Development and Characterization of Caspase Activatable GFP and a Family of Fluorescent Reporters

Samantha Elizabeth Bernard Nicholls

University of Massachusetts - Amherst, samebernard@gmail.com

Follow this and additional works at: http://scholarworks.umass.edu/open_access_dissertations

Recommended Citation
Development and Characterization of Caspase Activatable GFP and a Family of Fluorescent Reporters

A Dissertation Presented

by

SAMANTHA B. NICHOLLS

Submitted to the Graduate School of the University of Massachusetts Amherst in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

February 2013

Chemistry
Development and Characterization of Caspase Activatable GFP and a Family of Fluorescent Reporters

A Dissertation Presented

by

SAMANTHA B. NICHOLLS

Approved as to style and content by:

________________________________________
Jeanne A. Hardy, Chair

________________________________________
Lynmarie Thompson, Member

________________________________________
S. Thayumanavan, Member

________________________________________
Alejandro Heuck, Member

Craig Martin, Department Head
Chemistry
DEDICATION

For my husband Andrew, who has never faltered in his support; believing in me even when I did not believe in myself and giving me the confidence to accomplish whatever I set out to do.

For my parents John and Karen Bernard, who have always taught me to find my passion, whatever it may be and who have supported me through every disappointment and every accomplishment.
ACKNOWLEDGMENTS

To my advisor Jeanne Hardy; I thank you for all of your support and guidance throughout my PhD. Your enthusiasm for science and research was one of the reasons I chose to join the Hardy lab and has been a motivating factor throughout my research. I appreciate your dedication not only to helping your students grow as scientists but your genuine interest in our success as individuals.

To my committee; I thank you for all of your support and advice throughout the last five years. Your enthusiasm and guidance throughout my five years at UMass remind me how fortunate I am to have come to this program and how wonderful the community within the science departments is here in Amherst.

To my Hardy lab family; I am very grateful to the past and present Hardy lab members who have made my experience here at UMass so memorable. Sravanti Vaidya, Witold Witkowski, Kristen Huber, Elih Velasquez and Genevieve Abbruzzese were invaluable mentors not only in learning the science and techniques I needed to succeed in my project but also in helping to navigate being a graduate student here at UMass and for being true friends. In choosing a lab you not only choose the science you will learn but the people you will spend the next years of your life with and I have been very fortunate in my choice.

For my family and friends; I sincerely have the most supportive family and friends a person could ask for. I thank my husband, parents, brothers, uncles, and in-laws for their support not only over the last five years but always. To my parents for convincing me I could be whatever I wanted and to choose something that I would enjoy doing every day. To my brothers Donald and Ken, for taking all of my late night calls, for always making me smile when I’m stressed out, and for not letting me ever take myself too seriously. To my Uncle Tim; thank you for all of your support, I think reading all of the text books you bought me have paid off! To my in-laws Melinda, Allan, Nancy, John, Ashley, and David; I couldn’t ask for more supportive family, I appreciate all of your support and confused looks every time you listened to me try to explain what it is I do.
The cellular process of programmed cell death, or apoptosis, is critical in homeostasis and development. In addition it’s misfunction is implicated in an array of disease states from cancer to neurodegeneration, making it an attractive pathway for drug targeting. A family of proteases, known as caspases, plays a central role in the apoptotic cascade resulting in the ultimate destruction of the cell. We report a genetically encoded dark-to-bright reporter of caspase activity used in E.coli, mammalian cells, and whole organisms which can be used to monitor apoptosis. This reporter, caspase activatable green fluorescent protein (CA-GFP) consists of GFP fused through a flexible linker containing the caspase-3 and -7 recognition sequence, DEVD, to a hydrophobic peptide derived from the influenza A viral M2 protein. This fusion reporter shows a significant fluorescent response in the presence of active caspase. CA-GFP is unique in its ability to hold GFP in a dark state prior to cleavage by active protease. We investigate the mechanism of quenching, examining the structural characteristics which lead to the inability of the GFP chromophore to mature in the presence of the peptide. In better understanding the mechanism of quenching we can engineer CA-GFP to ultimately be used in transgenic animal models. This requires the development of a palette of protease-
activatable fluorescent proteins (PrA-FP) which would enable the monitoring of multiple proteolytic events within a cell or organism in real time. Our development of this palette of reporters, varying in their fluorescence and proteolytic response shows that CA-GFP has the potential to be a powerful tool for the study of the role of apoptosis during development in whole organism models and could be an important tool in understanding the role of individual proteases within the complex biochemical environment in the cell.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACKNOWLEDGMENTS</td>
<td></td>
<td>v</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td></td>
<td>vi</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td></td>
<td>xi</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td></td>
<td>xii</td>
</tr>
<tr>
<td>CHAPTER</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.</td>
<td>INTRODUCTION: PROTEASE REPORTING AND MOLECULAR IMAGING THROUGH FLUORESCENCE</td>
<td>1</td>
</tr>
<tr>
<td>1.1</td>
<td>Caspases and the Apoptotic Pathway</td>
<td>1</td>
</tr>
<tr>
<td>1.2</td>
<td>Existing caspase reporter technologies</td>
<td>4</td>
</tr>
<tr>
<td>1.3</td>
<td>Design of a genetically encoded dark-to-bright reporter</td>
<td>7</td>
</tr>
<tr>
<td>1.4</td>
<td>The fluorescent protein family</td>
<td>10</td>
</tr>
<tr>
<td>1.5</td>
<td>GFP structure and chromophore formation</td>
<td>13</td>
</tr>
<tr>
<td>1.6</td>
<td>The M2 transmembrane domain</td>
<td>15</td>
</tr>
<tr>
<td>2.</td>
<td>CA-GFP: A GENETICALLY ENCODED DARK-TO-BRIGHT REPORTER FOR CASPASE ACTIVITY</td>
<td>21</td>
</tr>
<tr>
<td>2.1</td>
<td>Introduction</td>
<td>22</td>
</tr>
<tr>
<td>2.2</td>
<td>Results</td>
<td>25</td>
</tr>
<tr>
<td>2.2.1</td>
<td>CA-GFP fluorescence in bacteria</td>
<td>25</td>
</tr>
<tr>
<td>2.2.2</td>
<td>CA-GFP can be activated and observed in cells</td>
<td>26</td>
</tr>
<tr>
<td>2.2.3</td>
<td>CA-GFP is an apoptosis reporter in mammalian cells</td>
<td>27</td>
</tr>
<tr>
<td>2.2.4</td>
<td>CA-GFP in whole animal models</td>
<td>31</td>
</tr>
<tr>
<td>2.3</td>
<td>Discussion</td>
<td>33</td>
</tr>
<tr>
<td>2.4</td>
<td>Materials and Methods</td>
<td>36</td>
</tr>
<tr>
<td>2.4.1</td>
<td>Molecular cloning</td>
<td>36</td>
</tr>
<tr>
<td>2.4.2</td>
<td>Lysate-based GFP fluorescence assay</td>
<td>37</td>
</tr>
<tr>
<td>2.4.3</td>
<td>Immunoblotting of bacterial cultures</td>
<td>38</td>
</tr>
<tr>
<td>2.4.4</td>
<td>Fluorescence microscopy</td>
<td>38</td>
</tr>
<tr>
<td>2.4.5</td>
<td>Flow cytometry</td>
<td>39</td>
</tr>
<tr>
<td>2.4.6</td>
<td>Cell culture and transfection</td>
<td>39</td>
</tr>
<tr>
<td>2.4.7</td>
<td>Immunoblotting of mammalian cultures</td>
<td>40</td>
</tr>
</tbody>
</table>
2.4.8 Live cell imaging .................................................................40
2.4.9 Statistical analysis for live cell imaging..............................41
2.4.10 Zebrafish injections of CA-GFP .......................................41

3. MECHANISTIC INSIGHTS INTO THE CA-GFP DARK STATE......44

3.1 Introduction ........................................................................45
3.2 Results ..............................................................................46
  3.2.1 CA-GFP requires no cofactors and can be activated \textit{in vitro} 47
  3.2.2 Chromophore maturation is prevented in dark CA-GFP ........48
  3.2.3 CA-GFP has secondary structure similar to GFP1-10 ..........50
  3.2.4 \(\beta\)-Barrel of CA-GFP not fully formed .........................54
  3.2.5 CA-GFP exists in 'pro-folding' conformation .......................57
3.3 Discussion ........................................................................60
3.4 Materials and Methods ..........................................................65
  3.4.1 Molecular cloning .........................................................65
  3.4.2 Protein expression and purification ................................66
  3.4.3 \textit{in vitro} CA-GFP cleavage and fluorescence assay ..........68
  3.4.4 Absorbance of GFP chromophore ..................................69
  3.4.5 Mass Spectrometry .......................................................69
  3.4.6 Circular dichroism spectroscopy ....................................70
  3.4.7 Size exclusion chromatography .....................................71
  3.4.8 Proteolysis .................................................................71
  3.4.9 Cleavage by Caspase-6 ................................................72
  3.4.10 TROSY NMR ............................................................72
  3.4.11 Lysate based CA-GFP fluorescence assays .....................73
  3.4.12 Western blotting ....................................................74

4. A VERSATILE PLATFORM FOR PROTEASE REPORTING ....77

4.1 Introduction ........................................................................78
4.2 Results and discussion .......................................................81
  4.2.1 Development of a caspase-6 activatable-GFP ....................81
  4.2.2 Profiling of caspase activity using C6A- and C7A-GFP ..........85
  4.2.3 Engineering a new proteolytic recognition: Human
      Rhinovirus 3C ................................................................88
  4.2.4 Engineering viral protease recognition: Dengue virus
      protease and HIV protease ..............................................91
  4.2.5 Engineering various colors of PrA-FPs for multi-protease
      reporting .................................................................95
4.3 Conclusion ..........................................................................99
4.4 Materials and Methods ........................................................101
  4.4.1 Molecular cloning ....................................................101
  4.4.2 Fluorescence assays in cell lysates .................................103
4.4.3 Fluorescence assays in whole cells ...................................................104
4.4.4 Protease profiling of C7A- and C6A-GFP e .....................................104
4.4.5 Western blotting .............................................................................105
4.4.6 in vitro analysis of DVPA-GFP ......................................................106
4.4.7 Two dimensional Gel Analysis .........................................................106

5. CA-GFP DEVELOPMENT AND APPLICATIONS ....................................110

5.1 CA-GFP: A model reporter of apoptosis ..........................................110
5.2 Analysis of background fluorescence ..............................................111
5.3 Peptide quenching ...........................................................................115
5.4 Expression and codon usage ............................................................116
5.5 Kinetic control ..................................................................................119
5.6 CA-FP in drug design: back to biology .............................................120

APPENDICES

A. MOLECULAR DYNAMICS SIMULATIONS OF PEPTIDE STABILITY ..........124

B. PROTOCOLS ..........................................................................................142

C. PROTEASE ACTIVATABLE-FLUORESCENT PROTEINS .........................155

BIBLIOGRAPHY ........................................................................................157
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 1.1 Comparison of Existing Technologies of CA-GFP</td>
<td>9</td>
</tr>
<tr>
<td>Table 4.1 Constructs used throughout the studies with the linker sequence as well as any additional mutations listed.</td>
<td>82</td>
</tr>
<tr>
<td>Table 4.2 Comparison of the linker sequence of CA-GFP and the DVPA-GFP and HIVA-GFP reporters ((\downarrow) indicates the site of cleavage by the intended protease).</td>
<td>92</td>
</tr>
<tr>
<td>Table 4.3 The values for the increase in fluorescence for each reporter as well as the reported quantum yield for each fluorescent protein and the brightness of each reporter after cleavage by active caspase-7 in <em>E.coli</em> lysates relative to CA-GFP.</td>
<td>98</td>
</tr>
<tr>
<td>Table A.1 Potential helical peptide sequences for regio-rigid scaffold</td>
<td>127</td>
</tr>
<tr>
<td>Table A.2 Log of simulations run using GROMACS</td>
<td>129</td>
</tr>
<tr>
<td>Table B.1 Buffer compositions for Co-NTA purifications of GFP proteins</td>
<td>144</td>
</tr>
<tr>
<td>Table B.2 Fluorescence Parameters for FPs</td>
<td>146</td>
</tr>
<tr>
<td>Table B.3 PCR reaction for overlapping primers amplification of full-length gene</td>
<td>149</td>
</tr>
<tr>
<td>Table B.4 PCR cycle for amplification for full-length gene, the method should be paused after 5 cycles and primers 2 and 3 added to the reaction.</td>
<td>149</td>
</tr>
<tr>
<td>Table C.1 Summary of constructs and fluorescent properties tested of existing protease activatable- fluorescent proteins.</td>
<td>155</td>
</tr>
</tbody>
</table>
### LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 1.1: Apoptosis can be initiated through either the caspase-8 dependent extrinsic pathway or the caspase-9 dependent intrinsic pathway.</td>
<td>1</td>
</tr>
<tr>
<td>Figure 1.2: Caspase share a common active-site cysteine-histidine diad in cleaving their substrates (A).</td>
<td>3</td>
</tr>
<tr>
<td>Figure 1.3: The substrate (S) binding pocket of the protease accommodates the peptide substrate (P) which is cleaved between the P1 and P1’ residues.</td>
<td>4</td>
</tr>
<tr>
<td>Figure 1.4: The CA-GFP is constructed of Green Fluorescent Protein fused through a linker containing the caspase-7 cleavage recognition sequence DEVD to the M2 transmembrane domain, which quenches GFP fluorescence.</td>
<td>8</td>
</tr>
<tr>
<td>Figure 1.5: GFP is formed by the assembly of 11 strands, shown as arrows, to form a β-barrel with a helix, shown as a cylinder at the center of the barrel, which contains the three residues that undergo posttranslational modification to form the fluorescent chromophore.</td>
<td>13</td>
</tr>
<tr>
<td>Figure 1.6: The three-dimensional structure of the GFP β-barrel (PDBID: 2YOG) with the chromophore shown in sticks in the center of the barrel (A).</td>
<td>14</td>
</tr>
<tr>
<td>Figure 1.7 M2 facilitates dissociation of the coat proteins and release of viral DNA.</td>
<td>15</td>
</tr>
<tr>
<td>Figure 1.8: Structure of the M2 transmembrane channel at pH 7.5-8 (PDBID: 2RLF) left, at pH 6.5 (PDBID: 3LBW) center, and at pH 5 with inhibitor rimantadine bound (PDBID: 3BKD) right.</td>
<td>16</td>
</tr>
<tr>
<td>Figure 2.1: CA-GFP consists of Green fluorescent protein (green) fused through a flexible linker containing the caspase-3 and -7 recognition sequence DEVD (blue) to a 27-amino acid quenching peptide on the C-terminus (gray).</td>
<td>25</td>
</tr>
<tr>
<td>Figure 2.2 CA-GFP becomes fluorescent only upon cleavage by active caspase.</td>
<td>26</td>
</tr>
<tr>
<td>Figure 2.3: CA-GFP fluorescence is robust enough for fluorescence microscopy and flow cytometry.</td>
<td>27</td>
</tr>
</tbody>
</table>
Figure 2.4: CA-GFP is fluorescently activated in mammalian cells undergoing apoptosis. ............................................................... 28

Figure 2.5: The response of CA-GFP can be measured in single cells undergoing apoptosis. ................................................................. 30

Figure 2.6: DNA injection of GFP and CA-GFP in zebrafish. ................................................................. 32

Figure 2.7: Quantification of fluorescence of the area of the imaged embryo shows a 10-fold increase in fluorescence signal in the treated CA-GFP injected embryos over those not treated with apoptosis inducer hydroxyurea. ................................................................. 32

Figure 3.1: CA-GFP is constructed from the Green Fluorescent Protein (S65T) GFP fused to a linker containing the caspase-3 and -7 recognition sequence DEVD to the 27 amino acid quenching peptide (QP) derived from the transmembrane domain of influenza M2. .......................... 45

Figure 3.2: CA-GFP can be activated in vitro ................................................................. 47

Figure 3.3: The CA-GFP silencing relies on prevention of chromophore maturation. ................................................................. 48

Figure 3.4: Size exclusion chromatogram for GFP, CA-GFP and cleaved CA-GFP. ................................................................. 49

Figure 3.5: Circular dichroism spectra of GFP and CA-GFP indicate that both proteins are folded into a predominantly β-sheet structure. ................................................................. 50

Figure 3.6: The structure of GFP β-barrel (PDB ID:2YOG, green) with the chromophore residues shown as sticks. ................................................................. 51

Figure 3.7: Circular Dichroism spectra. ................................................................. 52

Figure 3.8: The size exclusion chromatogram of GFP (MW 27.7 kDa), CA-GFP (32 kDa), cleaved CA-GFP (28.2 kDa), and GFP_{1-10} (20.5 kDa). ................................................................. 53

Figure 3.9: Protease susceptibility of GFP, CA-GFP, and GFP_{1-10}. ................................................................. 55

Figure 3.10: TROSY NMR HSQC spectra. ................................................................. 56

Figure 3.11: Constructs of CA-GFP and N-terminally tagged nCA-GFP with their relative fluorescence when co-expressed in E.coli with active caspase-7 or an inactive version of caspase-7 C186A. ................................................................. 58
Figure 3.12: Sequence alignment of linker regions of CA-GFP, nCA-GFP, and CA-sfGFP ................................................................. 60

Figure 4.1: Construct of CA-GFP showing the GFP (green) fused through the flexible linker, containing the protease recognition sequence, to the quenching peptide (gray). ................................................................. 79

Figure 4.2: Whole cell fluorescence of caspase-6 activatable reporter constructs. ........................................................................... 83

Figure 4.3: The three dimensional structure of GFP (PDBID: 2YOG) with the VELD (residues 24-27) highlighted in the strand. .................. 83

Figure 4.4: The relative fluorescence of C6A-GFPc (top) and C7A-GFPc (bottom) in response to co-transformation with inactive (white) or active (gray) caspases -3, -6, -7, and -8. .................................................. 86

Figure 4.5: Fluorescence response of PrA-GFPs containing various linker sequences to human Rhinovirus 3C (hR3C) ......................... 90

Figure 4.6: Two dimensional gel analysis of dengue virus protease 2 (DVP2) (top) with two DVP2 activatable reporters DVPA-GFP_v1 (middle) and DVPA-GFP_v2 (bottom). ....................................................... 92

Figure 4.7: The constructs for CA-CerFP, CA-CitFP, and CA-mNeptune. .......... 96

Figure 5.1: Comparison of the relative background fluorescence of reporters cotransformed with an empty pET vector and are therefore uncleaved. ......................................................................................... 113

Figure 5.2: Comparison of the relative background fluorescence in caspase activatable reporters of varying versions of GFP when co-transformed with inactive caspase-7 (C186A). ............................................... 113

Figure 5.3: Relative background fluorescence of C6A-GFPc and C7A-GFPc when co-transformed with the inactive versions of caspases -3, -6, -7, and -8. ................................................................. 119

Figure A.1: Design of a helical peptide regio-regular scaffold for proton transport. ........................................................................... 125

Figure A.2 Comparison of peptide 2MAG (initial coordinates shown in green) after a 1 ns simulation in H2O on left (cyan) vs. after 10 ns simulation in DMSO (blue). ................................................................. 126

Figure B.1: Primer design for overlapping molecular cloning method ............ 148
CHAPTER 1
INTRODUCTION: PROTEASE REPORTING AND MOLECULAR IMAGING THROUGH FLUORESCENCE

1.1. Caspases and the apoptotic pathway

Apoptosis, or programmed cell death, is an essential pathway in eukaryotes. It is critical for proper development and homeostasis. It is also one of the most widely targeted pathways for drug design in the search for treatments for a diverse array of diseases affected by its malfunction.

Apoptotic proteins are targeted for activation in the search for treatment of diseases involving rapidly dividing cells, such as cancer, while other groups work to find inhibitors to prevent unwarranted cell death in diseases such as Alzheimer's or to minimize cellular destruction after a heart attack or stroke. One of the most popular and plausible classes of proteins within the apoptotic pathway for successful drug targeting is

![Diagram of the apoptotic pathway]

Figure 1.1. Apoptosis can be initiated through either the caspase-8 dependent extrinsic pathway or the caspase-9 dependent intrinsic pathway. Both the extrinsic and intrinsic apoptotic pathways converge on the executioner caspases-3, -6, and -7 ultimately leading to cell death. (Thanks to KP for figure)
the family of proteases known as caspases. Caspases are expressed in an inactive or zymogen form in all cells and are held at bay until cell death is triggered. Apoptosis can proceed through one of two pathways, an intrinsic pathway and an extrinsic pathway (Fig. 1.1). The intrinsic pathway is initiated through the release of pro-apoptotic factors, such as cytochrome c, from the mitochondria. Cytochrome c then goes on to form a heptameric complex with Apoptotic Protease Activating Factor-1 (Apaf-1), dATP, and procaspase-9. The formation of this complex activates caspase-9 to cleave target enzymes, which include caspases-3, -6, and -7, known collectively as the executioner caspases. In the extrinsic pathway pro-apoptotic cytokines are bound. These receptors can then associate with Fas associated protein with death domain (FADD) and caspase-8 or -10 and form the Death-Inducing Signaling Complex (DISC) activating caspase-8, which can also activate the executioner caspases. Once the executioner caspases have been activated they go on to cleave specific targets in the cell leading to apoptosis. In healthy cells, however, executioner caspases (-3, -6, and -7) are antagonized by Inhibitor of Apoptosis Proteins or IAP’s. In cancer cells IAP’s are over-expressed, leading to the repression of apoptosis. The importance of caspases in apoptosis is central. They catalyze irreversible protein cleaving reactions leading to apoptosis, which can be inhibited or activated by small molecules and cellular factors. Understanding the molecular details of the intricacies of the caspase mechanism of cleavage is critical for the rational design of drugs for this class of proteins. It is also important to fully understand their collective and unique roles in developmental processes as well as any non-apoptotic roles they may have within normal cellular processes.
Caspases are a class of cysteine aspartate proteases and share a common active-site diad (cysteine-histidine) (1). They also share similar specificity for substrate proteins, cutting after aspartate residues by cleaving the amide bonds in a nucleophilic attack by the cysteine thiolate at the amide carbonyl in specific aspartate-containing motifs (2) (Fig. 1.2A). All caspases are expressed as inactive zymogens known as ‘procaspases’ that immediately dimerize post-translationally. The executioner caspases are constitutive dimers that are activated through cleavage by an initiator caspase at two sites releasing the pro-domain and generating a large and small subunit (Fig. 1.2B). The active form is a heterotetramer made of two large and two small subunits. Each active heterotetramer contains two active sites containing what is referred to in protease terminology as a subsite (S) pocket for substrate to bind. The specific binding site for each of the recognized residues N-terminal to the cleavage site is referred to as the S1-SX site with S1 being the residue after which the protease cleaves. The binding pocket which accommodates the residue immediately prior to which the protease cleaves is referred to

**Figure 1.2.** Caspases share a common active-site cysteine-histidine diad in cleaving their substrates (A). Active caspases are a heterotetramer of two large and two small subunits after cleavage of the pro-domain and intersubunit linker (B) (Figure generated by SN in Pymol).
as the S1’ site. The substrate or peptide (P) is named similarly, with the residue after which the protease cleaves being the P1 residue and the site prior to which the protease cleaves the P1’ residue (Fig. 1.3).

One of the preferred substrate cleavage sequences for the executioner caspases -3 and -7 is the amino acid sequence DEVD (P4-P1) (3). Synthetic substrates with this sequence are recognized by the executioner caspases. Since the study of caspases is so important to understanding both disease and development, the ability to monitor caspase activity both \textit{in vivo} and \textit{in vitro} is likewise critically important. For the investigation of caspase activity \textit{in vitro} the most common reporters are fluorogenic peptide substrates. These reporters are peptides containing a sequence that can be recognized by a particular caspase and a covalently attached fluorophore (3). The fluorophore is quenched while covalently coupled but gains fluorescence upon cleavage.

1.2. Existing caspase reporter technologies

Apoptosis is a critical cellular process in the regulation of homeostasis and is an important target in drug design. This significance has driven researchers to develop several reporters of apoptotic and caspase activity, with the most common technologies using small synthetic fluorogenic peptides that act as caspase substrates. While these substrates have an exceptional signal-to-noise ratio (50,000) they are primarily useful in
in vitro or cell-based applications but cannot be used in living organisms without sacrificing the animal. One of the most useful of these reagents is the tetrapeptide-aminoluciferin reagent Caspase-glo from Promega. In the presence of active caspase-3 or -7 the aminoluciferin, which is a substrate for luciferase, is cleaved from the peptide. When luciferase is added it produces a luminescent signal proportional to the caspase activity which can be used in high-throughput screening of in vitro caspase activity or whole cell caspase and apoptosis screens. This reagent is widely used and amenable to several systems (4). However recent studies from our own lab have also shown that there is a significant difference in the kinetics of cleavage of peptide substrates vs. protein substrates (5) indicating that these short peptide substrates may not accurately reflect the in vivo cleavage by all proteases. Therefore, there is a great need for a genetically encoded caspase reporter that is more similar to a protein substrate and thus more accurately reflects the activity of caspases against native substrates.

Several existing technologies use green fluorescent protein (GFP) and its variants to report caspase activity in live cells. Park et al. have developed a GST-DEVD-EGFP reporter, which is presented as a cost effective alternative for the expensive fluorescent substrates currently used. They use the GST tag as a mechanism for anchoring the reporter on GST-beads or by immobilization on a chip. After the addition of caspase and a washing step, they can see cleavage has occurred both through immunoblotting as well as by the loss of fluorescence which can be quantified (6). The obvious drawback is that this is an on-to-off reporter, so small changes in caspase activity cannot be discerned and this technology is not compatible with in vivo applications.

Many of these technologies use GFP or it’s derivative proteins, which fluoresce at
a variety of wavelengths in a fluorescence resonance energy transfer (FRET) pair, or in some combination with another fluorescent protein. FRET is a commonly used technique in which fluorophores of two distinct excitation and emission wavelengths are used to indicate proximity to one another. When the donor of the FRET pair is in close proximity to the acceptor the emission of the donor excites the acceptor and the fluorescence at the acceptor emission wavelength can be detected. When the pair of fluorophores is far apart the signal from the acceptor drops off and only the donor wavelength can be detected. Many of these FRET systems are based on reporters containing cyan and yellow fluorescent protein variants fused through the DEVD caspase cleavage site. One of these Cyan/Yellow reporters, SCAT3, which uses ECFP and Venus fluorescent proteins, has been successfully used to perform live imaging of caspase activation in both *Drosophila* and mice (7-9). Another FRET based caspase reporter is a real-time monitoring reporter using double FRET CFP (cyan fluorescent protein)/YFP (yellow fluorescent protein)/dsRFP (red fluorescent protein) reporting (10). This system was used to correlate an apoptotic timeline for multiple caspases in live cells under the induction of apoptosis. Though the system could report apoptosis, it also displayed a very high background from the multiple fluorescent proteins present in the cells at once, which is a limitation of all the FRET based reporters. More recent developments of these FRET reporters have developed FRET pairs which utilize green/red fluorescent proteins which decreases the background. One example is CaspeR3, which fuses TagRFP and TagGFP through a 13 amino acid linker containing the DEVD cleavage site in the linker (11). While this improves the signal over the Yellow/Cyan pair reporters, a dark-to-bright transition still has an advantage in terms of its low background signal.
One of the most notable of the existing caspase activity monitors is a reporter molecule named ‘Apoliner’. Apoliner is a reporter fusion of a membrane signaling sequence (mCD8), mRFP (monomeric red fluorescent protein), a caspase cleavage recognition site, and EGFP. Before apoptotic cleavage the EGFP and mRFP can be visualized in the membrane via confocal microscopy. When apoptosis is induced and Apoliner is cleaved the EGFP is trafficked to the nucleus, physically separating the EGFP and mRFP signals (12). However, Apoliner, like many of the technologies discussed above, can only be utilized via microscopy.

Another reporter utilizes a GFP with diminished fluorescence in a fusion construct that relies on N-rule cleavage, but this reporter in particular only reports a five-fold increase in the fluorescence after cleavage (13). Another reporter named the multimodality reporter takes advantage of the plethora of imaging technologies by making a reporter of a fluorescent protein, mRFP, a bioluminescence reporter, Firefly Luciferase, and a positron emission topography reporter, thymidine kinase (14). While this reporter was successful in tracking cleavage in vivo there was a very high background and the increase in signal of any of the three reporters was only 2-fold, which leads to a problematic signal to noise ratio.

1.3. Design of a genetically encoded dark-to-bright reporter

After assessment of the current technologies available for monitoring caspase activity in a variety of systems, we determined the criteria we deemed most important in the design of a novel reporter of apoptosis. First, it should be genetically encoded for non-invasive monitoring. Genetically encoded reporters have an advantage in that they
can be expressed in individual cells eliminating the need for cellular delivery, reducing potentially damaging effects on the cell. This also enables the reporter to be expressed and purified which is cost effective in comparison to many expensive fluorogenic peptide substrates.

