Title
Structural And Functional Characterization Of A Bacterial Photosensing Light-Oxygen-Voltage (LOV) Protein Domain From Rhizobium Leguminosarum

Permalink
https://escholarship.org/uc/item/23b297c7

Author
Mednick, Gabriel Franchesco

Publication Date
2016

Peer reviewed|Thesis/dissertation
UNIVERSITY OF CALIFORNIA
SANTA CRUZ

STRUCTURAL AND FUNCTIONAL CHARACTERIZATION OF A BACTERIAL PHOTOSENSING LIGHT-OXYGEN-VOLTAGE (LOV) PROTEIN DOMAIN FROM RHIZOBIUM LEGUMINOSARUM

A dissertation submitted in partial satisfaction of the requirements for the degree of

Doctor of Philosophy

in

CHEMISTRY

by

Gabriel Mednick

June 2016

The Dissertation of Gabriel Mednick is approved:

________________________________________________________
Professor Olóf Einarsson, Chair

________________________________________________________
Professor Roberto Bogomolni, PI

________________________________________________________
Professor Ted Holman

Dean Tyrus Miller
Vice Provost and Dean of Graduate Studies
# Table of Contents

List of Figures vi

List of Tables xii

Abstract xiii

Dedication xvii

Acknowledgments xviii

1 Introduction 1
   1.1 Light as Life energy ................................. 1
   1.1.1 Light harvesting mechanisms ......................... 4
   1.1.2 Light as a signal .................................... 8
   1.2 Light-Oxygen-Voltage domains (LOV) .................... 12
      1.2.1 Discovery ........................................... 12
      1.2.2 LOV domain Structure ............................... 17
      1.2.3 Mechanism of photochemistry ......................... 19
      1.2.4 Effector domains and signal transduction pathways (STPs) 23
   1.3 Research goals: a winding path .......................... 31

2 Spectroscopic characterization of a LOV-HK from *Rhizobium leguminosarum* 44
   2.1 Introduction ............................................. 44
      2.1.1 Photosensing in *Rhizobia* ........................... 46
      2.1.2 The general stress response (GSR) pathway ............ 52
      2.1.3 Optical spectroscopy in the study of LOV domains .... 55
   2.2 Materials and methods .................................... 56
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.2.1 Gene cloning and protein expression</td>
<td>56</td>
</tr>
<tr>
<td>2.2.2 Inclusion body problem</td>
<td>61</td>
</tr>
<tr>
<td>2.2.3 Protein expression and purification strategies</td>
<td>66</td>
</tr>
<tr>
<td>2.2.4 Instrumentation</td>
<td>68</td>
</tr>
<tr>
<td>2.3 Results</td>
<td>68</td>
</tr>
<tr>
<td>2.3.1 Absorption spectroscopy</td>
<td>69</td>
</tr>
<tr>
<td>2.3.2 Fluorescence kinetics</td>
<td>79</td>
</tr>
<tr>
<td>2.3.3 Circular dichroism (CD)</td>
<td>81</td>
</tr>
<tr>
<td>2.4 Discussion</td>
<td>82</td>
</tr>
<tr>
<td>3 Structural characterization of Rh-LOV using X-ray crystallography</td>
<td>92</td>
</tr>
<tr>
<td>3.1 Introduction</td>
<td>92</td>
</tr>
<tr>
<td>3.1.1 Requirements and procedures of crystallography</td>
<td>97</td>
</tr>
<tr>
<td>3.2 Materials and methods for crystallography</td>
<td>100</td>
</tr>
<tr>
<td>3.2.1 Protein preparation</td>
<td>100</td>
</tr>
<tr>
<td>3.2.2 Crystallization conditions</td>
<td>101</td>
</tr>
<tr>
<td>3.2.3 X-ray diffraction experiment</td>
<td>105</td>
</tr>
<tr>
<td>3.2.4 Solving and refining the structure</td>
<td>108</td>
</tr>
<tr>
<td>3.3 Results</td>
<td>113</td>
</tr>
<tr>
<td>3.3.1 Presentation of Rh-LOV structure at 1.89 Å</td>
<td>113</td>
</tr>
<tr>
<td>3.3.2 Protein-FMN interactions in the active pocket and connections to the C-terminus</td>
<td>120</td>
</tr>
<tr>
<td>3.3.3 Rh-LOV and Br-LOV dimerization interface differences</td>
<td>124</td>
</tr>
<tr>
<td>3.3.4 Rh-LOV vs lit state LOV structures</td>
<td>129</td>
</tr>
<tr>
<td>3.4 Discussion</td>
<td>134</td>
</tr>
<tr>
<td>4 Electron paramagnetic resonance (EPR) experiments</td>
<td>145</td>
</tr>
<tr>
<td>4.1 Introduction</td>
<td>145</td>
</tr>
<tr>
<td>4.2 Materials and methods</td>
<td>154</td>
</tr>
<tr>
<td>4.2.1 Design of mutations for DEER</td>
<td>154</td>
</tr>
<tr>
<td>4.2.2 Labelling reaction</td>
<td>155</td>
</tr>
<tr>
<td>4.2.3 Instrumentation</td>
<td>157</td>
</tr>
<tr>
<td>4.3 Results</td>
<td>158</td>
</tr>
<tr>
<td>4.3.1 Features of Rh-LOV surface secondary structure observed by CW EPR</td>
<td>158</td>
</tr>
<tr>
<td>4.3.2 Specific DEER distance measurements confirm that the dimer interface is conserved in solution.</td>
<td>162</td>
</tr>
<tr>
<td>4.3.3 Rh-LOV-S22R1 provides information about interactions between the N-termini in solution.</td>
<td>168</td>
</tr>
</tbody>
</table>
List of Figures

1.1 (A) The EM radiation spectrum is a range of discrete energies that exhibit wave properties, such as wavelength and frequency. (B) Dimeric PSII (PDB:2axt) contains 18 protein subunits per monomer and has 17 unique chromophores (over 100 total) that enhance its efficiency and absorption in the UV-IR regions of the EM spectrum. .......................................................... 6

1.2 (A) Bacteriorhodopsin’s (PDB: 1c3w) light sensitivity and signalling mechanism depend on the retinal chromophore. (B) Retinal undergoes an all-trans to 13-cis isomerization after green light absorption, which propagates to the protein structure and triggers proton pumping. ........................................................... 8

1.3 Bacterial photosensors and their chromophores: (A) cryptochrome (1np7); FAD. (B) LOV (1g28); FMN. (C) Bacteriophytochrome (2o9c); biliverdin. (D) Bacteriorhodopsin (1c3w); retinal. (E) PYP (2phy); p-coumaric acid. (F) BLUF (1yrx); FAD. .......................... 11

1.4 (A) Depiction of Thimann’s action spectrum for the phototropic effect in plants. (B) Demonstration of the phototropic effect in mouse-ear cress seedlings. After knocking out the LOV-containing genes, phot1 and phot2, the seedlings no longer bend towards the light (image adapted from Christie and Murphy, 2013). ............ 13

1.5 In the PAS subfamily, LOV domains include three unique sensors that detect light, oxygen or voltage. Light sensing LOV domains bind FMN, oxygen sensors bind heme, while voltage sensors do not have a bound chromophore. ................................. 15
1.6 Structural comparison of two PAS domains, LOV2 (left: 1G28) and PYP (right: 1NWZ), reveals the conservation of tertiary features, including the PAS core, the β-sheet scaffold and the helical connector. The N-terminal cap is a variable feature that is not present in all PAS domains.

1.7 The LOV photocycle has two major states, which are distinguished by noncovalent (dark) and covalent (S-C4a) attachment (adduct) of the FMN moiety.

1.8 A representation of the LOV2-RAC1 (2WKP) chimeric protein signalling mechanism occurring via dynamic changes in the Jα-helix—the ‘clam shell’ model. Light activation of RAC1 in vivo resulted in localized cell protrusion and changes in motility.

1.9 Four common downstream effector domains that can be regulated by LOV: (A) Diguanylate cyclase (blue), (B) Histidine kinase (purple); PhyR (pink), (C) STAS (hot pink), (D) helix-turn-helix (orange).

1.10 (A) GGDEF and EAL domains regulate the concentration of the secondary messenger, c-di-GMP, in response to environmental signals (adapted from Jenal and Malone 2006). (B) Two-component STPs involve kinase activation via autophosphorylation, followed by phosphotransfer to a cognate response regulator (RR) with specific targets in the cell.

1.11 (A) Proposed structural basis of signal propagation involves changes in the helical region linking sensor and kinase. (B) Regions of a canonical histidine kinase include N, G1, F and G2 boxes.

1.12 Research goals with Rh-LOV: (A) Aim 1. Characterize temporal photocycle properties. (B) Aim 2. Determine Rh-LOV structure and investigate its signalling dynamics. (C) Aim 3. Identify and characterize the downstream RR in this pathway.

2.1 Root nodules house Rhizobia, which fix nitrogen for the plant in exchange for a steady supply of sugars and nutrients.

2.2 (A) A comprehensive histogram of all STPs in Rle. (B) Rh-LOV-HK is categorized as a transmitter. (C) The Rh-LOV-HK gene (pRL110320) and the gene product for a PAS-HWE-HK (pRL110319) on the complementary strand share nearly identical secondary structure. Images generated in MiST2 (L. Ulrich and I. Zhulin, 2010).
2.3 In α-proteobacteria, the GSR pathway responds to different forms of stress, including environmental and chemical factors. This figure provides a model of how light stimulus leads to activation of the GSR pathway (image adapted from Sycz et al.).

2.4 LIC methodology uses sticky end cohesion, rather than annealing gene insert and vector using a ligase. This provides a more efficient approach to cloning (QB3 MacroLab).

2.5 Use of fluorescence emission to visualize LOV protein in the dark-state. Blue light irradiation causes formation of the photoadduct, which leads to fluorescence bleaching.

2.6 Full-length and Rh-LOV domain constructs with amino acid numbering. Rh-LOV-160 was the primary construct that was characterized in this thesis.

2.7 Rh-LOV absorption spectra (A) before and (B) after flash photolablation.

2.8 Rapid absorption kinetics at logarithmically spaced time intervals from 10 µs to 5 ms for (A) Rh-LOV and (B) Rh-LOV-HK.

2.9 Absorption kinetics between 1 µs and 5 ms was used to calculate the branching ratio for (A) Rh-LOV. (B) Rh-LOV-HK was observed to have similar time dependence but lacked stability after the laser excitation flash.

2.10 Distances between the active site cysteine and FMN: (A) Rh-LOV. (B) Br-LOV. (C) EL222. (D) Oat LOV2.

2.11 Adduct decay after photobleaching for (A) Rh-LOV and (B) Rh-LOV-HK. The dark state sample was used for the baseline spectrum. After flash irradiation, UV-vis absorption spectra were collected at 80 s intervals for ~7000 s.

2.12 (A) Rh-LOV-HK fluorescence before and after light activation. (B) Rh-LOV fluorescence before and after light activation. (C) Rh-LOV-HK time-resolved fluorescence return after photoexcitation. (D) Rh-LOV time-resolved fluorescence return after photoexcitation.

2.13 (A) Dark and excited state CD spectral comparison. (B) CD difference spectrum.

2.14 Rh-LOV-HK detailed photocycle model showing the dark, triplet-excited and adduct states with the duration of each step.

3.1 Example of a 24-well optimization tray: Starting with the crystal hit conditions, the pH and precipitate concentrations are varied. More than one protein concentration can also be tested by setting multiple drops in each well.
3.2 Image of Rh-LOV crystals from an optimization tray. The expanded view gives a scaled perspective of crystal size.

3.3 (A) A crystal loop was used to isolate a single crystal. The loop-crystal complex was then stored in a cryoprotectant, shipped to ALS and aligned remotely on the goniometer for X-ray diffraction. (B) A single diffraction image contains a regular array of spots, referred to as reflections. From the symmetry of the diffraction pattern, the space group can be determined.

3.4 (A) and (B) Rh-LOV crystallized in the *P*6₃ space group which is defined by a hexagonal repeating unit cell. The Rh-LOV dimer represents the basic repeating asymmetric unit. (C) This view shows the electron density fit over a portion of the structural model in the unit cell.

3.5 (A) Electron density of Rh-LOV (without structural model) shows high resolution details of FMN’s isoalloxazine ring structure. (B) A sample of the structural model (side chains not shown) demonstrating its fit into the electron density.

3.6 Presentation of Rh-LOV tertiary structure at 1.89 Å resolution (above), and the Rh-LOV primary sequence (below). The secondary structure is labeled alphabetically from the N- to C-terminus.

3.7 (A) Global structural comparison of Rh-LOV (green), Br-LOV (blue), EL222 (salmon) and Oat-LOV2 (purple) demonstrates the high degree of structural conservation. (B) Same view, emphasising conservation of protein interactions that bind FMN.

3.8 (A) Polar interactions with the FMN side chain. (B) Hydrophobic interactions with the isoalloxazine ring. (C) View of Gln 139 in the Iβ-strand leading to the C-terminus. (D) A view of residues in the C-termini (orange) interacting with a glycerol molecule (pink).

3.9 H-bonds that stabilize FMN in the active pocket: (A) Rh-LOV. (B) Br-LOV. (C) EL222.

3.10 (A) Dimer orientation and electrostatic distribution of charges on the surface of Br-LOV. (B) Dimer orientation and electrostatic distribution of charges on the surface of Rh-LOV.

3.11 (A) Rh-LOV dimerization interface (red: interaction interface). (B) Br-LOV dimerization interface (red: interaction interface). (C) Close up view of the N-terminal interactions that stabilize the Rh-LOV dimer.
3.12  (A) Structural comparison of Rh-LOV dark state (blue) with Ytva dark state (purple) and Ytva adduct state (brown). (B) Comparison of Rh-LOV dark state (blue) and Ytva adduct state (brown). (C) Close-up view of Rh-LOV dark state (blue) and Ytva adduct state (brown). (D) Same view but with emphasis only on the position of conserved residues that form the active pocket. 

3.13  A decrease in relative voltage after light irradiation of Rh-LOV crystals, followed by a slow return, corresponds to LOV’s photocycle characteristics of adduct formation and breakage. However, our instrumentation limited our ability to follow these changes and determine if they were on the same time scale as Rh-LOV’s photocycle.

4.1  (A) The basis of the EPR experiment involves microwave absorption in a strong magnetic field. This results in a spin state population inversion followed by precession back to the ground state. By convention, the acquired absorption spectrum is converted into a plot of its first derivative. (B) This schematic illustrates the MTSSL conjugation reaction with a cysteine residue, which is often mutated into the target region of the protein sequence.

4.2  (A) DEER experiments require the presence of two unpaired electrons to produce a dipolar coupling that has a range of $\sim$20-60 Å. (B) A pulse sequence is used to produce DEER modulation echo in time, which can be Fourier transformed into a function of frequency (Pake pattern) and normalized to produce a distance probability function.

4.3  The CW EPR spectra for different positions of the spin label (R1) in the protein correspond to the local environment of the label. The magnetic field (x-axis) for the different spectra are slightly offset for viewing. In each case, multiple conformations of the label are shown to illustrate its range of motion.

4.4  DEER distance measurement and distance measurement prediction based on the crystal structure between positions (A) S84R1-S84R1, (B) V88R1-V88R1 and (C) S102R1-S102R1.

4.5  (A) The Br-LOV DEER distance distribution is consistent with the distance between predicted label positions in the crystal structure. (B) In EL222, the DEER distance distribution lacks correspondence with the distance measurement in the crystal structure.
4.6 (A) Rh-LOV with model of the N-terminal extension based on the experimental DEER distance measurement for Rh-LOV-S22R1. (B) Rh-LOV-S22R1 DEER distance measurement. (C) Background subtracted DEER time trace. 

4.7 (A) Rh-LOV-S22R1 light vs dark state distance probability distributions and (B) light vs dark four-pulse DEER data. (C) Rh-LOV-V88R1 light vs dark state distance probability distributions and (D) light vs dark four-pulse DEER data. (E) Rh-LOV-S102R1 light vs dark state distance probability distributions and (F) light vs dark four-pulse DEER data. 

4.8 (A) EL222-Y125R1 light vs dark state distance probability distributions and (B) light vs dark four-pulse DEER data. (C) Br-LOV-L133R1 light vs dark state distance probability distributions and (D) light vs dark four-pulse DEER data. 

5.1 Summary of the proposed mechanism for light regulation of the GSR pathway in *Rle*. 

5.2 Overview of Rh-LOV experimental results: (A) The photocycle dynamics were temporally resolved. (B) The crystal structure was solved at 1.89 Å. (C) DEER was used to calculate select distances in the solution structure. 

5.3 Amino acid sequence comparison between Rh-LOV-HK and El-LOV-HK346 reveals conservation of secondary structure throughout both LOV and HK domains. 

5.4 Full-length Rh-LOV-HK models generated in Modprod using Rh-LOV and the HK from EL346: (A) Model A shows the linker region stretched out to spatially separate the LOV and HK domains. (B) Model B presents interactions between LOV and HK domains. (C) Model C shows another possible configuration through which the domain surfaces might interact. 

A.1 HSQC spectra for (A) Rh-LOV (160) dark state, (B) adduct state, and (C) Rh-LOV (160) vs (D) Rh-LOV (142).
## List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Soluble expression of various constructs of Rh-LOV and Rh-LOV-HK fusion proteins that were cloned for experimental characterization. Only soluble constructs were suitable for experimentation.</td>
<td>64</td>
</tr>
<tr>
<td>2.2</td>
<td>The rate of adduct formation relative to the average distance of S to C4a (FMN) in the dark state.</td>
<td>76</td>
</tr>
<tr>
<td>3.1</td>
<td>Conditions from the MCSG screen that were favorable to Rh-LOV crystal formation.</td>
<td>102</td>
</tr>
<tr>
<td>3.2</td>
<td>This list, generated in Phenix refine, includes pertinent experimental details and R values, which were used to assess the reliability and quality of the structural model that was produced from the diffraction data.</td>
<td>115</td>
</tr>
<tr>
<td>4.1</td>
<td>Cysteine substitution mutations were engineered for EPR and DEER experiments. Only the mutants listed in green were successful in DEER experiments. The reasons for this are discussed in the text.</td>
<td>155</td>
</tr>
</tbody>
</table>
Abstract

Structural and functional characterization of a bacterial photosensing Light-Oxygen-Voltage (LOV) protein domain from \textit{Rhizobium leguminosarum}

by

Gabriel Mednick

Light-Oxygen-Voltage (LOV) domains are an important family of sensors that can be found in bacteria, archaea and eukaryotes. This thesis focuses on the blue light-sensing LOV domains and their signal transduction pathways (STPs) in bacteria. Specifically, a LOV domain from \textit{Rhizobium leguminosarum (Rle)} is characterized. Biophysical characterization of the protein’s cyclic light response (photocycle) and structure were addressed in order to further our understanding of this important signalling pathway.

Chapter 1 introduces the fundamental characteristics of light sensing with a specific emphasis on bacterial LOV domains. After introducing these sensors and their associated signalling pathways, Chapter 2 focuses on a LOV-histidine kinase (HK) from \textit{Rle}. The importance of \textit{Rhizobium}-plant symbiosis and previous research on the \textit{in vivo} characterization of the Rh-LOV-HK STP are summarized. Then, the cloning, protein expression and purification methods that were used throughout this thesis research are described. In the results section, various
optical spectroscopy experiments that were used to characterize the photocycle properties of Rh-LOV-HK and the Rh-LOV domain in isolation are presented. Experiments included time-resolved flow-flash absorption measurements to measure early changes in the photocycle, and absorption and fluorescence spectroscopy to characterize the average light-activated (adduct) state lifetime. Chapter 2 concludes with circular dichroism (CD) experiments, which can provide insights into the conformational changes that lead to the active signalling state.

In Chapter 3, the experimental procedures that were used to crystallize and solve the structure of Rh-LOV at 1.89 Å resolution are described. Analysis of this high-resolution structure includes a close look at interactions with FMN in the active pocket, the dimerization interface and the position of the C-terminal region. For comparison, the structure of Rh-LOV is contrasted with previously crystallized bacterial LOV domain structures. These homologous LOV domains can have distinctly different temporal photocycle characteristics, which may arise from small differences in structure. Finally, we discuss the experimental procedures that were carried out with the intention of resolving the light-activated crystal structure of Rh-LOV.

In Chapter 4, the structural-functional characterization of Rh-LOV using electron paramagnetic resonance (EPR) techniques is presented. At various positions in the Rh-LOV structure, we show that changes in the continuous wave (CW)
EPR spectrum of a protein-bound paramagnetic label corresponds to the localized protein environment. Using the pulsed EPR technique, double electron-electron resonance (DEER), distance probability distributions were determined between various positions in the Rh-LOV solution structure. We show that it is a dimer and compare the solution and crystal structure dimer orientations. The results from DEER experiments that were undertaken to resolve conformational dynamics of the light-activated signalling state of Rh-LOV are also discussed.

Chapter 5 summarizes the results of this research and considers what we have learned about Rh-LOV-HK and its STP. We will also return to research obstacles that were encountered in this project and propose an alternative approach for isolating full-length Rh-LOV-HK. In conclusion, we consider the obstacles in producing an accurate model of full-length Rh-LOV-HK structure and discuss the trajectory of future LOV domain studies.
To Meher Baba

for your inner guidance in this outer play

xvi
Acknowledgments

Thanks to entropy for explaining why the uphill trek takes a little extra and the downhill gives a little less. Thanks to friends and family for making the hills in life worth traversing. Thanks to inspiring people for showing what life can be at its best. Thanks to the students in life who climb the hills simply to see the Mystery in each step. Thanks to the many professors who have shared plenty and inspired this entry into the path of knowledge in my life.

I would like to acknowledge my wife for her love, patience, support and great company during these graduate school years. Thank you, honey, for keeping a level head and leavened heart in life. I look forward to many more years of love and life adventures by your side.

