Functional Divergence of Photolyase and Cryptochrome-DASH: The Role of Loop Dynamics

by

Matthew A. Brolich
B.Sc., Simon Fraser University, 2008

Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy

in the Department of Molecular Biology and Biochemistry Faculty of Science

© Matthew A. Brolich 2015
SIMON FRASER UNIVERSITY
Fall 2015
Approval

Name: Matthew A. Brolich
Degree: Doctor of Philosophy
Title: Functional Divergence of Photolyase and Cryptochrome-DASH: The Role of Loop Dynamics

Examin ing Committee: Chair: Dr. Jonathan Choy
Associate Professor

Dr. Edgar Young
Senior Supervisor
Associate Professor

Dr. Mark Paetzel
Supervisor
Professor

Dr. Dipankar Sen
Supervisor
Professor

Dr. Claire Cupples
Internal Examiner
Professor

Dr. Chow Lee
External Examiner
Associate Professor
Biochemistry and Molecular Biology
University of Northern British Columbia

Date Defended/Approved: September 11, 2015
Abstract

Photolyase (PL) is an enzyme that repairs thymine dimers, a form of DNA damage caused by UV light. To do so, the damaged bases of the cyclobutane pyrimidine dimer (CPD) are displaced from the duplex into an extrahelical position. A “recognition loop” in PL must be displaced to allow it to bind substrate. Cryptochrome-DASH (CRY$_D$) is a structurally homologous protein to PL, with a high sequence identity, which cannot perform the same physiological function as PL. Here I ask how PL is able to bind to its substrate and how CRY$_D$ functionally diverged from PL. I hypothesize that there exist optimized conformational dynamics of these recognition loop regions that allow PL to bind CPD. Limited proteolysis experiments determined that the recognition loops of PL and CRY$_D$ are the most dynamic regions of each of these proteins. Furthermore, the conformational dynamics of the CRY$_D$ recognition loop are greater than that of PL. A difference between the dynamics is consistent with my hypothesis, but does not definitively prove it. The differences in recognition loop conformational dynamics could be due to primary, secondary or tertiary structures. To determine the impact sequence has on these conformational dynamics, end-to-end contact was measured in isolated recognition loop peptides using fluorescence quenching. It was found that the CRY$_D$ recognition loop peptide is more dynamic than that of PL. This indicates that sequence, at least in part, is responsible for conformational dynamic differences in these protein loops. To definitively determine the role the recognition loop plays in how PL and CRY$_D$ functionally diverged, I outlined a series of directed evolution experiments. To this end random mutagenesis was planned to be carried out only in the PL recognition loop region in search for partially functional PLs. If a correlation exists between conformational dynamics and function, then the recognition loop would be implicated in the functional divergence of these proteins. Technical development, including random mutagenic PCR, a mutation rate detection assay, in vivo functional assays and in vitro functional assay was undertaken to test this hypothesis.

Keywords: Photolyase; Cryptochrome-Dash; cyclobutane pyrimidine dimer; conformational dynamics; directed evolution; DNA repair
Dedication

This thesis is dedicated to everyone who has supported and helped me through my years at SFU. To Dr. Melanie O’Neill, thank you for everything; you are deeply missed.
Acknowledgements

My time at SFU as a graduate student has been very unusual, though I have been fortunate to have met many people who have helped me along the way. I would like to thank Dr. Melanie O’Neill for getting me started in science and helping me grow as a scientist and as a person. I would also like to thank Dr. Edgar Young for taking me as a student when I was in need of help and for keeping me on the right track when things seemed difficult.

My lab mates greatly enhanced my stay at SFU. My research was greatly helped by Neahlanna Mcleod and Dr. Michael Damiani, who both helped train me on the basics of protein work. I would also like to thank Michael Dahabieh, Rob Trasolini and Stephanie Luongo who helped me with various aspects of my project and numerous helpful conversations. I also want to thank Jennifer Kass, Dana Page, Kaylee Magee, and Collin Massey for help directly or indirectly with my project.

I would like to thank Dr. Dipankar Sen for his work on my committee and his insight into how I might grow as a scientist. I was also fortunate to have Dr. Mark Paetzel on my committee as well who provided insight into the direction of my project.

I would also like to thank Dr. Chow Lee and Dr. Claire Cupples for their time and expertise.

I was fortunate to be involved with two departments here at SFU, who both have helped me greatly. I would like to thank the Department of Chemistry and the Department of Molecular Biology and Biochemistry. Both faculty and staff have been incredibly supportive.
Table of Contents

Approval .......................................................................................................................... ii
Abstract .......................................................................................................................... iii
Dedication ...................................................................................................................... iv
Acknowledgements ......................................................................................................... v
Table of Contents ........................................................................................................... vi
List of Tables .................................................................................................................. ix
List of Figures ................................................................................................................ x
List of Acronyms ............................................................................................................ xii

Chapter 1. Introduction ............................................................................................. 1
  1.1. An introduction to photolyase ........................................................................... 1
    1.1.1. Photolyase repairs cyclobutane pyrimidine dimers .................................... 1
    1.1.2. Thymine dimers are an important product of UV damage .................... 2
    1.1.3. PL has two domains conserved in a Cryptochrome/Photolyase superfamily ............................................................. 3
    1.1.4. PL has importance in human health as well as other organisms .......... 6
    1.1.5. Detection of CPD ....................................................................................... 8
  1.2. PL repairs CPDs efficiently using light energy .................................................... 9
    1.2.1. PL uses UV light to overcome the activation energies of CPD repair and results in a high efficiency of energy transfer ................................................. 9
    1.2.2. Photolyase function relies on a FAD cofactor ....................................... 10
    1.2.3. FAD is a common biochemical cofactor ............................................. 11
    1.2.4. The active form of the FAD is fully reduced in PL ............................... 13
    1.2.5. The catalytic cycle of PL occurs stepwise in ~1 ns ............................. 15
  1.3. “Base-flipping” is required by PL and other enzymes ....................................... 18
    1.3.1. The PL binds to only a 4-6 bp region ................................................... 18
    1.3.2. The CPD is bound to PL in an extrahelical state and both the CPD and its complementary strand form interactions with PL ............................... 19
    1.3.3. The energy cost of displacing a CPD from the duplex is half that of a thymine pair ................................................................. 24
    1.3.4. The mechanism by which PL binds to and accesses the CPD is not well understood ................................................................. 25
    1.3.5. PL CPD binding involves conformational changes in a key recognition loop ................................................................. 27
  1.4. CRYD shares important structural and functional features with PL ............... 30
    1.4.1. Cryptochromes have a wide array of functions that are poorly understood ................................................................. 30
  1.5. The recognition loop is a promising region for studying PL-dsDNA binding ...... 37
  1.6. Multiple assays can be used to probe conformational dynamics ..................... 39
    1.6.1. Detecting conformationally dynamic regions using limited proteolysis and mass spectrometry ................................................................. 39
    1.6.2. End-to-end contact of peptides is a measure of conformational dynamics ................................................................. 41
  1.7. Multiple functional assays can be used to detect repair capability .......... 45
1.7.1. Functional assays chosen for this thesis investigate PL repair capability in different ways

Chapter 2. The recognition loop of CRY \textsubscript{D} is more flexible than that of PL

2.1. PL and CRY \textsubscript{D} have distinct recognition loops which may be conformationally dynamic

2.2. Limited proteolysis identifies the recognition loop as the most conformationally dynamic region of PL and CRY \textsubscript{D}

2.2.1. ProteinaseK and Chymotrypsin produce multiple PL and CRY \textsubscript{D} fragments

2.2.2. Preferential cleavage products were observed in PL and CRY \textsubscript{D} after limited proteolysis with Chy or PK

2.2.3. PL and CRY \textsubscript{D} have homologous proteolytic sites

2.3. The recognition loop of PL is less conformationally dynamic than that of CRY \textsubscript{D}

2.3.1. I would expect the CRY \textsubscript{D} recognition loop to be cleaved more readily than that of PL

2.4. Global stability differences between PL and CRY \textsubscript{D} could not be assessed due to irreversible folding

2.5. Differences in the recognition loop dynamics of PL and CRY \textsubscript{D} are, at least in part, due to sequence

2.5.1. Understanding the underlying mechanism for these differences in the conformational dynamics can help our understanding of the functional divergence of PL and CRY \textsubscript{D}

2.5.2. The fluorescence decay of folding CRY \textsubscript{D} and PL recognition loop peptides are not mono exponential

2.5.3. Under native conditions, end-to-end contacts occur faster in the CRY \textsubscript{D} loop peptide than that of PL

2.5.4. Intermolecular quenching is unlikely

2.5.5. It is likely there are secondary structures forming in the peptide

2.5.6. Double exponential lifetimes changes to mono exponential under denaturing conditions

2.5.7. Summary: CRY \textsubscript{D} loop peptide is more dynamic than that of PL

2.5.8. The energy of folding is mainly influenced by enthalpy

2.5.9. Conclusion: increased folding dynamics could impact binding due to increase in conformational space

Chapter 3. Technical Development for Directed Evolution of a Loop Mutant Library

3.1. A plan to test my hypothesis

3.2. Technical development summary

3.3. Mutagenic PCR methods were validated

3.3.1. A cassette was designed and created to hold the PL loop coding region for creating random mutations

3.3.2. Degenerate PCR was validated

3.3.3. Error-prone PCR was validated
3.3.4. A procedure for generating a random mutant library was developed using epPCR .............................................................. 88
3.4. PL in vivo functional assays was developed using repair deficient cells .......................................................... 89
  3.4.1. PL in vivo functional assay was validated ........................................................................................................ 89
  3.4.2. An improved plasmid system for PL was designed ........................................................................ 91
  3.4.3. A plasmid, pPLL, was designed to contain the mutant and wild type PLs .................................................... 92
  3.4.4. Western blot detection of PL was established .................................................................................. 93
3.5. PL in vitro functional assay was established for monitoring of enzyme activity and comparison between enzymes .......................................................... 95
  3.5.1. Changes in absorbance at 260 nm allows for detection of damaged DNA in ssDNA ........................................ 95
  3.5.2. Repair of ssDNA can be monitored by the change in absorbance at 260 nm .................................................. 96
  3.5.3. Monitoring of the absorbance at 260 nm allows for detection of damaged DNA in dsDNA ..................... 98
  3.5.4. Repair of damaged DNA in dsDNA ........................................................................................................ 99
  3.5.5. PL W392Y repairs comparably to PL ....................................................................................................... 102
  3.5.6. Development of an endonuclease detection method for thymine dimers ........................................ 105
3.6. Future directions and summary .................................................................................................................. 108
  3.6.1. Future Directions ........................................................................................................................................ 108
  3.6.2. Summary .................................................................................................................................................. 109

Chapter 4. Methods .............................................................................................................................. 110
  4.1. Protein purification and UV/Vis/CS spectroscopy .............................................................................................. 110
    4.1.1. PL/PL W392Y purification ....................................................................................................................... 110
    4.1.2. CRY Purification ........................................... 113
    4.1.3. CD-spectroscopy stability assay ........................................................................................................... 116
  4.2. Limited proteolysis method ........................................................................................................................... 116
  4.3. Fluorescence methods ......................................................................................................................................... 117
    4.3.1. Fluorescent peptide generation .................................................................................................................... 117
    4.3.2. Fluorescence TC-SPC method ..................................................................................................................... 117
  4.4. Mutagenic PCR .................................................................................................................................................. 118
    4.4.1. Subcloning pm242-cas ............................................................................................................................. 118
    4.4.2. Degenerate primer PCR ........................................................................................................................... 118
    4.4.3. epPCR ....................................................................................................................................................... 118
  4.5. Tagged PL expression under TAC promoter .................................................................................................. 119
    4.5.1. Subcloning pPLeL – performed by MBB Service Centre (Ziwei Ding) .......................................................... 119
    4.5.2. Western blot .......................................................................................................................................... 119
  4.6. In vivo functional assays .................................................................................................................................... 120
  4.7. In vitro A260 functional assays ........................................................................................................................ 120
  4.8. In vitro endonuclease protection functional assays ........................................................................................... 121

References ......................................................................................................................................................... 122
List of Tables

Table 1.1. Hydrogen bonds interactions of PL recognition loop without substrate ................................................................. 29
Table 1.2. Hydrogen bonds interactions of PL recognition loop with substrate bound ......................................................................................................................... 30
Table 1.3. Hydrogen bonds interactions of CRY$_D$ recognition loop without ssDNA .................................................................................................................................. 35
Table 1.4. Hydrogen bonds interactions of CRY$_D$ recognition loop with ssDNA bound .................................................................................................................................. 36
Table 2.1. Limited proteolysis fragment identification ................................................................................................................................. 54
Table 2.2. Fluorescence lifetimes of DBO peptides ................................................................................................................................. 77
Table 3.1. Thymine dimer synthesis ....................................................................................................................................................... 101
Table 3.2. Thymine dimer repair ....................................................................................................................................................... 101
Figure 2.14. Intermolecular quenching is not present in the peptide folding experiments. ................................................................. 74
Figure 2.15. Fluorescence lifetimes from TC-SPC experiment under denaturing conditions................................................................. 76
Figure 2.16. Eyring plots of PL and CRYD recognition loop peptides................................. 79
Figure 3.1. Flow chart of planned experiments................................................................. 83
Figure 3.2. The plasmid map of pm242-cas ................................................................. 84
Figure 3.3. Degenerate primer PCR amplification of the PL loop region......................... 86
Figure 3.4. Error-prone PCR protection assay................................................................. 88
Figure 3.5. PL in vitro functional assay........................................................................ 91
Figure 3.6. Plasmid map of PLeL ............................................................................... 92
Figure 3.7. Plasmid map of PLL ............................................................................... 93
Figure 3.8. Western blot of an expressed his-tagged PL from whole cell lysate. ...................... 94
Figure 3.9. UV-vis spectra of thymine dimer generation in (dT)18 .................................... 96
Figure 3.10. ssDNA repair by PL .................................................................................. 97
Figure 3.11. Double stranded DNA substrate .............................................................. 98
Figure 3.12. UV-vis spectra of thymine dimer generation in dsDNA .................................. 99
Figure 3.13. Repair of dsDNA substrate ....................................................................... 100
Figure 3.14. CD spectra of PL W392Y ........................................................................ 103
Figure 3.15. ssDNA repair by PL W392Y ...................................................................... 104
Figure 3.16. Repair of dsDNA substrate ....................................................................... 105
Figure 3.18. Autoradiogram of 32P end labeled TT96 PCR product ............................ 107
Figure 4.1 PL purification monitored by electrophoresis.................................................. 111
Figure 4.2. PL W392Y purification monitored by electrophoresis.................................... 112
Figure 4.3. UV-vis spectra after DEAE purification of PL .............................................. 113
Figure 4.4. Purification of CRYD monitored by gel electrophoresis ................................ 114
Figure 4.5. UV-vis spectra after purification of CRYD .................................................. 115
### List of Acronyms

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>(C&lt;&gt;C)</td>
<td>Cytosine dimer</td>
</tr>
<tr>
<td>(T&lt;&gt;T)</td>
<td>Thymine dimer</td>
</tr>
<tr>
<td>(U&lt;&gt;T)</td>
<td>Uracil-thymine dimer</td>
</tr>
<tr>
<td>(U&lt;&gt;U)</td>
<td>Uracil dimer</td>
</tr>
<tr>
<td>2-Ap</td>
<td>2-aminopurine</td>
</tr>
<tr>
<td>8-HDF</td>
<td>8-hydroxy-7-deazaflavin</td>
</tr>
<tr>
<td>Å</td>
<td>Angstroms ($10^{-10}$ m)</td>
</tr>
<tr>
<td>B-factors</td>
<td>Approximation of the displacement of the atomic position from a mean value</td>
</tr>
<tr>
<td>B-form DNA</td>
<td>Right-handed helical DNA having 10.5 bp/turn</td>
</tr>
<tr>
<td>bp</td>
<td>Base pairs</td>
</tr>
<tr>
<td>Chy</td>
<td>Chymotrypsin</td>
</tr>
<tr>
<td>CPD</td>
<td>Cyclobutane pyrimidine dimer</td>
</tr>
<tr>
<td>CRY$_D$</td>
<td>Cryptochrome-DASH</td>
</tr>
<tr>
<td>$C_{\alpha}$</td>
<td>Alpha carbon</td>
</tr>
<tr>
<td>Da</td>
<td>Dalton; 1 Dalton = 1 atomic mass unit</td>
</tr>
<tr>
<td>DBO</td>
<td>2,3-diazabicyclo[2.2.2]oct-2-ene</td>
</tr>
<tr>
<td>DNA</td>
<td>Deoxyribonucleic acid</td>
</tr>
<tr>
<td>epPCR</td>
<td>Error Prone PCR</td>
</tr>
<tr>
<td>EPR</td>
<td>Electron paramagnetic resonance</td>
</tr>
<tr>
<td>FAD</td>
<td>Flavin adenine dinucleotide</td>
</tr>
<tr>
<td>FADhq</td>
<td>Reduced flavin adenine dinucleotide</td>
</tr>
<tr>
<td>FADox</td>
<td>Oxidized flavin adenine dinucleotide</td>
</tr>
<tr>
<td>FADsq</td>
<td>Semiquinone flavin adenine dinucleotide</td>
</tr>
<tr>
<td>FMN</td>
<td>Flavin Mononucleotide</td>
</tr>
<tr>
<td>FRET</td>
<td>Förster resonance energy transfer</td>
</tr>
<tr>
<td>ITC</td>
<td>Isothermal calorimetry</td>
</tr>
<tr>
<td>MALDI-TOF</td>
<td>Matrix-assisted laser desorption/ionization – time of flight</td>
</tr>
<tr>
<td>MD</td>
<td>Molecular Dynamics</td>
</tr>
<tr>
<td>MTHF</td>
<td>5,10-methenyltetrahydrofolate</td>
</tr>
<tr>
<td>NER</td>
<td>Nucleotide excision repair</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>NMR</td>
<td>Nuclear Magnetic Resonance</td>
</tr>
<tr>
<td>PK</td>
<td>Proteinase K</td>
</tr>
<tr>
<td>PL</td>
<td>Photolyase that repairs cyclobutane pyrimidine dimers</td>
</tr>
<tr>
<td>rf</td>
<td>Retardation value; Ratio of distance traveled by component relative to solvent front</td>
</tr>
<tr>
<td>RP-HPLC</td>
<td>Reverse-phase high pressure liquid chromatography</td>
</tr>
<tr>
<td>SDS-PAGE</td>
<td>Sodium dodecyl sulphate polyacrylamide gel electrophoresis</td>
</tr>
<tr>
<td>UV</td>
<td>Ultraviolet</td>
</tr>
<tr>
<td>UVA</td>
<td>Ultraviolet A</td>
</tr>
<tr>
<td>UVC</td>
<td>Ultraviolet C</td>
</tr>
<tr>
<td>ε</td>
<td>Extinction coefficient</td>
</tr>
</tbody>
</table>
Chapter 1.

Introduction

1.1. An introduction to photolyase

1.1.1. Photolyase repairs cyclobutane pyrimidine dimers

Photolyases are a family of enzymes that catalyze the repair of a specific type of ultraviolet (UV) induced deoxyribonucleic acid (DNA) lesion called thymine dimers. There are two types of photolyases, which are categorized on which type of thymine dimer they repair. The cyclobutane pyrimidine dimer photolyases (PL) repair cyclobutane pyrimidine dimers (CPDs), while the 6-4 pyrimidine-pyrimidone photoproduct photolyases, repair (6-4) photoproducts.

Figure 1.1. Chemical structures of a thymine pair and its UV photoproducts, CPD and [6-4] photoproduct.

Note: A pair of thymines, each with double bonds between their own number 5 and 6 carbons. Given UV light, it can be converted to a CPD, with C5-C5 and C6-C6 linking bonds. Alternatively, a C4-C6 bond can be formed with a hydroxyl group addition to the C6 of the 5’ thymine. Reprinted with permission from (4). Copyright (2003) American Chemical Society.
1.1.2. Thymine dimers are an important product of UV damage

Thymine dimers are products of ultraviolet (UV) light damage in DNA.(Fig. 1.1) They are developed predominantly through a photoexcited electron singlet state. The reason that thymines are the target of the UV damage is that, compared to other bases, the emission and absorption spectra of thymines have the lowest excited triplet levels of the 5 nucleotides. Thymine dimers can take on different forms. Irradiation of mammalian cells with Ultraviolet C (UVC) causes DNA lesions, at least 80% of which are CPD. This was found using a mouse cell line carrying transgenic mutation reporter genes, specifically, lacI and cII. Of the CPD generated in double stranded DNA, 99% are cis-syn and 1% are trans-syn.(Fig. 1.2)

![Figure 1.2. Structures of two isoforms of CPD.](image)

**Note:** Two isoforms of CPD can be formed, a cis-syn or a trans-syn-I. The substituent ring groups are oriented in the same direction in the cis-syn isoform, whereas they are oriented in different directions in the trans-syn-I isoform. Both are syn, as the N-1 atoms are oriented in the same direction. Adapted from Kim.

CPD lesions distort the DNA locally about the lesion. There are crystal structures for UV induced DNA damage, which show a 30° bend about the CPD lesion. Consistent with this data, quantitative electron microscopy was also used to show that
dimers produce a 30° bend in DNA. A 30-35° bend was also calculated based on gel migration. While the thymine dimer is localized in the duplex in the crystal structure, it is currently not known if this is true in solution.

The structure of the CPD lesion has an impact on the global DNA structure. Atomic force microscopy shows that there is unwinding of the supercoil in plasmids and progressive linearization of plasmids due to UV exposure. Solid state fluorescence studies of CPD opposite a fluorescent cytidine analog was used to determine the extent of deformation in DNA. Consistent with the crystal structure, the degree of base destacking in the bases flanking the adenines on the complementary strand was minimal.

Both the local and global structures of DNA are affected by CPD formation. The structure of the CPD itself is not central to this thesis. However, the CPD structure is of great importance to binding function of PL.

UV exposure to cells lacking thymine dimer repair pathways can result in cell death. The effect of UV damage to DNA and cells can be extensive, including single stranded breaks in the DNA, apoptosis in cells, damage to the nucleus of the cell, or even tumour development. It is proposed that photorepair was one of the first DNA repair pathways to evolve.

1.1.3. PL has two domains conserved in a Cryptochrome/Photolyase superfamily

PL contains two domains, an α-helical domain and an α/β-domain. In the first PL crystal structure, these domains were outlined for E. coli PL. In E. coli, residues 1-131 are the α/β-domain, while 204-472 are the C-terminal α-helical domain (132-203 is the interdomain loop). The binding domain is defined as the positively charged cleft that accommodates the substrate and is contained in the C-terminal domain. The secondary structure of PL remains largely unchanged upon substrate binding, even up to fluences of UVC of 170 µW/cm², as seen in CD spectra.

3
Figure 1.3. PL domains.

Note: Top, two main domains can be identified in *A. nidulans* PL (PDB: 1qnf). The N-terminal $\alpha/\beta$ domain includes $\alpha$-helices 1-5 and $\beta$ sheets 1-5. The C-terminal $\alpha$-domain includes $\alpha$-helices 8-18 along with three $3_{10}$ helices 6-7 which forms an interdomain loop. Where $3_{10}$ helices are polypeptide helices with backbone hydrogen bonding 3 residues apart. $\alpha14$ and $\alpha15$, purple, and $\alpha9$, yellow, are shown for orientation in use with future figures. Bottom, the amino acid sequence with secondary structure elements denoted in red lines, $\beta$ denotes $\beta$-sheets, L denotes loops, and $\alpha$ denotes $\alpha$-helices.
PLs can be putatively identified by their sequences. PLs contain certain signature sequences in the last 150 carboxy-terminal amino acids, which generally share around 30% sequence identity. The N-terminal region is less conserved and is not as useful in the identification of PLs\textsuperscript{22}.