Secondly, the response should be a dark-to-bright fluorescence signal as fluorescence is by far one of them most widely used methods of monitoring signal. By having a dark-to-bright response the signal to noise ratio is maximized, allowing for unambiguous monitoring of activity. It should also respond independently of any cellular co-factors for ease in detection. As the system we wish to study is directly responsible for the degradation of protein substrates within the cell, the response should be independent of any cellular pathways which will ultimately be disrupted. Lastly, our ideal reporter should

**Figure 1.4.** The CA-GFP is constructed of Green Fluorescent Protein fused through a linker containing the caspase-7 cleavage recognition sequence DEVD to the M2 transmembrane domain, which quenches GFP fluorescence. After cleavage by active caspase the GFP fluoresces yielding a useful reporter of caspase activity. (Figure generated using manual modeling of linker and peptide fusion to GFP by SBN using Pymol)
Table 1.1. Comparison of Existing Technologies to CA-GFP.

<table>
<thead>
<tr>
<th>Reporter</th>
<th>Genetically Encodable</th>
<th>Multicolor</th>
<th>Organism</th>
<th>Temporal Resolution</th>
<th>S/N Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ac-DEVD-AFC</td>
<td>No</td>
<td>No</td>
<td>In Vitro</td>
<td>No</td>
<td>High</td>
</tr>
<tr>
<td>FRET</td>
<td>Yes</td>
<td>Yes</td>
<td>In Vivo/In Vitro</td>
<td>Yes</td>
<td>Low</td>
</tr>
<tr>
<td>BRET</td>
<td>Yes</td>
<td>Yes</td>
<td>In Vivo/In Vitro</td>
<td>Yes</td>
<td>Low</td>
</tr>
<tr>
<td>NES-DEVD-EYFP-NLS</td>
<td>Yes</td>
<td>Yes</td>
<td>In Vivo/In Vitro</td>
<td>No</td>
<td>Low</td>
</tr>
<tr>
<td>Mem-mRFP-DEVD-NLS-EGFP</td>
<td>Yes</td>
<td>Yes</td>
<td>In Vivo</td>
<td>No</td>
<td>Low</td>
</tr>
<tr>
<td>Ubiquitin-LETID-EGFP</td>
<td>Yes</td>
<td>No</td>
<td>In Vivo</td>
<td>No</td>
<td>Low</td>
</tr>
<tr>
<td>PepA-Nluc-DEVD-Cluc-PepB</td>
<td>Yes</td>
<td>Yes</td>
<td>In Vivo/In Vitro</td>
<td>No</td>
<td>High</td>
</tr>
<tr>
<td>CA-GFP</td>
<td>Yes</td>
<td>Yes</td>
<td>In Vivo/In Vitro</td>
<td>Yes</td>
<td>High</td>
</tr>
</tbody>
</table>

be amenable for in vitro, E.coli, FACS, mammalian cell and whole organism

applications. Each of these systems has distinct advantages in the study of apoptosis and in the development of effectors of caspase activity.

In this work, Green Fluorescent Protein (GFP S65T), a mutant of GFP wherein the chromophore is mutated to achieve an increase in fluorescence (15), was fused through a linker to the transmembrane region of the influenza A protein M2 (Fig. 1.4). The resulting fusion protein lacked the intrinsic fluorescence of GFP while still being stably expressed in E. coli. To take advantage of this dark state the cleavage recognition sequence for effector proteases caspase-3 and -7 (DEVD) was inserted into the linker region of the construct. In the presence of active caspase the M2 peptide is cleaved and the protein is able to fluoresce. Because of the low fluorescent background of the dark state of caspase-activatable green fluorescent protein (CA-GFP) there is a clear signal of caspase activity after the peptide has been cleaved. CA-GFP fulfills our criteria for an ideal reporter and offers several advantages in its properties over existing technologies (Table 1.1). We have sought to take advantage of this novel reporter of apoptotic activity
as well as understand the mechanism through which the M2 peptide is able to quench the GFP fluorescence. Here, we put forward a versatile and tunable platform for proteolytic activity reporting and give mechanistic insight to help in the development and expansion of this family of protease-activatable fluorescent proteins (PrA-FP’s).

1.4. The Fluorescent Protein Family

The importance of GFP and its many derivatives has been highlighted by the 2008 Nobel Prize in Chemistry being awarded to three of the most prominent researchers responsible for the discovery of GFP and the development of numerous techniques utilizing GFP. GFP was identified in the early 1960’s by Osamu Shimomura when the protein was isolated from the jellyfish *Aequorea victoria* (16). Though Douglas Prasher was the individual responsible for first cloning the GFP gene (17), it was further investigated and expanded upon by the groups of Martin Chalfie and Roger Tsein with whom Shimomura shares the Nobel Prize. Structural characterization by crystallography was central to understanding how GFP functions (18). The fluorescent capabilities of the protein are the result of a post-translational modification of three residues, Ser65-Thr66-Gly67, in the center of the β-barrel structure of the protein, which form the chromophore. When GFP is properly folded, the conformation of the helix in the center of the β-barrel induces formation of the chromophore.

The research community did not just stop with a green fluorescent protein but have expanded the color palette through extensive engineering to include other green, yellow and cyan proteins. This was accomplished by rational design as well as through directed evolution, performing point mutations on and around the chromophore region of
GFP (19). The discovery of a homologous protein dsRed in a coral species by Sergey Lukyanov (20) led to the development of a family of red fluorescent proteins extending the range of the color palette into the near IR region and increasing the utility even further of the fluorescent family of proteins. The Tsein lab has gone on to engineer a protein, not related to the GFP or dsRed proteins, but with a very different structure which emits light in the infrared range (21), further extending the capabilities of this large class of molecular tools. Fluorescent probes with a further red-shifted emission wavelength have a greater tissue depth penetrance and are therefore more useful in whole organism imaging (22,23). The near IR and IR shifted variants are increasingly in demand as imaging technology targeting cancer and other cell types of interest with fluorescent labels improves.

Wild-type GFP is rarely used today; enhanced GFP (EGFP) is a commonly used version that contains two fluorescence enhancing mutations, S65T and F64L. The combination of these two mutations enhances the fluorescence intensity at the 488 nm peak, normally the minor excitation wavelength, to 35 times that of wild type (15). This allows for the use of FITC (Fluorescein IsoThioCyanate) filters, common in many fluorescence instruments, to be used in exciting GFP. Also used commonly are the GFP cycle 3 mutant (24) and a superfolder GFP (25). These mutants were engineered to have a faster folding and chromophore maturation and to reduce aggregation.

Several groups have independently developed versions of GFP for protein complementation assays. The general design of these reporters is that a portion of GFP sequence is expressed independently from the rest, with one or both of the pieces fused to a protein of interest. The two sections can range from being approximately equal in size
as the versions developed by the Regan (26) and Kerppola (27) groups to the ‘split-GFP’
developed by the Waldo group which only removes the final strand of the β-barrel (28).
When the two independently expressed portions of GFP are drawn within close proximity
to each other by the interactions of their fusion partners they are able to assemble into the
β-barrel geometry, allowing for chromophore maturation. These split or bi-furcated
constructs are useful in determining if and where specific proteins of interest interact in a
cell.

Besides developing a variety of different colored fluorescent proteins, researchers
have also gone on to develop various reporter or ‘optical highlighter’ proteins. These
include proteins such as a photo-activatable GFP(29), Eos(30), Kaede(31), Dronpa(32),
and Kindling(33) proteins that have the ability to switch from either fluorescently ‘off’ to
‘on’ or green to red fluorescence when activated with UV light (for review see (34)) .
Kindling and Dronpa have the added benefit of being reversible in their transitions.
Kindling Fluorescent protein-1 (KFP1) can convert to red fluorescence upon irradiation
with green light. This state can relax back to non-fluorescence in the dark or can be
quenched back to the non-fluorescent state by irradiation of blue light (31,34). Dropna,
named after the Japanese word for the disappearance of a ninja, initially has green
fluorescence. This green fluorescence can be quenched by blue light to a non-fluorescent
state. It can then be activated to fluorescence by 400 nm irradiation in a reversible
process that can be repeated several times (32,34). These photoactivatable fluorescent
proteins are an excellent tool for precise labeling and tracking of a variety of targets in
living systems as well as a great tool in kinetic microscopy, a field previously dominated
by fluorescence recovery after photobleaching (FRAP) techniques. However,
photactivatable GFP variants have limited applicability outside of confocal microscopy because these proteins only respond to exogenous visible or UV illumination but can not directly respond to cellular or enzymatic cues.

**1.5. GFP Structure and Chromophore Formation**

GFP is formed by a stable β-barrel through the assembly of eleven β strands (Fig. 1.5) with a kinked helix formed on the interior of the barrel containing the three sequential residues which make up the fluorescent chromophore (Fig. 1.6A). The three amino-acid chromophore of wild-type GFP is formed by a cyclization reaction in which the amide nitrogen of Gly67 initiates a nucleophilic attack on the carbonyl carbon of Ser65, forming an imidazolone ring. The ring is then oxidized to a cyclic imine and the carbonyl oxygen of Ser65 is dehydrated, fully conjugating the system and forming the benzyldenedimethylimidazolinone chromophore (Fig. 1.6B). The chromophore is protected from solvent within the stable β-barrel of GFP, allowing it to fluoresce. This barrel is a very stable fold which is resistant to proteolysis (35) and has denaturation temperatures nearing 80°C (36,37). The fold and ability of the chromophore to mature is
also retained in several circularly permuted versions of the protein (38).

![Figure 1.6](image)

**Figure 1.6.** The three-dimensional structure of the GFP β-barrel (PDBID: 2YOG) with the chromophore shown in sticks in the center of the barrel (A). The chromophore undergoes an oxidation and dehydration to form the mature fluorescent form of the protein (B). (Figure generated by SN using Pymol and Chemdraw)

Chromophore maturation is essentially irreversible, however, fluorescence can be quenched through denaturation (39), exposing the chromophore to solvent, as well as changes in pH (39,40) and physical compression (41). In each case quenching is reversible by restoring the GFP to initial conditions, reestablishing the protection of the chromophore and fluorescence is nearly immediately restored. As the fluorescence of GFP is closely tied to the fully folded form of the protein it is often used as a model for protein folding. It has been seen that the fluorescence is recovered much more quickly in the case of fully-mature protein which has been denatured *in vitro* vs. that of protein which has never attained the correct fold and therefore never matured (42). The *in vivo* folding kinetics have been shown to be much longer than refolding kinetics (42). It is hypothesized that the oxidation step is the slow step in the *in vivo* chromophore maturation kinetics (43,44).
1.6. The M2 Transmembrane Domain

The second protein in the fusion described in this document is the transmembrane domain of the M2 protein, from influenza A virus. The selection of the M2 transmembrane domain was serendipitous when we discovered that fusing of the peptide to the C-terminus of GFP resulted in a ‘dark’ fusion protein. The function of this protein is to selectively channel protons. Mechanisms for transporting and channeling protons through native membranes exist in nature. Several of these proteins are large integral membrane proteins that contain several domains. M2 is a four helix bundle, the transmembrane domain of which has been shown to be responsible for proton transport (45). M2 also has a very high selectivity for protons over other ions and water (46). In order to fully understand the role that the M2 peptide is playing in the quenching mechanism of CA-GFP, it is important to recognize what is known about the structure and properties of this transmembrane domain.

This region has been extensively studied by several groups in order to determine M2’s structure and mechanism. (47-49). M2 is a proton channel that has been shown to be essential for the infection of the influenza virus. It has two main functions in the viral life cycle. After the virus has been internalized by the host cell, it is delivered to the secondary endosome. The mildly acidic pH of the endosomal compartment activates the M2 channel, which permits protons to flow into the virion interior, decreasing the pH.
within the virion. This allows dissociation of the coat proteins, which can then release the viral DNA for entry in the nucleus (Fig. 1.7). The second function of M2 is at a later stage of the virus life cycle, when it is used to reduce the pH of the Trans-Golgi Network (TGN). This function is necessary to increase the pH in vesicular compartments in order to protect acid sensitive proteins prior to virus assembly (50,51). The transmembrane region of this channel has been widely investigated as a possible drug target for the flu since inhibition of the channel should prevent viral replication (52,53).

The full-length M2 protein is 97 amino acids long with a 25-amino acid transmembrane region located between a shorter N-terminal extracellular sequence, and a larger intracellular C-terminal sequence. The transmembrane region forms a four-helix bundle in which the α-helices pack with a 30-37° crossing angle (49,54) (Fig.1.8). This transmembrane domain is studied independently of the full-length protein in most studies and synthetic versions of the peptide have been seen to tetramerize and function as a proton channel in this truncated form. There are currently several structures of the M2 transmembrane domain both with and without inhibitors amantidine and rimantadine (Fig. 1.8). Two residues, His37 and Trp41, located at the C-terminus of the peptide are responsible for the ‘gating’ of the proton channel’s function. It is widely accepted that when the four His37’s in the bundle are protonated they cause a

![Figure 1.8. Structure of the M2 transmembrane channel at pH 7.5-8 (PDBID: 2RLF) left, at pH 6.5 (PDBID: 3LBW) center, and at pH 5 with inhibitor rimantadine bound (PDBID :3BKD) right. The gating tryptophan residues are shown as spheres. (Figure generated by SN using Pymol)](image)
conformational change in the nearby Trp41’s at the base of the channel (50,55). This conformational change causes the channel to go from a ‘closed’ to an ‘open’ conformation which allows protons to flow through the channel (Fig. 1.8) (55-57). Another proposed mechanism is that the helices themselves shift to allow protons to pass in the open confirmation (47). Still another is that the channel itself is large enough to hold water molecules and that the conformational change in the tryptophans allow only the protons to be passed via a water bridge, but not entire water molecules (47). A recent analysis of crystal structures of the pore at increasing pH values suggests that the mechanism is based on both a conformational rearrangement of the helices as well as proton movement through the pore (58). Kinetic data on the M2 current can fit either Grotthus (proton hopping) or proton shuttling mechanisms, adding to the debate (59,60).

The extensive characterization of both the M2 transmembrane domain as well as GFP in terms of their structure have been pivotal in understanding the underlying mechanism of the dark state of CA-GFP. Throughout this thesis we show the variety of potential applications of this unique dark-to-bright reporter (chapter I), our model for the quenching mechanism of the dark state (chapter II), and our efforts to expand upon and understand the limits of this reporter for broader applications (chapters IV and V). Our goal is that through a detailed understanding of the mechanistic details of CA-GFP we are able to better engineer a reporter which can be used not only for deconvolution of individual caspase roles in the apoptotic pathway, but be useful in monitoring the activity of any enzymatic process, such as phosphorylation and ubiquitination, in the cell.
References

CHAPTER 2

CA-GFP: A GENETICALLY ENCODED DARK-TO-BRIGHT REPORTER FOR CASPASE ACTIVITY

This chapter was published in part as: Nicholls, S.B., Chu, J., Abbruzzese, G., Trembley, K.D., and Hardy, J.A., 2011. “Mechanism of a Genetically Encoded Dark-To-Bright Reporter for Caspase Activity.” The Journal of Biological Chemistry, Vol. 286, No. 28, 24977-24986. SBN performed all purifications and biochemical assays as well as experiments in bacterial cells and lysate. JC performed experiments and fluorescent microscopy in mammalian cells. GA performed FACS experiments and analysis.

Abstract

Fluorescent proteins have revolutionized modern biology with their ability to report the presence of tagged proteins in living systems. Although several fluorescent proteins have been described in which the excitation and emission properties can be modulated by external triggers, no fluorescent proteins have been described that can be activated from a silent dark state to a bright fluorescent state directly by the activity of an enzyme. We have developed a version of GFP in which fluorescence is completely quenched by the appendage of a hydrophobic quenching peptide. The fluorescence can be fully restored by catalytic removal of the quenching peptide, making it a robust reporter of proteolysis. We have demonstrated the utility of this uniquely dark state of GFP as a genetically encoded apoptosis reporter that monitors the function of caspases, which catalyze the fate-determining step in programmed cell death. Caspase Activatable-GFP (CA-GFP) can be activated both in vitro and in vivo, resulting in up to a 45-fold increase in fluorescent signal in bacteria and a 3-fold increase in mammalian cells. We used CA-GFP successfully to monitor real-time apoptosis in mammalian cells.
2.1 Introduction

The ever-growing palette of fluorescent proteins has arguably become the most widely used set of tools in cell and developmental biology. These fluorescent proteins act as markers for visualization of cellular processes, localization, gene expression patterns, and protein function. A spectrum of fluorescent proteins are now available, further expanding the multi-factorial imaging possibilities [for review see Ref. (1)]. The variously colored, split and photo-activatable fluorescent proteins [for review see Ref. (2)] are excellent tools for precise labeling and tracking of a variety of targets in living systems, but there is still a need for tools that can report on the functional state of the protein they interrogate. Many proteins are only activated by a post-translational event, so genetically encoded fluorophores that undergo a change in fluorescent properties in direct response to enzymatic action would meet this need. A dark-to-bright fluorescent reporter would be especially useful for reporting enzymatic activity, as the low intrinsic background would allow robust detection of even relatively rare events. Such a reporter could enable the same level of real-time, noninvasive, longitudinal studies of enzymatic activity as parent fluorescent proteins have enabled on expression and localization.

In an effort to develop and characterize a dark-to-bright activatable GFP, we sought to apply our reporter to a complex biological pathway that would allow thorough characterization of the quenching mechanism and kinetics of activation both in vitro and in vivo. We ultimately selected apoptosis, the process of programmed cell death. Apoptosis is essential for embryonic patterning and vertebrate development at all stages and is causally involved in up to half of human diseases lacking suitable treatment (3).
Thus, methods for monitoring apoptosis longitudinally would provide a full temporal understanding of the contributions of apoptosis to tissue formation, remodeling, response to drug treatment, and disease progression.

Apoptosis can be activated through an intrinsic caspase-9/apoptosome-dependent pathway or through an extrinsic, caspase-8/DISC complex-dependent pathway. Caspase-8 and -9 are apical initiators in the proteolytic cascade cleaving and thereby activating the downstream executioners, caspase-3 and -7. Cleavage of caspase-3 or -7 removes a prodomain and cuts an intersubunit linker, generating a small and a large subunit from each half of the dimer. Activated executioner caspases then cleave a select group of substrates (4), sentencing the cell to death. Caspase activation is the fate-determining step in the irreversible onset of apoptotic cell death, so they offer the most accurate and sensitive enzymatic indicator of apoptosis. Caspases show exquisite specificity for cleaving after aspartic acid residues in well-defined recognition sequences. The recognition specificities of the executioner caspases have been extensively mapped \textit{in vitro} (5-9) and \textit{in vivo} (10,11). The canonical and widely used recognition sequence comprising the amino acids DEVD remains an appropriate and selective sensor of executioner caspase activity (12).

Executioner caspase activity can be quantified using fluorogenic peptide substrates containing a caspase-recognition sequence linked to a synthetic fluorophore, whose fluorescence is completely quenched prior to cleavage (13). The synthetic nature of these peptide reporters clearly precludes their expression in living systems. A great deal of effort to date has focused on developing genetically encoded fluorescent apoptosis reporters based on cleavage by caspases. Because of the pervasive need for monitoring
apoptotic cell death, several clever genetically encoded apoptosis reporters have been developed.

The largest class of genetically encoded apoptosis reporters uses fluorescence resonance energy transfer (FRET) pairs separated by caspase cleavage sites. These reporters monitor a change in the FRET intensity upon caspase cleavage (14-16). The multimodality reporter uses copies of the caspase-recognition sequence to link a fluorescent protein, a bioluminescent protein, and a positron emission tomography reporter, which can each be more readily detected after caspase cleavage (17). The Apoliner and ApoAlert™ reporters use nuclear localization sequences to target a fluorescent protein to the nucleus following caspase cleavage (18). Apoptosis in these cells can be visualized based on changes in fluorescent protein localization. Another reporter relies on destruction of green fluorescent protein (GFP) using a tag for ubiquitin-independent degradation, which can be proteolytically removed by active caspase (19). The continued development of apoptosis reporters suggests that none to date have fully optimized brightness, sensitivity, and reporting mechanism.

An ideal apoptosis reporter would be genetically encodable, yet function in vitro, have very low background and high signal, consist of a single activatable fluorophore requiring no cofactors or additional biological processes, and be amenable to use in whole organisms, cells, microscopy, and flow cytometry. The caspase activatable-green fluorescent protein we have developed meets all of these criteria. The dark-to-bright transition in direct response to enzymatic activation constitutes a new class of fluorescent reporter.
2.2 Results

2.2.1 Design of CA-GFP

Within the goal of developing a new dark-to-bright version of GFP, we sought an application area that would provide stringent activity requirements and a useful and interesting framework for characterization of the quenching and dark-to-bright transition both \textit{in vitro} and intracellularly. We selected the proteolytically controlled pathway of apoptotic cell death as a first application for the dark-to-bright version of GFP.

Fusion of a variety of proteins has been observed to silence GFP fluorescence (20) which led us to design a fluorescent reporter based on this quenching principle (Fig. 2.1). We discovered that fluorescence of GFP (S65T) is quenched completely (0.9–1.5\% of GFP fluorescence remained) when our first candidate, a 27-amino acid peptide derived from the tetrameric proton channel domain of influenza M2 protein, was fused to the GFP carboxy terminus (Fig. 2.1). This full quenching proved to be an important component in maximizing sensitivity. To make a caspase-activatable GFP (CA-GFP) the caspase recognition sequence (DEVD) was inserted between GFP and the quenching peptide.

\textbf{Figure 2.1.} CA-GFP consists of Green fluorescent protein (green) fused through a flexible linker containing the caspase-3 and -7 recognition sequence DEVD (blue) to a 27-amino acid quenching peptide on the C-terminus (gray). After cleavage of the peptide by active caspase GFP fluorescence is recovered, shown in the manually rendered models (bottom)
2.2.2. CA-GFP Can Be Activated and Observed in Cells

To assess the response of CA-GFP to active caspase, CA-GFP was expressed in bacterial cells in the presence or absence of full-length active caspase-7 or the catalytically inactive C186A mutant. Only co-expression with active caspase-7 resulted in a 45-fold increase in CA-GFP fluorescence (Fig. 2.2), suggesting that caspase activity is responsible for the fluorescence increase.

Fluorescence in cells and lysates was observed within 1 h and was bright after 18 h. In bacterial cells, zymogen activation of full-length procaspase-7 to mature (active, two-chain) caspase-7 occurs with similar kinetics (21), further supporting the conclusion that the increase in fluorescence is due to caspase proteolytic activity. The gain of GFP fluorescence in the presence of wild-type caspase-7 (Fig. 2.2 top) correlated with cleavage of CA-GFP by caspase.

**Figure 2.2.** CA-GFP becomes fluorescent only upon cleavage by active caspase. Cleavage and fluorescence of bacterially co-expressed CA-GFP with active (WT) or inactive (C186A) caspase-7. Full-length caspase-7 undergoes auto-zymogen processing to generate the active, two-chain form of caspase-7. Mutation of the catalytic cysteine in the C186A mutant yields the inactive single chain pro-caspase-7 zymogen. Fluorescence of CA-GFP in lysates is shown on top; there is a 45-fold increase in fluorescence in the presence of active caspase-7. Immunoblot probed with anti-caspase-7 large subunit antibody is shown at center with an immunoblot probed with anti-GFP antibody shown at the bottom. CA-GFP is only fluorescent after cleavage in the presence of active caspase.
Figure 2.3. CA-GFP fluorescence is robust enough for fluorescence microscopy and flow cytometry. BL21 (DE3) E. coli cells expressing (A) CA-GFP alone or (B) CA-GFP with inactive caspase-7 C186A or (C) CA-GFP with active wild-type caspase-7. Differential interference contrast (DIC) images (left). Fluorescent microscopic images (center) of live cells. Intact cell flow cytometric analysis of GFP fluorescence (right) listing the fraction of cells with GFP fluorescence. Scale bars represent 10 µM.

CA-GFP expressed with wild-type caspase-7 yielded cells that were visibly green (Fig. 2.3C). These populations are also distinguishable by flow cytometry. 90% of cells expressing CA-GFP with active caspase-7 are fluorescent, whereas fewer than 1% are green in the absence of caspase activity.

2.2.3. CA-GFP is an apoptosis reporter in mammalian cells

While CA-GFP responds robustly to the activity of human caspase expressed...
heterologously in bacteria, perhaps the most significant application for CA-GFP is as a reporter of apoptosis in mammalian cells, where caspase activity governs apoptotic cell death. Treatment of CA-GFP-transfected NIH 3T3 cells with staurosporine (STS), a documented apoptosis inducer (22-24), resulted in a time-dependent change in cell morphology, which is consistently observed during apoptosis (Fig. 2.4A). A concomitant increase in cellular CA-GFP fluorescence was also observed as expected.

**Figure 2.4.** CA-GFP is fluorescently activated in mammalian cells undergoing apoptosis. A, fluorescent response of CA-GFP-transfected NIH 3T3 cells at indicated times after induction of apoptosis with staurosporine (STS). Scale bar represents 50 µm. Br: Brightfield microscopy. G, GFP fluorescence. B, CA-GFP cleavage after induction of apoptosis is observed as a function of time by immunoblotting with an anti-GFP antibody. C, appearance of cleaved caspase-3 was observed after induction of apoptosis by immunoblotting with an anti-caspase-3 antibody. Tubulin was probed as a loading control. D, fraction of transfected cells that are fluorescent (white bars) and fraction of cleaved CA-GFP (black bars) in cells induced to undergo apoptosis. *, significance level; p < 0.05 relative to zero time point.
Cleavage of CA-GFP was monitored by Western blot for GFP at various times after induction of apoptosis. A smaller fragment of cleaved CA-GFP appeared as a function of time after treatment with staurosporine (Fig. 2.4B). Cleavage of CA-GFP was consistent with the appearance of cleaved (active) caspase-3 (Fig. 2.4C). The observation that both GFP fluorescence and CA-GFP cleavage increase during apoptosis (Fig. 2.4D) indicates that CA-GFP functions as a mammalian apoptosis reporter on a population level. In control cells, increases in fluorescence, CA-GFP cleavage and the appearance of active (cleaved) caspase-3 were all correlated suggesting that CA-GFP is sensitive enough to report on even low levels of active caspases.

To uncover the details of CA-GFP expression and activation in mammalian cells, we performed time-lapse imaging of single cells expressing CA-GFP. To afford an internal, optically distinct fluorescent control, we generated the expression construct CA-GFP-IRES-mLumin that constitutively expresses both CA-GFP and the red fluorescent protein mLumin, which is a brighter, more photostable derivative of mKate (25). After addition of staurosporine to induce apoptosis, GFP fluorescence increased markedly (Fig. 2.5A). CA-GFP begins to be activated immediately and appears to remain localized in the cytoplasm for 3 h until the nuclear membrane is permeabilized. This is consistent with previously reported executioner caspase activity, which is cytoplasmic in the early stages of apoptosis (26), however, this is the first observation of the direct relationship of caspase activity to the loss of the nuclear membrane integrity. Notably, the red fluorescence of mLumin is present throughout the cell, including inside the nucleus prior to apoptosis. In untreated control cells, green and red fluorescence increased minimally and to approximately the same degree over the time course of the observation (Fig. 2.5B).
Red fluorescence also increases during apoptosis, due to shrinking of the volume of the
Figure 2.5. The response of CA-GFP can be measured in single cells undergoing apoptosis. *A*, time-lapse confocal images of NIH 3T3 cells co-expressing CA-GFP and mLumin were recorded at the indicated times following treatment with STS (*A*) to induce apoptosis or without STS treatment as a control (*B*). DIC; differential interference contrast images showing changes in cell morphology. *G*, green channel monitoring CA-GFP fluorescence. *R*, red channel, monitoring mLumin (control) fluorescence. Scale bar represents 25 µm. *B* and *C*, ratio of green (GFP)/red (mLumin) fluorescence for (*C*) untreated control or (*D*) staurosporine-treated cells. Ratios for six (*C*) or four (*D*) independent cells were measured. The G/R ratio of cells 1 and 2 shown in panel (*A*) are indicated as 1 or 2 in plot (*D*).
cell, which is a hallmark of apoptosis (27). Quantification of the green and red signals showed that only CA-GFP responds to apoptotic stimuli.

The ratios of green/red fluorescence for the untreated control cells (Fig. 2.5B) and for the staurosporine-treated cells induced to undergo apoptosis (Fig. 2.5A) were quantified. The appearance of GFP fluorescence in apoptotic cells appeared with related kinetics in all cells, however, cell-to-cell variations were observed. The variation in the kinetics of GFP fluorescence occurred in a manner that is consistent with cell-to-cell differences in morphological changes and cell death kinetics. One of the apoptotic cells observed had very low fluorescence, which may be due to lower efficiency of transfection in that cell (the lowest green/red ratio Fig. 2.5C and 2.5D). The ratio of green/red fluorescence in the apoptotic cells was 3-fold higher than in control untreated cells, indicating that CA-GFP responds to apoptotic signals with a 3-fold increase in signal over background.