I would like to acknowledge my father’s mother who was most eager to see me finish this PhD. I am sorry that you could not be here to call me doctor. I miss you grandma Mini. I would like to acknowledge my two big brothers, Flint and Virgil, who have taught me a lot. My mom and dad, Judy and Joel, I love you very much and thank you for the support and experiences you have shared with me over the years.

I would like to acknowledge my friends who have helped me to stay social and well rounded, even while immersed in my research, teaching and studies. I would like to give special shout out to my many beach volleyball friends who have
helped me stay active and gave me a chance to practice team work. Without the spirit of competition, practice and play, graduate school would have been much less balanced and a lot less fun.

I would like to thank my PI, Roberto Bogomolni for encouraging me to dream and giving me the opportunity to do so. Your life-long dedication to science and teaching is an inspiration. I also owe a very special thanks to my basement lab mate, Rafael Silverman y de la Vega, for many humorous interludes, life discussions, and technology startup brainstorming sessions to plan ‘positive change in the world’. I would also like to acknowledge and thank all of the undergraduates who entrusted me with their research training and participated in my research in both big ways and small—I hope you learned as much from me as I did from you.

I would like to thank the Partch lab, the Rubin lab and the Millhauser lab. I would especially like to thank Dr Eric Evans, Dr Chie Funatogawa, Dr Eefei Chen, Dr Sarvind Tripathi and Professors Seth Rubin, Carrie Partch, Bill Scott and Chad Saltikov for your invaluable guidance. I know I asked for a lot of assistance and borrowed a lot of reagents—I apologize. And thank you in the same breath for your patience and generosity. Your assistance in my research and learning process was invaluable.

I would like to thank my chair, Professor Olöf Einarsdóttir and my committee member, Ted Holman. You were both very patient and supportive. I also want to
thank presiding chemistry chair, Professor Ilan Benjamin for inspiring me to go for a PhD. I also have to give a very special thanks to Janet Jones, graduate adviser extraordinaire. Without you, there would be fewer PhD’s in chemistry. Lastly, I would like to take a moment of silence for the late Professor Bob Ludwig—thank you—your life and powerful teaching approach inspired myself and many others to strive to be modern day Renaissance scientists who are curious about all aspects of science.

Finally, I would like to acknowledge Hazratbai and Wustaji Nadel, and Devapa Modadda for being part of my family and for encouraging me on this educational journey. You are always in my heart and memories. I also want to thank Scott and Anne Makeig for your encouragement and love—you are incredible inspirations. Lastly, I would like to thank Nature in all of Her manifestations. The purpose of this research endeavor was, ultimately, to know and understand You better.
Chapter 1

Introduction

1.1 Light as Life energy

What is Life? In its most basic form, Life can be simplified to a single cell. As the fundamental unit of Life, the cell’s many internal processes are the parameters that define Life. From the perspective of science, we strive to build an understanding of Life based on the details of its internal processes. Through these small window glasses, we try to understand and make predictions about something greater—how the parts work together as a whole. Taking the cell apart is a challenging work-in-progress that requires countless detailed studies. Putting it back together is an even greater, possibly insurmountable, task.
Every aspect of Nature is a curiosity, but Life holds a unique and personal mystery. An almost insatiable curiosity about what and who we are has fueled philosophical debates and scientific experiments alike. The discipline of biophysical chemistry is one example. Using the rational approach of the scientific method and tools from chemistry, molecular biology and spectroscopy, the biochemist designs experiments to resolve the minute details of life’s biomolecules—capturing their form and function. Through illuminating the cell’s molecular machinery and its thermodynamic feats, the biochemist delves into the fundamental details that makes life and all of its transformations possible.

Three to four billion years ago, the spontaneous upwelling of life required an almost unimaginable serendipity. One possibility is that the precursor molecules of life—nucleotides, amino acids and lipids came together in an aqueous medium to form early protocells (Deamer 2005). These biological precursors would need a mechanism for replication, and a source of energy to facilitate the increasing organization and complexity that life exhibits (Urey 1952). With its mechanisms to create greater complexity, life seemingly defies the second law of thermodynamics. A closer look reveals that the cell’s ability to decrease its entropy is achieved by extracting energy from sources that are available in the cell’s environment. Chemical energy is inherently stored in high energy molecular bonds, and complex metabolic processes, such as glycolysis, can generate biological energy
through a series of oxidation-reduction reactions. This process requires a limited resource, such as a carbohydrate to oxidize, and the energy extraction requires many enzymes that must work in tandem or as part of a macromolecular complex.

Light is another form of energy that could have fueled early cellular life. Electromagnetic (EM) radiation, the most prevalent energy source on the planet, is characterized by its wave-particle duality. As a particle, the photon’s discrete energy quantity is proportional to its frequency (Figure 1.1 A). For Life to capture and convert this energy, the first requirement is a mechanism to absorb radiation from specific regions of the EM spectrum. All known biological mechanisms of light energy conversion are protein based—what makes light-harvesting proteins unique is that they require the attachment of one or more aromatic small molecules (Figure 1.1 B). These prosthetic groups, referred to as chromophores, have π electrons that are delocalized across the network of π bonds. Chromophores with unique π bond configurations undergo electronic energy level transitions by absorbing specific EM frequencies ranging from ultraviolet (UV) to infrared (IR) (Purcell and Crosson 2008). Light harvesting-energy conversion protein complexes are highly conserved and may be as old as Life itself. Today, light-harvesting energy conversion remains essential for all life to succeed, develop and diversify (Deamer 1997).
1.1.1 Light harvesting mechanisms

Photosynthesis, a mechanism that is conserved in algae, bacteria and plants, uses several different light-harvesting chromophores to absorb and funnel photons to a reaction center of one or both photosystems (PSI and PSII). Photosystems are nearly perfect in their efficiency of photon capture and conversion (Calhoun et al. 2009). After the absorption of a photon, an excited electronic state is generated at the reaction center. This excited electron is strategically removed from the reaction center and transferred to an electron carrier or energy storage molecule such as $NADP^+$ or plastoquinone (PQ), which contributes to the proton gradient that is used to produce ATP from ADP + $P_i$. The NADPH and ATP products are then utilized by enzymes, which couple the oxidation of these molecules to the many anabolic processes of the cell. In this way, the free energy of these transformations can be harnessed to drive the many uphill reactions of the cell.

Photosynthetic organisms have evolved in their specific light absorption patterns, allowing them to exploit unique wavelengths of the EM spectrum. This selectivity is based on whether both PSI and PSII are present and working in tandem. It also depends on the variety and abundance of different chromophore molecules that are bound to the photosystems. PSII absorbs at a slightly higher frequency (and energy) than PSI. As a result of PSII’s stronger ionization energy, it can extract electrons from water. This is achieved by the associated water
splitting-oxygen evolving complex, which is responsible for the majority of atmospheric oxygen (Figure 1.1 B). Some organisms, such as purple bacteria, use a single PSII-like photosystem that lacks the water splitting complex. In this case, small molecules are oxidized to provide the replenishing electron source.
Figure 1.1: (A) The EM radiation spectrum is a range of discrete energies that exhibit wave properties, such as wavelength and frequency. (B) Dimeric PSII (PDB:2axt) contains 18 protein subunits per monomer and has 17 unique chromophores (over 100 total) that enhance its efficiency and absorption in the UV-IR regions of the EM spectrum.

The light harvesting energy mechanisms of PSI and PSII are universal—with only one known exception. In Halobacteria, from the domain archaea, a unique form of light-energy harvesting has evolved. In this process, a membrane protein,
bacteriorhodopsin (Figure 1.2 A), acts as a light driven proton pump. Bacteriorhodopsin consists of an integral membrane protein and a protonated Schiff base linkage to its chromophore, retinal. When protein-bound retinal absorbs green light, it undergoes an all-trans to 13-cis isomerization (Figure 1.2 B). This transformation alters the protein in such a way that protons get pumped unidirectionally to the outside of the membrane. Like PSII, this production of a proton gradient is then used for chemiosmotic energy generation via ATPases. The ATPases are integral membrane protein complexes that generate ATP from ADP and P\textsubscript{i} using the controlled flow of protons in the direction of lower chemical potential as the driving force.
1.1.2 Light as a signal

Not only is light the most abundant energy source in the biosphere, it is also an important signal that allows living organisms to sense and respond to their electromagnetic environment. In some archaea, a bacteriorhodopsin homologue, sensory
rhodopsin, uses light as a signal to regulate the cell’s flagellar motor and initiate a phototaxis response (Möglicher et al. 2010). For organisms with the sense of sight, light sensing is even more profound—literally illuminating the world around us. The initial step in the visual process occurs through a similar mechanism to that of bacteriorhodopsin. In the retina of the eye, the rhodopsin proteins bind 11-cis retinal. Upon light absorption, retinal isomerizes to the all-trans isomer (Ernst et al. 2014). After this step, the rhodopsin protein mechanism is dissimilar to bacteriorhodopsin because of differing transmembrane protein folds.

Sensory rhodopsin is just one example of many light sensors that are distributed across the three domains of life. The photoreceptor is an apoprotein that requires the binding of a chromophore to become a holoprotein—the active light sensor form. The light sensor holoprotein is defined by its fold and chromophore, which determines its unique absorption pattern and sensitivity to specific regions of the EM spectrum. After light absorption, the light sensor must convert the signal into a precise physical change in the protein and, ultimately, trigger a specific physiological response. This comes about via physical or chemical changes that are initiated at the chromophore, such as the isomerization of retinal.

Although photoreceptors are not essential for life, bacteria often have one or more encoded in their genome. Bacterial light sensing proteins include LOV domains, cryptochromes, phytochromes, rhodopsins, photoactive yellow protein
(PYP) and BLUF domains (Figure 1.3). On light activation, these sensors influence decision making pathways that control bacterial development, behavior and virulence (Herrou and Crosson 2011). Most often, light sensors directly regulate the activity of a second ‘effector’ domain in the same protein. Typically, the effector domain will have enzymatic properties or regulatory functions in the cell. There are hundreds of such protein pathways that allow bacteria to sense and respond to their environment for optimum survival—collectively, they are referred to as signal transduction pathways (STPs).

The light sensor must alternate between light sensing and light activated signalling states. This arises naturally from the signalling mechanism, which involves a reversible change in the sensor protein. During the time that the protein remains in the light activated signalling state, its characteristic absorption pattern will be bleached, but after some time, the light sensor spontaneously returns to the dark state. In this way, it regains light sensitivity and is ready to undergo another round of the photocycle. Although chromophore specificity and absorption patterns are shared traits among photoreceptors of the same family, slight variations in protein sequence can lead to dramatic changes in the time-dependent behavior of the photocycle. In this way, homologous light sensors exhibit unique photocycle durations, which correspond to the optimal response time of their associated STPs (Zoltowski, Vaccaro, and Crane 2009; Circolone et al. 2012).
Figure 1.3: Bacterial photosensors and their chromophores: (A) cryptochrome (1np7); FAD. (B) LOV (1g28); FMN. (C) Bacteriophytochrome (2o9c); biliverdin. (D) Bacteriorhodopsin (1c3w); retinal. (E) PYP (2phy); p-coumaric acid. (F) BLUF (1yrx); FAD.
1.2 Light-Oxygen-Voltage domains (LOV)

1.2.1 Discovery

Phototropism, the bending of plants toward a light source (Figure 1.4 B), has been observed for centuries. Many notable scientists, including Charles Darwin, carried out experiments and recorded observations on the phototropic phenomenon (Whippo and Hangarter 2006; Christie and Murphy 2013). In the 1960’s, Kenneth Thimann and G. Curry investigated the phototropic response at different wavelengths. In this way, they produced an action spectrum that showed the response strength versus wavelength (Figure 1.4 A) (Briggs 2014; Christie and Murphy 2013). It would later be discovered that the protein responsible for phototropism causes other changes in plants as well, including chloroplast migration and stomatal opening (Sakai et al. 2001; Christie and Briggs 2001).
In 1995, the renowned plant scientist, Professor Winslow Briggs, discovered the genes (Phot1 and Phot2) that give rise to Thimann’s action spectrum (Figure 1.4 A and B) (Huala 1997). They isolated the N-terminal region of these phototropin genes, expressed the recombinant protein product and characterized its basic properties. Based on sequence similarity with two other flavoproteins, NIFL
and Aer, they were able to predict that this domain belonged to the LOV family. Initially, speculation about its function as a redox sensor was put forth but UV-vis absorption spectroscopy revealed its sensitivity to certain wavelengths of light, establishing its role as a light sensor (Christie et al. 1999). Lokhandwala et al. recently reported that a subset of these versatile sensors are capable of sensing both light and oxidative stress, adding a fourth possible category of sensing and confirming the early prediction by Huala et al. (Lokhandwala et al. 2015). Although LOV domains include three categories of sensors—light, oxygen or voltage (Figure 1.5)—LOV will be used to refer exclusively to the light-sensing receptors throughout this thesis.
Figure 1.5: In the PAS subfamily, LOV domains include three unique sensors that detect light, oxygen or voltage. Light sensing LOV domains bind FMN, oxygen sensors bind heme, while voltage sensors do not have a bound chromophore.
Professor Briggs was able to isolate and characterize the two photosensing LOV domain proteins (LOV1 and LOV2) that are present in tandem in both Phot genes (Phot1 and Phot2). However, the downstream enzymatic effector domains, which are serine/threonine kinases that give rise to the phototropic effect and other cellular responses in plants, are not amenable to *in vitro* isolation because of protein stability issues. This has limited the success of *in vitro* activity assays and prevented the preparation of concentrated samples that are necessary for biophysical characterization (Briggs and Christie 2002). Because of this limitation, many important questions still remain unclear regarding the details of the Phot-LOV1-LOV2-kinase structure and function. For instance, we do not know how the phototropin proteins are able to induce such a versatile number of responses in plants (Christie and Murphy 2013; Briggs 2014) or how the LOV1 and LOV2 domains can elicit independent responses even though they regulate the same downstream serine/threonine kinase (Briggs and Christie 2002; Herrou and Crosson 2011). Recombinant protein expression and stability issues are common problems that were also encountered in the bacterial LOV-histidine kinase (HK) that is described in this thesis.
1.2.2 LOV domain Structure

In addition to plants and fungi, LOV domains can be found across a broad phylogenetic range of bacteria and archaea (Crosson, Rajagopal, and Moffat 2003; Herrou and Crosson 2011). This unique sensor is derived from the Period-ARNT-Singleminded (PAS) superfamily. Members of the PAS superfamily share a conserved tertiary core structure (Figure 1.6). Structurally, the PAS domain family is characterized by five anti-parallel $\beta$-strands that twist into a curved shape. This $\beta$-sheet structure is partially enclosed in several helices belonging to the PAS core that, together, form the active pocket (Figure 1.6). In addition to sensing various forms of stimuli, the PAS structural fold can also serve as a dimerization interface, joining and orienting other domains that work as functional dimers (Huang, Edery, and Rosbash 1993; Ballario et al. 1998).

Although the PAS fold has a conserved tertiary structure, variations in the primary amino acid sequence can give rise to unique chromophore or ligand binding interactions. In this way, the versatile PAS fold allows for many specialized functions. However, the function and sensing specificity of the majority of PAS domains are not known (Möglich, Ayers, and Moffat 2009). In part, this is due to the fact that PAS domains are the most prevalent sensor module found in nature (Upadhyay et al. 2016). It is also a laborious and challenging task to identify the
functional roles of individual PAS sensors, since there are no throughput strategies for this type of analysis (Zoltowski, Vaccaro, and Crane 2009).

![Figure 1.6: Structural comparison of two PAS domains, LOV2 (left: 1G28) and PYP (right: 1NWZ), reveals the conservation of tertiary features, including the PAS core, the β-sheet scaffold and the helical connector. The N-terminal cap is a variable feature that is not present in all PAS domains.](image)

Light-sensing LOV domains share the characteristic PAS structure—what makes them unique is their active site affinity and steric complementarity with FMN. Light sensor proteins of a given family can easily be identified by conserved residues in the active site that are necessary for the binding of its chromophore. In LOV domains, a conserved sequence in the active site, GXNCRFLQ, is necessary for light activation via reversible, covalent adduct formation with FMN. Chemi-
cal bond formation causes a conformational change that typically propagates to a second effector domain in the same protein. This change activates the effector domain and elicits a cellular response. In most cases, this domain is part of the same protein but there are exceptions, such as short LOV domains that lack a connected effector domain (Hendrischk et al. 2009; Endres et al. 2015).

1.2.3 Mechanism of photochemistry

This study focuses on photosensing LOV domains that noncovalently bind flavin mononucleotide (FMN), which has absorption bands in the UV and visible regions of the EM spectrum. The chromophore absorption spectrum has a minor peak at 375 nm and a major peak centered at 450 nm. Absorption at these wavelengths results in an electronic excited state. After light absorption, the excited electron undergoes a spin-flip, forming a triplet excited state and delaying its return to the ground electronic state. During this prolonged excited state, a conserved cysteine thiol in the active site of the LOV protein forms a covalent bond with carbon-4a (C4a) of the FMN isoalloxazine ring (Figure 1.7) (Salomon et al. 2000; Swartz et al. 2001).

Typically, the covalent S-C adduct bond breaks within seconds to minutes, completing one round of the photocycle and restoring the light-sensing state of LOV and the ‘off’ state of the effector domain. Once the LOV domain has returned
to the dark state, it can undergo another round of the photocycle. In this way, the cyclic nature of the LOV domain’s photocycle allows for a dynamic response to the presence or absence of light.

Figure 1.7: The LOV photocycle has two major states, which are distinguished by noncovalent (dark) and covalent (S-C4a) attachment (adduct) of the FMN moiety.

Many LOV domains have an adjoining region, referred to as the Jα-helix, between the LOV domain and its effector domain. This helical region is thought to play a central role in the mechanism of signal propagation from the LOV sensor to the effector (Crosson, Rajagopal, and Moffat 2003; Conrad, Manahan, and Crane 2014). A partial unfolding or detachment of the Jα-helix could initiate a spatial rearrangement between the two domains. This signal propagation mechanism has been demonstrated by designing chimeric LOV-effector proteins where the effector can be any protein that one wishes to control by light regulation. As a general method, this approach belongs to the growing field of optogenetics which
seeks to use light to control cellular processes and study the spatio-temporal role of macromolecules in living cells (Deisseroth et al. 2006). Yi Wu et al. demonstrated this concept by fusing Rac1, a GTPase that regulates cytoskeleton dynamics, with the LOV domain from *Avena sativa* (LOV2-Ja). The chimeric protein exhibited blue-light activated GTPase activity in living cells (Wu et al. 2009).
Figure 1.8: A representation of the LOV2-RAC1 (2WKP) chimeric protein signalling mechanism occurring via dynamic changes in the Jα-helix—the ‘clam shell’ model. Light activation of RAC1 in vivo resulted in localized cell protrusion and changes in motility.
Sequence analysis indicates that not all LOV domains signal through an identical mechanism. Notable exceptions are the short LOV domains, LOV domains that have an N-terminal effector domain and LOV domains that lack the adjoining \( \alpha \)-helix. The bacterial LOV-histidine kinase (HK) from \textit{Rhizobium leguminosarum} is an example of a signalling pathway that lacks an intervening helix (Bonomi et al. 2012). With these different arrangements, the mechanism of signal propagation can be imagined but we still lack sufficient biophysical data to correctly describe the similarities and differences between them. In general, the conformational dynamics of LOV domains are not well understood and their conformational changes remain a topic of intense speculation (Herrou and Crosson 2011).

### 1.2.4 Effector domains and signal transduction pathways (STPs)

One- and two-component STPs are the predominant mechanisms by which bacteria sense and respond to their environment (Sambasivarao 2013; Ulrich, Koonin, and Zhulin 2005). These pathways serve as the decision making networks of the single-celled organism and determine bacterial behavior. Beyond their scientific importance, an understanding of an organism’s STPs could lead to novel strategies for controlling bacterial virulence. As antibiotic-resistant strains
of bacteria become more common, the development of novel ways to control bac-
terial virulence is of great importance (Wright 2007).

In any given bacterium, there is a multitude of transduction pathways that
elicit overlapping or conflicting responses. The integration of these signals can
make it difficult to untangle the phenotype and characterize the response of in-
dividual pathways (Circolone et al. 2012). The function of several LOV effec-
tor pathways have been successfully identified by phenotype analysis of in-frame
deletion mutants (ΔLOV-effector). As the name of this method implies, a gene
of interest is excised from the genome while keeping the surrounding regions un-
changed. Removing the entire gene prevents any polar effects that could ar-
tificially change the phenotype. Notably, the ΔLOV-HK knockout in several
α-proteobacteria, including Rle, has been shown to regulate the general stress
response (GSR) pathway (Gaidenko et al. 2006; Swartz 2006; Circolone et al.
2012; Bonomi et al. 2012; Foreman, Fiebig, and Crosson 2012). The GSR is the
master stress response pathway in α-proteobacteria and regulates up to 150 dif-
ferent genes in response to signals of stress, such as temperature, pH or starvation
(Kaczmarczyk et al. 2014). In bacteria, the canonical one- and two-component
effector domains are diguanylate cyclases (GGDEF or DGC) and histidine kinases
(HKs), respectively.
Figure 1.9: Four common downstream effector domains that can be regulated by LOV: (A) Diguanylate cyclase (blue), (B) Histidine kinase (purple); PhyR (pink), (C) STAS (hot pink), (D) helix-turn-helix (orange).
One-component signalling pathways are defined as having an input (sensor) and effector (signalling) domain in a single protein (Figure 1.9 A). Sensor and effector domains are modular and can be found in unique arrangements in different proteins. An important subset, the GGDEF output domains, work in conjunction with a phosphodiesterase (PDE or EAL), which can be present in the same multi-domain protein or co-expressed as a separate protein. These enzymes catalyze the making and breaking of cyclic-di-GMP, respectively (Figure 1.10 A) (Hengge 2009). C-di-GMP is a signalling cascade molecule that, with the exception of a few fungi, is unique to prokaryotes. The GGDEF signalling pathway has been shown to regulate many important bacterial processes, including changes in biofilm production, flagella motility, the cell cycle and development, and virulence (Jenal and Malone 2006; Hengge et al. 2016).