PL can be categorized into class I and class II. Class I PLs have a 25-43% sequence identity within this group, exemplified by PL from \textit{Escherichia coli}. Class II PLs have a 38-72% sequence identity within this group, exemplified by PL from \textit{Myxococcus xanthus}. Between the classes, there is no more than 17% sequence identity found within the whole sequence, not just in the C-terminal region as described earlier\textsuperscript{23,24}. Interestingly, most species do not encode both class I and class II PLs, as determined by whole genomes analysis\textsuperscript{19}.

Cryptochromes are proteins related to photolyases; together, they form the CRY/photolyase superfamily. Cryptochromes are defined as blue light photoreceptors which generally (see section 1.4.1 for exceptions) have no photolyase activity, but share a high sequence identity, 25-40%, and structural homology to photolyase\textsuperscript{25}. Cryptochromes have developed a novel function\textsuperscript{26–28}.

Photolyases are thought to have evolved first and diverged into CRY. Photolyase is believed to be one of the first DNA repair pathways\textsuperscript{19}. Class I and class II photolyases are found in each of the major domains of life, which is consistent with the claim that photolyase is one of the first DNA repair pathways to have evolved. Cryptochromes also have representatives in all kingdoms of life, with the exception of Archaea\textsuperscript{25}.

Photolyase has however been lost or evolved different functions in many eukaryotic organisms, including humans. The loss of photolyase in eukaryotic genomes may be due to weak natural selection and results in deleterious increase in mutations rates, as seen in a phylogenetic analysis\textsuperscript{23}. This is in contrast to the nucleotide excision repair pathways, which are prevalent in eukaryotic organisms exclusively.

Understanding how similar family members diverged in function will help in understanding how PL is able to bind DNA efficiently. As binding is a central question in
this thesis, a comparison of these areas is important. This work uses a representative member from the PL and CRY families in comparative studies.

1.1.4. PL has importance in human health as well as other organisms

PL, while not found naturally in humans, has applications in health and science. PL is not detected in human cells, as human cell lysate is unable to repair CPD in an 

*in vitro* activity assay. There have been developments that have made it possible for PL to be used to supplement human DNA repair pathways. If PL is delivered to non-native cells, they are then able to supplement the host with additional resistance to UV damage. In addition to creating transgenic organisms, two unique ways to do this have been developed. This includes a specialized liposome containing PL and using a film forming device.

Studies have been undertaken to determine the effects of PL in transgenic organisms. Transgenic mouse lines made to contain PL were subjected to UV irradiation. Transgenic mice containing PL were significantly more resistant to tumour formation compared to wildtype mice.

A unique method of transferring PL to non-native cells is the photosome. The photosome is a liposome that contains purified PL. The photosome is able to fuse with the cell membrane, delivering the PL to the cytoplasm. The supplementary PL can then aid the host's CPD repair.

The beneficial effects of the photosome have been quantified. PL transferred to keratinocytes via photosomes can improve UV resistance, as monitored by a decrease in single stranded DNA breaks 2.6 to 3.3-fold. Reduction in apoptosis and CPDs in human skin cells are also observed. Photosomes in sunscreen reduces p53 expression, as a measure of nuclear damage, up to 2.5 times the amount compared to standard SPF 50 sunscreen.

PL can aid in the treatment of actinic keratosis patients when PL is delivered via a film-forming device. This method was used in clinical trials by monitoring actinic
keratosis, thick patches of pre-cancerous skin. These results, while preliminary, indicate that this method would be effective in treatment of actinic keratosis.

These methods for delivering PLs to supplement mammalian repair activity of UV-damaged DNA are a significant medical advance. While this is not a direct focus of this thesis, further understanding of how this enzyme binds may be important for further advancements in this area.

PL is important for DNA repair in several plant species including rice, Arabidopsis, and spinach. In rice, PL is an integral protein required for survival and is the main factor in UV resistance. PL transgenic rice plants, made to have a greater amount of PL than wild type, are more resistant to UVB-induced growth inhibition. Arabidopsis with overexpressed PL were shown to have increased DNA repair as well. PL is also found in the spinach, Spinacia oleracea. Replacing the PL gene with that of E. coli PL only partially restored photoreactivation function.

PL has members in several species which has been of scientific interest. Fowlpox virus deficient in PL were not infective until PL was restored to them under UV conditions. PL was also found in a baculovirus from Chrysodeixis chalcites, through PCR using degenerate primers. The pathogenic bacteria, Y. Enterolytica, also has the ability to photoreactivate. E. coli was the first organism to be discovered to have the ability to repair UV damaged DNA.

Examples from PL are also found in fish and amphibians. PL from goldfish, Carassius auratus, has PL that can be induced by visible light by 10-fold. Through isolating cDNA, PL was found in and purified from Xenopus laevis. The abilities to withstand UV exposure are different among species; amphibian embryos had significantly higher sensitivity to UV irradiation.

Understanding the different PLs from a variety of organisms is of interest to many researchers. While this thesis only focuses on a single, previously discovered PL, understanding the history of what has been accomplished and the current interest in this field is important.
1.1.5. Detection of CPD

There are different ways in which CPDs can be detected. Crude assays employing plasmid transformation were the initial in vitro detection method reported in the literature\(^{44}\). Chemical degradation of substrate has also been used\(^{48}\). Reverse-phase high pressure liquid chromatography (RP-HPLC) can be used to separate DNA containing thymine-dimers and undamaged DNA\(^{49}\). More modern methods include the use of endonucleases, UV-absorbance and gel electrophoresis\(^{21,50,51}\).

An early and crude method to verify CPDs is the plasmid transformation assay. This assay can be used to detect thymine dimers contained in plasmids. The basis of this assay is that thymine dimer containing plasmids do not transform as efficiently as those with no damage\(^{44}\). This assay has the downfall of not being as quantitative as other assays available. Another limitation of this assay is the requirement for the substrate to be covalently closed circular DNA, as linear DNA will not transform.

A later advance in CPD detection took advantage of chemical degradation. The C4 of thymine dimers is more susceptible to attack by nucleophilic reagents. The products of this degradation can be detected using gel electrophoresis. This assay allowed for use of short oligos\(^{48}\). This method requires the oligonucleotide to be resolved on a gel. Short, linear oligonucleotides work best, since circular plasmid DNA may make the analysis more complex due to supercoiling.

More modern CPD detection techniques take advantage of different properties of CPDs. One example is using restriction enzyme site protection to detect the CPD. This process relies on the resistant nature of CPDs to cleavage by restriction endonucleases. The substrate is a DNA strand containing a pair of thymines situated in an endonuclease site. After irradiation, DNA strands with CPDs generated in this site will be resistant to that endonuclease, while undamaged DNA strands will be cleaved. This can be detected by gel electrophoresis\(^{50}\). The DNA here may be linear or plasmid, but the substrate and product must resolve separately by gel electrophoresis. The limitation here is that forms of DNA other than B-DNA may not be cleaved by certain enzymes\(^{52}\).
A gel shift assay can also be used to detect binding of CPD by PL. Here, PL can bind to CPDs, and will remain bound to the substrate to be detected by gel electrophoresis\textsuperscript{21}. This method relies on the ability of PL to bind CPD with higher affinity than undamaged DNA\textsuperscript{53,54}. Similarly to other methods that require gel electrophoresis for visualization, this assay is limited by the complex of PL to DNA being able to resolve from the substrate.

Another method for detection relies on the differences in the UV light absorbances between CPD and native DNA. The conjugated pi electrons in the pyrimidine ring are the main cause for the absorbance in the 260 nm region. CPD-containing DNA has a lower extinction coefficient than undamaged DNA. Monitoring the amount of CPDs in a poly(dT) DNA strand is possible by monitoring the decrease in the absorbance after CPDs are induced\textsuperscript{55}.

### 1.2. PL repairs CPDs efficiently using light energy

#### 1.2.1. PL uses UV light to overcome the activation energies of CPD repair and results in a high efficiency of energy transfer

PL repairs DNA \textit{in vitro} using UVA and blue light. PL requires light between approximately 300 nm to 500 nm in order to initiate DNA repair. In this first series of experiments\textsuperscript{56}, PL repair was monitored by the plasmid transformation assay. This was also monitored by using restriction endonucleases and gel electrophoresis to determine if the CPD was repaired\textsuperscript{9} (see section 1.1.5).

PL can repair dimers between thymines (T<>T), uracil and thymine (U<>T), uracils (U<>U) and cytosines (C<>C). PL repairs different types of dimers at 366 nm light with quantum yields of 0.9 (T<>T), 0.8 (U<>T), 0.6 (U<>U) and 0.5 (C<>C)\textsuperscript{55}. Activation energies for PL ranged from 1.9 kJ mol\textsuperscript{-1} to 0.9 kJ mol\textsuperscript{-1} for thymine dimers or uracil dimers\textsuperscript{57}. PL also is capable of repairing CPD dimers in the trans-syn conformation, though at a 10\textsuperscript{4}-fold lower efficiency than that of cis-syn dimers\textsuperscript{9}. Notably, PL is reasonably efficient at repairing cytosine dimers which have several different functional groups compared to thymine dimers.
The turnover rate of PL has previously been calculated. In the first purification of PL, the turnover rate was calculated to be 2.4 dimers/molecule/min for *E. coli* PL. This was measured by a transformation assay, where UV-damaged plasmids are unable to transform\textsuperscript{58}. Consistent with this result, RP-HPLC monitoring of the product showed a PL turnover rate of 2.4 dimers/molecule/min\textsuperscript{49}.

The catalytic chemistry is efficient and fast, occurring on the nanosecond timescale. The quantum efficiency of PL is 0.92, determined by fluorescence lifetimes of PL and apoenzyme variants of PL\textsuperscript{59}. Using subpicosecond UV transient absorption spectroscopy, the repair of the carbon-carbon bond of the CPD was initiated in ~60 ps and the bond is broken by 1,500 ps\textsuperscript{60}.

The catalytic chemistry occurs much faster than that of the actual turnover rate. This is likely affected by the rate at which PL comes into contact with its substrate. The turnover rate is at least in part limited by Brownian motion. While the focus of this thesis is on binding, the rate at which both the enzyme can perform and at what efficiency it can perform is crucial to understanding how this enzyme functions.

Larger fluences of light can inactivate PL. PL is inactivated at UVC fluences above 25 mW/cm\textsuperscript{2}, seen in an *in vitro* assay. Though the binding was unaffected, as determined by a gel shift assay, the PL was unable to dissociate from the substrate. The inactivation of PL is not caused by changes in secondary structure, as monitored by CD spectroscopy. Fluorescence evidence suggests that changes in tryptophan residues are a factor, possibly by being more exposed to solvent\textsuperscript{21}.

During this thesis, fluences of light are necessarily applied to PL. Having a good understanding of the limitations of the amounts of fluences that can be used is important in preventing inactivation of the enzyme.

1.2.2. Photolyase function relies on a FAD cofactor

The reaction of CPD repair by photolyase involves a single photolyase molecule which catalyzes the splitting of a cyclobutane ring in the CPD by transferring a single electron. This subsequently reforms two thymines and regenerates the enzyme through
back electron transfer. All PLs have a Flavin Adenine Dinucleotide (FAD) cofactor, which contains the isoalloxazine group from where the single electron is abstracted. PLs also often contain a secondary cofactor (see section 1.2.5). In order for this transfer to occur, this electron needs to be in a high energy state, facilitated by absorbance of a single photon of UVC light. In this thesis, I ask how photolyase is capable of binding DNA efficiently and how photolyase can maintain this bound state. Since part of binding involves conformational change in photolyase, I further ask which conformational change and what aspects of conformation change are important for CPD binding. The catalysis of CPD has been studied extensively previously, and while important to the function of this enzyme, this is a separate step from binding, which I will not be covering extensively.

### 1.2.3. FAD is a common biochemical cofactor

FAD contain specific moieties. The FAD consists of a riboflavin moiety bound to a phosphate group of an adenosine diphosphate molecule. The riboflavin consists of a ribose bound to an isoalloxazine group.

FAD is a common cofactor in several classes of enzymes. Seven classes have been described based on activity; the oxidation of carbon-heteroatom bonds, oxidation and reduction of carbon-carbon bonds, thio/disulphide chemistry, electron transfer reactions, oxygen reactions, non-redox reactions, and complex flavoenzymes.

The “oxidation of carbon-heteroatom bonds” is the class of flavoenzymes that catalyzes the oxidation between bonds of carbon-heteroatoms, such as D-amino acid oxidase. Another class is the “oxidation and reduction of carbon-carbon bonds” which uses the N5 of FAD to transfer a hydride, such as succinate dehydrogenase. The class of “thiol/disulphide chemistry” utilizes cystines to reduce the FAD and produce a thiol, such as glutathione reductase. “Electron transfer reactions” are another class of FAD containing enzymes. They transfer single electrons and require a source of electrons to regenerate their reduced state; an example of this type of enzyme is cytochromes P-450 reductase. Another class is the “oxygen reactions,” where oxygen reactivity is modulated by the FAD cofactor. An example of this are oxidases. “Complex flavoenzymes” are a
class which doesn’t fit into any of these categories. An example is glutamate synthase. Finally, the class of enzymes that PL is in is the “nonredox reactions.” This class differs from the electron transfer reactions in that there is no net redox change.

FAD can exist in multiple redox states. FAD is generally used by enzymes for its redox bioenergetics, which is active in the 7,8-dimethyl isoalloxazine moiety. The redox states include fully oxidized (FADox) the free radical semiquinone (FADsq) and fully reduced (FADhq). A single electron can be removed from the FADhq to form the FADsq, which is stable in PL. (Fig. 1.4)

**Figure 1.4. Structures of three FAD redox states and their pathways.**

Note: FAD can be observed in three states. The fully reduced, FADH is covalently attached to a hydrogen on the N5 of the isoalloxazine moiety (shown). A single electron can be lost to molecular oxygen to produce the FADsq, which still maintains the attached hydrogen on the N5, but now has a single unpaired electron. The hydrogen can be abstracted to produce FADox. Adapted from O’Neill.

The FAD cofactor plays a central role in catalysis in PL. While my thesis is not focused on catalysis, it is important to understand how PL repairs its substrate.
Specifically, this information would be central to help distinguish between a binding compromised PL mutant and a catalytically inert PL mutant.

There are various methods by which the different redox states of FAD can be determined. One method is by measuring the absorbance of the FAD containing enzyme. A common method is the use of steady state spectroscopy. Femtosecond spectroscopy can also be used in a similar way\textsuperscript{66,68}.

Each of the FAD states has a characteristic absorbance spectrum. FADhq does not absorb above 500 nm, while FADsq has absorbance peaks at 500 and 580nm. The neutral semiquinone of PL, FADsq, has absorbance peaks at 360 and 470nm, while FADox has absorbance peaks at 380 and 450 nm\textsuperscript{68}.

Using femtosecond spectroscopy, specific states of the FAD can be excited at different wavelengths and intramolecular quenching can be tracked\textsuperscript{68}. Different FAD states can be monitored using femtosecond spectroscopy to monitor rates of decay of each state\textsuperscript{66}.

Electron paramagnetic resonance (EPR) spectroscopy has also been used to monitor the flavin when it had unpaired electrons. EPR spectroscopy detects the electron spin, similar to nuclear magnetic resonance detecting the nuclei spin\textsuperscript{62,69}. EPR was also used to detect the presence of an iron-sulfur cluster in a PL\textsuperscript{70}.

While understanding the PL states is important, it is not central to this thesis. However, these properties are helpful for identifying the PL redox states. It is also helpful for identifying if there is a homogeneous or heterogenous mix of FAD states.

1.2.4. The active form of the FAD is fully reduced in PL

The FAD cofactor is the catalytic region of the enzyme\textsuperscript{58} which is active when it is fully reduced, both \textit{in vivo} and \textit{in vitro}. \textit{In vitro}, only the fully reduced form is able to efficiently repair CPD lesions induced in whole plasmid\textsuperscript{71,72}. Oxidation of FAD leads to a reversible loss of photoreactivation activity\textsuperscript{72}. The functional state in PL is FADhq due to
it having the slowest electron transfer dynamics of 2 ns with the adenine moiety and the fastest ET dynamics of 250 ps with the substrate.\textsuperscript{68}

PL is in the fully reduced state under physiological conditions. \textit{In vivo}, the semiquinone form (FADhq) cannot be detected by electron paramagenetic resonance until lysis and column purification indicating that the FADhq occurs naturally \textit{in vivo} and oxidizes during purification.\textsuperscript{71} Mutations that make the FADhq less stable make the enzyme catalytically inert \textit{in vivo}.\textsuperscript{17}

\textit{E. coli} PL in the fully reduced form repairs CPD with a quantum efficiency of close to 1 \textit{in vitro} and \textit{in vivo}. This was determined using a cross section of an \textit{in vivo} action spectrum (1.5-2.5 x 10\textsuperscript{4} M\textsuperscript{-1}cm\textsuperscript{-1} at 384 nm), using a killing assay and UV sensitive \textit{E. coli} cells, and the reported extinction coefficient (\(\varepsilon = 18.1 \times 10^4\) M\textsuperscript{-1}cm\textsuperscript{-1} at 384 nm \textit{in vitro}) at this cross section. The FADsq form of PL has a significantly lower quantum yield of 0.07, \textit{in vitro}.\textsuperscript{71,73}

Photoreduction is the process of using light to induce reduction of PL with the use of an electron donor.\textsuperscript{17} Transient absorption spectroscopy elucidated the electron transfer chain FAD-W382-W359-W306 in \textit{E. coli} PL during photoreduction.\textsuperscript{74} W306 is also important since it is the hydrogen atom donor for FAD photoreduction.\textsuperscript{74–76} Mutating W306 to phenylalanine disables PL electron transfer in this electron transfer chain \textit{in vitro}.\textsuperscript{75} (Fig. 1.5) However, \textit{in vivo}, intraprotein electron transfer is not part of the catalysis of PL.\textsuperscript{77}
The work in this thesis includes repair assays. Therefore it is important to understand under which conditions PL is active. It is also important to understand in what ways mutations might affect the active state, as mutations are also made for this thesis.

1.2.5. **The catalytic cycle of PL occurs stepwise in ~1 ns.**

The rate of repair after CPD is oriented in the PL active site is found to be on the ultrafast time scale of 560 ps. Time scales and efficiency were monitored through absorbance spectroscopy and ultrafast fluorescence spectroscopy. The full electron transfer reaction is ~1 ns. The catalytic cycle is reviewed by Sancar.
The catalytic cycle of PL occurs in four steps. This full electron transfer mechanism requires: 1) Excitation of FADhq to FADhq*, 2) transfer of a single electron from FADhq* to the CPD, forming FADsq and an excited CPD 3) Breaking of bonds between C6-C6 and C5-C5 in the CPD, 4) back electron transfer from the repaired CPD to the FADsq to regenerate the initial FADhq state.

Understanding the catalytic cycle is important for performing repair assays, as is undertaken in this thesis. While the actual catalytic cycle is not central to this thesis, it is necessary for the undertaking of this work.

Secondary cofactors are important in PL for expanding the range of possible wavelengths that can be used in the repair reaction. The secondary cofactor has a different absorbance profile than the catalytic cofactor. This allows a range of wavelengths to excite an electron in the secondary cofactor. This excitation energy can be passed to the catalytic cofactor through intermediates using Förster resonance energy transfer (FRET).

Typically, PL has one of two different possible secondary cofactors: MTHF or 8-HDF, which expand the range of wavelengths of light PL can absorb. Interestingly, PL from Sulfolobus tokodaii contains two FAD cofactors, one in the secondary cofactor location, as seen in crystal structures. A fourth type of possible co-factor was found, which is an iron-sulfur cluster. Homology modelling and EPR spectroscopy suggests that a Fe-S cluster is a co-factor in a PL from Agrobacterium tumefaciens. Finally, a fifth cofactor, flavin mononucleotide, was found in Thermus thermophilus to be the secondary cofactor after PL purification, spectroscopic analysis, reverse-phase HPLC and Nuclear Magnetic Resonance (NMR) analysis of the released chromophore.

The catalytic cycle starts with the isoalloxazine subunit of the FAD in its fully reduced state, FADhq. The FADhq gets excited by light to FADhq* or gets to this state using energy transferred from the secondary cofactor. Using time resolved fluorescence and absorption spectroscopy, it was found that the energy transfer from MTHF to FADhq and FADHsq occurred at rates of $6.4 \times 10^9$ and $3.0 \times 10^{10}$ s$^{-1}$, respectively. These numbers imply that the energy transfer step is not the limiting factor in catalysis, as it is on the nanosecond timescale, similar to catalytic repair.
In the second step of catalysis, the CPD abstracts an electron from the FADhq* to form two radicals, FADsq and a CPD radical\textsuperscript{80}. In this state, the CPD bases are in the form of the radical anion, thymidine. This process happens in less than 35 ns, as shown through radiolysis\textsuperscript{88}.

This process is facilitated by the unusual bent conformation from the FAD cofactor\textsuperscript{12}. The localization of the adenine moiety in close proximity to the isoalloxazine moiety is thought to help the electron transfer between the FAD and the substrate, as seen in CRY\textsubscript{D} and (6-4) photolyase\textsuperscript{89–92}. The role of adenine is important, as shown in Flavin Mononucleotide (FMN) being less able to maintain an excited state than FAD\textsuperscript{91}.

Once an electron is transferred from FAD to the CPD, the FADsq is stabilized by PL to be long-lived, unlike most radicals. This can be observed during PL purification, as it purifies in the blue semiquinone state due to oxidation by O\textsubscript{2}\textsuperscript{12,83}. In \textit{E. coli} PL, the stability of FADsq is, in part, due to N387, which forms a hydrogen bond with the N5 hydrogen on the isoalloxazine moiety of FAD\textsuperscript{17}. The stabilization of the FADsq is important for preventing back electron transfer from the anionic CPD free radical\textsuperscript{17}.

In the third step, the cyclobutane ring of the CPD is split. CPD splitting occurs on the nanosecond time scale, as seen by transient absorption spectroscopy\textsuperscript{93}. Molecular dynamics suggest that there is a C5-C5' linked intermediate during photoreactivation\textsuperscript{94}. Molecular dynamics calculation estimated the time for the C5-C5 bond of the CPD to break was between 100 to 150 fs after electron transfer from the FAD\textsuperscript{95}. Theoretical analysis suggests that the high performance of PL is in part due to its ability to suppress the back reaction\textsuperscript{96,97}.

The radical electron on the thymine is, in the last step, transferred back to the FADsq radical to regenerate FADhq. The back electron transfer happens with a \( \tau \) of 560 ps in \textit{E. coli} PL\textsuperscript{78,98}. The long-lived FADsq allows for back electron transfer to the active FADhq state\textsuperscript{12}. 

1.3. “Base-flipping” is required by PL and other enzymes

1.3.1. The PL binds to only a 4-6 bp region

DNA footprinting and mutagenesis of PL has also helped to elucidate the way in which PL binds to substrate. The PL binds to only a 6 bp region, as seen by DNA footprinting, with the CPD in the center. The footprint of DNase I yields 12 and 10 bp footprints, around the CPD, in dsDNA and ssDNA, respectively. The high resolution Methidiumpropyl-EDTA-Fe(II) footprinting shows the specific location to be between a 4-6 bp region\(^9\). The difference in the numbers is likely due to DNase I being sterically hindered from certain bases near the PL active site by PL.

A light flash that allows PL to repair this DNA removes any protection previously given to the DNA from DNase I. While DNA footprinting is consistent with the binding region that is seen in the crystal structure, it does not provide evidence for the state or orientation of the CPD relative to the duplex before, during, and after binding to PL.