### 2.2.4. CA-GFP in whole animal models

As mentioned previously our ideally designed reporter would be a useful tool not only in vitro, in bacterial systems, in mammalian systems, and in FACS sorting, but also in whole animal models. One animal model which has been well studied in terms of developmental apoptosis and is growing in popularity in drug screening applications is the zebrafish model. Zebrafish are a useful model because of their transparency, allowing for effective microscopy throughout the development of the embryo, as well as their rapid development. Zebrafish are also a very well characterized model system, having a fully sequenced genome as well as extensive mapping of the early stages of development. We
first began by injecting DNA constructs of GFP (S65T) and CA-GFP into zebrafish embryos. After extensive trials of STS treatment of wild-type (uninjected) and injected embryos varying STS concentrations and the developmental stage of treatment, it was determined that STS treatment was too harsh for the injected embryos. As the developmental apoptosis map is well mapped in zebrafish (28) it was determined that areas known to undergo apoptosis at specific life cycle stages would be observed for fluorescence in GFP, CA-GFP, and buffer injected control embryos. While there was some observable difference in CA-GFP vs. the two controls the expected punctate pattern of apoptosis was not observed.

To address the issue of mosaicity in expression our collaborators, the lab of Gerald Downes at UMass Amherst in the Biology department, generated a new construct for DNA injection, a Tol2-CMV-
CA-GFP and Tol2-CMV-GFP, as well as constructs for each CA-GFP and GFP for generating capped and poly-A-tailed mRNA for injection. They also determined that treatment of the zebrafish with 30 mM hydroxyurea (HU) triggered effective widespread apoptosis within the embryos. After injection of embryos with the Tol2-CMV constructs, fluorescence images of treated and untreated organisms did not appear to show a difference in the treated or untreated embryos as the fluorescence signal was weak (Fig. 2.6). Fluorescence was quantified over the area of the embryo to show a 10-fold increase in fluorescence in treated embryos over the untreated control (Fig. 2.7). Though these results are encouraging as to the success of CA-GFP in zebrafish it appears as though CA-GFP is not effective in transient expression assays. To be effectively used for single cell studies the signal-to-noise ratio needs to be improved. Efforts to generate a transgenic line of zebrafish stably expressing CA-GFP are ongoing.

2.3 Discussion

CA-GFP is directly activated by caspase activity and no other cellular components are essential. On the other hand, one possible reason for the more effective fluorescent conversion of CA-GFP in bacterial cells over other systems may be the presence of additional cellular factors, such as bacterial chaperone proteins. In mammalian cells CA-GFP conversion to the bright state can be seen in 20–40 min, which is consistent with the timing of caspase activation and the irreversible initiation of apoptosis (29), suggesting that some cellular factors (potentially macromolecules or metabolites) accelerate the maturation process over in vitro conditions where maturation of purified protein is slower (60 min). Therefore, although some maturation time is
required, CA-GFP still is able to accurately report caspase activity during apoptosis.

Similarly in split GFP, association of the two halves of the protein is not the rate-limiting step but rather subsequent maturation of the chromophore dictates the kinetics of the appearance of fluorescence (30). The maturation of CA-GFP in vitro is similar to that of split GFP (30), which has also proven to be an incredibly useful technology.

CA-GFP is unique among fluorescent protease reporters in its transition from a quenched dark state to a bright state upon cleavage. This type of reporter was inspired by small molecule fluorogenic substrates that transition from dark to bright fluorescence upon protease activity. The utility of a dark-to-bright transition stems from its low fluorescent background and use of a single fluorophore. The caspase-mediated increase in CA-GFP fluorescence is 45-fold in bacteria compared with the next most sensitive reporter, CyPet-YPet, in which a 3.3-fold increase in FRET fluorescence is observed (30). This gain in fluorescence makes CA-GFP readily applicable to protease profiling in bacteria and for engineering protease specificity by directed evolution methods using sorting by flow cytometry. In mammalian cells, the increase in CA-GFP fluorescence is also higher than other apoptosis reporters. CA-GFP fluorescence increases 3-fold upon induction of apoptosis and activation of caspases, whereas the next two strongest reporters give 0.6-fold (19) and 2-fold (17) increases in fluorescence upon caspase cleavage.

The response of CA-GFP to caspases is not reliant on any other biological processes, and is therefore suitable for monitoring caspase activity under any cellular conditions. Other classes of caspase reporters rely on nuclear localization of fluorescent proteins (18,19) or on proteosome-based degradation and thus are dependent on processes
such as nuclear translocation, the integrity of the nuclear pores and membrane or proper functioning of the proteosome to observe apoptosis. During apoptotic cell death all biological processes ultimately shut down, potentially impacting the utility of reporters that rely on other biological pathways. The fact that localization of CA-GFP is not essential for monitoring caspase activity also allows CA-GFP to report on the integrity of the organellar membranes, as we observed for the first time using our genetically-encoded reporter and time-lapse imaging. In mammalian cells CA-GFP has proven to be an effective tool for time-resolved observation of subcellular events in apoptosis. The fact that CA-GFP shows no fluorescence in the nucleus in early apoptosis suggests that it is working accurately. The executioner caspases have been shown to localize in the cytoplasm in the early stages of apoptosis, only entering the nucleus in later stages after the nuclear membrane has been compromised (26). Caspases have been shown to target two of the twenty nuclear pore components, Nup93 and Nup96 (31). Once the nuclear pore is compromised, larger proteins such as caspases and CA-GFP tetramers are able to pass into the nucleus. CA-GFP requires neither translocation nor additional cofactors and thus is useful for a wider array of applications than other reporters. These include monitoring apoptosis in intact living organisms without the need for sectioning or high-resolution microscopy and flow-cytometric sorting of apoptosing cells from healthy cells.

The fact that several caspase-reporting technologies have been developed prior to CA-GFP reflects the importance of apoptosis in modern biomedical research and the demand for new, better apoptosis reporting methods. CA-GFP combines dark-to-bright fluorescence, the most positive property of the synthetic peptide-based fluorophores, with the important advantage of being genetically encoded, to produce a new class of
apoptosis reporter. We recognize that the CA-GFP platform is also amenable to substitution of both the protease recognition site and the fluorescent protein leading to reporters for many proteases in a range of colors (See Chapter IV). A panel of CA-GFP derivatives of various colors and with various protease sensitivities would enable simultaneous imaging of a number of proteolytic cascades. It is also possible to envision engineering reporters based on the silenced state of GFP that respond to the action of other enzyme families. As we move forward with zebrafish and ultimately mouse models incorporating CA-GFP we hope to create an important contribution to the study of not only developmental apoptosis in organisms, but a tool for in vivo drug screening and toxicity assessment.

2.4. Materials and Methods

2.4.1. Molecular cloning

The gene for dark GFP (S65T) (referred to as GFP from this point forward) was created by PCR, amplifying the GFP gene using a reverse primer that also encoded the 27 amino acid quenching peptide derived from the transmembrane region of influenza M2. This PCR product was ligated into the XhoI and NdeI sites of pET21b (Novagen). A sequence encoding the linker (LEVLFQGP) was then inserted between the GFP and M2 genes using a site-directed mutagenesis using overlapping inverted primers and amplification of the entire plasmid similar to the QuikChange® (Agilent) approach. The newly inserted linker sequence was then mutated to encode the caspase-7 cleavage recognition site (DEVDFQGP). The final sequence of CA-GFP protein is GFP with the fusion of DEVDFQGPCNDSDDPLVVAASIIGILHLILWILDRL at the C-terminus. This construct
was used for expression and purification of CA-GFP. The CA-GFP gene was then
amplified by PCR and ligated into NdeI and XhoI of the pBB75 vector (a gift from
Adrien Batchelor), which has a p15 origin of replication, different than the pET family of
vectors which have CoLE1 origins of replication. This plasmid was used for co-expression
of CA-GFP with caspase-7. The mammalian expression vectors pCA-GFP and
pGFPSTOP were constructed by insertion of the CA-GFP or GFP, obtained by PCR into
the NheI and EcoRI sites of pmKate2-C (Clontech, Palo Alto, CA, USA). CA-GFP from
pCA-GFP was isolated and ligated into the NheI and EcoRI sites of pT-CD8- IRES
SCN2B (a kind gift from Dr. Alfred George, Vanderbilt University) to produce pT-CA-
GFP-IRES-SCN2B. The resulting plasmid was cut with SalI and NotI and the coding
sequence for mLumin (25) (mKate2-S158A) was inserted, generating pTCA- GFP-IRES-
mLumin. All constructs were confirmed by sequencing.

### 2.4.2. Lysate-based GFP fluorescence assay

CA-GFP was cotransformed with either wild-type caspase-7 or the C186A caspase-7
mutant into the BL21(DE3) strain of *Escherichia coli*. 50-mL cultures were inoculated
from a 5 mL dense overnight culture, allowed to grow to an OD$_{600}$ of 0.6 at 37 °C and
induced with 1 mM IPTG for 18 h at 25 °C. One mL of the culture was centrifuged using
a tabletop centrifuge, the media was decanted and the pellet resuspended in 400 µL of a
lysis buffer (0.5 mg/mL lysozyme, 2 units DNAase). The suspension was lysed by freeze-
thaw and the lysate was then centrifuged using a tabletop centrifuge. 100 µL of the
supernatant was analyzed on a Molecular Devices Spectramax M5 spectrophotometer,
measuring fluorescence (Ex. 475 nm/Em. 512 nm) in a costar 96-well flat bottom black
plate.

2.4.3. Immunoblotting of bacterial cultures

Aliquots of the cultured and lysed CA-GFP/caspase-7 expressing *E. coli* samples described above were reserved for immunoblotting analysis. SDS-loading buffer was added to lysate samples and the samples were boiled for 10 min. Three identical SDS-PAGE minigels were run. One gel was stained with Coomassie dye and imaged with a Gel Doc work station (Syngene). A second gel was transferred to a nitrocellulose membrane and immunoblotted using rat monoclonal anti-mouse caspase-7 antibody (Sigma) that is specific for the caspase-7 large subunit. The third gel was also transferred to a nitrocellulose membrane and immunoblotted using a mouse IgG monoclonal anti-GFP primary antibody (Abgent). Gels immunoblotted with either anti-caspase-7 or anti-GFP primary antibodies were treated with anti-mouse IgG alkaline phosphatase produced in goat (Sigma) and visualized using 1-StepTM NBT/BCIP (Thermo Scientific).

2.4.4. Fluorescence microscopy

To prepare samples of *E. coli* for imaging, 50 mL LB cultures of CA-GFP, CA-GFP co-transformed with inactive caspase-7, and CA-GFP co-transformed with wild-type caspase-7 were inoculated from dense 5 ml overnight cultures and grown to an OD$_{600}$ of 0.6. They were induced with 1 mM IPTG and grown 16–18 h at 25 °C. From these cultures 1-mL aliquots were spun down for 1 min at 13,000 rpm in an Eppendorf tabletop centrifuge. The media was decanted and the pellet was resuspended in PBS. Slides were prepared from these samples for DIC and fluorescence images taken using a Nikon
Spot/E600 fluorescent microscope. All fluorescence images used the same exposure time of 150 ms.

### 2.4.5. Flow cytometry

Flow cytometric analysis was performed on a Becton Dickinson (BD) LSR II configured with a 488 nm blue laser and a 530/30 nm bandpass filter. Samples were prepared by harvesting 300 µL of cultures expressing CA-GFP with or without wild-type or C186A caspase-7, induced overnight as described above. Cells were washed twice with PBS and resuspended at a final concentration of 5x10^7 cells/mL. With BD FACS Diva Software, measurements were recorded for 10,000 cells analyzed at a rate of 200 events s^-1. A gate was set to define a target population of fluorescent cells by including only events with a higher GFP intensity than cells expressing CA-GFP alone (uncleaved CA-GFP).

### 2.4.6. Cell culture and transfection

NIH 3T3 cells were obtained from ATCC and cultured in Dulbecco’s modified Eagle’s medium (DMEM) supplemented with 100 units/ml penicillin, 100g/ml streptomycin, 2mM glutaMAX and 10% fetal bovine serum (Atlanta Biologicals). Cultures were maintained in a humidified atmosphere of 5% CO₂ in air at 37 °C. For transfection, cells were plated in 12-well plates or 35-mm dishes (70– 80% confluence), cultured for 12 h and transfected with 1.5 g of plasmid DNA per 35mm well using 6l Fugene HD (Roche) according to the manufacturer’s instructions. For induction of apoptosis, cells were incubated with 1 M staurosporine (STS, Ascent Scientific NJ) for a period of time as indicated in “Results.” Controls were treated with vehicle alone (0.05% DMSO).
2.4.7. Immunoblotting of mammalian cultures

NIH 3T3 cells were transiently transfected with the indicated expression vectors and treated with 1 M staurosporine (STS) or left untreated. Cells were harvested at the indicated time points and solubilized in loading buffer (62.5 mM Tris-HCl pH 6.8, 2% SDS, 41.7 mM dithiothreitol (DTT), 10% glycerol, 0.01% bromphenol blue). After being boiled for 5 min, proteins were resolved on a 5–20% gradient gel or 12% homogenous SDS-PAGE gel and electrotransferred to PVDF membranes for immunoblotting using enhanced chemiluminescence detection. Primary antibodies against GFP (1:5000), caspase-3 (1:1000), and -tubulin (1:5000) were applied. Mouse monoclonal anti-GFP (Clone 3F8.2) and anti-tubulin (Clone AA2) antibodies were purchased from Millipore, rabbit polyclonal anti-caspase-3 antibody from Cell Signaling (San Diego, CA) and HRP-conjugated goat anti-mouse IgG and goat anti-rabbit IgG from Jackson ImmunoResearch Labs (West Grove, PA).

2.4.8. Live cell imaging

Epifluorescence was performed on an inverted wide-field fluorescence microscope (TE2000; Nikon, Japan) equipped with a cooled CCD camera (Qimaging, Canada) and a Plan Fluor 10/0.30 objective using Elements imaging software (Nikon). GFP fluorescence was captured using a 515/30 filter. Time-lapse, confocal imaging was carried out on a Zeiss LSM 510 Meta Confocal Microscope (Zeiss, Germany) over a 6-h period. GFP and mLumin (20) were excited at 488 nm and 543 nm, respectively, and fluorescence detected at 520/30 nm (for GFP) and 585LP nm (for mLumin) using a 40
(1.3 N.A.) oil immersion objective lens. The live cells were maintained in 12.5 mM Hepes buffered DMEM and kept at 37 °C with a 5% CO₂ atmosphere using heaters and a CO₂ regulator. Images from the confocal microscope were manually focused and captured every 20 min using LSM510 software (version 4.2; Zeiss). All fluorescence images were corrected by subtracting the background and analyzed using ImageJ and Adobe Photoshop 9.0. The ratio of green fluorescence intensity at each time point (IGt) was compared with the red fluorescence intensity (IRt) corrected for the initial intensity green:red ratio, IGt/IRt:IG0/IR0. The ratio was calculated based on the average pixel value for the whole cell, wherein IGt and IRt are the fluorescence intensities of GFP and mLumin at the indicated time point. As cells changed their morphology during imaging, the whole cell region was determined separately in each image.

2.4.9. Statistical analysis for live cell imaging

Manual cell counting was performed with ImageJ software using the Cell Counter plugin. More than 1000 cells were counted for both treatment and control groups. Statistics were calculated using Excel. Statistical significance was assessed by an unpaired Student two-tailed t test. Values were considered statistically significant at p 0.05.

2.4.10. Zebrafish injections of CA-GFP

To generate the DNA constructs for the initial DNA injections the genes for CA-GFP and GFP were amplified using PCR and ligated into the ApaI and EcoRI restriction sites of the pBluescript vector (Stratagene). The Tol2-CMV-GFP and Tol2-CMV-CA-GFP
constructs as well as the constructs to generate the capped and poly-A-tailed mRNA were generated in the Downes lab. Zebrafish were injected at the 1-4 cell stage and incubated at 25°C overnight until embryos reached the 24-27 hours past fertilization (hpf) stage. Embryos which were to be treated were then decorionated and treated with either STS (in initial trials) or 30 mM hydroxyurea to induce apoptosis. Embryos treated with STS were then moved to a neutral buffer for an additional incubation period and then observed. Fluorescence was quantified over the area of the embryo using a confocal microscope.

References

CHAPTER 3

MECHANISTIC INSIGHTS INTO THE CA-GFP DARK STATE


Protein Science. SBN performed all experiments in this chapter.

Abstract

Apoptosis is a critical process involved not only homeostasis and development but a wide variety of disease states as well. Caspases are the ultimate executors of the programmed cell death pathway. As caspases play such a central role in this widely studied process there is demand for technologies with which to study caspase activity. Our reporter of caspase activity, caspase activatable-GFP (CA-GFP), is unique in its advantageous ‘dark’ state, where chromophore maturation of the GFP is inhibited by the presence of a C-terminal hydrophobic quenching peptide. Here we show not only that the lack of fluorescence is due to an immature chromophore, but also that CA-GFP does not fold into the robust β-barrel of GFP until the peptide has been cleaved by active caspase. Both CA-GFP and GFP₁₋₁₀, a truncated version of the wild-type protein lacking the final strand of the barrel, appear to have similar secondary structure, which differs from that of the mature GFP. A similar susceptibility to proteolytic digestion indicates that this shared structure is not the robust, fully formed GFP β-barrel. We have developed a model that as CA-GFP is translated *in vivo* it follows the same folding path as wild-type GFP, however the presence of the peptide does not allow for CA-GFP to form the barrel of the fully matured GFP. CA-GFP is therefore held in a ‘pro-folding’
intermediate state until the peptide is released, allowing it to continue folding into the mature barrel geometry. The understanding of the structural basis of the dark state of the reporter will enable manipulation of this mechanism in the development of reporter systems for any number of cellular processes.

3.1 Introduction

In the previous chapter (chapter II) our reporter of caspase activity, CA-GFP, was introduced. This reporter is composed of the green fluorescent protein (GFP), the caspase recognition sequence, DEVD, and a quenching peptide derived from the tetramerization domain of influenza M2 protein (Fig. 3.1). When CA-GFP is expressed under conditions where active caspases are present, CA-GFP is cleaved and the quenching peptide is released and fluorescence is gained. Due to its response to caspases and the dark state prior to cleavage, CA-GFP has been a robust reporter of caspase activity. In bacteria heterologously expressing caspases, we observe a 50-fold increase in fluorescence after caspase cleavage(1). In mammalian systems, where low levels of active caspases are present constitutively, the fluorescent background is higher, and we observe a 3-fold increase in fluorescence upon induction of apoptosis. Although this increase in fluorescence is lower than in bacterial systems, it is still superior to any other genetically-encoded caspase reporter systems.
Caspases are cysteine proteases that control a variety of biological cascades including apoptosis, inflammation and neurodegeneration. Development of the caspase reporter was motivated by the fact that caspases play central roles in many processes in development and human disease. Monitoring caspases contributes to our ability to observe and understand apoptosis in the context of organismal development and predicting drug toxicity mediated by apoptosis. Although a number of caspase reporters had been developed previously, the genetically encoded characteristic of CA-GFP makes it amenable to a number of applications that cannot be fulfilled by existing reporters. The most striking characteristic of CA-GFP is its profoundly dark state prior to caspase cleavage. As a consequence the signal-to-noise for CA-GFP is strikingly improved compared to other previously reported caspase reporters.

The goal of this chapter is to understand the underlying structural implications of the peptide fusion which leads to the dark state of CA-GFP. Understanding the dark state of CA-GFP would enable us to improve the response time of CA-GFP, optimize the fluorescent increase upon cleavage of CA-GFP, and finally enable us to design reporters for other enzymatic processes.

3.2 Results

The most broadly applicable attribute of CA-GFP is the fully quenched nature of the uncleaved state. To fully harness the potential of CA-GFP for other uses, it is essential to understand the structural basis of the dark state at a mechanistic level. The goal of this work is to assess the overall folded state of CA-GFP to determine what structural factors prevent fluorescence of GFP prior to removal of the peptide.
3.2.1. CA-GFP requires no cofactors and can be activated \textit{in vitro}

To better understand the mechanism by which CA-GFP is activated, and to determine whether activation of CA-GFP was dependent on some intracellular components, cleavage of CA-GFP \textit{in vitro} was tested. Purified active caspase-7 and purified CA-GFP were incubated together in a minimal buffer and cleavage of CA-GFP by wild-type caspase-7 was observed as a function of time (Fig. 3.2A). During the cleavage assay, a gain in fluorescence was also observed for CA-GFP (Fig. 3.2B). This suggests that the transition from the dark state to the brightly fluorescent state is dependent only on cleavage and not on any intracellular condition such as cotranslational cleavage or the activity of chaperones to refold GFP. Cleavage of CA-GFP preceded fluorescence by 60 min (Fig. 3.2C). The appearance of fluorescence upon cleavage of purified CA-GFP seems to be slower than in mammalian cells where CA-GFP fluorescence appears within 20 min after activation of caspases (chapter II). This lag in acquisition of fluorescence for purified

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure3.2}
\caption{CA-GFP can be activated \textit{in vitro}. \textit{A}, cleavage of CA-GFP by wildtype caspase-7 from 0 min to 18 h. (overnight). \textit{B}, gain of fluorescence during \textit{in vitro} cleavage of CA-GFP by caspase-7 (\textit{black line}) with error bars (\textit{gray}). \textit{C}, CA-GFP cleavage precedes appearance of fluorescence. Percentage of total fluorescence (\textit{white}) and cleaved CA-GFP (\textit{black}) at each time point.}
\end{figure}
CA-GFP is similar to the *in vitro* chromophore maturation time for purified GFP S65T, which occurs on the timescale of 27 to 122 min depending on the starting state of the protein (2,3).

### 3.2.2. Chromophore maturation is prevented in dark CA-GFP

CA-GFP cleavage and fluorescence kinetics suggest that the GFP chromophore is not mature in the dark state. The absorbance spectra of CA-GFP does not have the characteristic 488 nm peak of GFP (Fig. 3.3A), also suggesting that the CA-GFP chromophore has not undergone chromophore maturation. In GFP the chromophore is formed by a cyclization reaction in which the amide nitrogen of Gly-67 initiates nucleophilic attack on the carbonyl carbon of Thr-65, forming an imidazolone ring. Next the ring is oxidized (loss of 2 Da) to a cyclic imine and the carbonyl oxygen of Thr65 is dehydrated (loss of 18 Da), fully conjugating the system (Fig. 3.3B) (2,4). Thus the maturation state of the chromophore can be observed as a change in the mass of the protein (loss of 20 Da). We compared the mass spectra

---

**Figure 3.3.** CA-GFP silencing relies on prevention of chromophore maturation. *A*, absorbance spectra of GFP and CA-GFP suggest that CA-GFP chromophore is non-functional. *B*, Chromophore maturation is a post-translational reaction which occurs in three sequential residues at the center of the GFP barrel. *C*, mass spectra of GFP and CA-GFP tryptic fragments indicate that the CA-GFP chromophore has not undergone the maturation reaction.
of tryptic peptides from dark CA-GFP and mature, green GFP (Fig. 3.3C). The mature chromophore-containing fragment (2449.2 Da) was observed for GFP as expected. The partially reacted intermediate in which only ring oxidation has occurred was also observed (2467.2 Da). This direct observation of the oxidized intermediate is the first observation for an S65T intermediate and supports other mechanistic studies on maturation of other versions of GFP (5) in which the oxidation step occurs prior to dehydration. In dark CA-GFP only the unreacted, immature peptide (2469.3 Da) was observed, indicating that the CA-GFP quenching peptide prevents chromophore maturation. The gain in the intensity of circular dichroism signal is consistent with an increase in the multimeric state of the protein. The oligomeric state of CA-GFP, GFP, and cleaved CA-GFP were assessed by size exclusion chromatography (Fig. 3.4). The majority of the GFP and cleaved CA-GFP in solution is monomeric (Fig. 3.4). Uncleaved CA-GFP exists predominantly as a species which resembles a soluble trimer or tetramer, which is consistent with the oligomeric state of the parent quenching peptide. The apparent tetramers may have exposed hydrophobic patches, as some portion of the CA-GFP is also found in higher

![Figure 3.4](image)

**Figure 3.4.** Size exclusion chromatogram for GFP, CA-GFP and cleaved CA-GFP. The observed molecular weights demonstrate that GFP and cleaved CA-GFP are monomers, while uncleaved CA-GFP predominantly resembles a tetramer. The molecular weights for the standards are marked as diamonds. The expected and observed molecular weights by size exclusion chromatography for GFP, CA-GFP and cleaved CA-GFP are shown below.
order oligomer, presumably clusters of this smaller order oligimer. Based on this observation oligimerization even to small ordered oligomers such as trimer/tetramer of CA-GFP appears to be sufficient for quenching GFP fluorescence.

3.2.3. CA-GFP has secondary structure similar to GFP

Like native GFP, CA-GFP is soluble when over-expressed in bacterial cells (3), suggesting that the GFP -barrel domain of CA-GFP is globally folded. To assess the fold of CA-GFP in the dark state, purified CA-GFP and GFP were analyzed by circular dichroism spectroscopy. While the spectra of non-fluorescent CA-GFP and fluorescent wild-type GFP are not superimposable (Fig. 3.5), the similarity in the spectral features were coincident with published spectra (6) and suggest that CA-GFP forms the β-barrel observed in all GFPs. The fact that GFP is known to be a stable and well-folded β–barrel protein suggests that CA-GFP is likewise a folded protein despite the fact that the CD spectra are not superimposable, particularly at wavelengths greater than 225 nm. Several other factors also suggest that CA-GFP is in a folded state. CA-GFP is soluble in solution for an extended period of time and is not prone to aggregation-induced insolubility. CA-GFP can also be cleaved rapidly by active caspases, suggesting that it is not massively aggregated and therefore resistant to proteolytic cleavage. On the other hand, the fact that CA-GFP does form some high-order

Figure 3.5. Circular dichroism spectra of GFP and CA-GFP indicate that both proteins are folded into a predominantly β-sheet structure.
oligomers suggests that the stability of the protein has been compromised by the appendage of the quenching peptide. Thus, we sought to determine whether CA-GFP is folded into the β-barrel observed in mature wild-type GFP or whether CA-GFP forms a stable, partially-folded intermediate.

The folding pathway for GFP has been studied previously by a number of groups (7-13). It is now clear that folding of GFP occurs co-translationally and that proper folding into the GFP β-barrel only occurs once the entire protein (strands 1 through 11) has been translated (14,15). Despite the fact that all 11 strands must be available for proper folding of GFP, it is possible to cut GFP following the 10th strand to make split GFP (16). In the split-GFP system, co-expression of GFP strands 1-10 (GFP1-10) with the 11th strand allows reassociation of the 11th strand with GFP1-10 and folding of the GFP β-barrel (Fig. 3.6A). Once the full GFP β-barrel is formed the chromophore is competent to mature, and split GFP becomes fluorescent. Chromophore maturation in split GFP occurs with similar kinetics to CA-GFP (16,17). We hypothesized that CA-GFP might adopt a stable, partially folded structure similar to that of GFP1-10. Indeed, the overall shape of the GFP1-10 spectrum was also indicative of a

**Figure 3.6.** A. The structure of the GFP β-barrel (PDB ID:2YOG, green) with the chromophore residues shown as sticks. The 11th strand of the barrel, removed in the GFP1-10 construct is highlighted in blue. B. The CD spectra of GFP, CA-GFP and GFP1-10 are superimposed to allow assessment of the overall shape of the three spectra. The maximum intensities of the three spectra have been scaled for the relative concentration of each protein. The non-scaled molar intensities of these proteins can be seen in Figure 3.7.
predominantly β-strand structure (Fig. 3.6B). More importantly, the spectra for GFP$_{1-10}$ and CA-GFP are nearly superimposable indicating a high degree of similarity between the two that differs somewhat from mature GFP.

Figure 3.7. Circular Dichroism spectra. A. The CD spectra of GFP from 200-250 nm collected at five degree increments from 20-90°C. The 20°C spectra is shown in the darkest shading; the 90°C spectra in the lightest shading and intermediate temperatures are shown in increasingly light color as a function of temperature. Spectra are plotted as the mean residual ellipticity (MRE x 10$^3$: degree•cm$^2$•dmol$^{-1}$•number of residues$^{-1}$) as a function of wavelength B. The CD spectra of CA-GFP collected and shown as in A. C. The CD spectra of GFP$_{1-10}$ collected and shown as in A. D. The melting profile of GFP as monitored by CD signal at 214 nm plotted as a function of temperature. E. Melting profile of CA-GFP. F. Melting profile of GFP$_{1-10}$.

In addition to the shape of the CD spectra, the thermal stability of CA-GFP also provides insight into the structure relative to mature GFP and GFP$_{1-10}$. To assess stability, we collected spectra from 200-250 nm of each protein every five degrees as the temperature was raised from 20-90°C (Fig. 3.7A-C). As was the case with the CD spectra, the denaturation properties of CA-GFP are more similar to GFP$_{1-10}$ than to GFP. As the temperature was increased for mature GFP the molar ellipticity decreased. Previous work
has suggested that an increase in CD signal is a sign of aggregation (18). Thus it appears that as GFP becomes unfolded it is prone to aggregation. In the cases of both GFP1-10 and CA-GFP the signal very gradually increases as a function of temperature up to 85°C. After this temperature the signal precipitously drops, suggesting a significant aggregation occurring at the highest temperatures. The melting profile can be illustrated by plotting the molar ellipticity at a selected wavelength as a function of temperature. We observed the greatest temperature-dependent change in CD signal at 214nm, so we have reported those data, (Fig. 3.7D-E), but similar trends are seen at wavelengths > 210 nm. The measured melting temperature (T_m) for GFP of 76°C, closely agrees with published value of 78°C based on loss of fluorescence (17,19). The curves for both GFP1-10 and CA-GFP cannot be fit to a two-state melting model but show similar stability profiles to one another. These data suggest that CA-GFP adopts an ensemble of structures that have more similar thermodynamic properties to GFP1-10 than to mature GFP.

