three-dimensional database to identify seven compounds capable of inhibiting the interaction between the Bak BH3 domain and Bcl-2 with IC\textsubscript{50} values ranging from 1.6 to 14 μM\textsuperscript{201}. Of these, one compound was found to be the most potent \textit{in vivo}, inhibiting cell viability in the Bcl-2 over expressing cell line HL-60 with an IC\textsubscript{50} of 10.4 μM. In addition to targeting Bcl-2 interactions, computer-based compound screening has also been used to develop small-molecule inhibitors against other protein-protein interactions including HIV-1 Nef/SH3\textsuperscript{202}, Arf1-GDP/ARNO\textsuperscript{203} and CD4/MHC Class II\textsuperscript{204}.

An alternative strategy for designing interaction inhibitors is the use of ‘fragment assembly’ to construct high affinity small-molecules targeted against a particular protein-protein interface. This technique was recently used in conjunction with an NMR-based methodology dubbed ‘structure-activity relationship or SAR by NMR’\textsuperscript{205}. Using NMR,
Oltersdorf and co-workers screened a library of chemical compounds to identify compounds capable of binding the hydrophobic BH3 domain-binding groove of Bcl-X\textsubscript{L}.\textsuperscript{206} From this screen two low affinity compounds were identified that were capable of binding hydrophobic pockets on Bcl-X\textsubscript{L} which are occupied by conserved residues of the BH3-domain of Bak. By assembling these low affinity fragments together, a first generation high affinity binder was created that was capable of inhibiting the Bak/Bcl-X\textsubscript{L} interaction with a K\textsubscript{i} of 36 nM. Subsequent structure-based design methods to reduce binding to human serum albumin (HSA),\textsuperscript{207} followed by compound optimization, resulted in the compound ABT-737, which is capable of high affinity binding to Bcl-X\textsubscript{L}, Bcl-2, and Bcl-w (K\textsubscript{i} ≤1 nM) while binding with reduced affinity to the less homologous proteins Bcl-B, Mcl-1, and A1 (K\textsubscript{i} = 460 nM, >1 μM, and >1μM, respectively) (Figure 5.1d).

Though effective at discovering small-molecule inhibitors of protein-protein interactions, the above mention approaches do possess drawbacks. Structure-based design approaches such as mimicry rely on available structural information of the interaction being targeted, limiting the number of interactions for which inhibitors can be developed, as too does \textit{in silico} screening of compound libraries. The use of SAR by NMR and fragment assembly can result in highly selective and high affinity molecules, but can be very resource and labor intensive. An alternative, and perhaps the more widely used approach, is the identification of interaction inhibitors via the \textit{in vitro} screening of large and chemically diverse compound libraries. This method has been utilized to discover small-molecule inhibitors against a number of interactions, including p53/hDM2 and Bcl-2 family interactions. For example, Vassilev \textit{et. al.} used this
approach to screen a chemically diverse collection of compounds against the interaction between p53 and hDM2, discovering a class of *cis*-imidazoline compounds termed “nutlins”, the most potent of which, Nutlin-3 (Figure 5.1e), inhibits the p53/hDM2 interaction with an IC$_{50}$ of 90 nM. Similarly, this screening approach was used to evaluate a natural products library for the ability to inhibit Bcl-2 interactions, identifying a series of compounds called progidiosins. Lead optimization lead to the development of GX15-070 (Figure 5.1f), which is capable of binding multiple Bcl-2 family proteins, including the unique family member MCL-1.$^{208}$ This molecule, also known as Obatoclax, is currently undergoing phase I/II clinical trial for a variety of cancer types.$^{209}$ Thus, in summary, a combination of *in silico* guided compound synthesis, followed by rapid *in vitro* screening may provide a facile approach to target protein-protein interactions. Initial hits may be further refined using high-resolution structures when possible. The targeting of protein-protein interactions is still in its infancy; very few compounds have entered clinical trials.

5.1.2 Methods for the Interrogation of Protein-Protein Interactions and Inhibitors

Therefore

In order to efficiently and effectively screen large compound libraries, methods must be available which allow the interrogation of protein-protein interactions in a high throughput manner. A number of techniques are available which allow the interrogation of protein-protein interactions, including surface plasmon resonance (SPR), analytical ultracentrifugation (AUC), and isothermal titration calorimetry (ITC). Though extremely effective at providing binding information, these methods can be time consuming and not
easily adaptable for the simultaneous analysis of large numbers of potential interaction inhibitors. Instead, methods are needed which require small amounts of reagent proteins, provide a simple readout such as absorbance or fluorescence, and are able to be performed in a microtiter plate-based format. One example of such a technique was the use of a recent solid phase receptor-binding assay\textsuperscript{210} by Boger and co-workers for the identification of small-molecule inhibitors of the interaction between matrix metalloproteinase 2 (MMP2) and integrin $\alpha_\text{V}\beta_3$.\textsuperscript{211,212} Though effective, assays of this type are non-homogeneous, often utilizing purified components captured upon a solid support, and requiring blocking and washing steps followed by secondary recognition by an enzyme-antibody conjugate. As a result these assays can be time consuming and are limited to interactions between proteins amenable to purification.

Alternatively, more homogenous methods of interrogating protein-protein interactions can also be adapted to the high-throughput analysis of chemical libraries. For example, Fluorescence Resonance Energy Transfer (FRET) can be used to detect the interaction between two labeled proteins of interest. As FRET efficiency is dependent upon the inverse sixth power of the distance between two fluorophores, the direct binding, and therefore the disruption of binding between two proteins by an inhibitor can be monitored with the fusion of two proteins of interest to appropriate donor and acceptor fluorophores such as CFP and YFP (Figure 5.2a).\textsuperscript{213} Additionally, inhibitors of protein-protein interactions can also be identified through the use of fluorescence polarization (FP),\textsuperscript{214} a technique which is particularly suited to the interrogation of interactions between proteins and smaller protein or peptide binding partners. In this technique one member of the interaction pair is fluorescently labeled with the interaction being
Figure 5.2. Strategies for the identification and interrogation of inhibitors of protein-protein interactions. a) FRET-based strategies utilize the interaction-dependent energy transfer between two attached complementary fluorophores. In the presence of a positive protein-protein interaction excitation of a donor fluorophore results in emission from an acceptor fluorophore. In the presence of an interaction inhibitor donor and acceptor fluorophores are separated, decreasing the efficiency of energy transfer from the donor to the acceptor fluorophore. b) FP-based strategies utilize a chemically labeled protein conjugate, the association of which with a larger protein partner results in high levels of polarized fluorescence upon excitation with polarized light. Introduction of an interaction inhibitor releases the labeled component from the complex, resulting in a reduction of polarized emission upon excitation with polarized light.
determined by the analysis of polarized fluorescence emission upon excitation with polarized light. In the absence of an inhibitor the fluorophore containing partner is directly bound to the larger protein resulting in the emission of highly polarized fluorescence. Addition of an inhibitor disrupts this interaction, liberating the fluorophore containing partner, resulting in reduced fluorescence polarization (Figure 5.2b).

Given their homogeneous nature, FRET and FP-based strategies are well suited for the high throughput screening of protein-protein interaction inhibitors, but are limited to interaction pairs amenable to purification and chemical labeling, techniques which can often be both challenging and time consuming. Additionally, almost all of these assays are performed primarily under purified conditions which do not accurately represent the native cellular environment in which the interaction occurs, an aspect which may fail to account for off-target effects caused by endogenous cellular components. Conceivably, any homogenous interaction detection strategy could be used to develop platforms for the high throughput screening of interaction inhibitors. Akin to the FRET based strategy described above, split-protein reassembly could easily be adapted to the high throughput screening of interaction inhibitors through the use of reversible split-protein reporters such as firefly luciferase.73

In one example utilizing a similar approach, complementing deletion mutants of β-galactosidase162 were employed to screen potential inhibitors of EGFR dimerization. In this study the authors utilized stably transfected cells harboring EGFR-β-galactosidase fusions. Though this required time consuming cell culture techniques, a diverse library of 19,776 compounds were screened in a 384-well plate format resulting in the
identification of 31 potential inhibitors without the need for protein purification or chemical labeling.\textsuperscript{215}

We have recently described a cell-free strategy in which the luminescence from split-luciferase,\textsuperscript{14} fused to interacting proteins, provides a means for the direct measurement of heterodimeric protein-protein interactions.\textsuperscript{73} This methodology utilizes \textit{in vitro} translation methodologies to produce protein fusions \textit{in situ}, eliminating the need for laborious cell culture techniques or protein purification and chemical labeling methodologies. Given its rapid, reversible, and homogeneous nature, we can imagine the adaptation of this cell-free split-luciferase assay for the screening of protein-protein interactions and inhibitors thereof (Figure 5.3). To demonstrate this application we sought to develop a panel of interaction pairs representing clinically relevant protein-protein interactions followed by the analysis of their specificity and disruption in the presence of peptide and small-molecule antagonists.

5.2 Results and Discussion

5.2.1 Cell-Free Interrogation of Protein-Protein Interactions

Towards demonstrating the ability of our cell-free split-luciferase assay to interrogate protein-protein interaction specificity we constructed a panel of split-luciferase fusions representing a class of protein-protein interactions comprising a so-called “helix-cleft” or “helix-receptor” interaction. This class of protein-protein interfaces are primarily characterized by the binding of a short $\alpha$-helical domain (\textasciitilde{}15-25 amino acids in length) with a larger protein domain which presents a relatively hydrophobic groove or cleft where the hydrophobic pocket recognizes specific residues presented on the surface of
Figure 5.3. General schematic for the cell-free interrogation of protein-protein interaction inhibitors. *In vitro* transcribed mRNA corresponding to split-luciferase fusions of the interaction of interest are used as templates for *in vitro* translation resulting in interaction dependent luminescence. Addition of an interaction inhibitor postreassembly results in luciferase disassociation and abolishment of luciferase activity.
the helix (Figure 5.4a). These helix-receptor interactions have been particularly amenable to targeting by small-molecules, likely due to the relatively dense yet small surface interface\(^2,195\) (Figure 5.4b). To construct a panel of helix-cleft interactions (referred to here as helix-receptor interactions) we made a series of split-firefly luciferase fusions comprising members of two clinically relevant groups of protein-protein interactions, namely the interaction between the homologous proteins hDM2 and hDM4 and the activation domain of p53, and interactions amongst pro- and anti-apoptotic members of the Bcl-2 protein family.

The transcriptional activator p53 is best known as a tumor suppressor that can up regulate genes that mediate cell-cycle arrest in response to various cellular stresses, including DNA damage and oncogene activation, and is found to have abrogated function in approx 50% of all human cancers. Aside from somatic mutations or deletions, the down-regulation of p53 activity has also been correlated to the over-expression of the human protein double minute 2 (hDM2) and its homologue hDM4. This down-regulation of p53 activity is achieved by the binding of a 15 amino-acid \(\alpha\)-helix present in the N-terminal activation domain of p53 to the hydrophobic cleft of either hDM2 or hDM4. This binding event sequesters p53 and blocks tumor suppressive transcriptional activity (Figure 5.5a and b).

A large set of helix-receptor interactions are observed within the members of the B-cell lymphoma (Bcl-2) family of proteins, which are crucial regulators of programmed cell death. This family of proteins can be divided into two categories, namely anti-apoptotic members such as Bcl-2 and Bcl-X\(_L\), and pro-apoptotic members such as BAK (Bcl-2 antagonist/killer) and BAD (Bcl-2 antagonists of cell death). To affect their
Figure 5.4. Surface interface of helix-receptor interactions (Bcl-X<sub>L</sub>/BAK). a) Hydrophobic binding cleft of Bcl-X<sub>L</sub>. Surface contact area between receptor and the helix BAK is shown in red. Hydrophobic pockets which accommodate conserved BAK residues are highlighted. b) Interaction between Bcl-X<sub>L</sub> and the BH3 domain of BAK. Surface contact residues are highlighted in red (Bcl-X<sub>L</sub>).
Figure 5.5. Structures of hDM2 and hDM4 with p53. a) 15 residue portion of the p53 activation domain bound to the p53 binding domain of hDM2 (a) and hDM4 (b).
various pro- and anti-apoptotic functions, these proteins, in addition to other Bcl-2 family members, can form homodimers as well as heterodimers resulting in various combinations of protein-protein interactions (Figure 5.6). For example the anti-apoptotic members Bcl-2 and Bcl-x<sub>L</sub> inhibit apoptosis by binding α-helical portions of either BAK or BAD (16 and 26 residues respectively) called BH3 domains.

To test the specificity of these helix-receptor interactions a series of firefly luciferase fusions were created in which each helix was N-terminally fused to the N-terminal fragment of firefly luciferase (NFLuc) while several receptor domains were fused C-terminally to the C-terminal half of firefly luciferase (CFIuc). Our helix panel consisted of nine helices; the activation domain of p53, and the BH3 homology domains of the pro-apoptotic Bcl-2 family members BAD, BIM, BAK, BIK, BID, BMF, HRK, and PUMA. Additionally, our panel consisted of six receptor domains: the p53 binding domain of hDM2 and hDM4, and the BH3 binding domain of the anti-apoptotic proteins Bcl-2, Bcl-x<sub>L</sub>, Bcl-w, and BFL.

Utilizing our cell-free split-luciferase assay, each CFluc-receptor fusion was co-translated <em>in vitro</em> from mRNA with mRNA corresponding to each helix-NFluc fusion (Figure 5.7a). From our cell-free assay we were able to determine the binding ability of each receptor domain to each helix tested. The results of these assays are summarized in Figure 5.7b. hDM2 and hDM4 were the only receptors to bind the activation domain of p53, with all other receptors showing little or no binding to this helix. For the BAD BH3 domain, only three receptors showed any significant binding, those being Bcl-x<sub>L</sub>, Bcl-2, and Bcl-w. These results correspond well with literature, which report binding constants of &lt;10, 11, and 60 nM respectively<sup>216</sup> for these helix-receptor interactions. For the BIM
**Figure 5.6.** Bcl-2 family interactions amongst the anti-apoptotic members Bcl-X_L (green), BFL (cyan) and Bcl-w (grey) and the pro-apoptotic members BAD (blue), BAK (red), BIM (brown), BMF (purple), PUMA (orange) and BID (yellow).\textsuperscript{217-221}
Figure 5.7. Specificity of helix-receptor interactions. a) Schematic for the cell-free interrogation of helix-receptor type protein-protein interactions. *In vitro* transcribed mRNA corresponding to helix-NFluc and CFluc-receptor fusions of the interaction of interest are used as templates for *in vitro* translation resulting in interaction dependent luminescence. b) Specificity of helix-receptor interactions. c) $^{35}$S-Methionine labeled helix-NFluc fusions. Helix fusions are as follows: 1) p53, 2) BAD, 3) BIM, 4) BAK, 5) BIK, 6) BID, 7) BMF, 8) PUMA, 9) HRK. d) $^{35}$S-Methionine labeled CFluc-receptor fusions. Receptor fusions are as follows: 1) hDM2, 2) hDM4, 3) Bcl-2, 4) Bcl-XL, 5) Bcl-w, 6) BFL. Additional fusions include p300-NFluc (lane 7) and CFluc-HIF-1α (lane 8).
BH3 domain binding was seen for all anti-apoptotic Bcl-2 receptors, with no binding to either hDM2 or hDM4 being observed. For the case of the BAK BH3 domain binding was observed for Bcl-2, Bcl-X\textsubscript{L}, Bcl-w and BFL. In the case of BIK, binding was seen for all Bcl-2 family receptors. In this case these results correspond well with literature, except for BFL, which shows appreciable binding despite literature reports which show no binding of BIK by BFL\textsuperscript{216}. As expected from literature reports, the BID BH3 domain is bound by all Bcl-2 receptors while again, as expected, no binding is observed for either hDM2 or hDM4. Again, with BMF, our results are consistent with previous reports, which show strong binding for Bcl-2, Bcl-X\textsubscript{L}, and Bcl-w while showing little affinity for BFL. Binding by all Bcl-2 family receptors was seen for the PUMA BH3 domain, again, consistent with previous literature results. Lastly, binding of the HRK BH3 domain was only seen by Bcl-X\textsubscript{L}. Of all of the pro-apoptotic BH3 domains tested, HRK was seen to be the most selective, agreeing with previously reported results which also show exclusive binding of HRK by Bcl-X\textsubscript{L}. Expression of comparable levels of all helix and receptor luciferase fusions were further confirmed by \textsuperscript{35}S Methionine labeled polyacrylamide gel electrophoresis (Figure 5.7c and d).

5.2.2 A Cell-Free Split-Luciferase Assay for the Detection of Peptide and Small-Molecule Antagonists of Protein-Protein Interactions

As a first test of the ability of our split-firefly luciferase assay to evaluate antagonists of protein-protein interactions we sought to inhibit the well characterized interaction between the catalytic subunit of cAMP-dependent protein kinase (PKA) with the naturally occurring inhibitor PKI\textsuperscript{172}, an interaction which we had previously
monitored utilizing split-firefly luciferase. To test the inhibition of PKA/PKI heterodimerization increasing concentrations of a PKI peptide were added to *in vitro* translated PKI-NFluc and CFluc-PKA fusions after complex formation. This addition produced a distinct inhibitory/dominant negative effect, resulting in a concentration dependent decrease in luminescence and yielding an observed IC$_{50}$ value of 11 nM (Figure 5.8a). As a further example, we interrogated the inhibition of the interaction between hypoxia inducible factor-1α (HIF-1α) and the CH1 domain of the transcriptional coactivator p300. Recently, the small-molecule chetomin (Figure 5.8b) was identified from an *in vitro* screen of over 600,000 compounds as a first-in-class inhibitor of the interaction between HIF-1α and p300. To determine if our cell-free split-luciferase assay was amenable to the interrogation of small-molecule inhibitors of protein-protein interactions, increasing concentrations of chetomin were added to the p300/ HIF-1α complex post firefly luciferase reassembly. As with the addition of PKI above, a concentration dependent decrease in luminescence was observed, yielding and IC$_{50}$ value of 290 nM (Figure 5.8c). Importantly, control experiments involving excess PKI-peptide or chetomin showed no decrease in luminescence when added to a preassembled zinc finger/dsDNA complex (Figure 5.8d), verifying that the observed loss in signal was due to the disruption of specific protein-protein interactions rather than other off-target effects such as the inhibition of luciferase activity. Thus, these two examples demonstrate that both peptide and small-molecule inhibitors of protein-protein interactions can be rapidly evaluated within a cell-free context without the need for laborious transfection and cell-culture techniques or through the use of purified or chemically labeled assay components.
Figure 5.8. Interrogation of peptide and small-molecule inhibitors of protein-protein interactions utilizing the cell-free split-firefly luciferase assay. a) Concentration-dependent dissociation of the PKI-NFluc/CFluc-PKA complex by PKI peptide. b) Structure of the small-molecule p300/HIF-1α interaction inhibitor chetomin. c) Concentration-dependent dissociation of the p300/HIF-1α complex by chetomin. d) Interrogation of possible off-target effects of PKI peptide and chetomin on dsDNA-dependent luciferase reassembly (PBSII-NFluc, CFluc-Zif268, and 5nM target dsDNA oligonucleotide).
5.2.3 Cell-Free Split-Luciferase Assays for the Interrogation of Inhibitors of Helix-Receptor Interactions

Having shown that our cell-free split-luciferase assay has the ability to report upon the activity of peptide and small-molecule inhibitors of protein-protein interactions, we next sought to utilize our helix-receptor panel to evaluate inhibitors of these types of interactions. To test the ability of our split-luciferase assay to interrogate the activity of inhibitors of hDM2/p53 and hDM4/p53 complex formation, luminescence was evaluated in the presence and absence of the hDM2/p53 inhibitor nutlin-3 (Figure 5.9a). Addition of 5 μM of the enantiomer (-)-nutlin-3 post-reassembly resulted in the specific disruption of the hDM2/p53 complex while leaving the interaction between hDM4 and p53 unaffected (Figure 5.9b). This specificity is consistent with previous studies in which (-)-nutlin-3 was found to have a 20-fold greater ability to inhibit the interaction between hDM2 and p53 over hDM4.225 To further test the specificity of this sensor pair the (-)-nutlin-3 enantiomer (+)-nutlin-3 was shown to be unable to disrupt p53 binding by either hDM2 or hDM4 at a concentration of 5μM, a result consistent with previous studies showing a 150-fold decrease in binding ability of (+)-nutlin-3 when compared to (-)-nutlin-3226 (Figure 5.9c).

To test inhibitors of Bcl-2 family interactions, inhibition of the interaction of CFLuc-Bcl-2 and CFLuc-Bcl-XL with BIM-NFluc by the post-reassembly addition of BIM peptide (residues 142-161) was attempted. Surprisingly, unlike the interactions described above, the addition of concentrations as high as 100 μM BIM peptide resulted in minimal decreases in luminescence (Figure 5.10a). This is most likely due to formation of an
Figure 5.9. Inhibition of the interaction of the p53 activation domain with hDM2 and hDM4. a) Structure of the small-molecule hDM2/p53 interaction inhibitor Nutlin-3. b) (-)-Nutlin-3 inhibition of the interaction of p53 with hDM2 and hDM4. c) (+)-Nutlin-3 inhibition of the interaction of p53 with hDM2 and hDM4.
extremely high affinity complex between the BIM-NFluc and the CF\text{Luc}-Bcl-2 and CF\text{Luc}-Bcl-XL fusions, a phenomenon which has previously been overcome by working under a kinetic rather than a thermodynamic regime, where the inhibitor is preincubated with the receptor being tested followed by subsequent addition of the peptide interaction partner.\textsuperscript{227} To facilitate the disruption of Bcl-2 family complexes, luciferase fusions were translated separately from corresponding mRNA followed by the preincubation of receptors with either 10 μM BIM peptide or buffer (Figure 5.10b). In the absence of BIM peptide, observed luminescence steadily increased, with maximum luminescence being reached within 30 minutes following addition of BIM-NFluc. In the presence of BIM peptide luminescence increase is significantly perturbed for a time period of approximately 30 minutes after BIM-NFluc addition, followed by a significant increase in signal, suggesting the eventual formation of a tight complex between BIM-NFluc and its CF\text{Luc}-receptor binding partner, which is not reversible within the timeframe of the experiment (5.10c). From this, the maximal sensitivity was determined to occur at approximately 20 minutes after BIM-NFLuc addition. Utilizing this modified inhibition strategy relying on delaying protein complex formation, we again tested the ability of BIM peptide to disrupt the interaction of BIM-NFluc with CF\text{Luc}-Bcl-2 and CF\text{Luc}-Bcl-XL. The results demonstrated a greater than 90% decrease in luminescence upon preincubation of both Bcl-2 and Bcl-XL with 100 μM BIM peptide (Figure 5.10d). These results, combined with those above, demonstrate the applicability of our cell-free split-luciferase assay to report upon the activity of inhibitors of helix-receptor type protein-protein interactions.
Figure 5.10. Inhibition of the interaction of the BIM BH3 domain with Bcl-2 family receptors. a) Initial inhibition of Bcl-2 and Bcl-X\textsubscript{L} binding to the BH3 domain of BIM by a BIM peptide postreassembly. b) Preincubation strategy for the interrogation of interaction inhibitors. Separately translated CF\textsubscript{Luc}-receptor fusion are preincubated with either vehicle (Top) or inhibitor (Bottom) followed by the addition of Helix-N\textsubscript{Fluc}. c) Time dependent luminescence utilizing our modified preincubation strategy in the presence and absence of 10 \( \mu \text{M} \) BIM peptide. d) BIM peptide inhibition of the interaction of BIM BH3 domain with Bcl-2 and Bcl-X\textsubscript{L} utilizing our preincubation strategy.
5.2.4 Specificity Determination of Peptide Inhibitors of Protein-Protein Interactions

A variety of strategies have been employed to develop synthetic inhibitors of protein-protein interactions, including the creation of peptides and miniature protein motifs comprised of the specific residues and features which contribute most to the interaction being targeted\textsuperscript{228}. One attractive approach is the use of synthetic $\alpha$-helical motifs which mimic either a particular domain (such as p53 or a BH3 derived peptide), or present specific residues known to contribute to the interaction being targeted. Though theoretically effective, helical motifs excised from their native protein context often lack secondary structure, resulting in binding, cell permeability, and high susceptibility to proteolytic degradation.\textsuperscript{229,230} Many approaches have been employed to stabilize or mimic the $\alpha$-helix, including the recent development of hydrogen bond surrogate derived $\alpha$-helices. This strategy consists of replacing an N-terminal main-chain $i$ to $i+4$ hydrogen bond with a carbon-carbon bond through a ring closing metathesis reaction (Figure 5.11a). This hydrogen bond surrogate preorganizes one turn of an $\alpha$-helix or “$\alpha$-turn”, overcoming intrinsic nucleation propensities and aiding in stable helix formation.\textsuperscript{228}

To test the ability of the hDM2/hDM4 and Bcl-2 family interaction sensors described above to identify specific inhibitors of protein-protein interactions a panel of hydrogen bond surrogate (HBS) $\alpha$-helices targeted against either the hDM2/p53 or Bcl-X\textsubscript{L}/BH3 peptide interaction were tested using our cell-free split-luciferase assay via the preincubation strategy described above. To truly test the ability of our method to identify interaction inhibitors, assays were performed utilizing a panel of HBS helices whose specificity were unknown to us at the time of testing (Figure 5.11b). 10 $\mu$M of each compound was preincubated separately against each receptor described above (hDM2,
Figure 5.11. Interrogation of peptide inhibitors of protein-protein interactions. a) Hydrogen-bond-surrogate (HBS) α-helices feature a carbon-carbon bond in place of the $i$ to $i+4$ hydrogen bond. $R =$ amino acid side chain. b) Fluorescence polarization derived binding affinity data for tested HBS helices. c) Inhibition profile of HBS helices tested against a panel of seven protein-protein interactions (six helix-receptor and one control).
hDM4, Bcl-2, Bcl-xL, Bcl-w, and BFL) followed by the addition of either p53 (hDM2 and hMD4) or the BIM BH3 domain (Bcl-2, Bcl-xL, Bcl-w, and BFL), for 20 min and luminescence measured. Additionally, each compound was tested against a control interaction between p300 and HIF-1α (Figure 5.11c).