Mutations in the binding cleft of PL reduce the ability for PL to bind and discriminate between substrates. Discrimination is the ratio of binding affinity for substrate relative to non-substrate for PL\(^9\). The decrease in discrimination between substrates and the loss of recognition, measured in binding affinity, in these mutants indicates that the mutated residues are important in helping to stabilize the substrate. The specific mutations made were in *Saccharomyces cerevisiae* PL at residues K383, R452, and Q514 and showed both a loss of affinity for CPD and ability to discriminate between damaged and undamaged DNA\(^1\). In *A. nidulans* numbering, these residues are Q282, R350 and Q411. The mutant R311A from *T. thermophiles*, aligned to *A. nidulans* R350, is not able to bind to DNA\(^1\). These residues are predicted to interact with the CPD, but the *A. nidulans* Q282 interaction was not found in the substrate bound PL co-crystal structure.(Fig. 1.6)

PL recognizes specifically the CPD itself and the changes in the DNA structure. The configuration of the DNA backbone and the cyclobutane ring are the important structural determinants of PL recognition. PL binds in the major groove of the substrate, as shown by a DNA protection assay, and confirmed with crystal structure\(^1,9\). However,
major groove contacts cannot be important in substrate recognition, as the enzyme binds equally well to a ssDNA CPD. Therefore, the structures that may be important in substrate recognition must be the cyclobutane ring or the phosphodiester backbone of the CPD containing strand.

1.3.2. The CPD is bound to PL in an extrahelical state and both the CPD and its complementary strand form interactions with PL

Three important crystal structures exist that show states of PL by itself, CPD by itself and PL bound to CPD. Comparison of these structures helps to elucidate important features of CPD binding. A limitation of these studies is that only a snapshot of a single state is observed. For example, the repaired state provides limited information as to the mechanism by which it got there. Possible mechanisms for this include an extrahelical CPD binding to the PL or the PL initiating a movement of the CPD into an extrahelical position.

A 1.8 Å crystal structure of PL was solved with substrate. The substrate used was a 14-mer of dsDNA with a single thymine pair linked by a formacetal link instead of the typical phosphate linkage.

Based on the cocrystal structure, the positively charged groove in PL and certain PL residues form interactions with the substrate. The phosphodeoxyribose backbone of the lesion strand, specifically the P-1, P+1, P+2, and P+3 phosphates from the central phosphate connecting the two thymines from the 5’ to 3’ direction, makes contacts with a positively charged groove in PL. (Fig. 1.6)
Figure 1.6. Contacts between PL and CPD.

Note: Contacts, based on crystal structure from *A. nidulans* PL bound to CPD (PDB: 1tez), includes hydrogen bonds, ionic interactions, and Van der Waals interactions. Arrows indicate interactions with residue side chains, dashed arrows indicate protein backbone interactions. Blue circles indicate residues from the PL recognition loop. Red boxes indicate residues that have lost CPD binding affinity and the ability to discriminate between damaged and undamaged DNA in previous mutation\(^{10}\). Phosphates are numbered upstream or downstream of the dimer. Adapted from Mees\(^{10}\).
The CPD has local structural changes when bound to PL. The CPD itself is displaced 180° outside of the DNA duplex and faces the FAD cofactor. The CPD bound to PL was repaired in this structure\textsuperscript{10}. This is in contrast to the crystal structure of CPD-containing DNA, which is located in the interior of the duplex\textsuperscript{14}. (Fig. 1.7) In the CPD crystal structure, there are bonds forming the cyclobutane ring of the CPD\textsuperscript{14}.

\textbf{Figure 1.7. Angle changes in bound and unbound CPD-containing DNA.}

Note: Structure of CPD (PDB: 1n4e)\textsuperscript{14} and structure of CPD bound to PL (PDB: 1tez)\textsuperscript{10}, DNA shown in stick: purple carbons, CPD containing strand; yellow carbons, complementary strand, PL shown as cartoon. \(\alpha14\) and \(\alpha15\), purple, and \(\alpha9\), yellow, are highlighted for orientation. The angle of the unbound CPD is 22°, right, which changes to 50°, left, in the final bound state\textsuperscript{10}. Dots are non-connected atoms in the crystal structure.

The thymine dimer forms interactions with the FAD and PL residues due to its specific bound position. The thymine dimer is oriented to face the isoalloxazaine moiety of the FAD cofactor. The C4 carbonyls of both the repaired thymines hydrogen bond with the N6 amide of the adenine moiety in FAD. Hydrogen bonds between the 3' thymine and N349 and 5' thymine and E283 also occur. This could potentially help stabilize the CPD radical anion\textsuperscript{10}. 
Some of the CPD global structure changes when bound to PL. As mentioned earlier (see section 1.1.2), CPD creates a 30-35° bend in the DNA\textsuperscript{13,14}. This angle increases to 50° after being bound and repaired by PL, as seen in the PL co-crystal structure.\textsuperscript{(Fig. 1.7)} These angles are measured by drawing a line in along the center of the duplex from the 5´ and 3´ ends and measuring the change in angle where they meet. The CPD substrate maintains a B-type conformation outside of the thymine dimer region, as it does in its unbound state. Again, similar to the unbound substrate, the adenines complementary to the CPD stack with their adjacent bases, but not with each other due to the change in angle\textsuperscript{10}.

PL forms non-covalent bonds to the non-lesion strand. PL forms van-der-Waals interactions between P402 and the adenines in the complementary strand. A hydrogen bond forms between the amide of L403 and the phosphate groups linking the two adenines. The hole left from the extrahelical CPD is filled by PL residues. The 10 by 10 Å hole left due to the movement of the CPD is partially occupied by a loop of residues G397 to F406, which is part of the recognition loop\textsuperscript{12}.\textsuperscript{(Fig. 1.8)}
Figure 1.8. PL bound to CPD.

Note: PL, shown as cartoon, not bound (1qnf) or bound (1tez) to CPD lesion, shown as sticks: purple carbons, lesion strand; yellow carbons, complementary strand. FAD shown at sticks with green carbons. The bases outside the CPD region form a duplex. The CPD is in an extrahelical position, 180° outside of the duplex, positioned in the active site of PL. α14 and α15, purple, and α9, yellow, are shown for orientation.
1.3.3. The energy cost of displacing a CPD from the duplex is half that of a thymine pair

Binding of the CPD requires a certain amount of energy to remove the CPD from the duplex into an extrahelical position. Moving the CPD into the extrahelical position is not spontaneous, which means that PL would need to stabilize this structure. Molecular Dynamics (MD) simulations determined that the energies of flipping a CPD would be between 6.25 and 7.5 kcal/mol, much lower than that of undamaged thymines, ~15 kcal/mol\textsuperscript{102}. Interestingly, MD simulations determined that the energy of a flipped out CPD is roughly 6.5 kcal/mol higher than one contained in the duplex: a lower barrier compared to undamaged DNA ranging between 13-19 kcal/mol\textsuperscript{103}.

The surrounding bases and the orientation of flipping of the CPD are important for the energies of base flipping. Through MD simulations, it was shown that these immediately surrounding bases altered the energies of flipping by a total of ~2 kcal/mol\textsuperscript{102}. Molecular mechanics and MD determined free energy differences of base flipping between major and minor grooves to be ~3 kcal/mol, from barrier heights of 21.3 kcal/mol, major groove, and 18.7 kcal/mol, minor groove, for GC pairs\textsuperscript{104}.

These numbers imply that the energy cost of flipping a CPD is approximately half the cost of that of a thymine pair. This may be a way in which a repaired thymine dimer removes itself from the PL active site, as the repaired thymines are more energetically unfavourable in an extrahelical position. The surrounding bases do play a role in the energetics as well. This implies that some substrates may be more challenging for PL to bind.

This phenomenon, known as "base flipping," is observed in other enzymes. These include methyltransferases, DNA glycosylases, T4 Endo V, Exonucleases, DNA Ligase and Polymerase, as seen in cocrystal structures\textsuperscript{105}. The methyltransferase, M. HhaI, has a flexible loop which is also observed to reorganize to facilitate binding of the DNA substrate, as seen in crystal structures with and without substrate\textsuperscript{106}. This enzyme has interesting parallels to PL.
A fluorophore in the loop of M. HhaI was used to compare the change in fluorescence in the 2-AP across from the CPD upon addition of substrate to HhaI. This experiment showed that the flexible loop of M. HhaI reorganized upon binding DNA. The fluorescence of this loop is coupled to base flipping in this enzyme and loop motion precedes base flipping.

1.3.4. The mechanism by which PL binds to and accesses the CPD is not well understood.

There are two possible scenarios for CPD binding. One scenario is that the CPD is in an extrahelical position before it enters the PL active site. The other scenario is where the CPD is on the interior of the duplex when PL binds and the CPD is flipped after the binding to PL. To distinguish the two, information on the changes that happen to both the PL and the CPD upon binding is required. It is still not known whether PL binds to the CPD in duplex or non-duplex DNA.

There is support in the literature for two states for bound CPD to PL. The CPD lesion on its own forms a 30-35° angle about the lesion. However, in the cocrystal structure of PL and CPD, the CPD is bent at a 45° angle. The difference in the angle between these two states could come about by two potential mechanisms. The first mechanism would be in which the CPD binds to the PL, then increases its angle. The second mechanism would be if the CPD changes its angle first, then binds to the PL. The second mechanism is not likely as there would be a need for a spontaneous conformation change in the DNA. Thus the first mechanism seems to be more likely. The change in this angle suggests that there would be at least two states for the CPD to bind to PL.

There are two possible mechanisms by which PL might bind to CPD. The first mechanism would be in which the CPD, in an extrahelical state, binds to the PL after the recognition loop changes conformation. The second is where the DNA, with CPD in the duplex, is bound by PL, followed by recognition loop rearrangement and CPD base flipping.
The difficulty of studying DNA binding and base flipping by PL is the current weak ability to experimentally study the initial protein-DNA complex. While methods like crystallography, fluorescence, DNA footprinting have been used to characterize the state of binding of PL, they have many drawbacks in characterizing the state of the substrate before contacting the protein in solution. Specifically, co-crystal structures of PL bound to substrate only provide an image of the end state, and not the mechanism by which PL binds to CPD. Fluorescence can give real time data for very short timescales, however probing base flipping is challenging because assigning a signal to the initiating stage compared to the bound stage would be challenging to do. DNA footprinting requires binding of substrate to occur before endonucleases are added, so probing the state initiating binding is out of the scope of this technique.

An experiment was performed, using heat capacity ($C_p$), to help determine if the CPD enters PL in an extrahelical state. Heat capacity is the amount of heat required to increase the temperature of a substance one degree and can be measured by isothermal titration calorimetry (ITC). Heat capacity changes ($\Delta C_p$) upon DNA binding was estimated from the accessible surface area from the crystal structures of PL with and without substrate, see equation (1).

$$\Delta C_p = 1.34(\Delta A_{\text{nonpolar}}) - 0.586 (\Delta A_{\text{polar}}) \text{ J/Kmol}$$

Where $\Delta C_p$ is the difference in the heat capacity between PL bound and unbound to DNA, $\Delta A_{\text{nonpolar}}$ is the difference in surface area of water accessible non-polar residues in $\text{Å}^2$ between PL bound and unbound to DNA. Similarly, $\Delta A_{\text{polar}}$ is the difference in surface area of water accessible polar residues in $\text{Å}^2$ between PL bound and unbound to DNA. The $C_p$ was measured for PL with and without substrate using ITC. The estimation and measurement of the $\Delta C_p$ are similar, indicating similar changes in the accessible surface area upon substrate binding. This suggests that the CPD in the extrahelical position is recognized by PL, and not flipped by PL.

Fluorescence evidence has helped to identify part of the nature of CPD binding by PL. Using 2-aminopurine (2-Ap), the duplex structure of the DNA can be monitored.
DNA duplex, π-stacked 2-Ap is quenched due to a relaxation in the excited state from the π-stacking $^{110}$.

In ssDNA, the emission of 2-Ap is significantly large. Annealing this to a complementary strand decreases the emission. Upon PL binding with a CPD that is paired with 2-AP, the fluorescence increases. This indicates that the CPD structure is significantly perturbed upon, not before, interaction with PL, consistent with the model that the CPD is displaced out of the duplex after binding to PL $^{111}$ (see section 1.3).

NMR suggests that PL plays a role in allowing the majority of the CPDs to change to an extrahelical state. NMR analysis, measured by $^{31}$P NMR spectra of CPD with and without PL, showed a large chemical shift upon the addition of PL $^{101}$. This result is consistent with the 2-AP study.

1.3.5. **PL CPD binding involves conformational changes in a key recognition loop**

The secondary and tertiary structure of PL remains mainly in the same position both when substrate is bound and unbound $^{10,12}$. The overlay of these two structures shows very little deviation in the secondary structures as well as the position of the backbone. The region that significantly moves is located in the loop region between position 392 and 409 in PL.

Comparison of the loop W392 to A409 in the structures with and without substrate shows that it moves 4 to 10 Å to occupy the space left by the flipped thymines $^{10,12}$. The loop moves out of the way to avoid steric hindrances with the CPD and interacts with the complementary strand.

In the PL crystal structure, before DNA has bound, the loop residues are interacting with several residues. (Fig. 1.9)(Table 1.1 and 1.2) The list shows that 13 hydrogen bonding interactions occur between the recognition loop and other residues in PL. These interactions may be important in stabilizing the loop region when CPD is not bound.
Figure 1.9. Hydrogen bonds in the PL recognition loop region.
Note: Hydrogen bonds formed by the A. nidulans PL recognition loop with substrate free (PDB: 1qnf), above, and bound states (PDB: 1tez), below. Top, PL recognition loop sequence. Asterisks indicate residues that interact with DNA. Inset is whole PL (PDB: 1qnf) for orientation. Hydrogen bonding partners are shown in sticks, loop residues, green; non-loop residues, grey; DNA, blue. α14 and α15, purple, are shown for orientation.
<table>
<thead>
<tr>
<th>Loop Residue</th>
<th>Loop Atom</th>
<th>Partner Residue</th>
<th>Partner Atom</th>
<th>Distance (Å)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ala 394</td>
<td>O</td>
<td>Trp 309</td>
<td>N (side chain)</td>
<td>3.3</td>
</tr>
<tr>
<td>Ala 394</td>
<td>N</td>
<td>Trp 392</td>
<td>O</td>
<td>3.6</td>
</tr>
<tr>
<td>Ser 395</td>
<td>O (side chain)</td>
<td>Trp 367</td>
<td>N (side chain)</td>
<td>3.3</td>
</tr>
<tr>
<td>Ser 395</td>
<td>O</td>
<td>Arg 306</td>
<td>N (side chain)</td>
<td>3.0</td>
</tr>
<tr>
<td>Ser 395</td>
<td>N</td>
<td>Gln 391</td>
<td>O</td>
<td>3.6</td>
</tr>
<tr>
<td>Ser 396</td>
<td>O</td>
<td>Gln 391</td>
<td>N (side chain)</td>
<td>3.2</td>
</tr>
<tr>
<td>Asp 399</td>
<td>O (side chain)</td>
<td>Tyr 149</td>
<td>N</td>
<td>3.9</td>
</tr>
<tr>
<td>Asp 399</td>
<td>N</td>
<td>Ser 147</td>
<td>O</td>
<td>2.6</td>
</tr>
<tr>
<td>Arg 404</td>
<td>O</td>
<td>Lys 361</td>
<td>N (side chain)</td>
<td>3.0</td>
</tr>
<tr>
<td>Phe 406</td>
<td>O</td>
<td>His 457</td>
<td>N (side chain)</td>
<td>3.3</td>
</tr>
<tr>
<td>Asn 407</td>
<td>O</td>
<td>Ser 410</td>
<td>N</td>
<td>3.8</td>
</tr>
<tr>
<td>Pro 408</td>
<td>N</td>
<td>Asn 407</td>
<td>O (side chain)</td>
<td>3.1</td>
</tr>
<tr>
<td>Pro 408</td>
<td>O</td>
<td>Ala 412</td>
<td>N</td>
<td>3.7</td>
</tr>
</tbody>
</table>

Note: Hydrogen bonds from PL recognition loop without substrate. PL from *A. nidulans* (pdb: 1qnf)\textsuperscript{12} Atoms are in polypeptide backbone unless otherwise noted. Yellow highlighted interactions are conserved during substrate binding.
Table 1.2. Hydrogen bonds interactions of PL recognition loop with substrate bound

<table>
<thead>
<tr>
<th>Loop Residue</th>
<th>Loop Atom</th>
<th>Partner Residue</th>
<th>Partner Atom</th>
<th>Distance (Å)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ala 394</td>
<td>N</td>
<td>Trp 390</td>
<td>O</td>
<td>3.1</td>
</tr>
<tr>
<td>Ser 395</td>
<td>O</td>
<td>Trp 309</td>
<td>N (side chain)</td>
<td>3.1</td>
</tr>
<tr>
<td>Ser 395</td>
<td>N</td>
<td>Gln 391</td>
<td>O</td>
<td>2.6</td>
</tr>
<tr>
<td>Gly 397</td>
<td>O</td>
<td>Arg 306</td>
<td>N</td>
<td>3.4</td>
</tr>
<tr>
<td>Leu 403</td>
<td>N</td>
<td>DA9</td>
<td>O (Phosphate)</td>
<td>2.7</td>
</tr>
<tr>
<td>Arg 404</td>
<td>N (side chain)</td>
<td>Trp 392</td>
<td>O</td>
<td>2.8</td>
</tr>
<tr>
<td>Phe 406</td>
<td>N</td>
<td>Ser 357</td>
<td>O (side chain)</td>
<td>3.2</td>
</tr>
<tr>
<td>Asn 407</td>
<td>N</td>
<td>G13</td>
<td>O (Phosphate)</td>
<td>3.1</td>
</tr>
<tr>
<td>Asn 407</td>
<td>N (side chain)</td>
<td>C15</td>
<td>O (Phosphate)</td>
<td>3.0</td>
</tr>
<tr>
<td>Asn 407</td>
<td>O</td>
<td>Ser 410</td>
<td>N</td>
<td>3.5</td>
</tr>
<tr>
<td>Pro 408</td>
<td>N</td>
<td>Asn 407</td>
<td>O (side chain)</td>
<td>4.1</td>
</tr>
<tr>
<td>Pro 408</td>
<td>O</td>
<td>Ala 412</td>
<td>N</td>
<td>3.7</td>
</tr>
</tbody>
</table>

Note: Hydrogen bonds from PL recognition loop with substrate. PL from A. nidulans (pdb: 1tez). Atoms are in polypeptide backbone unless otherwise noted. Yellow highlighted interactions are conserved when substrate is unbound.

1.4. **CRY<sub>D</sub> shares important structural and functional features with PL**

1.4.1. **Cryptochromes have a wide array of functions that are poorly understood**

The CRY family is defined by having a high sequence identity to photolyases, but lacking the DNA repair function of photolyases. Cryptochromes have a wide array of functions in various species. Plant and animal Cryptochromes have light dependent and independent roles in entraining their circadian rhythms. Cryptochrome has also been linked to being the magnetoreceptor in birds and insects, as seen in Cryptochrome knock out experiments in *Drosophila melanogaster*. It is thought that the mechanism for this requires an electron radical pair. Under a magnetic field, *E. coli* PL
yields a flavin-tryptophan radical pair, the basic requirement for the function of a magnetoreceptor\textsuperscript{117}. A more recently discovered class of the family is CRY-DASH (CRY\textsubscript{D})\textsuperscript{27}.

The sequence of these enzymes groups them into a number of subfamilies: animal CRY and (6-4) photolyase, class I CPD photolyase, plant CRY, and class II CPD photolyase, and CRY\textsubscript{D}. This analysis suggests that the cryptochrome ancestor evolved before the divergence of prokaryotes and eukaryotes\textsuperscript{27}. Sequence alignment of PL indicates that a gene duplication was followed by the functional divergence of the (6-4) photolyase\textsuperscript{118}.(Fig. 1.10)

Of the CRY/PL family, CRY\textsubscript{D} is arguably the most similar in structure to that of PL.(Fig. 1.10) CRY\textsubscript{D} from \textit{Synechosystis}, superimposed with PL from \textit{A. nidulans} has a 1.83 Å root mean square deviation, RMSD, about the $\alpha$ carbons, as determined in pymol.(Fig. 1.11) The largest differences occur in the recognition loop region\textsuperscript{27}. A homologous loop region in CRY\textsubscript{D} was also found from Y398 to I414\textsuperscript{119}.(Fig. 1.12)(Table 1.3 and 1.4) This loop does not overlay well with that of PL. In contrast, the other CRY proteins have an additional C-terminal domain that is lacking in PL and in CRY\textsubscript{D}. The tryptophan triad in PL is also conserved in CRY\textsubscript{D}, in Trp320, Trp373, and Trp396. In comparison, the PL has a tryptophan triad in the residues Trp306, Trp359, and Trp382\textsuperscript{76}. CRY\textsubscript{D} also contains a secondary cofactor, MTHF, similar to PL\textsuperscript{50}.

The physiological role of CRY\textsubscript{D} is currently not well defined\textsuperscript{27}. CRY\textsubscript{D} has been shown to be a transcriptional repressor of certain clock genes\textsuperscript{27}. CRY\textsubscript{D} was also shown to repair damage in ssDNA\textsuperscript{50}. However, this is not thought to be its main physiological function, as this would strongly limit repair capability to moments when the DNA is single-stranded. This function may be an evolutionary remnant. Pokorny proposed that the ssDNA repair by PL and CRY progenitors might have been relevant in an evolutionary role in repairing distorted DNA sequences during replication\textsuperscript{120}. The method used to detect this function was repair of ssDNA CPDs in the context of an endonuclease site. UV lesions prevent the endonuclease from cleaving the substrate, and repair by CRY\textsubscript{D} was detected by cleavage of this site. The authors suggest that since the chemical activity of CRY\textsubscript{D} was functional, the enzyme's ability to access the CPD in duplex must
be the region that is compromised\textsuperscript{50}. It is suggested that flipping the CPD is important in some way to the ability of PL to bind dsDNA containing dimers, as CRY\textsubscript{D} is only able to bind to structures with bases predisposed to be in the extrahelical position\textsuperscript{120}.

Figure 1.10. CRY\textsubscript{D} domains.

Note: Top, phylogenetic tree of CRY/PL members. Adapted from Brudler\textsuperscript{27}. Left, two main domains can be identified in *Synechocystis* sp. PCC 6803 CRY\textsubscript{D} (PDB: 1np7)\textsuperscript{27}. The N-terminal $\alpha/\beta$ domain includes $\alpha$-helices 1-5 and $\beta$ sheets 1-5. The C-terminal $\alpha$-domain includes $\alpha$-helices 8-12\textsuperscript{12}. $\alpha$14 and $\alpha$15, purple, and $\alpha$9, yellow, are shown for orientation. Right, the amino acid sequence with secondary structure elements denoted in red lines, $\beta$ denotes $\beta$-sheets, L denotes loops, and $\alpha$ denotes $\alpha$-helices.
Note: Structural superimposition of *A. nidulans* PL (grey, PDB: 1qnf) and *Synechosystis* CRYD (teal, PDB: 1np7), top. $\alpha 14$ and $\alpha 15$, purple, and $\alpha 9$, yellow, are shown for orientation for PL. $\alpha 14$ and $\alpha 15$, pink, and $\alpha 9$, orange, are shown for orientation of CRYD. Both these proteins share a similar overall fold, along with the same overall “bent” orientation in the FAD$^{12,119}$. Bottom, sequence alignment of the overlayed proteins relative to the PL secondary structure, right.
Figure 1.12. Hydrogen bond formation in the CRY\textsubscript{D} recognition loop.