CA-GFP adopts an oligomeric state greater than a monomer and consistent with a trimer or tetramer by size exclusion chromatography. We aimed to compare the oligomeric state of CA-GFP to mature

**Figure 3.8.** The size exclusion chromatogram of GFP (MW 27.7 kDa), CA-GFP (32 kDa), cleaved CA-GFP (28.2 kDa), and GFP1-10 (20.5 kDa). GFP and cleaved CA-GFP elute at a retention volume consistent with a monomeric form of the protein. The dark CA-GFP elutes as two peaks with retention volumes consistent with a population that is 50% lower order oligomer (putatively trimer or tetramer) and 50% higher oligomer. GFP1-10 elutes in the void volume indicating a largely aggregated population
GFP and GFP_{1-10} as further insight into the organization of CA-GFP. As expected, GFP elutes as a monomer. Approximately 50% of CA-GFP being in a low order oligomeric state and 50% higher-order oligomer (Fig. 3.4). In contrast, GFP_{1-10} elutes in a single peak at the void volume of the Superdex 200 column (Fig. 3.8), indicating a highly aggregated state. The size exclusion chromatography analysis corroborates the observations from CD, suggesting that GFP_{1-10} exists in a higher oligomeric state than CA-GFP, which is more oligomeric than GFP. Thus, CA-GFP appears to have oligomerization properties that are intermediate between mature GFP and GFP_{1-10}.

We reasoned that structural studies of CA-GFP would be dramatically eased if we attained oligomerically homogeneous preparations of CA-GFP. We also isolated the low-order oligomeric form of CA-GFP following size exclusion chromatography and allowed it to re-equilibrate at room temperature for 1 or 18 hours. Following incubation CA-GFP had re-equilibrated to a mixture that contained both low-order and high-order oligomers in time-dependent manner. We tried extensively to identify detergent-containing conditions that would shift CA-GFP entirely to a homogenous low oligomeric state, but were unsuccessful in this pursuit. Despite this we performed extensive crystallization trials, reasoning that perhaps crystallization solely of the low oligomeric state could occur. These crystallization trials were also uniformly unsuccessful. Thus, we think it likely that CA-GFP does not exist in a single homogenous conformation in solution but is indeed a mixture of highly oligomeric species as observed by size exclusion chromatography.

3.2.4. β-Barrel of CA-GFP not fully formed
To further investigate whether the solution structure of CA-GFP more closely resembles the mature GFP or GFP\textsubscript{1-10}, we subjected all three proteins to limited proteolysis by proteinase K (Fig. 3.9A). Proteinase K cleaves proteins relatively non-specifically following aromatic and aliphatic amino acids and is thus a useful means of achieving nearly full proteolysis. The propensity of GFP\textsubscript{1-10} towards aggregation has been suggested to indicate the presence of a less ordered folding intermediate of GFP. We expect that such a folding intermediate should be relatively more susceptible to proteolysis. In contrast, fully folded GFP is notoriously resistant to proteolytic cleavage due to the stability of the $\beta$-barrel (20). As predicted, GFP was resistant to digestion up to 10 minutes while both CA-GFP and GFP\textsubscript{1-10} were nearly fully digested within one minute (Fig. 3.9A). This strong susceptibility to digestion suggests that the $\beta$-barrel structure, which renders GFP resistant to the proteinase K treatment, is not present in CA-GFP.

![Figure 3.9](image)

**Figure 3.9.** Protease susceptibility of GFP, CA-GFP and GFP\textsubscript{1-10}. A. GFP, GFP\textsubscript{1-10}, and CA-GFP were subjected to digestion by proteinase K. Samples were collected and analyzed at 0, 1, 2, 5, and 10 minutes. GFP is highly resistant to digestion while GFP\textsubscript{1-10} and CA-GFP are nearly completely degraded after 1 min. B. The GFP $\beta$-barrel is drawn (green) with the predicted caspase-6 VELD cleavage sequence (blue) highlighted. The sequence falls in the center of one of the first $\beta$-strand while caspases are predicted to cleave in loop regions. C. GFP, GFP\textsubscript{1-10}, and CA-GFP were subjected to digestion by caspase-6. Caspase-6 is capable of cleaving GFP at a single site near the $N$-terminus. GFP is resistant to digestion while GFP\textsubscript{1-10} and CA-GFP undergo partial cleavage after a two-hour incubation.
Whereas Proteinase K has broad specificity and cleaves most proteins at multiple sites, caspase are very specific proteases, recognizing particular aspartate-terminated tetrapeptide motifs. Caspase-6 has a preference for cleaving protein substrates within loop regions at the recognition sequence VEXD (21). Thus, caspase-6 cleavage serves as a useful indicator of the ordered structure of a particular region of a protein containing a recognition motif. A caspase-6 compatible sequence, VELD (residues 16-19), is present natively in the N-terminal region of GFP, CA-GFP and GFP1-10. This VELD sequence is present in a fully folded GFP in the center of the first β-strand (Fig. 3.9B). We hypothesized that if this region of CA-GFP and GFP1-10 is organized into a protected β-strand conformation like that in mature GFP, caspase-6 would be unable to access the site and cleave the protein. As anticipated, GFP is resistant to cleavage by caspase-6 after incubation for 2 hours at room temperature (Fig. 3.9C). In contrast, caspase-6 is able to partially cleave both CA-GFP and GFP1-10 within 5 minutes. This suggests the first strand region of CA-GFP and GFP1-10 (Fig. 3.9B) is not likely in a β-strand and likely exists in a disordered or loop conformation.

Together, limited proteolysis by proteinase K and caspase-6 indicate that GFP1.

**Figure 3.10.** TROSY NMR HSQC spectra. A. Spectra of GFP collected at 37°C. B. Spectra of CA-GFP also at 37°C. The spectra of GFP is indicative of a well-folded protein while the spectra of CA-GFP resembles that of an unfolded protein.
and CA-GFP lack the stable β barrel structure present in mature GFP. If prevention of chromophore maturation is caused by the presence of a less folded stable folding intermediate or by small perturbations in a mostly formed β-barrel geometry then TROSY-HSQC NMR spectroscopy with $^{15}$N labeled GFP and CA-GFP would indicate a spectral shift for specific residues. A CA-GFP spectrum with chemical shifts similar to those of GFP could indicate which residues were being perturbed if the differences were subtle. The spectrum of GFP (Fig. 3.10A) showed several well-defined and well distributed peaks, consistent with a well folded protein (22). The spectra of CA-GFP (Fig. 3.10B) looked much different than the control spectra and the signal resembled that of an unfolded protein with little chemical shift dispersion (23). In sum, these experiments suggest that CA-GFP does not fold into a stable β-barrel conformation and is more similar to GFP$_{1-10}$ than to the mature GFP.

3.2.5. CA-GFP exists in ‘pro-folding’ conformation.

The similarity of CA-GFP to the GFP$_{1-10}$ conformation as assessed by CD and proteolysis suggests a number of potential models for how the quenching peptide prevents chromophore maturation. Due to the sequential proximity of the quenching peptide to the C-terminal 11$^{th}$ strand of GFP, we favor the model that fusing the quenching peptide to the GFP C-terminus may prevent proper folding of the final 11$^{th}$ strand in CA-GFP. With the 11$^{th}$ strand unable to complete the GFP barrel, the environment for chromophore maturation is never attained so CA-GFP remains in the dark state with an immature chromophore until the quenching peptide is cleaved and released. Dissociation of the quenching peptide allows the GFP barrel to properly fold,
enabling maturation of the chromophore. We reasoned that if this model were correct, fusion of the quenching peptide to other parts of GFP would likely have a differential effect on the behavior of the resulting ‘reporter’. We made a fusion protein with the quenching peptide on the GFP N-terminus (nCA-GFP) (Fig. 3.11A).

To compare the effectiveness of nCA-GFP with that of CA-GFP, we co-transformed E. coli with nCA-GFP and either active caspase-7 or an inactive variant where the active site cysteine has been mutated to an alanine (C186A).

In the absence of active caspase, nCA-GFP was equally ‘dark’ as the C-terminal fusion, CA-GFP, suggesting that the quenching peptide was equally effective at disrupting assembly of the GFP barrel thereby preventing chromophore maturation. However, when nCA-GFP was co-expressed with active caspase-7, it showed less than a 3.8-fold increase in fluorescence in E. coli lysates (Fig. 3.11B). This is in contrast to the C-terminal construct, CA-GFP, which shows a greater than 40-fold
increase in fluorescence when cleaved by active caspase. Western blot analysis using an anti-GFP antibody shows that nCA-GFP is only partially processed in comparison to CA-GFP (Fig. 3.11C), indicating the linker region is less accessible to proteolytic cleavage than in the C-terminal fusion. Partial digestion of nCA-GFP when co-expressed with active caspase indicates that the linker region may be less accessible to the protease. The low fluorescence recovery is then due to a smaller fraction of nCA-GFP being cleaved and a smaller portion of the cleaved protein attaining the mature barrel geometry to allow for chromophore maturation. Thus, it is clear that the conformational state attained by CA-GFP is better poised to allow cleavage, refolding, chromophore maturation and regaining fluorescence. Given the greater propensity for the C-terminal fusions of the quenching peptide to be cleaved and recover fluorescence, we can conclude that the dark state can be attained by a variety of unfolded and partially folded states. Nevertheless, the partially folded, partially aggregated state that CA-GFP obtains has more optimal properties for cleavage and refolding than that induced by adding the quenching peptide to other parts of GFP, such as the N-terminus.

The notion that folding of GFP is integral to the quenching mechanism is further supported by our observation that the quenching peptide is not capable of preventing fluorescence when appended to superfolder GFP (sfGFP). Superfolder GFP was developed by Waldo and coworkers to fold with faster kinetics and therefore is less prone to aggregation (24). When the quenching peptide is fused to the C-terminus the fluorescence of GFP is not quenched (data not shown). We predicted that a shorter linker would be a more stringent test of the quenching ability of the peptide. A CA-superfolder GFP (CA-sfGFP) construct was made with a shorter linker (DELD) rather than the longer
but the resulting CA-sfGFP was still brightly fluorescent suggesting that the kinetics of folding is involved in the quenching mechanism that is at work in CA-GFP. Thus, we venture to state that CA-GFP exists in a relatively optimal folding competent or ‘pro-folding’ state. Although we cannot firmly conclude that CA-GFP is quenched by prevention of the 11th strand from inserting into the partially formed barrel, this model is compatible with our observations.

3.3. Discussion

A number of studies have reported other versions of GFP in which maturation of the chromophore is prevented (16,25,26). GFP is a β-barrel protein in which all the 11 β-strands must be properly assembled to attain the fluorescent state. In each of these darkened states the GFP chromophore is not mature and maturation of the chromophore only occurs when the full GFP barrel assembles. Only when the chromophore is in precisely the correct structural environment can the reaction to form the chromophore occur. For all of the split GFP variants a substantial number of amino acids (10-50% of the protein) are missing, so it is not surprising that these proteins do not attain the properly folded state. In contrast, CA-GFP is composed of all the amino acids necessary for GFP to fold. We hypothesize that some aspect of CA-GFP folding may be sub-optimal as chromophore maturation does not occur. Thus we expect that structural
changes within the GFP portion of CA-GFP must be present, preventing the chromophore from maturing. These structural perturbations could range from very minor, similar to the R96A or R96M mutants (27), to more catastrophic as is the case with the split versions of GFP. While CA-GFP is robust at reporting global apoptosis events a dark reporter that housed a more rapidly maturing chromophore would enable us to capture more rapid kinetic details of the apoptotic cascade.

One of the key features of CA-GFP is the mechanism by which the quenching peptide silences fluorescence. The M2 protein, from which the CA-GFP quenching peptide is derived, forms active tetramers. The dark state of CA-GFP appears to be folded by circular dichroism and is soluble, forming low order oligomers under identical conditions in which GFP is fully fluorescent. The oligomerization of CA-GFP prevents formation of the mature chromophore and thus fluorescence. Cleavage of CA-GFP releases GFP, which can then undergo the necessary conformational rearrangements enabling chromophore maturation. Thus, it initially appeared that even low order oligomerization was sufficient to induce a slightly strained, but folded conformation of GFP that disallowed chromophore formation. It is not unprecedented that subtle changes to GFP, similar to oligomerization, prevent chromophore maturation. Although GFP is a stable and independently folding protein, subtle changes in the environment around the chromophore can dramatically affect fluorescence and the rate of chromophore maturation. For example, maturation of the chromophore in the S65T version of GFP, which adds but a single methyl group to the chromophore cavity is 4.4-fold faster than wild-type GFP. In the R96M variant, the lack of a positive charge drastically changes maturation from hours to months (27). When GFP is compressed with an atomic force
microscope tip, the hydrogen-bond network inside the GFP barrel is broken, resulting in a non-fluorescent chromophore (28). Even a subtle change in pH (to pH 6.5) is sufficient to quench fluorescence (29,30). When GFP is split to remove the last of the 11 β-strands in the GFP barrel, fluorescence is silenced. Association of GFP1–10 (amino acids 1–214) with GFP11 containing only the 11th strand (amino acids 215–230) results in recovered fluorescence (38).

Based on these early observations our initial model of the structure of CA-GFP in the dark state was that of a well-formed β-barrel conformation similar to that of mature GFP, with the presence of the quenching peptide inducing subtle perturbations, potentially a result of oligomerization, inhibiting chromophore maturation. The fully formed β-barrel of GFP has been shown to be impressively robust. Once the barrel has folded and the chromophore is mature it requires fairly extreme conditions to disrupt the fluorescence and unfold the protein. Melnik et al showed that even after partial digestion the fluorescence of the cycle-3 mutant remained nearly unchanged. They hypothesized that the proteases may cleave in loops but that the barrel remained formed and ‘sticks’ to the chromophore to retain fluorescence. They also observed that GFP fluorescence was unaffected in concentrations of urea up to 4M (31). For example, when a trypsin site was engineered into a loop to remove the 11th strand, the protein still had to be denatured and refolded to remove the cleaved strand from the barrel (32). As we observed throughout this study, CA-GFP lacks the remarkable stability of the mature fold, leading us to speculate that it does not exist in the fully formed β-barrel conformation but potentially as a stable intermediate along the native folding pathway.
Several recent protein-folding studies have pointed out that the canonical models of protein folding fail to take into account the kinetics and steric of co-translational folding (15,33-35). One recent study has shown that a stable folding intermediate consisting of the first 10 strands of GFP forms prior to the 11th strand’s release from the ribosome tunnel (15). It follows that if CA-GFP is folding co-translationally, as GFP does, it will sample the same folding intermediates as the WT protein prior to release of the 11th strand. As the peptide is the last portion to be translated and released it is possible that the quenching peptide interferes with the folding of the final stave of the barrel. Since CA-GFP is in a ‘pro-folding’ state, after the peptide is released a majority of the individual CA-GFP molecules can proceed to the native GFP fold.

This is slightly different from the case of nCA-GFP where the very hydrophobic peptide is being translated and released from the ribosome first, leading to a very different sampling of folding states as it is being translated. The grouping of long stretches of hydrophobic residues has been shown to lead to an increase in aggregation (36), which could lead to very different folding intermediates being sampled than the native GFP. This type of behavior has been harnessed to use GFP as a folding reporter. When peptides or proteins have been fused to the N-terminus of GFP, fluorescence is an indication of the stably folded state of the fused peptide or protein, whereas unstably folded protein or peptide fusions prevent GFP fluorescence (37,38). Our results suggest that the quenching peptide is not well folded and therefore leading to a less stably folded intermediate of the GFP barrel in nCA-GFP.

The differences in the oligomeric states of CA-GFP and GFP1-10 suggest that CA-GFP is in a more advanced folding intermediate conformation than GFP1-10. CA-GFP has
a SEC profile consistent with 50% of the protein being a trimer/tetramer. The heterogeneous mixture suggest that there may be several intermediate folding states present as these conformations tend towards self-aggregation (33) with the majority being in a similar stable state that forms a low order oligomer. This is in contrast to the largely aggregated GFP\textsubscript{1-10}. It is not surprising that GFP\textsubscript{1-10} is in an aggregated state; even the robust mature GFP can be coaxed to aggregate upon the addition of 2,2,2-trifluoroethanol (TFE), a chemical commonly used to induce aggregation in proteins (39). The observation that CA-GFP is not as aggregated then supports the hypothesis that it is more ordered than GFP\textsubscript{1-10}.

If the dark state of CA-GFP is dictated by the co-translational folding of the GFP barrel then faster folding variants of GFP should change the fold in the presence of the quenching peptide. The fact that superfolder GFP remains fluorescent in the presence of the quenching peptide indicates that folding kinetics control the dark or bright properties of this class of reporters. The improved folding kinetics of sfGFP prevent populating kinetic traps required for the quenching peptide to function. Thus, in order to develop the next generation CA-GFPs with faster response times, it will also be important to develop quenching peptides that can interact with GFP folding intermediates on a more rapid time scale.

In consideration of the structural insights we report above we propose a working model of the dark state of CA-GFP. We envision that as CA-GFP is being translated the N-terminus begins sampling the native GFP folding states. The presence of the quenching peptide on the C-terminus prevents it from attaining the fully mature β barrel conformation and it is trapped in a stable, ‘pro-folding’ intermediate. After the
quenching peptide is cleaved and released the GFP barrel is then allowed to fold into the robust fold and the chromophore matures, yielding fluorescence. Controlling the folding path of CA-GFP to attain a rapidly maturing chromophore yet maintaining the dark state will be critical in further development of the next generation of CA-GFP reporters. Once a more detailed understanding of the ‘pro-folding’ state is attained the ability to engineer further fluorescent protease reporters as well as expand the utility of the platform to respond to other enzymatic processes will be greatly eased.

3.4. Materials and methods

3.4.1. Molecular cloning

CA-GFP was generated by amplification of GFP (S65T) by PCR using a reverse primer encoding the 27 amino acid transmembrane domain of the M2 protein of the influenza A virus. After amplification of the new gene it was ligated into the XhoI and Ndel sites of pET21b. A linker sequence containing the caspase-3 and -7 cleavage recognition site DEVD was then inserted between GFP and the peptide by site-directed mutagenesis using overlapping inverted primers and amplification of the entire plasmid similar to the QuikChange® (Agilent) approach for a final fusion sequence of DEVDFQGPCNDSSDPLVVAASIIGILHLILWILDRL at the C-terminus of GFP. The expression construct for GFP (S65T) GFP_{1-10} was generated by inserting a stop codon (UAA) after residue K214 using the same site-directed mutagenesis approach using GFP (S65T) in pET21b as the template. CAsf-GFP was generated in a similar approach by amplification of the sfGFP (a kind gift from the Waldo group) gene by PCR using a
reverse primer encoding the linker region as well as the 27 amino acid transmembrane
domain of M2. This gene was then ligated into the XhoI and NdeI sites of pET21b vector.

The N-terminal peptide version of CA-GFP (nCA-GFP) was generated by
separately amplifying the peptide (M2) region from the GFP region of the gene using two
primers to amplify the M2 and two primers to amplify GFP. The first primer (P1)
aannealed to the N-terminal region of GFP with the linker sequence included N-terminally
to the GFP. The second primer (P2) annealed to the C-terminal of GFP and included a
stop codon (UAA) and the restriction site for XhoI. The third primer (P3) Included an
NdeI restriction enzyme site as well as a 6His sequence and annealed to the N-terminal
region of the M2 portion of the CA-GFP gene. The last primer (P4) annealed to the C-
terminus of M2 and included the same sequence for the linker as P1, giving primers P1
and P4 a 24 bp overlapping region. The GFP fragment was then amplified using primers
P1 and P2 while the M2 portion was amplified using primers P3 and P4. After gel
purification of the amplified fragments they were combined and allowed to anneal
through the overlapping region for five PCR cycles before the addition of primers P2 and
P3 which then amplified the full length gene. The gene was then ligated into pET21b
into NdeI and XhoI sites. The final sequence of the N-terminal peptide and linker is
MHHHHHHMCNDSDPLVVAASIIGILHLWILDRLDEVDFQGP.

3.4.2. Protein expression and purification

CA-GFP, CA-sfGFP, nCA-GFP, GFP1-10, and GFP were transformed into E. coli
strain BL21(DE3) for expression. One liter cultures of 2xYT media were inoculated with
1 mL of dense overnight culture and grown at 37°C to an OD600 of 0.6. The cultures
were then induced with 1mM isopropyl-β-D-1-thiogalactopyranoside (IPTG) at 25°C for three hours. Cells were then harvested by centrifugation and disrupted by microfluidization. After centrifugation at 15,000 x g for 45 min. the proteins were purified from supernatant using Co^{2+} affinity chromatography (HiTrap Chelating HP, GE). The column was washed with a buffer of 50 mM imidazole, 300 mM NaCl, 50 mM NaH_2PO_4 pH 8.0 and eluted in a buffer of 300 mM imidazole, 300 mM NaCl, 50 mM NaH_2PO_4 pH 8.0. Protein purity was assessed by gel electrophoresis; proteins were estimated to be at least 95% pure.

Cleaved CA-GFP was obtained by co-expression of CA-GFP and wild type full-length caspase-7 lacking the His6 tag. Flasks containing 1 liter of 2xYT media were inoculated with 1 mL of dense culture and grown at 37 °C to an OD_{600} of 0.6. The cultures were then induced using 1 mM isopropyl -D-1-thiogalactopyranoside (IPTG) and the temperature reduced to 25 °C for three hours of expression. Cells were harvested by centrifugation and disrupted by microfluidization. Clarified lysates were prepared by centrifugation at 15,000 x g for 45 min.

Wild-type caspase-7 was expressed from a plasmid comprising pET23b with full-length caspase-7 gene (40) (gift of Guy Salvesen). This plasmid was mutagenized by QuikChange (Stratagene) to introduce the C186A mutation. Wild-type or C186A caspase-7 were expressed in *E. coli* strain BL21(DE3). Protein expression was induced with 1mM IPTG and allowed to proceed for 12–18 h at 14 °C. Wild-type caspase-7 was purified by affinity chromatography on a Ni-NTA superflow column (Qiagen) developed with a step gradient of 250 mM imidazole. Eluted caspases were further purified using ion exchange chromatography on a 5 mL High-Q column (Bio-Rad) with a linear
gradient from 50mM to 750mM NaCl. The purity of all mutants was assessed by SDS-PAGE gel stained with Coomassie Blue (Bio-Rad) and found to be 95% pure.

The caspase-6 E. coli codon-optimized sequence gene construct in pET11a was transformed into the BL21(DE3) T7 express strain of E. coli (NEB). The cultures were grown in 2xYT media with Amp (100 mg/L, Sigma-Aldrich) at 37 °C until they reached OD_{600}=0.6. The temperature was reduced to 20 °C and cells were induced with 1 mM IPTG (Anatrace) to express soluble His-tagged protein. Cells were harvested after 18 h to ensure complete processing. Cell pellets stored at -20 °C were freeze-thawed and lysed in a microfluidizer (Microfluidics, Inc.) in 300 mM NaCl, 2 mM imidazole, and 50 mM Tris (pH 8.5). Lysed cells were centrifuged at 18,000 g to remove cellular debris. The filtered supernatant was loaded onto a 5-mL HiTrap Ni-affinity column (GE Healthcare). The column was washed with 300 mM NaCl, 50 mM imidazole, and 50 mM Tris (pH 8.5), and the protein was eluted with 300 mM NaCl, 250 mM imidazole, and 50 mM Tris (pH 8.5). The eluted fraction was diluted by five-fold into 2 mM DTT and 20 mM Tris (pH 8.5) buffer to reduce the salt concentration. This protein sample was loaded onto a 5-mL Macro-Prep High Q column (Bio-Rad Laboratories, Inc.). The column was developed with a linear NaCl gradient and eluted in 120 mM NaCl, 2 mM DTT, and 20 mM Tris (pH 8.5) buffer. The eluted protein was stored at -80 °C in the above buffer conditions. The identity and purity of the purified caspase-6 was analyzed by SDS-PAGE.

3.4.3. in vitro CA-GFP cleavage and fluorescence assay

Samples containing 10 µM CA-GFP with or without 9 µM wild-type caspase-7 were prepared at time zero to assess the ability of caspase-7 to digest CA-GFP directly.
At initiation of the reaction 100 µL aliquots of the digest and controls were added to a costar 96-well black plate, and the fluorescence was measured every 5 min (Ex. 475 nm/Em. 512 nm) for 15 h at 27 °C. The remainder of the sample was incubated at 27 °C in 15 µL aliquots to which SDS loading buffer was added at time points of 0, 5, 10, 15, 20, 30, 60, 90, 120 min, and at 15 h. These samples were then run on an SDS-PAGE gel to determine the ratio of cleaved to uncleaved CA-GFP. The percent of cleaved product was determined by quantification using GeneTools software (product version 4.00) on gels imaged with a Gel Doc work station (Syngen).

3.4.4. Absorbance of GFP chromophore

To measure the absorbance spectra of the GFP chromophore, 100 µL of purified CA-GFP and GFP were buffer exchanged into a buffer containing 10 mM NaH2PO4, pH 7 using Millipore Ultra-free 5K NMWL membrane concentrators and diluted to a concentration of 10 µM. The absorbance spectra were collected on a Molecular Devices Spectramax M5 spectrophotometer measuring absorbance from 350 –550 nm in a Costar UV, flatbottom 96-well plate (product number 3635).

3.4.5. Mass spectrometry

Purified CA-GFP and GFP samples were run on a one-dimensional SDS-PAGE gel and Coomassie stained. The bands of interest were excised and cut into 1x1mM pieces and incubated in water for 1 h. The water was removed and 250 mM ammonium bicarbonate was added. For reduction the gel slices were incubated with 45 mM DTT at 50 °C for 30 min. After cooling to room temperature the cysteines were then alkylated by
incubation with 100 mM iodoacetamide for 30 min at room temperature. The gel slices were washed twice with water, which was then removed and a 50:50 (50 mM ammonium bicarbonate: acetonitrile) mixture was placed in each tube, and samples were incubated at room temperature for 1 h. The solution was then removed and 200 µL of acetonitrile was added to each tube at which point the gels slices turned opaque white. The acetonitrile was removed, and gel slices were further dried in a SpeedVac. The gel slices were then rehydrated in a 2 ng/L solution of trypsin (Sigma) in 0.01% ProteaseMAX Surfactant (Promega): 50mM ammonium bicarbonate for 21 h at 37 °C. The supernatant of each sample was then removed and placed in a separate 0.5 mL Eppendorf tube. Gel slices were further dehydrated with 60 mL of 80:20 (acetonitrile: 1% formic acid). The extract was combined with the previous supernatants of each sample and further purified using a micro Zip Tip (Millipore). The tryptic fragments were analyzed using matrix-assisted-laser desorption/ionization Time-of-Flight (MALDI-TOF) and subsequent MS/MS using a Shimadzu Biotech Axima TOF2 (Shimadzu Instruments) mass spectrometer to determine the maturity of the chromophore. The observed mass loss in GFP is due to oxidation of the Tyr-66 (loss of two daltons) and dehydration (loss of eighteen daltons) during the chromophore maturation process.

3.4.6. Circular dichroism spectroscopy

Repeating structures such as helices and -sheets rotate circularly polarized light differently such that the degree of protein secondary structure can be estimated from a circular dichroism spectra. Purified CA-GFP and GFP were prepared at a 10 µM concentration as described for absorbance measurements. CD spectra were measured on a
J-715 circular dichroism spectrometer (Jasco) at 25 °C. For temperature-wavelength scans purified proteins were buffer exchanged into a buffer containing 10 mM NaH₂PO₄, pH 7 using Millipore Ultra-free 10K NMWL membrane concentrators and diluted to ~10 µM as determined by A₂₈₀ nm using a Nanodrop 2000C Spectrophotometer. CD spectra were measured on a J-715 circular dichroism spectrometer at intervals of every 5 degrees from 20-90 °C at a rate of temperature increase of 1°C per minute in a quartz cuvette with a 0.1 cm pathlength.

3.4.7. Size exclusion chromatography

The size of CA-GFP (1 mg/mL), GFP (1 mg/mL), GFP₁₋₁₀ (1.5 mg/mL) and cleaved CA-GFP (0.6 mg/mL) was determined by SEC using a Superdex 200 10/300 GL column (GE Healthcare). The molecular weight of each was determined by comparison to molecular weight standards albumin (66 kDa), carbonic anhydrase (29 kDa), ovalbumin (45 kDa), and ribonuclease A (14.7 kDa) with blue dextran 2000 used to determine the void volume of the column (Sigma product No. MW-GF-1000). The curves were normalized for the respective protein concentrations.

3.4.8. Proteolysis

Proteolysis of 20 µM CA-GFP, GFP₁₋₁₀, and GFP by the addition of 500 nM Proteinase K (Sigma) was performed in a buffer of 50 mM Tris pH 7.5, 5 mM CaCl₂. Aliquots were taken from the reaction at 0, 1, 2, 5, and 10 minutes and added to protease inhibitor phenylmethylsulfonyl fluoride (PMSF) to stop the reaction. SDS loading buffer was then added to the samples, which were boiled for 10 minutes and analyzed by SDS-
PAGE. The presence of the PMSF made the loading of samples onto the SDS-PAGE gel challenging but was aided by an additional 10% glycerol added to the SDS-PAGE running buffer, the addition of 1 µL of 10X SDS-PAGE running buffer to each reaction prior to loading to ensure the appropriate pH, the chilling of the samples, gel and running buffer prior to loading. It was also necessary to load the gel with a low voltage (50V) current being applied so that the sample would be drawn into the gel quickly.