Of the six receptors tested, three were shown to be selectively inhibited by their target compounds, namely LKH-VII-18 for hDM2 and hDM4, and MRW-310 and MRW-311 for Bcl-XL (Figure 5.12a-c). These results are consistent with binding data obtained from traditional FP experiments consisting of purified and chemically labeled components. Interestingly, the three compounds found not to inhibit in the above cell-free assay were shown to bind with affinities relative to those that were identified as inhibitors. For example, LKH-VII-19 was shown to bind hDM2 with a modest affinity of $1.8 \mu M \pm 0.169$ while the binding of LKH-VII-19 to hDM2 is approximately 5-fold weaker than that of LKH-VII-18 ($K_d = 345 \pm 161 \text{nM}$). In the case of the MRW-309, which was expected to bind Bcl-XL (231 $\pm$ 38 nM), failed to inhibit the interaction between Bcl-XL and the BH3 domain of BIM. Additionally, MRW-308, which is reported to binds Bcl-XL with an affinity of 902 $\pm$ 70 nM also failed to inhibit the interaction between Bcl-XL and BIM in our assay. Of the other receptors tested (Bcl-2, Bcl-w, and BFL) along with the control p300/HIF-1α, no inhibitory effects were observed. This result is not surprising given that the compounds tested were designed to target either hDM2/hMD4 or Bcl-XL. Taken together, these results demonstrate the ability of our cell-free split-luciferase assay to rapidly identify specific inhibitors of protein-protein interactions without the need for purified or chemically labeled components and within a context which more closely resembles a native cellular
**Figure 5.12.** Interaction inhibition by select HBS-Helices.  a) Interaction inhibition by LKH-VII-18. Specifically, the interactions between hDM2/p53 and hDM4/p53 are inhibited by LKH-VII-18.  b) Interaction inhibition by MRW-310. Specifically, the interaction between Bcl-X\(_L\) is inhibited by MRW-310.  c) Interaction inhibition by MRW-311. Specifically, the interaction between Bcl-X\(_L\) is inhibited by MRW-311. * denotes interactions inhibited by either LKH-VII-18, MRW-310, or MRW-311.
environment. Future experiments will aim to reengineer the helix/receptor pairs, for example by utilizing a shorter helix, such that lower affinity interactions can be measured under thermodynamic control.

5.2.5 Cell-Free Interrogation of inhibitors of the hDM2/p53 and hDM4/p53 Interactions

Having shown that we can identify selective peptide inhibitors of the hDM2/p53 and hDM4/p53 interactions, we next sought to utilize our cell-free split-luciferase assay to compare the activities of three analogous p53 derived inhibitors, LKH-VII-113 and LKH-VII-114, and LKH-VII-115 (Figure 5.13a). Specifically, analysis of these three peptides allows the analysis of various modifications believed to either stabilize helix formation or abolish hDM2 and hDM4 binding. For example, LKH-VII-114 contains a helix stabilizing HBS modification while LKH-VII-113 corresponds to the unmodified sequence. Comparison of these two peptides will allow us to determine the effects if any, of this modification on peptide specificity and inhibition. Additionally, LKH-VII-115 contains this same modification, but lacks two key residues necessary for effective hDM2/hDM4 binding. Utilizing our panel of helix-receptor interactions the specificity of these peptides was determined. 10 μM of each compound was preincubated separately against each receptor followed by the addition of either p53-NFLuc or BIM-NFluc and subsequent luminescence measurement. Additionally, each compound was tested against a control interaction between p300 and HIF-1α (Figure 5.13b). Inhibition was only seen for the hDM2/p53 and hDM4/p53 interactions, while none of the Bcl-2 family interactions or the interaction between p300 and HIF-1α showed any inhibition. Of the
Figure 5.13. Inhibition of hDM2/p53 and hDM4/p53 interaction by various p53 derived peptides.  


b) Inhibition profile of LKH-VII-113, LKH-VII-114, and LKH-VII-115 against a panel of seven protein-protein interactions (six receptor-helix and one control). 

c) Concentration dependent inhibition of the hDM2/p53 interaction by peptides LKH-VII-113 and LKH-VII-114. 

compounds tested, both LKH-VII-113 and LKH-VII-114 showed similar abilities to inhibit the interaction of both hDM2 and hDM4 with p53, while LKH-VII-115 showed no inhibition. Additionally, both LKH-VII-113 and LKH-VII-114 show a concentration dependent decrease in luminescence when titrated against either hDM2 (IC$_{50}$ LKH-VII-113 = 1.5 ± 0.3 μM, IC$_{50}$ LKH-VII-114 = 1.8 ± 0.1 μM) or hDM4 (IC$_{50}$ LKH-VII-113 = 11.7 ± 5.4 μM, IC$_{50}$ LKH-VII-114 = 7.8 ± 2.7 μM), again showing similar abilities of each peptide to inhibit the interaction of p53 with hDM2 and hDM4 (Figure 5.13c and d).

5.2.6 Specificity Determination of Small-Molecule Inhibitors of Protein-Protein Interactions

To determine the ability of our cell-free luciferase assay to interrogate small-molecule inhibitors of protein-protein interactions we evaluated a panel of small-molecules for their ability to specifically inhibit receptor-helix type protein-protein interactions. To achieve this, an initial panel of three small-molecules designed to inhibit Bcl-2 family interactions was tested, namely gossypol, which is a natural product identified from a product screen, GX15-070, and HA14-1 (Figure 5.14a). Based on previously determined inhibition data, an initial screen demonstrated the ability of two compounds, Gossypol and HA14-1, to inhibit the interaction between Bcl-X$_L$ and the BH3 domain of BIM by greater than 80% and 40%, respectively, while GX15-070 showed no inhibition at the concentration tested. Due to poor solubility, higher concentrations of GX15-070 were not tested. To further determine the specificity of these molecules our cell-free split-luciferase assay was used to profile each compound against our six helix-receptor interaction panel, including a control p300/HIF-1α.
Figure 5.14. Interrogation of small-molecule inhibitors of Bcl-2 family interactions. a) Structure of the Bcl-2 family inhibitors Gossypol, HA14-1, and GX15-070. b) Inhibition profile of gossypol (80 μM) against a panel of seven protein-protein interactions (six receptor-helix and one control). c) Inhibition profile of HA14-1 (400 μM) against a panel of seven protein-protein interactions (six receptor-helix and one control). d) Inhibition profile of GX15-070 (25 μM) against a panel of seven protein-protein interactions (six receptor-helix and one control). e) Interrogation of possible off-target effects by gossypol. f) Interrogation of possible off-target effects of HA14-1.
interaction (Figure 5.14b-d). Of the compounds tested, gossypol and HA14-1 showed significant inhibition, but little specificity, with gossypol inhibiting all interactions by greater than 80%, including almost complete inhibition of the interaction between p300 and HIF-1α (Figure 5.14b). Additionally, HA14-1 showed little specificity, with similar amounts of inhibition (greater than 40%) for all interactions (Figure 5.14c). GX15-070 showed no inhibition at the tested concentrations (Figure 5.14d).

With respect to gossypol and HA14-1, this pan inhibition of all interactions tested was quite surprising. To determine if the observed decrease in luminescence was due to inhibition of luciferase activity, a series of non-helix-receptor like interactions were tested for signal generation in the presence of varying concentrations of gossypol and 400 µM HA14-1 (Figure 5.14e and f) in addition to a tethered luciferase containing a covalent N- and C-terminal linkage designed to mimic the post-reassembled split-firefly luciferase. In the presence of gossypol, a concentration dependent decrease in luminescence was observed for all control reactions as well as for the tethered luciferase, likely resulting from the inhibition of luciferase activity. Significantly, the presence of HA14-1 had no effect on tethered luciferase activity or on the interaction between the leucine zipper proteins Fos and Jun, or on the interaction between a target RNA and pumilio RNA-binding proteins. This suggests no inhibitory effects of HA14-1 on luciferase activity. On the other hand, a 50% decrease in luminescence signal was observed for the interaction between the leucine zipper pair RR and EE, suggesting that HA14-1 has the ability to inhibit this particular leucine zipper interaction.

In an additional attempt to demonstrate the utility of our cell-free assay for the interrogation of Bcl-2 family interaction inhibitors, we turned to the interrogation of the
Bcl-2 family small-molecule inhibitor BH3I-1.\textsuperscript{232} Initial experiments showed modest
decreases in luminescence, with 50 μM BH3I-1 inhibiting the interaction of the BIM
BH3 domain with both Bcl-X\textsubscript{L} and Bcl-2 by 25% and 40%, respectively. To further
examine the specificity of BH3I-1, our cell-free split-luciferase assay was used to profile
100 μM BH3I-1 against our six helix-receptor panel, the related interacting pairs
p300/HIF-1α and Fos/Jun, and finally tethered luciferase (Figure 5.15a). Of the Bcl-2
family interactions tested, the greatest inhibition (70%) was observed for the interaction
between Bcl-2 and the BIM BH3 domain, followed by Bcl-X\textsubscript{L} with 43% inhibition.
Similar amounts of inhibition were observed for the interaction of Bcl-w and BFL with
the BIM BH3 domain. With respect to Bcl-2, these relative levels of inhibition
 correspond well with two previously reported studies in which Bcl-2 is reported as being
the most susceptible to inhibition by BH3I-1.\textsuperscript{227,233} Unlike Bcl-2, previously reported
levels of inhibition between Bcl-X\textsubscript{L} and Bcl-w are inconsistent, in that one study
demonstrates an approximately 2-fold greater susceptibility towards inhibition by BH3I-1
for Bcl-w over Bcl-X\textsubscript{L} (IC\textsubscript{50} Bcl-w = 2.33 μM versus IC\textsubscript{50} Bcl-X\textsubscript{L} = 5.86 μM)\textsuperscript{227} while a
second study shows virtually comparable levels of inhibition.\textsuperscript{233} In the case of BFL,
relative levels of inhibition were found to be comparable with one previously reported
study.\textsuperscript{227} Surprisingly, significant levels of inhibition were observed for the interactions
between hDM2 and p53 in addition to the interaction between p300 and HIF-1α, while
modest levels of inhibition were observed for the interaction between hDM4 and p53.
Meanwhile, no inhibition of the interaction between the leucine zippers Fos and Jun was
observed, nor did BH3I-1 show any ability to inhibit tethered luciferase activity,
Figure 5.15. Interrogation of small-molecule inhibitor specificity. a) Left: Structure of the small-molecule Bcl-2 family inhibitor BH3I-1. Right: Inhibition profile of BH3I-1 against a panel of eight protein-protein interactions (six receptor-helix and two control) in addition to tethered luciferase activity. b) Left: Structure of the small-molecule hDM2 inhibitor Nutlin-3. Right: Inhibition profile of (-)-nutlin-3 against a panel of eight protein-protein interactions (six receptor-helix and two control) in addition to tethered luciferase activity. c) Left: Structure of the small-molecule p300/HIF-1α inhibitor chetomin. Right: Inhibition profile of chetomin against a panel of eight protein-protein interactions (six receptor-helix and two control) in addition to tethered luciferase activity.
eliminating the obvious alternative mechanisms for the decrease in luminescence observed for non-target interactions.

Lastly, to further establish the utility of our cell-free split-luciferase assay, specificity profiles utilizing our helix-receptor interaction panel and control interactions were performed to evaluate the specificity of the small-molecules (-)-nutlin-3 and chetomin at concentrations of 100 μM (Figure 5.15b and c). For (-)-nutlin-3, significant levels of inhibition were observed for the target interactions of p53 with hDM2 and hDM4. Surprisingly, modest levels of inhibition were also observed for the additional Bcl-2 family interactions as well as the interaction between p300 and HIF-1α. Though appreciable, these observations could likely be a result of off-target effects as evidenced by a modest decrease (8%) in tethered luciferase activity as well as an apparent 10% inhibition of the interaction between Fos and Jun. In light of this, minimal off-target inhibition of Bcl-2 family interactions at concentrations of 100 μM cannot be ruled out given the greater than 10% decrease in luminescence observed for the Bcl-2 interactions as well as p300/HIF-1α. As expected, 100 μM chetomin resulted in almost complete inhibition of the interaction between p300 and HIF-1α. Additionally, insignificant decreases in luminescence were observed for the panel and control interactions as well as for the tethered luciferase.

5.2.7 Conclusion

The identification and large-scale profiling of peptide and small-molecule inhibitors of protein-protein interactions is an important area of research for the study of these interactions as well as for the development of viable therapeutic strategies. To
achieve this, assays are needed which allow the rapid identification of interaction inhibitors in a high throughput manner. Herein we demonstrate a first step towards the application of our previously developed cell-free split-luciferase assay to the identification and interrogation of protein-protein interactions inhibitors. We have demonstrated the ability of this approach to report upon the interaction of a variety of protein-protein interactions, to identify specific peptide inhibitors of therapeutically relevant protein-protein interactions, and its utilization for the rapid interrogation of inhibitor specificity. The homogenous nature, generality, and rapid signal generation afforded by this assay present a number of positive attributes allowing the adaptation of this assay to high throughput applications, providing an additional and improved strategy for the screening of drug-like inhibitors of therapeutically relevant protein-protein interactions.

5.3 Materials and Methods

5.3.1 General Reagents

All materials were obtained from Sigma-Aldrich unless otherwise noted. ZnCl$_2$ was obtained from EM Sciences. Restriction enzymes were obtained from NEB. BIM BH3 peptides was obtained from AnaSpec. Chetomin, HA14-1, and BH3I-1 were obtained from CalBiochem. (+)-Nutlin-3, (-)-Nutlin-3, and GX15-070 were obtained from Cayman Chemical. Gossypol was obtained from Sigma-Aldrich. The stapled helices used in this study were kindly provided by P. S. Arora. Rabbit reticulocyte lysates were prepared as previously described.\textsuperscript{234}
5.3.2 Cloning of Fusion Proteins and In Vitro Transcription

The fusion protein constructs used in this study are shown in Table 5.1. DNA coding for NFluc(2-416) and CFluc(398-550) were generated by PCR with appropriate primers and subsequently cloned into the pETDuet-1 vector using standard techniques with verification by dideoxynucleotide sequencing. Fragments encoding helices were generated by over-lapping primer extension using standard techniques. Fragments encoding the receptors used in this study were generated by PCR from specific plasmids. Plasmids encoding hDM2, hDM4, Bcl-2, Bcl-x₅, Bcl-w, and BFL were obtained from Open Biosystems. PKA was provided by S. S. Taylor. p300 was provided by B. Z. Olenyuk. The fusion protein constructs were generated using standard cloning techniques and verified by dideoxynucleotide sequencing. The mRNA necessary for cell-free assays was generated as follows: PCR fragments corresponding to the desired fusion constructs were generated using a forward primer containing a T7 RNA polymerase promoter and a Kozak sequence and a reverse primer containing a 3’ stem loop. The purified PCR products were subsequently used as templates for in vitro transcription using the RiboMAX Large Scale RNA Production System-T7 (Promega) following the manufacturer’s protocols.
<table>
<thead>
<tr>
<th>Receptor</th>
<th>Helix</th>
</tr>
</thead>
<tbody>
<tr>
<td>CFLuc-hDM2 (1-183)</td>
<td>p53 (7-36)-NFluc</td>
</tr>
<tr>
<td>CFLuc-hDM4 (1-185)</td>
<td>BAD (103-127)-NFluc</td>
</tr>
<tr>
<td>CFLuc-Bcl-2 (1-239)</td>
<td>BIM (142-161)-NFluc</td>
</tr>
<tr>
<td>CFLuc-Bcl-x(1-233)</td>
<td>BAK (71-90)-NFluc</td>
</tr>
<tr>
<td>CFLuc-Bcl-w (1-193)</td>
<td>BIK (51-70)-NFluc</td>
</tr>
<tr>
<td>CFLuc-BFL (1-175)</td>
<td>BID (126-145)-NFluc</td>
</tr>
<tr>
<td>CFLuc-PKACα</td>
<td>BMF (127-146)-NFluc</td>
</tr>
<tr>
<td>p300 (323-423)-NFluc</td>
<td>HRK (27-46)-NFluc</td>
</tr>
<tr>
<td></td>
<td>PUMA (137-165)-NFluc</td>
</tr>
<tr>
<td></td>
<td>PKI (5-24)-NFluc</td>
</tr>
<tr>
<td></td>
<td>CFLuc-HIF-1α-CTAD</td>
</tr>
</tbody>
</table>

**Table 5.1.** Fusion constructs used in this study.
5.3.3 Cell-Free Interrogation of Protein-Protein Interactions

25 μL translations were carried out in treated rabbit reticulocyte lysate using 0.4 pmols of each CFLuc-receptor mRNA and 2 pmols of each Helix-NFluc mRNA being analyzed, and 0.5 μL of RNasin Plus (Promega). Translations were incubated at 30 °C for 90 minutes and assayed by combining 10 μL of translation solution with 40 μL of Steady-Glo Luciferase Assay System (Promega). Light emission was monitored 1 minute after Steady-Glo addition using a Turner Biosystems 20/20 Luminometer with a 3 second delay and a 10 second integration time. Relative levels of luminescence were compared to a background translation containing no mRNA.

5.3.4 Cell-free interrogation of peptide and small-molecule inhibitors of protein-protein interactions

Duplicate 25 μL translation reactions were carried out in Rabbit Reticulocyte Lysate (Promega) according to the manufacturer’s protocol using 2 pmols of mRNA encoding either FRB-NFluc and CFLuc-FKBP, PKI-NFluc and CFLuc-PKA, or p300-NFluc and CFLuc-HIF-1α and 0.5 μL of RNasin Plus (Promega) and allowed to incubate for 90 min at 30 °C. For analysis of the p300/HIF-1α interaction 10 μM ZnCl₂ was added to the translation reaction. Following translation, lysates were diluted 1:4 with PBS containing 1% BSA followed by the addition of increasing concentrations of either rapamycin in DMSO, PKI, or chetomin in DMSO followed by a 30 min incubation at room temperature. Samples were assayed for luciferase activity by combining 20 μL of translation solution with 80 μL of Steady-Glo Luciferase Assay System (Promega).
Light emission was monitored 1 min after Steady-Glo addition using a Wallac 1420 VICTOR 3 V luminometer with a 1 s integration time.

To determine the inhibition of the interactions between p53 and hDM2 or hDM4 25 μL translations were carried out in treated rabbit reticulocyte lysate using 0.4 pmols of either CF Luc-hDM2 or CF Luc-hDM4 mRNA and 2 pmols of p53-NFluc mRNA and 0.5 μL of RNasin Plus (Promega) followed by an incubated at 30 °C for 90 minutes. Following incubation either (+)-nutlin-3 or (-)-nutlin-3 was added post-reassembly to a final concentration of 5 μM and allowed to incubate for 30 minutes at room temperature. Inhibition was assayed by combining 10 μL of translation solution with 40 μL of Steady-Glo Luciferase Assay System (Promega). Light emission was monitored 1 minute after Steady-Glo addition using a Turner Biosystems 20/20 n Luminometer with a 3 second delay and a 10 second integration time.

5.3.5 Cell-Free Interrogation of Bcl-2 Family Inhibitors: Modified Protocol

Separate 25 μL translations were carried out in treated rabbit reticulocyte lysate using 0.4 pmols of either CF Luc-Bcl-2 or CF Luc-Bcl-xL mRNA and 2 pmols of BIM-NFluc mRNA and 0.5 μL of RNasin Plus (Promega) followed by an incubated at 30 °C for 90 minutes. Following incubation 100 μM or DMSO was allowed to incubate with CF Luc-receptor translations for 20 minutes at room temperature followed by the subsequent addition of an equal volume of BIM-NFluc translation and incubation for an additional 20 minutes. Samples were assayed for luciferase activity by combining 10 μL of translation solution with 40 μL of Steady-Glo Luciferase Assay System (Promega).
Light emission was monitored 1 minute after Steady-Glo addition using a Turner Biosystems 20/20 Luminometer with a 3 second delay and a 10 second integration time.

5.3.6 Cell-Free Interrogation of Inhibitor Specificity

Separate 25 μL translations were carried out in treated rabbit reticulocyte lysate using 0.4 pmols of each CFluc-receptor mRNA and 2 pmols of each Helix-NFluc mRNA being analyzed, and 0.5 μL of RNasin Plus (Promega) followed by an incubated at 30 °C for 90 minutes. Following incubation 10 μL of each CFluc-receptor translation being analyzed was incubated with either 10 μM of the stapled helix being analyzed or 100 μM of the small-molecule being analyzed followed by a 20 minute incubation at room temperature. After 20 minutes 10 μL of either BIM-NFluc (for Bel-2 family interactions) or p53-NFluc (for hDM2 and hDM4 interactions) translation was added followed by an additional 20 minute incubation at room temperature. For tethered luciferase control inhibition 25 μL translations were carried out in treated rabbit reticulocyte lysate using 0.4 pmols mRNA and 0.5 μL of RNasin Plus (Promega) followed by an incubated at 30 °C for 90 minutes. Following translation, inhibitors were incubated with tethered luciferase translation for 40 minutes at room temperature. Samples were assayed for luciferase activity by combining 10 μL of translation solution with 40 μL of Steady-Glo Luciferase Assay System (Promega). Light emission was monitored 1 minute after Steady-Glo addition using a Turner Biosystems 20/20 Luminometer with a 3 second delay and a 10 second integration time. Relative levels of inhibition were determined by comparison with assays containing no inhibitor.
To determine the concentration dependent inhibition of p53/hDM2 and p53/hDM4 interactions by LKH-VII-113 and LKH-VII-114 assays were performed using the modified protocol described above utilizing increasing concentrations of either LKH-VII-113 and LKH-VII-114. Samples were assayed for luciferase activity by combining 10 μL of translation solution with 40 μL of Steady-Glo Luciferase Assay System (Promega). Light emission was monitored 1 minute after Steady-Glo addition using a Turner Biosystems 20/20^n Luminometer with a 3 second delay and a 10 second integration time. Relative levels of inhibition were determined by comparison with assays containing no inhibitor.
CHAPTER 6
NEW DIRECTIONS IN THE CELL-FREE SPLIT-FIREFLY LUCIFERASE ASSAY

6.1 Introduction


Given the rapid nature of the cell-free split-firefly luciferase assay format described within the previous chapters, and the versatility of split-protein reassembly, it was anticipated that our cell-free split-firefly luciferase assay can have applications beyond the study of heterodimeric protein-protein interactions or ternary protein-nucleic acid complexes. Ongoing work within our laboratory has lead to the application of this cell-free split-firefly luciferase assay format to the development of sensors for the generalized detection of specific RNA sequences, the visualization of oligomeric protein species, the detection of protease activity, and the solution phase detection of a target protein. These efforts are briefly described below.

Generalized cell-free detection of ssRNA sequences: In addition to dsDNA, the ssRNA products of cellular transcription also provide a suitable means for the detection of disease specific genetic mutations. To develop a generalized method for the detection of ssRNA sequences two new split-firefly luciferase sensor platforms have been developed utilizing sequence-independent RNA binding Argonaute proteins and high-affinity sequence specific zinc finger proteins in conjunction with user-defined ssDNA probes (Figure 6.1a). These sensor platforms, developed by Jennifer Furman have been shown to enable the detection of specific sequences of ssRNA, providing two new general and sensitive methods for the visualization of clinically relevant genes.
**Cell-free visualization of oligomeric protein species:** Protein misfolding diseases such as Alzheimer’s, Parkinson’s, and Huntington’s disease are becoming increasingly prevalent. As a result, new methods are needed to study mechanistic features of disease progression as well to provide methods for drug development and evaluation. To affect this, split-firefly luciferase reassembly was utilized in conjunction with the recently described mini-protein TJ10, serving as a structural probe, to visualize intermediates of β-amyloid (Aβ40) fibril formation (Figure 6.1b), a phenomenon believed to be a key component to the development of Alzheimer’s disease. This work, performed by Kalyani Mondal resulted in a sensor capable of visualizing previously unobserved intermediate oligomeric states of Aβ40, Aβ42, prion, and AIPP. Thus this new approach provides a new method for the study of protein misfolding diseases.