GGDEF and EAL enzymes are often highly redundant. For instance, *Escherichia coli* contains 19 proteins with a GGDEF domain and 17 with an EAL domain. Since c-di-GMP has many targets in the cell, its production is thought to be spatially and temporally regulated to prevent cross talk between separate pathways that rely on the same signalling molecule, c-di-GMP (Galperin, Nikolskaya, and Koonin 2001; Römling, Gomelsky, and Galperin 2005). The protagonist EAL domain plays a major part in the spatio-temporal control of these redundant signalling pathways.
Figure 1.10: (A) GGDEF and EAL domains regulate the concentration of the secondary messenger, c-di-GMP, in response to environmental signals (adapted from Jenal and Malone 2006). (B) Two-component STPs involve kinase activation via autophosphorylation, followed by phosphotransfer to a cognate response regulator (RR) with specific targets in the cell.
In the case of two-component signalling pathways, the signal transduction pathway utilizes a sensor-effector protein to activate a downstream response regulator (RR) that goes on to regulate various activities of the cell (Figure 1.10 B). The most prevalent two-component STPs involve a kinase homodimer that uses a phosphotransfer mechanism to activate its RR (Casino, Miguel-Romero, and Marina 2014). Detailed mechanisms for kinase signalling pathways have proven elusive. We know that conformational shifts in the sensor module propagate to the kinase and induce autophosphorylation, turning the kinase on. Autophosphorylation by the kinase homodimer may occur in cis or trans. It has been proposed that conformational changes propagate to the kinase via a twist in the helical coiled-coil region connecting the sensor and kinase (Figure 1.11 A). This twisting motion changes the orientation of the ATP binding site such that the $\gamma$-phosphate can be transferred to the phosphorylation site (histidine or serine/threonine) of the kinase. This is followed by phosphotransfer to an aspartic residue in a cognate RR.

HKs have an affinity for specific RR(s), resulting in selective phosphotransfer. Phosphorylation of the RR increases its activity and generates a physiological response in the organism (Figure 1.10 B). The nature of this response depends on the RR’s activity, which may be enzymatic (e.g., kinases), involve protein-protein interactions or control gene expression (e.g., HTH domains). Common
RRs include cheA, a regulator of the flagellar apparatus, and PhyR, an anti-anti-σ factor that leads to activation of gene expression. In *E. coli*, there are 25 canonical HKs, 4 HWE HKs and 31 RRs. This near 1:1 ratio indicates that highly degenerate bacterial kinase signalling pathways maintain independence by interacting with a specific RR.

The HWE histidine kinases are a subset of HKs that have been identified in α- and γ-proteobacteria. They are named for their distinct pattern of H and W-E residues that are often found in the N and G1 regions (Figure 1.11 B). In addition to this characteristic, they completely lack the F box region (Karniol and Vierstra 2004).
Figure 1.11: (A) Proposed structural basis of signal propagation involves changes in the helical region linking sensor and kinase. (B) Regions of a canonical histidine kinase include N, G1, F and G2 boxes.
1.3 Research goals: a winding path

Initially, our research plan was to study a hypothetical LOV domain from the γ-proteobacterium, *Shewanella* sp. ANA-3, and its associated one-component signal transduction pathway. In order to pursue the most comprehensive study, we planned to carry out both *in vivo* and *in vitro* experiments. In *Shewanella*, we were interested in learning whether a phenotypic change occurs in response to light. We produced an in-frame deletion of *Shewanella*’s only light sensor, a LOV-GGDEF-EAL integral membrane protein, in order to determine the function of its signalling pathway (Link and Phillips 1997).

The *in vitro* component of this research would utilize recombinant expression of proteins followed by molecular and biophysical characterization to provide insights into the structure of the protein and the dynamics associated with its signalling state. Specifically, we wanted to learn how light-activated conformational changes regulate the downstream effector domain. This project focused on a LOV domain that linked one-component signalling to light sensing. With both GGDEF and EAL domains present in the same protein, the nature of regulatory control in this system promised to be interesting and unique. There has yet to be a comprehensive study of a LOV-GGDEF or LOV-GGDEF-EAL protein, adding to the novelty of this regulatory pathway as a research project. After making an in-frame dele-
tion mutant (Baba et al. 2006, (procedure)) of the She-LOV-GGDEF-EAL gene in *Shewanella*, we used similar microbiology strategies and molecular techniques to design and clone several protein constructs. Cloning strategies proved to be the first major roadblock towards reaching our experimental objectives. After overcoming this issue, protein expression problems were encountered and so the She-LOV-GGDEF-EAL project was disbanded as a necessity of time.

In its place, a LOV-HK protein from *Rhizobium leguminosarum* (*Rle*) became the central focus of our research endeavors. Because the *in vivo* phenotype characterization for *Rle* had already been established (Bonomi et al. 2012), the aims of this research project focused on biophysical characterization of Rh-LOV-HK. Rh-LOV-HK appears to have an unusual mechanism of signal propagation from LOV to the effector, since the histidine phosphorylation site of the HK effector domain directly follows the LOV core. Given this difference relative to LOV domains that have an intervening Jα linker, it is plausible that a twisting or sliding motion could be sufficient to increase autophosphorylation in the kinase. If we were successful with this central aim, a second goal would be to identify and characterize the cognate Rh-RR partner in this two-component STP.

The LOV domain and full LOV-HK were cloned into a His-tag and other fusion tag vectors and overexpressed in *E. coli*. The LOV domain constructs expressed as soluble proteins with functional activity but we encountered expression and
stability issues with the full-length LOV-HK. Recombinant protein expression in
_Ε. coli_ is efficient and affordable but has a major shortcoming—up to 80% of
recombinant proteins end up sequestered into proteinaceous inclusion bodies and
are rendered useless (Esposito and Chatterjee 2006). This problem is not well
understood, nor is there a definitive solution. However, several strategies have
been developed to overcome this problem. We will discuss approaches that were
used to express soluble Rh-LOV-HK with limited success in Chapter 2.

Ultimately, because Rh-LOV was soluble and amenable to biophysical charac-
terization, it was utilized for most of the experiments described in this thesis. With
this shift in focus, research aims were adjusted accordingly: 1.) Basic optical char-
terization of Rh-LOV and Rh-LOV-HK’s photocycle properties, 2.) Characteri-
zation of Rh-LOV’s structure and conformational dynamics, and 3.) Identification
and characterization of the cognate RR in the Rh-LOV-HK signalling pathway
(Figure 1.12). To keep research moving forward between the many ‘not work-
ing’ stages, we also initiated biophysical experiments with a LOV domain from
_Brucella abortus_ and a LOV-helix-turn-helix (HTH) protein from _Erythrobacter
litoralis_. These two LOV domains were originally cloned into pET100D expression
vectors by Dr. Trevor Swartz. We will use these two well-studied LOV domains
[Br-LOV-PAS-HK] (Swartz 2006; Rinaldi et al. 2012; Sycz et al. 2015) and [EL222]
(Nash et al. 2012; Rivera-cancel, Motta-mena, and Gardner 2012) for comparative
analysis with Rh-LOV. We will also discuss original EPR experiments that were
carried out with these LOV domains in Chapter 4.

Through this research, I was exposed to many complex problems and a bat-
tery of molecular and biophysical techniques that will be discussed in the following
chapters. But the real ‘essential skills’ that I have acquired from this PhD process
go beyond the list of protocols, procedures and biological spectroscopy exper-
iments that follow—I have gained confidence in myself and believe that I can
address any biochemical question, regardless of the problem solving and learn-
ing that is required along the way. Furthermore, I believe that these skills are
broadly translatable across the fields of research, teaching and, quite possibly into
all aspects of life.
Figure 1.12: Research goals with Rh-LOV: (A) Aim 1. Characterize temporal photocycle properties. (B) Aim 2. Determine Rh-LOV structure and investigate its signalling dynamics. (C) Aim 3. Identify and characterize the downstream RR in this pathway.
References


Circolone, Franco et al. (2012). “Structural basis for the slow dark recovery of a full-length LOV protein from Pseudomonas putida”. In: *Journal of Molecular Biology* 417.4, pp. 362–374.


Rivera-cancel, Giomar, Laura B Motta-mena, and Kevin H Gardner (2012). “Identification of Natural and Artificial DNA Substrates for Light-Activated LOV HTH Transcription Factor EL222”. In:


Sakai, T et al. (2001). “Arabidopsis nph1 and npl1: blue light receptors that mediate both phototropism and chloroplast relocation.” In: Proceedings of the


Chapter 2

Spectroscopic characterization of a LOV-HK from \textit{Rhizobium leguminosarum}

2.1 Introduction

LOV domains are an important class of photosensing proteins that convert light stimulus into a biological signal. Initiation of this light-to-signal transformation starts with the ability to sense specific frequencies of the EM spectrum. LOV domains’ unique photosensitivity is derived from a noncovalently bound
chromophore, FMN, which has absorption peaks in the UV and blue regions of the spectrum. In the LOV active site, FMN is positioned in proximity to a cysteine thiol functional group. Photon absorption by the chromophore leads to the formation of a covalent cysteinyI-FMN (S-C4a) adduct. This internal chemical change drives a conformational change in the protein’s structure. Typically, this light-induced change initiates signal transduction by altering the activity of a C-terminal effector domain in the same protein. Although the details and extent of these conformational dynamics are still an area of active research, these changes are substantial enough to alter the activity of a downstream effector domain and trigger a cellular response (Conrad, Manahan, and Crane 2014).

The covalent S-C4a adduct, which defines LOV’s photoexcited state, eventually breaks and returns the sensor to its original dark state. The cyclic nature of the photocycle allows the LOV domain to return to the signalling state if light persists. It is not clear what disrupts the covalent adduct, but the duration of the excited state of LOV domains from different organisms have been observed to vary dramatically (Zoltowski, Vaccaro, and Crane 2009). This characteristic is significant since it determines the duration of the output (effector) domain’s signalling response. In cases where the LOV-effector response has been determined, light activation has been observed to upregulate the effector domain’s activity (Herrou and Crosson 2011). However, it is conceivable that the LOV domain signalling
state could lead to downregulation of an output domain as well.

2.1.1 Photosensing in \textit{Rhizobia}

The order \textit{Rhizobiales} belongs to the \(\alpha\)-proteobacteria class of Proteobacteria. Although technically they are soil bacteria, \textit{Rhizobia} are often found in a unique commensal relationship with plants of the legume family. In a form of mutually beneficial symbiosis, the plant supplies \textit{Rhizobia} with sufficient nutrients for survival in root nodules. Nodule formation localizes the bacterial colonization and acts as a nitrogen fixing organelle for the plant. The conversion of atmospheric N\(_2\) to ammonia provides the plant with an abundant source of this precursor, which is required for the production of nucleotides and amino acids (Figure 2.1).

To establish successful symbiosis, the initial infection by \textit{Rhizobia} requires the exchange of signalling molecules between the plant and bacterium (Krehenbrink and Downie 2008). The signals that initiate infection are highly selective—in many cases only one strain of \textit{Rhizobia} will infect one member of the legume family. For instance, the \textit{R. leguminosarum} strain only infects pea plants (Wang et al. 2012). Recent research suggests that light may be an important signal for initiating the conditions leading to bacterial infection in some \textit{Rhizobia}. 

46
Figure 2.1: Root nodules house *Rhizobia*, which fix nitrogen for the plant in exchange for a steady supply of sugars and nutrients.

Although most *Rhizobia* have a life cycle that takes place in soil or root nodules, 5 of the 25 genomes sequenced have a LOV domain. The strain that will be discussed here, *Rhizobium leguminosarum* bv. *viciae* 3841 (*Rle*), has two light sensors: a phytochrome and a LOV domain. The LOV sensor is connected to an HWE histidine kinase (HK). Light activation presumably leads to HK autophosphorylation, followed by phosphotransfer from the HK to a cognate response regulator (RR) protein with specific targets in the cell. Although the identity of the interacting RR has not been identified, Bonomi et al. produced an in-frame deletion of the LOV-HK gene in *Rle* and characterized the phenotype for ∆Rh-LOV-HK strain. Their results indicate that Rh-LOV-HK regulates the general stress response (GSR) pathway. Changes in the ∆Rh-LOV-HK strain included decreased...
extracellular polymeric substance (EPS) and biofilm production, downregulation of flagella synthesis, and attenuated root nodulation (Bonomi et al. 2012). These observations, indicate that light is an important signal for establishing *Rhizobium*-legume symbiosis. More recently, several other LOV-kinase pathways in α-proteobacteria have been linked to infection and the GSR pathway (Rambow-Larsen et al. 2009; Foreman, Fiebig, and Crosson 2012; Kaczmarczyk et al. 2014; Sycz et al. 2015).

*Rle*’s STPs are extensive, with over 800 predicted signal transduction proteins total (Figure 2.2 A). The genes coding for these pathways are spread out over the central chromosome and six smaller plasmids (Young et al. 2006). *Rle* has 58 HKs and 66 RR. Two of the four HWE HKs (pRL110319, pRL110320) are found on the pRL11 plasmid (Figure 2.2 C). The two genes are on adjacent DNA strands—since they are transcribed in opposite directions, their promoter regions are in close proximity. This suggests that they might be controlled by common regulatory factors and possibly share the same signalling pathway. PRL110320 codes for a 345 amino acid protein (Rh-LOV-HK) but pRL110319 is a 341 amino acid protein that is regulated by a PAS domain of unknown sensing specificity. Both pRL110319 and pRL110320 genes are followed by a small gene with the designation, DUF (domain of unknown function). This may be purely coincidental, but it is suggestive since it is common for proteins to function as subunits
of a higher order complex. Could this small protein play a role in Rh-LOV-HK’s signalling pathway and stability? Alternatively, could the PAS-HK and LOV-HK form a heterodimer? Although we were not able to explore these possibilities, we will return to a discussion of strategies for isolating stable Rh-LOV-HK in Chapter 5.
Figure 2.2: (A) A comprehensive histogram of all STPs in *Rle*. (B) Rh-LOV-HK is categorized as a transmitter. (C) The Rh-LOV-HK gene (pRL110320) and the gene product for a PAS-HWE-HK (pRL110319) on the complementary strand share nearly identical secondary structure. Images generated in MiST2 (L. Ulrich and I. Zhulin, 2010).
Rh-LOV-HK shares a high degree of homology with the LOV-PAS-HK from *Brucella*, a closely related α-proteobacterium. The Br-LOV domain has been studied extensively. However, similar to Rh-LOV-HK, expression problems have prevented biophysical studies with the full-length protein. A zoonotic pathogen, *Brucella*, is capable of infecting both animals and humans. In animals, *Brucella* causes brucellosis—characterized by miscarriage and infertility. In humans, there are many pathological conditions associated with infection, including acute undulating fever, headaches, malaise, sweats and arthritis (Gorvel 2008). A host of more complicated, chronic conditions can occur as well (Gorvel 2008). After infection, *Brucella* can remain alive in the host’s macrophage cells. Using a macrophage infection assay, it was shown that *Brucella*’s ability to infect these cells increased after light exposure (Swartz 2006). This discovery is notably similar to the light response that was observed in *Rle*.

Gaston Paris et al. cloned 22 of the 24 RRs from *Brucella* and used a two-hybrid assay to identify two RRs that interact with Br-LOV-PAS-HK. One RR is cheY-like, a well known protein involved in chemotaxis regulation. *Brucella* does not have all the genes necessary for flagellar propulsion, so cheY’s functional role is not understood in this context (Sycz et al. 2015). One possibility is that it serves as a phosphorylation sink, which regulates the intensity and duration of phosphotransfer. The second RR that undergoes phosphotransfer with Rh-LOV-
HK is a PhyR homologue. PhyR is an anti-anti-σ factor that has been established as a regulator of the GSR (Staroń and Mascher 2010; Galperin 2010). These results are consistent with the phenotype-based observation in Rle and could link Rle’s PhyR protein to the Rh-LOV-HK signalling pathway.

2.1.2 The general stress response (GSR) pathway

The GSR pathway protects bacteria against various types of stress and starvation conditions. Stimuli that can activate this pathway includes UV light, temperature extremes, pH, salt, ethanol and cell wall stress to name a few (Hecker, Pané-Farré, and Völker 2007). In this context, the GSR acts as master switch that determines bacterial lifestyle and survival. Under ideal conditions, growth, development and motility are all encouraged. However, when conditions are poor for survival, biofilm formation and sessility are the dominating characteristics. Many different STPs may influence the GSR’s activity at any given moment, so the GSR must integrate this information and respond appropriately. One important regulator of the GSR is a RR protein, PhyR, which functions as an anti-anti σ (STAS) factor. The active PhyR binds extracytoplasmic-function (ECF) σ factor (σECF), releasing it from the transcription complex in order to initiate gene transcription of proteins involved in the GSR response (Foreman, Fiebig, and Crosson 2012).

Figure 2.3 provides a model for the GSR’s response to light in many α-
proteobacteria. The activity of the two-component signalling pathway is initiated by light activation of a LOV-HK protein. This leads to phosphorylation of a histidine residue in the kinase, followed by phosphotransfer to an aspartate of its cognate RR, PhyR. Phospho-activated PhyR exhibits increased affinity for the anti-σ factor, NepR. The release of NepR from the σ\textsuperscript{ECF} factor allows for activation of the RNA polymerase, which transcribes stress response genes to mount an appropriate response. There are as many as a 150 stress response genes whose promoters are regulated by the GSR (Hecker, Pané-Farré, and Völker 2007; Sycz et al. 2015).
Figure 2.3: In α-protobacteria, the GSR pathway responds to different forms of stress, including environmental and chemical factors. This figure provides a model of how light stimulus leads to activation of the GSR pathway (image adapted from Sycz et al.).
2.1.3 Optical spectroscopy in the study of LOV domains

Optical spectroscopy plays a special role in the study of light sensors. Since these proteins are triggered by light absorption, optical spectroscopy can be used to observe their unique absorption and fluorescent properties and resolve their photocycle changes in time. In the field of LOV domains, the general mechanism of adduct formation and breakage is well established. However, with respect to the mechanistic details, difficult questions persist. For instance, it is uncertain whether adduct formation occurs through a thiol, thiolate or possibly even a thiolate radical. All three mechanisms have been proposed (Sato et al. 2005; Schleicher et al. 2004; Swartz et al. 2001).

In the results section, we will present optical spectroscopy spectral data that was acquired in order to characterize the photocycle of Rh-LOV and Rh-LOV-HK. We will consider early changes in the photocycle to determine the lifetime ($\tau_{\text{forward}}$) of adduct formation using time-resolved flow-flash absorption spectroscopy. We will also present the complete time-resolved photocycle, which was monitored by absorption and fluorescent changes after photo-illumination. Finally, using circular dichroism (CD), light-induced conformational changes will be considered in Rh-LOV.
2.2 Materials and methods

This section will outline the cloning, expression and purification techniques that were used throughout this thesis. Although considerable time and effort were invested in troubleshooting the cloning and expression processes, we will only touch on the challenges that were encountered with these methodologies and place greater emphasis on the steps that were taken to move our research objectives forward.

2.2.1 Gene cloning and protein expression

Initial cloning methods were based on Invitrogen’s pET directional TOPO® cloning methodology. This traditional cloning technique used blunt-end primer design. Topoisomerase I is used to cleave the phosphodiester backbone of a single strand in the cut vector, which creates a four base-pair overhang. This overhang intercalates into the complementary sequence on the forward primer, which allows for the correct orientation of the gene insert into the vector (Kits 2010). PCR reactions were performed in a PTC-100 (MJ research) using genomic DNA from \textit{R. leguminosarum bv. viciae} 3841 as template and Phusion® high fidelity polymerase. The resulting PCR products were verified by running 5 \(\mu\)L of PCR product on an agarose gel with EZvision stain—a non-toxic alternative to
the DNA intercalator dye, ethidium bromide, that has the same excitation and fluorescence emission properties. Concentrations were determined by NanoDrop (Thermo Scientific) analysis. After verification, PCR products were purified using a Qiagen PCR purification kit. The gene insert was mixed in 3:1 and 5:1 molar ratios with approximately 100-150 ng of the cut vector. After transformation into TOP10 competent cells, individual colonies were selected from the Luria broth (LB)/ampicillin (Amp) agar plates and grown up in LB/Amp media. The plasmid was isolated and purified using a Qiagen plasmid purification kit and the gene insert was sent for sequencing using the T7 forward primer (Sequetech Corp.).

This cloning strategy was eventually replaced with ligation independent cloning (LIC). The ligation step is reported to be the most error prone step in the cloning process, so eliminating this step significantly increases the efficiency of plasmid generation. After acquiring a large LIC vector collection from Berkeley’s QB3 MacroLab, our selection of vectors and cloning efficiency were much improved. The LIC vector set, designed by Dr. Scott Gradia, includes vectors with various fusion tags for affinity purification and several fusion tags that have been shown to enhance soluble protein expression. In many cases, fusion protein expression has been shown to overcome or decrease protein sequestration in inclusion bodies (Esposito and Chatterjee 2006).