Note: Hydrogen bonds in CRY\textsubscript{D} recognition loop with ssDNA free, top, from Synechocystis, and ssDNA bound, bottom, from A. thaliana (pdb: 1np7 and 2vtb, respectively)\textsuperscript{50,120}. \(\alpha\)14 and \(\alpha\)15, purple, and \(\alpha\)9, yellow, are shown for orientation. Inset is whole CRY\textsubscript{D} for orientation. Hydrogen bonding partners are shown in sticks, loop residues, green; non-loop residues, blue; DNA, purple sticks. \(\alpha\)14 and \(\alpha\)15, purple, are shown for orientation. Asterisks indicate residues that interact with DNA.
Table 1.3. Hydrogen bonds interactions of CRY$_D$ recognition loop without ssDNA.

<table>
<thead>
<tr>
<th>PL Numbering</th>
<th>Loop Residue</th>
<th>Loop Atom</th>
<th>Partner Residue</th>
<th>Partner Atom</th>
<th>Distance (Å)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ala 394</td>
<td>Ala 400</td>
<td>N</td>
<td>Asn 395</td>
<td>O</td>
<td>5.0</td>
</tr>
<tr>
<td>Ser 395</td>
<td>Gly 401</td>
<td>N</td>
<td>Trp 396</td>
<td>O</td>
<td>5.1</td>
</tr>
<tr>
<td>Ser 396</td>
<td>Ile 402</td>
<td>N</td>
<td>Asn 397</td>
<td>O</td>
<td>3.1</td>
</tr>
<tr>
<td>Ser 396</td>
<td>Ile 402</td>
<td>O</td>
<td>Leu 314</td>
<td>N</td>
<td>2.8</td>
</tr>
<tr>
<td>Arg 404</td>
<td>Arg 410</td>
<td>O</td>
<td>Lys 367</td>
<td>N (side chain)</td>
<td>3.1</td>
</tr>
<tr>
<td>Phe 406</td>
<td>Phe 412</td>
<td>N</td>
<td>Ser 363</td>
<td>O</td>
<td>3.0</td>
</tr>
<tr>
<td>Asn 407</td>
<td>Asn 413</td>
<td>N</td>
<td>Gln 417</td>
<td>O</td>
<td>3.7</td>
</tr>
<tr>
<td>Pro 408</td>
<td>Ile 414</td>
<td>N</td>
<td>Ser 418</td>
<td>O</td>
<td>3.3</td>
</tr>
</tbody>
</table>

Note: Hydrogen bonds from CRY$_D$ loop without ssDNA bound. Atoms are in polypeptide backbone unless otherwise noted. Based on *Synechosystis* CRY$_D$ crystal structure (pdb: 1np7)\textsuperscript{27}.  

35
Table 1.4. Hydrogen bonds interactions of CRY$_D$ recognition loop with ssDNA bound.

<table>
<thead>
<tr>
<th>PL Numbering</th>
<th>Loop Residue</th>
<th>Loop Atom</th>
<th>Partner Residue</th>
<th>Partner Atom</th>
<th>Distance (Å)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ala 394</td>
<td>Ala 436</td>
<td>O</td>
<td>Gly 437</td>
<td>N</td>
<td>3.4</td>
</tr>
<tr>
<td>Ser 395</td>
<td>Gly 437</td>
<td>O</td>
<td>Arg 350</td>
<td>N (side chain)</td>
<td>3.0</td>
</tr>
<tr>
<td>Gly 397</td>
<td>Gly 439</td>
<td>N</td>
<td>Thr 433</td>
<td>O</td>
<td>3.2</td>
</tr>
<tr>
<td>Pro 402</td>
<td>Glu 444</td>
<td>O (side chain)</td>
<td>DT 1</td>
<td>N (base)</td>
<td>3.3</td>
</tr>
<tr>
<td>Pro 402</td>
<td>Glu 444</td>
<td>O (side chain)</td>
<td>DT 4</td>
<td>N (base)</td>
<td>3.4</td>
</tr>
<tr>
<td>Leu 403</td>
<td>Asp 445</td>
<td>O (side chain)</td>
<td>Arg 403</td>
<td>N (side chain)</td>
<td>3.0</td>
</tr>
<tr>
<td>Arg 404</td>
<td>Arg 446</td>
<td>O</td>
<td>Arg 403</td>
<td>N (side chain)</td>
<td>4.1</td>
</tr>
<tr>
<td>Arg 404</td>
<td>Arg 446</td>
<td>N (side chain)</td>
<td>DT 1</td>
<td>O (phosphate)</td>
<td>4.9</td>
</tr>
<tr>
<td>Asn 407</td>
<td>Ser 449</td>
<td>N</td>
<td>DT5</td>
<td>O (phosphate)</td>
<td>2.9</td>
</tr>
<tr>
<td>Asn 407</td>
<td>Ser 449</td>
<td>O</td>
<td>Lys 452</td>
<td>N</td>
<td>3.2</td>
</tr>
<tr>
<td>Asn 407</td>
<td>Ser 449</td>
<td>O (side chain)</td>
<td>Ile 450</td>
<td>N</td>
<td>4.4</td>
</tr>
</tbody>
</table>

Note: Hydrogen bonds from CRY$_D$ loop with ssDNA bound. Atoms are in polypeptide backbone unless otherwise noted. Based on A. thaliana CRY$_D$ crystal structure (pdb: 2VTB).$^{120}$

In a cocrystal of CRY$_D$ with ssDNA, the electrostatics of the binding cleft of CRY$_D$ has a net positive charge$^{27,120}$. This is similar to that of PL, as the cavity at which the dsDNA binds is also positively changed. It was suggested that there exists little difference here for binding DNA.$^{119}$

Crystallographic evidence suggests that the CRY$_D$ loop is reorganized, less so than PL, to accommodate CPD in ssDNA$^{119,120}$. The loop in CRY$_D$ moves by 4.6 Å at the maximum, with R446 (R404 in PL numbering), which could help stabilize the structure through a salt bridge with the phosphate between the thymine pair.$^{120}$ This loop also moves up to 5 Å upon binding to CPD in ssDNA, compared to the 10 Å movement of PL binding to dsDNA.$^{10,120}$
1.5. The recognition loop is a promising region for studying PL-dsDNA binding

The recognition loop region of PL moves 10 Å to relieve steric hindrance between it and the CPD and to make interactions with the complementary strand of the CPD. It is the main observed mobile element upon substrate binding\textsuperscript{10}. This loop has a homologous region in CRY\textsubscript{D} which is also displaced upon binding to CPD in ssDNA, though to a lesser degree of 5 Å\textsuperscript{120}. It is possible that the differences in the displacement of this loop region could account for the differences in the ability for PL and CRY\textsubscript{D} to repair dsDNA.

I suggest that the PL recognition loop has two states: A productive state that allows CPD to access the PL active site, and an unproductive state that sterically hinders CPD from accessing the PL active site. The ability for PL to repair would depend on the number of productive states available. The number of productive and unproductive states the recognition loop can form, and the interconversion between them, are the recognition loop conformational dynamics. I hypothesize that the PL recognition loop has optimized conformational dynamics that allows CPD to access the active site of productive PL, and these dynamics are different in CRY\textsubscript{D}.(Fig. 1.13)

Figure 1.13. Model for hypothesis.
Note: U\textsubscript{1} onwards represent unproductive states in which PL cannot form an active complex with substrate and E\textsubscript{1} onwards represent productive states which are not prohibited from forming the active complex with substrate.

The term optimized refers to a range of loop dynamics which are not too large and not too small. Some degree of loop dynamics is required for the PL loop to reorganize itself during DNA binding. In addition, the loop is required to be not so dynamic as to move during the binding of CPD.
The term "too dynamic" refers to heightened loop dynamics, greater than optimal. This loop region must be able to sample a number of different states, including the initial position observed in the unbound crystal structure and the final position with bound DNA along with any intermediate positions between the two. Increasing the conformational space will increase the number of total states and likely increase the number of unproductive states. This increase in unproductive states should then shift the equilibrium towards the unproductive states. Therefore, this would decrease DNA binding.

The term "some dynamics are required" means the loop of PL remains dynamic enough to allow for reordering during CPD binding. In other words, if the recognition loop were too rigid, it would be unable to move out of the way and sterically interfere with the incoming CPD lesion.

During work in my thesis to explore conformational dynamics of PL (see Results Chapter 2), I found the recognition loop to be an important site with higher dynamics than the rest of the protein but lower than that of the CRY_D recognition loop. I have also devised an experimental strategy (See Chapter 3) for testing the relationship between loop dynamics and enzyme function.

It has been hypothesized that the differences between the ability of PL and CRY_D are due to differences in the electrostatics of the binding cleft. The hypothesis is that the CRY_D binding pocket has an overall lower hydrophobicity relative to PL, which prevents it from expending the 5-7 kcal/mol\(^{103}\) required for the movement of the CPD to an extrahelical position\(^{119}\). If this hypothesis is true, then I would expect a decrease in the hydrophobicity of this cleft to reduce binding affinity of CPD and, subsequently, repair rate. I would also expect that mutations in other regions, such as the recognition loop, would not change CPD binding.

Of these residues in the binding cleft, Huang\(^{119}\) predicted three to be important in PL: Met353, Trp286, and Trp392.(See Fig. 1.6) This prediction was based on a comparison of the PL crystal structure to that of CRY_D. He also suggested that the differences of DNA repair activity between CRY_D and PL are not related to DNA binding, but related to making the CPD extrahelical\(^{120}\).
I predict that if the latter hypothesis were true, changes in the dynamics of the recognition loop would not change the ability for PL to bind CPD. I suggest that the functional difference between CRY$_D$ and PL is rooted in the inability for CRY$_D$ to stably bind CPD in dsDNA. Specifically, I suggest that the recognition loop in PL has optimized conformational dynamics, different from CRY$_D$. If this is true, I would predict that changes in the dynamics of the recognition loop would impact the ability for PL and CRY$_D$ to bind CPD.

1.6. Multiple assays can be used to probe conformational dynamics

1.6.1. Detecting conformationally dynamic regions using limited proteolysis and mass spectrometry

Probing the recognition loop of PL is somewhat challenging in that there needs to be a property of this region that is significantly different from the rest of the protein to probe in the context of the full protein. One promising assay for this purpose is limited proteolysis$^{121}$. Limited proteolysis in conjunction with mass spectrometry can be used to probe the conformational dynamics of flexible regions in the protein and map these sites to the location of cleavage. (Fig. 1.14)
Figure 1.14. Scheme for identifying a conformationally dynamic region of protein.

Limited proteolysis utilizes the predisposition of proteases to preferentially cleave more conformationally dynamic regions of a protein$^{121,122}$. Given small concentrations of protease and limiting time constraints, it would be expected that cleavage of whole native protein would result in one or fewer cleavage events. Only the most conformationally dynamic regions would be cleaved. The regions of the protein that are able to sample a larger conformational space than others are more likely to achieve the correct conformation to enter the protease active site, and as a result, are more likely to be cleaved. This cleavage can be monitored by SDS-PAGE.

The model proposed for the direct link between limited proteolysis rates and cleavage site conformational dynamics is one where an induced fit mechanism is required for the stretch of 8 to 10 amino acids from the globular protein to fit into the protease active site for hydrolysis. Under this model, the rigid native structure cannot act as a substrate without reorganization, which is consistent with the difference between rates of proteolysis in folded compared to unfolded proteins. Results from limited proteolysis correlates with conformational dynamic information derived from other spectroscopic techniques$^{121}$.
To identify the amino acid sequences of the resulting PL fragments generated from limited proteolysis, mass spectrometry can be employed. Matrix-assisted laser desorption/ionization-time of flight (MALDI-TOF) mass spectrometry ionizes the protein fragments of interest from a matrix, and then monitors the results by mass spectrometry. First, a laser is used to desorb the protein from the matrix. The desorbed particles are then ionized and sent to a detector, which determines the molecular weight of the products based on the time it takes for the desorbed particle to reach the detector. To probe the peptide fragments further, a trypsin digestion can be performed prior to the MALDI-TOF. This will produce fragments of known sizes, which can be identified by comparing these fragments to the expected molecular weight of the predicted fragments.

1.6.2. End-to-end contact of peptides is a measure of conformational dynamics

Further analysis of conformational dynamics of specific regions of proteins can be performed through the use of a peptide, containing the sequence of interest, flanked by a fluorophore and a quencher. The rate at which the fluorophore becomes quenched, as measured by fluorescence spectroscopy, will positively correlate to the end-to-end contact between the fluorophore and the quencher, which is a direct gauge of conformational dynamics of the region of interest. The dynamics measured using this method are due to the intrinsic sequence and not due to other interactions made through the other regions of the protein. Further, to ensure secondary structures are not involved in the measurement, denaturant can be added in this experiment.

Fluorophores are excited by a photon of light and have an electron change from a resting to an excited state. This excited state is thermodynamically unfavourable and the electron will then spontaneously return to the ground state. The energy from this will be released as light energy of a longer wavelength. The rate at which the fluorophore returns to ground state and releases this light energy depends on several factors including the fluorophore itself and the surrounding environment. This rate of decay can be quantified as a lifetime, \( \tau \), which is the time constant associated with the exponential
decay. Quenching of this process increases the rate of decay and therefore decreases the lifetime of the fluorophore. (Fig. 1.15)

The rate of decay of fluorophores is often in the picosecond to nanosecond timescales. This makes monitoring the decay of fluorophores challenging. One method to measure fluorophore decay is Time-correlated single photon counting (TC-SPC). This technique quickly pulses a laser through the sample with low energy to excite one or fewer fluorophores in the sample. The time at which the fluorophore emits a photon and returns to its ground state is measured. This measurement then increases the count in a bin that encompasses that time frame. This procedure is then repeated and a profile that is equivalent to the fluorescence decay profile of the fluorophore is generated. A long decay will have proportionally more instances placed into bins at further away from the excitation time, as it will, on average, take longer for the fluorophore to emit and return to ground state. A short decay will proportionally have more of these instances placed into bins closer to the excitation time, since the fluorophore will, on average return to ground state sooner. It is important to excite one or fewer fluorophores at a time so as to not mix the signal from multiple excited molecules.
The peptide of interest has a fluorophore at one end and a quencher at the other end. Upon excitation, the fluorophore will emit light until it comes into quenching range of the quencher. Under these conditions, the fluorophore will not emit light\textsuperscript{126}.

The intramolecular quenching rate for small peptides containing fluorophores and quenchers at opposite ends is, by definition, equal to the rate of folding of these peptides\textsuperscript{126}. The rate of observed decay, $k_o$, is the sum of all the types of decays. In a situation with no intermolecular quenching, the rate of intramolecular quenching can be calculated as follows:

$$k_o = k_{em} + k_s + k_{fold}$$  \hspace{1cm} (2)

Where $k_{em}$ is the typically slow, intrinsic decay emission of the fluorophore. $k_s$ is the decay of the fluorophore due to solvent. $k_{fold}$ is the decay of the fluorophore from end-to-end contact, which equals the rate of quenching, $k_q$. This equation is a form of a chemical branching reaction where the reagent can form multiple products (decay by different methods).

The value of $k_{fold}$ can be calculated from the fluorescence lifetimes of the peptide with and without the quencher, DA and D, related by the following equations\textsuperscript{126}:

$$k_{fold} = 1/\tau_{DA} - 1/\tau_D$$  \hspace{1cm} (3)

$$1/\tau_D = k_{em} + k_s$$  \hspace{1cm} (4)

$$1/\tau_{DA} = k_o$$  \hspace{1cm} (5)

Here, it is assumed that $\tau_D$ is not strongly dependent on amino acid composition. It is reasonable to assume this, as varying the composition of this type of peptide does not have a strong influence on the lifetimes of the fluorophore\textsuperscript{127}.

For flexible peptides, it would be expected that the fluorophore and quencher will come into contact more frequently. Therefore, a faster decay would be expected with a shorter lifetime. For more rigid peptides of the same length, it would be expected that the
fluorophore and the quencher will come into contact less frequently. A slower decay would be expected here with a longer lifetime.

Average end-to-end distance can be measured using Förster resonance energy transfer (FRET) energy transfer for short peptides with a FRET donor and acceptor pair. This method has been verified by molecular dynamics modeling and gaussian chain modelling\textsuperscript{125}. To calculate end-to-end distances, the difference in FRET efficiencies of Time-resolved measurement FRET efficiencies, $E_{tr}$, must be measured for the peptide containing the end FRET donor and acceptor as well as the peptide containing the end FRET donor only. One way to do this is using fluorescence lifetimes ($\tau_{DA}$ and $\tau_D$) in equation (6)\textsuperscript{125}.

$$E_{tr} = 1 - (\tau_{DA}/\tau_D) \quad (6)$$

The donor-acceptor distance, $R$, can then be calculated using equation (7)\textsuperscript{124}. The Förster radius, $R_o$, of the fluorophore must be known.

$$R^6 = R_o^6 (1-E_{tr})/E_{tr} \quad (7)$$

Eyring plots can be used to quantitatively measure activation paramaters of peptide folding by measuring the rate of folding at different temperatures. The Eyring plot can be described in this case as the $k_{fold}$ divided by the temperature, $T$, which gives the activation enthalpy ($\Delta H^\ddagger$) and activation entropy ($\Delta S^\ddagger$) as per the following equation:

$$\ln(k_{fold}/T) = -(\Delta H^\ddagger/R)(1/T) + \ln(k_B/h) + \Delta S^\ddagger/R \quad (8)$$

Where R is the gas constant, $k_B$ is the Boltzmann constant, and $h$ is Planck’s constant.
1.7. Multiple functional assays can be used to detect repair capability

1.7.1. Functional assays chosen for this thesis investigate PL repair capability in different ways

There are a number of ways to monitor function in PL. These include monitoring repair by DNA absorbance of UV light, detecting thymine dimers using a restriction-site protection assay, or using a bacterial survival assay. The UV light absorbance assay allows for monitoring of substrate in real time in a simplistic way. Monitoring thymine dimers by using a restriction-site protection assay allows for a sensitive quantitation of the reaction progress. Measuring thymine dimers by the bacterial survival assay allows for monitoring CPD repair in vivo.

The bacterial survival assay allows for detection of UV damaged DNA in vivo. An E. coli strain devoid of DNA repair function, UNC1085, was initially generated for this specific assay. The genotype of UNC1085 is: F-, pro-82, glnX44(AS), phr-1, ΔnadA71, λ-, Δ(recA-srl)306, relA1?, endA1, rpsL286, thiE1, uvrA281, hsdR17. These cells are light sensitive and die when exposed to UVC light. The CPD generation is monitored by decrease in cell survivability. These cells can be transformed with a plasmid containing the PL gene. If this gene is expressed, these cells have a much higher probability of surviving UVC treatment after being exposed to photoreactivating light.

This assay has been performed to test the relative repair rate of PL in E. coli, and how mutations affect PL in these cells. This assay has been used quantitatively by plotting cell survival and fluence of UVC to generate an exponential death time course. This assay does, however, have limitations. This assay is unable to monitor directly how many UV lesions are generated or what types of lesions are generated. It also is not capable of detecting differences in enzyme concentrations between different constructs, which may skew results and can only be addressed by other assays. Published data using this assay requires some form of in vitro data for verification or a Western blot to control for enzyme concentrations.
A *tac* promoter induction system for the enzyme would allow for controlled expression of PL in cells to allow for consistent amount of enzyme to be expressed in each experiment to allow for comparison across data sets. A Western blot of the lysate of these cells would allow for measurement of the enzyme concentrations in each cell. The enzyme concentrations could be adjusted using the *tac* promoter, given the data from the Western blots.

One simple way to detect thymine dimer formation and repair is by monitoring the UV absorbance of DNA. Upon thymine dimer formation between two thymines, the typical absorbance at 260 nm is lost\(^{60,131}\). Upon repair of the dimer, the lost absorbance is recovered. This assay is relatively simple and has the advantage of being monitored in real time. This assay is limited by the need for poly deoxythymidine substrates to allow for the detection of the change in absorbance. The larger amount of thymines is required for a significant change in the UV absorbance, and thus a detectable signal.

This assay was used to determine the effects of the DNA backbone on repair of dimers by PL\(^{55}\). The substrates used in this study was a combination of dinucleotides or oligo (dT)\(_{12-18}\), along with oligo (dC)\(_{12-18}\) and poly(rU) of close to 200 bases. PL had the highest quantum yields when repairing thymine dimers, and less so for uracil and cytosine dimers\(^{55}\).

A more quantitative assay is the restriction-site protection assay. This assay relies on the inability of endonucleases to cleave CPD\(^{50}\). Historically, *MseI* and *VspI* have been used as the endonuclease sites\(^{50,120}\). Here, a thymine pair is engineered at the cleavage site or in the recognition site of an endonuclease. If a CPD is formed from the thymine pair, the DNA will become resistant to endonuclease cleavage. This can then be resolved on a gel to determine the amount of CPD contained in the DNA quantitatively. In the case of wild type PL, cleavage of the DNA occurs after photoreactivating light is given. However, in the case of CRY\(_D\), DNA does not cleave after these same conditions\(^{50}\). Alternatively, an endonuclease which cleaves CPD specifically can also be used in a similar way. However, in this case, cleaved DNA represents the number of lesions still present\(^{15}\).
Chapter 2.

The recognition loop of CRY_D is more flexible than that of PL


The comparison of the PL crystal structures with and without bound substrate has revealed the mobile nature of the recognition loop, which moves to prevent steric hindrance with the substrate and forms interactions with the complementary DNA strand. While CRY_D appeared to have a comparable loop region, it wasn’t shown until 2008 that this loop also changes position upon binding to ssDNA substrate. My experiments have shown that this loop region is the most conformationally dynamic region of each of these proteins. I have also shown that the heightened conformational dynamic nature of these regions is, at least in part, due to the sequences of these regions.

Contributions of co-authors were as follows: Neahlanna Mcleod contributed the initial generation of the limited proteolysis experimental scheme for PL and CRY_D. She also initially tested the proteases to determine the proteases to be used. Michael Damiani helped with the purification of PL and CRY_D. He was also involved with consultations regarding handling of these proteins. Lisa Wang contributed in the development of the fluorescent peptide design and experimental scheme. She also began the initial fluorescence testing of the fluorescent peptides. Melanie O’Neill
generated the experimental schemes for limited proteolysis for PL and CRY<sub>D</sub>. She also generated the experimental schemes for the fluorescent peptide experiments. Melanie also provided a significant amount of help and advice.

2.1. PL and CRY<sub>D</sub> have distinct recognition loops which may be conformationally dynamic

While the study of PL binding has been undertaken in the past, the mechanism by which PL gains access to the CPD is still debated. Two possible models could exist. One model for binding is where the CPD exists in an extrahelical position before binding to PL. The competing model is where the CPD enters the extrahelical position after binding by PL (see section 1.3). While the recognition loop region of PL has been noted and suggested to have importance, specifically in its conformational change upon CPD binding and its interactions with the complementary strand of the CPD, its conformational dynamics have not been previously probed.

The recognition loop region changes conformation the most dramatically from before and after CPD is bound. I suggest that this region is important to CPD binding. I considered that the ability of the recognition loop to move is the property that is important for the access of CPD.

An initial method of monitoring differences between the PL and CRY<sub>D</sub> recognition loop conformational properties was to look at the B-factors in their respective crystal structures. A B-factor is the approximation of the displacement of the atomic position from a mean value. The differences between the position of the atom and the given coordinates could be due to thermal motion or due to imperfections in the crystal. Since the recognition loop of PL is the region that moves the most between the substrate bound and unbound states, initial B-factor analysis of this region was performed to identify differences between PL and CRY<sub>D</sub>.