**Cell-free interrogation of protease activity:** As a class of enzymes, proteases are essential to normal cellular function as well as disease progression and cell death. Given their importance and association with disease, methods are needed which allow the study of protease substrate specificity as well as the interrogation of inhibitors of disease related proteases, preferably within a cellular context. To achieve this, split-firefly luciferase reassembly has been utilized in conjunction with a pair of designed antiparallel coiled coils to create a genetically-encoded turn-on sensor for functional protease activity (Figure 6.1c). This sensor, developed by Sujan Shekhawat demonstrates firefly luciferase reassembly as a response to protease function and provides a new method for the interrogation of protease substrate specificity and activation within a cellular context.
Figure 6.1. Additional applications of the cell-free split-firefly luciferase assay. a) Generalized detection of ssRNA. **Left:** ssRNA detection utilizing firefly luciferase fusions to the PAZ domain of human argonaute 2 (Argo), is capable of binding to 2-nucleotide overhangs of dsRNA. The use of short ssRNA guides produces dsRNA with the necessary 2-nucleotide overhang, facilitating argonaute binding and firefly luciferase reassembly. **Right:** ssRNA detection utilizing high affinity ZF-firefly luciferase fusions in conjunction with a guide DNA hairpin containing a ssDNA guide sequence. Firefly luciferase reassembly is accomplished through the formation of a pentameric complex, with sequence specificity being provided by hybridization of the ssDNA portion of the guide DNA hairpin. b) Visualization of Aβ40 oligomeric intermediates utilizing firefly luciferase fusions to the mini-protein TJ10, a designed protein capable of inhibiting Aβ40 fibril formation and is believed to bind to large oligomeric intermediate species of Aβ40. The use of TJ10 fusions results in the specific visualization of large intermediates species, with the presence of either smaller intermediates of Aβ40 fibrils failing to facilitate firefly luciferase reassembly. c) Detection of protease function utilizing anti-parallel coiled coils. The coiled coil pair acid and base (designated A and B) are tethered together by a protease cleavable linker, which is then attached to the C-terminal half of firefly luciferase (CFluc), while one half of the coiled coil pair (base, B) is attached to the N-terminal half of firefly luciferase (NFluc). Under these configurations the intramolecular interactions between A and B attached to CFluc occludes complex formation and results in low levels of firefly luciferase reassembly. The addition of the protease TEV cleaves the linker between CFluc A and B, allowing B-NFluc to dimerize with CFluc-A and facilitate firefly luciferase reassembly. d) Schematic of solution phase detection of VEGF utilizing Flt-1-firefly luciferase fusions. Addition of free Flt-1 facilitates the displacement of Flt-1-firefly luciferase fusions, resulting in complex disruption and a decrease in luminescence.
Cell-free solution phase detection of target proteins: As previously illustrated, our approach for the detection of specific dsDNA and ssRNA sequences relies on a three-hybrid type approach. We have conceptually utilized this methodology to create a split-firefly luciferase sensor for the solution phase detection of the clinically relevant growth factor VEGF in its native state. Utilizing the VEGF binding domain of the extracellular receptor Flt-1 an assay was created which can not only report upon the presence of VEGF but also molecules capable of disrupting the interaction between VEGF and Flt-1, an interaction which is implicated in tumorigenesis and a validated drug target (Figure 6.1d).

6.1.2 Cell-Free Interrogation of Kinase Inhibitor Specificity

In addition to utilizing macromolecules such as nucleic acids or proteins, we have also devised a three-hybrid system utilizing a synthetic small-molecule to create an assay platform for the interrogation of small-molecule binding to the ATP site of kinases. With over 500 members, half of the total predicted number, protein kinases make up a very large and important class of eukaryotic enzymes. Given their role in mediating most signal transduction within cells, through the covalent modification of target substrate proteins, and their control over many cellular processes including cell proliferation, transcription, and apoptosis, there exists an increasing need for kinase specific reagents to elucidate the distinct regulatory roles played by individual protein kinases. Additionally, the mutation and misregulation of protein kinases has been implicated in a number of human diseases including cancer and inflammation, making kinases one of the most important target classes for drug development.
Commonly, protein kinases are specifically targeted, either for functional study or disease treatment, through the use of small-molecule inhibitors targeted against the ATP binding pocket, a feature which is highly conserved amongst a large majority of protein kinases.\textsuperscript{235,238,239} Though effective inhibitors, these small-molecules are often plagued by target promiscuity, resulting in cellular toxicity and potentially raising doubts about studies employing falsely specific inhibitors.\textsuperscript{240} To address this problem, recent efforts have resulted in a number of new strategies to create highly selective kinase inhibitors, including the targeting of unique allosteric sites and the creation of bivalent inhibitors which target both the active site in addition to a unique surface feature of the kinase in question.\textsuperscript{222,241-244}

To efficiently determine specificity, compounds must be tested against a large and preferably diverse panel of protein kinases to determine off-target inhibition, a property that cannot be readily predicted based on available sequence or structural information.\textsuperscript{239} Traditionally, this is achieved through the use of individual binding or activity assays utilizing recombinant protein sources and radioactive outputs, making the screening of a large number of kinases and molecules difficult and time consuming. For example, Cohen and co-workers have frequently utilized \textit{E. coli} or insect Sf21 cells to express full length human kinases as glutathione S-transferase (GST) fusions followed by a solution based assay to determine the amount of $^{33}$P transfer from [$\gamma^{33}$P]ATP to a kinase specific substrate peptide.\textsuperscript{238,240,245} Kinase inhibition is observed as a decrease in $^{33}$P transfer as determined by radio-imaging of spotted filter papers. Though effective, this technique relies upon the use of purified kinases, which could potentially be misfolded and inactive. Additionally, this assay relies on the determination of kinases activity. Though this
information is preferable over simple binding information, there are few methods for determining kinase activity \textit{in vitro}, most of which rely on the use of radioactive substrates.

Alternatively, Fabian and co-workers have described a magnetic bead capture approach utilizing phage-expressed human kinase domains and immobilized ATP active site binding ligands.\textsuperscript{239} In this approach, kinase domains are expressed as fusions to the T7 bacteriophage allowing for target kinase amplification and identification utilizing either traditional phage plaque assays or by quantitative PCR (qPCR). Phage-tagged kinases were then bound to probe ligands, such as the promiscuous kinase inhibitor staurosporine,\textsuperscript{246} and immobilized on a solid support, resulting in a ‘bait’ and ‘prey’ type assay format (Figure 6.2). To determine small-molecule binding, tagged kinases and immobilized ‘bait’ ligands were combined with non-immobilized ‘prey’ molecules. If the non-immobilized ‘prey’ compound interferes with ATP site binding by the immobilized ‘bait’ ligand, phage-tagged kinases are released from the solid support. If the ‘prey’ molecule does not interfere with the ATP binding site immobilized kinases are retained and not detected in the output. Unlike the technique described above, this method only reports on small-molecule binding, with the assumption that ATP site occlusion potentially results in kinase inhibition. Additionally, this assay is non-homogenous, requiring the use of ‘bait’ molecules immobilized on a solid support. Lastly, though they are tried and true methods of detection, the use of phage as an output is not very straightforward, with either time consuming phage-display in \textit{E. coli} hosts and qPCR based techniques being required for the determination of small-molecule binding.
Figure 6.2. Phage-based kinase ATP site competition strategy developed by Fabian et al. Top: In the absence of an ATP site competitor the phage-tagged kinase is bound to the immobilized ‘bait’ molecule. Middle: The presence of an ATP site occluding ‘prey’ molecule competes with immobilized ‘bait’ molecule kinase binding resulting in phage detection by *E. coli* plaque analysis or qPCR. Bottom: The presence of a non-ATP site occluding ‘prey’ molecule fails to compete for kinase binding by the immobilized ‘bait’ molecule resulting in no phage detection. T7 phage is shown in tan, the attached kinase is shown in red, and ‘bait’ and ‘prey’ molecules are shown in light and dark blue respectively.
To address the need for simple and high throughput assay technologies for the
determination of kinase inhibitor specificity we devised a ‘bait’ and ‘prey’ methodology
which combines favorable aspects of the assays described above, namely the use of a
homogenous solution phase platform which is capable of determining the relative binding
of small-molecules to kinases. This approach has the potential to be far more facile than
traditional radioactive activity assays as well as assay configurations utilizing phage-
display/magnetic beads.

Utilizing our recently describe cell-free split-firefly luciferase we envisioned a
strategy which allows us to determine kinase-binding by a small-molecule ATP site
competitor. To achieve this, we devised a chemically induced dimerization (CID)
strategy in which reassembly of split-firefly luciferase is facilitated by the formation of a
ternary complex between a kinase-firefly luciferase fusion, a second firefly luciferase-
fusion protein, and a bifunctional small-molecule containing an ATP site binding
molecule (Figure 6.3a). To interrogate small-molecule binding, the addition of an ATP
site competitor would displace the kinase bound bifunctional molecule and prevent split-
firefly luciferase reassembly (Figure 6.3b). In this manner we can take advantage of the
rapid rates of protein synthesis, the elimination of protein purification, and the simple
luminescent output provided by our cell-free split-firefly luciferase assay format.
Figure 6.3. A general schematic for the CID split-firefly luciferase assay for interrogating ATP site binding small molecules. a) Split-firefly luciferase fusions to protein A (a non-kinase protein, blue) and a kinase of interest (tan). Addition of a synthetic bifunctional molecule containing a non-kinase binding moiety (green) and an ATP site binding moiety (red) results in ternary complex formation and split-firefly luciferase reassembly. b) Interrogation of ATP site binding small molecules. Addition of an ATP site occluding molecule (yellow) competes for kinase binding by the bifunctional molecule resulting in no split-firefly luciferase reassembly.
6.2 Results and Discussion

6.2.1 Design of a Split-Firefly Luciferase Chemically Induced Dimerization Assay

The chemically induced dimerization (CID) of two proteins was elegantly implemented by Schreiber and coworkers utilizing a dimeric version of the FKBP binding molecule FK506 to activate the TCR signaling pathway in human T-cells.\textsuperscript{247} In principle, the CID strategy utilizes fusion proteins consisting of two small-molecule binding domains attached to two proteins of interest which normally do not interact with each other in addition to a synthetic small-molecule capable of simultaneously binding both fusion proteins. Fusion protein binding to the synthetic small-molecule brings the two fusion proteins into close proximity resulting in signal transduction,\textsuperscript{247,248} transcriptional activation,\textsuperscript{249,250} or split-protein reassembly\textsuperscript{64} depending upon the fusion proteins used.

To design a proof of concept CID-based split-firefly luciferase assay for the detection of kinase-binding molecules we turned to two previously describe small-molecule-protein interaction pairs, namely \textit{E. coli} dehydrofolate reductase (DHFR) and the small-molecule inhibitor trimethoprim (TMP) and the $\alpha$-catalytic subunit of PKA (PKAC$\alpha$ or PKA) with the promiscuous kinase inhibitor staurosporine.\textsuperscript{246,251} To create our sensor we fused DHFR to the N-terminal fragment of firefly luciferase creating the fusion DHFR-NFluc(2-416) and fused the C-terminal fragment of firefly luciferase to PKA creating CFluc(398-550)-PKA. To create the bifunctional synthetic dimerizer the DHFR inhibitor TMP was conjugated through a tetra-ethylene glycol linker to the kinase inhibitor staurosporine to create compound 6 (TMP-Staur) (Figure 6.4a). A key aspect of this design is the use of the promiscuous pan-kinase inhibitor staurosporine, which has
been shown to bind numerous kinases with relatively high affinity ($K_d < 100 \text{ nM}$), allowing the use of TMP-Staur to potentially study small-molecule binding by a large and diverse panel of firefly luciferase-kinase fusions.

6.2.2 Evaluation of Small-Molecule-Dependent Firefly Luciferase Reassembly

We reasoned that the simultaneous binding of DHFR-NF\text{Luc} and CFluc-PKA to TMP-Staur would result in the small-molecule-dependent reassembly of functional firefly luciferase (Figure 6.4b). Utilizing 125 nM TMP-Staur, an initial test of this approach resulted in an approximately 9 fold-increase in luminescence upon TMP-Staur addition to \textit{in vitro} translated DHFR-NF\text{Luc} and CFluc-PKA, demonstrating the first small-molecule CID facilitated split-firefly luciferase reassembly within our cell-free system (Figure 6.4c). In order to evaluate the ability of our cell-free firefly luciferase CID assay to interrogate small-molecule ATP site binders, we envision that the addition of a second ATP site binder would compete for PKA binding with TMP-Staur and result in a concentration-dependent decrease in firefly luciferase reassembly. To achieve this we sought to interrogate the dominant negative inhibition of firefly luciferase reassembly by non-conjugated staurosporine. Utilizing 125 nM TMP-Staur added to DHFR-NF\text{Luc} and CFluc-PKA post-translation, the addition of staurosporine produced a distinct dominant negative effect, resulting in a staurosporine concentration-dependent decrease in luminescence and yielding an observed IC$_{50}$ value of 181 nM (Figure 6.4d), demonstrating that our assay is capable of reporting small-molecules which are capable of competing for ATP site binding with our TMP-Staur dimerizer.
Figure 6.4. A CID sensor for the interrogation of kinase ATP site binding small-molecules. a) Structure of TMP-Staur (compound 6), the CID used in this study. b) Representation of TNP-Staur induced firefly luciferase reassembly. c) Initial test of TMP-Staur-dependent firefly luciferase reassembly utilizing DHFR-NFluc and CFlic-PKA. A 9-fold increase in luminescence is observed in the presence of 125 nM TMP-Staur. d) Concentration-dependent decrease in luminescence as a result of staurosporine competition with TMP-staur for CFlic-PKA ATP site binding. e) TMP-Staur-dependent firefly luciferase reassembly utilizing the fusion proteins DHFR-NFluc and either CFlic-CDK2, CFlic-MSK2, or CFlic-PDGFRB.
Having successfully demonstrated our assay design, we next sought to evaluate the ability of TMP-Staur to facilitate firefly luciferase reassembly utilizing a small panel of kinases tethered to CFluc, namely the human kinases CDK2, MSK2, and the tyrosine kinase domain of PDGFRB (Figure 6.4e). Though previous studies report strong staurosporine binding by all three of these kinases, only CFluc-MSK2 showed significant TMP-Staur-dependent luminescence (11-fold), while PDGFRB showed minimal binding with only a 1.5-fold enhancement over no TMP-Staur, and CDK2 showed no TMP-Staur-dependent luminescence.

6.2.3 Evaluation of Alternative Orientations of Split-Firefly Luciferase Fusions

Previous studies have shown significant enhancements in split-firefly luciferase reassembly due to alternative attachment geometries of firefly luciferase fusions. To determine if the opposite orientation of attached firefly luciferase fragments benefits reassembly we constructed an alternative pair of fusions consisting of CFluc(398-550)-DHFR and a small panel of kinase-NFluc(2-416) fusions. An initial comparison using PKA resulted in a greater than 2-fold improvement in signal-to-noise (16-fold versus 36-fold) for our newly constructed CFluc-DHFR and PKA-NFluc fusion pair (Figure 6.5a). Given this enhancement in sensitivity, the kinases CDK2 and PDGFRB, in the form of CDK2-NFluc and PDGFRB-NFluc, were tested with CFluc-DHFR for TMP-Staur-dependent firefly luciferase reassembly (Figure 6.5b). Unfortunately, as with our previous fusions above, PDGFRB again produced minimal TMP-Staur-dependent signal, with only a 1.33-fold enhancement being observed in the presence of 125 nM TMP-Staurosporine, while the CDK2 was still inactive in this format.
Figure 6.5. Evaluation of alternative orientation DHFR and kinase firefly luciferase fusions. a) Initial evaluation of alternative orientations using DNFR- and PKA-firefly luciferase fusions. The use of the alternative fusions PKA-NFluc and CFluc-DHFR provides a 2-fold increase in sensitivity over the original fusions DHFR-NFluc and CFluc-PKA. b) TMP-Staur-dependent firefly luciferase reassembly utilizing the fusion proteins CFluc-DHFR and either PDGFR-NFluc or CDK2-NFluc.
6.2.4 Conclusion

Split-protein reassembly sensors are not limited to the interrogation of protein-protein or protein-nucleic acid interactions. Utilizing a variety of sensor architectures, I and other members of our laboratory have significantly expanded the utility of the cell-free split-firefly luciferase assay to the detection of virtually any ssRNA sequence, the visualization of intermediates and products in amyloid aggregation pathways, the evaluation of enzyme function, and the solution phase detection of an unlabeled target protein. Moreover, we have described a cell-free split-firefly luciferase assay based on chemically induced dimerization for the interrogation of small-molecule binders to the ATP site of kinases. We believe this assay provides a number of advantages over previously described methods for interrogating small-molecule inhibitors of protein kinases including rapid and simple protein synthesis, the elimination of protein purification, a homogeneous solution phase format, and a simple and sensitive luminescence readout. The straightforward cell-free production of fusion proteins allows the rapid production of a large number of kinase fusions, with the only experimental bottle-neck being the creation of kinase fusion genes. We have shown that the synthetic small-molecule TMP-Staur is capable of facilitating firefly luciferase reassembly utilizing at least two different human kinases PKA and MSK2, and also allows the interrogation of ATP site binding molecules through a simple competition assay. Unfortunately all kinase fusions tested did not display a small-molecule-dependent signal. One explanation for this could be non-activation of target kinases, which for some kinases may be essential for ATP or staurosporine binding, a problem which could be overcome by the co-translation of certain kinase fusions with known activating
kinases. Additionally, the use of a synthetic small-molecule dimerizer offers distinct advantages as the potentially cell-permeable small-molecule could facilitate the interrogation of kinase inhibitors *in vivo*, allowing one to interrogate not only kinase specificity but also cell-permeability of test compounds.

6.3 Material and Methods

6.3.1 Cloning of DHFR and Kinase Split-Firefly Luciferase Fusions and In Vitro Transcription

All materials were obtained from Sigma-Aldrich unless otherwise noted. Trimethoprim was purchased from Research Products International (RPI). Restriction enzymes were obtained from NEB and in vitro translational products from Promega. Oligonucleotide primers and targets were from IDT.

The following primers were used to clone DHFR from the *E. coli* stain XL-1 Blue into an existing cassette containing a 15-amino acid linker followed by the N-terminal fragment of firefly luciferase, residues 2-416: Fwd, 5’-GCTATCGGATCCGATCAGTC TGATTGCGGCCTAGCG-3’; and Rev, 5’- GCTAGCACCAGGTCCGCGCTCCAGA ATCTCAAAAG-3’.

Fragments encoding the kinases MSK2, CDK2, PDGFRB, and PKA used in this study were generated by PCR from specific plasmids. Plasmids encoding the kinases MSK2 and CDK2 were obtained from AddGene, a plasmid encoding PDGFRB was obtained from R. Vallaincourt, and a plasmid encoding PKA was obtained from S. S. Taylor. The mRNA necessary for cell-free assays was generated as follows: PCR fragments corresponding to the desired fusion constructs were generated using a forward
primer containing a T7 RNA polymerase promoter and a Kozak sequence and a reverse primer containing a 3’ stem loop. The purified PCR products were subsequently used as templates for *in vitro* transcription using the RibomAX Large Scale RNA Production System-T7 (Promega) following the manufacturer’s protocols.

6.3.2 Synthesis of TMP-Staur

The small-molecule CID TMP-Staur was synthesized using the follow synthetic scheme\textsuperscript{252} (Scheme 6.1):

**Synthesis of compound 2:** Trimethoprim (1; 1 g, $M_w$: 290) was added to a round bottom flask containing HBr (12.5 mL, 48%) refluxed at 95 °C. The solution was stirred under air for 6 minutes with the internal temperature maintained at 95 °C. The solution was partially quenched with 5 mL of 10 M NaOH added prowise while stirring under heat. Agitation was stopped and the solution was allowed to cool to room temperature and then placed at 4 °C overnight; this resulted in the formation of a white precipitate. The precipitate was filtered from solution and rinsed with a minimal amount of ice cold H$_2$O. The precipitate was then added to a minimal amount of boiling H$_2$O until it was dissolved followed by neutralization with 50% NH$_4$OH. The solution was allowed to sit at room temperature and the resulting crystals filtered to yield 2 (400 mg, 1.44 mmol, 42%). $^1$H NMR (500 MHz, CD$_3$OD, 25 °C): $\delta = 7.507$ (s, 1H, ArH), 6.52 (s, 1H, ArH), 3.79 (s, 6H, ArOCH$_3$), 3.73 (ArCH$_2$Ar).
Scheme 6.1. Synthesis of TMP-Staur.  a) 48% HBr in H₂O;  b) methyl bromoacetate, DBU;  c) NaOH, MeOH;  d)PyBOP, mono-BOC diaminotertaethylene glycol; e) TFA, DCM; f) carboxylated staurosporine, PyBOP, DIEA.
Synthesis of compound 3: Compound 2 (400 mg, $M_w$: 276.29, 1.448 mmol) was dissolved in 4.344 mL of dry-solv DMSO in a dry 25 mL flask and 1.1 molar equivalents of DBU was added. After compound 2 was completely dissolved, the solution turned a deep red and 1.1 molar equivalents of methyl bromoacetate was added. The solution was incubated overnight and room temperature. Water (15 mL) was added and the product was extracted with EtOAc (4 x 15 mL). The combined EtOAc solution was wased with water (30 mL), dried over MgSO₄, and evaporated. The residue was dissolved in a small amount of methanol (20% v/v) in dichloromethane, and precipitated by the addition of ether/pentane (25 mL, 4:1, v/v), filtered, washed with ether, and dried to yield compound 3 (135 mg, 0.45 mmol, 26.9%). $^1$H NMR (500 MHz, CD$_3$OD, 25 °C): $\delta$ = 7.49 (s, 1H, ArH), 6.52 (s, 1H, ArH), 4.50 (s, 2H, ArCH₂), 3.75 (s, 6H, ArOCH₃), 3.72 (s, 3H, COOCH₃), 3.63 (ArCH₂Ar).

Synthesis of compound 4: Compound 3 (164 mg, $M_w$: 348.35, 0.47 mmol) was dissolved in methanol (12 mL). The solution was then diluted with 3 mL of NaOh in water (50 % w/w) until a small amount of beige precipitate was formed. The solution was then stirred under air for 3 hours until the precipitate disappeared. The reaction mixture was then titrated to a pH of 4 with HCl (1 N) and then methanol evaporated off leaving a beige precipitate which was filtered and washed with brine and water to yield 4 (65 mg, 0.195 mmol, 41.4%). $^1$H NMR (500 MHz, CD$_3$OD, 25 °C): $\delta$ = 7.28 (s, 1H, ArH), 6.46 (s, 1H, ArH), 4.27 (s, 2H, ArOCH₂), 3.72 (s, 6H, ArOCH₃), 3.63 (ArCH₂Ar).
Synthesis of compound 5: Compound 4 (50 mg, $M_w$: 334.33, 0.15 mmol) and PyBOP (155 mg, 0.3 mmol, 2 equiv) were added to a dried, flamed flask and placed under vacuum. After the flask had cooled DMF (10 mL) and mono-BOC protected diamino-tetraethylene glycol (TEG-BOC; 260 mg, 0.75 mmol, 5 equiv) were added to the flask under Ar. The mixture was stirred overnight. The solvent was removed by rotary evaporation and the crude mixture separated by HPLC. HPLC purification (0% - 100% acetonitrile gradient in water with 0.1% TFA). The BOC group was deprotected by mixing with 50% TFA in DCM for 6 hours followed by evaporation of the DCM solvent to yield 5 which was characterized by mass spectrometry (5: expected 536.7 g/mol, found 536.4 m/z).

Synthesis of compound 6 (TMP-Staur): Compound 5 (8.6 mg, $M_w$: 536.7, 16 μmol), PyBOP (16.6 mg, 32 μmol, 2 equiv), DIEA (4 equiv), and carboxylated staurosporine$^{353}$ (10 mg, $M_w$: 566.7, 17.6 μmol) were mixed together in 744 μL of DMF and allowed to react overnight protected from light. The reaction mixture was added to 18 mL of water (0.1% TFA) and crude product was isolated by centrifugation. The crude precipitate was dissolved in acetonitrile and separated by HPLC to yield 6 which was characterized by high resolution mass spectrometry (6: predicted 1084.5091 m/z, found 1084.5099 m/z).

6.3.3 Evaluation of TMP-Staur-Dependent Firefly Luciferase Reassembly

To initially evaluate TMP-Staur-dependent firefly luciferase reassembly duplicate 25 μL translation reactions were carried out in Rabbit Reticulocyte Lysate (Promega) according to the manufacturer’s protocol using mRNA encoding DHFR-NFluc (0.5
pmols) and CFluc-PKA (0.2 pmols) and 0.5 μL of RNasin Plus (Promega) and allowed to incubate for 90 min at 30 °C. Following translation 1 μL of either TMP-Staur (125 nM final) or DMSO was added to 24 μL of translation mixture and allowed to incubate for 30 minutes at room temperature. Samples were assayed for luciferase activity by combining 20 μL of translation solution with 80 μL of Steady-Glo Luciferase Assay System (Promega). Light emission was monitored 1 minute after Steady-Glo addition using a Turner Biosystems 20/20 n Luminometer with a 3 second delay and a 10 second integration time.