Primers were designed based on the QB3 MacroLab protocol (UC Berkeley).
This method allows for the 5' and 3' ends of the cut vector to anneal with the insert via complementarity of base pairs rather than using a ligase (Figure 2.4). T4 DNA polymerase was used for its 3'→ 5' exomuclease activity to digest the complementary sticky end regions of the insert and vector. Following T4 digestion, the vector and insert were mixed in a 1:3 molar ratio. After a few minutes incubation, the annealed insert and vector were transformed into DH5α E. coli cells. The plasmid initially has nicks in its phosphodiester linkages but after transformation, the host ligase repairs these linkages. Hence, this method does not require the use of an external ligase.
Figure 2.4: LIC methodology uses sticky end cohesion, rather than annealing gene insert and vector using a ligase. This provides a more efficient approach to cloning (QB3 MacroLab).
The transformants were selected by plating on LB/Amp agar plates and incubated overnight at 37 °C. Individual colonies were selected from the plate and grown in liquid culture overnight. Plasmids were then isolated and purified with an Axygen® plasmid purification kit and sent for sequencing (Sequetech Corp.). Test expressions were typically done with 50 mL of LB/Amp. The cells were grown to an optimal optical density (0.6-0.8 OD at 600 nm), then induced with 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG). IPTG is a synthetic allolactose analogue that binds to the LAC repressor and induces expression of genes that are regulated by the LAC operon. All expression vectors that were used in this research utilized the LAC operon. Time points were taken by harvesting and pelleting 1-2 mL of cells at 0, 1, 3 and 5 hours in a microcentrifuge. SDS PAGE gels were used for analysis when appropriate but more often, the visual fluorescence of the LOV domain was used as a direct indicator of protein expression (Figure 2.5). With a long-wave UV (365 nm) light source for excitation, a 1-2 mL aliquot of cells was sufficient to visualize the LOV fluorescence properties. Furthermore, separation of soluble and insoluble cellular fractions were analyzed to determine whether the expressed protein was soluble or insoluble. Separation was achieved by sonicating the cells, followed by centrifugation at 14,000 rpm for 30 minutes in a microcentrifuge.
2.2.2 Inclusion body problem

Inclusion body expression has been identified as the bottleneck in the process of recombinant protein overexpression in *E. coli* host cells. The reasons why so many proteins trigger inclusion body formation in *E. coli* have not been identified. It is thought that the host expression system recognizes that the protein is unstable or harmful to its survival and sequesters it in proteinaceous (inclusion) bodies. The protein can be isolated from inclusion bodies using 6 M guanidinium hydrochloride (GdmCl) or an 8 M urea solution, purified by traditional affinity methods and then refolded by removing the denaturant (Sørensen and Mortensen 2005). It has been
reported that sub-denaturing concentrations of GdmCL (2-4 M), non-denaturing detergents (e.g., salicylic acid) or physical methods (e.g., sonication and high pressure French press) can also be used to release folded protein from inclusion bodies (Peternel and Komel 2010; Peternel et al. 2008).

Since inclusion bodies are predominantly the over-expressed protein, isolation of protein in this way allows for a jump-start on the purification process (Batas, Schiraldi, and Chaudhuri 1999). In the case of Rh-LOV-HK, it was observed that the LOV domain exhibited a normal photocycle in inclusion bodies, i.e. activation followed by return to the dark-state. However, we were unable to use non-denaturing approaches to obtain soluble, functionally active protein. We were also unable to refold Rh-LOV-HK after isolating it under denaturing conditions. A complication of the refolding process for LOV domains is the chromophore, which must be included in the buffer during the refolding process. We were able to refold Rh-LOV with about 20% efficiency but we had no success with Rh-LOV-HK. The issue of full-length Rh-LOV-HK stability most likely further complicated the folding process.

To surmount this problem, various solubility enhancing protein fusion-tags were tested (Cabrita, Dai, and Bottomley 2006). These protein tags have been empirically determined to decrease inclusion body protein sequestration during recombinant protein expression in E. coli. However, only modest soluble protein
expression was obtained using this strategy. Furthermore, the low yield of soluble protein that was harvested from these fusion constructs lacked stability and precipitated out of solution over time. With TEV cleavage, the protein became even less stable and much of it remained trapped on the column, indicating problematic TEV cleavage. This was also true when full-length protein was obtained by insect cell expression following the Life Technologies’ cloning and expression protocol (technologies 2002). Ample soluble protein was achieved from this expression system. Unfortunately, the majority of the protein obtained from the mammalian expression system remained trapped in the nickel NTA resin matrix rather than eluting off the column. We speculate that these problems were due to the lack of protein stability which resulted in oligomerization. After testing Rh-LOV-HK soluble expression with several different fusion tags, the fusion construct with the sumo protein tag (2St vector) was empirically determined to be the best approach to produce soluble full-length Rh-LOV-HK. The apparent lack of stability of Rh-LOV-HK raises questions as to how it may be stabilized \textit{in vivo}. For instance, there may be stabilizing partner proteins, such as the the associated RR or the DUF domain previously mentioned. This problem is revisited in Chapter 5 and a possible solution is proposed.
Eventually, shorter constructs were designed and cloned. In some cases, small alterations of the N- or C-terminus can improve soluble expression (Gräslund et al. 2008). Using this strategy, we found that inclusion body sequestration persisted in these slightly altered full-length constructs. We also produced truncated constructs that gradually included more of the kinase. However, constructs that included the histidine (His) phosphorylation site at sequence position 163 triggered inclusion body sequestration. Because of limited success in obtaining the full-length Rh-LOV-HK protein, Rh-LOV(160) and Rh-LOV(142)-core were cloned and studied independently of the kinase. Although this approach limited our ability to isolate and study the protein signalling pathway \textit{in vitro}, it

<table>
<thead>
<tr>
<th>Construct and fusion</th>
<th>Soluble expression</th>
<th>Insoluble expression</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rh-LOV-HK-2St (Sumo)</td>
<td>Moderate</td>
<td>Strong</td>
</tr>
<tr>
<td>Rh-LOV-HK-2Nt (NusA)</td>
<td>Weak</td>
<td>Strong</td>
</tr>
<tr>
<td>Rh-LOV-HK-2Ct (MBP)</td>
<td>Moderate</td>
<td>Strong</td>
</tr>
<tr>
<td>Rh-LOV-HK-2Bt (N-His)</td>
<td>Weak</td>
<td>Strong</td>
</tr>
<tr>
<td>Rh-LOV-HK-2Bct (C-His)</td>
<td>Weak</td>
<td>Strong</td>
</tr>
<tr>
<td>Rh-LOV+His-2Bct</td>
<td>Weak</td>
<td>Weak</td>
</tr>
<tr>
<td>Rh-LOV+J-2Bct</td>
<td>Strong</td>
<td>None</td>
</tr>
<tr>
<td>Rh-LOV-J-2Bct</td>
<td>Strong</td>
<td>None</td>
</tr>
</tbody>
</table>

Table 2.1: Soluble expression of various constructs of Rh-LOV and Rh-LOV-HK fusion proteins that were cloned for experimental characterization. Only soluble constructs were suitable for experimentation.
allowed for immediate biophysical characterization of a novel LOV sensor. Since Rh-LOV(160) was used in all experiments, we will refer to it as Rh-LOV (Figure 2.6).

![Diagram of LOV and HWE HK domains with amino acid numbering]

Figure 2.6: Full-length and Rh-LOV domain constructs with amino acid numbering. Rh-LOV-160 was the primary construct that was characterized in this thesis.

Rh-LOV constructs were cloned into a vector that expressed the protein with a C-terminal histidine (His) tag and a TEV cleavage site (2Bct vector). For all biophysical studies, the C-terminal His-tag was left intact. Out of convenience, it is common to leave the His-tag intact for biophysical protein research, and this strategy has been successfully utilized in other LOV domain studies (Rinaldi et al. 2012; Circolone et al. 2012).
2.2.3 Protein expression and purification strategies

*E. coli* (BL21 or DE3 strain) host cells were grown in LB media with 50 µg/mL Amp. Starting with 100 mL overnight, 14 mL of cells were transferred to each of six 2 L flasks containing 800 mL of LB/Amp media. Protein expression was initiated after growing cells to an optical density (OD) of 0.6-0.8 at 37 °C and shaking at 200 rpm. IPTG was used at 1 mM concentration to induce protein expression. The expression was carried out for 5-12 hours at room temperature (rt). The cells were harvested by spinning at 8000 rpm for 10 minutes in a Sorvall-4B centrifuge. The liquid media was decanted and the cell pellets were stored at −20 °C until the purification process.

2.2.3.1 Affinity purification

Various affinity purification methods were used in this research, including glutathione-s-transferase (GST), amylose (MBP-tag), calmodulin resin, as well as cobalt and nickel resins. Since nickel purification was the predominant method used, we will describe this approach as a general methodology. Other methods may vary in terms of the wash and elution strategy. Cell pellets from the previous step were thawed and solubilized in 20 mM Tris buffer (pH 7.5) with 250 mM NaCl and 10 mM imidizole. The cells were then lysed using a hydraulic French press under 15,000-20,000 psi. After lysis, the cellular milieu was spun at 17,000...
rpm for 20 minutes to pellet the insoluble fraction. The lysate was then filtered and bound to nickel NTA resin for 30 minutes before washing. The imidazole concentration was increased in 20 mM increments in each column volume (25-100 mL) of wash buffer until a final concentration of 70-90 mM was reached. After several column volumes of the final wash buffer, the protein was eluted with 300 mM imidazole in the same Tris buffer.

### 2.2.3.2 FPLC

Rh-LOV samples were concentrated in a 20 kDa spin concentrator at 7000 rpm and further purified over a Sephadex S75 (GE Healthcare) column on an Acta prime FPLC. For smaller purifications, ÄKTAprime was used with a S75/200 (GE Healthcare). For larger purifications, a more modern ÄKTApurifier was used with an S75 16/600 column (GE Healthcare). The major differences between these two columns are their void volumes (24 mL and 120 mL, respectively) and loading capacity (0.5 mL and 5 mL, respectively). Elution fractions were collected with a fraction collector. The final buffer conditions were 20 mM Tris buffer (pH ∼7.5) with 250 mM NaCl. The protein eluted as one minor peak in the void volume and a major peak that corresponded to the mass of Rh-LOV as a dimer (37 kDa). Both fractions contained Rh-LOV, but the void volume fraction appeared to be due to oligomerization and was not used. Pure fractions were combined and the protein was concentrated according to the needs of the experiment.
2.2.4 Instrumentation

Rapid absorption kinetics were carried out using a custom experimental setup that was built by Dr. Jim Lewis. This system utilizes 477 nm laser light excitation. UV-vis absorption spectra were acquired with a Hewlett Packard H4582A diode array spectrometer to resolve adduct decay kinetics. The dark state UV-vis absorption spectrum was also used to predict protein concentration. Fluorescence experiments were carried out on a Varian Cary Eclipse fluorescence spectrophotometer at room temperature (rt). Circular dichroism (CD) experiments were performed on a Jasco J-1500 CD spectrometer at rt.

2.3 Results

Optical spectroscopy covers a broad range of techniques that use visible light to acquire details pertaining to the structure and function of molecules. Light-sensing proteins undergo absorption changes at different stages of the photocycle—these changes can be characterized by absorption or fluorescence spectroscopy. The timing of these spectral observations can provide direct insight into the behavior of the signalling response (Corchnoy et al. 2003). The basic characterization of Rh-LOV discussed in this chapter will include monitoring time-resolved changes of the Rh-LOV and Rh-LOV-HK’s photocycle by absorption and fluorescence
changes. We will also monitor secondary structural changes in Rh-LOV’s dark and adduct states using CD.

2.3.1 Absorption spectroscopy

2.3.1.1 Photosensor activity

A LOV domain protein sample has a yellow hue, owing to its strong absorption in the blue. Formation of the LOV photoexcited state can be observed by visual inspection in a concentrated sample—after blue light irradiation, the sample will bleach and appear colorless. To understand these changes, fluorescence (Figure 2.5) or absorption (Figure 2.7) spectroscopy can be used to follow the intricate details of the photocycle in time.

UV-vis spectroscopy was used to show the changes in absorption before and after Rh-LOV was exposed to flash irradiation with white light, which contains all frequencies of the visible spectrum. The characteristic dark and photoexcited absorption spectra provide immediate feedback about the activity of a LOV domain (Figure 2.7). Protein-bound FMN has an absorption peak at 375 nm and three absorption peaks centered at 450 nm (Figure 2.7 (red = dark state absorption spectrum)). After light exposure, these absorption peaks are lost (blue = adduct state absorption spectrum), resulting in the visible bleaching effect. These absorption changes are due to a loss of aromaticity in the isoalloxazine ring after
adduct formation.

The absorption spectrum of the adduct in Figure 2.7 shows incomplete bleaching after flash irradiation—the dark state spectral envelope characteristics are still partially present. The quantum efficiency (QE) of adduct formation for different LOV domains can range many fold (Kasahara et al. 2002). Once light absorption occurs and the triplet state is formed, the excited state electron has a probability of relaxing back to the ground state or forming the S-C4a adduct. The back reaction is the reason for LOV’s less than perfect QE. Although we can calculate the branching ratio, there is not an obvious reason why different LOV domains vary in their rates of forward (adduct formation) and back (relaxation) reactions.

![Figure 2.7: Rh-LOV absorption spectra (A) before and (B) after flash photo-illumination.](image)

Figure 2.7: Rh-LOV absorption spectra (A) before and (B) after flash photo-illumination.
2.3.1.2 Intermediate states in the photocycle

Early changes in the absorption spectra of a photoreceptor can signify intermediates in the transition to the photoadduct state. It is well known that after light absorption, FMN forms a triplet-excited state, which is defined by an electron spin-flip and results in a long lived excited state due to spectroscopy selection rules that arise from the Pauli exclusion principle (Song 1967; Swartz et al. 2001). The triplet state is quenched either by formation of the S-C adduct or by the back reaction, which occurs via electron relaxation to the ground state (Figure 2.14). The following equation shows the relationship to calculate the lifetime for adduct formation $\tau_{forward}$.

$$k_{app} = \frac{1}{\tau_{app}} = \frac{1}{\tau_{forward}} + \frac{1}{\tau_{reverse}} \quad (2.1)$$

Equation 2.1 can be estimated from the rapid time-resolved absorption changes. Figure 2.8 shows the absorption spectra after laser excitation at time intervals ranging from 10 $\mu$s to 5 ms. From these data, $\tau_{app}$ can be calculated. Then, by taking the ratio of the absorption difference (Figure 2.9) between $\Delta A_{650}$ and $\Delta A_{450}$ at two different times, the extent of the forward and back reactions can be determined. With this result, equation 2.1 simplifies to one unknown variable, which allowed us to solve for Rh-LOV’s $\tau_{forward}$. 

71
Figure 2.8: Rapid absorption kinetics at logarithmically spaced time intervals from 10 µs to 5 ms for (A) Rh-LOV and (B) Rh-LOV-HK.
Figure 2.9: Absorption kinetics between 1 \( \mu s \) and 5 ms was used to calculate the branching ratio for (A) Rh-LOV. (B) Rh-LOV-HK was observed to have similar time dependence but lacked stability after the laser excitation flash.
Rh-LOV has a triplet state lifetime that lasts for 6.5 µs before the adduct state is formed. This formation time is longer than previously recorded for other LOV domains (Table 2.1). Both Rh-LOV and Rh-LOV-HK transitions appear to be on a similar time scale (Figure 2.9 A and B). Due to the protein’s instability issues and the power of the laser, instantaneous denaturation of Rh-LOV-HK was observed. Although the signal-to-noise ratio is much more significant, we note that the time scales for $\tau_{forward}$ and $\tau_{reverse}$ are similar to Rh-LOV (Figure 2.9).

The variability in the lifetime of the triplet state for different LOV domains may be an effect of proximity between the thiol and the C4a carbon of FMN. In Figure 2.10, we consider the S-C distance in several LOV domain crystal structures. The thiol functional group often exists in more than one conformation in the crystal structure (Figure 2.10 B and D). Table 2.1 compares the average S-C distance of four different LOV domains with their corresponding lifetime of adduct formation, $\tau_{forward}$. This small sample of LOV domains indicates that there may be a correspondence between $\tau_{forward}$ and S-C bond length.
Figure 2.10: Distances between the active site cysteine and FMN: (A) Rh-LOV. (B) Br-LOV. (C) EL222. (D) Oat LOV2.
<table>
<thead>
<tr>
<th>LOV domain</th>
<th>$\tau_{forward}$</th>
<th>S-C distance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rh-LOV</td>
<td>6.5 $\mu$s</td>
<td>4.3 Å</td>
</tr>
<tr>
<td>Br-LOV</td>
<td>3 $\mu$s</td>
<td>3.6 Å</td>
</tr>
<tr>
<td>EL222</td>
<td>1.9 $\mu$s</td>
<td>3.8 Å</td>
</tr>
<tr>
<td>Oat LOV2</td>
<td>4 $\mu$s</td>
<td>4.1 Å</td>
</tr>
</tbody>
</table>

Table 2.2: The rate of adduct formation relative to the average distance of S to C4a (FMN) in the dark state.

### 2.3.1.3 Rh-LOV and full-length Rh-LOV-HK photocycle half-life

One round of the photocycle, dark state $\rightarrow$ adduct state $\rightarrow$ dark state, determines the active state duration for the downstream effector domain. The LOV photoadduct state can last from seconds to hours. EL222 and Br-LOV represent the two extremes—with photocycle half-lives ($\tau_{1/2}$) lasting for $\sim$30 seconds and a duration of many hours, respectively (Zoltowski, Motta-Mena, and Gardner 2013; Swartz 2006).

Figure 2.11 shows the UV-vis absorption changes for Rh-LOV and Rh-LOV-HK at 80 second intervals after photobleaching for $\sim$2 hours. The insert follows the adduct breakage and return to the dark state with time. Rh-LOV’s photocycle was observed to occur on an intermediate time scale with a $\tau_{1/2}$ $\cong$ 22 minutes and complete return to the ground state in $\sim$2 hours. The photocycle characteristics of Rh-LOV’s adduct state lifetime are notably similar to that of full-length Rh-
LOV-HK (Figure 2.11 B). This is an important observation, providing evidence that the kinase domain does not influence the LOV domain’s return to the dark state. If the photocycle duration is a property of the LOV domain alone, then the observed time variance of different LOV domains must be due to differing protein interactions with the chromophore or possibly, different interactions associated with the conformational change of the adduct state that allows for signal propagation to the effector domain.
Figure 2.11: Adduct decay after photobleaching for (A) Rh-LOV and (B) Rh-LOV-HK. The dark state sample was used for the baseline spectrum. After flash irradiation, UV-vis absorption spectra were collected at 80 s intervals for ~7000 s.
2.3.2 Fluorescence kinetics

FMN is inherently fluorescent in the long-wave UV (∼310 nm) region (Figure 2.5). Far UV irradiation of the dark state of a LOV domain, causes visible fluorescence emission in the green without any other physical change in the protein (Figure 2.5). Figure 2.12 A and B show the fluorescence emission of dark and adduct states of Rh-LOV-HK and Rh-LOV, respectively. In the dark state, fluorescence emission occurs with a maximum at 500 nm. On light activation, the fluorescence emission of FMN is temporarily bleached. Because fluorescence is a property of the chromophore, the observed changes are conserved in all LOV domains. However, LOV domain fluorescence may have unique characteristics as well resulting from the presence of quencher molecules, such as nearby tryptophans (W) in proximity to the chromophore.

Fluorescence spectroscopy can be monitored in real time and this approach was used to follow Rh-LOV and Rh-LOV-HK’s respective photocycles. To resolve fluorescence changes in time, Rh-LOV-HK and Rh-LOV samples were irradiated with a camera flash to initiate adduct formation. In Figure 2.12 A and B, a loss of fluorescence is observed at 500 nm. The fluorescence emission peak centered around 345 nm is due to Ws in the sequence of Rh-LOV-HK (W = 4) and Rh-LOV (W = 1) (Figure 2.12 A and B). After light exposure, the fluorescence return was monitored at 500 nm for 2.5 hours (Figure 2.12 C and D). Like the time-resolved
absorption changes, both Rh-LOV-HK and Rh-LOV returned almost completely to the dark state in this time period.

Figure 2.12: (A) Rh-LOV-HK fluorescence before and after light activation. (B) Rh-LOV fluorescence before and after light activation. (C) Rh-LOV-HK time-resolved fluorescence return after photoexcitation. (D) Rh-LOV time-resolved fluorescence return after photoexcitation.
2.3.3 Circular dichroism (CD)

CD measures optical activity when a molecule has an absorption difference between left- and right-circularly polarized light. Formation or the loss of a chiral center can provide important information about chemical bond rearrangements. In protein studies, CD can be used to observe changes in protein secondary structure, ligand binding or protein-protein interactions (Greenfield 2007).

In Rh-LOV dark vs adduct CD difference spectrum, we observed an increase centered at 220 and 286 nm, and a decrease at 195 and 240 nm. The changes at 195 and 220 nm signify a loss in beta-strand secondary structure.

Figure 2.13: (A) Dark and excited state CD spectral comparison. (B) CD difference spectrum.
2.4 Discussion

LOV sensor domains are light sensing modules that can regulate a variety of effector domains in bacteria, including diguanylate cyclases, histidine kinases, helix-turn-helix DNA binding domains and STAS domains. The defining characteristic of a LOV domain is its photocycle with two predominant states. The dark state is defined by a noncovalently bound FMN with the ability to absorb light. The light activated or adduct state is defined by the formation of a covalent S-C bond with the C4a atom in the isoalloxazine ring. Although there is still some debate regarding the precise mechanistic details of cysteine-adduct formation, the general scheme is well defined (Conrad, Manahan, and Crane 2014). It is less well understood how different LOV domains exhibit unique temporal photocycle changes but the residues of the binding pocket are thought to be important since they potentially influence adduct stability and solvent exposure (Brosi et al. 2010; Conrad, Manahan, and Crane 2014). We set out to better understand Rh-LOV’s unique photocycle properties using optical spectroscopy.

Spectroscopy is defined as the study of the interaction of light and matter. This interaction results in several possible outcomes, including scattering, absorption, fluorescence and chemistry. In Rh-LOV, our primary goal was to observe the properties associated with the photocycle, which include light absorption, triplet-
state formation, photochemistry (adduct formation) and eventual return of the protein to its dark state. Among the many types of sensors that control STPs, light sensors have unique properties of absorption and fluorescence, which undergo dynamic changes during the course of the photocycle. These changes provide immediate insight into the physical changes that lead to activation of the C-terminal effector domain. Because a pulse of light induces instantaneous activation of the LOV signalling state, each step of the photocycle, including intermediate states, can be identified by time-resolved spectroscopy. This is an important contrast to ligand binding sensors where the activation is limited by diffusion processes. In the latter case, there is syncopation in the sensor activation, which will result in an ensemble effect. This complicates the ability to resolve intermediate states and determine the sequence of events in the protein’s mechanism. As a result, light sensing proteins lend themselves to mechanistic studies by optical spectroscopy.