The B-factors for the PL recognition loop were analyzed by comparing the individual B-factors of the C<sub>a</sub> of the amino acids in this loop. The average B-factor of the PL recognition loop (residues 393 – 407) is 44.7 Å<sup>2</sup> (Fig. 2.1) This is among the highest
in PL, which has an average B-factor for the Cα of 27.8 Å². The B-factors in this loop range from 29.2 Å² to 60.4 Å², with the highest B-factors closer to the C-terminus of the loop (PL residue number 392 – 400) and the lowest ones closer to the N-terminal end of the loop (PL residue number 401-410).

**Figure 2.1. B-factors of alpha carbons of loop residues for PL and CRYD.**

Note: PL (PDB: 1qnf) and CRYD (PDB: 1np7) alpha carbon B factors are plotted, aligned to PL numbering, to show the overall differences in certainty of position of these residues as an initial measure of differences in flexibility.12,119

The B-factors of the Cα of the amino acids in the CRYD recognition loop (CRYD residues 399 – 413) averaged 35 Å². (Fig. 2.1) This is also among the highest B-factors in CRYD, which has an average B-factor for the Cα of 23.08 Å². The B-factors in this loop range from 26.4 Å² to 53.7 Å², with the highest B-factors closer to the middle of the loop. (Fig. 2.1)

The B-factors in the recognition loops of PL and CRYD are among the highest in each of their respective structures. However, they are dissimilar in their profiles of B-factor distribution in that the B-factors of the PL loop are higher closer to the C-terminal of the loop, while that of CRYD is highest closer to the middle of the loop. This difference
in B-factors could indicate differences in the thermal motion of select regions of these proteins potentially caused by dissimilar loop dynamics.

Based on these B-factor values, I would predict that these loops are the most conformationally dynamic regions of these proteins. I would conclude that PL and CRY$_D$ have different conformational dynamics in their loop regions.

The hydrogen bonds made by PL are distributed among most of the loop residues when PL is without substrate. When CPD is bound, more bonds are formed closer to the middle of the loop and fewer on the N-terminal region. In CRY$_D$ without substrate, hydrogen bonds are distributed closer to the ends of the loop region. But, similarly to PL, CRY$_D$ has a shift in bond formation to the middle of the loop away from the N-terminal when CPD from ssDNA is bound.

2.2. Limited proteolysis identifies the recognition loop as the most conformationally dynamic region of PL and CRY$_D$

2.2.1. ProteinaseK and Chymotrypsin produce multiple PL and CRY$_D$ fragments

I next employed limited proteolysis to assay the conformational dynamics of this loop region. I hypothesized that the recognition loop regions of both PL and CRY$_D$ would be preferentially cleaved, relative to the rest of the protein, by proteases.

The proteases I chose were Chymotrypsin (Chy) and Proteinase K (PK)$^{132,133}$. A series of trials with varying concentrations of protease was performed to determine an appropriate level of cleavage for the desired timeframe. While PK is non-specific in the sequence it cleaves, Chy cleaves preferentially after a large hydrophobic amino acid, for example, phenylalanine, tryptophan, or tyrosine.
Figure 2.2. Limited proteolysis of PL.

Note: 7% SDS-PAGE, stained with Sypro Ruby protein gel stain, of limited proteolysis of a ratio 2000:1 PL:PK over 77 h, upper panel. A total of 20 µL was loaded for each sample. The retardation factor (rf) was calculated for bands produced by the ladder and plotted, lower panel. Molecular masses were assigned to bands relative to the molecular mass of the molecular mass standard.
Figure 2.3. Limited proteolysis of CRY<sub>D</sub>.  
Note: 7% SDS-PAGE, stained with Sypro Ruby protein gel stain, of limited proteolysis using a 2000:1 ratio of CRY<sub>D</sub>:PK over 20 h, upper panel. A total of 20 µL was loaded for each sample. The retardation factor (rf) was calculated for bands produced by the ladder and plotted, lower panel. Molecular masses were assigned to bands relative to the molecular mass of the molecular mass standard.
2.2.2. **Preferential cleavage products were observed in PL and CRY\textsubscript{D} after limited proteolysis with Chy or PK**

If the protease cleaves in the recognition loop of PL, I would expect to see fragments around 42 kDa and 13 kDa in size. Since the molecular mass of PL is 54.5 kDa, and the loop is situated on the C-terminal side of PL. If the protease cleaves in the recognition loop of CRY\textsubscript{D}, I would expect to see fragments around 45 kDa and 13 kDa in size. CRY\textsubscript{D} is 58 kDa, and the loop is also situated on the C-terminal side of this protein.

Uncleaved PL and CRY\textsubscript{D} show migration on SDS-PAGE consistent with their calculated molecular weights of 54.5 kDa and 58 kDa. Using Chy or PK, both PL and CRY\textsubscript{D} were cleaved within a 2 minute timeframe. PL digested with Chy yielded only one visible fragment of 45 kDa. Cleavage of PL with PK yielded a total of 4 main bands, two that migrate at 45.5 and 42.3 kDa, and two lower ones that migrate around 29 and 27.9, named pk1, pk2, pk3, and pk4, respectively. (Fig. 2.2 and Table 2.1) The two lowest migrating bands are likely due to secondary proteolysis of the first generated fragments. CRY\textsubscript{D} digested with PK yielded a fragment of 45 kDa, along with three smaller fragments of 29 kDa, 28 kDa, and 25 kDa, named ck1, ck2, ck3, and ck4, respectively. The three lowest bands are likely due to secondary proteolysis of the first generated fragment. CRY\textsubscript{D} digested with Chy produced a 48 kDa fragment. (Table 2.1)
### Table 2.1. Limited proteolysis fragment identification.

<table>
<thead>
<tr>
<th>Name</th>
<th>Parent</th>
<th>Enzyme</th>
<th>Molecular Weight (kDa)</th>
<th>Fragment</th>
<th>Identification Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>PL parent</td>
<td>PL</td>
<td>None</td>
<td>53.2</td>
<td>1-484</td>
<td>Gel Electrophoresis</td>
</tr>
<tr>
<td>Pk1</td>
<td>PL</td>
<td>PK</td>
<td>45.5</td>
<td>(94-95)-484</td>
<td>Mass Spectrometry</td>
</tr>
<tr>
<td>Pk2</td>
<td>PL</td>
<td>PK</td>
<td>42.3</td>
<td>1-(402-404)</td>
<td>Mass Spectrometry</td>
</tr>
<tr>
<td>Pk3</td>
<td>PL</td>
<td>PK</td>
<td>29.0</td>
<td>(94-95)-(402-404)</td>
<td>Hypothesized</td>
</tr>
<tr>
<td>Pk4</td>
<td>PL</td>
<td>PK</td>
<td>27.9</td>
<td>105-(402-404)</td>
<td>Hypothesized</td>
</tr>
<tr>
<td>PL-chy</td>
<td>PL</td>
<td>Chy</td>
<td>45.0</td>
<td>(98-99)-484</td>
<td>Mass Spectrometry</td>
</tr>
<tr>
<td>CRY₀ parent</td>
<td>CRY₀</td>
<td>None</td>
<td>56.3</td>
<td>1-489</td>
<td>Gel Electrophoresis</td>
</tr>
<tr>
<td>Ck1</td>
<td>CRY₀</td>
<td>PK</td>
<td>44.5</td>
<td>1-(401-410)</td>
<td>Mass Spectrometry</td>
</tr>
<tr>
<td>Ck2</td>
<td>CRY₀</td>
<td>PK</td>
<td>29.2</td>
<td>95-(401-410)</td>
<td>Hypothesized</td>
</tr>
<tr>
<td>Ck3</td>
<td>CRY₀</td>
<td>PK</td>
<td>27.5</td>
<td>110-(401-410)</td>
<td>Hypothesized</td>
</tr>
<tr>
<td>Ck4</td>
<td>CRY₀</td>
<td>PK</td>
<td>24.8</td>
<td>140-(401-410)</td>
<td>Hypothesized</td>
</tr>
<tr>
<td>Cry-chy</td>
<td>CRY₀</td>
<td>Chy</td>
<td>48.0</td>
<td>(80-81)-489</td>
<td>Mass Spectrometry</td>
</tr>
</tbody>
</table>

Note: Identification method by gel electrophoresis or mass spectrometry is identified experimentally. Identification method that is hypothesized is based on possible cleavage sites to generate the observed fragment molecular weight.

These results are consistent with my expectations of a single cleavage site being produced. Additional bands for both PL and CRY₀ fall below the anticipated 45 kDa fragment. These bands are likely due to secondary proteolysis of the 45 kDa product. However, it is possible that they could be produced from primary cleavage reactions. Based on these results, I would conclude that the most flexible region of the proteins could be in the loop region.

#### 2.2.3. PL and CRY₀ have homologous proteolytic sites

The molecular weight difference between the product bands of PL-chy and Cry-chy are ~3 kDa apart in molecular weight. This is the same difference between the molecular weight of uncleaved PL and uncleaved CRY₀. This suggests that there is a homologous site of proteolysis between these two proteins for Chy. Similarly, pk2 and
ck1 are ~3kDa apart in molecular weight. This is suggestive of a homologous site of proteolysis between these two proteins.

The PK experiment does not indicate whether the 42 kDa, pk2, or the 45 kDa, ck1 fragment contains the N-terminus or C-terminus of PL or CRYD, respectively. However, if it does contain the N-terminus, then the cleavage site would be approximately at position 400 in either protein, which is consistent with cleavage of the recognition loop region. This was tested by LC/MS/MS.

Samples tested were generated from single excised gel bands of fragments treated with trypsin. Both PK and Chy fragments were tested in a similar way. The weight of the cleaved fragments was measured and assigned to different theoretical trypsin cleaved peptides from PL.(Fig. 2.4) The differences between assigned cleaved peptides in whole PL and truncated pk2 and ck1 fragments was used to determine which end of the protein was missing.
Figure 2.4. Mass spectrometry alignment.

Note: Alignment of sequences from mass spectrometry for PL and CRY\textsubscript{D}. Light blue indicates regions of positive matches. Both PL and CRY\textsubscript{D} fragments cleave in the loop region, as only the N-terminal regions are positively matched.

The cleavage site was mapped to between residues 402 and 404 for PL for pk2. A similar procedure was performed for CRY\textsubscript{D} and the cleavage site was mapped to be between residues 401 to 410, for ck1. The Chy cleavage sites for both PL and CRY\textsubscript{D} occurred in the N-terminal region at residues 98-99 and 80-81, respectively. Pk1 also was mapped to a N-terminal cleavage site at position 94-95. (Table 2.1)

This uncertainty is generated through the lack of coverage due to the trypsin cleavage sites. As there is coverage up to residue 402 and after 404 in parent PL, and there is only coverage up to residue 402 in pk2, it can be concluded that the cleavage took place after residue 402, but not after residue 404.

The uncertainty is greater in CRY\textsubscript{D} than PL in this case due to a lack of coverage of the mass spectrometry of the whole CRY\textsubscript{D} protein. The whole parent CRY\textsubscript{D} was mapped up to residue 401 and after 410. As a result, since the ck1 fragment was
mapped up until residue 401, it can be concluded that the cleavage took place after residue 401, but not after residue 410. (Fig. 2.4)

From the location of the cleavage site in PL and CRY\textsubscript{D}, I conclude that the cleavage site of pk2 and ck1 occurs within the recognition loop. While the data carries some uncertainty, it is within the error range to make this conclusion. The uncertainty of the cleavage site extends within the entirety of the loop region for PL, positions 393-408, and CRY\textsubscript{D}, positions 399-414.

2.3. The recognition loop of PL is less conformationally dynamic than that of CRY\textsubscript{D}

2.3.1. I would expect the CRY\textsubscript{D} recognition loop to be cleaved more readily than that of PL

I hypothesized that the recognition loop of PL is less conformationally dynamic than that of CRY\textsubscript{D}. As a result, I would expect the PL loop to be cleaved by protease at a slower rate than that of CRY\textsubscript{D}.

To probe these proteins, limited proteolysis was performed as in section 2.2 but following cleavage over time. The primary cleavage rate was monitored by gel electrophoresis and densitometry.

PK was used for this assay for a number of reasons. First, it was known that it would cleave within the loop region of PL and CRY\textsubscript{D}. It also lacks sequence dependent recognition sites for its cleavage reaction, making it less susceptible to skewing the result based on amino acid sequence.

Under limited proteolysis conditions, PL degrades slower than CRY\textsubscript{D}. Under the same low E:S ratio conditions, of 1:750, whole PL persisted for more than 80 h. (Fig. 2.2) while all of whole CRY\textsubscript{D} degraded in less than 20 h. (Fig. 2.3) The main products produced from these reactions are pk2 and ck1. As mentioned earlier (see section 2.2.2), this is an N-terminal product whereby the recognition loop has been cleaved. PK2
and CK1 concentrations initially increase during the first several hours, but subsequently decay, presumably due to subsequent proteolytic cleavage.

The decay of whole protein due to proteolysis does not obey first order kinetics, as expected. The decay is initially quick, then slows, which is apparent in both PL and CRY_D, but is more severe in PL. The initial decay for both proteins is complete within 30 minutes, which accounts for about 40% of the total reaction in both PL and CRY_D. The time course for this fast component was not quantified, as the standard error of the initial points in this component is significantly higher than the other regions and might not be as informative.

Each data point was collected in triplicate. The individual runs were very similar overall, but the initial component, less than 30 minutes was significantly more variable. The second part of the decay can be fit to a first order exponential decay. Rate constants were determined for each run then averaged.(Fig. 2.5) The average rate constants were (6 ± 1) x 10^{-5} s^{-1} and (5 ± 1) x 10^{-6} s^{-1} for CRY_D and PL, respectively.
Figure 2.5. Fitted plots of PL and CRY\textsubscript{D} limited proteolysis (1:750 E:S ratio).

Note: Lines are fit to first order exponential decays for times less than 0.5 h and linear for times greater than 1 h. Inset is the same plot on a linear scale; fits are to a double first-order exponential decay.

A possible explanation of the biphasic decay is that there is interchange between a cleavable, “open,” state and an uncleavable “closed” state in the loop. Addition of protease could allow for rapid degradation of substrate in the “open” state. The following first order reaction would then be reflective of an interchange between the “closed” and “open” states. I propose that the first order reaction progress after 30 minutes is reflective of the rate at which these regions change their conformation.

A second, less likely, possible explanation is that the products of proteolysis degradation are inhibiting the protease though interaction with the protease itself or through interactions with the native PL or CRY\textsubscript{D}. Inhibition of protease by products would lead to biphasic decay, as the amount of product increases as the reaction proceeds, which would cause a decrease in the rate. Since this biphasic decay is more pronounced
for PL, it is possible that certain cleavage products that exist in PL, but not as abundant in CRY$_D$, are the cause of the inhibition. However, the cleavage products also are degraded by PK, making it unlikely that this would be the main cause of the biphasic decay.

Based on the observations, I conclude that the recognition loop of PL is less conformationally dynamic than that of CRY$_D$. The decay was approximately 10-fold slower for PL than for CRY$_D$.

Under native conditions, PL is cleaved slower than CRY$_D$ in the presence of PK. Normally, slow or specific cleavage rates of proteinases are associated with rigid structures in their substrates.$^{121,122}$ Proteolysis occurs only at specific sites in PL and CRY$_D$, implicating these proteins in having differences in structures or conformational dynamics.

It is possible that the cleavage of CRY$_D$ and PL is due to partial denaturation of certain regions of these proteins, followed by subsequent proteolysis. In this case, the increased availability of reactive conformations in CRY$_D$ might reflect a localized folding instability in one region, or an overall lower global folding stability. This is a property that can be determined by probing the folding stability of these proteins, for example by using circular dichroism (CD) spectroscopy.

The results from limited proteolysis, however, are not consistent with global unfolding of either of these proteins. It would be expected that in a reaction with PL or CRY$_D$ and PK, many fragments would be generated if global unfolding of the substrates occurred instead of specific primary cleavage sites. PL and CRY$_D$ were incubated with BSA to ensure that loss of parent is due to cleavage, and not protein precipitation. Incubation of PL and CRY$_D$ with BSA for the same duration showed little loss of soluble protein, less than 10%, as seen by UV-vis spectroscopy and electrophoresis.

It is also possible that differences in the ability of PK to access the recognition loop sites could contribute to the differences observed in cleavage rates. This seems unlikely, as PL and CRY$_D$ share more than 90% homology in secondary structure with a 1.83 Å RMSD in their tertiary structures about their $\alpha$ carbons. These proteins also share
60% identical or closely related residues. The accessible surface area for PL is 20,454 Å² compared to 21,293 Å² for CRYD, which are also very similar.

The two main possibilities that exist for the differences in PL and CRYD cleavage rates are that 1) greater access to the CRYD recognition loop conformation by PK allow for increased cleavage or 2) the PL loop is chemically more challenging for PK to cleave. I hypothesized that the loop region of PL is not significantly more challenging for PK to cleave than CRYD.

To test this, the limited proteolysis experiment was repeated under mild denaturing conditions using 0.5% SDS. If the first case was true, then I would expect less differences in cleavage rate of partially or fully denatured PL and CRYD. If the second case was true, I would expect CRYD to maintain its susceptibility to protease in its loop compared to PL.

Under denaturing conditions, more bands appeared for both peptides. (Fig. 2.6) A parent band of ~55 kDa appeared for both PL and CRYD. Both the pk1 and pk2 bands appeared faintly in PL. Ck1 also appeared for CRYD. Both gels with the additional SDS had an artifact which smeared the lanes in both the gels, more prominently in that of PL. Additional bands of lower than parent molecular weights appeared smeary and faint.
Overall cleavage was faster for both PL and CRYD under denaturing conditions than under native conditions, as expected. Under denaturing conditions of 0.5% SDS, cleavage due to PK happens more rapidly at $\sim 1 \times 10^{-3}$ s$^{-1}$ for both PL and CRYD. Since no differences are seen, then this indicates that the rigidity of PL remaining in 0.5% SDS no longer is enough to prevent PK cleavage. Therefore, access to cleavable CRYD recognition loop conformations is the likely factor that contributes to the CRYD 10-fold higher rate constant in the experiments under native condition, not the ease of bond cleavage by PK.

### 2.4. Global stability differences between PL and CRYD could not be assessed due to irreversible folding

The overall secondary and tertiary structures of PL and CRYD may play a pivotal role in tuning the dynamics of these loop regions. CD spectroscopy was used to probe the overall secondary structure fold of each of these purified proteins.
CD spectroscopy was used as it gives information on the secondary structure in proteins. $\alpha$-helices generate troughs at 222 nm and 208 nm and a peak at 193 nm. $\beta$-sheets typically have troughs at 218 nm and a peak at 195 nm\textsuperscript{134}.

The limitation of CD spectroscopy is that it probes secondary structure, not tertiary. However, the spectra of PL and CRY\textsubscript{D} appeared to be very similar and comparable to that of the PL spectra from the literature\textsuperscript{21}. Troughs were found in similar intensity and shape between both proteins between 210 nm and 220 nm. This is indicative of a largely $\alpha$-helical secondary structure, which is observed in all three proteins from their crystal structure.

To help better understand the difference between PL and CRY\textsubscript{D}, the secondary structural stability was probed. PL and CRY\textsubscript{D} were subjected to denaturing and non-denaturing conditions to attempt to determine their overall folding stability. Heating PL and CRY\textsubscript{D} produce a sigmoidal unfolding curve. PL and CRY\textsubscript{D} unfolded under both conditions in a sigmoidal fashion, as expected.

Folding stability of both secondary and tertiary structures were measured by monitoring denaturing progress between various proteins. Heat or urea were used to denature the protein. It was expected that the troughs of the PL and CRY\textsubscript{D} CD spectra would be lost after denaturation. The rate was expected to be similar between these two proteins if unfolding is similar. The rate was expected to be faster for a less stable protein.

PL and CRY\textsubscript{D} denatured in a sigmoidal fashion. This indicates a loss of secondary structure, and therefore, tertiary structure in a cooperative manner. The signal reaches zero for PL past 60 °C.(Fig. 2.7) The signal of CRY\textsubscript{D} reached zero past 50 °C.(Fig. 2.7) The profiles of these spectra were different. The PL signal changed slower than that of CRY\textsubscript{D}. The midpoints for unfolding of these two proteins were 44.5 °C for PL and 40 °C for CRY\textsubscript{D}.
Figure 2.7. PL and CRY\textsubscript{D} denature with heat but are unable to refold when cooled.

Note: CD (222 nm) of PL, left, and CRY\textsubscript{D}, right, at varying temperatures either increasing, blue, or decreasing, purple. Extreme hysteresis is observed, indicating a lack of refolding.

These results indicate that PL and CRY\textsubscript{D} denature and lose secondary structure, and its original tertiary structure, by definition. CRY\textsubscript{D} denatures faster than PL, which indicates a difference in stability. The resistance of the PL signal to change relative to CRY\textsubscript{D} indicates that PL denatures slower than CRY\textsubscript{D}. Therefore, qualitatively, PL has a higher folding stability than that of CRY\textsubscript{D}.

To quantify the previous results, the spectra must necessarily return to normal at the same rate when the denaturant is removed. If the CD spectra of PL and CRY\textsubscript{D} return to normal when the denaturant is removed, then it becomes possible to quantify the protein folding stability. However, if PL and CRY\textsubscript{D} do not refold under these conditions or refold at different rates than their unfolding (hysteresis), then quantification of these results cannot be performed accurately.

After complete denaturation with heat, the CD for both PL and CRY\textsubscript{D} did not significantly change upon cooling.(Fig. 2.7) This is an indicator that the proteins did not regain any secondary structures upon cooling. Moreover, this is strong evidence that both of these proteins are unable to refold by themselves after heat denaturation.
The differences in the midpoints indicate that CRY$_D$ is overall slightly less stable than PL. However, the extreme hysteresis between the unfolding and folding reactions makes it impossible to accurately quantify the differences here, as equilibrium would not be reached. It is likely that a portion of the proteins formed aggregates that are thermostable and do not refold.

To probe the stability of these proteins to search for conditions for refolding, chemical denaturation was used. Increasing urea concentrations were used to denature both PL and CRY$_D$.

Both proteins produce a sigmoidal curve when denatured with increasing concentrations of urea. The midpoints of each of these are 2.6 M for CRY$_D$ and 3.0 M for PL. (Fig 2.8)

Figure 2.8. PL denatures in urea.

Note: Denaturation of PL, left, and CRY$_D$, right, monitored by fluorescence spectroscopy. Each data point is the total area derived from a single spectra at a particular urea concentration. Denaturation similarly forms a sigmoidal curve, much like temperature denaturation.

After complete denaturation with urea, the CD for both PL and CRY$_D$ again do not significantly change upon dilution of the urea. This suggests that PL and CRY$_D$ are not able to refold once denatured by urea.
2.5. Differences in the recognition loop dynamics of PL and CRY$_D$ are, at least in part, due to sequence

2.5.1. Understanding the underlying mechanism for these differences in the conformational dynamics can help our understanding of the functional divergence of PL and CRY$_D$

The differences in recognition loop conformational dynamic between PL and CRY$_D$ have been shown in the previous section (See section 2.3). While this observation has implications for PL CPD binding function, understanding the underlying properties that confer these dynamics may help in understanding both the functional implications and the functional divergence of PL and CRY$_D$.

It is plausible that the loop dynamics have an impact on function. The loop displacement when substrate is bound is necessary to prevent steric hindrance and helps by forming interactions with the complementary strand,(Fig. 1.9) making loop dynamics of this region potentially important. If loop dynamics do impact function, then understanding the dynamics of these loops would be important for understanding functional divergence in these proteins.

The differences observed in the reaction rates require local conformational changes of the cleaved region, so the differences observed relates to the differences in the energy landscapes of PL and CRY$_D$. The CRY$_D$ loop thus is shifted towards conformations that are more accessible compared to that of PL.

It is possible that the conformational dynamics of the recognition loop has to do with the environment it is in and tertiary structure of the protein, specifically the local structure. It is also possible that the loop dynamics are encoded in the primary amino acid sequence of these regions.