3.4.9. Cleavage by Caspase-6

Purified CA-GFP, GFP1-10, and GFP were incubated with purified active caspase-6 at room temperature in a buffer of 100 mM HEPES pH 7.5, 10% sucrose, 0.1% CHAPS, 30 mM NaCl, and 5 mM DTT. Aliquots of each reaction were taken at 0, 5, 10, 30, 60, and 120 minute time intervals and added to SDS loading buffer with DTT and immediately boiled for 10 min to stop cleavage. The samples were then assessed using SDS-PAGE to determine cleavage at each time point.

3.4.10. TROSY NMR

$^{15}$N labeled CA-GFP and GFP were obtained by growth in M9 minimal media using $^{15}$N ammonium chloride as its sole nitrogen source (Cambridge isotopes) (41). A 20 mL LB culture was inoculated from a dense overnight culture. After reaching an OD$_{600}$ of 0.5, the cells were spun down and washed once with PBS. The cells were resuspended in 100 ml $^{15}$N labeled M9 media. The culture was incubated about 3.5 hours at 37°C until reaching an OD$_{600}$ of 0.5. The doubling time of the cells in the M9 media was approximately 90 min in agreement with published values (42). The culture was then
diluted into 200 mL of labeled M9. After incubation at 37°C until reaching an OD$_{600}$ of 0.5 once more the culture was diluted to 1L of labeled M9 and then induced with 1mM isopropyl $\beta$-D-1- Thiogalctopyranoside (IPTG) at 25°C for three hours. The $^{15}$N labeled proteins were purified as described above. The proteins were buffer exchanged using Millipore Ultra-free 10K NMWL membrane concentrators into a buffer of 10 mM NaCl, 2.7 mM KCl, and 6 mM NaH$_2$PO$_4$ pH 7.2 at a concentration of 60 µM. $^1$H-$^{15}$N Transverse relaxation optimized spectroscopy (TROSY) spectra in this study were obtained at 37 °C on a 700-MHz Varian NMR system equipped with a cryogenically cooled triple-resonance probe. Spectra were processed using NMRpipe (43) and analyzed using Cara(44).

3.4.11. Lysate-based CA-GFP fluorescence assays

Expression constructs for CA-GFP with the peptide on either the C- or N-terminus, CA-GFP and nCA-GFP respectively, in pET21b (Amp) was co-transformed in the BL21(DE3) strain of E. coli with either a constitutive two chain version of caspase-7 (C7 CT) or full-length inactive caspase-7 where the active site cysteine has been mutate to an alanine (C186A). Both versions of caspase-7 were contained in the vector pBB75 (Kan). 50 mL cultures were inoculated from dense overnight cultures and grown at 37°C to an OD$_{600}$ of 0.6. Cultures were then induced with 1mM IPTG at 25°C for 18 hours. Two mL of each culture was centrifuged and resuspended in 800 µL a buffer of 0.5 mg/mL lysozyme (Sigma) and 2 units DNAase (NEB). Cells were lysed using four cycles of freeze-thaw and the supernatant (100 µL) was analyzed for fluorescence (Ex. 475 nm/Em. 512 nm) in a co-star 96-well black plate on a Molecular Devices Spectramax M5.
spectrophotometer. Fluorescence values were normalized based on the relative OD of each culture.

3.4.12. Western blotting

Samples of supernatant prepared as described for the lysate fluorescence assays of CA-GFP or nCA-GFP each co-transformed with either the active WT caspase-7 or the inactive C186A caspase-7 (above) were taken. Three identical SDS-PAGE gels were run of the four samples. One was stained with Coomassie, two gels were transferred onto Hybond-ECL™ nitrocellulose membrane (GE Healthcare) and blocked overnight in a solution of 0.03 mg/mL BSA in TBS. One was then blotted with an anti-GFP (Millipore) monoclonal mouse primary antibody, and the other was blotted with a mouse primary antibody recognizing the large subunit of caspase-7 (Sigma). The western blots were then treated with anti-mouse IgG alkaline phosphatase produced in goat (Sigma) and visualized using 1-Step™ NBT/BCIP (Thermo Scientific).

References

44. Keller, R. (2004) Optimizing the process of nuclear magnetic resonance spectrum analysis and computer aided resonance assignment. ETH, Zurich
CHAPTER 4

A VERSATILE PLATFORM FOR PROTEASE REPORTING

This chapter will be submitted in part as: Wu, P., Nicholls, S.B. and Hardy, J.A., 2012. “A Versatile Platform for Protease Reporting.” Biophysical Journal. SBN performed profiling of caspase activity against caspase-activatable reporters as well as development and characterization of CA-CerFP, CA-CitFP, CA-mNeptune, DVPA-GFP, and HIVA-GFP. PW developed and characterized C6A-GFP and HRA-GFP and their respective derivatives including all constructs with a stop codon inserted in the linker region.

Abstract

Proteases are one of the most important and historical utilized classes of drug targets. To effectively study this class of proteins, which encodes nearly 2% of the human proteome, it is necessary to develop effective and cost-efficient methods to report on their activity both in vitro and in vivo. One of the most important classes of proteases both in terms of homeostatic regulation and drug design is the caspase family. Caspases are responsible for carrying out the process of programmed cell death, a function necessary for the regulation of cells. This important class of proteases is a relevant drug target for a variety of diseases such as cancer and neurodegenerative diseases. We have previously discussed our reporter of caspase activity, CA-GFP (chapters II and III). CA-GFP is a genetically encoded dark-to-bright reporter of caspase activity in a variety of systems. Here we show the utility of CA-GFP as a platform for the creation of a family of protease reporters by engineering the linker region, which includes the protease recognition site, as well as the fluorescent protein used.
4.1. Introduction

The use of genetically encoded fluorescent reporters has become increasingly important in the study of proteolytic activity both *in vitro* and *in vivo* [for review see (1)]. Proteases are at the center of the most critical cellular cycles and pathways and are one of the most common drug targets for drug design [for review see (2)]. Several metalloproteases have been successful drug targets for treatment of hypertension while proteases involved in blood coagulation have been the target for treatment of blood disorders for over half a century. Viral proteases are an important drug target as well in wide spread viruses such as HIV, Dengue, and West Nile Virus which affect millions of people worldwide. One important class of enzymes, the caspase family, is central to the critical process of apoptosis, or programmed cell death, and therefore has become an important drug target for treating a variety of diseases. Caspases are a family of cysteine-aspartate proteases that cleave cellular substrates in response to cellular stress and inflammation ultimately leading to cell death. Once the apoptotic cascade is activated, initiator caspases (-8 and -9) are triggered to cleave the zymogen forms of the executioner caspases (-3, -6, and -7) into the large and small subunits, which make up the active form of the protease. Activation of the executioners is the last irreversible step in the apoptotic cascade ultimately leading to cell death. The search for activators of caspases could lead to new cancer therapies, initiating cell death in rapidly dividing cancer cells, while inhibitors of caspases could provide treatments for neurodegenerative diseases or control of tissue damage in instances such as post heart attack. To validate any protease as an appropriate drug target, we first need the ability to easily and cost
effectively assess the activity of the protease in its native biological setting as well as in the presence and absence of drug candidates.

It is important to understand not only the \textit{in vitro} activity of these drug targets but also the biologically relevant \textit{in vivo} function. Current technologies for studying protease activity in a native cellular context include FRET, luciferase, and positron emission (PET) reporters similar to those discussed specifically for caspases in Chapter I and II. We have recently reported the development of a genetically encoded, dark-to-bright reporter of caspase activity(3). This reporter is multi-functional in that it can be used both \textit{in vitro} and \textit{in vivo} and in both bacterial and mammalian systems to indicate the activity of executioner caspases (chapter II). Our caspase-activatable reporter was designed by fusing GFP (S65T) through a flexible linker containing the caspase-3, and -7 recognition sequence DEVD to a hydrophobic 27-amino-acid peptide on the C-terminus (Fig. 4.1). The presence of this peptide inhibits chromophore maturation in GFP until it is removed in the presence of active caspase.

As our CA–GFP has been shown to be useful in the reporting of caspase-7 activity we sought to use this as a model for the further design of other proteolytic reporters including other caspases as well as additional proteases. Caspase-6 has recently been shown to play a unique role in both Alzheimer’s (4) and Huntington diseases (5), cleaving APP and Huntingtin proteins respectively at sites unique from
caspase-3 cleavage. It has therefore been deemed an important drug target in the prevention and treatment of neurodegenerative diseases. In order to effectively target caspase-6 with therapeutics it is also necessary to assess the effect that such an inhibitor would have on homeostasis throughout the body. One example is a recent study showing another role of caspase-6 in the maintenance on the prostate tissue(6). In order to assess the efficacy and off target effects of any caspase-6 specific therapeutics it is necessary to specifically monitor caspase-6 activity in a biological context. A caspase-6 specific CA-GFP would enable longitudinal monitoring in cells and whole organism models of caspase-6 activity both during homeostasis as well as in response to potential drug candidates.

The recently revealed non-apoptotic role of caspase-6 in neurodegenerative diseases and in prostate function implies the possibility that other caspases may have roles independent of the apoptotic cascade as well. Though caspases have been profiled to have differences in their substrate preferences, despite their similar active sites, by profiling of peptide libraries (7) much is still unknown about the differences in their cellular substrates. In order to effectively assess the unique roles of each caspase it is essential to have the ability to simultaneously monitor the activity of multiple caspases. To do so using CA-GFP requires the ability to change not only the protease recognition sequence in the linker region but the fluorescence color as well in order to differentiate the activities. While GFP is a ubiquitous tool in molecular biology an entire palette of fluorescent proteins has been developed [for review see (8)]. By expanding the color palette of our reporter it could become a useful tool in imaging multiple proteolytic events within a cell. As the fluorescence from the red fluorescent protein family has been
shown to have a greater tissue depth penetrance (9,10) a red-shifted version would dramatically improve our ability to monitor proteolytic events, particularly in transgenic animals. The first half of this chapter focuses on our engineering of the linker region to respond to caspase-6 as well as other biomedically relevant proteases while the second half describes the development of reporters which fluoresce at a variety of wavelengths to enable robust simultaneous reporting of multiple proteases.

4.2. Results and discussion

We hypothesize that CA-GFP can act as a tunable platform for a variety of applications and that by substitution of other protease recognition sites in the linker region we can develop a family of Protease Activatable (PrA) reporters. To develop new versions of CA-GFP that could respond to additional proteases, we found it was first necessary to understand the essential characteristics and parameters of the linker region that allow fluorescence recovery after proteolytic cleavage. In addition, the similarity in structure of the various colors within the GFP family of fluorescent proteins leads us to believe that the color of the reporter is also tunable.

4.2.1 Development of a caspase-6 activatable-GFP

Until very recently caspase-6 has been viewed as being very structurally and functionally similar to the two other executioner caspases -3 and -7. More recent studies have shown that caspase-6 not only has structurally unique states from the canonical caspase structure (11-13) but that it also may have unique non-apoptotic roles namely in neurodegenerative diseases (4,5). CA-GFP has been shown to be a robust in vivo reporter
for executioner caspases both in *E. coli.* and in mammalian cells (3). In order to build a Caspase-6 Activatable GFP reporter (C6A-GFP) that would allow us to distinguish the activity of caspase-6 from that of caspases-3 and -7, we replaced the DEVD linker sequence in CA-GFP with the caspase-6 recognition sequence VEID (C6A-GFP or VEIDF-GFP) (Table 4.1).

**Table 4.1.** Constructs used throughout the studies with the linker sequence as well as any additional mutations listed. ‡ indicates the site of cleavage by the intended protease; * indicates a stop codon.

<table>
<thead>
<tr>
<th>Version of –GFP reporter</th>
<th>Linker Sequence</th>
<th>Additional Mutations</th>
<th>Intended Protease</th>
</tr>
</thead>
<tbody>
<tr>
<td>CA-GFP</td>
<td>DEVD↓FQGP</td>
<td></td>
<td>Caspase-7</td>
</tr>
<tr>
<td>C6A-GFP or VEIDF-GFP</td>
<td>VEID↓FQGP</td>
<td></td>
<td>Caspase-6</td>
</tr>
<tr>
<td>VEIDF-GFP‡</td>
<td>VEID↓FQGP</td>
<td>D27E</td>
<td>Caspase-6</td>
</tr>
<tr>
<td>C6A-GFP or VEIDG-GFP‡</td>
<td>VEID↓QGP</td>
<td>D27E</td>
<td>Caspase-6</td>
</tr>
<tr>
<td>VEIDE-GFP‡</td>
<td>VEID↓EQGP</td>
<td>D27E</td>
<td>Caspase-6</td>
</tr>
<tr>
<td>VEIDK-GFP‡</td>
<td>VEID↓KQGP</td>
<td>D27E</td>
<td>Caspase-6</td>
</tr>
<tr>
<td>C7A-GFP or DEVDF-GFP‡</td>
<td>DEVDF↓QGP</td>
<td>D27E</td>
<td>Caspase-7</td>
</tr>
<tr>
<td>LEVLF-GFP or LEVLF-GFP‡</td>
<td>LEVLFQ↓GP</td>
<td></td>
<td>hR3C Protease</td>
</tr>
<tr>
<td>HRA-ΔYKGFP</td>
<td>LEVLFQ↓GP</td>
<td>Deletion of Y273 and K274 of GFP</td>
<td>hR3C Protease</td>
</tr>
<tr>
<td>HRA-GFP or LEVLG-GFP‡</td>
<td>LEVLQ↓GP</td>
<td></td>
<td>hR3C Protease</td>
</tr>
<tr>
<td>DEVLF-GFP‡</td>
<td>DEVLFQ↓GP</td>
<td></td>
<td>hR3C Protease</td>
</tr>
<tr>
<td>DEVLG -GFP‡</td>
<td>DEVLGQ↓GP</td>
<td></td>
<td>hR3C Protease</td>
</tr>
<tr>
<td>EEVLF-GFP‡</td>
<td>EEVLFQ↓GP</td>
<td></td>
<td>hR3C Protease</td>
</tr>
<tr>
<td>DEVEF-GFP‡</td>
<td>DEVEFQ↓GP</td>
<td></td>
<td>hR3C Protease</td>
</tr>
<tr>
<td>DEVEG-GFP</td>
<td>DEVEGQ↓GP</td>
<td></td>
<td>hR3C Protease</td>
</tr>
<tr>
<td>DEVCF-GFP‡</td>
<td>DEVCFQ↓GP</td>
<td></td>
<td>hR3C Protease</td>
</tr>
<tr>
<td>DEVSF-GFP‡</td>
<td>DEVSFQ↓GP</td>
<td></td>
<td>hR3C Protease</td>
</tr>
<tr>
<td>VEIDF-GFP‡</td>
<td>VEIDFQ↓GP</td>
<td></td>
<td>hR3C Protease</td>
</tr>
<tr>
<td>PEHDF-GFP</td>
<td>PEHDFQ↓GP</td>
<td></td>
<td>hR3C Protease</td>
</tr>
<tr>
<td>LEHDF-GFP</td>
<td>LEHDFQ↓GP</td>
<td></td>
<td>hR3C Protease</td>
</tr>
<tr>
<td>DEVD*</td>
<td>DEVD*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DEVDFQ*</td>
<td>DEVDFQ*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LEVL*</td>
<td>LEVL*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LEVLFQ*</td>
<td>LEVLFQ*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DEVL*</td>
<td>DEVL*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DEVLFQ*</td>
<td>DEVLFQ*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ELEVLFQ*</td>
<td>ELEVLFQ*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EEVEFQ*</td>
<td>EEVEFQ*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>VEID*</td>
<td>VEID*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CA-CerFP</td>
<td>DEVD↓FQGP</td>
<td>F64L, Y66W, S72A, Y145A, H148D</td>
<td>Caspase-7</td>
</tr>
<tr>
<td>CA-CitFP</td>
<td>DEVD↓FQGP</td>
<td>T65G, V68L, Q69M, S72A, T203Y</td>
<td>Caspase-7</td>
</tr>
<tr>
<td>CA-mNeptune</td>
<td>DEVD↓FQGP</td>
<td>mNeptune Fluorescent protein; M2 peptide on N-terminal</td>
<td>Caspase-7</td>
</tr>
</tbody>
</table>
The resulting construct was dark before cleavage by caspase-6 as expected, indicating that the quenching peptide functioned analogously as in CA-GFP. However, when co-transformed with active caspase-6 in E. coli there was no observable fluorescence (Fig. 4.2A), despite the fact that there was observable cleavage by western blot analysis (Fig. 4.2B). After comparison of the cleavage pattern observed with that of a construct in which a stop codon was inserted after the VEID in the linker (Table 4.1), we recognized that the cleaved fragments assessed by western blot using anti-GFP antibody were not at the expected molecular weight. Closer analysis of the sequence of GFP revealed a VELD sequence at residues 24-27 near the N-terminus of the fluorescent protein (Fig. 4.3). As GFP is very sensitive to truncations at either termini (14), it

**Figure 4.2.** Whole cell fluorescence of caspase-6 activatable reporter constructs. VEIDG<sup>e</sup> has the greatest fluorescence response in the presence of active caspase-6 (gray bars) vs. cotransformation with an empty pET vector (white bars). Western blot analysis showed that the VEID version of the reporter was being cleaved at an N-terminal site.

**Figure 4.3.** The three dimensional structure of GFP (PDBID: 2YOG) with the VELD (residues 24-27) highlighted in the strand. (Figure modeled in Pymol and Adobe Illustrator courtesy of PW)
was not surprising that no fluorescence was observed. We then mutated Asp27 into Glu (D27E) to eliminate the internal VELDG sequence within GFP. The resulting VEIDF-GFP\textsuperscript{e} was also dark prior to cleavage but a small amount of fluorescence was observable when co-transformed with active caspase-6 (Fig. 4.2A). Unfortunately, VEIDF-GFP\textsuperscript{e} did not show as strong a fluorescence increase as C7A-GFP. We had previously found that CA-GFP and derivatives could be optimized best when expression constructs that express the mature, cleaved form of the caspases were used. In our lab we call these the CT or constitutively two-chain versions of the caspases. In these experiments we had used caspase-6 CT (pC6 FL D179 CT), so we could expect that differences in recovery of fluorescence were likely due to the intrinsic amount of cleavage not to differences in the levels of active caspase-6.

Western blotting with an anti-GFP antibody showed that VEIDF-GFP\textsuperscript{e} is not well cleaved in the co-expression (Fig. 4.2B) although mature caspase-6 is expressed in appropriate levels, as assessed by western blotting (data not shown). This is not due to a shortage of caspase-6 activity, as in a control VEID-GFP construct that only has VEID linker but not the D27E mutation, a majority of the expressed VEID-GFP reporter is clearly cleaved. VEID is the reported caspase-6 preferred substrate sequence. From literature, caspase-6, unlike caspase-7 and -3, does not have strong residue preference at its P1’ subsite (15). However interestingly, in VEID-GFP co-expressed with caspase-6 the majority of the reporter is apparently cleaved at the N-terminal VELDG site, but not at the linker site. While western blotting results suggest that there is also a fraction of VEID-GFP reporter cleaved at both the N-terminal VELDG site and the VEID linker, we did not observe any bands representing a single cleavage at the VEID linker only. This
strongly suggests that caspase-6 cleaves the N-terminal VELDG site first and that
caspase-6 prefers a glycine at the P1’ subsite in our reporter construct. In light of this, we
further modified the VEIDFQ linker in VEIDF-GFP into VEIDGQ. The resulting
VEIDG-GFP reporter yields a strong fluorescent signal increase upon co-expression
with active caspase-6. As caspase-7 and -3 have P1’ preference of small residues
especially glycine, a low level increase on fluorescent signal upon caspase-7 co-
expression is expected. Residues including Glu and Lys are frequently found in the P1’
subsite of substrates of caspase-6 but not in casp-3 and -7 substrates (7). However,
introducing Glu or Lys at the P1’ does not improve the signal strength of the reporter (Fig.
4.2A). Therefore, although P1’ Gly increases the basal fluorescent level in our C6A-GFP reporter co-expression with both empty pET vector and casp-7, VEIDG-GFP is still
overall our best reporter considering the strong signal increase upon caspase-6 activity.

4.2.2. Profiling of caspase activity using C6A- and C7A-GFP

While all caspases have very similar catalytic diad geometries during substrate
cleavage, they have been shown to have differences in their cleavage site preferences due
to differences in the residues that comprise the peptide recognition sub-sites (S1-S4) (7).
To profile the activity of caspases-3, -6, -7, and -8 against our caspase-6 and -7 reporters
C7A-GFP (DEVDF-GFP) and C6A-GFP (VEIDG-GFP) co-transformations with both
the active and inactive forms each of the caspases was performed and the resulting
fluorescence in E. coli lysates was determined for each case (Fig. 4.4). In the case of
C6A-GFP all of the caspases tested showed at least a 3-fold increase in activity in the
presence of the active caspase. Additional investigation is necessary in the case of C6A-
GFP<sup>e</sup> co-transformed with the inactive caspase-8 to ensure efficient expression, as it was not observed by western blot (Fig. 4.4 top), however co-expression with the active caspase-8 demonstrated efficient cleavage. As expected the strongest response was from caspase-6 (greater than 14-fold) (Fig. 4.4 top). Western blotting using an anti-GFP antibody shows that in the presence of caspase-7 C6A-GFP<sup>e</sup> is only partially processed, indicating that while it is able to cleave the caspase-6 cleavage sequence VEID, it does not prefer this recognition sequence.

Thus, it appears that C6A-GFP<sup>e</sup> is a reasonably good reporter for caspase-6 activity, since caspase-6 triggers 3-4 fold greater response than other related caspases.

In the case of the C7A-GFP<sup>e</sup>, caspase-7 clearly triggers the greatest response (Fig. 4.4 bottom). It was surprising to us that though caspases -3 and -8 were able to efficiently cleave C7A-GFP<sup>e</sup>, they did not result in a similar level of fluorescence.
reporting as caspase-7. This observation remains an open question in the lab. Caspase-3 is generally considered a more robust protease than caspase-7. Caspase-3 has a catalytic rate ($k_{cat}/K_M=1.8 \mu M^{-1}s^{-1}$) nearly one order of magnitude higher than caspase-7 ($k_{cat}/K_M =0.4 \mu M^{-1}s^{-1}$)(16). Although caspase-3 and -7 share the same preferred recognition sequence when assessed with peptide substrates (DEVD) it is possible that due to it’s higher turnover number caspase-3 cleaves other *E. coli* proteins and therefore more detrimental to the cell over the 18 hour induction period. It is also possible that it is cleaving the reporter at additional sites, as was observed at the N-terminal VELDG site in the original C6A-GFP. We do not observe any additional bands by western blot, however, it is possible that if the additional sites cleave the reporter into sufficiently small fragments the antibody would no longer be able to recognize them. It preliminarily appears that the GFP portion of C7A-GFP may house a second recognition site that decreases activity. In cells co-expressing DEVD*, which expressed GFP with DEVD at the C-terminus followed by a stop codon (*), with inactive caspase-3 (C163A) the fluorescence is more than seven-fold of that with the active caspase-3. Though these results are preliminary this suggests that active caspase-3 is cleaving GFP in a location that leads to its decreased fluorescence. This site has not yet been identified but is the subject of ongoing work in the lab. Although caspase-8 prefers the recognition sequence LETD (7), it has previously been reported to cleave at the caspase-7 recognition sequence DEVD(17). We also observe efficient cleavage of C7A-GFP though the fluorescence recovery is lower than one might expect for this level of cleavage. The observation that multiple caspases have the ability to cleave the substrates typically thought to be predominantly recognized by other caspases supports the idea that while
caspases are reported to have very specific cleavage preferences they are able to recognize the substrate of the other caspases in the absence of the intended caspase (18). The efficient processing of the reporter with the notable lack of fluorescence as seen in the cases of caspase-8 cleavage of both C6A-GFP and C7A-GFP and caspase-7 cleavage of C6A-GFP is somewhat confounding. In the latter case there is only partial processing observed, however, the fluorescence is approximately half that expected for 50% processing. Further investigation is necessary to ensure processing is occurring at precisely the expected linker site and no other. Sequence analysis does not suggest any compelling secondary cleavage sites, as was the case with caspase-6 cleavage of the first C6A-GFP (VEIDF-GFP) construct. On the other hand, secondary cleavage sites that decrease the response to other caspases could have some advantages. In particular we have taken advantage of this characteristic when we have used CA-GFP variants for directed evolution experiments.

4.2.3 Engineering new proteolytic recognition: human Rhinovirus 3C

By successfully engineering a caspase-6 reporter we gained some insight into how the CA-GFP concept could be adapted to report on other biomedically important proteases. In addition to building a relevant reporter, we also wanted to target a protease with a similar recognition sequence to the one in CA-GFP but that would cleave in a different position so that we could assess whether the length of the remaining C-terminal tail after cleavage would impact fluorescent recovery. We varied the linker sequence in our new protease activatable (PrA-GFP) reporter to yield a specific reporter for human rhinovirus 3C (hR3C) protease which also is commercially available as PreScission®
Protease (GE Healthcare). hR3C has a preferred substrate sequence of LEVLFQGP (19). Our initial design changed the DEVDFQGP sequence in CA-GFP to LEVLFQGP, yielding a reporter with a moderate 5-8 fold increase in fluorescence upon hR3C protease co-expression. Not only is the fluorescence increase much less than for CA-GFP co-expression with caspase-7, the relative fluorescence after cleavage is also very low in comparison to CA-GFP or C6A-GFP (Fig. 4.5A). This is not due to different expression level of the reporter nor is it related to the completeness off the proteases digestion (data not shown). Sequences of CA-GFP and LEVLFQ-GFP are identical other than the substitution of two aspartates (D) for the two leucines (L) in the linker. Due to the position of the cleavage site within the recognition sequence, caspase-7 generates a DEVD C-terminus in CA-GFP while hR3C protease yields LEVLFQ at the C-terminus, giving a C-terminal remnant two residues longer in length. In order to increase the reporter detectability while keeping its specificity against hR3C protease, we made a spectrum of variations by point mutations of several linker residues. We first investigated the impact of length of C-terminal remnant. We removed Tyr and Lys residues just preceding the LEVLFQ sequence (HRA-ΔYKGFp, Table 1) to yield a C-terminus with equal length to DEVD, and we did not observe significant improvement in signal. We then focused on the LEVLFQ linker, especially the two leucines which represent the only sequence difference between CA-GFP and LEVLFQ-GFP. We replaced the leucines in LEVLFQ-GFP with negatively charged or hydrophilic residues. Replacing the two leucines indeed leads to significant increase in fluorescent signal in all the new reporters (Fig. 4.5A). Both DEVLFQ-GFP and EEVLFQ-GFP show a strong increase in fluorescence upon co-expression of hR3C protease while still maintaining recognition for
cleavage. Other linker variants either have lower signal strength or higher background in the dark state.

LEHDFQ-GFP shows a similar signal to noise ratio and even stronger signal strength, however, LEHD is a caspase-9 recognized sequence and therefore it is not specific to hR3C protease.

In our efforts to investigate why the LEVLFQ-GFP reporter has a lower fluorescent signal, we introduced stop codons either after the glutamine in the linker to mimic the C-terminus generated by hR3C cleavage, or before the phenylalanine to mimic the length of the remnant in CA-GFP cleavage by caspase-7. Our results clearly show that the C-terminal LEVLFQ remnant strongly affects the fluorescent signal of GFP (Fig. 4.5B), likely due to the clustering of four hydrophobic residues which may lead to aggregation and prevent GFP chromophore maturation. As expected, when we replaced the phenylalanine with glycine in the reporters, we also observed further fluorescent increases, both before and after cleavage. In LEHDG-GFP the linker sequence may be
too different from an optimized hR3C substrate and therefore is not well recognized (Fig. 4.5A).