Rh-LOV’s photocycle was initially reported in the PNAS publication, Light regulates attachment, exopolysaccharide production, and nodulation in Rhizobium leguminosarum through a LOV-histidine kinase photoreceptor, which reported that the photcycle only partially returned after 7 hours (Bonomi et al. 2012). We observed a near complete return in both Rh-LOV and full-length Rh-LOV-HK within ∼2.5 hours. Unfortunately, we do not have an explanation for this discrepancy.

From our results, we produced a detailed photocycle model for Rh-LOV (Fig-
This model includes the triplet excited intermediate state and the temporal component for each step. Even though the overall photocycle scheme is the same for all LOV photosensors, the rate constants of each step can vary. The duration of the adduct state is especially important since it determines the duration of the signalling phase.

The photocycle of Rh-LOV has a unique time-scale, which is intermediate between the fast cycling EL222 and the truncated cycle observed in Br-LOV. With $\tau_{1/2} \sim 22$ minutes and a near full return to the dark state in $\sim 2.5$ hours, the change in Rh-LOV’s kinase activity is expected to follow a similar time trajectory. The photocycle was not significantly altered in the Rh-LOV domain construct, giving rise to an interesting question: If not the downstream effector domain, what controls the unique cycling behaviour of different LOV domains? We will consider this question further in the next chapter by comparing the Rh-LOV crystal structure with those of EL222 and Br-LOV.
Figure 2.14: Rh-LOV-HK detailed photocycle model showing the dark, triplet-excited and adduct states with the duration of each step.

The unique time scale of Rh-LOV’s photocycle has potential as an optogenetic tool. Optogenetics is a budding field of study that uses light to control cellular
processes, most often in neural tissue (Mitra, Yang, and Moffat 2012). For designing optogenetic experiments, having control of the duration of the signalling response is an important consideration. Using different LOV domains can provide this flexibility. LOV domains have already proven to be effective in the emerging field of optogenetics (Mitra, Yang, and Moffat 2012). As the role of optogenetic applications continues to grow, LOV domains that have unique photocycle durations or different signalling mechanisms, may be used as part of a complete tool kit for the needs of different experiments.

The CD measurements of Rh-LOV in dark and photo-excited states reveal that light activation triggers conformational changes. The CD difference spectrum changes indicate that there is a loss of $\beta$-strand secondary structure on light activation. In the next two chapters, we will investigate the structure and observable dynamics using crystallography and EPR experiments.

References


Bonomi, H. R. et al. (2012). “Light regulates attachment, exopolysaccharide production, and nodulation in Rhizobium leguminosarum through a LOV-histidine


Circolone, Franco et al. (2012). “Structural basis for the slow dark recovery of a full-length LOV protein from Pseudomonas putida”. In: Journal of Molecular Biology 417.4, pp. 362–374.


Corchnoy, Stephanie B. et al. (2003). “Intramolecular proton transfers and structural changes during the photocycle of the LOV2 domain of phototropin 1”. In: Journal of Biological Chemistry 278.2, pp. 724–731.


Peternel, Spela and Radovan Komel (2010). “Isolation of biologically active nanomaterial (inclusion bodies) from bacterial cells.” In: Microbial cell factories 9, p. 66.

Peternel, Spela et al. (2008). “Engineering inclusion bodies for non denaturing extraction of functional proteins.” In: Microbial cell factories 7, p. 34.


Sato, Yoshiaki et al. (2005). “Reactive cysteine is protonated in the triplet excited state of the LOV2 domain in Adiantum phytochrome3”. In: *Journal of the American Chemical Society* 127.4, pp. 1088–1089.


Sørensen, Hans Peter and Kim Kusk Mortensen (2005). “Soluble expression of recombinant proteins in the cytoplasm of Escherichia coli”. In: *Microbial cell factories* 4.Figure 1, p. 1.


Young, J Peter W et al. (2006). “The genome of Rhizobium leguminosarum has recognizable core and accessory components.” In: *Genome biology* 7.4, R34.


Chapter 3

Structural characterization of Rh-LOV using X-ray crystallography

3.1 Introduction

In bacteria, light can be utilized as a source of energy, provide rhythmic stimulation to mediate circadian rhythms or act as a signal that regulates the activity of various signal transduction pathways (STPs) (Purcell and Crosson 2008; Gomelsky and Hoff 2011). STPs can lead to a multitude of cellular decisions.
However, identifying the function of individual STPs \textit{in vivo} is complicated by the fact that at any particular moment, the cell must integrate many forms of sensory input into a unified response (Circolone et al. 2012). Even with these limitations, a steady accumulation of evidence has established the functional role of several STPs. Notably, in $\alpha$-proteobacteria the LOV-HK two-component pathway has been identified as a regulator of the general stress response (GSR) pathway (Kaczmarczyk et al. 2014). Given the large variety of organisms that use LOV domains and the modularity of their associated effector domains, it is clear that LOV sensors are utilized for many specialized signalling purposes (Glantz et al. 2016).

Protein crystallography provides a powerful approach to understand LOV domains from a structural-functional perspective. In the year 2000, the first crystal structure of a LOV domain derived from \textit{Avena Sativa} (oat) was solved (Crosson and Moffat 2001). Since then, many homologous LOV domain structures from different bacteria have been deposited in the Research Collaboratory for Structural Bioinformatics Protein Data Bank (RCSB PDB or PDB). As expected, LOV domains share the characteristic fold of the PAS protein superfamily. Although this basic architecture is essentially unchanged, LOV domains do have unique differences from the PAS superfamily (Möglich, Ayers, and Moffat 2009). Most notably, they have the ability to noncovalently bind FMN. Atomic level resolution
of the LOV domain’s 3-dimensional structure has allowed us to identify the unique protein-chromophore interactions that give rise to this binding specificity.

Within the LOV family of photosensors, there are also distinct differences, making each LOV domain unique. For example, the N- and C-terminal regions often exhibit variability in length and orientation. Most LOV domains initiate a cellular response via a C-terminal effector domain, so the details of this region are important to the signalling mechanism (Möglich and Moffat 2007; Banerjee et al. 2016). However, genomic analysis has revealed that there are classes of LOV domains which lack an effector domain altogether and there are even cases where the effector domain has been found at the N-terminus (Endres et al. 2015; Glantz et al. 2016). There are also examples where the LOV and C-terminal HK effector domains are separated by an intervening PAS domain. In the most common C-terminal signalling LOV domains there appear to be differences as well. For instance, the $\alpha$ linker region can significantly vary in length (Möglich, Ayers, and Moffat 2009). In all of these examples, the connecting region and other interactions will be important for understanding different mechanisms of signal propagation in LOV domains.

The dimerization interface is another characteristic that varies between LOV domains. Most LOV domains crystallize as a dimer. There is supporting evidence that many LOV domains are stable dimers in solution as well (Endres et al. 2015;
Herrou and Crosson 2011; Banerjee et al. 2016). The dimer interface orients the effector domains and has important consequences regarding regulation and function of the effector domains. In some cases, light-dependent dimerization has been reported—in this scenario, dimerization leads to activation of the effector domain (Zoltowski, Motta-Mena, and Gardner 2013). Crystallography can be used to identify the dimer interface providing that the crystal dimer reflects the physiological orientation. Dimerization in the crystal phase can occur in response to crystal packing restraints, so it is important to verify that the dimer orientation in the crystal matches that of the solution phase (Bahadur et al. 2004).

Several LOV domain crystal structures have been solved in both dark and adduct states (Mitra, Yang, and Moffat 2012; Möglich and Moffat 2007; Sambasivarao 2013; Endres et al. 2015). Although global conformational changes were not apparent in these experiments, changes included formation of the S-C4a adduct that results in the sp² → sp³ puckering of the isoalloxazine ring. CD and NMR spectroscopy of LOV domains in the solution phase have demonstrated that there are more global conformational changes associated with light activated state (Harper et al. 2004; Harper, Christie, and Gardner 2004; Corchnoy et al. 2003). From crystal structure analysis, it has been proposed that H-bond disruption between the FMN’s isoalloxazine ring (C4 carbonyl) and a conserved glutamine residue in the Iβ-strand could lead to spatial rearrangements between LOV and the
effector. This rearrangement would allow the light signal to initiate the activity of the effector domain (Conrad, Manahan, and Crane 2014). However, the extent of these changes may vary and the details of signal propagation are still a topic of intense speculation. Crystal structures of LOV’s photoactivated state have not revealed significant conformational changes for reasons that will be discussed in section 3.3.4.

In this chapter, an outline of the procedures that were used to produce Rh-LOV protein crystals will be presented. A basic description of the X-ray diffraction experiment and an outline of the steps that were used to solve the structure of Rh-LOV will also be included. Then we will discuss the procedures that were used for in silico manipulation and refinement of the structure to produce the final model of Rh-LOV, which was uploaded to the PDB. Finally, presentation and analysis of the structure will cover protein interactions with the chromophore, interactions at the dimer interface and the role of the N-terminus. The photo-excited state diffraction experiments will also be discussed.

When appropriate, Br-LOV (3T50) (Rinaldi et al. 2012) and EL222 (3P7N) (Nash et al. 2012), two LOV domain structures that were previously solved and uploaded to the PDB, will be used for structural comparison. Although these LOV domains share many common features, they exhibit different dimer orientations and temporal photocycle dynamics relative to Rh-LOV and each other. These
differences will be taken into consideration when comparing Rh-LOV with these structures.

### 3.1.1 Requirements and procedures of crystallography

The first requirement of protein crystallography is that the protein of interest can be purified to a high level and remain stable way above physiological concentrations, typically ranging between 5-10 mg/ml. Once this requirement has been met, the protein must be empirically tested for favorable crystallization conditions. Crystals will form when conditions are ideal for the protein to precipitate out of solution in an ordered fashion such that a repeating crystal lattice forms. Because it is impossible to accurately predict the ideal crystallization conditions for a given protein, the most common strategy uses high-throughput technology to test many precipitants and conditions simultaneously.

In the process of producing high quality crystals, acquiring crystal hits from a high-throughput screen is only the starting point. Once a crystal hit has been identified, optimization around those conditions can result in better diffracting crystals that result in higher resolution data. This can be done by varying the pH and the concentration of precipitant. Scaling up the drop size can also help to form larger crystals that can produce better quality diffraction data. Optimization conditions are typically set by hand in 24-well hanging-drop trays (Figure 3.1).
Figure 3.1: Example of a 24-well optimization tray: Starting with the crystal hit conditions, the pH and precipitate concentrations are varied. More than one protein concentration can also be tested by setting multiple drops in each well.

Crystals are a repeating lattice that form from an ordered precipitation. The atomic structure of the molecule in the crystal lattice will scatter X-rays and result in a unique diffraction pattern. The scattered X-rays that result from constructive and destructive interference contain information about the intensities and position of each reflection in the lattice. The quality and extent of these data determines the resolution. However, the diffraction pattern is only half of what is necessary to solve a crystal structure. The second and more challenging part of determining
a novel crystal structure, is solving the so-called ‘phase problem’. To solve for the \(x, y\) and \(z\) position of each reflection in the unit cell, we need to solve the following electron density equation:

\[
\rho(xyz) = \frac{1}{V} \sum_{h} \sum_{k} \sum_{l} |F_{hkl}| \exp[2\pi i (hx + ky + lz) + i\phi_{hkl}] \quad (3.1)
\]

where \(F_{hkl}\), the structure factor, is proportional to the square root of the scattering intensities \((I_{hkl})\). The intensities and positions of the reflections are measured in the diffraction experiment but the phase, \(\phi_{hkl}\), must be solved in order to fit the data to its source in the atomic structure. Several strategies for solving the phase problem have been developed. The simplest method, referred to as molecular replacement, utilizes a homologous crystal structure that has previously been solved. When a homologous structure is available, the molecular replacement technique computationally matches the fit of the known structure onto the unsolved electron density, resulting in a solution to the \(\phi_{hkl}\) of the reflections. With a complete diffraction data set and a correct solution to the phase, the 3-dimensional structure of any molecule, even macromolecular complexes, can be solved. The following outline provides a description of the procedures and steps that were taken to crystallize and solve the structure of Rh-LOV.
3.2 Materials and methods for crystallography

3.2.1 Protein preparation

The purification methods outlined in Chapter 2, which included His-tag affinity purification and running a size-exclusion column, were carried out in the same fashion. Rh-LOV was then concentrated by centrifugation in a vivaspin-20 (Sartorius AG) with a 10 kDa pore size until the desired concentration (5-10 mg/ml) was achieved. The concentration was predicted by diluting the sample (1:19 protein to buffer ratio), followed by a measurement of the optical density at 450 nm using UV-vis spectrophotometry. The concentration was determined by Beer’s law (equation 3.2) using the established assumption that the extinction coefficient for the protein-bound FMN is approximately equal to that of free FMN. Trays were set with the concentrated protein shortly after preparation to prevent precipitation—a common side-effect of storing concentrated protein samples. An additional complication of working with LOV domains is the chromophore, which
can dissociate from the protein leading to inactivation.

\[
A = \varepsilon lC
\]

where:

- \( A \) is the optical density
- \( \varepsilon \) is the extinction coefficient of free FMN at 450 nm (12,500 L mol\(^{-1}\)cm\(^{-1}\))
- \( l \) is the cuvette thickness (1 cm)
- \( C \) is the concentration

\[ (3.2) \]

### 3.2.2 Crystallization conditions

Rh-LOV crystallized under a variety of conditions and precipitants (Table 3.1). In the case of Rh-LOV, a disordered precipitate formed immediately and then reorganized into an ordered crystal lattice over a period of two months at 4\(^\circ\)C. Initially, the JCSG (QIAGEN) and MCSG (Microlytic) crystallography suites were used to test for favorable crystallization conditions. Both suites consist of conditions for setting four 96-well plates that allow for the simultaneous sampling of hundreds of precipitants, buffers and pH combinations. These crystallography suites are based on conditions that have been empirically determined to produce favorable results for the crystallization of proteins.

Using a 3 \times 96 well tray, 1:1, 2:1, and 1:2 ratios of protein to precipitant conditions were set with a 200 nL total volume. Art Robbins’ Phoenix crystal tray setting robotic system was used to set the different crystallization conditions and
**MCSG screen: conditions of crystal hits**

<table>
<thead>
<tr>
<th>Screen</th>
<th>Well A2: 0.1 M Na$_2$HPO$_4$: Citric Acid (pH = 4.2)</th>
<th>2.0 M Ammonium Sulfate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Screen</td>
<td>Well E11: No buffer or salt, 25% (w/v) PEG 2000 MME</td>
<td></td>
</tr>
</tbody>
</table>

Table 3.1: Conditions from the MCSG screen that were favorable to Rh-LOV crystal formation.
protein in each well. These trays were stored in a 4 °C cold room and monitored weekly. A microscope and a plate reader were used to search for crystals. Crystals first appeared \(\sim2\) months after setting the trays (Figure 3.2). All crystal hits came from the MCSG suite (Table 3.1). In the first diffraction experiment, the Rh-LOV crystals were isolated directly from the MCSG plates. Higher resolution diffraction data sets were acquired from crystals (Figure 3.2) that were further optimized using the method described above (Figure 3.1).
Figure 3.2: Image of Rh-LOV crystals from an optimization tray. The expanded view gives a scaled perspective of crystal size.
3.2.3 X-ray diffraction experiment

Crystals were isolated under a microscope with a crystal loop (Figure 3.3 A) and transferred to the protein’s buffer conditions with 20% glycerol added for cryoprotectant. The loop was secured in a puck and stored in liquid N\textsubscript{2} until diffraction. All X-ray diffraction experiments were carried out remotely from the UCSC campus. Three separate beam lines—Stanford Synchrotron Radiation Lightsource (SSRL) at SLAC, Advanced Light Source (ALS) at LBNL and Advanced Photon Source (APS) at ANL were used to collect diffraction data.

Diffraction with a resolution of 2.84 Å was achieved in the first X-ray diffraction experiment. In a second diffraction experiment, we attempted to characterize the structure of the photoadduct state. Although adduct formation was not observed in the resulting electron density, we did get better diffraction data with a resolution limit of 1.89 Å. This experiment was repeated with more intense light irradiation before cryo-treatment but the adduct state was not observed—instead, two more high resolution (∼1.86 Å) dark state data sets were obtained (Figure 3.3 B).

After data were collected, the phase problem and refinement of the structure were addressed with computational programs designed for specific steps in the structure solving process. Most of these programs are available through the CCP4 (Collaborative Computational Project Number 4) suite—a collection of open source software tools for solving macromolecular X-ray structures. Exten-
sive tutorials are available on the CCP4 website (http://www.ccp4.ac.uk) and there are many other resources on the topic, such as CCP4 workshops, the CCP4 manual and additional informative publications (Winn 2003; Winn et al. 2011). These learning resources were supplemental to the invaluable guidance of Professor Seth Rubin and Dr. Sarvind Tripathi. The next section will cover a basic outline of the specific computational programs and how they were used in the structure solving process.
Figure 3.3: (A) A crystal loop was used to isolate a single crystal. The loop-crystal complex was then stored in a cryoprotectant, shipped to ALS and aligned remotely on the goniometer for X-ray diffraction. (B) A single diffraction image contains a regular array of spots, referred to as reflections. From the symmetry of the diffraction pattern, the space group can be determined.
3.2.4 Solving and refining the structure

After obtaining the initial diffraction pattern, MOSFLM was used to run an autoindexing algorithm. MOSFLM can predict the dimensions of the unit cell and the space group of the crystal from a single diffraction image. This is necessary for determining how many diffraction patterns (in 1° increments) are needed to generate a complete data set for producing a 3-dimensional model. We were able to predict that the geometry was hexagonal and the space group was P6$_3$ (Figure 3.4), which required collecting 120 diffraction patterns for a complete data set. After data collection, MOSFLM was used to integrate the diffraction data into a single electron density map (Leslie and Powell 2007).
Figure 3.4: (A) and (B) Rh-LOV crystallized in the $P6_3$ space group which is defined by a hexagonal repeating unit cell. The Rh-LOV dimer represents the basic repeating asymmetric unit. (C) This view shows the electron density fit over a portion of the structural model in the unit cell.

This data set was then processed in Phaser, a molecular replacement fitting program. Molecular replacement utilizes a solved structure that is predicted to be
structurally homologous to the molecule of interest. By placing the known structure into the unit cell, the phase and structure factors (F) can be calculated. Phaser then uses the phase and experimental structure factors to generate the electron density map (McCoy et al. 2007). We used the LOV domain structure from B. abortus (PDB ID: 3T50) to solve the phase of Rh-LOV. For some molecules, solving the phase requires complex strategies but for Rh-LOV, the molecular replacement approach provided us with an almost instantaneous solution to the phase problem.

Further manipulation of the structure was done using COOT, a flexible platform for building accurate structural models (Emsley et al. 2010). After opening the electron density map file and the structural model (Br-LOV) that was used to solve the phase problem, the sequence was mutated from Br-LOV to that of Rh-LOV using COOT. Detailed manipulations of fitting the protein side chains and backbone into the electron density were also carried out. COOT uses the electron density combined with restraints for bond angles and steric clashes to guide the manual manipulations of the structure.

The structural changes made in COOT were intermittently exposed to rounds of refinement using the Phenix refine program. Phenix, a stand-alone software program, refines the electron density map and produces the table of structural parameters (Table 3.2) that are used to gauge the quality of the structural model.
(Adams et al. 2010). In addition to refining the model, Phenix produces a new electron density map, which uses a color scheme to indicate regions where the model properly fits the density (blue), regions where the model is conflicting with the density (red) and regions of density that have not been fit by the model (green). The COOT-Phenix cycle was repeated several times to improve the model. The success with molecular replacement, combined with the high-resolution electron density map (Figure 3.5 A and B), simplified the manipulations that were required to generate the final structural model of Rh-LOV.
Figure 3.5: (A) Electron density of Rh-LOV (without structural model) shows high resolution details of FMN’s isalloxazine ring structure. (B) A sample of the structural model (side chains not shown) demonstrating its fit into the electron density.
3.3 Results

3.3.1 Presentation of Rh-LOV structure at 1.89 Å

The results in Table 3.2 describe the quality and completeness of the diffraction data acquired with a 1 Å X-ray beam. The crystals diffracted at a resolution of 1.89 Å. Rh-LOV crystallized as a dimer in the asymmetric unit. The space group was P6_3 and the unit cell had hexagonal geometry. The data set was nearly 100% complete based on the number of reflections that were accounted for in the structural model.

The R-work and R-free were 0.19 and 0.17, respectively; the reliability factor (R) is a measure of the agreement between the structural model and the experimental X-ray diffraction data. To calculate R, the structure factor (F_{hkl}) is required. As noted, F_{hkl} is proportional to the square root of the intensity (I) of the individual reflections and can be used in the following way to calculate R:

\[
R = \frac{\sum |F_{\text{obs}}| - |F_{\text{calc}}|}{\sum |F_{\text{obs}}|} \tag{3.3}
\]

where F_{obs} and F_{calc} are the structure factor magnitudes for the diffraction data and the model, respectively. R-work uses the data set that is exposed to rounds of refinement while the R-free factor uses the same formula on a subset of the data. About \sim 5-10% of the diffraction data is set aside and used to calculate
For this purpose. Since this test data set is not changed in the refinement of
the model, it provides an independent measure to gauge the quality of the model
during the refinement process.

The molecular visualization program, Chimera (UCSF), was used to visualize
the Rh-LOV structure with atomic level detail. Chimera was also used to con-
trast and compare Rh-LOV with other LOV domains available in the PDB data
base. Most often, the ribbon model setting was used to visualize these protein
structures. Although this setting produces a cartoon-like appearance, it is im-
portant to recognize that each model was solved by carefully fitting the primary
sequence of amino acid residues into its respective electron density map. There-
fore, these models are an accurate representation of the 1°, 2°, 3° crystal phase
protein structures.