Five unique conserved residues in PL and adjacent to or in the recognition loop, W392, S395, P402, I405, and P408, are distinct in CRY$_D$. (Fig. 2.9) These residues may be important in contributing to the differences in the energy landscapes due to their
limited $\phi$ and $\psi$ angle restrictions, such as W392, which is at the edge of a helix and is bulky, thereby restricting its torsion angle.

```
392 WSASSGMD - PKP - LRIFNP  A. m. PL
384 WAASTGTD - AAPYFRIFNP  E. c. PL
398 YTAGIGND - ARD - FRYFNI  Syn. CRY - DASH
434 YGAGVGNL - PRE - DRYFSTI  A. t. CRY - DASH
385 YLAGVGAD - PRG - SRQFNL  V. c. CRY - DASH
402 YSAIGND - PRE - NKFNM  X. l. CRY - DASH
402 YITGTLPD - SRE - FDRID  A. t. CRY1
422 WVS S SAEFRLLDSSLVT  D. m. CRY1
399 WLS CS SF QOF FH - - CY  H. s. CRY1
432 WLS S A SFHQYFR - - VY  D. m. (6-4) PL
```

**Figure 2.9. Sequence alignment of the recognition loop region of PL and CRY.**

Note: A.n.: *A. nidulans*, E.c.: *E. coli*, Syn.: *Synechocystis* sp. PCC 6803, A.t.: *A. thaliana*, V.c.: *V. cholerae*, X.l.: *X. laevis*, D.m.: *D. melanogaster*, H.s.: *H. Sapien*. Highlighted residues indicating conserved amino acids in CPD-PL (green), residues common to both CPD-PL and CRYD (blue), conserved residues in CRYD (orange), and residues shared by animal CRY and 6-4 PL (pink and purple).

Shared residues between the PL and CRYD recognition loops include, numbered as PL residues, A394, G397, D399, R404, and F406. It is less likely that these residues have a large impact on the differences observed in the conformational dynamics of the recognition loop. These residues likely would contribute the same or similar effects to their respective loops. (Fig. 2.10)

I hypothesized that the difference observed between the conformation dynamics of the PL and CRYD recognition loops is due in part to sequence. Loop dynamics of isolated PL and CRYD loop peptides were measured to test this hypothesis.
2.5.2. The fluorescence decay of folding CRY$_D$ and PL recognition loop peptides are not mono exponential

To probe loop dynamics, PL and CRY$_D$ recognition loop peptides were created with a 2,3-diazabicyclo[2.2.2]oct-2-ene (DBO) fluorophore at one end with or without a Trp quencher at the other (Fig. 2.11). Absorbance spectroscopy was also used to verify the concentrations of peptides by monitoring Trp absorbances (Fig. 2.12). These peptides were then subjected to TC-SPC (see section 1.6.2), and the fluorescence lifetimes were measured. Ideally, I will observe a single population of substrate become a single population of product creating a single exponential fluorescence decay.
Figure 2.11. Structure of DBO Peptides.

Note: General chemical structure of recognition loop peptides, top; Sequence of recognition loop peptides, bottom. In CRLP’, Phe was used in place of Tyr to prevent any undesired absorbance or quenching from Phe that might affect the experiment.
Figure 2.12. Absorbance spectra of DBO peptides.
Note: Absorbance peak at 280 nm is from Trp residue (65 μM in aerated phosphate buffer). Narrower wavelength selection is shown to better illustrate Trp peaks, inset. Deviations in the spectra around 350 nm may be due to an artifact caused when the light source changes.

If the flexibility differences are due to sequence, then I would expect the CRYD loop to be more conformationally dynamic than that of the PL loop. The fluorophore quenches upon end-to-end contact, so the more dynamic peptide will have a shorter fluorescence lifetime. Therefore, I would expect CRYD loop peptide to have a shorter fluorescence lifetime and a faster rate of fluorescence decay than that of the PL loop peptide. Without quencher present, I would expect similar fluorescence lifetimes and
decay. I would expect the DBO peptides with quencher to have a shorter lifetime than DBO without quencher.

The fluorescence decay of the PL recognition loop lacking quencher is faster than that of free DBO. Fluorescence measurements were performed in low concentrations at 65 µM. The fluorescence decay of DBO in the PL recognition loop without a quencher had an exponential lifetime of 241 ± 5 ns. (Fig. 2.13) This lifetime is shorter than the lifetime of free DBO at 323 ± 5 ns. (Fig. 2.13) but in range of other peptides conjugated to DBO. The lower lifetime is not likely due to quenching from amino acids, as the peptides prepared do not have DBO quenching residues. Instead, it is more likely that the covalent attachment or the environment at the N-terminus plays a role in this decrease.

The peptides with quencher show a double first-order exponential decay. (Fig. 2.13) This equation uses two first-order reactions additively to represent two starting reagents. The intensities were fit to equation (9).
Figure 2.13. Fluorescence lifetimes from TC-SPC.

Note: DBO peptides for PL or CRYD loop (65 μM in aerated phosphate buffer), PRLP’ and CRLP’ decays from TC-SPC experiments. PRLP, blue, PRLP’, green, and CRLP’, teal. Fluorescence lifetimes were fitted to a double or a single exponential curve, equation (9) or equation (10).

\[
I = I_1 \exp(-t/\tau_{D1}) + I_2 \exp(-t/\tau_{D2})
\]  \hspace{1cm} (9)

\[
I = I_0 \exp(-t/\tau_D)
\]  \hspace{1cm} (10)

Where I is the intensity observed, I₁ is the initial intensity contribution from state 1, I₂ is the initial intensity contribution from state 2, t is the time, and τ₁ is the lifetime of state 1 and τ₂ is the lifetime of state 2. Because the decay kinetics of DBO were not monoexponential when quenched by tryptophan, I would conclude, that there is more than one folding reaction taking place, possibly due to internal peptide secondary structure. To determine the fluorescence of the non monoexponential decays, amplitude-
weighted averages of $\tau_{DA}$ were used from fitted double first-order exponential decays, equation (11).

$$\tau_{DA(average)} = \frac{l_1\tau_{DA1} + l_2\tau_{DA2}}{l_1 + l_2}$$

(11)

2.5.3. Under native conditions, end-to-end contacts occur faster in the CRY$_D$ loop peptide than that of PL

The PL and CRY$_D$ loop peptides conjugated to DBO and Trp had shorter lifetimes than without quencher, $\tau_{DA} = 92 \pm 3$ and $47 \pm 2$ ns, respectively. The $k_{fold}$ of the CRY$_D$ loop was found, from equation (3), section 1.62, to be $1.72 \times 10^7$ s$^{-1}$, which is 2.5 fold faster than that of the PL loop, with a $k_{fold}$ of $6.8 \times 10^6$ s$^{-1}$. I would conclude that conformational dynamics of the CRY$_D$ loop is higher than that of the PL loop. Therefore, the heightened conformational dynamics of the CRY$_D$ recognition loop is, at least in part, due to sequence.

2.5.4. Intermolecular quenching is unlikely

Intermolecular quenching is not expected due to the low concentrations of the peptides.$^{126}$ Lifetimes of DBO, as measured in the fluorescence lifetimes experiments appear to be independent of peptide concentrations between 50 to 100 $\mu$M. A simple experiment was used to rule out intermolecular quenching. CP, a PL loop peptide that has the Trp quencher, but lacks DBO was used.(Fig. 2.11) In this experiment, samples of PRLP and CP were mixed in 50:50 ratio at the same total peptide concentration as before. Lifetimes were then measured.

If intermolecular quenching is present, I would expect a lower intensity of fluorescence in this experiment due to the concentration of PRLP being decreased. I would, more importantly, expect a faster rate of decay in this mixed sample compared to PRLP alone due to quencher being present (trans quenching). If intermolecular quenching is not present, I would still expect a lower intensity of fluorescence due to the concentration of PRLP being diluted. I would also expect the same rate of decay compared to the tests of homogeneous PRLP, “trans” quenching would not occur.
The latter is observed in this experiment, so it can be concluded that intermolecular quenching is not present in this experiment at the concentrations used. (Fig. 2.14) While this experiment was only carried out for the PL peptide, we make the assumption that the case is similar for that of the CRY_D peptide. This assumption is likely valid, as the length and concentrations of each peptide are identical. There have also been experiments that show the intermolecular quenching for this type of experiment is negligible under these concentrations\textsuperscript{124,126}.

![Figure 2.14. Intermolecular quenching is not present in the peptide folding experiments.](image)

**Notes:** Fluorescence peptides PRLP and CP were mixed and exposed to a TC-SPC experiment (65 \( \mu \)M in aerated phosphate buffer). PRLP, blue; PRLP mixed with CP, Green. The decay rates of the mixed sample are comparable to that of the non-mixed samples indicating that intermolecular interactions are not significant.
2.5.5. It is likely there are secondary structures forming in the peptide

A possible reason for the non-monoexponential decay could be due to incomplete conformational averaging. Interchain interactions would only be able to produce single exponential folding kinetics if they were able to interconvert rapidly compared to the folding timescale. Thus I propose that the double exponential behaviour reflects two distinct conformations which do not interconvert on the timescale of the experiment.

To test whether this difference in secondary structure formation exists, the peptide fluorescence lifetimes were measured under a denaturing condition of 8 M urea. I expect that under these conditions, both the PL and CRY\textsubscript{D} recognition loop peptides will be mainly a single population of unfolded peptides. This will allow for a single type of unfolded reagent to produce a single type of folded product. I would expect the lifetime of the fluorophores to be monoexponential or closer to mono exponential due to loss of interchain interactions in each peptide.

I would also expect an increase of the lifetimes due to the increase of the solvent viscosity from the denaturant. When viscosity of the solution is increased, the excited fluorophore may lose less kinetic energy to the environment. Lowering the ability of the molecule to transfer energy as heat shifts the equilibrium to increase the amount of energy lost as light energy in the form of emission fluorescence. In addition, due to the viscosity of the solvent, the end-to-end collisions are also slowed. Therefore, the fluorescence of DBO should increase in urea.

A number of possible outcomes exist. If the peptide secondary structure completely denatures, the fluorescence lifetime should be mono exponential. If the secondary structure only partially denatures, the decay should be closer to, but not entirely single exponential. If the peptide secondary structure is not caused by secondary structure or if that secondary structure did not denature, the same double exponential decay from non-denaturing conditions would be maintained. I would also predict, that the CRY\textsubscript{D} loop peptide will maintain its faster rate of folding over that of PL.
2.5.6. **Double exponential lifetimes changes to mono exponential under denaturing conditions**

Under 8 M urea denaturing conditions, the decay for loop peptides with and without quencher became mono exponential. (Fig. 2.15 and Table 2.2) This suggests that there was in fact more than one conformation in slow exchange in the non-denaturing experiment. The decay lifetimes increased by 1.3-fold, 2-fold, and 2.6-fold to 304, 188 and 122 ± 2 ns for the loop peptide without quencher, the PL loop with quencher and the CRY loop with quencher, respectively. The slight increase of the peptide lacking quencher, relative to the quenched peptides, indicates that the increase of fluorescence due to increased viscosity is a minor component.

![Fluorescence lifetimes from TC-SPC experiment under denaturing conditions.](image)

**Figure 2.15. Fluorescence lifetimes from TC-SPC experiment under denaturing conditions.**

Note: The peptides PRLP, PRLP' and CRLP' were monitored in a TC-SPC experiment in the presence of 8M urea (65μM in aerated phosphate buffer). PRLP, blue, PRLP', green, and CRLP', teal. Fluorescence lifetimes were fitted to a single exponential curve, equation (10).
Table 2.2. Fluorescence lifetimes of DBO peptides.

<table>
<thead>
<tr>
<th>Peptide</th>
<th>Quencher</th>
<th>Urea</th>
<th>$\tau_1$ (ns)</th>
<th>$\tau_2$ (ns)</th>
<th>Average $\tau$ (ns)</th>
<th>Category of $\tau$</th>
</tr>
</thead>
<tbody>
<tr>
<td>DBO</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>323 ± 5</td>
<td>$\tau_D$</td>
</tr>
<tr>
<td>PRLP</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>241 ± 5</td>
<td>$\tau_D$</td>
</tr>
<tr>
<td>PRLP’</td>
<td>+</td>
<td>-</td>
<td>85</td>
<td>303</td>
<td>92 ± 5</td>
<td>$\tau_D\alpha$</td>
</tr>
<tr>
<td>CRLP’</td>
<td>+</td>
<td>-</td>
<td>44</td>
<td>110</td>
<td>47 ± 5</td>
<td>$\tau_D\alpha$</td>
</tr>
<tr>
<td>PRLP</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>304 ± 2</td>
<td>$\tau_D$</td>
</tr>
<tr>
<td>PRLP’</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>188 ± 2</td>
<td>$\tau_D\alpha$</td>
</tr>
<tr>
<td>CRLP’</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>122 ± 2</td>
<td>$\tau_D\alpha$</td>
</tr>
</tbody>
</table>

Note: Fluorescence decays ($\lambda_{ex} = 372$ nm, $\lambda_{em} = 430$ nm) of recognition loop peptides (65 µM) in aerated phosphate buffer or 8 M urea.

Under denaturing conditions, the lifetimes for both PL and CRY$_D$ loop peptides become single exponential. The overall difference in the folding is maintained for both the CRY$_D$ and PL recognition loop peptides under denaturing conditions. The $k_{\text{fold}}$ for CRY$_D$ is $4.9 \times 10^6$ s$^{-1}$ and the $k_{\text{fold}}$ of PL is $2.0 \times 10^6$ s$^{-1}$. The difference between these mono exponential rates is 2.5-fold, which is comparable to the difference observed under non-denaturing conditions. Based on this data, I would conclude that the peptides lose their secondary structure under denaturing conditions. I would also conclude that the CRY$_D$ recognition loop is more conformationally dynamic than that of PL.

2.5.7. Summary: CRY$_D$ loop peptide is more dynamic than that of PL

The change of the fit to monoexponential under denaturing conditions supports the validity of my analysis for the double exponential fits under non-denaturing conditions. As mentioned earlier (see Section 2.5.6), observing a change to monoexponential decay would suggest that the peptide lost secondary structures under denaturing conditions. Therefore, under native conditions, the peptides have a population that contain some secondary structure, while under denaturing conditions, the peptides do not contain secondary structure.
These results suggest that the dynamic nature of these peptides is, at least in part, rooted in sequence. Isolated peptide sequences only depend on the sequence and their environment for determination of dynamics. This is in contrast to the loop sequences in the context of the whole protein whose dynamics can be affected by the protein scaffold.

These results are also consistent with the limited proteolysis experiments. This gives two independent methods in which the CRYD recognition loop is more conformationally dynamic than that of PL. Since both results confirm one another, the confidence in each is increased.

2.5.8. **The energy of folding is mainly influenced by enthalpy**

I considered whether the dynamics of the isolated loop peptides was primarily driven by enthalpy or by entropy. I hypothesized that the peptide folding dynamics is mainly influenced by enthalpy. The PL structure has more amino acids, like proline, that would typically contribute a higher enthalpic component to folding activation, while CRYD has residues, like glycine, that typically contribute lower enthalpic barrier. I would expect to see a high enthalpy component and a small entropy component. If the increase in loop dynamics in CRLP can largely be attributed to the decrease in the enthalpic barrier to conformational rearrangement observed in PRLP, then the larger PRLP enthalpic barrier may be reflective of preorganization of this loop in PL in the absence of substrate.

The $k_{\text{fold}}$ of the PL and CRYD loop peptides were measured over a range of temperatures under native and denaturing conditions. The lifetime for the DBO conjugated loop increased with decreasing temperature. This is expected because as thermal motion decreases, the ability for end-to-end contact decreases since movement is the basic requirement for quenching to occur if the ends are not already in contact.

Eyring plots of the CRYD and PL loops were used to determine the barrier free energies of folding, which were about 34.3 and 36.4 kJ/mol, respectively, under denaturing conditions (Fig. 2.18). Under denaturing conditions, the entropic components were found to be similar within experimental error, with only 0.1 kJmol\(^{-1}\) difference.
However, the enthalpic component showed a 2 kJ/mol difference, with CRLP being lower. The entropic parameter doesn’t change significantly between the two peptides and thus does not play a large role in the folding rate differences.

Figure 2.16. Eyring plots of PL and CRY\textsubscript{D} recognition loop peptides.
Note: Temperature-depending folding rate constants under native, N (phosphate buffer pH 7.5), or denaturing, U (8 M urea), conditions of the recognition loop peptides. Lines are fit to equation (8).

2.5.9. Conclusion: increased folding dynamics could impact binding due to increase in conformational space

The increased conformational dynamics of the CRY\textsubscript{D} recognition loop is consistent with my hypothesis that loop dynamics influence the ability for CRY\textsubscript{D} to form an active complex with dsDNA. The increased conformational dynamics of the CRY\textsubscript{D} recognition loop likely allows for a larger conformational space. It is also likely that the added conformational states are unproductive states, since there would be few states
that would allow for formation of the active complex. This would cause a shift in the equilibrium towards more unproductive states and preventing CRY$_D$ from forming the active complex.
3.1. A plan to test my hypothesis

I will prepare a series of compromised PL recognition loop mutants to determine if a correlation exists between conformational dynamics and function. These mutants will be assayed by limited proteolysis as well as functional assays. A correlation between conformational recognition loop dynamics and function will implicate conformational dynamics of this region in DNA repair.

Correlating loop conformational dynamics with dsDNA repair capability in PL will determine if the recognition loop functions in CPD binding. A correlation between rate of enzyme function and recognition loop conformational dynamics will be consistent with the hypothesis that loop dynamics are important for facilitating CPD binding. The most active PL recognition loop sequences of CPD repair would indicate the optimal amount of conformational dynamics for the loop.

I propose for this outcome that decreasing the conformational dynamics relative to PL would increase the barrier to changing conformation and prevent recognition loop movement out of the way of the CPD. Increasing the conformational dynamics of the loop region would increase the conformational space that the recognition loop could sample shifting the equilibrium towards unproductive states, preventing the recognition loop from entering a productive state to allow for the active complex to form.

Alternatively, should there be no correlation between conformational dynamics and dsDNA repair, the hypothesis would be rejected. If this were the case, I would
conclude that loop dynamics play no role in CPD binding. This would be an interesting implication since there are conserved residues in the loop which do not participate in DNA binding. This would indicate that another region of PL is important for the binding of CPD.

To test this hypothesis, an accurate quantitative measure of loop dynamics will be needed. Based on the discussion from chapter 1 of this thesis, limited proteolysis is the ideal choice. This assay’s sensitivity is only limited by the ability to detect and measure cleavage products. Another limitation of this assay is that it requires the mutants to be purified, which may or may not be possible. Alternatively, the end-to-end contact fluorescence method could be employed (section 1.6). This assay is limited by the fact that the loops are not measured in the context of the proteins. It is desirable to measure the loop dynamics and how they change in the context of the enzyme’s tertiary structure.

To accurately quantify the repair rates of the mutants, the restriction-site protection assay will be used. This assay will allow for detection of small changes in rate of function between mutants. The minimum number of mutants required to draw a correlation would depend on the type of fit observed in the correlation and would depend on how much each mutant could deviate from the trend.

A loop mutant library with functionally compromised mutants is necessary to test this hypothesis that the recognition loop conformational dynamics is important in PL function. Preexisting mutations from the literature are not in the loop\textsuperscript{17,122}. One method for finding compromised versions of PL is through directed evolution\textsuperscript{136,137}. After performing the selection of partial function PL clones (see section 1.7), an error-prone PCR (epPCR) process can then be repeated on these mutants to gain different mutations, potentially with increased or decreased levels of compromised function. In a separate example, 6 rounds of directed evolution were used to generate 7 different variations in function in fluorescent proteins\textsuperscript{137}. 

82
3.2. Technical development summary

The work presented in this section includes progress and plans for the technical developments made in this thesis. The overall scheme for testing my hypothesis was outlined. (Fig. 3.1) Degenerate PCR was used to generate a library of loop mutant sequences. Following this, epPCR would enable fine tuning of the mutation rate. An \textit{in vivo} selection screen would allow for identification of compromised PL loop mutants. The expression could be confirmed by Western blot. The mutant PL would then be purified and have its repair rate determined by an \textit{in vitro} repair assay. Limited proteolysis could then be used to test loop conformational dynamics.

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{flowchart}
\caption{Flow chart of planned experiments.}
\end{figure}
3.3. Mutagenic PCR methods were validated

3.3.1. A cassette was designed and created to hold the PL loop coding region for creating random mutations

To aid in the generation of a library of PL mutants, a plasmid was used as a cassette to hold the recognition loop region of PL for further rounds of mutagenesis. This plasmid, pm242, contains many features which includes a single NheI cleavage site to insert the loop region, being only ~4 kbp long, as well as an amp resistance gene for selection. Insertion of the loop in the NheI site would generate the plasmid pm242-cas. (Fig. 3.2) The PL loop was successfully inserted into the NheI site of this plasmid, verified by endonuclease digests and gel electrophoresis, as well as DNA sequencing. The next step was to perform degenerate primer PCR on the recognition loop coding region of this construct.

![Figure 3.2. The plasmid map of pm242-cas.](image)

3.3.2. Degenerate PCR was validated

For the initial generation of the PL loop random mutant library, a mutagenic technique must be employed. Degenerate primer PCR was used to generate a large
initial pool of recognition loop coding regions with a high mutation rate, along with NheI endonuclease sites at either end for future subcloning. The degenerate primer contains an 18 bp region where each base has a 30% probability to mutate to a different base.

Given this mutation rate of 30%, the average number of mutations can be calculated using binomial or Poisson distribution. Binomial distribution was selected as a model of distribution since this will determine the probability mass function of a sample size, \( n \) = number of degenerate primers used, without replacement, where the population size, \( N \) = number of degenerate primers available, is much greater than the sample size. Poisson distribution was selected to determine the probability mass function of this system since there is a set number of replication intervals and an average number of expected mutations. As the sample size for the binomial distribution becomes large, it begins to closely approximate the probability mass function of the Poisson distribution. In this case, the binomial distribution is a reasonably close approximation to that of the Poisson distribution.

The mean for both Poisson and Binomial is equivalent. The mean for the rate of 30% mutation probability of each of the bases in the 18 bp region is 5.4 expected base mutations per PCR amplified strand. Based on the probability mass function, I would expect more than 95% of the PCR products to contain no fewer than 1 mutation and no more than 11 mutations per fragment.(Fig. 3.3)
Figure 3.3. Degenerate primer PCR amplification of the PL loop region.

Note: Template and primers shown with in-frame wild type translation sequence, top. Graphic shows total length of PCR product. Key position numbers shown, from left to right, 5' of PCR product from beginning of forward primer, 5' of degenerate sequence (blue highlight in primer sequence), 5' of BamHI site in wild type, end of loop region to be cloned (yellow highlight in template sequence), end of sequence shown, end of PCR product. Bottom left, PCR and Degenerate PCR products with and without BamHI digestion. Degenerate PCR yields near 100% protection from BamHI, bottom. Poisson and Binomial distributions of the number of errors expected after degenerate primer PCR amplification of the PL loop region, bottom right. Where Y is the probability of having N mutations, M is the length of the degenerate region of the primer, N is the number of successful mutations, \( \lambda \) is the expected average number of mutations, and e is Euler's number.

To determine the approximate number of random errors generated in a single PCR product, an assay which utilizes the specificity of endonuclease recognition sites was developed. The region that is PCR amplified contains a naturally occurring BamHI endonuclease restriction site, which upon mutation will no longer be recognized by BamHI. The proportion of PCR amplified fragments that are resistant to BamHI digest,
will be proportional to the number of total random mutations if the mutations are normally distributed.