To determine TMP-Staur-dependent firefly luciferase reassembly for CDK2, MSK2, and PDGFRB duplicate 25 μL translation reactions were carried out in Rabbit Reticulocyte Lysate (Promega) according to the manufacturer’s protocol using mRNA encoding DHFR-NFluc (0.5 pmols) and CFluc-kinase (0.2 pmols) and 0.5 μL of RNasin Plus (Promega) and allowed to incubate for 90 min at 30 °C. Following translation 1 μL of either TMP-Staur (125 nM final) or DMSO was added to 24 μL of translation mixture and allowed to incubate for 30 minutes at room temperature. Samples were assayed for luciferase activity by combining 20 μL of translation solution with 80 μL of Steady-Glo Luciferase Assay System (Promega). Light emission was monitored 1 minute after Steady-Glo addition using a Turner Biosystems 20/20 n Luminometer with a 3 second delay and a 10 second integration time.

To determine TMP-Staur-dependent firefly luciferase reassembly utilizing alternative orientation firefly luciferase fusions of PKA, CDK2, and PDGFRB duplicate 25 μL translation reactions were carried out in Rabbit Reticulocyte Lysate (Promega) according to the manufacturer’s protocol using mRNA encoding Kinase-NFluc (0.2
pmols) and CFluc-DHFR (0.5 pmols) and 0.5 μL of RNasin Plus (Promega) and allowed to incubate for 90 min at 30 °C. Following translation 1 μL of either TMP-Staur (125 nM final) or DMSO was added to 24 μL of translation mixture and allowed to incubate for 30 minutes at room temperature. Samples were assayed for luciferase activity by combining 20 μL of translation solution with 80 μL of Steady-Glo Luciferase Assay System (Promega). Light emission was monitored 1 minute after Steady-Glo addition using a Turner Biosystems 20/20 Luminometer with a 3 second delay and a 10 second integration time.

6.3.4 Staurosporine Competition Assay

To determine kinase ATP site competition by staurosporine duplicate 25 μL translation reactions were carried out in Rabbit Reticulocyte Lysate (Promega) according to the manufacturer’s protocol using mRNA encoding DHFR-NFluc (0.5 pmols) and CFluc-PKA (0.2 pmols) and 0.5 μL of RNasin Plus (Promega) and allowed to incubate for 90 min at 30 °C. Following translation 1 μL of TMP-Staur (125 nM final) and 1 μL of increasing stock concentrations of staurosporine were added to 23 μL of translation mixture and allowed to incubate for 30 minutes at room temperature. Samples were assayed for luciferase activity by combining 20 μL of translation solution with 80 μL of Steady-Glo Luciferase Assay System (Promega). Light emission was monitored 1 minute after Steady-Glo addition using a Turner Biosystems 20/20 Luminometer with a 3 second delay and a 10 second integration time.
APPENDIX A

PLASMID MAPS
pMAL-NGluc-Zif268:

2651 ACAATAACAA TAACAACAAC CTCCGGATCG AGGGAAGGAT TTCAAGATT
       N N N N N N L G I E G R I S E F Frame 1

BamHI
2701 GGATCCCAAGC CCACCGAGAA CAACGAAGAC TTCAACATCG TGGCCGTGGC
       G S K P T E N N F N I V A V A Frame 1

2751 CAGCAACTTC GCGACACCGG ATCTCGATGC TGACCACGGG AAGTTGCCC
       S N F A T T D L A D R G K L P G Frame 1

NruI                        SacII
2801 CAGCAACTTC GCGACACCGG ATCTCGATGC TGACCACGGG AAGTTGCCC
       K K L P E V L K E M E A N R Frame 1

2851 AAAGCTGGCT GCACCAAGGG CTGTCTGATG TGCCGCTCCC ACATCAAGTG
       K A G C T R G C I L S H I K C Frame 1

2901 CACGCCCAAG ATGAAGAAGT TCATCCCAGG ACGCTGCCAC ACCTACGAAG
       T P K M K K F I P G R C H T Y E G Frame 1

PstI
2951 GCGACAAAGA GTCCGCACAG GCGGCATAG GCCTGCAGGG CGGTTCAAGG
       D K E S A Q G G I G L Q G G S Frame 1

KpnI
3001 GTGTTGGGTTT CTGCGCGGGGG TTGGTACCCCG GGGGAGAAGC CCTACGCTTG
       G G G S G G G G Y P G E K P Y A C Frame 1

3051 CCCAGTGGAG ATCCTTGATG GCCGTCGCTTG GAGCTCACCA
       P V E S C D R F S R D E L T R Frame 1

3101 GGCACATCCG CATCCACACA GGCAGAAGGC CCTCCCCAGT GGCATCTCGC
       H I R I T G Q K F F Q C R I C Frame 1

AgeI         HindIII
3151 ATGCCGCAACT TCACGCCCGAG CGACCACCTC ACCACCCACA TCCGCACCCA
       M R N F S R D H L T T H I R T H Frame 1

3201 CACAGCAGAA AAGCCCTTTG CCTGCGACAT CTGGAAAGAG AAGTTTGC
       T G E K P F A C D I C G R K F A R Frame 1

3251 GGAGCGATGA ACGCAAGAGG CATCCACAGG TCCAGCAGGG TGACGAGG
       S D E R K R H T K I H T G E Q K Frame 1

3301 CTGGGCCACTG GCGTGTGTTT TCAACGCTCG TGACTGGLAGG AACCTGCGGC
       L G T G R F T T S * Frame 1
**pMAL-PBSII-CGluc:**

**BamHI**

```
2701  GGATCCCCCG GGGAGAAGCC CTATGCTTGT CCGGAATGTG GTAAGTCCCTT
     G S P G E K P Y A C P E C G K S F  Frame 1
2751  CAGCCAGGCG GCAAACTGTC GCAGCCACCA ACAGTACCAC AGCAGCTGAAA
     S Q R A N L R A H Q R T H T G E K  Frame 1
2801  AACCCTTAAA ATGCCCATGAG TGCCGCAAAT CTTTATAGCG CAGGGATCAC
     P Y K C P E C G K S F S R S D H  Frame 1
2851  CTGACTACCC ATCAAGCCAC TCATACTGGC GAGAAGCCAT ACAATGTCC
     L T T H Q R T H T G E K P Y K C P  Frame 1
2901  AGAATGTGCG AACCTCTTTCA GTCGACAGCG TGTGCTGGTG CGCCACCAC
     E C G K S F S R S D V L V R H Q R  Frame 1
```

**AgeI**

```
2951  GTACTCACAC CGGTGGGGGT GCCGGTTCAG GCCGGGGGT TTCTGGTGGR
     T H T G G G G S G G G G S G  Frame 1
```

**KpnI**

```
3001  GGTGGTACCG AGGCGATCGT CGACATTCCT GAGATTCCTG GGTTCAAGGA
     G G T E A I V D I P E I P G F K D  Frame 1
3051  CTTGGAGCCC ATGGGACAGT TCATCGCACA GGTGAGCTCG TGTGCTGGAC
     L E P M E Q F I A Q V D L C V D C  Frame 1
3101  GCACAACCGG CTGCCTCAA GGGCTTGCCA ACGTGCAGTG TTCTGACCTG
     T T G C L K G L A N V Q C S D L  Frame 1
3151  CTCAAGAAGT GGCTGCAGCG ACGCTTCGCG ACCTTTGCCA GCAAGATCCA
     L K K W L P Q R C A T F A S K I Q  Frame 1
```

**HindIII**

```
3201  GGGCCAGGTTG GACAAGATCA AGGGGGCCGG TGTTGACTAA AAGCGTTGGCA
     G Q V D K I K G A G G D *  Frame 1
```
pMAL-MBD2-CGluc:

EcoRI

2651 ACAATAACAA TAACAAACAC CTGGGGATCG AGGGAGGAT TTCAGAATTG
N N N N N N N L G I E G R I S E F Frame 1

2701 GAAAGCGGCA AAGCGATCGA TTGCCGGGCGG CTGGCCGGGG GTTGGAAAAA
E S G K R M D C P A L P P G W K K Frame 1

2751 AGAAGAGTAGT ATTCGTAADA GCCGGCGCTAG CGCGGGCCAA AGCGATGCTGT
E E V I R K S G L S A G K S D V Y Frame 1

2801 ATTTATTTTAG CCCGACCGCG AAAAATATTC GTAGCAAACC GCAGCTGGCG
Y F S P S G K K F R S K P Q L A Frame 1

2851 CGTTATCTGG GCAACACCGT GGATCTGAGC AGCTTTGATTT TTGGTACCGG
R Y L G N T V D L S S F D F R T G Frame 1

AgeI

2901 CAAAATGACC GGTGGGGGTG GCGGTTCAGG GGCGGGGGGG TCTGGTGGGG
K M T G G G G G G G G G G G G G G Frame 1

KpnI

2951 GTGGTACCGA GGCGATCGTC GACATTCCTG AGATTCCTGG GTTCAAGGAC
G T E A I V D I P E I P G F K D Frame 1

3001 TTGGAGCCCA TGGACATGAT CATCGCACAG GTCGATCTGT GTGTGGACTG
L E P M E Q F I A Q V D L C V D C Frame 1

3051 CACAACCGGC TGCCCCTAGCA GGCTGGCCAA CGTGCAGTGT TCTGACCTGC
T T G C L K G L A N V Q C S D L L Frame 1

3101 TCAAGAAGTG GGCAACGCGCA CGCTGACCGA CCTTTGCCAG CAAGATCCAG
K K W L P Q R C A T F A S K I Q Frame 1

HindIII

3151 GGCCAGGTGG ACAAGATCAG GGCTGGCGGT GGTGACTAAA AGCTTGCGCAG
G Q V D K I K G A G G D * Frame 1
pD-FFlux-N416-C398
7667 bp
PBSII-NFluc (2-416):

51  CTTTAAGAAG GAGATATACC ATGGGCAGCA GCCATCACCA TCATCACACC
  M  G  S  S  H  H  H  H  H  H  H  H  H  H  Frame 2

BamHI  EcoRI
101  AGCCAGGAGT CGAATTCGGA GAAGCCCTAT GCTTGTCGCC AATGGTGTAAT
  S  Q  D  P  N  S  E  K  P  Y  A  C  P  E  C  G  K  Frame 2

151  GTCCCTTCAGC CAGGCGCGAA ACCTGCGCGC CCACCAGCGT ACCCACACGG
  S  F  S  Q  R  A  N  L  R  A  H  Q  R  T  H  T  G  Frame 2

201  GTGAAGGAAG CTGAAAGTGG CGAGGAATTGC CTTTCAGTGG CAGGAGATGT CAGGGTGCC
  E  K  P  Y  K  C  P  E  C  G  K  S  F  S  R  S  Frame 2

251  GATCACCTGA CTACCACCAT AAGCAGCTGA AAGCAGCTGA CTGGTGCGCC
  D  H  L  T  S  Q  R  A  N  L  R  A  H  Q  R  T  H  T  G  Frame 2

AgeI
301  ATGTCCAGAAA CTGGAAGAAG CTGGTGCCG GTGGGGGTTCT CAGGCGCGCC
  C  P  E   C  G  K  S   F  S  R   S  D  V  L  V  R  H  Frame 2

KpnI
351  ACCAACGTAC TCACACCGGT GGGGGTGCGA GTCCAGGCGG TGGGGGTTCT
  Q  R  T   H  T  G   G  G  G  G   S  G  G   G  G  S  Frame 2

401  GGTGGGGGTTG GTACCGAAGA CGCCAAAAAC ATAAAGAAAG GCCCGGGCGCC
  G  G  G  T  G  D  A  K  N  I  K  K  G  P  A  P  Frame 2

451  ATTCTATCCT CTAGAGGATG GAACCGCTGG AGAGCAACTG CATAAGGCTA
  F  Y  P  L  E  D  G  T  A  G  E  Q  L  H  K  A  M  Frame 2

501  TGAAGAGATA CGCCCTGGTGT CTTGGAAACAA TTGCTTTTAC AGATGACATAC
  K  R  Y   A  L  V   P  G  T  I  A  F  T  D  A  H  Frame 2

551  ATCGAGGTTGA ACATCACTGA CGCGGAATAC TTCGAAATGT CCGTTCGTT
  I  E  V  N  I  T  Y   A  E  Y  F  E  M  S  V  R  L  Frame 2

601  GCAGAAGCT ATGAGACAT ATGGGCTGAA TACAAATCAC AGAATGCGTCG
  A  E  A  M  K  R  Y  G  L  N  T  N  H  R  I  V  Frame 2

651  TATGCAATGTA AAACTCTCTT CAATTCCTTA TGCCGCTGCT TGCGCCTGAT
  C  S  E  N  S  L  Q  F  F  M  P  V  L  G  A  L  Frame 2

701  TTTATCGGAG TTGGCAGTTGC GCCGGCGGAAC GACATTTTATA ATGAACGTGA
  F  I  G  V  A  V  A  P  A  N  D  I  Y  N  E  R  Frame 2

751  ATGCTCAAC AGTATGAGCA ATGAGGTGGG ATCCGCTAGG TGATTGGCTG
  L  L  N  S  M  N  I  S  Q  P  T  V  V  F  V  S  K  Frame 2

801  AAAAAAAATTT TGAAAGCTGC AAAAAAATT ACACACTAC KGLQKILNVQK KLPII
  Frame 2

851  CAGAATAATT TTTATCGGAG TTCCTAAAAAG TGATTACCAG GATTCAGTGC
  Q  K  I  I  I  M  D  S  K  T  D  Y  Q  G  F  Q  S  Frame 2
CFluc(398-550)-Zif268:

**NdeI**

1801 AAGAAGGAGA TATACATATG ATGTCCGTTT ATGTAAACAA TCCGGAAGCG

\[\text{Frame 3}\]

1851 ACCAACGCCCT TGATGACCAA GGATGGATGG CTACATCTCTG GAGACATAGC

\[\text{Frame 3}\]

1901 TTACTGGGAC GAAAGACGAA ACCACTCCCAT AGTTGACCIC TTGAAGCTCTT

\[\text{Frame 3}\]

**EcoRV**

1951 TAATTAAATA CAAAGGATAT CAGGTGGCCC CCGCTGAATT GGAATCGATA

\[\text{Frame 3}\]

2001 TTGTTACAAC ACCCAACAT CTTCGACCGG GGGGTGCGAG GCTCTCCCCGA

\[\text{Frame 3}\]

2051 CGATGACCCTG GTGAACCTTC CCGCCGCCCT TTGTTTGTGG GAGACCGGAA

\[\text{Frame 3}\]

2101 AGACGATGAC GAAAGAAAGG ATCGTGGGAT TTCAAGGGAAC TGAAGATGACA

\[\text{Frame 3}\]

2151 ACCGCAGAAA AGTGCGCGGG AGAGTTTGTG TTTGTGGACG AAGTACCGAA

\[\text{Frame 3}\]

2201 AGGTCTTACC GGAAAACTCG ACGCAAGAAA AATCAGAGAG ATCCTCATAA

\[\text{Frame 3}\]

**PstI**

2251 AGGCCAAGAA GGGCGGAAAG TCCAAATTGG GCCTGCCAAGCG GTCTTCCCGA

\[\text{Frame 3}\]

**XmaI**

2301 GGTTGCGCTT CTGCGCGCGG TGCCGGCGCC GGGGATCCGG CTTAGCTTTG

\[\text{Frame 3}\]

2351 CCCAGTGGAG CTCTGTGATC GCCGCTCTCG CCGCTCCGAC GAGCTCACCC

\[\text{Frame 3}\]

2401 GCCACATCCG CATACCCACT GCCAGAGAAC CCTTCCAGTG CCGCATCTGC

\[\text{Frame 3}\]

2451 ATGCGCAACT TCAGCCGCAG CGACCACCTC ACCACCCACA TCCGCACCCA

\[\text{Frame 3}\]

2501 GCAAGAGCCAA GGCCGGCGCG GCCCGAGACAC CTTTCCAGTG CCGCATCTGC

\[\text{Frame 3}\]

2551 GGAGCGATGA AGCAAGAGG CAGGCAAGAC TCCACTPTGC GCAGAAGAC

\[\text{Frame 3}\]
XhoI
2601  CTCGAGTCTG GTAAAGAAAC CGCTGCTGCG AAATTTGAAC GCCAGCACAT
       L E S G K E T A A A K F E R Q H M  Frame 3

PacI
2651  GGACTCGTCT ACTAGCGCAG CTTAATTAAC CTAGGCTGCT GCCACCGCTG
       D S S T S A A *  Frame 3
PBSII-CFluc(394-550):

51 CTTTAAGAAG GAGATATACC ATGGGCAGCA GCCATCACCA TCATCACCAC
   M G S S H H H H H H H Frame 2

BamHI EcoRI
101 AGCCAGGATC CGAATTCGGA GAAGCCCTAT GCTTGTCGCC AATGTGGTAA
   S Q D P N S E K P Y A C P E C G K Frame 2

151 GTCTTTTCAGC CAGCGCGGCA ACCTGCGCGC CCACCAGCGT ACCCACACGG
   S F S Q R A N L R A H Q R T H T G Frame 2

201 GTGAAAACC GTATAAGTCG CCAGAATTTG GCCAAATCTTT TAGCCCGGAG
   E K P Y K C P E C G K S F S R S Frame 2

251 GATCACCTGA CTACCACCTGA AGCGAATCAT ACTTGCGGAGA AGCCACATCA
   D H L T T H Q R T H T G E K P Y K Frame 2

301 ATGTCCAGAA TGTTGCGGAGA CTTTCACTCG CTACGATTTG CTGGTGCCGC
   C P E C G K S F S R S D V L V R H Frame 2

AgeI
351 ACCAACGTAC TCACACCGGT GGGGGTGGCG GTTCAGGGAG TGGGGGTTCT
   Q R T H T G G G G S G G G G S Frame 2

KpnI
401 GGTGGGGGTG GTACCGGACC TATGATTATG TCCGGTTATG TAAACAATCC
   G G G G T G P M I M S G Y V N N P Frame 2

451 GGAAGCGACC AACGCCCTTTA TTGACAAGGA TGGATGGCTA CATTCTGGAG
   E A T N A L I D K D G W L H S G D Frame 2

501 ACATAGCTTA CTGGGACGGA GACGAACACT TCTTCACTGA TGACCACCTG
   I A Y W D E D E H F I V D R L Frame 2

551 AAGTCTTAAA TTAAATACAA AGGATATCG AGTGCCCGTG CTGAATGGGA
   K S L I K Y G Y Q V A P A E L Frame 2

601 ACTCGATATPG TTACAACACCA CCAACATCTT CGAGGCAGGC GTGGCGAGTC
   S I L L Q H F N I F D A G V A G L Frame 2

651 TTCCCGACGA TGACACCGGT GAACCTCCCG CCGCGCATTTG TGGTTTGGAG
   P D D A G E L P A A V V V L E Frame 2

701 CAGGAAAGA CGATGACGGA AAAAGAGATC GTGGATTACG TGGCCAGTCA
   H G K T M T E K I V D Y V A S Q Frame 2

751 AGTAAACACC GCGGAAAAAGT TGCGCGGGAGG AGTGTGGTTT GTGGACGGAG
   V T T A K K L R G G V V F V D E V Frame 2

801 TACGGAAAGG CCTTACCGGA AAACCTCAGC CAGAAAAGAT CAGAGAGATC
   P K G L T G K L D A R K I R E I Frame 2

NotI
851 CTCTAAAGG CCAAGAAGGG CGGAAATACC AAATTGTGAT AAGCGGCGCCGC
   L I K A K K G G K S K L * Frame 2
NFluc(1-398)-Zif268:

NdeI

<table>
<thead>
<tr>
<th>Gene</th>
<th>Start</th>
<th>End</th>
<th>Sequence</th>
<th>Frame</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1001</td>
<td>1051</td>
<td>CCCATCTTAG TATATTAGTT AAGTATAAGA AGGAGATATA CATATGGAGG</td>
<td>M E D</td>
</tr>
<tr>
<td></td>
<td>1051</td>
<td>1101</td>
<td>ACGCCAATAA CATAAAGAAAG GCCGCCGGC GCATTCTATCC TCTAGAGGAT</td>
<td>AKN I K K P A P F Y P L E D</td>
</tr>
<tr>
<td></td>
<td>1101</td>
<td>1151</td>
<td>GGAACCGCTGT GAGAGCAACT GCATAAAGCT ATGAAGAGAT ACGCCCTGTTG</td>
<td>G T A G E Q L H K A M K Y A L V</td>
</tr>
</tbody>
</table>

BsiWI

<table>
<thead>
<tr>
<th>Gene</th>
<th>Start</th>
<th>End</th>
<th>Sequence</th>
<th>Frame</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1151</td>
<td>1201</td>
<td>TCTTGGAACA ATTTGCTTTTA CAGATGCACA TATCGAGGTT ACGCCAGAGC</td>
<td>P G T I A T D A H I E V N I T Y</td>
</tr>
<tr>
<td></td>
<td>1201</td>
<td>1251</td>
<td>ACGCGGAATA CTTCCGAAATG TCCTCGGCTGT TGGCAGAAAGC TATGAAACGA</td>
<td>A E Y F E M S V R L A E A M K R</td>
</tr>
<tr>
<td></td>
<td>1251</td>
<td>1301</td>
<td>TATGGGCTGA ATACAAATCA CAGAATCGTC GTATGCAGTG AAAACTCTCT</td>
<td>Y G L N T N H R I V V C S E N S L</td>
</tr>
<tr>
<td></td>
<td>1301</td>
<td>1351</td>
<td>CGCCCGCGAA CGACATTATAT ATGAACGTG ATTTGCTCAA CAGTATGAAC</td>
<td>G T A G E Q L H K A M K Y A L V</td>
</tr>
<tr>
<td></td>
<td>1351</td>
<td>1401</td>
<td>ATTTCGCAGC CTACCGTAGT GTTTGTTTCC AAAAAGGGGT TGCAAAAAAT</td>
<td>I S Q P T V V F V S K K G L Q K I</td>
</tr>
<tr>
<td></td>
<td>1401</td>
<td>1451</td>
<td>TGGATCGCTGAC AAAACAATTG CACTGATAAT GAATTCCTCT GGATCTACTG</td>
<td>D R D K T I A L I M N S S G S T G</td>
</tr>
<tr>
<td></td>
<td>1451</td>
<td>1501</td>
<td>GGTTACCTAA GGGTGCCGGC CTTCCGCGATA GAACCTGCTGT CGTCAGCTCG</td>
<td>L P K G V A L P H R T A C V R F</td>
</tr>
<tr>
<td></td>
<td>1501</td>
<td>1551</td>
<td>TCGCATGCCA GAGATCCTAT TTTTGGCAAT CAAATCATTC CGGAAGTACAC</td>
<td>S H L P P G F N E Y D F V P E S F</td>
</tr>
</tbody>
</table>

EcoRI

<table>
<thead>
<tr>
<th>Gene</th>
<th>Start</th>
<th>End</th>
<th>Sequence</th>
<th>Frame</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1551</td>
<td>1601</td>
<td>TGTGCTGCA AAAACAATTG CACTGATAAT GAATTCCTCT GGATCTACTG</td>
<td>D R D K T I A L I M N S S G S T G</td>
</tr>
<tr>
<td></td>
<td>1601</td>
<td>1651</td>
<td>GGTGCCCTAA GGGTGCCGGC CTTCCGCGATA GAACCTGCTGT CGTCAGCTCG</td>
<td>L P K G V A L P H R T A C V R F</td>
</tr>
<tr>
<td></td>
<td>1651</td>
<td>1701</td>
<td>TCGGATGCGA GAGATCCTAT TTTTGGCAAT CAAATCATTC CGGAAGTACAC</td>
<td>S H L P P G F N E Y D F V P E S F</td>
</tr>
<tr>
<td></td>
<td>1701</td>
<td>1751</td>
<td>GATTTTAAGT GTTTGTCCAT TCCCTACACG TTTTGGAAGAT TTTTACTCAC</td>
<td>G Y L I C G F R V V L M Y R F E</td>
</tr>
<tr>
<td></td>
<td>1751</td>
<td>1801</td>
<td>TCGGATATTT GATATGCTGA TTTCGAGTCG TCTTAATGTA TAGATTGAAA</td>
<td>G Y L I C G F R V V L M Y R F E</td>
</tr>
</tbody>
</table>
1851 GAAGAGCTGT TTTTACGATC CCTTCAGGAT TACAAAATTC AAAGTGCGTT
  E  E  L  F   L  R  S   L  Q  D   Y  K  I  Q   S  A  L    Frame 3
1901 GCTAGTACCA ACCCTATTTTT CATTCTTCGC CAAAAGCACT CTGATTGACA
  L  V  P   T  L  F  S   F  F  A   K  S  T   L  I  D  K    Frame 3
1951 AATACGATT TTCTAAATTTA CACGAAATGG CTTCCTGGGGG CGCACCTCTT
  Y  D  L  S  N  L  H  E  I  A  S  G  G  A  P  L    Frame 3
2001 TCGAAAGAAG TCGGGGAAAG GGTTGCAAAA CGCTTCCATC TTCCAGGGAT
  S  K  E  V   G  E  A   V  A  K   R  F  H  L  P  G  I    Frame 3
2051 ACGACAAAGGA TATGGGCTCA CTGAGACTAC ATCAGCTATT CTGATTACAC
  R  Q  G   Y  G  L  T   E  T  T   S  A  I  T  P   Frame 3
2101 CCGAGGGGGA TGATAAACCG GGCGCGTCGC GTAAAGTTGT TCCATTTTTT
  E  G  D  K  P  G  A  V  G  K  V  V   P  F  F  Frame 3
2151 GAAGCGAAGG TTGTGGATCT GGATACCGGG AAAACGCTGG GCGTTATTTT
  G  G  D   D  K  P   G  A  V  G   K  V  V   P  F  F     Frame 3
2201 GAGAGGGCGA TTATGTGTCTA GAGGACCTAT GATTATGCTG CAGGGCGGTT
  R  G  E   L  C  V  R  G  P  M  I  M  L  Q  G  G  S   Frame 3
2251 CAGGCCGGTG GGGTTTCTTG GGGGTTGGGA GCCCCGGGGA ACGCCCTTAC
  G  G  G   G  S  G   G  G  G   P  G  E  R  P  Y    Frame 3
2301 GCTTGCCCAG TGGAGTCTTG TGATCGCCGC TTCTCCCGCT CCGACCGGCT
  A  C  P  V   E  S  C   D  R  R   F  S  R  S  D  E  L   Frame 3
2351 CACCCCGCAC ATCCCCATCC ACAAAGCCCAA GAAGCCCTTC CAGTGCGGCA
  T  R  H  I  R  H  T  Q  K  P  F  Q  C  R  I    Frame 3
2401 TCTGCATGCG CAACTTCAGC CGCAGCGACC ACCTCACCAC CCACATCCGC
  C  M  R  N  F  S  R  S  D  H  L  T  T  H  I  R   Frame 3
2451 ACCACACAG GGCAGAAAGCC CTTTGCTCTGC GACATCTGTG GAAGAAAGTT
  T  T  T  G  E  K  P  F  A  C  D  I  C  G  R  K  F  Frame 3
2501 TGCCAGGAGC GATAGACCGA AGAGGCATAC CAAGATCCAC TTGCGGCCGA
  A  R  S  D  E  R  K  R  H  T  K  I  H  L  R  Q  K    Frame 3
2551 AGGACCTCGA GTCTGGGAAA GAACCGGCTG CTGCCGAATTT TGAGCGCCAG
  D  L  E  S  G  K  E  T  A  A  A  K  F  E  R  Q  Frame 3
PstI
2601 CACATGGACT CGTCATACCG CGCAGCTTTA AAGACCTTAG CTGCTGCCAC
  H  M  D  S  S  T  S  A  A  *    Frame 3
pRSF-CFluc-Pum
5353 bp

lad
CFluc
Pum
Kan
CFluc(398-550)-Pum:

NcoI

<table>
<thead>
<tr>
<th>Frame 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>51</td>
</tr>
<tr>
<td>CTTTATAAACGATGATTACCATGGGTATGTCCGGTTATGTAAACAATCCG</td>
</tr>
<tr>
<td>M G M S G Y V N N P</td>
</tr>
<tr>
<td>101</td>
</tr>
<tr>
<td>GAAGCGCACCAGCGCTTGTATGACAAGGATGGATGCTACATTCTGGAGA</td>
</tr>
<tr>
<td>E A T N A L I D K D G W L H S G D</td>
</tr>
<tr>
<td>151</td>
</tr>
<tr>
<td>CATAGCTTACTGGGACGAAGACGAACACTTCTTCATAGTTGACCGCTTGA</td>
</tr>
<tr>
<td>I A Y W D E D E H F F I V D R L K</td>
</tr>
<tr>
<td>201</td>
</tr>
<tr>
<td>AGTCTTTTAATTAATAACGAGATACCGTGATCCCGGCCCCTGTAATAGGAA</td>
</tr>
<tr>
<td>S L I K Y G Q V A P A E L E</td>
</tr>
<tr>
<td>251</td>
</tr>
<tr>
<td>TCGATATTGTACCAACACCACACATCTTGTACGCGGGCGTGACGAGGTCT</td>
</tr>
<tr>
<td>S I L L Q H P N I F D A G V A G L</td>
</tr>
<tr>
<td>301</td>
</tr>
<tr>
<td>TCCCGACGATGACGCGGGTGAACTTCCCCGCGCCGGTTGTGGTTTGGAGC</td>
</tr>
<tr>
<td>P D D D G L P A A V V V L E H</td>
</tr>
<tr>
<td>351</td>
</tr>
<tr>
<td>ACGAAAGACGATGAACGGAAAAAGACAGTAGTGAGACGCCGAGCTCACGAGGAAG</td>
</tr>
<tr>
<td>G K T M T E K E I V D V A S Q</td>
</tr>
<tr>
<td>401</td>
</tr>
<tr>
<td>GTAACCAACCCCGAAAAAGGTGCCCGGAGGATGTGTGTTTGTGGACGAAGT</td>
</tr>
<tr>
<td>V T T A K K L R G G V V F V D E V</td>
</tr>
<tr>
<td>451</td>
</tr>
<tr>
<td>ACCAAAGGTTCTACCGGAAAGCTCAGGCGAAACGAAACTCCACAGAGAGATCC</td>
</tr>
<tr>
<td>P K G L T G K L D A R K I R E I L</td>
</tr>
</tbody>
</table>

BglII

<table>
<thead>
<tr>
<th>Frame 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>501</td>
</tr>
<tr>
<td>TCATAAAGGCGCAAGAAAGGCCGAAAGTCCCAAATTTGGCCTGCCAGGGCGGT</td>
</tr>
<tr>
<td>I K A K K K G K S K L G L Q G G</td>
</tr>
</tbody>
</table>

SalI

<table>
<thead>
<tr>
<th>Frame 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>551</td>
</tr>
<tr>
<td>TCAGGCGGTGCGGTTCTGGCCGGGCGGTGGAGCCCGTCAAGCCGCGGCG</td>
</tr>
<tr>
<td>S G G G G G G S V D G R S R</td>
</tr>
<tr>
<td>601</td>
</tr>
<tr>
<td>CCTGCTGGACAATTTTCTGCAACAACCGCTAAGCGACCGAATCAGGCC</td>
</tr>
<tr>
<td>L I E D F R N R Y P N L Q L R E</td>
</tr>
<tr>
<td>651</td>
</tr>
<tr>
<td>AAATTCGCAGGCCATTATTAGGAATATCGGAGATCGAGCCATGCAAGAGCG</td>
</tr>
<tr>
<td>I A G H I M E F S Q D H G S R</td>
</tr>
<tr>
<td>701</td>
</tr>
<tr>
<td>TTATACGGCTCAAGCTCAGGACCGCGGGAACGCGGCGGTGACGCGTCAAG</td>
</tr>
<tr>
<td>F I Q L K L E R A T P A E R Q L V</td>
</tr>
<tr>
<td>751</td>
</tr>
<tr>
<td>GTTATACGGAATTCTGACCGGCGGTATCAGCTGTATGGTGATTTGTG</td>
</tr>
<tr>
<td>F N E I L Q A A Y Q L M V D V F G</td>
</tr>
<tr>
<td>801</td>
</tr>
<tr>
<td>GCAACTATGGAGATCAGAAATGTTCGGACTGCTGACGGAAGTATAAGGAG</td>
</tr>
<tr>
<td>N Y V I Q K F F E F G S L E Q K</td>
</tr>
<tr>
<td>851</td>
</tr>
<tr>
<td>CTGCGGCGTGGCGGACGCGCTCGCGGCGGACGCGTCAAGCGGCGGCGGTCA</td>
</tr>
<tr>
<td>L A L A E R I R G H V L S L A L Q</td>
</tr>
</tbody>
</table>
901 GATGTATG GC TGCCGC GTGA T T C AGAA A GC GCTGGA A T T ATTCCGAGCG M Y G C R V I Q K A L E F I P S D Frame 2

951 ATCAGCAGAA CGAAATGTTG CGCGAAC TGG ATGGCCATGT GCTGAAATGC Q Q N E M V R E L D G H V L K C Frame 2

1001 GTGAAAGATC AGAACGGCAA CCATGTTG CG CAGAAATGCA TTGAATGCGT V K D Q N G N H V V Q K C I E C V Frame 2

1051 GCAGCCGCAG AGCCTGACT GTATTATTTGA TGCGTTTAAA GGCAGGTGTQ Q P Q S L Q F I I D A F K G Q V F Frame 2

1101 TTGCGCTGAG CACCCATCGG TATGGCTGCC GCCTGATTCA GC CGATCTCG AL S T H P Y G C R V I Q R I L Frame 2

1151 GAACAT TGGCC T G C GATATGACA G ACCCTGCCG ATTCTGGAAG AACTGCATCA E H C L P D Q T L P I L E E L H Q Frame 2

1201 GCATACCGAA CAGCTGGTGC AGGATC AGA T G C A C T AT GTGATTCAGC H T E Q L V Q D Q Y G N V I Q H Frame 2

1251 ATGTGCTGGA ACATGGGC GC CCGGAAGATA AAAGCAAAAT T GTGCGCGAA V L E H G R P E D K S K I V A E Frame 2

1301 ATGCGCGCCA ACGTGCTGTTG GCTGAGCCAG CATAAAATTG CGAGCAACGT I R G N V L V L S Q H K F A S N V Frame 2

1351 GTGGGAAAAA TGCGTGACC TATGCAGACGC CACCCGACGC GC GGATGCTGA V E K C V T H A S R T E R A V L Frame 2

BstZ17I

1401 TTGATGAA GT GTGCACCATG A ACGATGGCC G C C A T GC G GCTGTATACC D E V C T M N D G P H S A L Y T Frame 2

1451 ATGATGAAAG ATCAGTATGC GAACTATGTG GTGCAGAAAA TGATGTGATGT M M K D Q Y A N Y V V Q K M I D V Frame 2

1501 GCGCGAACC GC GGCAGCGCA AAATTTGT G CATAAATTG CCGCGCAT A E P G R K I V M H K I R P H I Frame 2

1551 TTGCACC C TT GC A AT GCA AT CATATGGCA AACATATCTTG CGCGAAACTG A T L R K Y T Y G K H I L A K L Frame 2

EcoRI

1601 AAAAAATATT ATATGAAAAA CGCGCTGGA T CTGCGCGAAT T C C AT CACCA E K Y Y M K N G V D L G E F H H H Frame 2

NotI

1651 TCATCAC C C A T A A A A A C TTGCACC CGC AT AATGCTTA AGCGAACAG H H H * Frame 2

HindIII
pETDuet-PumMut-NFluc
7733 bp

AMP
PumMut-NFluc:

\[ \text{NcoI} \]

\[
\begin{align*}
51 & \quad \text{CTTTAAGAAG GAGATATACC ATGGGCAGCA GCCATCACCA TCATCACCAC} \quad \text{MGSSHHHHH} \quad \text{Frame 2} \\
\end{align*}
\]

\[ \text{BamHI} \]

\[
\begin{align*}
101 & \quad \text{AGGCAGGATTC CGGGCCGCAG CGGCCTGCTG GAAGATTTC GCAACAACCG} \quad \text{SQDFRERSHH} \quad \text{Frame 2} \\
151 & \quad \text{CTATCCGAAC CTGCACTGCTG CGCAATTTGA GGCCCATATT ATGGAATTTA} \quad \text{YPNLQLREIA} \quad \text{Frame 2} \\
201 & \quad \text{GCCAGGATCA GCATGGGCAGC CGCTTTATTC AGCTGAAACT GGAACGCGCG} \quad \text{QDPQRSSF} \quad \text{Frame 2} \\
251 & \quad \text{ACCCCGGGCGG AACGCCAGCT GGTGTTTAAC GAAATTCTGC AGGGCGCTTA} \quad \text{TPAESQLFYE} \quad \text{Frame 2} \\
301 & \quad \text{TCAGCTGATAG GTGGATTGTTG TTGCCAACTA TGTTGATTCA AAATTTTTTG} \quad \text{QLMVDFGNYVIQ} \quad \text{Frame 2} \\
351 & \quad \text{AATTTGGCAG CCTGGAACAC AAACCTGCGC TGCCGGAAACG CATTCCGGGC} \quad \text{FGSLEQL} \quad \text{Frame 2} \\
401 & \quad \text{CATGTGCTGA GCCTGGCGCT GCAGATGTAT GGCTGCCGCG TGATTCAGAA} \quad \text{HVLSLALQMY} \quad \text{Frame 2} \\
451 & \quad \text{AGGCAGGGAA TTTATTCCGA GCCTGGCGCT GCAGATGTAT GGCTGCCGCG} \quad \text{ALEFIQSDQ} \quad \text{Frame 2} \\
501 & \quad \text{TGGATGGCCA TGATGGCTAAA TGCGTGAAAAG ATCAGAACGG CAACCATGTG} \quad \text{DGHVLKCVKDQ} \quad \text{Frame 2} \\
551 & \quad \text{GTGCAGAAAT GCATTTGAAG CTGGCTGAGC CAACCATGTG CAACCATGTG} \quad \text{VQKCIIEC} \quad \text{Frame 2} \\
601 & \quad \text{TGATGCGTTT TAAAGCCACGG TGGTTGCAGC GAGCAGCTGA GCAGGCGCGTA} \quad \text{DFAFKGVFALS} \quad \text{Frame 2} \\
651 & \quad \text{GCCGGCGGTAT TCAGCCGATT CTGGAACAC GCTGGCCGGA TCAGACCTTG} \quad \text{RVIQRIELYHC} \quad \text{Frame 2} \\
701 & \quad \text{CCGATTCTGG AAGAAGCTCA TCAGCATTAC GAACAGCTGG TGCAAGATCA} \quad \text{PILEELHQET} \quad \text{Frame 2} \\
751 & \quad \text{GTATGCCAGC TATGGTATTG AACATGTGGT GGAACATGGC CGCCGGAAAG} \quad \text{YGSYVIEHVL} \quad \text{Frame 2} \\
801 & \quad \text{ATAAAAAGCAA AATTTGCGCG GAAGATCTCG CAAAGCTGCT GTGTCGCGC} \quad \text{KSFIGAVG} \quad \text{Frame 2} \\
851 & \quad \text{CAGCATAAT TTGCCGACAA CGTGGTGCAG AAATGCGCGA CCCATGCGAG} \quad \text{QHKFANNVQ} \quad \text{Frame 2}
\end{align*}
\]
pDMBD2-NFluc(-MCS2)
6635 bp
### MBD2-NFluc:

<table>
<thead>
<tr>
<th>Restriction Enzyme</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>NcoI</strong></td>
<td>CTTTAAGAAG GAGATATACC ATGGGCAGCA GCCATCACCA TCATCACCAC</td>
</tr>
<tr>
<td><strong>BamHI</strong></td>
<td>AGCCAGGGATC CGAATTCCGA AAGCGGCAAA CGCATGGATT GCCCGGCAGCT</td>
</tr>
<tr>
<td><strong>AgeI</strong></td>
<td>CGGGCAAAAG CGATGTGTAT TATTTTAGCC CGAGCGGCAA AAAATTTTGT</td>
</tr>
<tr>
<td><strong>KpnI</strong></td>
<td>GTGGGGGTTC TGGTGGGGGT GGTACCGAAG ACGCCAAAAA CATAAAGAAA</td>
</tr>
<tr>
<td><strong>BsiWI</strong></td>
<td>CAGATGCACA TATCGAGGTG AACATCACGT ACGCGGAATA CTTCGAAATG</td>
</tr>
</tbody>
</table>

Frame 2
851 GGATTTCAGT CGATGTACAC GTTCGTCACA TCTCATCTAC CTCCCGGTTT
   GFQSMYTFVTSHLPFPGF Frame 2
901 TAATGAATAC GATTCTGTAC CAGAGTACCTT TGATCTGTAC AAAACAATTG
   NEDFVPSFEDRDKTI Frame 2
951 CACTGATAAT GAATACCTCT CGATCTACTG GGTTACTTTA AAAGTGCCGG
   LGMNSSTGLPGBKVA Frame 2
1001 CTTCGCATAA AAGCTACGCT TGGTACGTCC TCGATCTGCA AGATCTCTAT
   LPHRACVGFSHARDPI Frame 2
1051 TTTTGCAAT CAAATCTTC CAGATACCTG GTTTTAAGT GTTGTCCAT
   FGNIPIFDTSAVVFS Frame 2
1101 TCCACAGCG TTTTGAATGC TTTACTACAC TCGGATTTTT GATATGTGGA
   HHGFMFTTLGLICG Frame 2
1151 TCGAGTGCAG TCTAAATGTA TAGATTTGGA GAAGACTCTG TTTACGATC
   FTRVLFSFEELFRS Frame 2
1201 CCTTCAAGAT TACAAAAATC AAATGTGGGT TCGAGTACCA ACCTATTTT
   LQDYKIQSALVLPSFL Frame 2
1251 CATTCTTCCG CAAAGACTC GTGACAGATC AAGACATTG ATCTAATTTA
   FFAKSTLIDKYLDSL Frame 2
1301 CAGAATAATTG CTTTCTGCGGT CGCATCCTTT TCGAAAGAGA TCGAGGAGAC
   HEIARGAALPSKEVGEA Frame 2
1351 GGTTGCAAAA CGTTTCCATC TTTCAAGGAT ACGACAAGGA TATGGGCTCA
   VAKRFTLHPGIRQGYGT Frame 2
1401 CTGAGACTAC ATCAAGCTTT CTGATACGCA CCGAGGAGGA TGATAAACCG
   ETTSAITPEGEDKPS Frame 2
1451 GCCCGGTTGCA GTAAAGTGTG TTCATTTCATTT GAAGCGAGG TACGAGACTC
   GAVGKVVPFFEAKVVDL Frame 2
1501 GAGTACCGGG AAAACGTCGT GCAGTCTAAT CAAGAGGCGA TATGGTGTCA
   DTGTKLVQRGELCVR Frame 2
1551 GAGACCTAT TACATTGTAC GGTATTACCA ACAATCGGGA AGCGACCAAC
   GPMIMSGYVNPNPEATN Frame 2

NotI BglII EcoRV
1601 GCCCTGATTTA ACAAGGATGG ATGATAAGGG GCCGAGATC TGATATCCCT
   ALIDKDG* Frame 2
The above vector map represents a series of helix containing vectors utilized to study the interactions between various helix and receptor proteins.
BMF-NFluc:

KpnI
HindIII
901 GTTAAGCTTG GTACCGAGCT CGGATCCCCA CCATGCATCA AGCAGAGGTA M H Q A E V

BamHI
HindIII
951 CAGATTGCAC GAAAGCTTCA GTGCATTGCA GACCAGTTCC ACCGGACGTA Q I A R K L Q C I A D Q F H R T Y

HindIII
BsiWI
1001 CGCGTCCCGG GGCggTGGCT CATCTGGCGG AGGTGAAGAC GCCAAAAACA

XmaI
SmaI
1051 TAAAGAAAGG CCCGGCGCCA TTCTATCCGC TGGAAGATGG AACCGCTGGA K K G P A P F Y P L E D G T A G

1101 GAGCAACTGC ATAAAGCTAT GAAGAGATAC GCCCTGGTTC CTGGAACAAT E Q L H K A M K R Y A L V P G T I

1151 TGCTTTTACA GATGCACATA TCGAGGTTGA CATCACTTAC GCTGAGTACT A F T D A H I E V D I T Y A E Y F

1201 TCAGAAATGTC CGTTCGGTTG GCAGAAGCTA TGAAACGATA TGGGCTGAAAT E M S V R L A E A M K R Y G L N

1251 ACAAAATCAACA GAATCGTGCGT ATGCAGTGAA AACTCTCTTC TTTTCTTTAT T  N H R I V V C S E N S L Q F F M

1301 GCCGGGTTGG GGCggGTGGAT TTATCGGAGT TGCAGTTGCG CCCGCGAACG P V L G A L F I G V A A P A N D

1351 ACATTATAAA TGAAGCTGAA TTGCTCAACA GTATGGCGAT TTCGCGAGCT I Y N R E E L L N S M G I S Q P

1401 ACCGTGGGTG TCGTTTCCAA AAAGGGTTG CAAAAATTT GGAAGCTGCA T V V F V S K K G L Q K I L N V Q

1451 AAAAAAGCTC CCAATCATCC AAAAAATATT TATCATGGAT TCTAAGACCG K K L P I I Q K I I I M D S K T D

1501 ATTCACAGGG ATTTCACTCG ATGTACACGT TCGTACATCG TCATCTACCT Y Q G F S M Y T F V T S H L P

1551 CCCGGGTTTG ATGAATACGA TTTTGTGCCA GAGTCCTTCG ATAGGGACAA P G F N E Y D F V P E S F D R D K

1601 GACAATTCGA CTGATCATGA ACTCCTCTGG ATCTACTGGT CTGACCTAAG T I A L I M N S S G S T G L P K G

1651 GTGTCGCTCT GCCTCATAGA ACTGCCTGCG TGAGATTCTC GCATGCCAGA V A L P H R T A C V R F S H A R

Frame 3
245

1701 GATCCTATTT TTGGCAATCA AATCATTCCG GATACTGCGA TTTTAAGTGT
     D P I F G N Q I I P D T A I L S V Frame 3

1751 TGTTCATTC CATCACGGTT TTGGAAATTT TACTACACTC GGATATTGTA
     V P F H H G F G M F T T L G Y L I Frame 3

1801 TATGTCGATT TCGAGTCGAT TTAATGTATA GATTGAAGA AGAGCTGTTT
     C G F R V V L M Y R F E E E L F Frame 3

1851 CTGAGGAGCC TTCAGGGATT CAAGATCTCA AGTGCAGTTC TGGTGCCAAC
     L R S L Q D Y K I Q S A L L V P T Frame 3

1901 CCTATTCTCC TTCTTCGCCA AAAGCACTCT GATTGACAAA TACGATTAT
     L F S F F A K S T L I D K Y D L S Frame 3

1951 CTAATTTCACA CGAAATTGCT TCTGGTGCGC CTGCCCTCTC TAAGGAAGTC
     C G F R V V K I I Q S A L L V P T Frame 3

2001 GGGGAAGCGG TTGCCAAGAG GTTCCATCTG CCAGGTATCA GGCAAGGATA
     G E A V A K R F H L P G I R Q G Y Frame 3

2051 TGGGCTCAGT GAGACTACAT CAGCTATTCT GATTACACCC GAGGGGGATG
     G L T E T T S A I L I T P E G D D Frame 3

2101 ATAAACCGGG CGGGGTGCGT AAAGTGTTTC CATTTTTTGA AGCGAAGGTT
     K P G A V G K V V P F F E A K V Frame 3

2151 GTGGATCTGG ATACCGGGAA AACGCTGGGC GTTAATCAAA GAGGCGAACT
     V D L D T G K T T L G V N Q R G E L Frame 3

2201 GTGTGTGAGA GGTCTTATGA TTATGTCCGG TTATGTAAAC AATCCGGAAG
     C V R G P M I M S G Y V N N P E A Frame 3

2251 CGACCAACGC CTTGATTGAC AAGGATGGAT GATAACATTG TGGTAGCTTAG
     T N A L I D K D * Frame 3
BAD-NFluc:

<table>
<thead>
<tr>
<th>Restriction Enzyme</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>KpnI</td>
<td>GTTAAGCTTTG GTACCGAGCT</td>
</tr>
<tr>
<td>HindIII</td>
<td>CGGATCCCCCA CCATGAACCT</td>
</tr>
<tr>
<td>BamHI</td>
<td>CTGGGCAGCA M N L W A A</td>
</tr>
<tr>
<td></td>
<td>Frame 3</td>
</tr>
<tr>
<td>KpnI</td>
<td>CAGCGCTATATG GCGCAGAGCT</td>
</tr>
<tr>
<td>HindIII</td>
<td>CCGGAGGATG AGTGACGAGT</td>
</tr>
<tr>
<td>BamHI</td>
<td>TTGTGGACTC Q R Y G R E L R R M S D E F V D S</td>
</tr>
<tr>
<td></td>
<td>Frame 3</td>
</tr>
<tr>
<td>XmaI</td>
<td>CTTTAAGAAG ATCTCGTACG</td>
</tr>
<tr>
<td>BglII</td>
<td>CGTTCCGGGG</td>
</tr>
<tr>
<td>BsiWI</td>
<td>CGGTGGCTCA TCTGGCGGAG</td>
</tr>
<tr>
<td>SmaI</td>
<td>Frame 3</td>
</tr>
<tr>
<td>XmaI</td>
<td>GTGAAGAGCGC CA AAAACATA</td>
</tr>
<tr>
<td>BglII</td>
<td>AAGAAAGGCC CGGCCCAATT</td>
</tr>
<tr>
<td>BsiWI</td>
<td>CTATCCGGCTG E D A K N I K K P A P F Y P L</td>
</tr>
<tr>
<td>SmaI</td>
<td>Frame 3</td>
</tr>
<tr>
<td>XmaI</td>
<td>GAAGATGGAA CCGCTGGAGA</td>
</tr>
<tr>
<td>BglII</td>
<td>GCAACTGCAT AAGGCTATGA</td>
</tr>
<tr>
<td>BsiWI</td>
<td>AGAGATACGC E D G T A G E Q L H K A M K R Y A</td>
</tr>
<tr>
<td>SmaI</td>
<td>Frame 3</td>
</tr>
<tr>
<td>XmaI</td>
<td>CCTGGTTTCT GTGAAGATCC</td>
</tr>
<tr>
<td>BglII</td>
<td>GGACCAATTG CTGCACTATAC</td>
</tr>
<tr>
<td>BsiWI</td>
<td>GGAGTGGACA L V P G T I A F T D A H I E V D I</td>
</tr>
<tr>
<td>SmaI</td>
<td>Frame 3</td>
</tr>
<tr>
<td>XmaI</td>
<td>TCACCTACGC TGAGTCTCC</td>
</tr>
<tr>
<td>BglII</td>
<td>GAATATGCGC TCGTGGTGG</td>
</tr>
<tr>
<td>BsiWI</td>
<td>ATGGGAACATG E D A K N I K K G P A P F Y P L</td>
</tr>
<tr>
<td>SmaI</td>
<td>Frame 3</td>
</tr>
<tr>
<td>XmaI</td>
<td>AAACGATATG GGCTGAATAC</td>
</tr>
<tr>
<td>BglII</td>
<td>AAATCACAGA ATCGTCGTAT</td>
</tr>
<tr>
<td>BsiWI</td>
<td>GCAGTGAAAA K R Y G L N T N H R I V V C S E N</td>
</tr>
<tr>
<td>SmaI</td>
<td>Frame 3</td>
</tr>
<tr>
<td>XmaI</td>
<td>CTCTCTCACA TTCTTATAG</td>
</tr>
<tr>
<td>BglII</td>
<td>CGGTGTTGGG CGGTGTTT</td>
</tr>
<tr>
<td>BsiWI</td>
<td>ATCGGAGTTG S L Q F F M P V L G A L F I G V A</td>
</tr>
<tr>
<td>SmaI</td>
<td>Frame 3</td>
</tr>
<tr>
<td>XmaI</td>
<td>CAGTTGCGCC CGCGAACAC</td>
</tr>
<tr>
<td>BglII</td>
<td>ATTTATAATG AAGTAAGATT</td>
</tr>
<tr>
<td>BsiWI</td>
<td>GCTCAACAGT V A P A N D I Y N E R E L L N S</td>
</tr>
<tr>
<td>SmaI</td>
<td>Frame 3</td>
</tr>
<tr>
<td>XmaI</td>
<td>ATGGGGCATT TTCCAGGTCAC</td>
</tr>
<tr>
<td>BglII</td>
<td>CGTTGGTTTC GTTTCCAAAA AGGGGTGGCA M G I S Q P T V V F V S K K G L Q</td>
</tr>
<tr>
<td>BsiWI</td>
<td>Frame 3</td>
</tr>
<tr>
<td>XmaI</td>
<td>AAAAATTTTG AACGTGCCAA AAAAGCTCACA AATCATCACA AAAATTATTA K I L N V Q K K L P I I Q K I I</td>
</tr>
<tr>
<td>BglII</td>
<td>Frame 3</td>
</tr>
<tr>
<td>XmaI</td>
<td>TCGAAGATTT GCCAAGGTCAT</td>
</tr>
<tr>
<td>BglII</td>
<td>CGTTGGTTTC GTTTCCAAAA AGGGGTGGCA M G I S Q P T V V F V S K K G L Q</td>
</tr>
<tr>
<td>BsiWI</td>
<td>Frame 3</td>
</tr>
<tr>
<td>XmaI</td>
<td>GTCACACATTC ATCTACCTCC</td>
</tr>
<tr>
<td>BglII</td>
<td>CGGTGGTTTC GTTTCCAAAA AGGGGTGGCA M G I S Q P T V V F V S K K G L Q</td>
</tr>
<tr>
<td>BsiWI</td>
<td>Frame 3</td>
</tr>
</tbody>
</table>