We used the high-resolution (1.89 Å) diffraction data and molecular replace-
ment technique to produce the 3-dimensional model of Rh-LOV (Figure 3.6). Its
tertiary structure shares the classic PAS domain architecture. Conservation of the
PAS domain superfamily’s core structure is noteworthy since a protein’s structure
gives rise to its functional characteristics. Figure 3.6 shows the canonical five anti-
parallel β-strands and four interwoven helices that form the PAS core. These sec-
secondary structural elements are labeled alphabetically from the N- to C-terminus.
The β-sheet core structure has a torsion that twists it into concave shape, which
<table>
<thead>
<tr>
<th>Wavelength (Å)</th>
<th>1.000 Å</th>
</tr>
</thead>
<tbody>
<tr>
<td>Resolution range (Å)</td>
<td>62.87 - 1.85 (1.95 - 1.85)</td>
</tr>
<tr>
<td>Space group</td>
<td>P 63</td>
</tr>
<tr>
<td>Unit cell (Å)</td>
<td>125.74 125.74 53.13 90 90 120</td>
</tr>
<tr>
<td>Total reflections</td>
<td>304586 (42894)</td>
</tr>
<tr>
<td>Unique reflections</td>
<td>41153 (5958)</td>
</tr>
<tr>
<td>Mosaicity</td>
<td>0.52</td>
</tr>
<tr>
<td>Multiplicity</td>
<td>7.4 (7.2)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>100 (100)</td>
</tr>
<tr>
<td>Mean I/σ (I)</td>
<td>10.2 (2.1)</td>
</tr>
<tr>
<td>Wilson B-factor (Å²)</td>
<td>27.05</td>
</tr>
<tr>
<td>R-merge (%)</td>
<td>11.3 (100.6)</td>
</tr>
<tr>
<td>R-pim (%)</td>
<td>4.4 (39.7)</td>
</tr>
<tr>
<td>CC1/2</td>
<td>0.997 (0.535)</td>
</tr>
</tbody>
</table>

Reflections used for R-free
R-work (%) 0.1702 (0.2713)
R-free (%) 0.1873 (0.2860)

Number of non-hydrogen atoms 2116
Macromolecules 1959
Ligands 63
Water 94
Protein residues 251
RMS (bonds) (Å) 0.012
RMS (angles) (°) 1.39
Ramachandran favored (%) 99
Ramachandran allowed (%)
Ramachandran outliers (%) 0.81
Clash score 1.99
Average B-factor 40.90
Macromolecules 41.10
Ligands 30.40
Solvent 43.60

Table 3.2: This list, generated in Phenix refine, includes pertinent experimental details and R values, which were used to assess the reliability and quality of the structural model that was produced from the diffraction data.
is further enclosed by the PAS core. In LOV domains, this pocket has steric and electrostatic complementarity to the FMN chromophore. These interactions allow for the noncovalent attachment of the chromophore, which is the basis of light sensing. In addition, formation of the covalent adduct provides a mechanism for the conformational changes that propagate the signal to the effector domain.

Figure 3.7 demonstrates the high level of conservation between different LOV domain structures. The overall tertiary structure and many active site residues are conserved. Although the core tertiary structure is conserved, the N- and C-termini of different LOV domains differ in length and orientation (Möglich and Moffat 2007; Banerjee et al. 2016). These regions are of particular functional interest since they have been implicated in the mechanism of signal propagation from the LOV sensor to its output domain. In the Rh-LOV crystal structure, the termini were not resolved beyond the initial few N- and C-terminal residues. Rh-LOV contains an N-terminal cap that is 24 amino acids in length and consists of helical secondary structure. The N-terminal cap is not present in all LOV domains and its functional role is not well understood. Two possible roles for the N-terminal cap are that it is involved in the signalling mechanism and/or contributes to the dimerization interface (Watts et al. 2006).

The C-terminal region of Rh-LOV connects the sensor to its effector HK domain. There is typically a C-terminal connector region referred to as the Jα-helix.
However, in Rh-LOV this adjoining region is technically part of the HK, and contains the histidine phosphorylation site. This is a unique difference from other LOV domains that regulate the effector domain via changes in the Jα-helical region. The absence of this helical connector suggests that Rh-LOV signals by a unique mechanism. The conformational dynamics for this region have yet to be elucidated, so the details of its signalling mechanism are open for speculation. We will consider the active site and the N- and C-terminal regions in more detail in the following sections.
Figure 3.6: Presentation of Rh-LOV tertiary structure at 1.89 Å resolution (above), and the Rh-LOV primary sequence (below). The secondary structure is labeled alphabetically from the N- to C-terminus.
Figure 3.7: (A) Global structural comparison of Rh-LOV (green), Br-LOV (blue), EL222 (salmon) and Oat-LOV2 (purple) demonstrates the high degree of structural conservation. (B) Same view, emphasising conservation of protein interactions that bind FMN.
3.3.2 Protein-FMN interactions in the active pocket and connections to the C-terminus

The conserved residues of the active pocket orient and bind the FMN’s isoalloxazine ring. The active pocket consists of polar side chains that H-bond with FMN (Figure 3.8 A) and the ribityl side chain. There are also conserved nonpolar side chains in the make-up of the pocket’s topology that share hydrophobic complementarity with regions of the chromophore (Figure 3.8 B). Incorporation of FMN occurs during the translation process of a LOV protein and it remains noncovalently attached throughout the protein’s functional lifetime.

In Rh-LOV, the H-bond between the C4 carbonyl of isoalloxazine and the side chain of Gln 139 provides a direct connection between FMN and the C-terminus (Figure 3.8 C). In other LOV domain studies, it has been proposed that photoadduct formation weakens this H-bond, leading to a change in the effector domain (Freddolino, Gardner, and Schulten 2013). Figure 3.8 D provides a close up of the C-terminus. In the crystal structure, a glycerol molecule was built into the electron density between the adjacent C-termini. Although the C-termini do not interact directly in the crystal structure, these helices are the target of autophosphorylation and must be sensitive to conformational changes that act as a switch. Currently, there is insufficient biophysical data for us to model the
details of this activation process.

Figure 3.8: (A) Polar interactions with the FMN side chain. (B) Hydrophobic interactions with the isoalloxazine ring. (C) View of Gln 139 in the Iβ-strand leading to the C-terminus. (D) A view of residues in the C-termini (orange) interacting with a glycerol molecule (pink).

Within the subset family of LOV domains, LOV structures can be identified
by the conservation of active site residues that interact with FMN and the conserved cysteine. Mutational studies have revealed that even similar amino acid substitutions of the residues involved in the FMN-H-bond network can alter the sensitivity and photocycle duration (Zayner and Sosnick 2014; Raffelberg et al. 2011). Although mutational analysis has identified residues that alter the photocycle duration for a single LOV domain, it is not fully understood what differences allow for the dramatic variation in photocycle duration between different LOV domains (Conrad, Manahan, and Crane 2014).

In Figure 3.9, we consider the H-bond distances involved in the protein-FMN interactions in Rh-LOV, Br-LOV and EL222 structures. Like other LOV domains, Rh-LOV binds FMN in close proximity to a cysteine. Rh-LOV has electron density for more than one conformer of the cysteine—the distance between the sulfur and the C4a carbon atom ranged from 3.5-4.2 Å. The H-bonds between conserved side chains and FMN’s isoalloxazine ring in Rh-LOV were compared with Br-LOV and EL222. As can be seen in Figure 3.9, the interactions and distances that secure the chromophore are largely conserved (Figure 3.9). There are two important H-bond interactions with the ribityl side chain (not shown) and three additional H-bond interactions secure the hydrophilic corner of the isoalloxazine ring from below. The small differences in these H-bond networks does not explain the large differences in their respective photocycle durations.
Figure 3.9: H-bonds that stabilize FMN in the active pocket: (A) Rh-LOV. (B) Br-LOV. (C) EL222.
3.3.3 Rh-LOV and Br-LOV dimerization interface differences

Rh-LOV crystallized as a dimer. It is common for LOV domains to dimerize—this is one functional role that has been suggested for PAS domains in general (Huang, Edery, and Rosbash 1993). Because HK domains typically function as dimers, the orientation of Rh-LOV as a parallel dimer has physiological merit. However, as we have noted, dimerization in the crystal phase can be an artifact of crystal packing. In the next two figures, we will take a closer look at the dimerization interface and compare it with the Br-LOV, which was also observed to crystallize as a dimer.

In PyMOL, the APBS plugin applies the Poisson-Boltzmann distribution to the surface of the protein to determine its electrostatic contours. It is apparent from Figure 3.10 that the electrostatic differences are more pronounced at the dimer interface. Other LOV domains, such as Br-LOV, have notably different dimer orientations and interacting residues that form the interface. Br-LOV was observed to dimerize in an anti-parallel orientation. Since full-length Br-LOV has an intervening PAS domain between LOV and HK, the orientation of these domains could be radically different. In the next chapter, we explore whether these dimer orientations are maintained in the solution phase.
Figure 3.11 A and B displays the dimerization interface using Chimera’s intersurf feature, which was set to predict interactions between residues that are within 4 Å of each other. Although LOV domains with different sequences share many common structural characteristics, the dimer interface can vary dramatically (Rinaldi et al. 2012; Circolone et al. 2012). This interface has important consequences since it orients the effector domains and it can play a role in LOV’s mechanism of light regulated signalling. For instance, experimental results with EL222 have shown that it is a monomer under dark state conditions but after photoactivation, EL222 dimerizes on its target DNA, initiating gene regulation (Nash et al. 2012).
As can be observed from Figure 3.11 A and B, Br- and Rh-LOV domains exhibit radically different dimerization interfaces. The dimerization of Br-LOV involves an anti-parallel orientation and many stabilizing interactions take place between the β-sheet regions (Rinaldi et al. 2012). Rh-LOV dimerization relies predominantly on interactions between the extended N-terminal helices. In Rh-
LOV, the N- and C-terminal regions were not completely resolved in the crystal structure. Protein termini typically have more freedom, which limits the ability to resolve these regions in a crystal structure. Secondary structure predictions show that both termini have helical character. We were able to place the initial few residues of the N-terminus into the electron density. The N-terminal helices interact and appear to contribute to its stability as a dimer. Notably, there is an ionic salt bridge (K-E) and hydrophobic interactions that stabilize the dimer interface (Figure 3.11 C). In shorter constructs of Rh-LOV that were designed for NMR (Appendix A.1), we observed that total truncation of this N-terminal cap resulted in nonviable protein. This further implicates the N-terminal cap’s role in dimerization and indicates that it may also be necessary for protein stability.
Figure 3.11: (A) Rh-LOV dimerization interface (red: interaction interface). (B) Br-LOV dimerization interface (red: interaction interface). (C) Close up view of the N-terminal interactions that stabilize the Rh-LOV dimer.
3.3.4 Rh-LOV vs lit state LOV structures

Currently there are ∼15 LOV structures published in the PDB. Of these structures, 7 have been solved in both dark and photo-excited states. Phy3 LOV2 was the first LOV domain that was crystallized in both states (Crosson and Moffat 2002). The adduct state structure was achieved by exposing the crystals to intense light irradiation, followed by cryo treatment with liquid $N_2$ and the X-ray diffraction experiment. Confirmation of the photoexcited state was verified by observing electron density for the S-C4a covalent adduct. The LOV structures that have been successfully resolved in the adduct state undergo extremely limited conformational changes. The changes in protein structure may be muted as a result of crystal packing restraints (Freddolino, Gardner, and Schulten 2013).

From previous crystallographic studies of LOV structures, a general mechanism of signal propagation between LOV and its effector domain has been proposed (Mitra, Yang, and Moffat 2012). However, since global conformational changes appear stifled in the crystal phase adduct state, these insights are limited to speculation based on careful analysis of contacts that are present in both adduct and dark state crystal structures. Sequence and structural analysis of the C-terminal regions of different LOV domains indicates that there may be several different classes of LOV signalling mechanisms (Möglic and Moffat 2007).
Figure 3.12: (A) Structural comparison of Rh-LOV dark state (blue) with Ytva dark state (purple) and Ytva adduct state (brown). (B) Comparison of Rh-LOV dark state (blue) and Ytva adduct state (brown). (C) Close-up view of Rh-LOV dark state (blue) and Ytva adduct state (brown). (D) Same view but with emphasis only on the position of conserved residues that form the active pocket.

Comparing Rh-LOV (dark state structure) with both dark and photoadduct state
structures of Ytva-LOV from Bacillus subtilis shows the limited extent of conformational changes observed (Figure 3.12 A and B). This was also observed when comparing Rh-LOV with other adduct state structures. The photoexcited state crystal structure of Rh-LOV was pursued on two separate occasions but we were unable to observe electron density for the covalent S-C bond. We tested Rh-LOV crystals for photoactivity using a custom microspectrophotometry setup. We observed a cyclic photobleaching effect after light irradiation by following absorbance changes at 450 nm (Figure 3.13). Although the Rh-LOV crystals appeared to absorb light and recover with time, the change in light transmission was recorded with an oscilloscope that was limited to a 4000 s acquisition time. Therefore, we could not verify that these changes were due adduct formation or that they occurred on a time scale corresponding to Rh-LOV’s photocycle.

In the X-ray diffraction experiment, the crystals were irradiated with both continuous and flash illumination. However, the S-C4a adduct formation was not visible in the electron density, indicating that the crystals were either inactive or they had cycled back to the dark state before the diffraction experiment. Difficulties have been reported in other cases, indicating that not all LOV domain crystals are amenable to stable adduct formation (Rinaldi et al. 2012; Nash et al. 2012). Recently, the crystal structure of a short LOV protein from D. shibae was reported in dark and photo-excited states. Endres et al. reported that although
the covalent adduct was not formed, other minor changes suggested that small conformational changes did occur after light activation (Endres et al. 2015). In *D. shibae*, the adduct is short lived (τ_{recovery} = 9.6 s), adding to the challenge of isolating this state. Rh-LOV has a relatively long photocycle (τ_{recovery} \sim 7000 s) and because the crystals were stored in liquid N₂ after photo-illumination, we assumed that the adduct state would remain locked in the crystals. However, it is possible that the adduct state is less stable in the crystal phase and that Rh-LOV cycled back to the dark state during the transfer and setup of the remote diffraction experiments.
Figure 3.13: A decrease in relative voltage after light irradiation of Rh-LOV crystals, followed by a slow return, corresponds to LOV’s photocycle characteristics of adduct formation and breakage. However, our instrumentation limited our ability to follow these changes and determine if they were on the same time scale as Rh-LOV’s photocycle.
3.4 Discussion

The details of the Rh-LOV crystal structure were presented at 1.89 Å resolution. This high resolution structure shares the tertiary core that is common to the PAS superfamily as a whole. The most notable differences between PAS domains occur in the 1° sequence of amino acids, as well as variation of length and position of the N- and C-termini. These small differences can alter the dimerization tendencies and determine a PAS domain’s affinity and specificity for ligand or chromophore interactions that give rise to specialized functions (Möglich, Ayers, and Moffat 2009). For instance, the LOV domain subset of the PAS superfamily share unique residues in the active pocket that are responsible for the noncovalent attachment of FMN. This protein-chromophore relationship gives rise to LOV domains’ photosensitivity and ability to undergo photochemistry.

Although LOV domains share a common photochemical mechanism—light absorption leading to a triplet excited state, followed by formation of the S-C4a adduct, the duration of the adduct state can vary from seconds to days in different LOV domains (Circolone et al. 2012). In previous studies, kinetic analysis of single substitution mutations of conserved amino acids has revealed that certain FMN-protein interactions can increase or decrease the QE of adduct formation and the photocycle duration (Zayner and Sosnick 2014; Raffelberg et al. 2011).
In this way, it is possible to change the photocycle in a given LOV domain. However, this does not explain how different LOV domains that share these conserved features can have radically different photocycle durations.

In Chapter 2, we observed that the photocycle duration was similar for Rh-LOV and Rh-LOV-HK, indicating that the LOV domain is sufficient for determining the photocycle duration. However, we were unable to identify any obvious features in the Rh-LOV structure that could explain its unique photocycle duration. In Figure 3.9, we compared Rh-LOV’s H-bond network to FMN with those in Br-LOV and EL222—two LOV domains that have been crystallized and studied previously. This comparison revealed only small differences in the protein-FMN H-bond network. Since these observations were inconclusive, the dynamics of LOV signalling and LOV temporal regulation will require further investigation.

The terminal Jα linker region between LOV and its effector is not well resolved in any of these three structures, so other strategies will be needed to investigate these questions.

Differences between LOV domains at the C-terminus have significant implications for the mechanism of signal propagation. This region typically connects to a downstream effector domain in the same protein. Signal propagation to the effector domain may occur via different mechanisms based on the length and position of the Jα-helical region. As noted in Chapter 1, many LOV domains are
thought to regulate their downstream effector domains’ activity by undergoing a spatial rearrangement where the Jα-helix partially unfolds or detaches to facilitate activation of the effector domain. Because Rh-LOV lacks this intervening helical linker, its signalling mechanism provides an interesting twist to this model. The Rh-LOV core is directly followed by a helical region of the HK, which includes the active site histidine of the HK domain. We designed various constructs to explore this region but encountered protein inclusion body expression problems with constructs that were longer than Rh-LOV (160). In the crystal structure we were only able to include the first few residues of this extended helical region. An understanding of how Rh-LOV regulates its effector domain will require a more detailed picture of domain organization. Ideally, a high resolution structure of Rh-LOV-HK will be attained in the future. With knowledge of the ATP binding site and active site histidine relative to the LOV domain, we will be able to unravel the mechanism of signal propagation and the details of autophosphorylation.

Dimerization is important in the function of many proteins. This is true for most histidine kinases, which undergo autophosphorylation as a homodimer—the first step in the activation of the two-component STP (Rivera-Cancel et al. 2014). Since we encountered expression and stability issues with Rh-LOV-HK, we were unable to resolve the orientation of the full-length structure. From the Rh-LOV crystal structure, it is apparent that the structure dimerizes because of interactions
between the N-terminal helices. If this parallel orientation between monomers is conserved in the full-length structure, it would place the kinase domains adjacent to each other. From this orientation, it is not hard to imagine how small changes in the LOV domains might allow for the autophosphorylation event to occur. To go beyond speculation we will need high-resolution structures of the full-length protein.

The Rh-LOV dimerization interface was analyzed using Chimera (UCSF), a structure graphics viewing program. The dimerization interface in Br-LOV, which crystallized as a head-to-tail dimer, was also visualized for comparison (Rinaldi et al. 2012). Questions still remain regarding whether these dimer interfaces are conserved in the full-length structure. The Br-LOV dimer orientation, though radically different from that of Rh-LOV, may be due to the fact that full-length Br-LOV has an intervening PAS domain between LOV and HK domains. We do not know if this orientation is conserved in the full-length protein. Like Rh-LOV-HK, recombinant Br-LOV-PAS-HK protein expression results in inclusion body formation. Novel strategies will be needed to experimentally determine the orientation of these signalling modules.

Conformational changes were observed in CD measurements of Rh-LOV. To resolve these changes, Rh-LOV crystals were irradiated and stored in liquid N$_2$ until X-ray diffraction. Although these crystals diffracted and produced high-
resolution data, there were no observed changes in electron density that could be attributed to covalent adduct formation or conformational rearrangement. The adduct state structures of some LOV domains have been solved but not all. Even in the cases where both dark and adduct state LOV structures have been resolved, insignificant conformational changes have been observed (Zoltowski 2007). The most significant change in the adduct state crystal structures is the isoalloxazine ring puckering at the C4a carbon. However, H-bonds between the protein and FMN are relatively unchanged.

From analysis of interactions between FMN and Jα-helix a glutamine in the Iβ-strand that H-bonds with the C4 carbonyl of the isoalloxazine ring has been identified (Herrou and Crosson 2011). Disruption of this H-bond has been proposed to be a central step in the mechanism of signal propagation from LOV to the effector. However, as noted, this H-bond is not disrupted in the adduct state crystal structures. One explanation for this observation is that the naturally occurring conformational changes are restricted in the crystal due to packing constraints. This is supported by the observation that solution techniques such as CD and NMR suggest significant conformational dynamics (Harper et al. 2004; Harper, Christie, and Gardner 2004; Corchnoy et al. 2003; Nash et al. 2012).

Rh-LOV crystals were tested for photoactivity using a custom UV-vis microspectrophotometry setup (Figure 3.13). Apparent changes were observed but
we could not verify that this change was due to the LOV photocycle since the extent and duration of these changes were not resolved. Some groups have reported growing the crystals under constant light irradiation, allowing the crystals to form while in the adduct state (Circolone et al. 2012). This strategy would be worth pursuing in future attempts to crystallize Rh-LOV in the adduct state.

In the next chapter, we will use electron paramagnetic resonance (EPR) spectroscopy to explore the details of Rh-LOV. Using the technique, double electron-electron resonance (DEER), we will determine whether the dimerization interface in the solution phase is consistent with the crystal structure dimer organization. As well, we will use DEER as an alternative approach to explore Rh-LOV’s conformational dynamics.

References


Bahadur, Ranjit Prasad et al. (2004). “A Dissection of Specific and Non-specific Protein-Protein Interfaces”. In: Journal of Molecular Biology 336.4, pp. 943–955.

Circolone, Franco et al. (2012). “Structural basis for the slow dark recovery of a full-length LOV protein from Pseudomonas putida”. In: *Journal of Molecular Biology* 417.4, pp. 362–374.


Corchnoy, Stephanie B. et al. (2003). “Intramolecular proton transfers and structural changes during the photocycle of the LOV2 domain of phototropin 1”. In: *Journal of Biological Chemistry* 278.2, pp. 724–731.


Watts, Kylie J. et al. (2006). “Function of the N-terminal cap of the PAS domain in signaling by the aerotaxis receptor Aer”. In: Journal of Bacteriology 188.6, pp. 2154–2162.