4.2.4. Engineering viral protease recognition: Dengue virus protease and HIV

Protease

We sought to further expand our PrA-FP family by the inclusion of reporters for two additional viral proteases of particular interest in drug design. The first is a reporter for Dengue virus protease. Worldwide there are 2.5 billion people at risk of infection with Dengue fever (20). Dengue is unique from other viruses, which are traditionally combated by vaccination, in that effective vaccination requires a tetravalent vaccine which has not been successfully developed (21). In an alternative to vaccination the viral protease, Dengue Virus Protease (DVP), is one of the targets for drug design in the development of new therapies for this global health problem. DVP2 is a serine protease responsible for the processing of the viral polyprotein and recognizes the substrate sequence GRR (22). Dengue virus has been shown to induce apoptosis in infected cells (23), therefore having a Dengue virus protease reporter (DVPA-GFP) would allow real-time imaging of the activation of the viral proteases leading to the activation of the apoptotic cascade. The first DVPA-GFP design simply replaced the first three residues of the CA-GFP linker (DEV) with the DVP2 recognition sequence GRR, leaving the remainder of the linker identical (Table 4.2).
Expression of the DVPA-GFP in *E. coli* showed that this alteration of the linker sequence did not alter the presence of the critical dark state of the reporter. Incubation of purified DVPA-GFP with active DVP2 did not yield any change in fluorescence even after 18-24 hr incubations. SDS-PAGE analysis of the reaction did not indicate any cleavage of the reporter, although analysis was complicated by the fact that DVP2 has a molecular weight of 30.5 kDa, very close to that of DVPA-GFP so it was difficult to accurately assess cleavage. Further investigation of the literature revealed that

<table>
<thead>
<tr>
<th>Version of GFP Reporter</th>
<th>Linker Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>CA-GFP</td>
<td>DEVD↓FQGP</td>
</tr>
<tr>
<td>DVPA-GFP</td>
<td>GRR↓DFQGP</td>
</tr>
<tr>
<td>DVPA-GFP_v1</td>
<td>GRR↓GFQGP</td>
</tr>
<tr>
<td>DVPA-GFP_v2</td>
<td>DEGRR↓GGP</td>
</tr>
<tr>
<td>HIVA-GFP</td>
<td>VSFNF↓PQITL</td>
</tr>
</tbody>
</table>

Table 4.2. Comparison of the linker sequence of CA-GFP and the DVPA-GFP and HIVA-GFP reporters (↓ indicates the site of cleavage by the intended protease).

**Figure 4.6.** Two-dimensional gel analysis of dengue virus protease 2 (DVP2) (top) with two DVP2 activatable reporters DVPA-GFP_v1 (middle) and DVPA-GFP_v2 (bottom). Each reporter was run after a 24 hr incubation with the protease (right) and independent from the protease (left). Though the bands are broad and not conclusive it does not appear as though there is any significant cleavage of the reporter by the protease.
DVP2 has a very low preference for charged residues in the P1’ substrate sub-site (22), suggesting that the linker sequence may not be ideal for cleavage. Two further DVPA-GFP constructs were cloned, changing the P1’ site to a more preferred residue, glycine (DVPA-GFP_v1), and in one construct moving the recognition sequence further from the C-terminus of GFP (DVPA-GFP_v2). Both of these additional constructs also exist in a dark state when expressed, however, once again neither showed any increase in fluorescence after a 24 hr incubation of purified reporter with active DVP2. Western blot analysis of the cleavage was attempted however the DVP2 appears to cross-react with the anti-GFP antibodies used further confounding this analysis. We undertook a two dimensional gel analysis to facilitate separation of DVP from DVPA-GFP. Though the results were not definitive due to the broad band of both proteins from the isoelectric focusing, the results did not indicate any cleavage (Fig. 4.6). A GST-tagged version of DVP2 was constructed in an effort to better separate the two proteins in SDS-PAGE analysis however the GST fusion further diminished the already low catalytic activity (on the order of minutes$^{-1}$, data not published) of the DV protease, so this approach also proved intractable and work on DVPA-GFPs has been suspended. Future development of a DVPA-GFP may be more complicated than other proteases because of the low catalytic activity of the existing protease construct. While caspases have evolved to cleave a variety of substrates and prefer to cut in loop regions DVP2 has evolved for a much more specific function in the viral life cycle. It is possible that there are exosites which help DVP2 recognize its polyprotein substrate specifically rather than only the active site geometry dictating cleavage. This is plausible given the fact that non-specific cleavage could lead to premature cell death or the triggering of an immune response, which would
be detrimental to the viral life cycle. A very slow protease may be sufficient for cleaving
the polyprotein, due to the proximity of the intramolecular cleavage reaction and this low
catalytic turnover may be beneficial in avoiding degradation of necessary host
components. Alternatively, DVP2 may simply be a very primitive protease due to its
origins.

Another important proteolytic drug target is HIV protease. There are currently
several FDA approved therapeutics that target HIV protease however the ability of the
virus to rapidly acquire resistance to these therapies has led to a continuing need for
candidate drugs. A cost effective and robust reporter such as CA-GFP would allow for
high throughput screens of activity in the presence of potential inhibitors as well as assist
in directed evolution approaches studying this rapid acquiring of resistance. The
substrate recognition sequence for HIV protease is significantly longer than that of the
previously discussed reporters. Though HIV protease recognizes several sequences we
chose the sequence of the transframe (TF) – protease (PR) domain cleavage site;
VSFNFPQITL (24) (Table 4.2). HIVA-GFP was also dark when expressed in \textit{E. coli}.
Co-transformations with the protease did not yield any increase in fluorescence. This is
most likely due to the fact that HIV protease forms inclusion bodies when expressed in
bacteria and is therefore not available in the cytosol as was the case for the previously
tested proteases on other PrA-GFPs.

Due to the difficulties of expressing active proteases, so far we were unable to
verify the successful activation of or DVPA or HIVA-GFP reporters. However,
development of these reporters was nonetheless informative. It shows that we can alter
and even extend the linker region while maintaining the dark state of the reporter. This
indicates that though the tail remnant after cleavage could affect fluorescence recovery, the length or the sequence of the linker itself is not as essential as the quenching peptide in maintaining the dark state of the PrA-GFPs. This further fortifies our hypothesis that CA-GFP is a flexible platform that can be engineered to fit into the need of a broad protease activity monitoring.

4.2.5. Engineering various colors of PrA-FPs for multi-protease reporting

As discussed above, caspase-6 has a unique role in neurodegeneration and is therefore a target of drug development independent of the other executioners. With an effective C6A-GFP we can monitor caspase-6 activity in neurons and other cell types to determine specific activity levels in comparison to the activity levels of caspase -3 and -7. We could also monitor the activity of executioner caspases simultaneously to ensure that caspase-6 is self-activating in these situations and not being activated by any of the initiator or even inflammatory caspases through an unknown mechanism. The ability to simultaneously monitor multiple caspases at once depends firstly on the ability to have multiple reporters with linker sequences correlating to the reporter preference of each caspase and secondly these reporters need to have minimally overlapping spectral properties. For example if C6A-CerFP, C7A-CitFP, and a C1A-mNeptune were all monitored simultaneously in neuron cells we would be able to observe caspase-6 activation independently of -7 as well as monitor any conditions which lead to the inflammatory caspases (caspase-1) contributing to caspase-6 activation (25). These experiments are currently being done using antibody recognition of full-length and cleaved caspases (26) however, this requires the expensive and time consuming process
of developing antibodies for each form of each caspase. It also is assuming that the activity of the caspase is strictly correlated with its form, which while true in vitro, may not necessarily be the case in all in vivo scenarios. This method also requires fixing and staining of samples for imaging. The CA-FP’s would allow for real-time imaging in whole organisms, allowing for a much more in depth analysis of the molecular processes.

As several of the various color shifted version of GFP result from only a few point mutations, we hypothesized that the color of the fluorescent protein in CA-GFP could be shifted to cyan and yellow versions without disrupting the mechanism of the dark state. We designed CA-Cerulean (27) (CA-CerFP) and a CA-Citrine (28) (CA-CitFP) versions of our reporter which both still contain the caspase-7 recognition sequence DEVD in the linker. These fluorescent proteins have minimally overlapping excitation and emission wavelengths and both have the highest

![Diagram](image-url)

**Figure 4.7.** The constructs for CA-CerFP, CA-CitFP, and CA-mNeptune. The constructs for CA-CerFP and CA-CitFP are identical to CA-GFP with the exception of the color shifting point mutations in the GFP portion of the reporter. CA-mNeptune has the quenching peptide fused to the N-terminus of the mNeptune fluorescent protein derived from *Entacmaea*. All three reporters have a significant response to active caspase (colored bar) when compared to the response inactive caspase (white bar).
quantum yield for their color class (29) giving them the most potential for the best signal
to noise ratio possible. Both versions are in a dark state when co-expressed in *E. coli*
with the inactive caspase-7 C186A and have a significant increase in fluorescence when
co-expressed with active caspase-7 (Fig. 4.7). This demonstrates that the CA-GFP
platform is adaptable and can be modified to produce various colors of reporters. We also
were motivated to see if the dark state existed in a version using a red fluorescent protein.
Red-shifted fluorescent proteins are becoming more widely used as their emission
properties fall within the ‘critical window’ for whole organism imaging due to the deep
tissue penetrance for red wavelengths of light and the absence of background
autofluorescence in tissues in this spectral region. The red fluorescent proteins (RFPs)
are derived from *Entacmaea quadricolor* (30). While they have a different protein
sequence to GFPs from *Aequorea*, the protein folding of the two families are strikingly
similar. We selected mNeptune (10) for its improved brightness and spectral properties in
the near IR region increasing its popularity for *in vivo* imaging. In our first design of CA-
mNeptune we fused the M2 peptide through an identical linker as CA-GFP to the C-
termius of mNeptune, however, this protein showed a considerable level of fluorescence
prior to cleavage. As we had already observed that an *N*-terminal fusion of M2 to GFP
resulted in a dark state prior to cleavage (nCA-GFP, chapter III) we reasoned that we may
see a similar effect with mNeptune. The resulting *N*-terminal fusion, CA-mNeptune has
a very low fluorescent background and over a thirty-fold increase in signal when cleaved
by active caspase after cotransformation in *E. coli*.

Though all three reporters have a very clear activation and are an improvement
over the signal to noise of comparable reporters the CA-mNeptune, CA-CerFP and CA-
CitFP reporters differ greatly in their overall fluorescence intensity as well as their fold-increase in fluorescence over the background from CA-GFP. We reasoned that this is due not only to the difference in published quantum yields of the mNeptune (10), Cerulean (27) and Citrine (28) variants but also due to the differences in folding kinetics of the different fluorescent proteins (31,32) (see Chapter III) (Table 4.3).

**Table 4.3.** The values for the increase in fluorescence for each reporter as well as the reported quantum yield for each fluorescent protein and the brightness of each reporter after cleavage by active caspase-7 in *E. coli* lysates relative to CA-GFP.

<table>
<thead>
<tr>
<th>Reporter</th>
<th>Fold-Increase upon Cleavage</th>
<th>Relative Brightness</th>
<th>Quantum Yield of Fluorescent Protein*</th>
<th>Ex. λ (nm)</th>
<th>Em. λ (nm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>C7A-GFP</td>
<td>45±4</td>
<td>1</td>
<td>0.66 (33)</td>
<td>475</td>
<td>512</td>
</tr>
<tr>
<td>C7A-CerFP</td>
<td>18±2</td>
<td>0.06</td>
<td>0.62 (27)</td>
<td>433</td>
<td>475</td>
</tr>
<tr>
<td>C7A-CitFP</td>
<td>12±7</td>
<td>12</td>
<td>0.76 (28)</td>
<td>513</td>
<td>529</td>
</tr>
<tr>
<td>C7A-mNeptune</td>
<td>32±5</td>
<td>0.002</td>
<td>0.20 (34)</td>
<td>530</td>
<td>650</td>
</tr>
</tbody>
</table>

In designing the CA-mNeptune we began with a C-terminal fusion of the quenching peptide similar to the CA-GFP constructs, however, we found that we see lower fluorescence background before cleavage when the peptide is fused to the N-terminus.

We designed N-terminal constructs in the GFP reporters as well but observed much lower fluorescence recovery (see Chapter III). We reason that this difference is due to the differences in the folding intermediates of the two different proteins and their respective abilities to fold into the conformation necessary for chromophore maturation after cleavage of the peptide. The four fluorescent proteins described here; GFP, Cerulean, Citrine, and mNeptune constitute a panel of spectrally distinct fluorescent proteins which have individually been engineered for improved brightness in their respective families of color (cyan, yellow and red). Each can be used individually to monitor specific proteolytic events or together to watch several events using multi-channel imaging. The CA-mNeptune in particular has potential for effective use in whole animal models as
mNeptune’s fluorescence falls in the ‘near-infrared window’ which minimizes background from water and hemoglobin in tissue.

4.3. Conclusion

CA-GFP has proven itself a useful and effective reporter of caspase activity and apoptosis in a variety of systems (Chapter II). Throughout this chapter we have shown that CA-GFP can be manipulated not only to change the protease specificity of the reporter but the color of fluorescence as well. By expanding the color palette PrA-FPs can be used in multiplex microscopy, FACS sorting, and directed evolution applications. While there are challenges in engineering PrA-FP’s they are similar to the challenges that exist in the development of any reporter. The reporter can only be as specific as the protease it is monitoring. The character of the protease is the most important factor in the design, all potential processing sites must be considered as well as the propensity of the protease to cleave in specific secondary structure conformations, such as a loop. The ability of multiple proteases to cleave a similar substrate is also a major limitation as in the example of the LEHDF-GFP, which can be cleaved by hR3C protease but is also the substrate for caspase-9. This overlapping specificity may be overcome by the use in conjunction with small molecule inhibitors for one class of proteases while monitoring the activity of the desired protein.

In order to be effective in isolating the activity of a specific caspase independently of the others it is imperative that the reporter only responds in the presence of the selected protease. Our profiling results indicate that while we see significantly higher fluorescence in the case of the caspase the reporter was designed for, there is still
signal from the other caspases as well and the level of selectivity is not ideal. The ability of caspases to cleave secondary sites, as we found in caspase-6 may be used to our advantage if we design a reporter which could be activated in the presence of the desired caspase but inactivated by secondary cleavage if multiple caspases are active. This is obviously not a useful reporter in the case of apoptosis where all caspases are activated, but could be useful in verifying the unique non-apoptotic activities of caspases, such as the activation of caspase-6 and only caspase-6 in neuronal cells; if caspases other then caspase-6 are active the we will not see any fluorescence. This reporter would also necessitate a multiple reporter system, similar to that used in mammalian cells in chapter II, where a constitutively active fluorescent protein is co-expressed to ensure expression of the reporter.

Our observation that caspases have not had the same affinity for their respective PrA-FP’s as reported for peptide substrates correlates with the idea that the small peptide substrates that have been widely used to characterize protease activity may not be reflecting the entire selectivity properties of proteases and that when presented with a protein substrate the sterics and kinetics are significantly different. The size and active site conformations of some proteases may not allow for efficient processing of the linker regardless of the sequence. A better understanding of the sterics and geometry of the linker region are necessary to ensure that it is fully accessible. It has also recently been recognized that small tetrapeptide substrates do not take into account any P1’ preferences of the proteases. Recently it has been reported that though caspases-3 and -7 share the same P1-P4 specificity that the P6, P5, P2’ and P3’ residues were able to contribute to caspase-7 specificity in substrates (35). These extended recognition properties may not
only suggest why we do not observe the same cleavage kinetics as with tetrapeptide substrates but also gives us an advantage in designing specificity as more is understood about the extended recognition preferences of caspases.

The study of proteases and their activity is an expansive and important genre of research and we hope that our examples of the versatility of the CA-GFP reporter platform encourage other researchers to utilize this technology in their respective fields.

4.4. Materials and methods

4.4.1. Molecular cloning

Caspase-6 and hR3C protease reporters: CA-GFP in pBB75 contains a DEVDFQ linker as previously described (3). All caspase-6 and hR3C protease reporter variations, unless specified, are based on this construct with the DEVDF sequence in CA-GFP mutated to corresponding sequences by site-directed mutagenesis using overlapping inverted primers and amplification of the entire plasmid similar to the QuikChange® (Agilent) approach. In all constructs denoted GFPΔ (Table 4.1), an additional D27E mutation has been introduced by site directed mutagenesis to eliminate the internal caspase-6 recognition site (VELDG, residues 24-28) in GFP.

DVPA-GFP and HIVA-GFP were generated using site-directed mutagenesis using overlapping inverted primers and amplification of the entire plasmid similar to the QuikChange® (Agilent) approach on the linker region using CA-GFP as the parent plasmid. DVPA-GFP_v1 and _v2 were also generated similarly by introducing mutation in the linker region of DVPA-GFP.
CA-CerFP and CA-CitFP: The CA-CerFP construct was developed by replacement of the codons for the following residues in CA-GFP F64L, S72A, Y66W, Y145A, and H148D. These substitutions were made through multiple rounds of single point mutagenesis using a modified QuikChange® (Agilent) protocol with CA-GFP in pET21b as the template plasmid. The CA-CitFP construct was developed also using CA-GFP in pET21b as the template plasmid for replacing residues T65G, V68L, Q69M, S72A, and T203Y through multiple rounds of modified Quikchange® mutagenesis.

The CA-mNeptune expression construct was generated by amplifying the peptide (M2) region from the CA-GFP gene separately from the mNeptune gene using two primers to amplify M2 and two primers to amplify mNeptune. The mNeptune parent was generated by replacement of the codons for M41G, A45V, S61C, A158C, and Y197F in the mKate2 gene in the pmKate2C vector (Evrogen). The first primer (P1) annealed to the N-terminal region of the mNeptune gene with the linker sequence included N-terminal to the mNeptune gene. The second primer (P2) annealed to the C-terminus of mNeptune and included a stop codon (UAA) and the restriction site for XhoI. The third primer (P3) included an NheI restriction enzyme site as well as a 6His sequence and annealed to the N-terminal region of the M2 portion of the CA-GFP gene. The last primer (P4) annealed to the C-terminus of M2 and included the same sequence for the linker as P1, giving primers P1 and P4 a 24 bp overlapping region. The mNeptune fragment was then amplified using primers P1 and P2 while the M2 portion was amplified using primers P3 and P4. After gel purification of the amplified fragments they were combined and allowed to anneal through the overlapping region for five PCR cycles before the addition of primers P2 and P3 which then amplified the full length gene. The gene was then
ligated into pACYC pETDuet vector into NheI and XhoI sites. The final sequence of the N-terminal peptide and linker is
MHHHHHHMCNDSSDPLVVAASIIGILHLILWILDRLDEVDFQGP which is appended to the N-terminus of mNeptune. The full-length construct was then further amplified using PCR and ligated into the EcoRI and NotI sites of pET21b. (We thank Jun Chu for construction of the CA-mNeptune gene in the pACYC vector).

4.4.2. Fluorescence assays in cell lysates

Expression constructs for the fluorescent protein reporters CA-GFP, CA-CerFP, CA-CitFP, and CA-mNeptune, each in the pET21b vector, which also contains an Ampicillin (Amp) resistance gene, were co-transformed via electroporation in BL21(DE3) strain *E. coli* with expression constructs for an active constitutively two-chain version of caspase-7 or a full-length inactive version of caspase-7 in which the active site cysteine has been mutated to alanine (C186A) in the vector pBB75, which also contains a Kanamycin (Kan) resistance gene. 50 mL LB cultures supplemented with 100 µg/mL ampicillin and 40 µg/mL kanamycin were inoculated with 50 µL from a dense 5 mL overnight culture and incubated until reaching an OD$_{600}$ of 0.6. The cultures were then induced with 1 mM IPTG for 18hrs at 25°C. From each sample 800 µL aliquots were taken in duplicate, pelleted by centrifugation and resuspended in a buffer of 0.5 mg/mL lysozyme and 2 units DNAase. After lysis with four cycles of freeze-thaw the supernatant of each sample was measured for fluorescence. CA-GFP (Ex. 475/ Em. 512 nm), CA-CerFP (Ex. 433/Em. 475 nm) and CA-mNeptune (Ex. 530 / Em. 650 nm) fluorescence measurements were taken in a co-star 96-well black plate on a Molecular
Devices Spectramax M5 spectrophotometer. CA-CitFP (Ex. 515 / Em. 529 nm)
fluorescence was measured on a JASCO FP-6500 spectrofluorometer using a quartz cell. Fluorescence values were normalized to the total number of cells in culture, which was estimated from the relative OD$_{600}$ of each culture.

4.4.3. Fluorescence assays in whole cells

Expression constructs for caspase-6 or hR3C protease reporters contained in the pBB75 vector (Kan) were co-transformed into the BL21(DE3) strain of *E. coli* with either an active constitutively two-chain version of caspase-6 (13) in pET11b (Amp), an hR3C protease in GEX- vector (Amp), a gift from M. Romanowski, or with an empty pET21b (Amp) vector as a control. 2 mL auto-induction media (AIM) cultures were inoculated from a dense 100 µL overnight LB(Amp/Kan) culture with 100 µg/mL Amp and 40 µg/mL Kan and incubated until reaching an OD$_{600}$ of 0.6-0.9. The cultures were then incubated for 18hrs at 16°C. From each sample, the cells are rinsed with PBS, resuspended and diluted in PBS to a final OD$_{600}$ of 2.0. 200 µL of this diluted suspension was measured for GFP fluorescence (Ex. 475/ Em. 512) in a co-star 96-well black plate on a Molecular Devices Spectramax M5 spectrophotometer.

4.4.4. Protease profiling of C7A- and C6A- GFP$^e$

Expression constructs for the reporters C7A-GFP$^e$ and C6A-GFP$^e$ in the vector pBB75 with Kan resistance were co-transformed into the BL21(DE3) strain of *E. coli* with expression constructs for either active or inactive variants of caspases -3, -6, -7, -8, and -9 in pET vectors, which also contain an Amp resistance gene. In the case of
caspase-3 expression construct, the active version was the wild-type human caspase-3 sequence in pET23b and the inactive version was the active site knock out C163A. For caspases-6 (pET11a) and -7 (pET23b), the active versions are constitutively two-chain constructs that express the large and small subunits separately and which require no further processing for activation. The inactive version of caspase-6 is a full-length uncleavable (FLUC) (pET11a) version in which all self-processing sites have been mutated to alanine (D23A, D179A, D193A) to prevent cleavage into the large and small subunits therefore yielding the protein inactive. The inactive version of caspase-7 is the full-length human sequence with the active site cysteine mutated to alanine to prevent activity. In the case of caspase-8 both the active and inactive versions are the human gene sequence which lack the N-terminal death effector domain (DED) in the vector pET15b, with the inactive mutant being the active site knock out C285A.

After co-transformation, 50-mL LB cultures were inoculated from a dense overnight culture with the appropriate double antibiotic selection. Cells were grown, induced, and lysed as described above in the lysate fluorescence assays section. Fluorescence measurements of the supernatant samples were taken in a co-star 96-well black plate on a Molecular Devices Spectramax M5 spectrophotometer (Ex. 475/ Em. 512). Fluorescence values were normalized based on the relative OD of each culture.

4.4.5. Western Blotting

Samples of supernatant were taken from *E. coli* cell lysate prepared as described above for the lysate fluorescence assays. Identical SDS-PAGE gels were run of each sample to be blotted with the required primary antibody. One of the identical SDS-
PAGE gels was transferred to nitrocellulose and blotted with an anti-GFP (Millipore) monoclonal mouse primary antibody. The western blots were then treated with anti-mouse IgG alkaline phosphatase produced in goat (Sigma) and visualized using 1-Step™ NBT/BCIP (Thermo Scientific). The identical gel was stained with Coomassie.

4.4.6 *in vitro* analysis of DVPA-GFP

DVPA-GFP, DVPA-GFP_v1 and DVPA-GFP_v2 were each incubated with DVP2 at a 1:1 molar ratio in 100mM Tris pH 8.0, 50 mM NaCl at room temperature overnight. Fluorescence was measured using a Molecular Devices Spectramax M5 spectrophotometer monitoring fluorescence (Ex. 475/ Em. 512) over 18-24 hours.

4.4.7 Two Dimensional Gel Analysis

DVPA-GFP_v1 and DVPA-GFP_v2 were incubated 18 hrs at room temperature in 100 mM Tris pH 8.0, 50 mM NaCl with the active DVP2 at a concentration of 5 mg/mL for each protein. After 18 hrs the samples were diluted 10-fold in a rehydration buffer of 8M Urea, 2% CHAPS, and 50 mM DTT. IPG strips (7 cm, Biorad) with a pH range of 4-7 were incubated at room temperature in a volume of ~130 µL per sample. After 1 hr the strips were covered with mineral oil (~1mL) and incubated at room temperature for ~12 hours. After rehydration was complete the mineral oil was removed. Isoelectric focusing was run on a Biorad Protean IEF Cell instrument on the ‘rapid’ ramping mode. Once the method was complete the samples were stored at -80°C overnight. The samples were thoroughly thawed and incubated in Equilibration buffer 1 (6M urea, 2% SDS, 0.05M Tris HCl pH 8.8, 20% glycerol, 2% DTT), in a volume of 2.5
mL. After a 10 min incubation the buffer was decanted and replaced with 2.5 mL of equilibration buffer 2 (6M urea, 2% SDS, 0.05M Tris HCl pH 8.8, 20% glycerol, 2.5% Iodoacetamide), and incubated for another 10 min. The IPG strips were then loaded onto a 16% acrylamide SDS-PAGE gel and overlaid in a solution of 0.5% agarose in 1X SDS-PAGE running buffer with 0.003% bromophenol blue as an indicator for the gel front. After the solidification of the agarose the SDS-PAGE gel was run to completion and stained with coomassie.

References


108

CHAPTER 5
CA-GFP DEVELOPMENT AND APPLICATIONS

5.1. CA-GFP: A model reporter of apoptosis

Apoptosis is arguably one of the most widely studied cellular processes. A PubMed search of the keyword ‘apoptosis’ yields over 92,000 references just in the last five years. It is critical that we understand each step of the cascade not only in disease models but also in normal cellular development and homeostasis. As we develop drug targets throughout the apoptotic pathway it is essential we understand the ramifications of treatment on healthy cells as well as those specifically targeted. In order to fully understand these intricacies the ability to study apoptosis longitudinally in whole animal models is essential.

In chapter I we set forth our criteria for an ideal apoptotic reporter. They were that it should be genetically encoded for non-invasive monitoring. It should respond independently of any cellular co-factors in a dark to bright mechanism for ease in detection. Lastly it should be amenable for in vitro, E.coli, FACS, and mammalian cell applications. Throughout this thesis we have underscored how CA-GFP fits each of these criteria. We have sought to characterize not only the function but the mechanism by which this reporter operates in an effort to emphasize the potential of our system. While CA-GFP has proven effective in it’s own right, as outlined in chapters II-IV there are challenges in engineering CA-GFP to respond to other proteases, however they are not insurmountable. The first challenge in making CA-GFP a more ideal reporter system is improving the kinetics of the fluorescent response to cleavage. By engineering a version in which the fluorescent response was near instantaneous it would allow for kinetic
analysis of protease activity *in vivo*, similarly to the method in which small fluorogenic substrates are currently used *in vitro*. The other major challenge in the development of CA-GFP remains the ability to thoroughly control the effects of the peptide. With a detailed understanding of the peptide interactions the quenching mechanism has the potential to allow for the activation of GFP fluorescence in response to other enzymatic reactions such as phosphorylation or protein conformational changes in signaling pathways. Lastly there is more work to be done in obtaining an optimal caspase activatable- fluorescent protein (CA-FP) for use in whole organism models as this application holds the greatest potential impact for CA-FP use.

**5.2. Analysis of background fluorescence**

The most unique and advantageous feature of CA-GFP and its derivatives is the presence of the dark state. As discussed in chapter III our model for the existence of the dark state is that the presence of the peptide stalls CA-GFP in it’s sample of native folding conformations leaving CA-GFP in a ‘pro-folding’ intermediate state. There is a noticeable difference in background between species, different colors, and different linker sequences though all background levels are significantly less than the fluorescence of the cleaved form. The assumption is that some level of background is from autofluorescent proteins in the bacterial cells or lysate as it is commonly accepted that there is significant background in this range of the spectrum. Some portion of the background is also presumably from some fraction of the CA-FP molecules which have achieved the mature conformation despite the presence of the peptide. The obvious explanation for the variation from one reporter to another is that a different portion of these molecules have
achieved this mature conformation. From our model in chapter III it would imply that the folding rates of each reporter is slightly different. It has been observed that different versions of the fluorescent protein family have widely different folding and maturation rates, even when differing only by a few point mutations. When comparing the dark state of the reporters with the range of linker sequences we have constructed (Fig. 1) we see that the constructs which have bulky hydrophobic residues in the linker have a lower background level than those with charged residues in the linker. As we observed previously these reporters have a lower fluorescence recovery when cleaved and so it correlates that the background of these versions is also lower. Lengthening the linker does not appear to have a significant effect on the background in the case of HIVA-GFP (linker sequence: VSFNFPQITL, Fig. 1), however this linker sequence also contains several hydrophobic residues which lead to a decrease in fluorescence in the original linker. It would be interesting to see if a longer, more charged linker shifted the background properties significantly. It would also be useful to test the effect of longer more standard fusion linkers such as a Gly-Ser linker (1) and determine if this made any difference when still fused to the peptide.
The background also varies greatly based on any point mutations within the fluorescent protein portion of the reporter as well. In Figure 2 we see that the introduction of the

**Figure 5.1.** Comparison of the relative background fluorescence of reporters cotransformed with an empty pET vector and are therefore uncleaved. The different reporters are identified by their linker sequence. Reporters with more hydrophobic resides in their linker tend towards lower background fluorescence.

**Figure 5.2.** Comparison of the relative background fluorescence in caspase activatable reporters of varying versions of GFP when co-transformed with inactive caspase-7 (C186A).