The probability that a fragment contains at least one mutation in the BamHI site can be predicted. Given this 6 bp endonuclease recognition site, the probability that none of these bases mutate is \((0.3)^6\) or 0.07%. Therefore the probability of fragments containing at least one mutation should be more than 99.93%. I would expect the same amount of BamHI protection in this PCR reaction. The degenerate PCR loop product showed complete protection from BamHI, consistent with my expectations. The expected 0.07% expected digestion was not detectable.(Fig. 3.3) Therefore, I would conclude that near 100% of the PCR product was mutated. The next step is to insert this PCR library into a plasmid (see section 3.4.3).

3.3.3. Error-prone PCR was validated

Error-prone PCR (epPCR) allows for a random mutant PL library to be created or enhanced. EpPCR utilizes a DNA polymerase that is more prone to creating errors\(^\text{138}\). The errors are accumulated during the PCR cycles and produce a random number of mutations in random positions. Typical error rates achieved can be greater than 9 mutations/kb over 100 to 10,000 fold PCR amplification. The method of epPCR using Mutazyme II has been used to generate a variety of GFP variants with different absorbance and emission spectra\(^\text{137}\).

There are other methods that can be used to produce random mutations. Degenerate primers can be used to create errors in specific positions with certain frequencies. PCR with limiting amounts of dideoxyribonucleic acid and magnesium concentrations can also be used to add errors.

For my purposes, PCR using an error-prone polymerase, Mutazyme II, will greatly increase the error rate to allow for subsequent rounds of error generation. The error rate can range from 0 to 16 mutations/kb under varying PCR conditions. In addition to this, there are two orientations the PCR fragment can be inserted to regenerate the loop in PL, a forward and a reverse direction, further increasing the mutational space.
3.3.4. **A procedure for generating a random mutant library was developed using epPCR**

Before using the epPCR on the degenerate primer product, the epPCR needed to be validated and the error rate quantified. To test this, normal PCR and epPCR of the a 189 bp long test construct containing three BamHI sites, at positions 52, 102, and 151, was performed. (Fig. 3.4) The error rate was compared between the two using the BamHI protection assay (see section 3.3.3).

![Test Construct Diagram](image1)

**Test Construct**

<table>
<thead>
<tr>
<th>Position #</th>
<th>1</th>
<th>52</th>
<th>102</th>
<th>151</th>
<th>189</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>GGATCC</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>GGATCC</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>GGATCC</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Figure 3.4. Error-prone PCR protection assay.**

Note: PCR and epPCR template, top. BamHI protection assay, bottom left. A small amount of DNA is protected from BamHI digestion after epPCR (red arrow).
In this case, given the 3 BamHI regions covering 18 bp, and given a 1% rate of mutation, the probability that no changes occur to the BamHI region is \((0.99)^{18}\) or 83%. Therefore, I would expect 17% of the total PCR products to be protected from BamHI. Conversely, the error rate can be determined experimentally by measuring the percent of protected PCR product and back calculating to determine the error rate.

A PCR amplified test fragment was generated, verified by gel electrophoresis. Similarly, an epPCR amplified fragment of the test construct was generated, also verified by gel electrophoresis. Digesting both products with BamHI yielded products at 50 bp or 100 and 50 bp products for the PCR and epPCR products, respectively.

The PCR BamHI digestion indicates that the PCR product was fully digested. However, while the vast majority of the epPCR product was also digested to 50 bp products, there was observable protection shown by the 100 bp product. This product is the result of one of the three BamHI sites being protected with the other two being digested. This small amount is expected, as only 17% is expected to be protected. I would conclude that the epPCR produced 1% mutations. The next step in the procedure is to perform rounds of epPCR after selection of PL mutants.

3.4. **PL *in vivo* functional assays was developed using repair deficient cells**

3.4.1. **PL *in vivo* functional assay was validated**

In order to screen for compromised PL loop mutants, a quick method for assaying function is needed. The bacterial survival assay allows for detection of mutants with compromised, but still significant, repair capability, without the need of purifying proteins. As explained previously (section 1.7) the assay relies on *E. coli* cells that are devoid of DNA repair function. A mutant library, as described earlier (section 3.3.4) can be transformed into these cells. These cells can then be irradiated with UV light. Cells harbouring fully functional or semi functional mutants will have a higher probability of surviving, and cells lacking PL function will have a higher probability of dying. Functional mutant clones can be tested and selected for using the survival assay. Using the
A degenerate primer mutation method, the probability of keeping wild type PL is less than 0.2%, according to the Poisson distribution. Therefore, non-functional mutants can be screened away, while compromised and fully functional mutants will be retained. The next step of this process would be to further characterize compromised PL mutant and perform progressive rounds of epPCR and selection.

A variant of the *in vivo* functional assay (section 1.7), called the dot spot assay, has been developed for rapid identification of compromised PLs from a library of PL mutants. This assay, similarly, monitors UNC1085 survival after exposure to UV and blue light in the *in vivo* functional assay. However, the key difference is in how these cells are plated. In the dot spot assay, cells are spotted on a plate from stocks of cultures of identical optical densities.

Multiple dilutions of the stock solution are made and are individually spotted on the same plate. Exposure of these spots to UV light will result in certain spots containing low colony counts that are unable to grow or spots containing high counts of colonies that merge together in the spotted area. The dilution factor at which the gradient begins to change from no colonies to many colonies is an indicator as to how functional the transformed plasmid is at rescuing cells. I would expect wild type PL to survive up to a higher dilution factor than non-functional or compromised mutants.

PL containing cells given no UVC, or UVC alone, or both UVC and UVA also showed results consistent with my expectations. PL containing cells without any treatment grew up to 4 log dilutions. These cells treated with UVC alone could only grow up to 2 log dilution, indicating cell death. These cells treated with both UVC and UVA grew up to 4 log dilutions, indicating photoreactivation.
Figure 3.5. PL in vitro functional assay.

Note: UNC1085[pet3AnPhr] cells containing A. nidulans PL grew up to 4 log dilutions. Exposure to 30 s UV light reduced the ability of these cells to grow to 2 log dilutions. However, given 30 mins UVA, this cell type can be rescued back to surviving up to 4 log dilutions.

After exposure to UVC and UVA, cells containing different plasmids show different results. Cells containing the PL plasmid allowed for cell growth of cells up 4 log dilution. Cells containing an empty vector showed cell growth of up to 2 log dilution. Cells containing the CRYD plasmid showed similar results to the empty vector and only allowed for cell growth of up to 2 log dilution as well. It was expected that the cells containing CRYD plasmid would have the same low repair activity as cells containing an empty vector, both significantly lower than cells containing the PL vector. This assay shows detectable differences between the function of PL and CRYD. It also has a relatively wide range of dilutions to detect functional, functionally compromised, and non-functional PL mutants. The next step is to repeat this assay with the pPLL plasmid transformed into UNC1085 to allow for control of expression.

3.4.2. An improved plasmid system for PL was designed

A plasmid was designed and developed to accept the PL loop mutant library to reform whole PL. This plasmid, pPLeL contains the PL coding region with a NheI site in place of the loop region. After being cleaved by NheI, a PL loop region with NheI sticky
ends may be inserted into this plasmid to regenerate the whole protein in frame. Immediately upstream of the PL gene, a poly-his tag is encoded to allow for nickel column purification of the protein and immunoblotting. The full protein is under the control of a Tac promoter to allow for control of expression. (Fig. 3.6)

Figure 3.6. Plasmid map of pPLeL.

To generate this plasmid, the PL coding region from the plasmid pet3AnPhr was subcloned into the plasmid pGEX-4T-2. This plasmid was verified by gel electrophoresis and endonuclease digestion. DNA sequencing confirmed the sequence of this plasmid’s DNA coding region. The next step is to insert the degenerate PCR product or the PL wild type loop into the Nhel site.

3.4.3. A plasmid, pPLL, was designed to contain the mutant and wild type PLs

The whole wild type and mutant PLs are the desired end products of cloning. The desired plasmid to achieve this is a combination of pPLeL (see section 3.4.2) and the
PCR amplified loop region of pm242-cas (see section 3.3.1) such that the PL recognition loop region is regenerated in pPLEL. The newly formed plasmid is named pPLL or pPLL* for wild type and mutant library, respectively.(Fig. 3.7)

**Figure 3.7.** Plasmid map of pPLL.

Previous plasmids containing PL could be overexpressed only in specific *E. coli* strains. The plasmid pPLL will allow for overexpression of PL and PL mutant libraries in any strain utilizing the TAC promoter and containing the lac repressor. It will also allow for quantification of cellular concentration of PL. The next step of this process would be to screen pPLL* for compromised PL mutants.

**3.4.4. Western blot detection of PL was established**

To address the issue of potential differences in cellular concentrations of PL and PL mutants, a Western blot was developed for quantification of PL cellular
concentrations. A single band of 54 kDa was expected to be detected by the Western blot.

A N-terminal his-tagged PL that lacks the recognition loop region, pPLEL,(Fig. 3.6) was expressed to validate the Western blot. Anti-his antibodies were used for binding to the PL. With no induction, no bands were observed, as expected. After 2 h and 3 h, the a 54 kDa band was observed in the similar intensities, but, unexpectedly, lower molecular weight products were also observed.(Fig. 3.8)

Figure 3.8. Western blot of an expressed his-tagged PL from whole cell lysate.
Note: Without induction, the his-tagged PL is not detected in the whole cell lysate. After 2 h, PL is detected. After 3 h, PL is detected with a significant minor lower molecular weight band. The 30 kDa control is an unrelated his-tagged protein.

These results suggest that PL is being expressed and is detected by this method. It is possible that these smaller bands are fragmented N-terminal products of degraded PL. These bands are consistent with the type of breakdown products seen in the earlier proteolysis experiments. It was observed previously that there are regions in PL which are more susceptible to proteolysis, as seen using a range of proteinases.(See section 1.7)
3.5. **PL in vitro** functional assay was established for monitoring of enzyme activity and comparison between enzymes

3.5.1. Changes in absorbance at 260 nm allows for detection of damaged DNA in ssDNA

To detect the presence of thymine dimers in real time, an *in vitro* assay was developed. This assay is dependent on the decrease in the absorbance at 260 nm upon forming thymine dimers. The formation of thymine dimers removes the aromatic nature of the bases, thereby removing its UV absorbing properties.

The extinction coefficient of thymines at wavelength 265 nm decreases from about 19,000 M$^{-1}$cm$^{-1}$ to 0 M$^{-1}$cm$^{-1}$ upon dimer formation$^{55}$. The substrate used was a (dT)$_{18}$ ssDNA substrate. This substrate was used to allow the majority of absorbance at 260 nm to be due to thymines. The majority of the change upon thymine dimer formation would then be attributed to repair of thymine dimers.

The procedure for this experiment was to use single stranded (dT)$_{18}$ or double stranded DNA oligos. Spectra from 200 nm to 600 nm were taken before and after irradiation with UV light to generate thymine dimer damage.

If 100% dimerization were complete, I would expect that there would be a 100% decrease in the absorbance at 260 nm. However, a thymine could dimerize with a thymine on its 5' or 3' end. Therefore, it is possible that a thymine pair could preclude a non-paired thymine from thymine dimer formation. If this situation happens maximally to exclude the largest number of single thymines from forming a dimer, a total of 33.3% of the dimers may be left undimerized. Therefore, under these conditions, I would expect at most a 66.6% decrease in the absorbance under these conditions.

After UV irradiation, the A$_{260}$, as a percentage of the non-irradiated, was 58.2% after 3.5 h,(Fig. 3.9) indicating 41.8% dimer formation. This is less than the 66.6% decrease, indicating that there are regions available for further dimerization. I would
conclude that thymine dimers were successfully created and monitored by this method. The next step is to repair these dimers in vitro.

**Figure 3.9.** UV-vis spectra of thymine dimer generation in (dT)\textsubscript{18}.

Note: (dT)\textsubscript{18} before (solid triangles), and with (open triangles) or without (covered during irradiation) 2.5 h UV irradiation (solid circles). Exposure to UV decreases the absorbance of the (dT)\textsubscript{18} substrate by more than 50%. This is consistent with thymine dimer generation. All DNA concentrations are 170 ng/\mu L.

3.5.2. Repair of ssDNA can be monitored by the change in absorbance at 260 nm

Measuring PL catalytic activity can be used to determine if PL compromised mutants are compromised due to binding and not due to compromised catalysis. One way of doing this is by monitoring ssDNA thymine dimer formation and repair using UV spectroscopy (section 3.5.1). Thymine dimers were induced in (dT)\textsubscript{18} as determined by
UV-vis spectroscopy. (Fig. 3.9) To repair this damage, the UV-damaged oligo was then incubated with photoreduced PL and exposed to UVA light.

The absorbance at 260 nm was used to monitor percent repair. It would be expected that this method would allow for a maximum of 90% of repair damage, as ~90% of photoproducts generated by UV irradiation are CPD. After irradiation for 2 hours of UVA, the absorbance increased indicating repair. The increase observed was from 0.65 to 1.13 absorbance units, with a theoretical expected maximum of 1.15, taking into account 90% CPD formation by UVC. (Fig. 3.10) Therefore, the total repair seen in this experiment was 96% over 125 mins. This is a rate of 0.23 absorbance units/h, or 46% repair/h.

Figure 3.10. ssDNA repair by PL.
Note: The durations shown are the amount of UVA exposure for PL (440 nM) incubated with (dT)₁₈ (65 ng/μL).

While ssDNA is not the physiological substrate of PL, this assay was designed to be used to help determine if PL mutants have compromised catalytic function. This
process is also a quick test, as the signal is strong and there are few inputs. It is also important to measure repair of dsDNA, which will be covered in following sections (section 3.5.3 and 3.5.4). The next step is to measure multiple PL mutants with this assay.

3.5.3. Monitoring of the absorbance at 260 nm allows for detection of damaged DNA in dsDNA

A second substrate was used to assay for dsDNA function. This substrate was a \((dT)_{18}\) flanked by 6 bases on either side, annealed to a \((dA)_{18}\) flanked by the complementary sequence to each of the 6 bases on the \((dT)_{18}\) strand. (Fig. 3.11)

![Double stranded DNA substrate.](image)

**Figure 3.11.** Double stranded DNA substrate.

*Note:* Sequence of dsDNA substrate containing 18 thymines, similar to the ssDNA \((dT)_{18}\) substrate. The chain of thymines are flanked by sequence for complementary pairing to the adenine rich strand.

The procedure for generating this substrate was to mix these two oligos and melt any structure through heat denaturation. Annealing of these complementary strands was facilitated by cooling the mixture to room temperature. A spectrum of the product was recorded at 80 °C, before and after irradiation. The product was irradiated with UV light to generate thymine dimers.

The expectation for this experiment was for there to be a decrease in the \(A_{260}\), similar to that of the ssDNA monitoring experiment, but not as intensely (see Section 3.5.2). The added absorbance at 260 nm from the adenine strand was expected to stay constant before and after the UV irradiation. It was also expected that the 260 nm absorbance would be comparable to that of the thymine strand before UV irradiation.

The ssDNA decreased by ~58% after irradiation of 2.5 h. The total absorbance contribution by the thymine strand is 43%, \(\varepsilon = 19,000 \text{ M}^{-1}\text{cm}^{-1}\) for thymine \((T_pT)_{55}\) and \(\varepsilon = 24,000 \text{ M}^{-1}\text{cm}^{-1}\) for adenine \((A_pA)\) when the dsDNA (5% excess adenine strand) is
melted. So for a 58% decrease in the absorbance of an oligo that contributes 43% of the total absorbance units of this construct, I would expect a total decrease of 25% in the absorbance over 2.5 h or 10% per hour.

After 1 h of UVC irradiation, the dsDNA showed an 8% decrease in its absorbance at 260 nm. (Fig. 3.12) The decrease of 8%, out of a possible 43% absorbance contributed by thymines, corresponds to approximately 19% generation of thymine dimers over 1 h.

Figure 3.12. UV-vis spectra of thymine dimer generation in dsDNA.
Note: dsDNA substrate before (solid triangles), and after 1 h UV irradiation (open triangles). Exposure to UV decreases the absorbance of the dsDNA substrate by about 10%. This is much less than anticipated.
3.5.4. Repair of damaged DNA in dsDNA

The difference in the absorbance between samples containing UVC treated and untreated DNA was 0.07 before repair. I would expect the thymine dimers of the UV-irradiated dsDNA to be repaired by a maximum of 90% of its original, as that is the proportion of CPD that is expected to be generated through UVC irradiation. Therefore, the theoretical maximum OD change would be 0.06. The observed absorbance increased a total of 67% of this 0.06 after 1 hour UVA treatment. (Fig. 3.13 and Table 3.1)

Figure 3.13. Repair of dsDNA substrate.
Note: Substrate with PL (320 nM) with damaged dsDNA (95 ng/µL) before UVA irradiation (solid triangles), with 60 min UVA irradiation (open triangles) or incubated at 60 min without UVA irradiation (solid circles). All measurement of Substrate was performed at 80 °C.
### Table 3.1. Thymine dimer synthesis.

<table>
<thead>
<tr>
<th>Substrate</th>
<th>Treatment</th>
<th>Enzyme</th>
<th>Max $\Delta OD_{260}$</th>
<th>Theoretical Max $\Delta OD_{260}$</th>
<th>Dimer Synthesis (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(dT)$_{18}$</td>
<td>No UV</td>
<td>-</td>
<td>0</td>
<td>3.26</td>
<td>0</td>
</tr>
<tr>
<td>(dT)$_{18}$</td>
<td>UVC, 2.5 h</td>
<td>-</td>
<td>1.90</td>
<td>3.26</td>
<td>42</td>
</tr>
<tr>
<td>dTdA-2</td>
<td>No UV</td>
<td>-</td>
<td>0</td>
<td>1.12</td>
<td>-</td>
</tr>
<tr>
<td>dTdA-2</td>
<td>UVC, 1 h</td>
<td>-</td>
<td>0.21</td>
<td>1.12</td>
<td>19</td>
</tr>
</tbody>
</table>

Note: Reactions were carried out in 20 mM Tris with 2 mM EDTA, pH 7.5. dTdA-2 absorbance was monitored at 80 °C to melt DNA.

### Table 3.2. Thymine dimer repair.

<table>
<thead>
<tr>
<th>Substrate</th>
<th>Treatment</th>
<th>Enzyme</th>
<th>Observed $\Delta OD_{260}$</th>
<th>Theoretical Max $\Delta OD_{260}$</th>
<th>Dimer Repair (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(dT)$_{18}$ + UVC, 2.5h</td>
<td>-</td>
<td>PL</td>
<td>0</td>
<td>0.5</td>
<td>0</td>
</tr>
<tr>
<td>(dT)$_{18}$ + UVC, 2.5h</td>
<td>UVA, 125 min</td>
<td>PL</td>
<td>0.48</td>
<td>0.5</td>
<td>96</td>
</tr>
<tr>
<td>(dT)$_{18}$ + UVC, 2.5h</td>
<td>-</td>
<td>PL W392Y</td>
<td>0</td>
<td>0.5</td>
<td>0</td>
</tr>
<tr>
<td>(dT)$_{18}$ + UVC, 2.5h</td>
<td>UVA, 125 min</td>
<td>PL W392Y</td>
<td>0.49</td>
<td>0.5</td>
<td>98</td>
</tr>
<tr>
<td>dTdA-2</td>
<td>-</td>
<td>PL</td>
<td>0</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>dTdA-2 + UVC, 1 h</td>
<td>-</td>
<td>PL</td>
<td>0</td>
<td>0.06</td>
<td>0</td>
</tr>
<tr>
<td>dTdA-2 + UVC, 1 h</td>
<td>UVA, 1 h</td>
<td>PL</td>
<td>0.04</td>
<td>0.06</td>
<td>67</td>
</tr>
<tr>
<td>dTdA-2</td>
<td>-</td>
<td>PL W392Y</td>
<td>0</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>dTdA-2 + UVC, 1 h</td>
<td>-</td>
<td>PL W392Y</td>
<td>0</td>
<td>0.07</td>
<td>0</td>
</tr>
<tr>
<td>dTdA-2 + UVC, 1 h</td>
<td>UVA, 1 h</td>
<td>PL W392Y</td>
<td>0.05</td>
<td>0.07</td>
<td>69</td>
</tr>
</tbody>
</table>

Note: Reactions were carried out in 20 mM Tris with 2 mM EDTA, pH 7.5. Protein concentrations were 80 nM when used. dTdA-2 absorbance was monitored at 80 °C to melt DNA.

The rate for repair is small due to the overall changes in absorbance but is about 0.04 - 0.05 absorbance units at 260 nm per hour for both PL and PL W392Y, which is approximately a 5% increase per hour. The total noise generated through the spectra would likely be above the signal generated by the changes of the incremental steps in
the repair. The next step for this assay would be to try to make this assay more sensitive by trying a different detection method, such as fluorescence.

### 3.5.5. PL W392Y repairs comparably to PL

The PL mutant, PL W392Y, was initially studied, among other mutants, to help identify which residues contributed to the unusual kinetic stability of the FADsq in PL. PL W392Y was found to destabilize the FADsq and allow for more rapid conversion to FADox. This mutant was also studied *in vivo* by Trasolini et al. and was found to compromise repair function of PL in the *E. coli* strain UNC1085 (see section 1.7). Due to the nature of the *in vivo* assay, it is unclear if this mutant is a functionally compromised form of PL, or simply not being produced or maintained at the same concentrations as wild type in the cells. Because of these unusual properties, this mutant was tested *in vitro* using absorbance to monitor repair (see section 1.7) and crystallized for x-ray crystallography by Dr. David Sanders.

PL W392Y was purified using heparin and DEAE chromatography. CD spectroscopy was performed to verify folding of the mutant. The spectrum was very similar to that of PL, with troughs between 210 nm and 220 nm. In collaboration with Dr. David Sanders, PL was crystallized (21.0 mg/mL), as replicated from a previous PL crystallization. Subsequently, PL W392Y was crystallized for the first time.
Figure 3.14. CD spectra of PL W392Y.
Note: A single broad trough between 210 nm and 218 nm is reminiscent of the CD spectra of PL and CRY. This trough occurs in the region expected for proteins with high α-helical content. PL W392Y in 18 μM concentration.

Using the A$_{260}$ *in vitro* assay (see section 3.5.2), the activity of PL W392Y was probed. PL W392Y was photoreduced before the experiment. The A$_{260}$ of the UV irradiated ssDNA was increased by photoreduced PL and PL W392Y by 0.48 and 0.49, respectively. (Fig. 3.15 and Table 3.1) This initial result was unexpected, as the catalytic site of PL W392Y was suggested to have compromised function.
Figure 3.15. ssDNA repair by PL W392Y.

Note: Left, (dT)$_{18}$ (65 ng/μL) repair by PL W392Y (440 nM). The durations shown are the amount of UVA exposure. Right, a comparison of repair over time by PL and PL W392Y indicates that the PL W392Y is not compromised for ssDNA repair.

In this same assay, irradiated dsDNA was used to probe the function of PL W392Y (see section 3.5.4). The increase in the 260 nm absorbance was 0.05, which is comparable to that of PL. (Fig. 3.16 and Table 3.1) This result was unexpected, as this suggests that the function of PL W392Y is not compromised relative to wild type PL.
Figure 3.16. Repair of dsDNA substrate.
Note: Substrate with PL W392Y (320 nM) with damaged dsDNA (95 ng/μL) before UVA irradiation (solid triangles), with 60 min UVA irradiation (open triangles) or incubated at 60 min without UVA irradiation (solid circles). All measurement of Substrate was performed at 80 °C.