Chapter 4

Electron paramagnetic resonance (EPR) experiments

4.1 Introduction

LOV domains are dynamic protein modules that sense light and convert this important environmental signal into a physiological response. LOV sensors belong to the PAS superfamily, which has a known characteristic structural fold. In the case of light sensing LOV domains, the conserved $\beta$-sheet and PAS core form a pocket with steric and electrostatic affinity for binding the FMN moiety. This chromophore interaction forms the basis of LOV domains’ light sensitivity and
ability to undergo light-dependent conformational changes in a cyclic fashion.

In bacteria, light-activation of a LOV domain can elicit a variety of cellular responses. The specificity of the cellular response depends on the context and the activity of the C-terminal effector domain. LOV domains typically regulate a downstream effector and its respective response pathway via conformational changes that occur after the protein and FMN form a covalent S-C(4a) adduct (Herrou and Crosson 2011). Although the details of adduct formation are conserved, there appears to be some flexibility in the mechanism of effector domain regulation. For instance, in the case of the aureochrome-LOV and the LOV-HTH (EL222) from *E. litoralis*, it has been proposed that light activation leads to dimerization (Herman and Kottke 2015; Nash et al. 2012). As a dimer, the HTH domains bind a target promoter DNA region and activate gene transcription. In other cases, it has been proposed that light-activation reorients the effector relative to the LOV domain through changes in the Jα-helix (Harper et al. 2004). However, the details and extent of these changes are not well understood and remain an area of active investigation (Herrou and Crosson 2011; Freddolino, Gardner, and Schulten 2013).

Extensive biophysical experiments have been carried out with varying degrees of success to resolve the LOV-effector conformational dynamics. In oat LOV2, NMR chemical shift differences between dark and adduct states of the protein
were resolved and mapped onto the static crystal structure (Harper, Christie, and Gardner 2004). Although precise distance changes could not be resolved in this way, the results indicate that signal propagation is directed by global conformational changes in the LOV structure. Adduct state crystal structures have also been solved for several LOV domains (Mitra, Yang, and Moffat 2012; Möglich and Moffat 2007; Sambasivarao 2013; Endres et al. 2015). Although significant physical changes were not observed in these adduct state crystal structures, a conserved glutamine in the Iβ-strand of the β-sheet has been identified and implicated in the mechanism of signal relay and regulation of the effector domain (Raffelberg et al. 2011). This glutamine’s H-bond with the C4 carbonyl of FMN directly connects the C-terminal Jα-helix to changes at the chromophore, making it a likely candidate for signal propagation. In the LOV adduct state crystal structures, minor changes occur in the H-bond orientation rather than total disruption.

In our research, EPR experiments were undertaken to better understand the nature of Rh-LOV’s solution structure. This approach was also used as an alternative method to investigate Rh-LOV’s light-induced conformational dynamics. Like the nuclear magnetic resonance (NMR) experiment, EPR detects changes in quantized spin populations. In a strong magnetic field (B₀), the electron’s spin state will orient with or against the B₀ field, with an energy of ±1/2gₑμᵦB₀. When the spin is aligned with the B₀ field, it is lower in energy (mₛ = −1/2) than
orienting against it \( (m_s = +1/2) \). This separation of energy states in the \( B_0 \) field is referred to as the Zeeman effect in quantum mechanics.

For NMR experiments, EM absorption by the paramagnetic nucleus occurs in the radio-wave region, while for EPR, the paramagnetic electron absorbs in the microwave region \( (\nu \sim 10^{10} \text{ Hz}) \). Both experiments are performed by placing the sample in a strong magnetic field and applying EM radiation at a frequency that is equivalent to the energy difference between the spin state orientations. Absorption of photon energy causes a spin state population inversion (Figure 4.1 A). As these spin states undergo relaxation, they precess (Larmor precession) around the magnetic field, producing a signal. The spin state energy difference and fundamental equation for EPR absorption is as follows:

\[
\Delta E = h\nu = g_e\mu_B B_0
\]

where \( g_e \) is the electron’s g-factor and \( \mu_B \) is the Bohr magneton.

Protein EPR methodologies can be used to determine structural and dynamical details about the system of interest (Altenbach et al. 1989; Borbat and Freed 2007). For these experiments, a paramagnetic center is commonly added by conjugating a stable radical label to specific amino acids in the protein. Most
often, this is done by mutating the selected position to a cysteine and labeling it with a nitroxide moiety such as (1-Oxyl-2,2,5,5-tetramethyl-\(\Delta^3\)-pyrroline-3-methyl) Methanethiosulfonate (MTSSL), which conjugates with the cysteine thiol via formation of a disulfide bond (Borbat and Freed 2007). Once conjugated to the protein, the MTSSL label is referred to as R1 (Figure 4.1 B). In protein studies, this approach is referred to as site-directed spin labeling (SDSL). For proper label conjugation, SDSL requires the selection of surface exposed positions to make cysteine substitutions into the amino acid sequence. In addition, any native cysteines should be removed by mutation if they are expected to be accessible during the labelling reaction (Klare 2012).
Figure 4.1: (A) The basis of the EPR experiment involves microwave absorption in a strong magnetic field. This results in a spin state population inversion followed by precession back to the ground state. By convention, the acquired absorption spectrum is converted into a plot of its first derivative. (B) This schematic illustrates the MTSSL conjugation reaction with a cysteine residue, which is often mutated into the target region of the protein sequence.

The characteristics of the local environment of the label will be reflected in the spectral width and regularity relative to the spectrum of free MTSSL (Hubbell et al. 1998; Columbus and Hubbell 2002). Spectral broadening occurs when the
label is inhibited by secondary structure or other interactions in the labeled region (Altenbach et al. 1989; Columbus and Hubbell 2002). Time-resolved CW EPR experiments can also be carried out to observe dynamic changes in the labeled region with microsecond temporal resolution. EPR measurements, combined with other biophysical data such as a static crystal structure, can contribute to a more dynamic model of a protein’s mechanism of action (Altenbach et al. 2001).

An extremely powerful application of EPR, double electron-electron resonance (DEER), involves the dipole-dipole coupling of two paramagnetic electrons. Through a series of pulses, a coupling echo can be established. By taking the Fourier transform of the modulated echo intensities as a function of pulse timing, the distance between the two labels can be determined (Jeschke 2012). The sensitivity of this approach depends on the inverse distance cubed between the paramagnetic centers, and distances from 15 to 80 Å can be resolved (G. Jeschke, M. Pannier and Spiess 2000). The precision of a DEER measurement depends on the distance between the labels and the mobility of the label. Due to the intrinsic flexibility of the MTSSL label, a distribution of distances is observed, even in the absence of protein backbone motions (Jeschke 2012).

DEER is notably important when other biophysical techniques, such as NMR and crystallography are not feasible. Unlike NMR (size limit ≈ 25 kDa), there are no size limitations for the sample in DEER experiments. The EPR experiment is
carried out under normal buffered conditions (with the addition of 25% glycerol) in which both the protein concentration and solution conditions more closely simulate the protein’s natural physiological environment relative to crystallography. Using the information from CW EPR, combined with distance restraints from multiple DEER measurements, a low resolution picture of a molecule’s structural topology can be generated (Bhatnagar, Freed, and Crane 2007).
Figure 4.2: (A) DEER experiments require the presence of two unpaired electrons to produce a dipolar coupling that has a range of ∼20-60 Å. (B) A pulse sequence is used to produce DEER modulation echo in time, which can be Fourier transformed into a function of frequency (Pake pattern) and normalized to produce a distance probability function.
This chapter will present the results of EPR experiments performed on spin labeled Rh-LOV constructs. The CW EPR spectra demonstrate that spectral variations reflect the secondary structure and local environment of the spin label. Using the DEER technique, we were able to show that Rh-LOV is a dimer in solution. Furthermore, we show that the dimer orientation is consistent with the crystal structure. This approach for determining solution state dimerization is also considered for Br-LOV and EL222. Lastly, DEER experiments were carried out to explore conformational changes associated with light activation. After adduct formation, some labeled positions showed an increase in the signal modulation depth. These observations will be discussed further in the Results section.

4.2 Materials and methods

4.2.1 Design of mutations for DEER

In Rh-LOV and Br-LOV, labeling positions were chosen by looking at the secondary structure and selecting mutations at ~10 residue intervals throughout the sequence without knowledge of whether the positions were surface exposed (Table 4.1). We hoped this strategy would allow for extensive characterization of the structure and global conformational dynamics. When a structural model is available, surface exposed labeling positions and distances can be predicted using
computer programs, such as the mtsslWizard plugin in PyMOL (Hageluken et al. 2012). Several single mutants were designed in EL225 (full-length EL-LOV-HTH), three by random selection and three using mtsslWizard. The three labeling positions that were chosen based on mtsslWizard readily conjugated with the label, while only one of three worked when the positions were chosen randomly (Table 4.2).

<table>
<thead>
<tr>
<th>Rhizobium leguminosarum</th>
</tr>
</thead>
<tbody>
<tr>
<td>S22C, A29C, V88C, R84C, S102C, S111C,</td>
</tr>
<tr>
<td>Erythrobacter litoralis</td>
</tr>
<tr>
<td>S43C, S140C, S213C, Q96C,</td>
</tr>
<tr>
<td>L109C, D130C, Y125</td>
</tr>
<tr>
<td>Brucella abortus</td>
</tr>
<tr>
<td>A52C, A90C, C69A, G76C, K55C,</td>
</tr>
<tr>
<td>R83C, S87C,</td>
</tr>
<tr>
<td>S105C, S140C, L142C, T151C,</td>
</tr>
<tr>
<td>L133</td>
</tr>
</tbody>
</table>

Table 4.1: Cysteine substitution mutations were engineered for EPR and DEER experiments. Only the mutants listed in green were successful in DEER experiments. The reasons for this are discussed in the text.

4.2.2 Labelling reaction

To incorporate cysteines into the sequence, primers were designed using the online program, PrimerX. The primer design consisted of two overlapping primers, with a single cysteine substitution flanked by 15 base pairs that are complementary
to the gene of interest. These primers were amplified in a PCR reaction using a plasmid template that contained the gene insert of interest. After amplification of the entire plasmid template, the PCR reaction was treated with the restriction enzyme, DPN1, which digests methylated DNA. Since *E. coli* methylates its DNA, the wild type plasmid was removed in this step. The remaining linear strands of amplified plasmid containing the mutation were transformed into DH5α *E. coli* cells. After transformation, *E. coli*’s ligase repairs these nicks, which regenerates the circular plasmid. The plasmid was then isolated, purified and transformed into BL21 cells.

After following the standard protein expression and purification process, the protein was incubated with label in a 5:1 molar ratio of MTSSL label to protein. Lyophilized aliquots of label were reconstituted in acetonitrile (0.5 mg MTSSL/30 µL acetonitrile). Incubation was carried out at room temperature for 30-150 minutes followed by desalting on a PD-10 desalting (3 kDa) column (GE Healthcare) to remove free label. After the labeling reaction with MTSSL, a small fraction of free FMN chromophore separated on the desalting column. This could have been caused by MTSSL out-competing FMN, resulting in R1 conjugation with the active site cysteine. After separation from the free label, the protein was concentrated to a final concentration of 200 µM, which included the addition of 25% glycerol. Because glycerol is a cryoprotectant, the protein could be stored at
-80°C under these conditions indefinitely.

Before attempting the DEER experiments, we needed to verify that the protein was properly labeled. Initially, we carried out an HPLC purification using a reverse phase column followed by electron spray ionization (ESI) mass spectrometry to determine whether the label was conjugated to the protein and present in a 1:1 ratio. Later, we followed a more time-efficient approach that involved CW EPR to test for bound label and predict the concentration of the spin label. With a measurement of both the spin label and protein concentration, we were able to predict the ratio of protein to label. The measurement of the spin label concentration allowed us to adjust the concentration of the sample for DEER experiments, which are ideally carried out at a concentration of ~200 μM spin label. Although the spin label to protein concentration could vary depending on the efficiency of the labeling process, this method produced reasonable results with most samples. After collecting the CW spectra of a protein sample, a custom Labview program, developed by Dr. Eric D. Walter, was used to subtract the background and integrate the spectra in order to calculate the concentration of the spin label.

4.2.3 Instrumentation

The room temperature CW EPR experiments were carried out on a Bruker EMX spectrometer at X-band frequency (9.8 GHz). There was 100 kHz field
modulation with typical modulation amplitudes of 1 Gauss.

The pulsed EPR experiments were carried out at 80 K on a Bruker E580 spectrometer at X-band with a second frequency DEER module and an MD5 dielectric resonator.

4.3 Results

4.3.1 Features of Rh-LOV surface secondary structure

observed by CW EPR

In an original study of T4 lysozyme, the crystal structure and select CW EPR measurements were used to show that the MTSSL spectral line widths of an EPR spectrum correspond with the localized protein secondary structure that restricts free motion of the spin label (Mchaourab et al. 1996). The extent of spectral broadening, relative to free MTSSL is indicative of label mobility. Furthermore, H. Mchaourab et al. were able to demonstrate that spectral changes can be observed in response to conformational dynamics (Mchaourab et al. 1996). In addition to using CW EPR for determining the spin label concentration for DEER experiments, we considered the qualitative differences between spectral broadening and label position on the Rh-LOV structure.

Figure 4.3 compares the CW EPR spectral envelope for several Rh-LOV sam-
amples with different positions of the R1 label. The CW EPR spectra for Rh-LOV-A29R1, V88R1 and S102R1 show significant broadening relative to free MTSSL. In the crystal structure, the label position of A29R1 is restricted by the N-terminal dimer interactions. Position S102R1 is similarly restricted by the C-terminal helix. The V88R1 label position appears to be restricted by steric and electrostatic repulsion from the ribityl side chain of FMN. In all three cases, the label position is in a secondary structure region (β-strand or α-helical). Position S84R1, on the other hand, is in an unstructured coil and its CW EPR spectrum reflects this. Positions T63R1 and S111R1 did not conjugate with substantial label, so the respective spectra predominantly represent free label. Using the mtsslWizard plugin, we did not find any conformers for these positions, indicating that they were not surface exposed. For other positions, mtsslWizard shows a predicted number of label conformers based on hard sphere analysis of the label’s environment in the static protein structure (Hagelueken et al. 2012).

Dynamic changes were not explored in Rh-LOV samples using time-resolved (tr)-CW EPR. We did test this approach with EL222-Y125R1 but time-dependent changes were not observed in the protein after Light. It is known that conformational dynamics occur in the LOV structure after adduct formation, yet the time scale and extent of these changes are not known. For tr-CW EPR spectral changes to occur in response to the photocycle, the label would need to be in a region of
changing secondary structure. This approach would be most effective when combined with other biophysical experiments, such as NMR, which could be used to select label positions in regions that undergo significant chemical shift changes.
Figure 4.3: The CW EPR spectra for different positions of the spin label (R1) in the protein correspond to the local environment of the label. The magnetic field (x-axis) for the different spectra are slightly offset for viewing. In each case, multiple conformations of the label are shown to illustrate its range of motion.
4.3.2 Specific DEER distance measurements confirm that the dimer interface is conserved in solution.

The dimer interface in a crystal structure can be an artifact of crystal packing, which will result in artificial interactions and oligomerization between protein monomers. Like many other LOV domain structures, Rh-LOV crystallized as a dimer (Circolone et al. 2012; Endres et al. 2015; Rinaldi et al. 2012). To determine if this was consistent with the dimer organization in the solution phase, select DEER measurements were compared with the respective distances between labeled positions in the crystal structure. Providing that the protein’s solution structure and dimerization interface are unchanged in the crystal structure, the predicted distances in the structural model should be consistent with the experimental DEER measurements (Hagelueken et al. 2012). To measure the corresponding distance between labels in the crystal structure, we utilized the mtsslWizard’s distance measurement prediction tool in PyMOL. MtsslWizard selects an average distance based on the position of predicted label conformers (Figure 4.3).

As can be observed in Figure 4.4, the experimental DEER coupling distances for the A29R1-A29R1, S84R1-S84R1 and S102R1-S102R1 positions in the solution dimer are consistent with the respective measurements in the crystal structure.
The two types of measurements give distances that are within experimental error of each other and indicate that the solution orientation is unchanged relative to the crystal dimer. We were not able to identify the source of the longer distance in the S102R1 spectrum. In Figure 4.7, we show that the shorter distance becomes more dominant after photo-illumination. In the next section we will show that even one DEER measurement may be sufficient for predicting dimerization and the solution dimer orientation.
Figure 4.4: DEER distance measurement and distance measurement prediction based on the crystal structure between positions (A) S84R1-S84R1, (B) V88R1-V88R1 and (C) S102R1-S102R1.
4.3.2.1 DEER measurements in Br-LOV and EL222 also validate this method for determining dimerization and dimer organization under buffered conditions.

To show that DEER spectroscopy is reliable for predicting dimer orientation, DEER measurements were made on Br-LOV-L133R1 and EL222-Y125R1 and the distances compared to those in their respective crystal structures. Br-LOV-L133R1 produced a well defined oscillation and the DEER distance between monomeric partners corresponds well with the predicted orientation and distance measurement based on the crystal structure. This single solution DEER measurement indicates that the anti-parallel dimer orientation observed in the crystal structure of Br-LOV is maintained under normal solution conditions. Many contacts were observed in the crystal structure, supporting the observations that it is a stable dimer. With the kinase present, the dimerization interface is expected to orient the monomers so that the cis or trans autophosphorylation mechanism can occur. Given these two possible mechanisms, there are many possible orientations that the full length protein could assume and it is feasible that the Br-LOV-PAS-HK solution dimer shares this anti-parallel orientation. It is also possible that the dimer interface in Br-LOV is unique to the LOV construct alone, since PAS domains and HKs are also known to contribute to dimerization (Huang, Edery, and Rosbash 1993; Casino, Miguel-Romero, and Marina 2014).
EL222 is not considered to be a dimer in solution until light activation (Zoltowski, Motta-Mena, and Gardner 2013). However, EL222-Y125R1 produced a DEER result in the dark state. We could not explain this result. Furthermore, other single mutants that we produced for DEER did not produce a dipolar coupling. We do not know if the EL222 crystal dimer represents the natural dimer configuration of EL222. Given that it is not considered to be stable dimer, it is not surprising that there is poor correspondence between the DEER distance distribution and the measured distance in the crystal structure (Figure 4.5 B).
Figure 4.5: (A) The Br-LOV DEER distance distribution is consistent with the distance between predicted label positions in the crystal structure. (B) In EL222, the DEER distance distribution lacks correspondence with the distance measurement in the crystal structure.
4.3.3 Rh-LOV-S22R1 provides information about interactions between the N-termini in solution.

The N- and C-termini tend to have more conformational flexibility than the rest of the protein and were not resolved in the electron density map for the Rh-LOV crystals. As a result, several residues were not included in the final crystal structure model. The S22R1 DEER distance measurement was used in combination with the predicted α-helical structure of the N-terminus to extend the structural model to include several additional residues of the N-terminus (Figure 4.6). The N-terminal helical cap is not present in all LOV domains and its function(s) is not fully understood. One possibility is that it facilitates protein-protein interactions. These interactions could be involved in the signalling state changes that alter the C-terminal effector domain or form a dimerization interface. For Rh-LOV, we designed shorter constructs that were partially or entirely missing the N-terminal helix. Complete removal of the cap resulted in non-viable protein expression. A closer look at the N-terminus in Chapter 3 revealed that it is important for dimerization in Rh-LOV.
Figure 4.6: (A) Rh-LOV with model of the N-terminal extension based on the experimental DEER distance measurement for Rh-LOV-S22R1. (B) Rh-LOV-S22R1 DEER distance measurement. (C) Background subtracted DEER time trace.
4.3.4 Rh-LOV light vs dark experiments reveal region specific conformational dynamics

After one minute of intense light irradiation of spin-labeled Rh-LOV samples, the protein was frozen in liquid nitrogen and DEER experiments were carried out in the same manner as the dark state samples. Although signal-to-noise averaging often required running the experiment for more than eight hours—a great deal longer than the Rh-LOV’s photocycle—the DEER experiment was carried out at 80 K, effectively locking the protein in the photoexcited state for the duration of the experiment. The temperature-dependent behavior can be predicted assuming Arrhenius-like kinetic dependence (Equation 4.2) with an activation energy \( E_A \) approximately equal to 55 kJ/mol (Corchnoy et al. 2003; Harper et al. 2004). Given this temperature dependence, the protein will remain in the adduct state indefinitely.

\[
k = A e^{-E_A/k_B T} \quad (4.2)
\]

The light and adduct state DEER experiments were compared but we did not observe distance changes in the adduct state measurements. However, there were changes in the DEER modulation depth. Notably, several mutation positions show an increased modulation depth in the adduct state (Figure 4.7 F and 4.8 B).
Because the modulation depth corresponds to the signal strength, this suggests that photo-excitation may increase dimer affinity. This observation is noteworthy, as it has been shown that LOV domains can undergo dimerization on light activation as part of the mechanism for regulating the activity of its downstream effector domain (Herman and Kottke 2015; Zoltowski, Motta-Mena, and Gardner 2013).
Figure 4.7: (A) Rh-LOV-S22R1 light vs dark state distance probability distributions and (B) light vs dark four-pulse DEER data. (C) Rh-LOV-V88R1 light vs dark state distance probability distributions and (D) light vs dark four-pulse DEER data. (E) Rh-LOV-S102R1 light vs dark state distance probability distributions and (F) light vs dark four-pulse DEER data.
4.3.4.1 Light activated changes in Br-LOV and EL222

Figure 4.8 shows the dark and adduct state DEER results for Br-LOV and EL222. Like Rh-LOV, a distance change was not observed in the adduct state measurements. However, the EL222 modulation depth increased significantly for the light-irradiated sample. In EL222, it has been proposed that dimerization occurs on its target DNA after light activation (Nash et al. 2012). However, we were able to observe a DEER signal with only one label position present in the monomer under dark state conditions (Figure 4.8 A and B). We followed up on this result by repeating the DEER experiment after light irradiation and the addition of EL222’s target DNA sequence, which was identified by Rivera-Cancel et al. in previous experiments (Rivera-cancel, Motta-mena, and Gardner 2012). There was an increase in modulation depth relative to both dark and adduct states. These results indicate that the monomer-dimer equilibrium is dynamic. However, in the presence of light and DNA, the dimer state of EL222 becomes more stable. Dimerization likely stabilizes the HTH interactions that form with the major groove of the target region in the double stranded DNA to recruit the polymerase.
Figure 4.8: (A) EL222-Y125R1 light vs dark state distance probability distributions and (B) light vs dark four-pulse DEER data. (C) Br-LOV-L133R1 light vs dark state distance probability distributions and (D) light vs dark four-pulse DEER data.
4.4 Discussion

Rh-LOV is the light sensor component of an important two-component signalling pathway. After light absorption and formation of the S-C4a adduct, the light signal propagates from the LOV domain to the HK domain and alters its activity. The HK undergoes autophosphorylation at a histidine residue, followed by phosphotransfer to a RR protein that acts on specific targets in the cell. For Rh-LOV-HK, it has been shown that the signalling pathway regulates the GSR (Bonomi et al. 2012).