The background also varies greatly based on any point mutations within the fluorescent protein portion of the reporter as well. In Figure 2 we see that the introduction of the
F64L mutation to the S65T background used in CA-GFP, which comprises the widely used EGFP version (2), drastically changes the background of fluorescence. F64L is known to have a much higher quantum yield than S65T, which is why it is widely used, and so it is possible that a similar portion of the uncleaved reporter is maturely folded however the fold increase of CA-EGFP after cleavage is only 4-fold, indicating that either there is a higher percentage already maturely folded or it is more difficult for this version to attain the mature conformation once it is cleaved, essentially a greater fraction of the cleaved CA-EGFP are trapped in the ‘pro-folding’ state rather than continuing on in the folding pathway. As the F64L mutation is only one residue from the chromophore it is possible that this mutation directly affects the chromophore’s folding geometry. In the case of CA-CerFP and CA-CitFP there are several point mutations near the chromophore which shift the fluorescent properties, these mutations could also affect the folding of the chromophore region. However some of these mutations are on the protein surface which could effect the overall folding pathway which these reporters sample, again shifting the small percentage of molecules which can attain the mature geometry accounting for this dark state. Though it does not account completely for the discrepancy Cerulean and Citrine fluorescent proteins also have very different quantum yields (chapter IV) from GFP S65T which again would play a role in the background level. The potential for the spectral shifting point mutations to change the folding efficiency and pathway may also explain the discrepancy we see in the quantum yield of the fluorescent protein vs. the brightness of the reporter. For example CA-CitFP is 12-fold brighter than CA-GFP however the quantum yield of Citrine is only a 10 percent increase from GFP.
These point mutations may also be having an effect on the percent conversion to the mature folded form.

In the case of superfolder GFP, which was specifically engineered to have a much faster folding and maturation rate, we do not observe a dark state. It is difficult to directly compare CA-sfGFP and CA-GFP as both the linker, which is shortened to DELD (chapter III) and the fluorescent protein have been changed simultaneously but it is obvious that this dual effect leads to the lack of the dark state. This supports the argument that the length and chemical character of the linker along with the folding efficiency of the fluorescent protein are two of the biggest contributing factors to the dark state.

5.3. Peptide quenching

Though it is clear that the linker plays a role in the folding and therefore fluorescence of GFP after cleavage one aspect of the quenching mechanism which needs to be fully understood is the specific interactions of the peptide. Our model of how the peptide quenches fluorescence (chapter II) is that it inhibits proper folding of the GFP barrel until it is removed by a protease. This does not address the specific interactions that the peptide has with the residues of GFP which leads to the lack of proper folding. One way to investigate these interactions more specifically, in the absence of structural coordinates which we have been unable to obtain, is to perform a series of truncations of the M2 peptide to determine a minimal length and sequence necessary for quenching. This may also help to address the low level of toxicity of the current CA-GFP construct in mammalian cells which has not allowed the development of a stable cell line.
containing any of the CA-FP constructs. While it is known that fluorescent proteins have some level of toxicity due to the generation of reactive oxygen species in the chromophore maturation reaction it does not prevent culturing of stable cell lines of the independent fluorescent proteins. Therefore it appears as though the peptide sequence may be contributing to this effect and a shorter length may decrease the detrimental effects.

Another possibility is to fuse other helical bundle domains through a similar linker to determine if it is the hydrophobic nature or the specific sequence that is playing a larger role. One such peptide is the KCNQ (Kv7) potassium channel A-domain. The Kv7 tail is a tetrameric coiled-coil in solution however unlike M2 it is not the transmembrane but the cytoplasmic portion of the protein (3). As this peptide has been shown to be tetrameric in solution even when fused to a larger protein, the probability is that it would also tetramerize when fused to GFP.

5.4. Expression and codon usage

Another case where we have observed differences in the dark state is based on the vector and expression conditions used for co-expression of the reporter. One of the major challenges in being able to fully achieve the potential impact of PrA-FP reporters will be to fully understand implications of vector and strain in terms of expression. Throughout our studies one of the recurring challenges we have faced was optimization of co-expression conditions of our reporter with the protease. We discovered that not only does co-expression require a double antibiotic selection but also incorporation of the two respective genes on plasmids containing different origins of replication. After trials of
several different plasmids, including a pET Duet vector specifically designed for co-expression applications, we found many instances where either the reporter or protease was not well expressed. More confounding was instances where caspase-7 would be co-expressed but not processed. This was remedied by using a constitutively two-chain version (or CT version) of the caspase in which the large and small subunits are expressed, negating the necessity for processing for activity. The ultimate co-expression combination we determined was optimal was a combination of the pBB75 vector and pET21b (or any pET vector). Using these vectors we were able to co-express the protease and reporters from either vector interchangeably, although there is some variability in the background from each of these as well. This may be due to some subtle expression differences such as the copy number of each plasmid. The mRNA sequences (from the T7 to T7terminal sites in the plasmid) when compared does not have significant differences indicating that subtle sequence differences are less likely than the inherent differences of the plasmids.

Another consideration is the codon sequence for the reporters. The codon sequence for CA-GFP used in all cases throughout this thesis was not optimized for \textit{E. coli} codon usage (4), but a synthetic version with S65T mutation originally designed for plant gateway vector use (5). However, there is a only low percentage (4\%) of rare codons for \textit{E. coli} and so expression is not drastically inhibited. This may mean that the overall kinetics of translation are slowed, which could change the landscape of the folded states sampled therefore effecting the dark state of the reporter. CA-mNeptune is also not codon optimized for \textit{E.coli} and has 8\% rare codons though we still see considerable expression levels in our experiments. In engineering of a reporter that has faster
maturation kinetics one of the challenges mentioned above was that these versions are also optimized for faster folding (6, 7), which could effect the background level. One possibility is that the introduction of codons less frequently used in E. coli may sufficiently slow the translation, allowing for the sampling of folding intermediates necessary for achieving the ‘pro-folding’ state (8). Previous studies have shown that codon usage has a significant effect on GFP. In a study by Kudla et al. it was determined that fluorescence levels varied 250-fold across 154 genes that all encoded the same GFP protein sequence (9). This could potentially allow for the advantageous dark state of the Pr-FP while allowing for much faster observation of cleavage by having a more quickly maturing version of the fluorescent protein.

Ultimately there are other sources for background as well. Though we see what appears to be similar expression levels of the overall amount of reporter in co-transformation of C7A- and C6A-GFP reporters with different caspases an analysis of the background in each case shows significant variations between the fluorescence of the ‘dark’ state. In the case of the caspase-6 full length- uncleavable construct, the active site cysteine is still present therefore though it is reported to have very low activity it is possible that this construct could self-activate (10) which could explain the higher background in this case (Fig. 5.3). However for caspases -3, -7, and -8 the inactive mutant is an active site mutation to alanine, preventing the possibility of catalytic activity and therefore cleavage.
Figure 5.3. Relative background fluorescence of C6A-GFP\(^e\) and C7A-GFP\(^e\) when co-transformed with the inactive versions of caspases -3, -6, -7, and -8.

### 5.5. Kinetic Control

Ultimately the kinetics of maturation is one of the aspects of CA-GFP most in need of improvement. While the fluorescence in mammalian cells appears within 20 min, \textit{in vitro} the maximal fluorescence isn’t achieved for four hours (chapter II). In a model where one can imagine watching the apoptotic cascade through a panel of reporters at different wavelengths for each caspase it is critical that the response to cleavage be essentially immediate, as is the case with small molecule fluorescent substrate probes. The ability to control this aspect of the reporter lines primarily in attaining a full understanding of the specific detailed interactions the linker and peptide have in the structural context of CA-GFP. While we understand the big picture of the dark state
mechanism, in that the presence of the peptide appears to sufficiently stall GFP folding in a ‘pro-folding’ state prior to chromophore maturation, we must understand these specific details in order to best engineer them. One example as given above is to control the folding of a faster maturing version of GFP, such as the superfolder GFP which matures within five minutes (7). Another example would be if we could mimic this ‘pro-folding’ conformation in a version of GFP in which the chromophore is allowed to mature but not fluoresce. To my knowledge the only examples of this in the literature to date are versions which can only be used in vitro due to the necessity of denaturing the protein to destabilize it once the chromophore is formed. However, one can imagine that if the interaction is specific that perhaps through fusion of the peptide to another strand or through circular permutations of the reporter this could be achieved.

5.6. CA-FP in Drug Design: Back to Biology

In an interview after receiving his Nobel prize for his work in the development of GFP and it’s ‘toolbox’ after being asked what he hoped to ultimately achieve from his research Roger Tsein’s response was he wanted to develop something ‘clinically beneficial’(11). Ultimately the same sentiment is echoed throughout this thesis. Before we can begin to understand how to fix biological processes (ie. Develop treatments) we first need to fully understand how they work. Then we need to understand the full effect, in whole organisms not only of the disease state but what happens when we perturb the healthy or unhealthy systems with potential therapies. GFP has provided us a method by which to effectively and relatively uninvasely do this since it’s development in the 1960’s, now more than 50 years later, we are still finding new ways to use GFP and it’s now
countless derivatives to uncover more information into the beautiful intricacies of cellular biology.

The ultimate vision for CA-FP is to see it used in transgenic model organisms. A CA-FP transgenic mouse model has the potential to be a powerful tool in the study of the role of apoptosis in development as well as a screening tool for drug development. With this tool researchers would have the potential to map out the developmental apoptosis map in mice as they have in more microscopy amenable organisms such as zebrafish (12) and C. elegans (13). This mapping of apoptosis would allow for interrogation of genetic factors in development, differentiating them from environmental factors. The role of metals and other co-factors in apoptosis could be observed in all organismal systems.

Such a model organism would also be a powerful tool in drug toxicity screening. Drug design for the past several decades has focused on the study of the effects of molecules on a pathway or system independent from the intricacies of how each pathway is connected in a cell or organism. Though this approach has been effective in some cases it is clear that for success moving forward a more detailed understanding of the underlying biology is a necessity. In the case of both mice and zebrafish a transgenic line of organisms could be used to longitudinally observe the effects of potential therapies, observing localized fluorescence in regions the drug was triggering apoptosis while being able to assess the overall viability of the organism. A zebrafish line could potentially be used in a high-throughput application for this purpose.

Each insight we gain into the intricacies of apoptosis underscores how interwoven and compensatory systems originally believed to operate independently are. For example apoptosis has long been a central area of study for cancer researchers working on
methods by which to induce cell death. Recent reviews (14-16) highlight the work showing how closely the cellular mechanisms for not only apoptosis but other methods of cell death such as autophagy and necrosis are with the web of metabolic pathways in the body, opening up whole new pharmacological and basic research directions. The ability to assess the function of a single protease in this complicated web of cellular function is critical to being able to tease out the independent role of critical proteins in these processes. If we had the ability to simultaneously monitor the activity of initiator, executioner, and inflammatory caspases we could better understand the delicate regulatory balance that exists within the cell. However, caspases are near the end of this cascade effect, in order to monitor processes more difficult to assess \textit{in vivo} such as post-translational modifications a more complete understanding of the dark state is necessary to enable appropriate engineering of this class of reporters. With complete control of the quenching mechanism a class of reporters could be designed to respond to any number of cellular reactions and interactions including phosphorylation, ubiquitination, and glycosylation. Each of these modifications is critical in the function of countless proteins in every cellular process. A more detailed understanding of the events which elicit these responses and their consequences in the cell is critical for understanding the ramifications of current and potential therapies both in terms of bolstering efficacy and minimizing adverse side effects.

\textbf{References}

APPENDIX A

MOLECULAR DYNAMICS SIMULATIONS OF PEPTIDE STABILITY

A.1. Introduction

In a collaborative project to design helical peptides to be used as regio-regular scaffolds for proton transfer we screened several helical peptides in silico. The focus of this project is to understand proton transport in order to develop better materials for proton exchange membranes in hydrogen fuel cells. Our goal was to build a helical peptide with a photo acid generator on the N-terminus and a proton sensitive probe on the C-terminus. By using a peptide as the scaffold the placement of the proton transferring moieties can be easily controlled. The proton transporting moiety we chose was the naturally occurring amino acid –histidine. Histidine is an ideal candidate as its pKa is typically 6 and can exist as the imidazole or the imidazolium moiety, easily accepting and donating a proton. Since we are using amino acids we can easily control the direction of the transfer either through a linear mechanism as shown here, or by a zigzag mechanism simply by changing the placement of the histidines in the sequence (Figure A.1). In order to place the histidines moieties in a specific placement we first needed a sequence that would remain helical under organic solvent conditions needed for single molecule measurements.

Our search for alpha helical proteins and peptides in the literature yielded many sequences, far to many to efficiently synthesize. The literature sequences also only report the structure in an aqueous environment as it is the biologically relevant condition,
though it is not an ideal media for the proton transport measurements we hoped to do using single molecule spectroscopy. We decided to use molecular dynamics as an *in silico* screen for helicity of peptide sequences in organic solvents.

We also designed a series of synthetic peptides using the non-natural amino acid amino isobutaric acid. When we synthesized these peptides with alternating alanine/AIB residues they showed a propensity for helicity in organic solvents as the AIB locks the phi/psi angles of the peptide into an alpha helical conformation. We also screened these sequences *in silico* to compare them to the natural helical sequences.

**Figure A.1.** Design of a helical peptide region-regular scaffold for proton transport. Our design included a $N$-terminal photo-acid generator and a $C$-terminal proton sensitive probe to indicate proton transfer across the histidine side chains of the helix.

### A.2. Results

The program we chose was GROningen Machine for Chemical Simulations (GROMACS (1)) which is an open source molecular dynamics program that is compatible with several force fields, was applicable to our system. We have employed molecular dynamics simulations in GROMACS to model the conformation of peptides as a first-order screening mechanism. We selected GROMACS because it is flexible, open source software designed specifically for molecular dynamics of biological molecules and accommodates a variety of force fields. Most importantly it allows modeling of explicit solvent molecules, which is critical for our analysis. Unlike many molecular
dynamics packages that cater to protein and peptide simulations, GROMACS can be parameterized for many different solvents.

As a control we chose GCN4 (PDBID: 2ZTA (2)), a leucine zipper transcription factor known to be helical in water and ran a control simulation in water for 5 ns. At the completion of this simulation the peptide had remained helical. We then moved to one of the peptide sequences we had selected, which had existing NMR coordinates in the protein data bank PDBID: 2MAG (3), an abbreviation for Magainin2 a peptide used as an antibiotic, and ran a simulation again using explicit water as a solvent (Table A.1). After confirming that 2MAG stayed helical in water as expected efforts were directed to being able to simulate the effects of DMSO, which was originally suggested as a potential solvent suitable for single molecule spectroscopy. After optimizing the parameters a 10 ns simulation of 2MAG in DMSO was run and, as seen in Figure A.2, it unfolds almost immediately and never regains any of its’ helicity. We used the ending coordinates of the now unfolded 2MAG and ran a 30 ns simulation (total, 10 ns and then 20ns from these coordinates, Table 2) in water as the solvent to determine if 2MAG could refold into a helix under these conditions. Though it attempts to refold into a helix 2MAG never regains it’s initial conformation.

![Figure A.2. Comparison of peptide 2MAG (initial coordinates shown in green) after a 1 ns simulation in H2O on left (cyan) vs. after 10 ns simulation in DMSO (blue).](image)
### Table A.1. Potential helical peptide sequences for region-rigid scaffold

<table>
<thead>
<tr>
<th>Peptide Sequence</th>
<th>Buffer</th>
<th>Oligomeric State</th>
<th>Method of structural determination</th>
<th>PDB ID</th>
<th>Length</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ac-WNA³LAAVAAÀ¹⁰LAAVAAÀ¹⁵LAAV(4)</td>
<td>Phosphate buffer/methanol</td>
<td>Monomer; aggregates</td>
<td>CD</td>
<td>21</td>
<td>Membrane helix</td>
<td></td>
</tr>
<tr>
<td>GIGAVLKVLTGLPALISWIKRRQO(5)</td>
<td>Aq</td>
<td>Monomer, tetramer</td>
<td>Fluorescence, CD, NMR, CD</td>
<td>26</td>
<td>Synthetic melittin</td>
<td></td>
</tr>
<tr>
<td>SLLSLIRKLIT(6)</td>
<td>TFE, SDS</td>
<td>CD</td>
<td>11</td>
<td>decoralin</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DTASDAAAAALTAA NAKAAAELTAAANAAA AAAATAR(7)</td>
<td>H₂O</td>
<td>Monomer or dimer</td>
<td>x-ray crystallography</td>
<td>1wfa</td>
<td>37</td>
<td>Antifreeze protein</td>
</tr>
<tr>
<td>Ac-EAEKEAKEAAKKKAA A-NH(8)</td>
<td>aq</td>
<td>Monomer</td>
<td>CD</td>
<td>16</td>
<td>74% Helical</td>
<td></td>
</tr>
<tr>
<td>Ac-NELKKKLELCKAKWL EAKKKLEALK-NH₂(9)</td>
<td>Aq</td>
<td>Monomer</td>
<td>CD &amp; analytical ultracentrifugation</td>
<td>25</td>
<td>89% alpha-helix at 0 °C; membrane binding</td>
<td></td>
</tr>
<tr>
<td>GAGKAAKGAAGAAK GAGK(10)</td>
<td>Aq</td>
<td>Monomer</td>
<td>CD, NMR</td>
<td>18</td>
<td>65% helical; amphipathic</td>
<td></td>
</tr>
<tr>
<td>WEAALAEALAEALAE HLAELAEALAEALAA(11)</td>
<td>Aq</td>
<td>Monomer</td>
<td>CD</td>
<td>30</td>
<td>GALA synthetic peptide; pH 7 = random coil; pH5 = alpha helix; mimetic of natural membrane proteins, W &amp; H are spectroscopic probes</td>
<td></td>
</tr>
<tr>
<td>GIGKFLHSAKKFGKAF VGEIMNS-NH₂(3)</td>
<td>25%TFE/H₂O; DPC micelles; SDS micelles</td>
<td>Monomer</td>
<td>NMR</td>
<td>2maga</td>
<td>23</td>
<td>Magainin2, antibiotic peptide; helical in all three solvent systems; membrane protein</td>
</tr>
<tr>
<td>QEQLEDARRLKA IYE(12)</td>
<td>Aq</td>
<td>Monomer</td>
<td>CD, NMR</td>
<td>15</td>
<td>40% helical, folds into helix as peptide; paper also</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1: Boc-V(\text{VAL})(\text{VAL})(\text{Lac})(\text{L(Aib)})(\text{VAL}) -OMe</td>
<td>1: acetonitrile &amp; H(2)O</td>
<td>Helices aggregate</td>
<td>x-ray</td>
<td>11, 11 &amp; 14</td>
<td>They substitute lactic acid for alanine into their helices to see if it will still accommodate this residue in the helical conformation</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>2: Boc-(\text{VAL})(\text{L(Aib)})(\text{VAL}) -OMe</td>
<td>2 &amp; 3: isopropanol and H(2)O</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3: Boc-(\text{VAL})(\text{L(Aib)})(\text{VAL})(\text{Lac})(\text{L(Aib)})(\text{VAL}) -OMe (\text{(13)})</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

includes mutants with 2 mutations
Table A.2. Log of simulations run using GROMACS. Files are stored on ‘Solitude’ as well as on the Hardy lab server. The File path of each simulation is indicated in the table as it appears on the ‘Solitude’ hard drive.

<table>
<thead>
<tr>
<th>Peptide</th>
<th>PDB ID</th>
<th>Date Simulation Completed</th>
<th>Time scale</th>
<th>RMS</th>
<th>File path</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>2mag</td>
<td>2MAG</td>
<td>7/14/08</td>
<td>1000 ps</td>
<td>6.61</td>
<td>~/gromacs/071408/DMSO</td>
<td>Completely Unfolded</td>
</tr>
<tr>
<td>2mag_final</td>
<td>2MAG</td>
<td>7/30/08</td>
<td>1000 ps</td>
<td></td>
<td>~/gromacs/072908/2magunfolded</td>
<td>did not completely refold</td>
</tr>
<tr>
<td>GCN4</td>
<td>2ZTA</td>
<td>7/31/08</td>
<td>1000 ps</td>
<td></td>
<td>~/.gromacs/072908/gcn4water</td>
<td>retained helicity</td>
</tr>
<tr>
<td>2mag</td>
<td>2MAG</td>
<td>8/7/08</td>
<td>10 ns</td>
<td>9.6</td>
<td>gromacs/080508_2mag_dmso</td>
<td></td>
</tr>
<tr>
<td>GCN4</td>
<td>2ZTA</td>
<td>8/14/08</td>
<td>5ns (5324ps)</td>
<td>0.793</td>
<td>~/.gromacs/080708_gcn4_water</td>
<td>retained helicity well</td>
</tr>
<tr>
<td>2magfinal</td>
<td>2MAG</td>
<td>8/21/08</td>
<td>10 ns</td>
<td>0.704 w/ first 8 aa; 7.36 Exec. RMS</td>
<td>~/.gromacs/082108_2magfinal_water/ions_water</td>
<td>In water; starts to reform helix</td>
</tr>
<tr>
<td>2magunfolded</td>
<td>2MAG</td>
<td>09/01/08</td>
<td>20 ns</td>
<td>5.491</td>
<td></td>
<td>Kinked helix</td>
</tr>
</tbody>
</table>

As mentioned previously peptides were also designed to maintain a helical conformation in organic solvent by the incorporation of the non-natural amino acid AIB. Though we were able to generate starting coordinates in a PDB file format using ChemDraw software there was difficulty interfacing with the GROMACS library for the AIB residues. The project was taken over by collaborators from UMass Amherst (Dr. Scott Auerbach) and University of Puerto Rico at Mayaguez, Mayaguez, Puerto Rico (Dr. Gustavo Lopez)(14).

GROMACS proved to be a fairly reliable tool in terms of our controls monitoring helical structure in water. Though we never experimentally tested peptide structure in...
DMSO of the peptides we ran in simulation the prediction that most native peptide sequences will unfold in organic solvent is as we would expect. If *in silico* screening had shown a peptide which stayed in a helical conformation in DMSO or another solvent I believe it would have been a valid potential structure as it would have been an anomaly. In terms of the effectiveness of GROMACS as an effective screen I would agree that with a basic understanding of Unix and Linux commands GROMCAS is a relatively straightforward program, making it accessible to a broader audience of users than many computational programs. For effective screening of large libraries of peptides use on a computational cluster is recommended as longer simulations take on the order of days to complete using a single node.

A. 3. Protocol for running a GROMACS simulation (adapted from: online MD GROMACS tutorial (15))

Molecular Dynamic (MD) Simulations with GROMACS

Performing MD Simulations:

Make a new directory for the particular run

Copy the MD-Data file folder into the new directory and make a new file using the command ‘*mkdir*’

Open the new directory using the command ‘*CD*’

Then, copy interesting files (including the PDB coordinate file your generated with pymol) into the current directory as following:
You can always get more information on a Gromacs statement by typing it followed by the '-h' flag. For simplicity, the name of the peptide is denoted ‘X’. Take care of extensions.

1. First of all, you need to create some starting files that are required by Gromacs, especially the topology file which gives the topology of your protein, i.e. the way atoms are linked, angles between atoms... For this, type:

```
pdb2gmx -f X.pdb -o X.gro -p X.top -i X.itp -ter -ignh
```

* -f input file
* -o output file that will be used in further statements
* -p topology file that will be created and used later
* -i restraint file that is used when you want to freeze a given moiety in the protein
* -ter to select the type of N and C termini
* -ignh to ignore hydrogens

Gromacs will ask you to choose the force field. Select the "Improved alkane force field".

Moreover, you will have to choose for (non-)charged termini. People who selected a charged peptide have to choose NH3+ and COO-. Others have to choose NH2 and COOH.
It is noticeable that the number of charges of your peptide is null.

2. Then, you need to specify what kind of water box you want to solvate your peptide in. Type (Note that you have to replace DIST by 0.7 if you started from an extended structure and 1.4 otherwise):

    **editconf** -f *X.gro* -o *X_box.gro* -bt cubic -d DIST -c

* -f input file
* -o output file that will be used in further statements
* -i restraint file that is used when you want to freeze a given moiety in the protein
* -bt is the box type
* -d is the distance between the water and the box
* -c centers the peptide in the box

For performing simulations in vacuum, just make a copy of the original Gromacs file by typing:

    **cp** X_box.gro X_solv.gro

Finally, skip the step 3 and go straight to the step 4.

3. After this, we can create water molecules (or other) in the box with the following statement:
genbox -cp X_box.gro -cs SOLVENT -o X_solv.gro -p X.top -seed 1515 >&
genbox.out

* -cp input file

* -cs structure file of the solvent model. Instead of SOLVENT, type:

  o spc216.gro for water (It contains coordinates of water molecules that have been equilibrated at 300K.

  o dmsol.gro for DMSO

  o methanol216.gro for methanol

* -o output file that will be used in further statements

* -p topology file

* -seed table of random number

For people who are performing simulations in methanol or DMSO, you have to modify manually the X.top file. For this, edit the file with an editor (you can use e.g. vi if you like challenges... or more simply kedit):

  kedit X.top

Change the following line:

  #include "spc.itp" should become (regarding the solvent you are using):

  #include "methanol.itp" or

  #include "dmsol.itp"
Also add at the very end of the file a new line containing the number of solvent molecules. You can find it in the output of the genbox statement (it should be written in the genbox.out file. Edit it!) or look at the X_solv.gro file (Take the total amount of residue and remove 4). Hence, the following lines:

```
[ molecules ]
; Compound       #mols
Protein          1
DMSO             497 or
Methanol         497
```

Again, do not put 497 (This value has been put randomly as example). Put the number that is specified in the genbox.out file!

4. It is time now to create what we call an index file, which allows the user to group different sets of atoms and to call them by a single name. That is very useful when you need to work on a part of your system, like only the backbone of the protein for instance. You will understand this concept a little bit later. To do this, type:
make_ndx -f X_solv.gro -o X.ndx

* -f input topology file (Note we are using a Gromacs coordinates file in this case)
* -o Output file

Gromacs shows you all groups defined automatically. Just type q (and 'enter') to accept. You are invited to have a look into the created index file.

5. Before launching a simulation, it is necessary to create a file that will contain all information about the simulation. We are talking in this case of preprocesion. This file can be created by typing:

    grompp -f dyna/em0.mdp -po X_em0.mdp -c X_solv.gro -r X_solv.gro -n X.ndx -p X.top -o X_em0.tpr

* -f input file that contains information about the simulation
* -po output file that contains information about the simulation
* -c conformation file generally used for constraining a protein
* -r conformation file generally used when restraining a protein
* -n index file
* -p topology file
* -o output file that Gromacs will use

6. Now, we can perform the first simulation with:

    mdrun -v -deffnm X_em0
Generally, a huge amount of flags could be used with the mdrun statement. Instead of typing all these flags, it is more convenient to use the -deffnm flag to give a general name. Hence, all output file will contain the name given followed by standard extensions.

The -v flag will give you the step of the run and the estimated remaining time.

7. Actually, the previous simulation was a simple minimization (using a steepest descent algorithm) to "fit the water molecules around the peptide". You have to perform three other minimizations. During them, the protein is frozen and only water molecules are allowed to move. For this, you are going to repeat steps 5 and 6 with different names as following:

```
grompp -f dyna/em1.mdp -po X_em1.mdp -c X_em0.gro -r X_em0.gro -t X_em0.trr -n X.ndx -p X.top -o X_em1.tpr
```

Maybe you noticed the new -t flag this time. This permits to read some information from the trajectory of the previous minimization (Of course, we could not do it in the first minimization knowing that no trajectory existed...)

```
mdrun -v -deffnm X_em1
```

Waiting for the first minimization to finish...

```
grompp -f dyna/em2.mdp -po X_em2.mdp -c X_em1.gro -r X_em1.gro -t X_em1.trr -n X.ndx -p X.top -o X_em2.tpr
```

```
mdrun -v -deffnm X_em2
```
Waiting for the second minimization to finish...

```
grompp -f dyna/em3.mdp -po X_em3.mdp -c X_em2.gro -r X_em2.gro -t X_em2.trr -n Xndx -p X.top -o X_em3.tpr

mdrun -v -deffnm X_em3
```

Waiting for the third minimization to finish...

8. Once all those minimizations have been done, we should not have any "wrong" interactions between water molecules and the peptide (such as bumps) Now, we are going to prepare the system using 5 MD simulations before launching the final one (this last one is the one we are interesting in).

In the same way than before, you are going to type:

```
grompp -f dyna/pr1_SOLV.mdp -po X_pr1.mdp -c X_em3.gro -r X_em3.gro -t X_em3.trr -n Xndx -p X.top -o X_pr1.tpr
```

In the previous statements (and in all next grompp statement, think about replacing SOLV by:

1. vacuum if you are performing simulations in vacuum
2. water if you are performing simulations in water
3. MeOH if you are performing simulations in methanol
4. DMSO if you are performing simulations in DMSO

    mdrun -v -deffnm X_pr1

Waiting for the first MD simulation to finish...

    grompp -f dyna/pr2_SOLV.mdp -po X_pr2.mdp -c X_pr1.gro -r X_pr1.gro -t X_pr1.trr -n X.ndx -p X.top -o X_pr2.tpr

    mdrun -v -deffnm X_pr2

Waiting for the second MD simulation to finish...

    grompp -f dyna/pr3_SOLV.mdp -po X_pr3.mdp -c X_pr2.gro -r X_pr2.gro -t X_pr2.trr -n X.ndx -p X.top -o X_pr3.tpr

    mdrun -v -deffnm X_pr3

Waiting for the third MD simulation to finish...

    grompp -f dyna/pr4_SOLV.mdp -po X_pr4.mdp -c X_pr3.gro -r X_pr3.gro -t X_pr3.trr -n X.ndx -p X.top -o X_pr4.tpr

    mdrun -v -deffnm X_pr4

Waiting for the fourth MD simulation to finish...

    grompp -f dyna/pr5_SOLV.mdp -po X_pr5.mdp -c X_pr4.gro -r X_pr4.gro -t X_pr4.trr -n X.ndx -p X.top -o X_pr5.tpr
mdrun -v -deffnm X_pr5

Waiting for the fifth MD simulation to finish...