3.5.6. Development of an endonuclease detection method for thymine dimers

To help with detection of rates of repair for PL, a more sensitive detection method was developed. This method involves the use of a 400 bp dsDNA substrate, TT96. This 400 bp dsDNA substrate is encoded on a plasmid, pl-TT96 (synthesized by IDT), which can be readily PCR amplified to produce the desired substrate. (Fig. 3.17)
Figure 3.17. Sequence of TT96 plasmid.

Note: TT96 PCR product and surrounding sequence, top, and pTT96 plasmid map, bottom. M13F and M13R sequences shown for orientation.
This 400 bp substrate can be irradiated by UV light to produce a thymine dimer at the only thymine pair site at position 96 and 97. This substrate contains a MlyI cleavage site at position 96 between the only paired thymines. Substrates containing thymine dimers would be resistant to cleavage by MlyI. Substrates not containing thymine dimers would be susceptible to cleavage by MlyI. The resultant fragments can be resolved by gel electrophoresis, followed by gel purification of the dimer containing 400 bp product. This CPD-TT96 can then be detected by monitoring the proportion of MlyI-cleaved to uncleaved TT96 after repair by PL or PL mutants. This substrate can then be radiolabeled for detection of small amounts of repair.

The substrate was prepared from the plasmid, pI-TT96, successfully by using PCR amplification of this specific region from this plasmid. A pure 400 bp band was resolved by gel electrophoresis. This substrate was radioactively end-labeled with \( ^{32} \text{P} \) and detected by an autoradiogram of the resolved gel electrophoresis. (Fig. 3.18)

**Figure 3.18. Autoradiogram of \( ^{32} \text{P} \) end labeled TT96 PCR product.**

Note: TT96 PCR product end labeled with \( ^{32} \text{P} \), left, or under the same conditions with a 10-fold lower concentration, right. These results show that the TT96 PCR product has been successfully end labeled with \( ^{32} \text{P} \). There are three contaminants, two below and one above the PCR product. The smaller products cannot be visualized at a 10-fold lower concentration.
3.6. Future directions and summary

3.6.1. Future Directions

There are two sets of future experiments that can be undertaken, one if my hypothesis is accepted, or another if my hypothesis is rejected. If the recognition loop is important for CPD binding, then it would be desirable to determine if more than just sequence plays a role in loop dynamics and what loop movements occur before, during and after CPD binding. If the loop is not important for CPD binding, then it would be interesting to see how loops that differ significantly from PL interact with the CPD and to further determine what is important for CPD binding.

If the PL recognition loop is important for binding, then it would be possible to probe, in part, how CRYD functionally diverged from PL. Directed evolution could be used in the CRYD recognition loop to potentially produce a functional PL. I would hypothesize that CRYD is unable to bind CPD solely due to the composition of its loop. I would expect that a functional PL could be created using directed evolution. If CRYD cannot produce a functional PL, then I would conclude that the recognition loop is not the sole property that inhibits CRYD from repairing CPD.

The loop flexibility could also be modulated without changing the sequence using a molecular crowding agent, such as Ficoll or polyethylene glycol\textsuperscript{140}. Here I would hypothesize that the flexibility of the loop is important independent of sequence. If the molecular crowding agent would lower the conformational dynamics of the loop, then it is possible that a difference in function would be observed. I would expect flexible mutants that were previously compromised will become more functional, and the previously optimally functional mutants will become less functional.

If loop is not important in CPD binding, then it would be desired to determine why this is and what is important in the divergence of CRYD from PL function. A crystal structure and a cocrystal structure of a functional PL mutant that shares no common loop residues with PL would be desired, which could be generated through directed evolution. I hypothesize that the loop will change positions, much like wild type PL, but only the backbone of the loop will be interacting with the CPD and other PL residues.
The interactions and position of the loop bound and unbound to substrate would be compared to that of PL.

Another way to determine how CRYD functionally diverged from PL would be to perform directed evolution on the entire CRYD sequence to generate a partially functional PL. I hypothesize that CRYD is compromised in PL function in its binding cleft, then I would expect mutations that give rise to CRYD function to be in the binding cleft.

3.6.2. Summary

The technical advances include in vivo and in vitro repair, in vivo selection, detection of whole protein in vivo and plasmid generation. Degenerate PCR has been designed and tested, which yielded an average of 5 mutations per loop insert. An assay to determine the number of mutations was created, which was consistent with estimations of mutation rates. The generation of mutations through epPCR was designed and tested and the number of mutants was quantified using the endonuclease error detection method. The in vivo selection screen was tested and speed of testing was improved in a spot assay. In vivo expression of a his-tagged PL was confirmed by Western blot. The absorbance monitoring in vitro repair assays were developed and tested on PL and PL W392Y. The endonuclease in vitro repair assay has partially been developed. Limited proteolysis has previously been developed (see chapter 2) to test loop dynamics.

While testing of this hypothesis is incomplete, the framework for this testing is complete. The technical development of many required assays is complete or near completion. There are significant differences in the conformational dynamics of the loop region of these proteins. The framework for testing the role that recognition loop conformational dynamics has in function has been developed.
Chapter 4. Methods

4.1. Protein purification and UV/Vis/CS spectroscopy

4.1.1. PL/PL W392Y purification

PL was expressed and purified using procedures by Todo\textsuperscript{141} and a plasmid provided by Tamada\textsuperscript{12}. PL was measured using an extinction coefficient\textsuperscript{67} of 141060 M\textsuperscript{-1}cm\textsuperscript{-1}. The plasmid pet3AnPhr contains PL under the control of a T7 promoter. This plasmid was transformed into BL21(DE3)pLysS cells which contains T7 polymerase and the T7 lysozyme that supresses the T7 promoter. Inoculation of 1 L of LB-amp was performed using 5 mL of overnight culture. PL was induced using a final concentration of 800 mM IPTG after the culture reached an OD\textsubscript{600} of 0.8 to 1.0. Harvesting was carried out after 3 h of induction by centrifuging the culture at 5,000 rpm for 10 minutes at 4 °C. The pellet was resuspended in 5 mL Buffer B (10 mM potassium phosphate, 100 mM NaCl, ph 7.0) per 1 L bacterial culture. The suspension was frozen in liquid nitrogen and stored at -80 °C.

Lysis of cells was performed by first thawing the frozen resuspension on ice. The resuspension was sonicated three times for 2 mins at 50% amplitude 1 second pulses on ice using a sonicator, Sonic Dismembrator Model 500, from Fisher Scientific. Cell debris was separated from soluble protein by centrifugation at 16,000 rpm for 1 h at 4 °C.

The first two rounds of column purification were performed using HiTrap Heparin HP from GE Healthcare Life Sciences. The supernatant was first filtered using a 0.22 µm syringe filter before loading it onto the column at a rate of 1 mL/min. Elutions were performed using a step gradient of 50 mM NaCl intervals. PL eluted at ~400 mM NaCl. Concentrations and purity were monitored by UV-vis spectroscopy and SDS-PAGE.(Fig.
4.1) Fractions containing PL were diluted with equal volumes of Buffer B and loaded and eluted on the same HiTrap Heparin column. An improvement to the protocol were increasing the gradient up to 400 mM NaCl and eluting PL in 5 column volumes of 400 mM NaCl to reduce the procedure to only 1 heparin column used. (Fig. 4.2)

Figure 4.1 PL purification monitored by electrophoresis.
Note: PL was purified using a series of 3 columns, 2 heparin and 1 DEAE. PL, expected at 55 kDa, was pooled from each set of elutions containing this band and subsequently loaded on to the following column. Each lane except the ladder was loaded with 20 µL, the ladder contains 5 µL PAGE Ruler (Fermantas).
Figure 4.2. **PL W392Y purification monitored by electrophoresis.**

Note: PL W392Y was purified using a series of 2 columns, heparin and DEAE. PL W392Y, expected at 55 kDa, was pooled from each set of elutions containing this band and subsequently loaded on to the proceeding column. Each lane except the ladder was loaded with 20 µL, the ladder contains 5 µL PAGE Ruler (Fermantas).

A single round of DEAE purification was then performed by PL diluting in Buffer A (10 mM potassium phosphate, ph 7.0) the pooled PL containing heparin fractions to ~80 mM NaCl followed by loading onto a HiTrap DEAE Sepharose FF from GE Healthcare Life Sciences. Elution was performed using a step gradient of 10 mM NaCl intervals. PL elutes around 100 mM NaCl. Fractions were assayed for quantity and purity using UV-vis spectroscopy and SDS-PAGE.(Fig. 4.1 and Fig. 4.3) Protein was stored by adding 10% glycerol, final concentration, to the protein in Buffer A, flash frozen, and stored at -80 °C.
Figure 4.3. UV-vis spectra after DEAE purification of PL.

Note: PL (0.639 mg/mL) in eluted from DEAE in column buffer Buffer E (~120 mM NaCl 10 mM KPO$_4$, 5 mM BME, pH 7.4). Peak at 280 nm is a quantitative measure of protein yield. Peaks at 440 nm and 580 nm indicate that there is FADox and FADsq present. Shoulders around 440 nm and 580 nm are characteristic of FAD bound to PL.

4.1.2. CRY$_D$ purification

CRY$_D$ was expressed and purified using a procedure from Todo$^{141}$. CRY$_D$ was measured using an extinction coefficient$^{67}$ of 120320 M$^{-1}$cm$^{-1}$. The plasmid containing the CRY$_D$ gene, pGEX-4t-2-CRY, was provided by Todo$^{141}$. This plasmid was transformed into JM109 which allows for the expression of genes under a TAC promoter. The GST linkage sequence contained a thrombin sequence to allow for removal of the GST. Inoculation was performed using a 5 mL overnight starter culture of pGEX-4t-2-CRY containing JM109 in a total volume of 1 L LB-amp. CRY$_D$ was induced for 5 h using a final concentration of 600 mM IPTG after the culture reached an OD$_{600}$ of 0.6 to 0.8. The culture was then harvested by centrifugation at 5,000 rpm at 4°C. Cell pellets were resuspended in 5 mL buffer CA (136 mM NaCl, 2.7 mM KCl, 0.01 mM Na$_2$PO$_4$, 0.02 mM KH$_2$PO$_4$, 2 mM DTT, and 5% Glycerol, pH 7.4) for every 1 L bacterial culture. The suspension was frozen in liquid nitrogen and stored at -80°C.
Cell lysis was performed by thawing the cell resuspension on ice, followed by three rounds of sonication for 2 mins at 50% amplitude using 1 second pulses on ice using the sonicator Sonic Dismembrator Model 500, from Fisher Scientific. Cell debris was removed by centrifugation at 16,000 rpm for 1 h at 4 °C.

The first round of column purification used a GST column, GSTrap HP from GE Healthcare Life Sciences. Supernatant was syringe filtered through a 0.22 µm filter before column loading at 1 mL/min. Unbound proteins were removed by washing with buffer CA. CRY$_D$ was eluted with Buffer CB (1x PBS, 2 mM DTT, 5% Glycerol, and 50 mM glutathione). Concentration and purity of elutions were verified by UV-vis spectroscopy and gel electrophoresis. (Fig. 4.4) CRY$_D$ was concentrated in Milipore centrifugal concentrator device (MWCO 30 kDa) to ~1 mL volume and cleaved with thrombin (2 U per 1 mg fusion protein) at room temperature for 2 h.

![Figure 4.4. Purification of CRY$_D$ monitored by gel electrophoresis.](image)

**Figure 4.4.** Purification of CRY$_D$ monitored by gel electrophoresis.

**Note:** CRY$_D$ was purified using a series of 2 columns, GST-sepharose and Heparin. GST-CRY$_D$, expected at 70 kDa, or CRY$_D$, expected at 55 kDa, were pooled from each set of elutions containing these bands and subsequently loaded on to the proceeding column. Each lane except the ladder was loaded with 20 µL, the ladder contains 5 µL PAGE Ruler (Fermantas).

The second round of column purification was performed using a heparin column, HiTrap Heparin HP from GE Healthcare Life Sciences. The thrombin cleavage reaction
was first filtered using a 0.22 µm syringe filter before loading onto the column at 1 mL/min. CRY₇ was eluted with NaCl step gradient of 50 mM NaCl intervals in buffer CA. CRY₇ eluted at ~350 mM NaCl. Concentrations and purity was monitored by UV-vis spectroscopy and SDS-PAGE.(Fig. 4.5)

**Figure 4.5.** UV-vis spectra after purification of CRY₇.

Note: CRY₇ (0.854 mg/mL) eluted from Heparin in CRY Elution Buffer (~300 mM NaCl, 10 mM Na₂PO₄, 2.7 mM KCl, 1.8 mM KH₂PO₄, 5 mM BME, pH 7.4). Peak at 280 nm is a quantitative measure of pure protein yield. Peak at 380 nm is typical of MTHF. Small shoulder at 440 nm is typical of FADox bound to CRY₇.
4.1.3. **CD-spectroscopy stability assay**

PL and CRY\(_D\) stability assays were carried out in 10 mM potassium phosphate, pH 7.4. A Jasco J-810 Spectropolarimeter was used with a 0.1 mm quartz cuvette. Spectra between 190 nm and 320 nm was measured. Temperature was varied using a Peltier temperature controller between 20 °C to 70 °C. Concentrations of Urea were varied with an 8 M stock urea in 10 mM potassium phosphate, pH 7.4.

4.2. **Limited proteolysis method**

Limited proteolysis was performed at room temperature in 10 mM Na\(_2\)PO\(_4\) and 150 mM NaCl at pH 7.4. PL and CRY\(_D\) concentrations were 1 µM for timecourse experiments. Protease to substrate ratios ranged from 1:200 to 1:10, while reaction times ranged between 30 seconds to 77 h. SDS experiments were carried out in the same buffer with 0.5% SDS. Reactions were stopped by addition of SDS loading dye (32.9 mM Tric-HCl, pH 6.8, 13.15% (w/v) glycerol, 1.05% SDS) and heating at 97 °C for 3 mins. Samples were analyzed by 7% SDS-PAGE, stained with Coomassie brilliant blue or Sypro ruby red. Gels were scanned using a conventional scanner or Typhoon imager by Amersham Bioscience. Densitometry was performed on gel bands using Unscan-it by Silk Scientific or by ImageQuant by Molecular Dynamics. For both programs, boxes were used to manually select bands to be quantified. Total loss was measured as a percent of a control protein band that lacked protease.

Fragment identification was carried out by LC/MS/MS by Midwest Bioservices. Bands were excised from SDS-PAGE and digested in gel by trypsin. Fragments were extracted from the gel and separated by nanoLC, reverse phase C-18. Detection was performed by electrospray MS/MS. Peptide identification was determined by comparing the predicted masses to the masses generated by the cleaved sequences.
4.3. Fluorescence methods

4.3.1. Fluorescent peptide generation

Fluorescence peptides were synthesized by Biosyntan. Sequences were confirmed by MALDI-TOF mass spectrometry. Purity was determined by HPLC to be greater than 95%. Samples were prepared in 10 mM KH$_2$PO$_4$/K$_2$HPO$_4$, 150 mM NaCl, pH 7.5, or 8 M urea. Concentrations were confirmed by UV-vis spectroscopy using an extinction coefficient at 280 nm of 5600 M$^{-1}$cm$^{-1}$. All samples were diluted to 65 µM.

4.3.2. Fluorescence TC-SPC method

Fluorescence measurements were measured using a Fluorolog 3 spectrometer by Horiba-JobinYvon, equipped with double excitation and emission monochromators, cooled photomultiplier detector (IBH) and electronics for TC-SPC. Steady state measurements used a 450 W xenon-arc lamp as a light source. Emission spectra were collected using the Trp excitation peak at 295 nm, using a 2 nm bandpass, or the DBO excitation peak at 375 nm, using an 8 nm bandpass.

TC-SPC light source was a 375 nm laser diode operating at ~70 ps full width, half maximum at 100 kHz repetition rate. The forward mode was used with 500 ns, 1 µs, or 2 µs total time bound to 2048 channels. Source emission light was passed through a double monochromator, 10-15 nm bandpass centered at 430 nm. Decays were collected after 10,000 counts accumulated in the peak channel. Temperatures were maintained to ±0.1 °C using a Peltier-controlled thermostated sample holder.

Channels were converted to time by dividing the total time by the number of channels. The number of counts in each channel was used as the relative intensity. The CB lifetime data was subtracted from each of the experimental data and fit to a mono or double monoexponential equation.
4.4. Mutagenic PCR

4.4.1. Subcloning pm242-cas

DNA fragments were analyzed by agarose gel electrophoresis, stained with ethidium bromide. Subcloning of pm242-cas, the template for mutagenic PCR, used the plasmid pm242 and the epPCR product or the PCR product using the forward primer: 5'-gtgaatgcagaattccactatggcgtgctgctagc-3' and the reverse primer 5'-gtgaatgcagtgctagccaggccaagaagtttgatgc-3' with the plasmid, pet3AnPhr as the template. Both were cleaved with NheI (10 U) for 1 H at 37 °C in NEB buffer #4. The large fragment of each was purified using QIAgen Gel Extraction Kit. Ligation was carried out using the Roche Rapid Ligation Kit. Ligation mix was transformed into DH5-α by electroporation and plated on ampicillin-LB agar plates. Plasmids were verified by NheI digestion and sequencing.

4.4.2. Degenerate primer PCR

Degenerate primer PCR used a primer purchased from Alpha DNA with the following degenerate sequence: 5'-GTGAATGCAGCTAGCXZEZZJXEZZXEXIJJXXXCCGCTGCGGATTTC-3', where X: 70% A, 10% G, 10% C, 10% T; Z: 10% A, 70% G, 10% C, 10% T; E: 10% A, 10% G, 10% C, 70% T; J: 10% A, 10% G, 70% C, 10% T. PCR was performed using this forward primer and the reverse primer, PL-rl-r (GTG AAT GCA GCT AGC AGG ATT GAA AAT CCG CAG CG). PCR conditions used 97 °C for 5 minutes, followed by 30 cycles of 55 °C for 60 seconds, 72 °C for 30 seconds, and 90 °C for 60 seconds.

The BamHI mutation rate assay was performed using BamHI (0.5 U/µL) from NEB. The PCR product was incubated with BamHI in NEB buffer #4 at 37 °C for 1 h. The digest was analyzed by agarose gel electrophoresis, stained with ethidium bromide.

4.4.3. epPCR

The GeneMorph II kit from Agilent was used for epPCR. PCR was performed using PL-rl-f (5'-GTGAATGCAGCTAGCAGTGGCATGGATCCAAAACC-3') and PL-rl-r
(5'-GTGAATGCAGCTAGCAGGATTGAAAATCCGCAGCG-3'). PCR conditions started with 97 °C for 5 minutes, followed by 30 cycles of 55 °C for 60 seconds, 72 °C for 30 seconds, and 90 °C for 60 seconds.

4.5. Tagged PL expression under TAC promoter

4.5.1. Subcloning pPLeL – performed by MBB Service Centre (Ziwei Ding)

The plasmid pPLeL was subcloned using pGEX-4t-2-CRY and pet3AnPhr. pGEX-4t-2-CRY was PCR amplified using the primers HindHispGEXEcoN3R (5' -CACGAAGCTTGAATGATGGTGATGGTGATGACCTAGTATAGGGGACATGAATACTG-3') andpGEX-Xho5F (5' -TAGACTCGAGCGGCCGCATC-3') to allow for HindIII and XhoI cleavage. Plasmid pet3AnPhr was PCR amplified using the primers XhoLPCterm3R (5' -CCGCTCGAGTCTAGAGGATCCCCAACCAAC-3') and Hind3thrombinLP5F (5' -GTTCAGCTTGATGCCACGCGTTCTATGGCGGCTCCGATTCTG-3') to allow for HindIII and XhoI cleavage. These products were cleaved with XhoI and HindIII and ligated to form pPLeL.

4.5.2. Western blot

Western blots were performed using whole cell lysate from pPLeL, expressed in the same way as in PL purification. The whole cell lysate was separated using 7% SDS-PAGE analysis and the proteins were transferred onto nitrocellulose paper using the Biorad Mini Transblot system at 4 °C and 80 V for 1 h. The membrane was blocked using 5% milk (w/v) powder and 3% BSA (w/v) in TBST (0.1% (v/v) Tween-20, 50 mM Tris, 150 mM NaCl, at pH 7.6) with 0.00002% (w/v) sodium azide for 1 h in the dark. This was followed by primary antibody incubation of antihis-tag Mouse mAb (Cell Signalling Technology) at 4 °C overnight. The membrane was washed 5 times with 1X TBST for 10 minutes each. Next the membrane was incubated for 40 minutes with secondary antibody, goat anti-mouse HRP (Pierce Antibodies). The membrane was then washed with 1X TBST three times for 5 minutes each. The membrane was visualized using west
pico, 1:1-peroxide and luminol from life technologies and imaged using Las-4000 Chemiluminescence Scanner (GE Life Sciences).

4.6. *In vivo* functional assays

The *in vivo* functional assays used UNC-1085 *E. coli* cells transformed with pet3AnPhr, pGEX-4t-2-CRY, pGEX-4t-2, or pPLeL. Cell culture was grown overnight in LB-amp without induction. The OD$_{600}$ was determined using a UV-vis spectrometer and the culture was serially diluted 10-fold, 8 times. The dilutions were spread on LB-amp plate, or spotted using a multi pipettor in 2 µL spots on LB-amp plates. The plates were exposed or not exposed to UVC (UVP UVGL-15 compact UV lamp 254 nm), at a distance of 65 cm with a 25% filter, for 30 seconds or less; the time was adjusted when changing fluence. These plates were then exposed or not exposed to UVA lamp (entela UVL56 Black-Ray Lamp 365 nm), at a distance of 32 cm with four, 10% filters, for 30 mins for photorepair. Plates were then shielded from light by aluminum foil and grown overnight at 37 °C.

4.7. *In vitro* A$_{260}$ functional assays

The *in vitro* A$_{260}$ functional assay was performed using (dT)$_{18}$ and dsDNA oligonucleotides from Invitrogen. To perform this assay with dsDNA, the dA strand was added in 5% excess concentration and absorbances were measured at 80 °C using a Cary Temperature Controller connected to a Cary Varian 300 spectrophotometer.

Oligos were dissolved in TE buffer (Sigma-Aldrich) and concentrations were verified by UV-vis spectroscopy. Induction of dimers was carried out in a 96 well plate for 3-4 h using a UV transilluminator by Fotodyne. Water was added to compensate for loss of volumes due to evaporation. Absorbance decreases were monitored by UV-vis spectroscopy.

PL and PL W392Y were photoreduced before use. Enzyme at a concentration of 5 µM was prepared in PBS (136 mM NaCl, 2.7 mM KCl, 0.010 mM Na$_2$PO$_4$, and 18 mM
KH$_2$PO$_4$, pH 7.4) with 12.5 mM DTT. It was then irradiated with white light from a projector 10 cm from the sample for 6 minutes on ice in a 1 cm pathlength quartz cuvette. Spectra of the sample were taken before and after photoreactivation and a loss of the 580 nm peak was used as an indicator of phororeduction.

Damaged DNA was incubated with PL in a well of a 96 well plate and exposed or not exposed to UVA lamp (Entela UVL56 Black-Ray Lamp 365 nm) at a distance of 32 cm with four, 10% filters. Absorbance increases were monitored by UV-vis spectroscopy.

4.8. In vitro endonuclease protection functional assays

TT96 was PCR amplified using the forward primer 5'-GGTACCTATGGAAGATGTAT-3' and the reverse primer 5'-AGATCTCCCCGATAGGAG-3' with the template as a plasmid pI-T96.

DNA OptiKinase (MJS Biolynx), at a final concentration of 1 U/µL, was used to radiolabel TT6 with 5 Mbq of $^{32}$P ATP. DNA was purified through ethanol precipitation. Samples were analyzed using a 5% acrylamide gel and exposed on a storage phosphor screen for 45 mins before imaging on a Typhoon imager by Amersham Bioscience.
References