EPR is a promising and versatile biophysical technique that can be used in protein studies, most commonly by the incorporation of spin labels. In this study of Rh-LOV, EPR methods were used to make several important observations including: 1) A distance measurement between the N-termini in a region that was not resolved in the crystal structure, 2) Identification of the Rh-LOV dimer orientation under solution phase conditions, and 3) Qualitative observations of light activated signal propagation dynamics.

Extensive mutants were produced in order to add spin labels to unique positions in the Rh-LOV structure. Without a structural model, the design of the mutants had two major setbacks—the selected positions were not always surface exposed, and we could not predict if the distance would be within the DEER
sensitive range. Because the crystal structure had not been solved at the time of the mutant design, there was no clear alternative to these shortcomings. If a crystal structure is available, the mtsslWizard tool, a labeling and distance prediction plugin that operates in PyMOL, can be used to consistently predict positions that are surface exposed and within the optimal distance for measurement. In this study, mtsslWizard was used to predict conformers of the label and measure average distances between spin label conformers.

CW EPR spectra of the labeled Rh-LOV samples allowed us to verify conjugation of the spin label and to predict the concentration of the spin label with Dr. Eric Walter’s custom Labview program. We also used CW EPR to show that there is good correspondence between the local region of the spin label and the predicted environment in the crystal structure. Spectral differences, particularly broadening, can be explained based on the local environment of the labeled region in the structural model.

Since a single label position in Rh-LOV and Br-LOV was sufficient for producing an echo oscillation resulting from dipolar coupling, we were able to verify that these domains are stable dimers in solution. In Rh-LOV, several DEER restraints were measured throughout the structure. The results of these measurements were compared with the distances in the crystal structure (Figure 4.4), which allowed us to conclude that Rh-LOV is a stable dimer in solution.
Only one DEER measurement was taken for Br-LOV. This measurement indicates that the dimerization interface is consistent with the crystal structure (Figure 4.5). Interestingly, the dimer orientation in Br-LOV is anti-parallel, rather than both monomers having the same orientation. We do not have any biophysical data to show that this orientation is conserved in the full-length protein. The dimer organization has the restriction that the histidine and bound ATP must be in proximity of each other in order to undergo autophosphorylation. It is not known if this event occurs in cis or trans configuration for Rh-LOV-HK.

DEER experiments were carried out at various positions throughout the Rh-LOV structure in both dark and photoadduct states. In several cases, there were changes in the modulation depth even though the measured distance did not change. There are several possible reasons why measurable conformational shifts did not occur. Conformational changes may simply be imperceptible or isolated to the C-terminus. It is unlikely that adduct formation results in a dramatic conformational change unless the Jα-helix detaches from the LOV core. Given the position of the histidine phosphorylation site in the extended helical region, it is unlikely that signal transduction requires large-scale changes. Unfortunately, we were unable to label positions in the C-terminal region of Rh-LOV or full-length Rh-LOV-HK. Therefore, more DEER experiments are needed to study potential conformational changes in the C-terminus and effector domain.
References


Circolone, Franco et al. (2012). “Structural basis for the slow dark recovery of a full-length LOV protein from Pseudomonas putida”. In: Journal of Molecular Biology 417.4, pp. 362–374.


Corchnoy, Stephanie B. et al. (2003). “Intramolecular proton transfers and structural changes during the photocycle of the LOV2 domain of phototropin 1”. In: Journal of Biological Chemistry 278.2, pp. 724–731.


Klare, Johann P (2012). “{Site-Directed} Spin Labeling and Electron Paramagnetic Resonance (EPR) Spectroscopy: A Versatile Tool to Study {Protein-Protein} Interactions”. In:


Rivera-cancel, Giomar, Laura B Motta-mena, and Kevin H Gardner (2012). “Identification of Natural and Artificial DNA Substrates for Light-Activated LOV HTH Transcription Factor EL222”. In:


Chapter 5

Conclusions

5.1 LOV domains: A growing body of knowledge

Since the discovery of LOV domains 20 years ago, we have learned a great deal about their structure and function. Intensive molecular and biophysical characterization has provided us with insights into the fundamental details regarding their photochemical mechanism and structure. However, there is still much progress to be made in understanding what controls the photocycle duration and the conformational changes that orchestrate changes in the activity of the effector domain. Although the mechanism of light activation and core tertiary structure are con-

183
served features, the photocycle duration and the mechanism of signal propagation can differ, making each LOV domain unique. In this thesis, the defining properties of the photocycle, structure and signalling dynamics were explored for a novel LOV domain from *R. Leguminosarum* (*Rle*). We hope that our results have made a contribution to the growing body of knowledge regarding LOV domains, which will lead to the next line of questions, future experiments and new discoveries.

Regarding the physiological role of LOV regulated signal transduction pathways (STPs), progress has been made in identifying both the components of the signal transduction pathway and the cellular response (Foreman, Fiebig, and Crosson 2012). In several LOV-HK containing α-proteobacteria, the HK domain has been shown to interact with the response regulator (RR), PhyR, which activates the general stress response (GSR) pathway (Kaczmarczyk et al. 2014; Sycz et al. 2015).

The *Rhizobium* family fix nitrogen for legume plants in a mutualistic relationship that is unique and important to planetary life. In the nitrogen cycle, these bacteria manage the energetically expensive process of converting atmospheric nitrogen into ammonia, which is used by legume plants for the production of nucleotides and amino acids. In the sustainable method of agricultural crop rotation, legume plants are grown as a cover crop to replenish the biologically usable nitrogen in the soil (Fageria, Baligar, and Bailey 2005). In lieu of this approach,
fertilizers are required to supply ammonia that has been produced by expensive processes which require the expenditure of fossil fuels.

In light of this important role for Rhizobia in nature, Bonomi et al. produced a genomic in-frame deletion of the LOV-HK gene product—one of Rle’s two light sensors. Through phenotype characterization of the ∆LOV-HK strain of Rle, they were able to directly link this two-component STP to the GSR (Bonomi et al. 2012). One significant effect of this knockout was a decrease in Rle’s ability to colonize its host. This surprising observation suggests that light is an important signal for establishing plant-bacterial symbiosis in certain Rhizobia. Although the precise RR has not been identified, Figure 5.1 shows a model of the Rh-LOV-HK signalling pathway. Light activation of the LOV domain leads autophosphorylation of the HK domain. This is followed by phosphotransfer to a RR (possibly PhyR, an anti-anti-σ factor). Phosphorylation of the RR increases its affinity for NepR, an anti-σ factor. Binding of NepR by PhyR allows σ^{ecf} to activate the transcription of various GSR associated genes.
Figure 5.1: Summary of the proposed mechanism for light regulation of the GSR pathway in *Rle*.
To investigate this important light-driven signalling pathway further, we pursued a complementary in vitro approach. Initially, our research goals were to study the structure and signalling dynamics that lead to regulation of the downstream HK domain. If we were successful with this first goal, a second objective was to identify the associated RR and study its interaction with the HK domain. Isolation of Rh-LOV-HK was problematic, so we cloned the LOV domain in isolation. It expressed solubly and we were able to purify sufficient amounts of protein for biophysical characterization. Time-resolved absorption and fluorescence spectroscopy were used to characterize the Rh-LOV’s and Rh-LOV-HK’s photocycle properties. Characterization of the Rh-LOV’s structure and dynamics included the use of paramagnetic techniques and crystallography.

Although we learned a great deal about Rh-LOV, there are limitations to all biophysical techniques. For example, with crystallization, conformational dynamics can be obstructed by crystal packing. Crystal packing can also lead to artificial dimerization contacts. With NMR, we can acquire structural details of a protein in solution but there are limits to the molecular weight of a molecule, which were exceeded by the Rh-LOV dimer (37 kDa). EPR and fluorescence techniques can be used as alternative methods for acquiring structural details and resolving conformational changes. Fluorescence labels are less rigid compared to the MTSSL nitroxide label commonly used for protein EPR studies. The size and rigidity of
MTSSL will result in a more accurate distance measurement but changes cannot be followed in real time. Given these experimental limitations, modern day characterization of a new protein or macromolecule requires a battery of approaches to produce the most complete model of the molecule’s structure and function. In the following section, we will briefly summarize the results from the experiments that were carried out to characterize Rh-LOV (Figure 5.2).

5.2 Summary of results

Rh-LOV’s unique photocycle duration was characterized, using optical spectroscopy. Both the full photocycle duration and the early triplet excited state intermediate were resolved—formation of the triplet excited state occurred in 6.5 $\mu$s and the half-life ($\tau_{1/2}$) was reached in $\sim$22 minutes. Complete return to the dark state occurred in $\sim$2.5 hours. The duration of the Rh-LOV’s photocycle is significantly slower than the fast cycling LOV domains from *E. litoralis* and *D. shibae*, which have photocycles on the order of seconds.

The crystal structure of Rh-LOV was solved at 1.89 Å resolution. This high resolution structure allowed us to produce a detailed model of the Rh-LOV structure. Although we were not able to explain the differences in photocycle duration by structural analysis, we were able to identify important contacts in the N-terminus that contribute to the formation of a stable dimer. Rh-LOV crystallized
as a parallel dimer. Because the histidine phosphorylation site is located in the helical region directly following the LOV core, this organization is suggestive of how kinase activation might occur. In many LOV domains, this linker region is thought to undergo dynamic changes. In Rh-LOV-HK, a dynamic change in the position of the histidine leads to autophosphorylation. However, we do not know whether autophosphorylation occurs in cis or trans configuration. We also do not know the spatial position of the ATP binding sites relative to the histidine. A high resolution, full-length LOV-HK structure is needed to help resolve these important details in the signalling mechanism.

Characterization of the Rh-LOV’s solution structure was carried out via EPR techniques. Pulsed EPR (DEER) was used to measure the dipolar coupling signal and generate a distance probability. We were able to determine that the crystal dimer and solution dimer have the same orientation. Dark and light state DEER distance measurements were pursued as an alternative approach for identifying Rh-LOV’s structural signalling dynamics. Although distance changes were not observed, several of the photoadduct state DEER measurements showed an increase in signal modulation, which suggests an increase in dimerization. Labeling experiments were also attempted with the full-length Rh-LOV-HK but were unsuccessful due to a lack of protein stability.
Figure 5.2: Overview of Rh-LOV experimental results: (A) The photocycle dynamics were temporally resolved. (B) The crystal structure was solved at 1.89 Å resolution. (C) DEER was used to calculate select distances in the solution structure.
In many cases, such as this study, experiments to elucidate the signalling mechanism have been limited to the LOV domain in isolation from its downstream effector. Even notable exceptions, such as EL222 (LOV-HTH), EL346 (LOV-HK) and YTVA (LOV-STAS), have only given us limited insights. For instance, in the EL222 crystal structure, the Jα-helix is not resolved and in EL346, the LOV and HK domains were crystallized separately. After light absorption and adduct formation, the details of signal propagation have yet to be solved with the clarity needed to understand the mechanism. Based on structural organization of different LOV-effectors, it is apparent that there is more than one mechanism for regulating the activity of the effector. Biophysical techniques such as CD and NMR, which observe the protein in a solution environment, have been used to demonstrate the presence of these conformational changes (Harper et al. 2004; Harper, Christie, and Gardner 2004; Corchnoy et al. 2003). However, these changes have proven hard to capture using crystallography.

5.3 Bridging in vivo and in vitro experiments

The shortcomings that currently limit our understanding of the bacterial STPs are often related to protein stability issues, which are common with recombinant protein expression in E. coli. Stability is a requirement for structural-functional protein characterization and high protein concentration is often required for bio-
physical characterization, which can further destabilize protein folding and affect function.

Protein expression issues were the major limitation of our initial research efforts. Although various strategies were used, including insect cell expression, solubilizing fusion tags and protein refolding—we were unable to isolate significant quantities of stable, full-length protein. Using the sumo fusion tag expression system, we were able to purify enough full-length protein for biophysical experiments. However, protein was lost during expression (inclusion body sequestration), purification (irreversible binding to the column) and during the TEV cleavage process (inefficient cleavage). Although the isolated protein had photocycle activity, it would lose the chromophore and become less viable over time. We were also unable to further purify the protein on a size-exclusion column via FPLC. These results indicate that many of the problems encountered were due to oligomerization of the protein.

Why is this protein not amenable to expression, purification and remaining stable in solution? One possibility is that Rh-LOV-HK requires other stabilizing factors, such as other proteins. Because Rh-LOV-HK interacts with a RR, it is possible that coexpression would stabilize Rh-LOV-HK expression; however, this has not been tested. One potential solution to this problem would be to overexpress Rh-LOV-HK in Rle with a purification tag. Broad range host vectors can
be used for this purpose. If Rh-LOV-HK depends on a stabilization factor such as its cognate response regulator or a DUF domain, this additional protein could be isolated while simultaneously solving the LOV-HK stability issues. Furthermore, Rh-LOV-HK overexpression in *Rle* could be used to accentuate the phenotype associated with this pathway.

### 5.4 A full-length model of Rh-LOV-HK

Rh-LOV-HK and other full-length HWE HK dimers remain an unsolved structural problem. If Rh-LOV-HK could be crystallized and characterized, what would its structural organization be like and could we uncover the details of signal propagation from the LOV domain to the kinase? Without any biophysical data to answer these questions, we can only speculate based on comparative analysis and modeling. Interactions that connect changes at the chromophore to the C-terminus are largely conserved and are currently the basis for predictions of how light-activation propagates from the LOV domain to the kinase (Crosson, Rajagopal, and Moffat 2003). However, it seems apparent that a full-model is needed to explore and understand how homologous LOV domains can have unique photocycle durations and different effector activation mechanisms.

As more structural data becomes available in the PDB, we can make a comparison to our system and produce more accurate models of the full Rh-LOV-HK
structure. Currently there are only two HWE HKs in the PDB (EL346 from *E. litoralis* and the HK from Br-PAS-HK), both of which were published in the last two years (Rivera-Cancel et al. 2014; Rinaldi et al. 2016). Using the modeling program, Modprod, we produced a model from Rh-LOV and El346 (LOV-HWE-HK). Although the secondary structure of the two proteins is conserved, EL346 is thought to function as a monomer, which is atypical for kinases (Rivera-Cancel et al. 2014).
Figure 5.3: Amino acid sequence comparison between Rh-LOV-HK and El-LOV-HK346 reveals conservation of secondary structure throughout both LOV and HK domains.
Figure 5.4: Full-length Rh-LOV-HK models generated in Modprod using Rh-LOV and the HK from EL346: (A) Model A shows the linker region stretched out to spatially separate the LOV and HK domains. (B) Model B presents interactions between LOV and HK domains. (C) Model C shows another possible configuration through which the domain surfaces might interact.

Given that these models are based on structural data, they likely provide an accurate representation of the two domains. However, their organization with
respect to each other remains unknown. Domain organization is critical to un-
derstanding signal propagation from the sensor to the enzyme. Rh-LOV-HK is
thought to be a stable dimer—prediction of a plausible dimerization interface
could lead to clues of the overall structural organization. The prediction of dimer
organization can be guided by known restraints, such as the ATP binding sites,
which need to be proximal to the target histidine residue in order to facilitate
autophosphorylation.

5.5 New frontiers: The future of LOV and STP pathways

We still have a great deal to learn about bacterial LOV domains and their asso-
ciated STPs. Structural details and photocycle characteristics have been worked
out for many LOV domains. However, the more subtle details have proven chal-
lenging to modern day approaches. We still do not know whether the adduct forms
via a thiolate or radical mechanism. We also do not know the details regarding
the precise conformational changes that activate the effector signalling pathway
or what controls the duration of the photocycle.

In the last few years, the focus of LOV domain publications has shifted towards
exploring their potential as an optogenetic tool (Herrou and Crosson 2011). In this
context, the LOV domain is fused to a protein of interest. The target protein and its cellular response can then be controlled by light illumination. Although LOV’s regulatory mechanism may not work with all proteins, it has been successfully demonstrated in several instances (Motta-Mena et al. 2014). Although the field of optogenetics has predominantly focused on the study of neural networks, it has many other potential applications. For instance, this approach could be used to study STPs in bacteria. For the majority of sensor domains, the sensing specificity is not known. By systematically producing LOV-effector fusions, these essential decision-making pathways could potentially be elucidated \textit{in vivo}.

### 5.6 Closing thoughts

Graduate school should be a time of exploration and expansive learning—by reaching beyond a single research endeavor, new ideas and possibilities may present themselves. Even through creative endeavors that are unrelated to science, inspiration and insight can cross barriers and spill into the scientific realm. Because graduate research typically does not define one’s future work, there should be a balance between academic endeavors and personal development. When I started graduate school, my P.I. told me that graduate school is a time to dream big. I took this advice to heart and have tried to approach my education and research with this mind-set. Although many of my projects were unsuccessful, the experi-
ential gain far outweighed any experimental failures.

In reflection, my PhD experience was filled with learning, growth and self-discovery. The independence that I was given for research and experimentation may have limited my ability to make scientific progress but it contributed greatly to my learning process. The guiding aims of this thesis research were often derailed by time and experimental limitations. In response, new ideas emerged and new goals were established. Sometimes this led to dissatisfaction but the primary objective—to challenge and train a PhD candidate—was accomplished through the navigation of challenges encountered in each step of the research process. This allowed me to develop powerful, efficient problem-solving skills and gain experience with many molecular and biophysical techniques. A vital part of my research experience was to guide undergraduates in their research endeavors. I was fortunate to work with many talented individuals and, although I cannot take credit for their enthusiasm and interest in science, I feel fortunate to have fueled it in some way. Given the well-rounded training and experiences that I have gained through this PhD process, I am stepping forward on a foundation of priceless experience.
References


Corchnoy, Stephanie B. et al. (2003). “Intramolecular proton transfers and structural changes during the photocycle of the LOV2 domain of phototropin 1”. In: Journal of Biological Chemistry 278.2, pp. 724–731.


Appendix A

Ancillary materials

A.1 Nuclear magnetic resonance (NMR)

Nuclear magnetic experiments can be used to determine many important properties of macromolecules, including their 3-dimensional structure. Because it is done in a buffered solution, NMR-derived structures are considered to be similar to those present under biological conditions. Although obtaining a structural solution based on NMR is more laborious than that based on X-ray crystallography, in some cases it may be the only amenable method to obtain a structural model for the protein of interest. NMR is a versatile alternative to crystallography but it is limited to molecular sizes of less than 25 kDa. In the experiment outlined
here, molecular size proved to be an obstacle. NMR is also important for obtaining information about protein conformational changes, including those in Br-LOV and oat LOV2 (Rinaldi et al. 2012; Harper, Christie, and Gardner 2004). In both cases, the changes in structure associated with photoexcitation were mapped onto the crystal structure model. This approach has provided insights into the subtle structural changes that allow for signal propagation from the LOV domain to the effector domain.

A.1.1 Materials and methods for NMR

To express protein with isotopic $^{15}$N, DE3 expression cells were transformed with Rh-LOV(142) and grown to an optimal cell density (OD $\cong$ 0.6-0.8) in four flasks with 250 mL of LB/Amp in each. After centrifugation, the cell pellets were transferred to M9 minimal media, which contained $^{15}$N. The cells incubated for 1 hour at room temperature and 200 rpm shaking, before initiating protein expression with 1 mM IPTG. Expression was carried out for 5 hours at room temperature. Basic purification protocols, outlined in chapter 2, were then used to isolate the protein. The sample was concentrated to 3 mM and a 300 $\mu$L sample was loaded into a Shegemi NMR tube. HSQC experiments were carried out with a 600 MHz Varian Inova NMR Spectrometer.

As the HSQC spectra show (Figure A.1), the resolution is not sufficient to
resolve each residue in the structure. We attempted to express shorter constructs of the LOV core to increase our resolution for Rh-LOV(160), which weighs 37 kDa as a dimer (including His tag and TEV cleavage site). The shortest structure that we were able to express, Rh-LOV (142), contained a portion of the N-terminal cap. This construct had a molecular weight of 34 kDa as a dimer and exhibited only slightly better resolution.

Several shorter constructs were cloned and expressed but functional, FMN-bound LOV protein was not isolated, most likely because the removal of the N-terminal cap disrupted dimerization. In turn, this may also have disrupted folding of the tertiary core structure.

A.1.2 NMR results

NMR experiments were pursued as a way to map conformational changes by assigning chemical shifts to specific residues in the HSQC NMR spectrum. We hoped to use this method as a complementary approach to DEER and crystallography but we were unable to produce a construct that was within the molecular weight cutoff (25 kDa) for NMR experiments.
Figure A.1: HSQC spectra for (A) Rh-LOV (160) dark state, (B) adduct state, and (C) Rh-LOV (160) vs (D) Rh-LOV (142).