Frame 3
PUMA-NFluc:

<table>
<thead>
<tr>
<th>901</th>
<th>GTTAAGCTTTG</th>
<th>GTACCGAGCT</th>
<th>CCGATCCCCA</th>
<th>CCATGGAACA</th>
<th>GTGGGCCCGG</th>
</tr>
</thead>
<tbody>
<tr>
<td>951</td>
<td>GAGATCGGGG</td>
<td>CCCAGCTGCG</td>
<td>GCGGATGGCG</td>
<td>GACCGGTCCC</td>
<td>GGGCCGGTGG</td>
</tr>
<tr>
<td>1001</td>
<td>CTCACTCTGCG</td>
<td>GGAGGTGAAG</td>
<td>AGGCCAAAAA</td>
<td>CATAAAAGAAA</td>
<td>GGCACGCGGC</td>
</tr>
<tr>
<td>1051</td>
<td>CATTCTATCC</td>
<td>GCTGGAAGAT</td>
<td>GGAACCGGCTG</td>
<td>GAGACCAACT</td>
<td>GCATAAGGCT</td>
</tr>
<tr>
<td>1101</td>
<td>ATGAAGAGAT</td>
<td>ACGCCCTGGT</td>
<td>TCCTGGAACA</td>
<td>ATTGCCTTTTA</td>
<td>CAGATGCACA</td>
</tr>
<tr>
<td>1151</td>
<td>TATCGAGGTG</td>
<td>GACATCUCTT</td>
<td>ACCTCAGGAT</td>
<td>TTCTCAGGTT</td>
<td>TGGGCGCGGT</td>
</tr>
<tr>
<td>1201</td>
<td>GTATCcAGTG</td>
<td>AAAAACTCTCT</td>
<td>TCAATATTTTT</td>
<td>ATGCCGCGTT</td>
<td>TGTCGTCACA</td>
</tr>
<tr>
<td>1251</td>
<td>ATTTATCGGA</td>
<td>GTTGCAGTTG</td>
<td>CGCCCGCGAA</td>
<td>CGACATTTAT</td>
<td>AATGAAGGCT</td>
</tr>
<tr>
<td>1301</td>
<td>ATTTATCGGA</td>
<td>CAGTATGGCC</td>
<td>ATTTTCTGGG</td>
<td>CTACCTGGGTC</td>
<td>GTTGTTTCC</td>
</tr>
<tr>
<td>1351</td>
<td>CAAATCATTC</td>
<td>CGGATACTGC</td>
<td>GATTTTAAGT</td>
<td>GTTGTTCCAT</td>
<td>TCCATCACGG</td>
</tr>
<tr>
<td>1401</td>
<td>AAAAAAGGGG</td>
<td>TGCAAAAAAT</td>
<td>TTTGAAGCT</td>
<td>CGATAATGGT</td>
<td>CACTGATCAT</td>
</tr>
<tr>
<td>1451</td>
<td>CCAAAAAATT</td>
<td>ATTATCATGG</td>
<td>ATTCTAAAAC</td>
<td>GGATTACCAG</td>
<td>GGATTTCAGT</td>
</tr>
<tr>
<td>1501</td>
<td>CGATGTACAC</td>
<td>GTCAGTACCT</td>
<td>CTCTCATCTAC</td>
<td>CTCCCGGTTT</td>
<td>TAATGAATAC</td>
</tr>
<tr>
<td>1551</td>
<td>GATTTTGTGC</td>
<td>CAGATGCTTT</td>
<td>CGATAGGGAC</td>
<td>AAGACAATTG</td>
<td>CACTGATCAT</td>
</tr>
<tr>
<td>1601</td>
<td>GAACCTCTCT</td>
<td>GGATCTACTG</td>
<td>GCTGCTACTA</td>
<td>AGGTGTGCT</td>
<td>CTGCTCTACA</td>
</tr>
<tr>
<td>1651</td>
<td>GAACCTCTTG</td>
<td>CGTCTAGTCC</td>
<td>TCCATGATCC</td>
<td>GAGATCTATAT</td>
<td>TTTTGGCAAAT</td>
</tr>
<tr>
<td>1701</td>
<td>CAAATCATTC</td>
<td>CCGATCTACTG</td>
<td>GATTTTAATG</td>
<td>GTGTTCCAT</td>
<td>TCCATACCGG</td>
</tr>
</tbody>
</table>
1751 TTTTGGATTG TTTACTACAC TCGGATATTTT GATATGTGGA TTTCGAGTGC
FGMF TTTLLG YLICGF R V V Frame 3
1801 TCTTAATGTA TAGATTTGAA GAAGAGCTGT TTCTGAGGAG CCTTCAGGAT
LMYR FE E E ELF LRS L Q D Frame 3
1851 TACAGATTC AAGTGGCGCT GCTGTG GCCA ACCCTATTTC CCTTCAGGAT
YKIQ S ALLV P T L F S FF A Frame 3
1901 CAAGATCACT CTGATACCA AATAGGATTT ATCTAATTAA CACGAAATG
KSTLIDKKYDLSNLEGIA Frame 3
1951 CTTCTGGTGG CGCTCCTCTC TCTAAGGAAG TCGGGAAAGC GGTGCAAG
SGAPL SKEVGEAVAK Frame 3
2001 AGGTTCCATC TGCCAGGTAT CAGGAAGGTA TATGGCTCA CTGAGACTAC
RFHLPGIRQGYYGLTETT Frame 3
2051 ACTGAGCTAT CTGATTACAC CCGAGGGGGA TGATAAACCG GGCGCGTGC
SAILITPEGD DKPGAVG Frame 3
2101 GTAAGTTGT TCCATTTTTT GAAGCGAAGG TTGTGGATCT GATAGCGGG
KVPPPFEAKVVDLDTG Frame 3
2151 AAGAGCTTGG GCGTATCTCA AAGAGCGGAA CTGATGTTGA GAGGTCTAT
KTGLVNRGERLCEVGRPM Frame 3
2201 GATATATGTC GGTTATGTTA AACAATCGGA AGCGACCAAC GCCTTGATG
IMSGYNPENPAATNALD Frame 3
2251 ACAAGGATGGA ATGATAACAT TCTGTTAGCT AGGTAATGCA TATCTAGCT
KDGE Frame 3
BAK-NFluc:

<table>
<thead>
<tr>
<th>Restriction Enzyme</th>
<th>Sequence</th>
<th>Frame</th>
</tr>
</thead>
<tbody>
<tr>
<td>KpnI, HindIII, BamHI</td>
<td>GTTAAGCTTG GTACCGAGCT CGGATCCCA CCATGGGGCA GGTGGGACGG</td>
<td>3</td>
</tr>
<tr>
<td>BsiWI</td>
<td>QLAIGDINRRYDTYA</td>
<td>3</td>
</tr>
<tr>
<td>XmaI, SmaI</td>
<td>GTCCCCGGGCG GTGGGGTCTCAT CTGGCGGAGG TGAAGACGCC AAAAACATAA</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>AGAAAGGGCCC GGCAGCTCTTC TATCCGCTGG AAGATGGGAC CGCTGGAGAG</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>CAACGTCAAT AGGCTATGAA GAGATACGCC CTGGTTCCCTG AAACATTGGA</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>AGATGTCGT GTGGTGGCGA GAAGCTATGA AACGATATGG GCTGAATACA</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>AATCAGCAGA TCGTGGTATG CAGTGGAAAC TCTCTTCAAT TCTTTATGCC</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>GTGGTGTTCG TTTCCAAAAA GGGGTTGCAA AAAATTTTGA ACGTGCAAAA</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>AAAAGCTCCCA ATCATCCAAA AAATTATTAT CATGGATTCT AAAACGGATT</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>ACCAGGGATT TCAGTGGACT TGAGCTGAGTC ACATCTCTCA TCACTGCTCC</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>GTGGTGGTGGC TTTCCAAAAA GGGGTTGCAA AAAATTTTGA ACGTGCAAAA</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>AAAGCTCAGA ATATCAGATG CTGCCTGGCAT TCCAGGGTCTG GGGACAGAC</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>CCTATTTTTTG GCAATCAAAT CATTCCGGAT ACTGCGATTT TAAGTGTTGT</td>
<td>3</td>
</tr>
</tbody>
</table>
PIFGNQIIPDTAILSVV Frame 3

P  I  F  G  N  Q  I  I  P  D  T  A  I  L  S  V  V Frame 3

1751 TCCATTTCCAT CACGGTTTTG GAATTTTAC TACACTCGGA TATTTGATAT
PFHGMFTTLGYLIC Frame 3

1801 GTGGATTTCCG AGTCGCCTTA ATGTATAGAT TTGAAGAAAGA GCTGTTTCTG
GFRVLMYRFEEELFL Frame 3

1851 AGGAGCCTTC AGGATTACAA GATTCAAAGT GCGCTGCTGG TGCCAACCCT
RSLDYKIQLSVLVPTRL Frame 3

1901 ATTCTCCTTC TTCCCAAAAA GCACTCTGAT TGACAATAC GATTTATCTA
FSFPAKSTLIDKYLDSLN Frame 3

1951 ATTTACACGA AATTGCTTCT GGTGGCGCTC CCCCCTCTAAGA GGAAGTCGGG
LHEIASGPALSLKEVG Frame 3

2001 GAAGCGGTTTG CCAAGAGGTT CCATCTGCAA GGTATCAGGC AAGGATATGG
EAVAKRFHLPGRQGYG Frame 3

2051 GCTCACTGAG ACTACATCAG CTATTCTGAT TACACCGAG GGGATGATA
LTETTSAILITPEGDDK Frame 3

2101 AACCGGCGGC GGTCCGTTAA GGTGTTCCTAT TTTTTGAAGC GAAGGTTTGT
PGAVKGKVVFPEAKVV Frame 3

2151 GATCTGGATA CCGGGAAACG GCTGGCGGTT AATCAAGAG GCGAACTGTT
DLDTGKTLGVPNQRGEKC Frame 3

2201 GTGGAAGGTT CCTATGATTA TGCCGGTTTA TGTAACAAT CCGGAAGCGA
VRGPMMISSGYSVNNPPEAT Frame 3

2251 CCAACGCCCTT GATGGCAAGA GATGGATGAT AACATTCTGG TAGCTAGGTA
NALIDKDGG* Frame 3
HRK-NFluc:

KpnI
HindIII
BamHI

901 GTTAAGCTTG GTACCGAGCT CGGATCCCAC CCATGTCTGC CGCCGGCGCAG
M S S A A Q Frame 3

BsiWI

951 CTCACCGCGG CCCGGCTCAA AGCGCTAGGC GACGAGCTGC ACCAGACGTA
L T A A R L K A L G D E L H Q T Y Frame 3

XmaI
SmaI

1001 CGCGTCTCCGG GCAGGCGTGGT CATCTGGCGG AGGTGAAGAC GCCAAAAACA
A S R G G S S G G G D A K N I Frame 3

1051 TAAAGAAAAG CCCGGCCTCGC TTCTATCCGC TGGAAGATGG AACCCTGGGA
K K G P A P F Y P L E D G T A Frame 3

1101 GACCAACTGC ATAAGCGCTAT GAAGAGAGAT GCCCCGGTTTC CTGAAACAT
E Q L H K M K R Y A L V P G T I Frame 3

1151 TGCTTTTATA GATGCGCACA TCGAGTGGGA CATCACTTAC GCTAGTACT
A F T D A H I E V D I T Y A E Y F Frame 3

1201 TCGAAATGTC CGTCTGGTTT GCCGAAGACT TGAGTACAT GCCTGACAT
E M S V R L A E M K R Y G L N Frame 3

1251 ACAATCTACA GAATCGTCGT ATGCAGTGAA AACTCTTCTC AATTCTTAT
T N H R I V V C S E N S L Q F F M Frame 3

1301 GCCGGTTTTG GCCGCGTTTT TTATCGGAGT TGCAGCTGCG CCCGCGAACG
P V L G A L F I G V A V A P A N D Frame 3

1351 ACATTTTATA TGGATCGTAA TTGCTCAACA GTATGGGCA GCCGCGACG
I Y N E R E L L N S M G I S Q P Frame 3

1401 ACCGTGTTGT TCGTTTCCAA AAGGGGTGGT CAAAAATTT TGACGTGCA
T V V F V S K K G L Q K I L N V Q Frame 3

1451 AAAAAAGCTC CCAATCATCC AAAAAATTAT TATCATGGAT TCTAAAACGG
K K L P I I Q K I I I M D S K T D Frame 3

1501 ATTACCAGGG ATTTACGTGA ATGTAGAGTC TCGTCACATC TCAACTACCT
Y Q G F Q S M Y T F V T S H L P Frame 3

1551 CCCGGTTTTA ATGAAATCAGG TTTTGTGCCA GAGTCCTCGG ATAGGGACAA
P G F N E D F V P E S F D R D K Frame 3

1601 GACAATTTCA CTGATCATGA ACTCCTCTGG ATCTACTGGT CTGCCTAAAAG
T I A L I M N S S G S T G L P K G Frame 3

1651 GTGTGTCCTCT GCCTCATAAG ACTGCGTGGG TGAGATTCTC GCATGCCAGA
V A L P H R T A C V R F S H A R Frame 3
1701 GATCCTATTT TTGGCAATCA ATTCATCCG GATACTGCAG TTTTAAGTGT
  D P I F G N Q I I P D T A I L S V Frame 3
1751 TGTTCCATTG CATCAGCGTGT TTGGGAATTT TACTACACTC GGATATTTGA
  V P F H H G F M F T T L G Y L I Frame 3
1801 TATATGCGATT TCGAGTCTGT TTAATGTATA GATTGAAAGA AGAGCTGTTTT
  C G F R V V L M Y R F E E E L F Frame 3
1851 CTGAAGGAGCC TTCAGGATTAA CAAGATTCAA AGTGCGCTGC TGGTGCCCAAC
  L R S L Q D Y K I Q S A L L V P T Frame 3
1901 CCTATTCTCC TTCTTCGCCA AAAGCACTCT GATTGACAAA TACGATTTAT
  L F S F A K S T L I D K Y D L S Frame 3
1951 CTAATTTCACA CGAAATGTCT TCTGCTGCGC CTCCCCCTCTC TAAGGAAGTC
  N L H E I A S G G A P L S K E V Frame 3
2001 GGGGAAGCGG TTGCCCAAGAG GTTCCATCTG CCAGGTATCA GGCAAGGATA
  G E A V A K R F H L P G I R Q G Y Frame 3
2051 TGGGTCTACAG GAGACAGCAG CTACCTTTCT GATTACACCAG GAGGGGGATG
  G L T E T T S A I L I T P E G D D Frame 3
2101 ATAAACCGGG CCGGGTCTCGT AAAGTTGTTC CATTTTTTGA AGCGAAGGTT
  K P G A V G K V V P F F E A K V Frame 3
2151 GTGATCTGGG ATACCCGGAA AAGCTGGGCA GTTAACTGAA GAGGGGAACCT
  V D L D T G K T L G V N Q R G E L Frame 3
2201 GTGTGTGAGA GGTCTATATGA TTATGTCCGG TTATGTAACAA ATCCGGAAG
  C V R G P M I M S G Y V N N P E A Frame 3
2251 CGACCAACGC CTTGATTGAC AAGGATGGAT GATAACATTC TTGTAAGCTAG
  T N A L I D K D G * Frame 3
BIM-NFluc:

KpnI

HindIII

BamHI

901 GTTAAGCTTG GTACCGAGCT CGGATCCCCA CCATGCAGCC AGAGATATGG
M R P E I W

Frame 3

BglII BsiWI

951 ATTCGCCCAAG AGTTGCGGCG TATTGGGACAC GAGTTTAACG CTAAGATCTC
I A Q E L R R I G D E F N A K I S

Frame 3

XmaI

SmaI

1001 GTACCGTGTC CCGGCCGCGG CGCTCATCTTG CGGAGGTGAA GACGCCAAAA
Y A S R G G S S G G E D A K N

Frame 3

1051 ACATAAAAGAA AGGCCCGGCG CCATTCTATC CGCTGGAAGA TGGAACGGCT
I A Q E L R R I G D E F N A K I S

Frame 3

1101 GGAGAGCAAC TGCAAAAGGC TAGAAAGAGA TACGCCCTTG TCTCTGGGAAC
G E Q L H K A M K R Y A L V P G T

Frame 3

1151 AATTACTTTT ATACATGACAT ATATCGAGGT GGACATCTG TACGCTGAGT
I A F T D A H I E V D I T Y A E Y

Frame 3

1201 ACTTCCGAAAT GTCCGGTCGC TTGGGAGAAG CTATGAAACG ATATGGCGCT
F E M S V R L A E A M K R Y G L

Frame 3

1251 AATACAAATCT ACAGAATGCA CTATGCAGA ATATGGGCGT CATTGCAGAG
N T N H R I V V C S E N S L Q F F

Frame 3

1301 TAGCCGGCTG TGCCGGCCTG TATTTATCGG ATACACGTACG CATTGCAGAG
M P V L G A L F I G V A V A P N

Frame 3

1351 AGCAGATTTTA TAATGAACGT GAATTGCGAT CAGTATGCG TATTTCTCTC
D I Y N E R E L L N S M G I S Q

Frame 3

1401 CCTACCGTGCC GTTACGGTCTT AAAAAAGGG TGCCAAAAAA TTTGGAACGT
P T V V F V S K K G L Q K I L N V

Frame 3

1451 GCAGAAAAAG TCCCAATCTA TCCAAAATAT TATTCATG GATTTTCATA
Q K K L P I I Q K I I I M D S K T

Frame 3

1501 CGGATTACCA GGGATTTCA TGCCGACTAC CATTGTCGAT CTTCTCATCA
D Y Q G F Q S M Y T F V T S H L

Frame 3

1551 CCTCCCCGTTT TTAATGGAAT CGATTGATCG CTCAGCTTCTT TCGATAGGGA
P P G F N E Y D F V P E S F D R D

Frame 3

1601 CAAGACAATT GCACTGATCA TGAACCTTCG TGGATCTACT GTGCTGCTTA
K T I A L I M N S S G T G L P K

Frame 3

1651 AGGTTGTCGC TCCGGCTCCTG AGACTGCTG GCCTGAGATT TCTGCGATGCC
G V A L P H R T A C V R F S H A

Frame 3
AGAGATCCTA TTTTGGCAA TCAAATCATT CCGGATACTG CGATTTTAAG
R D P I F G N Q I I P D T A I L S Frame 3

TGTTGTTCCA TTCCATCACG GTTTTGAAT GTTTACTACA CTCGGATATT
V V P F H H G F G M F T T L G Y L Frame 3

TGATATGTGG ATTTCGAGTG GTCTTAATGT ATAGATTTGA AGAAGAGCTG
I C G F R V V L M Y R F E E E L Frame 3

TTTCTGAGGA GCCTTCAGGA TTACAGATTT CAAAGTGCC TGCTGGTGCC
F L R S L Q D Y K I Q S A L L V P Frame 3

AACCCTATTC TCCTTCTTCG CAAAGAGCAC TCTGATTGAC AAATACGATT
T L F S F F A K S T L I D K Y D L Frame 3

TATCTAATTTC ACACAAGATT GCTTCTGGTG GCCGCTCCCCT CTCTAAGGAA
S N L H E I A S G G A P L S K E Frame 3

GTCGGGGAAAG CGGTTGCAA GAGGTTCAT CTGCCAGGTA TCAGGAAGGG
V G E A V A K R F H L P G I R Q G Frame 3

ATATGGGCTC ACTGAGACTA CATCAGCTAT TCTGATTACA CCCGAGGGGG
Y G L T E T T S A I L I T P E G D Frame 3

ATGATAAACCC GGGCGCGGTC GGTAAAAGTG TTCCATTGGT TGAAGCCAGG
D K P G A V G K V V P F E A K Frame 3

GTTGATGAGCT TGGATAACGG GAAACCGCTG GCAGGTTAATC AAAGAGGGGA
V V D L D T G K T L G V N Q R G E Frame 3

ACTGTTGATG AGAGGTCTTA TGATTATGTC CGGTTATGTA AACATCCGG
L C V R G F M I M S G Y V N N P E Frame 3

AAGCGACCAA CGCCTTGATT GACAGGATG GATGATAACA TTCTGATAGC
A T N A L I D K D G * Frame 3
1701 GATCCTATTT TTGGCAATCA AATCATTCGG GATACTGCGA TTTTAAGTGT
   D P I F G N Q I I P D T A I L S V Frame 3
1751 TGGTCCATTC CATCACGGTT TTGGGAATGTT TACTACACTC GGATATTTGA
   V P F H H G F G M F T T L G Y L I Frame 3
1801 TATGTCGATT TCGAGTCGTC TTATGTAATA GATTGAAGA AGAGCTGTGTT
   C G F R V V L M Y R F E E E L F Frame 3
1851 CTGAGGAGCC TTCAGGATTA CAAGATTCAA AGTGCGCTGC TTGTGCCAAC
   L R S L O D Y K I Q S A L L V P T Frame 3
1901 CCTATTCTCC TTCTTCGCCA AAAGCCTCTC GATTGACAAA TACGATTAT
   L F S F F A K S T L I D K Y D L S Frame 3
1951 CTAATTTTCA CGAAATTGCT TCTGGTGCGG CTCCCCCTCTC TAAGGAAGTC
   N L H E I A S G G A P L S K E V Frame 3
2001 GGGGAAGCGG TTGCGCAAGAG GTTCCATCTG CCAGGTATCA GGCAAGGATA
   G E A V A K R F H L P G I R Q G Y Frame 3
2051 TGGGCTCACT GAGACTACAT CAGCTATTCT GATTACACCC GAGGGGGATG
   G L T E T T S A I L I T P E G D D Frame 3
2101 ATAAACCGGG CGGGTGCGGT AAAGTTGTTG CATTTTTTGA AGCGAAGGTT
   K P G A V G K V V P F F E A K V Frame 3
2151 GTGGATCTGG ATACCGGGAA AACGCTGGGC GTTAATCAAA GAGGCGAACT
   V D L D T G K T L G V N Q R G E L Frame 3
2201 GTGTGTTGAGA GGTCTATATGA TTATGGCAGG TTATGCAAC AATCCGGAAG
   C V R G P M I M S G Y V N N P E A Frame 3
2251 CGACCAACGC CTTGATTGAC AAGGATGGAT GATAACATTG TGATTAGCTAG
   T N A L I D K D G * Frame 3
BIK-NFluc:

KpnI

<table>
<thead>
<tr>
<th>HindIII</th>
<th>BamHI</th>
</tr>
</thead>
<tbody>
<tr>
<td>901</td>
<td>GTTAAGCTTG GTACCGAGCT CCGATCCCCA CCATGGAGGG CAGTGACGCA</td>
</tr>
<tr>
<td></td>
<td>MEGSDA</td>
</tr>
</tbody>
</table>

BsiWI

<table>
<thead>
<tr>
<th>XmaI</th>
<th>SmaI</th>
</tr>
</thead>
<tbody>
<tr>
<td>951</td>
<td>TTGGCCCTGC GGCTGGCCTG CATCGGGGAC GAGATGGACG TGGCGTACGC</td>
</tr>
<tr>
<td></td>
<td>LALRLACIGDEMDVAYA</td>
</tr>
</tbody>
</table>

Frame 3
<table>
<thead>
<tr>
<th>Position</th>
<th>Sequence</th>
<th>Frame</th>
</tr>
</thead>
<tbody>
<tr>
<td>1701</td>
<td>CCTATTTTTG GCAATCAAAT CATTCGGAT ACTGCAGTTTT TAAGTGTGTTGT</td>
<td>Frame 3</td>
</tr>
<tr>
<td>1751</td>
<td>TCCATTTCCAT CACGTTTTTG GAATGTTTAC TACACTCGGA TATTGATAT</td>
<td>Frame 3</td>
</tr>
<tr>
<td>1801</td>
<td>GTGGATTTTCG AGTCTGTCTTA ATGTATAGAT TTGAAGAAGA GCTGTTTCTG</td>
<td>Frame 3</td>
</tr>
<tr>
<td>1851</td>
<td>AGGAGCCTTT AGGATTCAA GATTCAAAGT GCGCTGCTTG TGCCAACCCT</td>
<td>Frame 3</td>
</tr>
<tr>
<td>1901</td>
<td>ATTCTCTTTTC TCGCCAAAA GCACTCTGAT TGACAATAC GATTTATCTA</td>
<td>Frame 3</td>
</tr>
<tr>
<td>1951</td>
<td>ATTTACACGA AATTGCTTCT GTGGCGCGTC CCCTCTCTAA GGAAGTCGGG</td>
<td>Frame 3</td>
</tr>
<tr>
<td>2001</td>
<td>GAAGCGGTTT GCAAGAGGTT CCATCTGCCA GGTATCGAG ACCATATGGG</td>
<td>Frame 3</td>
</tr>
<tr>
<td>2051</td>
<td>GCTACTGAG ACTACATCAG CTATTCTGAT TACACCGAG GGGATGATA</td>
<td>Frame 3</td>
</tr>
<tr>
<td>2101</td>
<td>AAGCGGGCCG GGTGCTTCAA TTTTCTGAT TAAGTGTGTTG</td>
<td>Frame 3</td>
</tr>
<tr>
<td>2151</td>
<td>GATCTGGATA CCGGGAAAAAC GTGGGGCTTG AATCAAAAGG GCCAAGCTGTG</td>
<td>Frame 3</td>
</tr>
<tr>
<td>2201</td>
<td>GTGAGAGGTT CCTATGATTA TGTCGGTCTTA TGATACAAAT CGCGGAAGCGA</td>
<td>Frame 3</td>
</tr>
<tr>
<td>2251</td>
<td>CCAACGCCTT GATTGACAAG GATGGATGAT AACATCTCGG TAGCTAGGTA</td>
<td>Frame 3</td>
</tr>
</tbody>
</table>
### p53-NFluc:

<table>
<thead>
<tr>
<th>Position</th>
<th>Sequence</th>
<th>Frame</th>
<th>Protein Frame</th>
</tr>
</thead>
<tbody>
<tr>
<td>901</td>
<td>TAAGCTTTGGT</td>
<td>Frame 1</td>
<td>MDFSPVEP</td>
</tr>
<tr>
<td>951</td>
<td>CCGAAATCAC</td>
<td>Frame 1</td>
<td>PLSEQFSDLWKLPEN</td>
</tr>
<tr>
<td>1001</td>
<td>ACAACGTTCCTG</td>
<td>Frame 1</td>
<td>NVLSPLKPISYARSG</td>
</tr>
<tr>
<td>1051</td>
<td>GCCATCTGGAG</td>
<td>Frame 1</td>
<td>GSFLAGGEDAKNIKGP</td>
</tr>
<tr>
<td>1101</td>
<td>CTATGAAAGAG</td>
<td>Frame 1</td>
<td>FTCVGGLPNTNTNHIV</td>
</tr>
<tr>
<td>1151</td>
<td>TTATTCGAGG</td>
<td>Frame 1</td>
<td>LICFIAVAPANDYNER</td>
</tr>
<tr>
<td>1201</td>
<td>CATATCGAGG</td>
<td>Frame 1</td>
<td>HIEVDITYAYEYFMSVR</td>
</tr>
<tr>
<td>1251</td>
<td>GTTGGGCAGAA</td>
<td>Frame 1</td>
<td>LREAAMKRYGLNTNHIV</td>
</tr>
<tr>
<td>1301</td>
<td>TCGATGCAGC</td>
<td>Frame 1</td>
<td>VCSENSLQFPMPVLAG</td>
</tr>
<tr>
<td>1351</td>
<td>TTATTTATCG</td>
<td>Frame 1</td>
<td>LFIGVAVAPANDYNER</td>
</tr>
<tr>
<td>1401</td>
<td>TGAATTGTCTC</td>
<td>Frame 1</td>
<td>ELLNSMGISQPTTVFVS</td>
</tr>
<tr>
<td>1451</td>
<td>CCAAAAGGG</td>
<td>Frame 1</td>
<td>KKGILQKILNVQKKLPI</td>
</tr>
<tr>
<td>1501</td>
<td>ATCCAAAAAA</td>
<td>Frame 1</td>
<td>IQKIIIMDSKTQGYFQ</td>
</tr>
<tr>
<td>1551</td>
<td>GTGCATGTAC</td>
<td>Frame 1</td>
<td>SMTFTVTSHTLPPPFGENVY</td>
</tr>
<tr>
<td>1601</td>
<td>ACGATTTTGTG</td>
<td>Frame 1</td>
<td>DFVSEPFRDKTIALI</td>
</tr>
<tr>
<td>1651</td>
<td>ATGAACTCTCTGAGCTTCGCTTCTCTGCTCATTTCCTGACCT</td>
<td>Frame 1</td>
<td>MNSSTGGLPKGVALPH</td>
</tr>
<tr>
<td>1701</td>
<td>TAGAACTGCC</td>
<td>Frame 1</td>
<td>RTACVRFSHARDPIFGN</td>
</tr>
</tbody>
</table>
1751 ATCAAATCAT TCCGGATACT GCGATTTTAA GTGTTGTTCC ATTCACATCAC
   Q I I P D T A I L S V V P F H H Frame 1
1801 GGTTTTTGAA TGTTTACTAC ACTCGGATAT TTGATATGTG GATTTTCGAGT
   G F G M F T T L G Y L I C G F R V Frame 1
1851 CGTCTTTAATG TATAGATTG TAAAGAGAGCT GTTTCTGAGG AGCCTTCAGG
   V L M Y R F E E E L F L R S L Q D Frame 1
1901 ATTACAAAGAT TCAAAGTGCG CTGCTGGTGC CAACCCTAATT CTCCTTCTTC
   Y V K I Q S A L L V P T L F S F F Frame 1
1951 GCCAAAAGCA CTCTGATTGA CAAATACGAT TTATCTAATT TACACGAAAT
   A K S T L I D K Y D L S N L H E I Frame 1
2001 TGCTTCTGTT GGCCTCCTCC TCTCAAGGAA AGTCGGGGAA GCGGTTGCCA
   A S G G A P L S K E V G E A V A K Frame 1
2051 AGAGGTTCCA TCTGCCAAGGT ATCAGGCAAG GATATGGGCT CACTGAGACT
   R F H L P G I R Q G Y G L T E T Frame 1
2101 ACATCAAGCTA TTCTGATTAC ACCCGAGGGG GATGATAAAC CGGGCGCGGT
   T S A I A L T P E G D D K P G A V Frame 1
2151 CGGTAAAGTT GTTCCATTTT TTGAAGCGAA GGTTGTGGAT CTGGATACCG
   G K V V P F F E A K V V D L D T G Frame 1
2201 GGAACACGCT GGGCGTATAT CAAAGAGCGG AACTGTGTGT GAGAGTCTCT
   K T L G V N Q R G E L C V R G P Frame 1
2251 ATGATTATGT CCGTTTATGT AAACAAATCCG GAAGCGACCA ACGCCTTGAT
   M I M S G Y V N N P E A T N A L I Frame 1
2301 TGACAAGGAT GGATGATTAC ATTTGCTAGG TTAGTAAATG CATAATCTAG
   D K D G * Frame 1
The above vector map represents a series of receptor containing vectors utilized to study the interactions between various helix and receptor proteins.
CFluc-hDM2:

1751 GGAATTGCC CTTACCCACC ATGTCCGGTT ATGTAACAA TCCGGAAGCG
  M S G Y V N N P E A Frame 1
1801 ACCAACGCTT TGATTGACAA GGATGATGGA CTACATCTCTG GAGACATAGC
  T N A L I D G K W L H S G D I A Frame 1
1851 TTACTGGGAC GAAAGACGAAAC ACTTCTCTCAT CGGACCCGCCA CTGAAAGTTC
  Y W D E D E H F F I V D R L K S L Frame 1
1901 TGATTAAGTAA CAAAGGGCATC CAGGTGGGTCCT CAGCCTGAATTC GGAATCCATC
  I K Y K G Y Q V A P A E L E S I Frame 1
1951 TTGCTCCAAC ACCCAAACAT CTTGAGCCGCA GTGTTGCGCA GCTTTCGCGCA
  L L Q H P N I F D A G V A G L P D Frame 1
2001 CGATGACGCC GGTGAACTTC CGGCCGCCGT TGTTGTTTTG GAGCAGGCAA
  D D A G E L P A A V V V L E H K Frame 1
2051 AGACGATGAC GAAAAAAAGATGACGGATT AGCTGCGCCG AGCTAGTAACA
  T M T E K I V D Y V A S Q V T Frame 1
2101 ACCGCCAAAA AGTTGCGCGG AGGATGGGTGG TTTGTTGGACG AATACCGCAA
  T A K K L R G V V F V D E V P K Frame 1
2151 AGGTCTTACC GAAACAGCTG AACAGAAGAAA AATCAGAGAG ATCCTCATAA
  G L T G K L D A R K I R E I L I K Frame 1
2201 AGGCCAAGAA GGGCGGAAAG ATCGCCGTGGG GAGGTGGCTC ATCTGGCGGA
  A K K G G K I A V G G G S S G G Frame 1

XmaI

2251 GGTCAAGATCT CGTACGGTCC CGGGATGTGC AATACCAACA TGTTCTGCC
  G Q I S Y G P G M C N T N M S V P Frame 1
2301 TACTGATGTT GGTACGAAACA CCTAACAGAT TCCAGCTTCCG GAACAGAGA
  T D G A V T T S Q I P A S E Q E T Frame 1
2351 CCCTGTTAGG ACCAAAGCGA TTGGCTTTTAG AGTTATTTAGA GTCTGTGTTGG
  L V R P K F L L K L K S V G Frame 1
2401 GCCACAAAAAG ACACTATACACT TATGAAAGAG GTTCTTTTTTT ATCTGGCCAA
  A Q K D T Y T M K E V L F Y L G Q Frame 1
2451 GTATATTATG ACTAAAGCACT TATATGATGGA GAAGCAACAA CATATTGTAT
  Y I M T K R Y D E K Q Q H I V Y Frame 1
2501 ATGGGCTGAA TTGATCTTCTA GAGAGATTGGTT TGCGCTCTCCG AAGCTTACC
  C S N D L L G D L F G V P S F S Frame 1
2551 GTGAAAGAGC ACAGGAAAAAT ATATACCAATG ATCTACAGGA ACTTGGTATG
  V K E H R K I Y T M I Y R N L V V Frame 1
AGTCAATCAG CAGGAATCAT CGGACTCAGG TACATCTGTG AGTGAGAACA
V N Q Q E S S D S G T S V S E N R Frame 1

GGTGTCACCT TGAAGGTGGG AGTGATCAA A AGGACCTTGT ACAAGAGCTT
C H L E G G S D Q K D L V Q E L Frame 1

CAGGAAGAGA AACCTTCATC TTCACATTTG GTTTCTAGAC CATCTACCTC
Q E E K P S S S H L V S R P S T S Frame 1

ATCTAGAGGG AGAGCAATTA GTGAGACAGA AGAAAATTCA GATGAATTAT
S R R R A I S E T E E N S D E L S Frame 1

EcoRV NotI

CTGTTGAACG ACAAGAGAAA CGCTAAAGAG ATATCCAGCA CAGTGGCGGC
G E R Q R K R * Frame 1
CFluc-hDM4:

1751 GGAATTGCC TTACACCACC ATGTCCGTTT ATGTAAACAA TCCGGAAGCG
  Frame 1
  M S G Y V N N P E A

1801 ACCAACGCTT TGATGGACAA GGATGGATGG CTACATTCTG GAGACATAGC
  Frame 1
  T N A L I D K D G W L H S G D I A

1851 TTAATGGAAC GAAGACGCAACTTCTTTCAT GCCTGACGCGC CAGTCAATTC
  Frame 1
  Y W D E D E H F F I V D R L K S L

1901 TGATTAAGTA CAAGGCTAT CAGTTCGTCCT CCGTCAATT GGAATCCATC
  Frame 1
  I K Y K G Y Q V A P E L S I

1951 TTGCTCCAAC ACCCAACAT CTTCGAGCGCA GGTTGTCGCCG GTCTTCCCGA
  Frame 1
  L L Q H P N I F D A G V A G L P D

2001 CGATGACGCC GGTGAATTTCC CGGCCGCCT CGTTGATTGG GAGGCAGGAA
  Frame 1
  D D A G E L P A A V V L E H G K

2051 AGACGATGAC GAAGAAAGAG ATCGTGGATT AGTCGCGCAAG TCAAGTAAACA
  Frame 1
  T M T E K I V D Y V A S Q V T

2101 ACCGCCGAAAA AGTTGGCCCGG AGGAAGTGGTGG TTTGTGCGGAG AGATCCGAA
  Frame 1
  T A K K L R G V V F V D E V P K

2151 AGGTCTTACC GGAAGACTCG AGCGCAAGAAA AATCAGAGAG ATCCTCATAA
  Frame 1
  G L T G K L D A R K I R E I L K

2201 AGGCCAAGAAA GGGCGGAAAG ATCGCCGTGG GAGGTGGCTC ATCTGGCGGA
  Frame 1
  A K K G G K I A V G G S S G

 XmaI

BsiWI  SmaI

2251 GGTCACTTCT CGTACGGTCC CGGGATGACA TCATTTCCTA CCTCTGCTCA
  Frame 1
  G Q I S Y G P G M T S F S T S A Q

2301 GTGTCTAACCA TCTGACAGTG CTTGCAGGAT CTCTCCTGGA CAAATCAATT
  Frame 1
  C S T S D S A C R I S P G Q I N Q

2351 AGGTACGACC AAAACTCGCC CGTTTGGAAGA TTTTGATAGC AGCAGATGGC
  Frame 1
  V R P K L P L K I L H A A G A

2401 CAAGGGAAAGA TGGTCAGTCTTAAGAGGGTC ATGCACTATT TAGGTGATGA
  Frame 1
  Q G E M F T V K E V M H Y L G Q Y

 NdeI

2451 CATAAAGGTGTG AAGCAACTTT ATGATCATGCA GGAGAAGCAG ATGGATATATT
  Frame 1
  I M V K Q L Y D O Q E Q H M V Y C

2501 GTGGTGGAGA TCTGACATGG GAACTACTGG GACGTGAGG GAGTCGTGGT
  Frame 1
  G G D J L G E L G R Q S F S V

2551 AAAGACCCAA GCCCTCTCCTA TGATATGCTA AGAAAGAATC TTGTCACTTT
  Frame 1
  K D P S P L Y D M L R K N L V T L
2601 AGCCACTGCT ACTACAGATG CTGCTCAGAC TCTCGCTCTC GCACAGGATC
  ATATTDA AQTLALAQDH Frame 1

2651 ACAGTATGGA TATTCAAGT CAAGACCAAC TGAAGCAAAAG TGCAAGAGAA
  SMDIPSQDLKQS AEE Frame 1

2701 AGTTCCACTT CCAGAAAAAG AACTACAGAA GACGATATCC CCACACTGCC
  SSTSRTKRE DDIPTLP Frame 1

2751 TACCTCAGAG CATAAATGCA TACATTCTAG AGAAGATGAA GACTTAATTG
  TSEHKCISHSEREDDLIE Frame 1

NotI

2801 AAAATTTAGC CCAAGATGAA ACATCTAGGT AATGAGCGGC CGCTCGAGTC
  NLADETSRR * Frame 1
CFluc-Bcl-2:

1751 AATGCCCCTT ACACCACCAT GTCCGTTAT GTAAACAATC CGGAAGCGAC
   M S G Y V N N P E A T Frame 2

1801 CAACGCCTTG ATTGACAAAG ATGGATGGCT ACATTCTGGA GACATGCTT
   N A L I D K D G W L H S G D I A Y Frame 2

1851 ACTGGGACGGA AGACCGAAAC TTCTTCATCG TTGACCGCCT GAGTCTCTG
   W D E D E H F F I V D R L K S L Frame 2

1901 ATTAAGTACA AAGGCTATCA GGTGCGTCCC GCTGAATTGG AATCCATCTT
   I K Y G Y Q V A P A E L S I L Frame 2

1951 GCTCCAAACAC CCCAACATCT TCGACGCAGG TGTCGCGAGT CTTCGGCGACG
   L Q H P N I F D A G V A G L P D D Frame 2

2001 ATGACGCCCGG TGACTCTCCC GCCGCCGGTTG TTGTTTTGGA GCACGGAAG
   D A G E L P A A V V V L E H G K Frame 2

2051 ACGATGACGG AAAAAGAGAT CGTGAGATTAC GTGCGCGTTC GCAGAAACG
   T M T E K E I V D Y V A S Q V T T Frame 2

2101 CCCGAAAAAG TTCCGCGGGAG GAGTTGTGTTT TGTTGACGAA GTACGGAAG
   A K K L R G G V V F V D E V P K G Frame 2

2151 GTCTTACCGG AAAACTCGAC GCAAGAAAAA TCAGAGAGAT CCTCATAAAG
   L T G K L D A R I K R E I L K Frame 2

2201 GCCAAAGAAG GGCGAAGAGAT CGGCGGGGGA GGTGCGTCTC GTGCGGGAG
   A K K G G K I A V G G G S G G Frame 2

BsiWI

2251 TCGATCTCG TACGGATATGG CGCGCGTCTGG GAGAAACGGGG TACGATAACC
   Q I S Y G M A H A G R T G Y D N R Frame 2

2301 GGGAGATAGT GATGGAATAG ATCCATTATC AGCTGCGACA GAGGGGCTAC
   E I V M K Y I H Y K L S Q R G Y Frame 2

2351 AGGTGGGGATG CGGGAGATGT GGGCGCCGCG CCCCCCGGGG CGCCCCCGCG
   E W D A G D V G A A P P G A A P A Frame 2

2401 ACCGGGCATC TTCTCCCTCC AGCGCGCGGGCA CACGCCCACT CACGCCCGCAT
   P G I F S S Q P G H T P H P A A S Frame 2

2451 CCCGGAACCC GGTCGCGCGG ACCTCGCCGC TGCAGACCACGC CGCTCGCACC
   R D P V A R T S P L Q T P A A P Frame 2

2501 GGGCGCGCCG CGGGCGCTGC GCCTGCACCGG GTCGACCCGT TGTTCCACCT
   G A A A G P A L S P V P P V V H L Frame 2

2551 GACCGCTCGC CGGCGCGCGG CGACCTTCTC CGCGCGCTAC CGCGCGCGACT
   T L R Q A G D D F S R R Y R R D F Frame 2

2601 TCGCCGAGAT GTCCAGGGAG CTGCACCTGA CGCCCTTCGC CGGCCCGGGGAG
A E M S S Q L H L T P F T A R G  Frame 2

2651 CGCTTTGCCA CGGTGTTGGA GGAGCTCTTC AGGGACGGGG TGAACTGGGG
    R F A T V E E L F R D G V N W G  Frame 2

2701 GAGGATTGTG GCCTTCTTTG AGTTCGTTGG GGTCATGTGT GTGGAGAGCG
    R I V A F F E F G G V M C V E S V  Frame 2

2751 TCAACCGGGA GATGTGCCCC CTGGTGACA ACATCGCCCT GTGGATGACT
    N R E M S P L V D N I A L W M T  Frame 2

2801 GAGTACTCGA ACCGGCACCT GCACACCTGG ATCCAGGATA ACGGAGGTG
    E Y L N R H L H T W I Q D N G G W  Frame 2

2851 GGATGGCTTT GTGGAACTGT ACGGCCCAAG CATGGGCTCT GTTTTTGATT
    D A F V E L Y G P S M R P L F D F  Frame 2

2901 TCTCCTGGGT GTCTCTGAAAG ACTCTGCTCA GTTTGGCCCT GGTGGGAGCT
    S W L S L K T L L S L A L V G A  Frame 2

EcoRV

2951 TGCATACCCC TGGTGCTTA TCTGGGCCAC AAGTGATAAG ATATCCAGCA
    C I T L G A Y L G H K *  Frame 2
CFluc-Bcl-Xₐ:

1751 AATTGCCCTT ACACCACCAT GTCCCGTTAT GATAACAATC CGGAAACGAC
   M S G Y V N N P E A T Frame 2
1801 CAACGCCTTG ATGACAAAGG ATGGATGGCT ACATTCTGGA GACATAGCTT
   N A L I D K D G W L H S G D I A Y Frame 2
1851 ACTGGGACGA AGACGAAACAC TTCCTCATCG TTGACCGCCT GAAGTCTCTG
   W D E D E H F F I V D R L K S Frame 2
1901 ATTAAGTACA AAGGCTATCA GTGGGCTCCC GCTGAATTGG AATCCATCTT
   I K Y K G Y Q V A P A E L E S I L Frame 2
1951 GTTCAACAC CCCAACATCT TCGACGCAAG TGGCAGACTG CTTCCCGACG
   L Q H P N I F D A G V A G L P D D Frame 2
2001 ATGACGCCGG TGAACCTCCC GCGGCCGTTG TTGCTTTGGA GCACGGAAAG
   D A G E L P A A V V V L E H G K Frame 2
2051 ACGAGTGACGG AAAAAGAGAT CGTGGATTAC GTCGCCAGTC AAGTAACAAC
   T M T E K I V D Y V A S Q V T T Frame 2
2101 CGCGAAAAAG TTGCCGCGGAG GAGTTGTGTT TGTGGACGAA GTCCGGAAAG
   A K K L R G G V V F V D E V P K G Frame 2
2151 GTCTTACCGG AAAACTCGAC GTAAGAAAAA TCAGAGAGAT CCTCATAAAG
   L T G K L D A R K I R E I L K Frame 2
2201 GCCAAGAGGG GCGGAAAGAT GCGCCGGTGG GGTGGCTCAT CTGGCGGAGG
   A K K G G K I A V G G S S G G G Frame 2

BsiWI
2251 TCAGATCTCG TACGGTCCCG GGATGTCTCA GAGCAACCGG GAGCTGGTGG
   Q I S Y G P G M S Q N R E L V V Frame 2
2301 TTGACTTTTCT CTCCTCACAAG CTCTTCGAAA AAGGATACAG CTGGAGTTCG
   D F L S Y K L S Q K G Y S W S Q Frame 2
2351 TTTAGTGATG TGGAAGAGAA CAGGACTGAG GCCCAGAAAAG GTGAGCTGAATC
   F S D V E E N R T E A P E G T E S Frame 2
2401 GGAAGATGGAG ACCCCCATGTT CCATCAATGG CAACCCATCC TGGCACCCTGG
   E M E T P S A I N G N P S W H L A Frame 2
2451 CAGACACGCC CGCGGTTGAA GGAGCACCAGT GCCACACAGC CAGTTTGGAT
   D S P A V N G A T G H S S S L D Frame 2
2501 GGCAGGGAGG TGATCCCCAT GGCAGCAGTA AAGCAAGCCG TGAGGAGGCC
   A R E V I P M A A V K O A L R E A Frame 2
2551 AGGCGACGG AGTGAAGTGC GTTACCCGGCG GCCATTCAGT GCCCTGACAT
   G D E F E L R Y R R A F S D L T S Frame 2
2601 CCCACGCTCCA CATCAACCCCA GGAGACAGAT ATCAGAGCTT TGACACAGTTA
   Q L H I T P G T A Y Q S F E Q V Frame 2
2651  GTGAATGAAC TCTTCCGGGA TGGGGTA AAC TGGGGTCGCA TTGTGGCCTT
       V N E L F R D G V N W G R I V A F  Frame 2