9. Finally, you can perform the final MD simulation. To start, we are going to perform a
1ns simulation. You will be able to extend it later, if necessary or some computer time
left. Type:

    grompp -f dyna/md_final_SOLV.mdp -po X_md00.mdp -c X_pr5.gro -r
X_pr5.gro -t X_pr5.trr -n X.ndx -p X.top -o X_md00.tpr

    mdrun -v -deffnm X_md00

Waiting for the final MD simulation to finish...

To run the simulation in the background (so you can quit the terminal on your computer)
change the command to

    Mdrun –deffnm X_md00 &

10. As we said before, the longest the simulation is, the most accurate the estimation of
the free energy will be. You can extend the previous MD simulation by typing:

    tpbconv -f X_md00.trr -s X_md00.tpr -n X.ndx -o X_md01.tpr -extend 1000

    * -f input file that is the trajectory of the previous MD simulation
    * -s topology read from the previous preprocessed MD simulation file
* -n index file
* -o output file that Gromacs will use
* -extend time in ps of the extension (1ns in in this case)

Select the entire system if your are asked for a group.

And you launch the new simulation with:

```bash
mdrun -v -deffnm X_md01
```

Waiting for the MD simulation to finish...

11. Sometimes, your computer can crash while performing a MD simulation. In this case, it is not necessary to start again from the beginning, you can use the already performed simulation restarting it just after the crashing. In the case you are in situation like that, you can just type:

```bash
tpbconv -f X_md01.trr -s X_md01.tpr -n X.ndx -o X_md02.tpr
```

And you will relaunch again the simulation with:

```bash
mdrun -v -deffnm X_md01
```

Waiting for the MD simulation to finish...

Note that in this way you are not starting again the MD simulation rather than continuing it just after the crashing. This is possible because you write velocities in the trajectory files.

Once you decided all simulations are long enough, you will have to analyze results.
References

B1. Expression and Purification of CA-GFP and all GFP based constructs (all CA-FP’s, GFP, sfGFP, Cerulean, Citrine, and mNeptune)

Expression

Day 1:

In the latter part of the day set up a 50 mL overnight culture of CA-GFP in BL21(DE3) to obtain a dense culture

- Add 50 µl of 100 mg/mL ampicillin (or appropriate antibiotic) to 50 mL of LB (note* A 5 mL culture can also be used for this step if inoculating less than 5 flasks)
- Inoculate culture media with a single colony of CA-GFP in BL21(DE3)
- Shake at ~250 rpm at 37°C

Day 2:

Growth for induction

- Add 1mL of 100 mg/mL ampicillin to 1L of 2xYT media
- Inoculate culture with 1 mL of dense overnight culture
- Shake at ~250 rpm in 37°C

Induction conditions

- Monitor OD$_{600}$ until it reaches 0.6-0.7 (approx. 3 hrs)
- At OD 0.6-0.7 induce culture by adding 1mM IPTG (1 mL of 1M IPTG for 1L culture)
- Drop shaker temperature to 25°C
- Allow culture to shake at 25°C for 3 hours
- After 3 hrs harvest cells by centrifugation
- If using a fluorescent GFP construct at this point you would be able to see GFP fluorescence in the cell pellet by eye or when looked at under UV light

**Purification (Using a COBALT-NTA column)**

- Thaw and Resuspend cells in lysis buffer, ensure a homogenous mixture
- Break open cells (via microfluidizer), be sure to add cells into microfluidizer through a 25 or 50 mL pipet so as not to clog the instrument
- Spin lysed cells 15,000 rpm for 45 minutes (resulting supernatant should be very green for GFP)
- Filter supernatant using 0.45 µM Millipore filter in filter assembly
- Load onto Co-NTA column at 3 mL/min, run lysis buffer until baseline collecting flow-through in a 250 ml bottle
- Wash until baseline, collect wash fraction as one aliquot in a 50 mL tube
- Elute until baseline, collect elution fractions in 5-7 mL aliquots in 15 mL tubes
- 1M Imidazole Wash, collect in one 50 mL tube. This step is important for the life of your column as some fraction of GFP and CA-GFP proteins tend to adhere strongly to the column.
- After purification of a fluorescent GFP construct it is necessary to wash the column with 10 mM NaOH during recharging immediately after purification (Order of recharging: water, EDTA, water, NaOH, water, Co- solution, water). Be sure that there is no green color left on the column after the NaOH wash, if so it may also be necessary to wash in a separate step with GdnHCl.

Note: GFP and cleaved CA-GFP elutes in the Wash fractions

Table B1. Buffer compositions for Co-NTA purification of GFP proteins

<table>
<thead>
<tr>
<th>Lysis</th>
<th>Wash</th>
<th>Elution</th>
</tr>
</thead>
<tbody>
<tr>
<td>50 mM NaH₂PO₄ pH 8.0</td>
<td>50 mM NaH₂PO₄ pH 8.0</td>
<td>50 mM NaH₂PO₄ pH 8.0</td>
</tr>
<tr>
<td>300 mM NaCl</td>
<td>300 mM NaCl</td>
<td>300 mM NaCl</td>
</tr>
<tr>
<td>2 mM Imidazole</td>
<td>50 mM Imidazole</td>
<td>250 mM Imidazole</td>
</tr>
</tbody>
</table>

B2. Fluorescence Lysate Assay

Day 1: Co-transform reporter (ie. CA-GFP) and protease (ie. Caspase-7) in electrocompetent BI21(DE3) cells and plate on double antibiotic selection plate

Day 2: Inoculate 5 mL LB cultures with 5 μL 100 mg/ml Ampicillin and 40 mg/ml Kanamycin and a single colony from co-transformants. Incubate overnight at 37°C shaking at ~250 rpm. Prepare appropriate number of 50 ml LB media in baffled 250 ml flasks. Each co-transformation requires two flasks for replicates. Therefore to test one protease requires four flasks (two for the active form of the protease and two for the inactive form or for a co-transformation with an empty vector if no inactive version of the protease is available)
Day 3: For each co-transformation inoculate two 50 ml LB cultures with 50 µL of 100 mg/ml Ampicillin, 50 µL of 40 mg/ml Kanamycin, and 50 µL of the dense 5 mL overnight culture. Incubate at 37°C for approximately 3.5 hrs until cultures reach an OD$_{600}$ of approximately 0.6. Cultures can then be induced with 1ml of 1M IPTG and the temperature dropped to 25°C. (*Note: be sure to check OD of each flask as different co-transformations tend to grow at different rates, it may be necessary to use more than one shaker and move flasks into different shaker as they reach the appropriate OD) The induction time is 18hrs (overnight).

Day 4: For each flask (two for each sample) collect two 2 mL aliquots for lysate and one 1 mL aliquot for whole-cell fluorescence measurements. Spindown all aliquots in a tabletop centrifuge at max speed for two minutes and decant supernatant. Resuspend the lysate fractions in 800 µL of lysis buffer (0.5 mg/mL lysozyme, 2 units DNAase) and lyse using four cycles of freeze thaw alternating between a dry ice bath and the 37° waterbath. After lysing spin samples in tabletop centrifuge for 2 min at max speed. Resuspend 1mL aliquots in 1mL of PBS for whole cell measurements. A sample from each flask should also be taken to measure OD of the culture for normalization. As these samples are very dense I measure 500 µL of culture diluted with 500 µL LB compared to a 1mL LB blank (*Note when measuring bacterial cultures on the Nanodrop always use a cuvette, make sure that the cuvette function is checked in the software).

In a black 96-well co-star plate add 100 µL of supernatant from each lysed sample (each sample should have a 1A, 1B, 2A, and 2B sample the number refers to the flask while the letter refers to the duplicate aliquot taken from each flask). Add 100 µL of each
whole cell sample as well. Make sure to have 100 µL of lysis buffer and 100 µL of PBS as blanks. Measure the endpoint fluorescence on the Spectramax M5 plate reader at the wavelengths indicated in the table below for each fluorescent protein. Citrine fluorescent protein should be read on a fluorimeter due to the proximity of the excitation and emission wavelengths in order to avoid bleed through of signal.

Table B2. Fluorescence Parameters for FPs

<table>
<thead>
<tr>
<th>Fluorescent protein</th>
<th>Excitation (nm)</th>
<th>Emission (nm)</th>
<th>Instrument</th>
</tr>
</thead>
<tbody>
<tr>
<td>GFP</td>
<td>475</td>
<td>512</td>
<td>Spectramax M5 platereader</td>
</tr>
<tr>
<td>Cerulean</td>
<td>433</td>
<td>475</td>
<td>Spectramax M5 platereader</td>
</tr>
<tr>
<td>Citrine</td>
<td>513</td>
<td>529</td>
<td>JASCO FP-6500 spectrofluorimeter (Thai lab fluorometer)*</td>
</tr>
<tr>
<td>mNeptune</td>
<td>530</td>
<td>650</td>
<td>Spectramax M5 platereader</td>
</tr>
</tbody>
</table>

Raw fluorescence is normalized based on the OD of each culture using the equation below.

\[
(Raw \text{ Fluorescence} - Buffer \text{ Fluorescence}) \times \left(\frac{15}{OD}\right) = \text{Normalized Fluorescence}
\]

*Western Analysis of Lysate samples*

From each lysate sample collect a 60 µL aliquot, add 15 µL of SDS loading dye with DTT and boil samples for 10 minutes. For optimal blotting normalize the amount of each sample loaded using the equation below. Run three identical SDS-PAGE gels, two using the kaleidoscope marker and one using the NEB protein ladder. Transfer the two gels with kaleidoscope marker onto Hybond-ECL™ nitrocellulose membrane (GE Healthcare) in a two hour transfer at 100V. After transfer is complete follow standard Hardy lab western protocol. Blocking step is often performed overnight. For one
western blot with the Millipore anti-GFP primary antibody diluted 1:1000 in blocking solution. For the other blot with the protease specific antibody, in the case of caspase-7 the primary antibody for the C7 large subunit was used.

B3. In Vitro Cleavage Assay

Buffer exchange CA-GFP into a buffer of 135 mM NaCl, 4.5 mM NaH$_2$PO$_4$ pH 7, 8 mM KH$_2$PO$_4$ pH 7, 3 mM KCl to approximately 10 µM. For caspase-6 cleavage the most effective buffer was the caspase-6 activity assay buffer (100 mM HEPES pH 7.5, 10% sucrose, 0.1% CHAPS, 30 mM NaCl, and 5 mM DTT), optimization may be required for other protease/reporter combinations. Make sure to prepare enough volume to read 100 µL and for 15 µL gel sample aliquots at each of the timepoints with at least 20 µL extra to compensate for any pipetting errors and evaporation of the sample. Take 15 µL gel samples before the addition of caspase for the 0 min time point. At initiation of the reaction pipet 100 µL aliquots of the digest and controls into a costar 96-well black plate, and the fluorescence was measured in a kinetic assay format every 5 min (Ex. 475 nm/Em. 512 nm) for 15 h at 37 °C. Make sure to set up the method on the software before the addition of caspase in order to start reading as immediately after initiation as possible. The remainder of the sample was incubated at 37 °C in 15 µL aliquots to which SDS loading buffer was added at time points of 0, 5, 10, 15, 20, 30, 60, 90, 120 min, and at 15 h. These samples were then run on an SDS-PAGE gel to determine the ratio of cleaved to uncleaved CA-GFP.
B4. Stepwise cloning method (Overlapping primers method)

This method is useful when rearranging or piecing together a gene from already existing sequences. I will use moving the M2 peptide from the C-terminus of GFP to the N-terminus as an example. You will need to design four primers. The first two will amplify the GFP portion and the second two to amplify the M2 portion. P1 is designed to anneal to the N-terminal region of GFP with the linker region encoded in the primer (Fig. 1). P2 is designed to anneal to the C-terminus of GFP and includes a stop codon and the C-terminal restriction site for the new gene. P3 includes the N-terminal restriction site for the new gene and the N-terminal region of the M2 portion of the gene. P4 is designed to anneal to the C-terminus of M2 and includes the same linker sequence as P1, meaning P1 and P4 will generate an overlapping region. This overlapping region should be approximately 24 bp and have a $T_m$ greater than 60°C.

![Figure B1. Primer design for overlapping molecular cloning method](image)

Amplify each section in independent PCR reactions, run each reaction on an agarose gel, and gel purify the inserts using the Promega gel purification kit. After gel purification of the amplified fragments combine the two fragments in a single PCR reaction (see below for cycle parameters) and allowed to anneal through the overlapping region for five PCR
cycles before the addition of primers P2 and P3 which then amplified the full length gene. The full length insert and vector are then digested with the appropriate restriction enzymes, gel purified, and ligated into the desired vector (I use the Roche Rapid DNA Dephos & Ligation Kit, you can also use the NEB T7 ligase with this protocol so the kit is not necessary).

Table B3. PCR Reaction for Overlapping Primers amplification of full-length gene

<table>
<thead>
<tr>
<th>Component</th>
<th>Volume (µL)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fragment 1 (longer fragment)</td>
<td>Depends on Concentration ~160 ng</td>
</tr>
<tr>
<td>Fragment 2 (shorter fragment)</td>
<td>Depends on Concentration ~30 ng</td>
</tr>
<tr>
<td>Water</td>
<td>Up to final volume of 50 µL</td>
</tr>
<tr>
<td>10x Pfu Turbo buffer</td>
<td>5</td>
</tr>
<tr>
<td>10 mM dNTP’s</td>
<td>1</td>
</tr>
<tr>
<td>Pfu Turbo</td>
<td>1</td>
</tr>
<tr>
<td>P3 (added after 5 cycles)</td>
<td>1</td>
</tr>
<tr>
<td>P2 (added after 5 cycles)</td>
<td>1</td>
</tr>
</tbody>
</table>

Table B4. PCR cycle for amplification for full-length gene, the method should be paused after 5 cycles and primers 2 and 3 added to the reaction.

<table>
<thead>
<tr>
<th>Temperature</th>
<th>Time</th>
<th>Repeats</th>
</tr>
</thead>
<tbody>
<tr>
<td>95°C</td>
<td>1 min</td>
<td>1</td>
</tr>
<tr>
<td>95°C</td>
<td>30 sec</td>
<td>40</td>
</tr>
<tr>
<td>63°C</td>
<td>30 sec</td>
<td></td>
</tr>
<tr>
<td>72°C</td>
<td>1 min</td>
<td></td>
</tr>
<tr>
<td>72°C</td>
<td>10 min</td>
<td>1</td>
</tr>
<tr>
<td>4°C</td>
<td>∞</td>
<td></td>
</tr>
</tbody>
</table>

B5. NMR sample preparation and Analysis

This protocol was adapted from the Gierash lab protocol (1)

Day 1: Inoculate 5 ml overnight culture. Make minimal media and autoclave media and flasks

M9 minimal medium (for 1 L): (autoclave or filter sterilize and store at 4°C)
- 6.5 g Na$_2$HPO$_4$
- 3.0 g KH$_2$PO$_4$
- 0.5 g NaCl
- 1g NH$_4$Cl (for unlabeled medium, or 1.0 g $^{15}$NH$_4$Cl for $^{15}$N-labeled medium)
- 120 mg MgSO$_4$
- 11 mg CaCl$_2$

(filter sterilize)

4 g regular D-glucose (or 2 g D-$^{13}$C-glucose for $U^{13}$C samples only)

Vitamins (biotin and thiamine at a final concentration of 10 mg/mL for both)

(I make 10 X stocks of both glucose and vitamins)

For each L of media being grown autoclave:
- 20 mL LB in 250 mL Erlenmyer
- 100 mL M9 in 250 mL Erlenmyer (first resuspension)
- 100 mL M9 in 500 mL Erlenmyer (Dilution to 200 mL)
- 800 mL M9 in 1 L Fernbach (Final dilution to 1L)

Day 2: Add 10mL 10x vitamins and 10 ml 10x Glucose to each Erlenmyer as well as 10x Vitamins and 10X glucose and appropriate antibiotics to cooled media before inoculation.

- Start with a 5-ml culture of LB medium from a freshly colony (o/n @ 37°C)
- Add 50 uL of o/n culture in 20 ml LB medium; incubate the culture until the OD$_{600}$ is 0.5 @ 37°C.
- Spin the cells down and wash once with PBS, resuspend in 100 ml M9 $^{15}$N-labeled medium (prepared with $^{15}$NH$_4$Cl) (Erlenmyer 1, pre-warmed to 37°C). Starting OD should be about 0.1, incubate the culture until the OD$_{600}$ is 0.5-0.6.

- Dilute the culture to 200 ml M9 $^{15}$N-labeled medium (Erlenmyer 2, pre-warmed to 37°C) and grow until the OD$_{600}$ is 0.5-0.6.

- Dilute the culture to 1L M9 $^{15}$N-labeled medium (Fernbach, pre-warmed to 37°C) and grow until the OD$_{600}$ is 0.5-0.6.

- Induce protein overexpression.

- Continue the growth for induction time of protein.

Purify protein by normal method. The final protein concentration should be ~0.1 mM in the desired buffer with 5% D$_2$O (should be exactly the same as a reported spectra if comparing the coordinates). The minimum volume for the Shigemi BMS 005-V NMR tube recommended for the cryogenically cooled triple-resonance probe is 0.28 mL. Previous spectra were measured on the 700-MHz Varian NMR system at 37°C. Spectra were processed using NMRpipe (2) and analyzed using Cara (3).

**B6. Proteinase K Proteolysis Assay**

Buffer exchange protein into a buffer of 50 mM Tris pH 7.5, 5 mM CaCl$_2$ or if the protein is sufficiently concentrated dilution into the above buffer is acceptable. The final concentration of the protein should be between 10-30 µM. Prepare tubes for five aliquots with 1µL of X mM protease inhibitor phenylmethylsulfonyl fluoride (PMSF) in each tube.
Take 10 µL aliquot for 0 min timepoint and add to PMSF. Add 500 nM proteinase K to the protein and begin timing. Take 10 µL aliquots at each of the timepoints, adding the aliquot to the PMSF. After all timepoints have been collected add 3 µL SDS loading buffer with DTT to each sample and boil for 10 minutes. The samples may not load into SDS-PAGE gel easily, make sure that running buffer and gels have been incubated at 4°C prior to running the gel. Also ensure that the samples are completely cooled and spin down in tabletop centrifuge before loading.

B7. Two Dimensional Gel Analysis

Before performing experiment check the pI of the protein of interest and order the appropriate IPG strip with a pH range complementing the proteins you are trying to separate. Strips are ordered from NEB. The 7 cm size should be ordered to be compatible with the Normanlly lab apparatus. The manual that comes with the IPG strips contains photo instructions of many of the steps described here and is very helpful.

Day 1. Prepare sample, in my case it was an O/N digest of DVP2 with the DVPA-GFP reporters, by diluting in a rehydration buffer of 8M Urea, 2% CHAPS, and 50 mM DTT. I performed my digests at a high concentration (5 mg/mL of each protein) and diluted my reactions 10-fold in the rehydration buffer. Once the sample is prepared using a pipet dispense a continuous line of your protein solution (125-130 µL) in one well of the IPG strip tray. Carefully place the IPG strip face down in the protein solution making sure there are no gaps or bubbles under the strip. Repeat for each sample. Incubate at room temperature for 1 hr. After 1 hr most of the proteins solution should have be absorbed by
the strip. Cover the strips with mineral oil (~1mL) and incubate at room temperature for ~12 hours (approx. overnight).

Day 2. Gently blot off mineral oil with damp filter paper. Place wicks (small pieces of filter paper) over the electrodes of the isoelectric focusing apparatus and wet with ~6 µL of H₂O. Carefully load the strips into the apparatus making sure the gel side is facing down and that it is on top of the wicks on each side. Cover the samples with ~1 mL of mineral oil. Load the tray onto the instrument. There are several options for the voltage ramping mode. I selected the rapid ramping mode as I was only trying to separate two purified proteins. Be sure to start this early in the morning as the ‘rapid’ ramping mode took nearly 10 hours to complete. Once the ramping method is complete the samples can be placed in the IPG strip tray and be stored at -80°C until needed. While the isoelectric focusing step is running pour appropriate percent acrylamide gels. Only the resolving portion is necessary and should poured within ~1cm of the top of the short glass plate. Be sure to use a tall plate with 1 mm thickness to ensure that the IPG strip will fit.

Day 3. Thoroughly thaw the samples (~1 hr at room temp). Incubate strips in Equilibration buffer I (6M urea, 2% SDS, 0.05M Tris HCl pH 8.8, 20% glycerol, 2% DTT), a volume of 2.5 ml is needed for a 7 cm strip. After a 10 min incubation decant equilibration buffer 1 and replace with 2.5 ml of equilibration buffer 2 (6M urea, 2% SDS, 0.05M Tris HCl pH 8.8, 20% glycerol, 2.5% Iodoacetamide), incubating for another 10 min then decant. Carefully slide the IPG gel into the SDS gel with the plastic backing touching the tall plate. Carefully use forceps to make sure there are no air
bubbles under the IPG gel. At this point you can add a piece of filter paper on which the protein ladder has been loaded or place a small plastic divider on one side of the strip to make a well. Once all the strips are properly placed pour a solution of 0.5% agarose in 1X SDS-PAGE running buffer with 0.003% bromophenol blue as an indicator for the gel front and allow to solidify (~10 min). If you made a well for the protein ladder then add marker sample and run SDS-PAGE as normal.

References
# APPENDIX C
## PROTEASE ACTIVATABLE- FLUORESCENT PROTEINS

Table C.1: Summary of constructs and fluorescent properties tested of existing protease activatable- fluorescent proteins.

<table>
<thead>
<tr>
<th>Protein</th>
<th>Dark state in bacteria</th>
<th>Fluorescence Increase in bacteria</th>
<th>Cleavage in bacteria</th>
<th>in vitro cleavage</th>
<th>in vitro fluorescence</th>
<th>Dark state in Mammalian cells</th>
<th>Vectors</th>
</tr>
</thead>
<tbody>
<tr>
<td>CA-GFP</td>
<td>Yes</td>
<td>45x</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>DEVDFQ</td>
</tr>
<tr>
<td>NtermM2 CA-GFP (nCA-GFP)</td>
<td>Yes</td>
<td>6x</td>
<td>assumed</td>
<td>NT</td>
<td>NT</td>
<td>NT</td>
<td>DEVDFQ</td>
</tr>
<tr>
<td>F64L CA-GFP (CA-EGFP)</td>
<td>Yes*</td>
<td>4x</td>
<td>assumed</td>
<td>NT</td>
<td>NT</td>
<td>NT</td>
<td>DEVDFQ</td>
</tr>
<tr>
<td>CA-mKate2</td>
<td>NT</td>
<td>NT</td>
<td>NT</td>
<td>NT</td>
<td>No</td>
<td>No</td>
<td>DEVDFQ</td>
</tr>
<tr>
<td>cCA-mNeptune (C-terminal M2)</td>
<td>Yes</td>
<td>1.07x</td>
<td>Yes</td>
<td>NT</td>
<td>NT</td>
<td>No</td>
<td>DEVDFQ</td>
</tr>
<tr>
<td>cCA-mNeptune (C158A)</td>
<td>Yes</td>
<td>No (LD)</td>
<td>Yes</td>
<td>NT</td>
<td>NT</td>
<td>No</td>
<td>DEVDFQ</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Linker Sequence</th>
<th>DNA Archive Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>pET21b</td>
<td>B1-31</td>
</tr>
<tr>
<td>pET21b</td>
<td>B3-8</td>
</tr>
<tr>
<td>pET29b</td>
<td>B3-9</td>
</tr>
<tr>
<td>pET21b</td>
<td>B3-6</td>
</tr>
<tr>
<td>pET21b</td>
<td>B3-7</td>
</tr>
<tr>
<td>pET21b</td>
<td>B2-45</td>
</tr>
<tr>
<td>pAlter-EX2</td>
<td>B2-48</td>
</tr>
<tr>
<td>pmKate 2C</td>
<td>B3-16</td>
</tr>
<tr>
<td>pmKate 2C</td>
<td>B3-17</td>
</tr>
</tbody>
</table>

*F64L CA-GFP (CA-EGFP) is assumed to be in vitro cleavable based on its predecessor F64L CA-GFP (CA-GFP)
<table>
<thead>
<tr>
<th>NtermM2 CA-mNeptune (CA-mNeptune)</th>
<th>Yes</th>
<th>32x</th>
<th>Yes</th>
<th>NT</th>
<th>NT</th>
<th>Yes</th>
<th>DEVDFQ</th>
<th>pET21b</th>
<th>B3-18</th>
</tr>
</thead>
<tbody>
<tr>
<td>NtermM2 CA-mNeptune (C158A)</td>
<td>NT</td>
<td>NT</td>
<td>yes</td>
<td>NT</td>
<td>NT</td>
<td>Yes</td>
<td>DEVDFQ</td>
<td>pmKate 2C</td>
<td>B3-19</td>
</tr>
<tr>
<td>CA-IRFP</td>
<td>NT</td>
<td>NT</td>
<td>NT</td>
<td>NT</td>
<td>NT</td>
<td>No</td>
<td>DEVDFQ</td>
<td>pmKate2C</td>
<td>B2-52</td>
</tr>
<tr>
<td>MMPA-GFP</td>
<td>Yes</td>
<td>No (LD)</td>
<td>Yes</td>
<td>(&lt;100%)</td>
<td>No (LD)</td>
<td>NT</td>
<td>PLGLEEAP</td>
<td>pACYC Duet</td>
<td>B3-32</td>
</tr>
<tr>
<td>CA-sfGFP</td>
<td>No</td>
<td>NA</td>
<td>NT</td>
<td>Yes (50%)</td>
<td>NA</td>
<td>NT</td>
<td>DELD</td>
<td>pET16b</td>
<td>B1-26/B2-71</td>
</tr>
<tr>
<td>HRA-GFP</td>
<td>Yes</td>
<td>1.9x</td>
<td>assumed</td>
<td>NT</td>
<td>NT</td>
<td>No</td>
<td>LEVLFQ</td>
<td>pAlter-EX2</td>
<td>B2-60</td>
</tr>
<tr>
<td>HRA-GFP ΔYK</td>
<td>Yes</td>
<td>10.3x</td>
<td>assumed</td>
<td>NT</td>
<td>NT</td>
<td>NT</td>
<td>LEVLFQ</td>
<td>pACYC Duet</td>
<td>B2-65</td>
</tr>
<tr>
<td>CA-CFP</td>
<td>Yes</td>
<td>NT</td>
<td>NT</td>
<td>Yes</td>
<td>No (CRC)</td>
<td>NT</td>
<td>DEVDFQ</td>
<td>pET21b</td>
<td>B3-38/B4-6</td>
</tr>
<tr>
<td>CA-YFP (Citrine)</td>
<td>Yes</td>
<td>7x</td>
<td>Yes</td>
<td>NT</td>
<td>Yes</td>
<td>NT</td>
<td>DEVDFQ</td>
<td>pET21b</td>
<td>B3-40</td>
</tr>
<tr>
<td>CA-CerFP</td>
<td>Yes</td>
<td>24x</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>NT</td>
<td>DEVDFQ</td>
<td>pET21b</td>
<td>B3-50</td>
</tr>
<tr>
<td>HIVA-GFP</td>
<td>Yes</td>
<td>NA</td>
<td>ND</td>
<td>No</td>
<td>NA</td>
<td>NT</td>
<td>TVSFNFPQITLW</td>
<td>pET21b</td>
<td>B3-74</td>
</tr>
<tr>
<td>DVPA-GFP</td>
<td>Yes</td>
<td>NA</td>
<td>ND</td>
<td>No</td>
<td>NA</td>
<td>NT</td>
<td>GRRDFQGP</td>
<td>pET21b</td>
<td>B3-79</td>
</tr>
<tr>
<td>DVPA-GFP V1</td>
<td>Yes</td>
<td>NA</td>
<td>ND</td>
<td>No</td>
<td>NA</td>
<td>NT</td>
<td>GRRGFQGP</td>
<td>pET21b</td>
<td>B3-80</td>
</tr>
<tr>
<td>DVPA-GFP V2</td>
<td>Yes</td>
<td>NA</td>
<td>ND</td>
<td>No</td>
<td>NA</td>
<td>NT</td>
<td>DEGRRGGP</td>
<td>pET21b</td>
<td>B3-81</td>
</tr>
</tbody>
</table>

- background 12x higher than CA-GFP

Abbreviations: NT= not tested; NA= not applicable; ND= none detected

Initials: LD: Lindsay Dawson; JC: Jun Chu; CRC Charnell Chasten
BIBLIOGRAPHY

Chapter 1


Chapter 2


**Chapter 3**

44. Keller, R. (2004) Optimizing the process of nuclear magnetic resonance spectrum analysis and computer aided resonance assignment. ETH, Zurich

Chapter 4

Chapter 5


**Appendix A**


**Appendix B**