2701  TTTCTCCTTC GGCGGGC AC TGTGCGTGGG AAGCGTAGAC AAGGAGATGC
       F S F G G A L C V E S V D K E M Q  Frame 2

2751  AGGTATTGGT GAGTCGGGTC GCAGCTTGGGA TGGCCACTTA CCTGAATGAC
       V L V S R I A A W M A T Y L N D  Frame 2

2801  CACCTAGAGC CTTGGATCCA GGAGAACG GC GGCTGGGATA CTTTTGTGGA
       H L E P W I Q E N G G W D T F V E  Frame 2

2851  ACTCTATGGG AA CAATGCAG CAGCCGAGAG CCGAAAGGGC CAGGAACGCT
       L Y G N N A A A E S R K G Q E R F  Frame 2

2901  TCAACCGCTG GTTCTCGACG GGCATGACTG TGGC CGGC GTGTTCTGCTG
       N R W F L T G M T V A G V V L L  Frame 2

EcoRV NotI

2951  GGCTCAATCT TCA GTCGGAA ATGATAAGAT ATCCACACCA GTGGCGGCCG
       G S I F S R K *  Frame 2
CFluc-Bcl-w:

1751 AATTCGCCCTT ACACCACCAT GTCCCGGTAT GTAACAATC CGGAAGCGAC
        M S G Y V N N P E A T Frame 2

1801 CAACGCCTTG ATTGACAAGG ATGGATGGCT ACATTCTGGA GACATAGCTT
        N A L I D K G W L H S G D I A Y Frame 2

1851 ACTGGGACGA AGACGAACAC TTCTTCATCG TTGACCACCT AGAGTCTCTG
        W D E D E H F I V D R L K S Frame 2

1901 ATTAAGTACA AAGGCTATCA GTTGGCTCCC GCTGAATTGG AATCCATCTT
        I K Y K G Y Q V A P A E L E S I L Frame 2

1951 GCTCCAACAC CCCAACATCT TCGACGCAAG GTCGGCAAGT CTTCCCGACG
        L Q H P N I F D A G V A G L P D D Frame 2

2001 ATGACGCCGG TGAACCTCCC CGCGCGAAGT TTGTAAAAAA GCACGGAAAAG
        D A G E L P A A V V V L E H G K Frame 2

2051 ACAGTACGGG AAAAAAGAGT CGTGGATTAC GTCGCCAGTT CTTCCCGACG
        T M T E K I V D Y V A S Q V T T Frame 2

2101 CGCGAAAAAG TTGCGCGGAG GAGTTGTGTT TTGGGCAAGA GTACGGAAAAG
        A K K L R G G V V F V D E V P K G Frame 2

2151 GTCTTACCAGG AAAACTCGAC GCAAGAAAAA TCAGAGATG GCTCATAAAAG
        L T G K L D A R K I R E I L I K Frame 2

2201 GCCAAGAGGG CGCGAAGAGT CGCGTGCGGA GTTGACCTCA CTGCCGCGAGG
        A K K G G K I A V G G G S S G G Frame 2

      BsiWI

2251 TCAGATCCTCG TACGGGCTCCG CGATGACCGAC CCCAGGCTCG GCCCGAAGACA
        Q I S Y G P M A T P A S A P D T Frame 2

2301 CACGGGCTCCT CGGCGGCAAGG TTGGTAGGTT ATAAAGCTGA GCAGAAGGGTT
        R A L V A D F V G Y K L R Q K G Frame 2

2351 TATGCTCTGT GAGCTGCGCC CGGAGGAGGC CCAGCACCTG ACCGCTGCA
        Y V C G A G F G E G P A A D P L H Frame 2

2401 CCAAGCCATG CGCGAAGCTG GAGATGAGTT CGAGACCCGC TTCCGGCGCA
        Q A M R A A G D E F E T R F R T Frame 2

2451 CTTTCATGTA TCTGGCGGCT CAGCTGCTAG TGACCACAGG CTCAGCCCCA
        F S D L A A Q L H V T P G S A Q Frame 2

2501 CAACGCTTCA CCCAGCTTCG CGATGAACTT TTGGAGGGA GCACCCACTG
        Q R F T Q V S D E L F Q G G P N W Frame 2

2551 GGCGCGCCTT GTAGCTTCTG TTGGCTTTGG GGTGACTGT TGCTGCTGAGA
        G R L V A F F V F G A A L C A E S Frame 2

2601 GTGTCAACAA GGAGATGAAA CCACGTGGTTG GACAAGTGCA GAGTGGGATG
        V N K E M E P L V G Q V Q E W M Frame 2
BamHI

2651 GTGGGCTACC TGGAGACGCG GCTGGCTGAC TGGATCCACA GCAGTGGGGG

V A Y L E T R L A D W I H S S G G Frame 2

2701 CTGGGCGGAG TTCACAGCTC TATAACGGGA CGGGGCCCTG GAGGAGGC GC

W A E F T A L Y G D G A L E E A R Frame 2

2751 GGCGTCTGCG GGAGGGGAAC TGGGCATCAG TGAGGACAGT GCTGACGGGG

R L R E G N W A S V R T V L T G Frame 2

2801 GCCGTTGGCAC TGGGGGCCCT GGTAACTGTA GGGGCCTTTT TTGCTAGCAA

A V A L G A L V T V G A F F A S K Frame 2

EcoRV NotI

2851 GTGATAAGAT ATCCAGCACA GTGGCGGCGG CTCGAGTCTA GAGGGCCGC

* Frame 2
CFluc-BFL:

1751 GGAATTGCC CTACACCACC ATGTCGGGTT ATGTAACCA TCCGGAAGCG  Frame 1
     MSG Y VNNP EA

1801 ACCAACGCCT TGATTTGACAA GGAATGATGG TTACATCTCTG GAGACATAGC  Frame 1
     TNALIDKDGWLHSGDIA

1851 TTACTGGGAC AAAGACAGCA ACTTTCTTCAT CGTTGACGGCT CTGAAAGTCTC  Frame 1
     YWDEDEHFFIVDRLKL

1901 TGATTAAGTA CAAGGGCTAT CAGGTGCCCT CCGGTGAATTT GGAATCATC  Frame 1
     IKYKGYQVAPAELESI

1951 TTGCTCCAAC ACCCAACAT TTTCGACGCA GGTGTCGACG CTCTTCCCAGA  Frame 1
     LLQHPNIFDAVGALP

2001 CGATGACGCC GGTGAACCTTC CGCCGAGCGG TGTTGTTTGT GAGCAACGGAA  Frame 1
     DDA GEIAPAVVLEVHKG

2051 AGACGATGAC GGAaaaaAGATGCTGGGAT TTACGCGCCAG TCAAGTAACA  Frame 1
     TMTEKIPvDYVASQVT

2101 ACCGCCGAAA AGTGGCGCCGG AGGAGTTGTC TTGTGGACG GGTACCCGAA  Frame 1
     TAKKLRGVVFPDEVPK

2151 AGGTCTTACC GGAAAATCGT ACGCAAGAAA AACGCAAGGG AATCCAGAGG  Frame 1
     GLTGKLDARKIREILIK

2201 AGGCCAAGAA GGGCGGAAGA ATGCGCGTGG GAGGTGGGCT ATCTTGGCGA  Frame 1
     AKKGGKIIVGGSSSG

XmaI

2251 GGTCAGATCT CGTACGGTCC CGGGATGACA GACTGTTAAT TTGGATATAT  Frame 1
     GQISYGPGMTDCFEYG

BsiWI

2301 TTACAGGCGTG GTTACCGGACT ATCTGCGATG CGTCCTACAG ATACCACAC  Frame 1
     YRLAQDYLQCVLQIPQP

SmaI

2351 CTGGATCAGG TCCATGGCGAA ACGTCGAAAG TGCCTACAAA TGTTGGCTTC  Frame 1
     GSPLSKTRVLOQNVAF

2401 TCAGTCCAAA AAGAAGTGGG AAAGAAACCTG AAGTCATGCT TGGAACATGT  Frame 1
     SVQKEVENKLNKSLD NV

PflMI

2451 TAATGGTTGTG TCGGTAGACG CTGCCGAGAC ACTATTCGAC CAAATGATGG  Frame 1
     N V V SVIDTAR LFNQVME

2501 AAAAGGAGTT TGAAGACGGC ATCATTAACG GGGGAAAGAT TGTAACACATA  Frame 1
     KEFEDGIINWGRIVTI

2551 TTTGCTATTTTG AAGGTATTCT CATCAAGAAA CTTCCTCAGAC AGCATAATTGC  Frame 1
     FAFFEGILIKKLRLRQQAIA
2601 CCGGATGTG GATACTATA AGGAGATTTC ATATTTTGGT GCGGAGTTCA
  P D V D T Y K E I S Y F V A E F I Frame 1
2651 TAATGAATAA CACAGGAGAA TGGATAAGGC AAAACGGAGG CTGGGAAAAT
  M N N T G E W I R Q N G G W E N Frame 1
2701 GGCTTTGTAA AGAAGTTTGA ACCTAAATCT GGCTGGATGA CTTTTCTAGA
  G F V K K F E P K S G W M T F L E Frame 1
2751 AGTTACAGGA AAGATCTGTG AAATGCTATC TCTCCTGAAG CAATACTGTT
  V T G K I C E M L S L L K Q Y C * Frame 1
pcDNA3.1-DHFR-NFuc
7321 bp
DHFR-NFluc:

KpnI

<table>
<thead>
<tr>
<th>HindIII</th>
<th>BamHI</th>
</tr>
</thead>
<tbody>
<tr>
<td>901</td>
<td>GTTAAGCTTG</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>951</td>
<td>GCGTTAGGCG</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>1001</td>
<td>CCTGCTCTGCC</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>1051</td>
<td>TGATTATGCG</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>1101</td>
<td>GCGAAAAATA</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>1151</td>
<td>GTGGGTGAAAG</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>1201</td>
<td>AAATCAGGTT</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>1251</td>
<td>GCGCAAAAAC</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| Frame 3 |

EcoRI

| BglII BsiWI |
| XmaI |

<table>
<thead>
<tr>
<th>SmaI</th>
</tr>
</thead>
</table>

| Frame 3 |
| 1301 | CCATTTCGCC | GATTACGAGC | CGGATGAACGT | GFAATCGTA | TTCAGCGAAT |
|      |         |         |         | H F P D Y E P D W E S V F S E F |
| 1351 | TCCAGCATGC | TGATGCGAC | AACTCTCACA | GCTATTGCTT | TGAGATTCG |
|      |         |         |         | H D A D A Q N S H S Y C F E I L |

| Frame 3 |

| Frame 3 |
| 1401 | GAGCGCGGAGA | AGATCTCGTA | CGGCCTCCCGG | GGGGTGGCTG | CATCTGGCGG |
|      |         |         |         | E R R K I S Y A S R G G S S G G |
| 1451 | AGGTGAAGAC | GCCAAAAACA | TAAAGAAAGG | CCCGGCGCCA | TTCTATCCGC |
|      |         |         |         | G E D A K N I K K G P A P F Y P L |
| 1501 | TGGAAGATGG | AACCCTGGA | GAGCAACTGC | ATAAAGCTAT | GAAGAGATA |
|      |         |         |         | E D G T A G E Q L H K A M K R Y |
| 1551 | GCCCTGTTC | CTGGAACAAAT | TGCTTTTACA | GATGCACATA | TCGAGATGGA |
|      |         |         |         | A L V P G T I A F T D A H I E V D |
| 1601 | CATCACTTAC | GCTGAGTCTC | TGAAAAATGTC | CGTTCGGTGG | GCAGAAGCTA |
|      |         |         |         | I T Y A E Y F E M S V R L A E A M |
| 1651 | TGAAACGATA | TGGGCTGAAT | ACAATACACA | GAATCGTCGT | ATGCAGTGA |
|      |         |         |         | K R Y G L N T N H R I V V C S E |

| Frame 3 |
1701  AACTCTCTTC AATTCCTTAT GCGGCTGTGT GGGCGGTTAT TTATCGGAGT
   N S L Q F F M P V L G A L F I G V Frame 3
1751  TGCAATTCGCG CCCGCAAGCG ACATTTATAT TAACAGTGAAC TTGCTCAACA
   A V A P A N D I Y N E R E L L N S Frame 3
1801  GTATGGGCGAT TTCCGACGCT ACCGTTGGGT TGCTTCCCAA AAAAGGGTTG
   M G I S Q P T V V F V S K K G L Frame 3
1851  CAAAATTTTT GAACGTGTCG AAAAGAGGTC ACCATCATCC AAAAATTAT
   Q K I N V K K L P I I Q K I I Frame 3
1901  TATCATGGAT TCTAAAACGG ATATTACGGG ATTTGAGTCG ATGTGACCAGT
   I M D S K T D Y Q F Q S M Y T F Frame 3
1951  TCGTCACATC TCATCTACCT CCCGTTTTTA ATGAATACGAT TTTGTTGCCA
   V T S H L P P G F N E Y D F V P Frame 3
2001  GAGTCTCTCG ATAGGGGAAA GACAATTGCA CTGATCATAG ACTCCTCTGG
   E S F D R D R I A I L M N S G Frame 3
2051  ATCTACTGCTT CTGCTAATAAG GTGTGCTGCT CTGCTCATAGA ACTGCCCTGCG
   S T G L P K G V A L P R T A C V Frame 3
2101  TGAGATTCTTC GATGCCCAA AATATCCATA TTGGAATATC AATACCCCGG
   R F S H A R D P F G N Q I I P Frame 3
2151  GATACGTGCGA TTTTAAGGTG TGTTCCATCC CATCAGCGTT TTGGAATGTT
   D T A I L S V V P F H H G F G M Frame 3
2201  TACTACACTCT GGATATTTGA TTATGAGTTG TCTGGCTGTC TTATATGATA
   T T L G Y L I C G F R V V L M Y R Frame 3
2251  GATTTGAAGGA AGAGCTGTTTT CTGAGGAGCC TTTGGGATTG TCGAGGTTTT
   F E E E L F L R S L Q D Y K I Q Frame 3
2301  AGTGCCTGTGC TTGTTGCAAAC CCTATTCTCC TTCTTCGCCA AAAGGACCTCT
   S A L L V P T L F S F F K A S T L Frame 3
2351  GATGGACAAA TACGATTATTA TCPGATTTAA TCAGAATGATT CAGATTTCAA
   I D K Y D L S N H E I A S G G A Frame 3
2401  CTCCCTCTCT TAAAGAGATC GGGGAAGCGG TTCCCAAGAG GTTTCCATCTG
   P L S K E V G A V A K R F H L Frame 3
2451  CCAGGTATCTT GGCAAGATTAG TGGGCTCAGT GAAGACTACAT CAGCTATTCT
   P G I R Q G Y L T E T T S A I L Frame 3
2501  GATTACACCC GAGGCGGATG ATAAACCGGG CCGGTCAGGT AAAGTTTGGC
   I T P E G D D K P G A V G K V V P Frame 3
2551  CATTCTCTGA AGCGAAGGTT GTGGATCTGG ATACCGGCAA AAGCGCTGGC
   F E A K V V D L D T G K T L G Frame 3
2601  GATTAATCTAA GAGGCGAAGCT GTGGTGTGAGA GTTCACATTTA TTATGCGCG
   V N Q R G E L C V R G P M I M S G Frame 3
2651 TTATGTAAC AATCCGGAAG CGACCAACGC CTTGATTGAC AAGGATGGAT
     Y V N N P E A T N A L I D K D G *  Frame 3
PKA-NFluc:

901  TAAAGCTGGT ACCGAGCTCG GATCCCCACC ATGGGGAACG CCGCCGCCGC  
     MG NA AAA A  
   Frame 1

951  CAAGAAGGCGG AGCGAGCAGG AGAGCGTGAA AGAGTTCCTA GCCAAAGCCA  
     KK G S Q E V K E F L A A K  
   Frame 1

1001 AGGAAGATTT CCTGAAAAAAA TGGGAGACCCT TCTCTAGAA TACAGCCCAG  
     ED FL F K R W E P S Q N T A Q  
   Frame 1

1051 TTGATACGT TTGATAGAAT CAAGACCCTT GGCACCGGCT CCTTTGGCGG  
     LD Q F D R I K T L G T G S F G R  
   Frame 1

1101 AGTGATGCTG GTGAAGCACCA AGGAGATGCG GAACACTAC GAAGCTAGGA  
     V M L V K H E S G N H Y A M K I  
   Frame 1

1151 TCTTAGACAA GCAGAGAGTGT GTGAAGCTAA AGCAGATCAGA GCCACCTCTG  
     ED K Q K V R L K Q I E H T L  
   Frame 1

1201 AATGAGAAGGC GCATCCTGCA GGCAGCTCAC TTCCCCGTTCC TGTCAAAAC  
     NE K R I L Q A V N FP F L V K L  
   Frame 1

1251 TGAATTCTCC TTCAAGGACA ACTCAAACCT GTACATGGTC ATGGAGTATG  
     E F S F K D N S N L Y M V M E Y V  
   Frame 1

1301 TAGCTGGTGG CGAGATGGTC TCCCCACCTAC GGCAGAAGTG AAGCTCAGC  
     AG G E M F S H L R R I G R F S  
   Frame 1

1351 GAGCCCATG CGGCTTTCTTA CGGCGGCCAG ATCGTCTCTG CCGCTAGTA  
     E P H A R F Y A A Q I V L T F E Y  
   Frame 1

1401 TCTGCACCTC CTGGACCTCA TCTACCCGGA CCTGAAGCCG GAGAATCTTC  
     L H S L D L I Y R D L K P E N L L  
   Frame 1

1451 TCACTGACCA GCAGGGCTAT ATTCAGGTGA CAGACTTCCG TTTTGCCAAAG  
     I D Q Q Y I Q V T D F G F A K  
   Frame 1

1501 CGTGTGAAAG GCCGTTACTTG GACCTTGTGT GGCAGCCTCTG AGTACTTGGC  
     R V K G R T W T L C G T P E Y L A  
   Frame 1

1551 CCCCCGAGTT ATCTCGAGCA AAGGCTCACA CAAGGCTGTG GACTGGTGGG  
     PE I L S K G Y N K A V D W W A  
   Frame 1

1601 CTCTCGGAGT CTCATCTCTC GAGATGGCTG CTGTTACCACG ACCCTTCTTC  
     L G V L I Y E M A G Y P F P F  
   Frame 1

1651 GCTGACCGACG CTATCCGAGAT ATATGAGAAAA ATCGTCTCTG GGAAGGCTCG  
     A D Q P I Q I Y E K I V S G K V R  
   Frame 1

1701 GTTCCCCATCC CACTTACGCT GTGACTTGAA GGACCTCTGT CGGAACTTTC  
     F P S H F S S D L K D L L R N L L  
   Frame 1

1751 TGCAAGTGGG TCTCCACCAAG CGGTTTGGGA ACCTCAAGAA CGGGTCATAAT  
     Q V D L T K R F G N L K N G V N  
   Frame 1
1801 GACATCAAGA ACCACAAGTG GTTTGCCACG ACTGACTGGA TTGCCATCTA
DIKNHKWFAITTDWIAYI Frame 1
1851 TCAGAGAAGG GTGGAAGCTC CCTTCATACC AAAGTTTTAA GGCCCTGGGG
QRVKEAPFIPKFKGPD Frame 1
1901 ACACGAGTAA CTTTGAGCAG TATGAGGAGG AAGAGATCCG GTGTCTCCATC
TSNFDDYEEEIERVSI Frame 1

BsiWI
1951 AATGAGAAGT GTTGGCAAGGA GTTTACTGAG TTTAAGATCT CGTACGCGTC
NEKCGKEFTPKFISYAS Frame 1
2001 CCGGGCCGTC GTTCATCTTC GCGGAGGTGA AGACGCCAAA AACATAAAAGA
RGGGSSTGDAPSNDKIKK Frame 1
2051 AAGGCCCCGGC GCCATTCTAT CCAGCTGGAAAG ATGGGACCCT TGGAGACCAA
GPAPFYPLEDGTAGEQ Frame 1
2101 CTGCATAAGG CTATGAAAGA ATACGCCCCG GTTCTCTGGGCA CAATTGTCTT
LHKAMKRYALVPGTIAF Frame 1
2151 TACAGATGCA CATATCGAGG TGGACATCAC TTACGAGAGG TACTTCTGAAA
TDHIEDITYAYEFEM Frame 1
2201 TGGTCCATTCC GTTGGCAAGGA GTTTACTGAG TTTAAGATCT CGTACGCGTC
SVRLAEAMKRYGLNTN Frame 1
2251 CACAGAATCG TCCTATCTCG TGAAGAACCT CTCTCAATCTCTTATGCGGTC
HRIVVCSENSLOFFMPV Frame 1
2301 GTTGGGCCGCG TTATTTCTCG GAGTTGCGAGT TGCGCCCGCG AACGACATTT
LGAFLFIGVAPANDY Frame 1
2351 ATATATGAAAG TGAATGCGCT AACAGTATGG GCATCTCGCA GCCTACCGTG
NERELLSMGISQPTV Frame 1
2401 GTGGTCTCTTT CCAAAAAAGG GGTGCAAAAA ATTTTGAAGG TGCAAAAAAA
VFVSKGQLQIKILNVOKK Frame 1
2451 GCTCCAAATTC ATCCAAAAAA TTATTATTCC GGATTTCAGA ACGGATTACCA
LPIIQKIIIIIMDSKTQ Frame 1
2501 AGGGATTGCA GTGATGGTAC ACGTCTCAGA CATCTCATCT ACCTCCCGGT
GFSMYTVTSHLPG Frame 1
2551 TTTATAGAAT AGCAAGAGGG GCCAGAGCTT TCGATAGGG AACAGACAAT
FNYDFVPESDRFDKTI Frame 1
2601 TGCACTGATC ATGAAGTCTTG CGGATGCTCT TGGTCTGCCT AAAGGTGTCC
ALIMNSSGSGTGLPKGV A Frame 1
2651 CTCTGCCCTCA TAGGACTGCC TGGGTGAGTTCT GTCTCGCATGC CAGAGATCTCT
LPHRTACCVRSHPARPD Frame 1
2701 ATGGGCTGCA ACTCAATCAT CCGAGATCTG CGGATTTAAA GTGTGTTTCC
IFGNIIPDTAILSVP Frame 1
2751 ATTCCATCAC GGTGTTGGAA TGTTTACTAC ACTCGGATAT TTGATATGTG
   F H H G F G M F T T L G Y L I C G Frame 1
2801 GATTTCGAGT CGTCTTTATAG TATAGATTG AGAAGAGGCT GTTTTCTGAGG
   F R V V L M Y R F E E E L F L R Frame 1
2851 AGGCTTCAAG ATTACAAGAT TCAAGAGTCG CTGCTGTTCG CAACCTATT
   S L Q D Y K I Q S A L L V P T L F Frame 1
2901 CTGCTTCTTC GCACACAAAGCA CTCTGATTGA CAAATACGT TTATCTAATT
   S F F A K S T L I D K Y D L S N L Frame 1
2951 TACACGGAAGT TGCTTTCTGTT GCCGCTCCC TCTCTAAGGA AGTCGGGAA
   H E I A S G G A P L S K E V G E Frame 1
3001 CCAGTTGCCA AGAGTTTCCA CTGCGTCTAG TTCTAGCGAG GATATGGGCT
   A V A K R F H L P G I R Q G Y G L Frame 1
3051 CAGTCTGACT ACATCGCTTA TTCTGATTAC ACCCGAGGGG GATGATAAAC
   T E T T S A I L I T P E G D D K P Frame 1
3101 CGGGCGCGGT CGTGAAAGTT CTGCCATTTT TTGAAGCGAA GGTTGTGGAT
   G A V G K V V P F F E A K V V D Frame 1
3151 CTGGTATCCG GGAAAACGCT GGGCGTTAAT CAAAGAGGGC AACTGTGTGT
   L D T G K T L G V N Q R G E L C V Frame 1
3201 GAGAGGTTCT ATGATTTATT CGGTTATAGT AAACAATCCG GAAGCGACCA
   R G P M I M S G Y V N N P E A T N Frame 1
3251 ACAGCCTTTGT TGACAAGGAT GGATGATAAC ATCTGTGTAG CTAGTGAATG
   A L I D K D G * Frame 1
2651 ATACGAGCC GGATGACTGG GAATCGGTAT TCAGCGAATT CCACGATGCT
  YE PDWESVFSFEHDA Frame 1

2701 GATGCGCAGA ACTCTCACAG CTATTTGCTTT GAGATTCTGG AGCGGCGGTA
  DAQNSHYCFEILERR* Frame 1

EcoRV          NotI
2751 ATGAGATATC CAGCACAGTG GCCGCCGCTC GAGTCTAGAG GCCGCCGCGGT
REFERENCES

(4) Michnick, S. W.; Ear, P. H.; Manderson, E. N.; Remy, I.; Stefan, E. Nat Rev Drug Discov 2007, 6, 569-82.


(67) Rackham, O.; Brown, C. M. *EMBO J* 2004, 23, 3346-55.


(72) Ooi, A. T.; Stains, C. I.; Ghosh, I.; Segal, D. J. *Biochemistry* 2006, 45, 3620-5.


(209) Lessene, G.; Czabotar, P. E.; Colman, P. M. *Nat Rev Drug Discov* **2008**, *7*, 989-1000.


